Program for the Stern/Salomon Microstructure Meeting, Friday, May 24, 2024

Last updated: Wednesday May 22, 2024 9:34 AM

Program Committee	Tarun Chordia, Goizueta School, Emory University Joel Hasbrouck, Stern School, NYU Paolo Pasquariello, Ross School, University of Michigan Gideon Saar, Johnson School, Cornell University Clara Vega, Federal Reserve Board Liyan Yang, Rotman School, University of Toronto

Further details on the conference and registration will be posted on this page.

The Stern Microstructure Conference is open to everyone with an interest in market microstructure research. The conference will be held at the Kaufman Management Education Center (KMEC), 44 W. 4th St., NYC (near the southeast corner of Washington Square Park). For more complete directions see https://www.stern.nyu.edu/experience-stern/contact-us/visit-stern. There will be a registration desk in the lobby. Sessions will be held in KMEC 1-70.

Registration Link The registration link is no longer active. Email Joel Hasbrouck at jh4@stern.nyu.edu.

Please note: Hard copies of the papers will not be available at the conference. Links are available below. The schedule below is tentative and subject to revision. (Please don't make travel plans contingent on any particular ordering of the papers.)

For each paper: 30 min for the presenter; 20 min for the discussant; and 10 min general discussion.

Thursday, May 23	
6:00 pm	Conference dinner. Place: KMEC 5-50. (44 West 4th St. You can enter the building from the smaller plaza on the west side. Take the elevator to the 5th floor.) Open to program participants and a limited number of registered attendees. First-come first-served, preference for program participants from previous years. Hors d'oeuvres and drinks open at 6:00; dinner at 6:30.
Friday, May 24	
8:30 - 8:55	Continental Breakfast, Room KMEC 1-100.
8:55 - 9:00	Welcome. We usually pass the microphone around the room for introductions (like, "Joel Hasbrouck from Stern Finance.")
9:00 - 10:00	Measuring Interest Rate Risk Management by Financial Institutions Celso Brunetti (Federal Reserve Board) Nathan Foley-Fisher (Federal Reserve Board) Stephane Verani (Federal Reserve Board)

	Discussant: Matthew Pritsker (Federal Reserve Bank of Boston) Chair: Paolo Pasquariello, Ross School, University of Michigan
10:00 - 11:00	Fragmentation and optimal liquidity supply on decentralized exchanges Alfred Lehar (University of Calgary); Christine Parlour (Haas School of Business at UC Berkeley); Marius Zoican (University of Toronto) Discussant: Agostino Capponi (Columbia) Chair: Joel Hasbrouck, Stern School, NYU
11:00 - 11:15	Break
11:15 - 12:15	Liquidity Provision in a One-Sided Market: The Role of Dealer-Hedge Fund Relations Mathias Kruttli (Kelley School of Business, Indiana University); Marco Macchiavelli (Isenberg School of Management, UMass Amherst); Phillip Monin (Federal Reserve Board); Alex Zhou (Cox School of Business, Southern Methodist University) Discussant: Or Shachar (Federal Reserve Bank of New York) Chair: Tarun Chordia, Goizueta School, Emory University
12:15-1:15	Lunch, Room KMEC 1-100
1:15-2:15	Trading, Ambiguity and Information in the Options Market Azi Ben-Rephael (Rutgers University) Tony Cookson (University of Colorado at Boulder) Yehuda Izhakian (Zicklin School of Business, Baruch College) Discussant: Indira Puri (Stern School, New York University) Chair: Clara Vega, Federal Reserve Board
2:15-3:15	Detecting Informed Trading Risk from Undercutting Activity in Limit Order Markets Barardehi Yashar (Chapman University) Peter Dixon (SEC) Liu Qiyu (SEC) Discussant: Andriy Shkilko (Wilfrid Laurier University) Chair: Liyan Yang, Rotman School, University of Toronto
3:15-3:30	Break
3:30-4:30	AI-Powered Trading, Algorithmic Collusion, and Price Efficiency Winston Dou (Wharton) Itay Goldstein (Wharton) Yan Ji (HKUST) Discussant: Elena Asparouhova (Eccles School, University of Utah) Chair: Joel Hasbrouck, Stern School, NYU
4:30	Adjourn

Measuring Interest Rate Risk Management by Financial Institutions*

Celso Brunetti[1], Nathan Foley-Fisher[1], and Stéphane Verani[1]

[1]Federal Reserve Board

First version: June 2022; this version: October 2023

Abstract

Financial intermediaries manage myriad interest rate risk exposures. We propose a new method to measure financial intermediaries' residual interest rate risk using high-frequency financial market data. Our method exploits all available high-frequency information and is valid under extremely weak assumptions. Applying the method to U.S. life insurers, we find their interest rate risk management strategies are generally effective. However, life insurers are more sensitive to changes in long-term interest rates than property and casualty insurers. We show that the term premium helps to explain the difference in sensitivities between the two types of insurer.

JEL Codes: G20; C58

Keywords: financial institutions; interest rate risk management; high-frequency financial econometrics; subsampling; life insurers.

*For providing valuable comments, we would like to thank, without implication, Mark Carey, Burcu Duygan-Bump, Peter Hansen, Max Huber, Anastasia Kartasheva, Borghan Narajabad, Ali Ozdagli, Andrew Patton, Matt Pritsker, Roberto Renò, Rich Rosen, Oleg Sokolinskiy, Pavel Szerszen and participants in the Interest Rate Variability and the Financial Sector Conference, Society for Economic Measurement Annual Conference 2023, the International Risk Management Conference 2023, and seminars at the European Central Bank, St. Gallen University, and the Federal Reserve Board. We are grateful to Julia Silbert and Renee Garrow for exceptional research assistance. The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The views in this paper are solely the authors' and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System.

1 Introduction

Financial intermediaries are exposed to interest rate risk. They have multiple sources of exposure arising from cash flow differences across balance sheet components as well as contractual or embedded options with asymmetric payoff characteristics. Although intermediaries have a wide range of asset and liability management tools available to hedge interest rate risk, they do not fully insulate themselves from all potential changes in interest rates for several reasons.[1] Financial markets may be incomplete, fully hedging may be prohibited by its cost, and carrying interest rate risk may be a source of earnings.[2] Thus, financial intermediaries carry some *residual* exposure to interest rate risk, which could have significant consequences for financial stability and macroeconomic outcomes in bad states of the world (Holmstrom and Tirole, 1997; Brunnermeier and Sannikov, 2014).

In this paper, we propose a new method to measure the time-varying residual interest rate risk exposure of financial intermediaries using minute-

[1] Risk managers at financial institutions are expected to monitor and manage interest rate exposures at prudent levels, but not fully eliminate the risk. Supervisors provide detailed guidance on management practices and coordinate their standards. See, for example, the Office of the Comptroller of the Currency (OCC) Revised Handbook March 2020, the Federal Deposit Insurance Corporation (FDIC) Letter on Financial Institution Management of Interest Rate Risk 2010, the Federal Reserve Board (FRB) Supervisory Manual on Interest Rate Risk, the National Association of Insurance Commissioners (NAIC) Risk-Based Capital for Insurers Model Act, the OCC-FDIC-FRB Joint Policy Statement on Interest Rate Risk 1996, and the Basel Committee on Banking Supervision Guidance on Standards 2014.

[2] Even an established hedging strategy may be exposed to "basis risk"—that is, it might lose its effectiveness.

by-minute financial market data. We calculate the daily realized covariance of high-frequency stock returns for those intermediaries and Treasury security returns. We construct a conditional covariance by projecting out aggregate stock market returns from stock returns and Treasury security returns. We then introduce *realized gamma* as the ratio of the conditional covariance to the daily realized conditional variance of Treasury security returns. *Realized gamma* is a daily estimate of the sensitivity of an individual firm's stock price returns to realized changes in interest rates. We calculate returns at five-minute intervals using every possible five-minute grid point in a trading day, exploiting all available high-frequency information as described in Zhang, Mykland and Aït-Sahalia (2005).

High-frequency data provide a consistent estimate of time-varying interest rate risk, even when changes in financial institutions' exposure are slow moving. We carefully address well-known market microstructure concerns associated with high-frequency financial market data in Section 2.1.1. A feature of our realized volatility estimates is the ability to aggregate them over time (Corsi, 2009). In practice, we can consistently estimate a measure of longer-term interest rate risk over any horizon by averaging our daily estimates. For example, in our empirical application described in Section 3, we averaged the daily estimates over a period of two months.

We also propose a new statistical test of the daily residual interest rate risk exposure of financial intermediaries. We conduct statistical

inference on the *realized gamma* estimates by calculating asymptotically valid confidence intervals using subsampling (Politis, Romano and Wolf, 1999). The essence of the subsampling method is to approximate the sampling distribution of the daily *realized gamma* with the empirical distribution generated by estimating the *realized gamma* on an exhaustive set of intra-day subsamples.[3] Although computationally intensive, the method of subsampling behaves well under extremely weak, easily satisfied assumptions.[4] Our approach to statistical inference is crucial because it is by definition impossible to know everything about each financial intermediary's proprietary risk management framework.

Our new method provides a time-varying measure of residual interest rate risk exposure because it is based on financial intermediaries' publicly-traded equity values. As the owners of a financial intermediary, equity investors are the ultimate bearers of its interest rate risk. Equity values thus reflect intermediaries' exposure to interest rates *after* they have executed their interest rate risk management strategies. The correlation of equity values with interest rates reveals market participants' views on the effectiveness of financial intermediaries' hedging strategies in relation to the changes in the interest rates that occurred. The measure is a reflection

[3]Our limiting concept is the length of the time interval between two stock price observations going to zero. We provide the main theoretical results for our application in Appendix B.

[4]By contrast, bootstrapping the confidence intervals would require showing the time-series properties were preserved within samples or impose strong assumptions about the data generating process.

of the hedging strategy conditional on the actual changes in interest rates. A measure of zero doesn't necessarily mean that financial intermediaries are fully hedged. That said, intuitively, the stock price of a financial intermediary with fully hedged interest rate risk would be uncorrelated with all possible changes in interest rates (Allen, 1993).

Note that we will not address the question of *why* financial intermediaries bear interest rate risk. Importantly, we are not making any normative statement about how much interest rate risk financial intermediaries could or should carry. In particular, our notion of effectiveness does not imply that intermediaries should aim for zero residual interest rate risk. Nor does it imply that market participants think intermediaries should do so. Rather, our measure derives from the compensation for the interest rate risk borne by the ultimate owners of the intermediary, as in Allen (1993). When ownership is obtained through traded equity, the equity market price reflects that compensation.

Monitoring residual interest rate risk exposures is an important component in analysts, policymakers, and supervisors' evaluation of the financial conditions of intermediaries. Interest rate risk exposures are typically included as part of credit rating reports and investment analysis. As part of their financial stability discussions, central bankers are attuned to the potential effects on their decisions on financial intermediaries, e.g., Brainard (2022). Supervisors of financial institutions expect regular reports

concerning interest rate risk management and exposures. Monitoring is required because interest rates can change swiftly and significantly, with large potential effects. The profitability of entire financial sector industries has been threatened by interest rate exposures. For example, the life insurance industry struggled to cope with the sharp rise in interest rates in the late 1970s and early 1980s, when the Federal Reserve under Chairman Volcker fought inflation (NAIC, 2013).

We apply our new method to publicly-listed U.S. life insurers during the period from 2007 to 2022. Interest rate risk management is at the heart of the modern life insurer business model because the duration of life insurers' insurance liabilities, such as life insurance policies and annuity contracts, is typically much longer than the duration of the assets available in the economy.[5] This negative duration gap means that a decrease in the interest rate increases the present value of a life insurer's fixed-rate liabilities faster than the present value of its fixed income assets, which could lead to insolvency if left unmanaged. The same duration gap also means that persistently low interest rates depress life insurers' net investment spread on new business and forces them to reinvest the proceeds from maturing bonds into bonds paying lower coupon rates, which further depresses their overall net investment spread and, in turn, adversely affects their financial condition. In addition, explicit and implicit options on both

[5]For example, the duration of a typical life annuity is ten years, while the median corporate bond duration is around 5 years.

assets and liabilities contributes to life insurers' interest rate risk. Because the prospect of insolvency is incompatible with the sale of long-term life and longevity insurance, state insurance regulations, or both, life insurers must credibly manage interest rate risk.

We find that life insurer stock prices are largely *uncorrelated* with long-term (10-year) Treasury interest rates. This suggests that life insurers' interest rate risk management is effective most of the time. This finding is comforting given some of the largest life insurers in the U.S. have been managing interest rate risk for over a century. However, in some states of the world, realized gamma is statistically significant, revealing that after managing their interest rate risk—with liability driven investment, capital structure, and derivatives—life insurers remain exposed to changes in long-term interest rates in some states of the world.

We contrast our analysis of life insurance companies with an analysis of publicly-listed property and casualty (P&C) insurance companies. P&C insurers provide an ideal alternative to life insurers because the structure of their business means that they are relatively less exposed to interest rate risk. For example, the vast majority of P&C premiums are renewable every year and, therefore, P&C insurers do not need to actively manage a duration gap between their assets and insurance liabilities. Consistent with this difference in business model, we find that life insurers are more sensitive to changes in long-term interest rates than P&C insurers.

We then show that a measure of the term premium—the compensation for the risk associated with holding longer-term bonds—helps to explain the difference between the estimated sensitivities of life insurers and P&C insurers. We use the estimate of the term premium from the term structure model of Adrian, Crump and Moench (2013). We control for the funding cost of life insurers and a measure of the corporate credit return on life insurers' assets. Our finding likely reflects the outsized importance of longer-term debt in life insurers' investment portfolios. We use these results to illustrate how our measure provides information about the impact that rapidly changing interest rates may have on insurers.

Lastly, we show that our finding that life insurers' interest rate risk management is generally effective is *not* due to low long-term interest rate volatility. We provide two alternative approaches to address the potential endogeneity between realized gamma and long-term interest rate volatility. Both approaches are based on the exogenous increase in interest rate volatility that occurs on scheduled Federal Open Market Committee (FOMC) meeting days.

1.1 Related literature

Our paper connects to three distinct strands of literature. First, our method contributes to the high-frequency financial econometrics literature. Conceptually, our method is an extension of the single-factor realized beta

model of Andersen, Bollerslev, Diebold and Wu (2006) and Hansen, Lunde and Voev (2014). We include a second right-hand side variable, that is Treasury security returns, in the estimated regression specification. To the best of our knowledge we are the first to introduce a second right-hand side variable. Our computation of asymptotically valid standard errors using the subsampling approach is unusual in the high frequency financial econometrics literature because the approach is conservative and computationally intensive. Our realized gamma estimates do not suffer from bias due to non-synchronous trading—see, for example, Christensen, Kinnebrock and Podolskij (2010) and Barndorff-Nielsen, Hansen, Lunde and Shephard (2011)—since we use index data aggregated at the one-minute frequency. Our choice of five-minute sampling frequency and averaging immunizes our estimates from market microstructure noise biases, as described in Section 2.1.1.

Second, our method relates to—but is distinct from—studies of interest rate risk that measure the effects of realized changes in interest rates. These studies differ from other interest rate risk assessments that use balance sheet information to describe scenarios of potential effects associated with *hypothetical* changes in interest rates, e.g., Möhlmann (2021). Other papers that study *actual* changes in interest rates tend to focus on banks. Flannery and James (1984) studies the correlation between bank stock prices and interest rates using a similar regression model and weekly data. English,

Van den Heuvel and Zakrajšek (2018) identify the response of bank stock prices to FOMC interest rate shocks. Paul (2022) revisits the findings of English et al. (2018) by decomposing the effect of monetary policy surprises into changes in future expected short-term rates and changes in term premium. Hoffmann, Langfield, Pierobon and Vuillemey (2018) use supervisory bank balance sheet data to estimate interest rate risk and study its determinants in the cross section. Vuillemey (2019) and Begenau, Piazzesi and Schneider (2015) show that banks increase their exposure to interest rate risk using derivatives. Most of these papers study low-frequency data and, in some cases, attempt to identify interest rate shocks. By contrast, we exploit the information in high-frequency data and we use changes in interest rates rather than identified shocks.

Third, our paper adds to the extensive literature on risk management of financial institutions—e.g., Froot, Scharfstein and Stein (1993); Froot and Stein (1998). Our method is applicable to any financial intermediary. We chose to focus on the interest rate risk of life insurers, as they have received much less attention than, for example, banks. The theoretical foundation for our application to life insurers comes from recent work studying interest rate risk management at insurance companies (Foley-Fisher, Narajabad and Verani, 2016; Verani and Yu, 2021). In these papers, limited liability insurers manage the ex-ante risk of insolvency due to future movement in the interest rate by choosing an optimal insurance

price, asset portfolio, and capital structure. Our method is an ex-post statistical test of the performance of insurers' ex-ante interest rate risk management strategy. As such, our analysis is closely related to empirical work that measures the residual interest rate risk exposure of insurers using a two-variable regression model of stock prices and low-frequency data (Brewer III, Mondschean and Strahan, 1993; Berends, McMenamin, Plestis and Rosen, 2013; Hartley, Paulson and Rosen, 2016; Ozdagli and Wang, 2019; Sen, 2021; Koijen and Yogo, 2022; Huber, 2022). These estimates are weighted averages of the underlying time-varying interest rate risk parameter, where the weights depend on volatility that is potentially time-varying. If the time-varying parameter is correlated with the time-varying volatility, great care must be taken to avoid misspecifying the errors (Hamilton, 2008). Any analysis that compares estimates across time periods is subject to this concern. For example, we show in Appendix A that estimates obtained through low-frequency rolling window ordinary least squares (OLS) regressions are severely biased, inconsistent, and potentially misleading.[6] In contrast to our findings in this paper, incorrect inference on the OLS estimates suggests that life insurers are sensitive to any movement in long-term interest rates at almost all times in the post-crisis period.

[6]To be sure, not all the papers cited in this paragraph provide estimates use rolling windows, but all of them use low-frequency OLS. Some of the papers use stock prices only as a motivation for subsequent analysis of insurer balance sheet measures of interest rate risk.

The rest of our paper is structured as follows: Section 2 sets out the empirical framework for our estimation and explains how we construct our standard errors using subsampling. Section 3 describes our application to US life insurers, including institutional background and details on the data. We summarize our main findings in section 3.3 and offer some concluding remarks in section 4.

2 Methodology

2.1 A two-variable regression model

In this section, we introduce our new method to measure the residual interest rate risk exposure of financial intermediaries. Let r_{ijt} be the continuously compounded stock return of financial intermediary i indexed to minute j within day t. Let r_{mjt} be the continuously compounded return on aggregate market m and r_{yjt} be the continuously compounded return on Treasury security y.

Our framework is a regression model with two right-hand side variables using minute-by-minute financial market returns:

$$r_{ijt} = \alpha_t + \beta_t r_{mjt} + \gamma_t r_{yjt} + \epsilon_{ijt} \tag{1}$$

where $\{\alpha_t,\ \beta_t,\ \gamma_t\}$ are day-specific coefficients estimated using within-day

returns. Our regression with the restriction $\gamma_t = 0$ is well established in the finance literature and is referred to as the one-factor capital asset pricing model (CAPM). In the CAPM regression, the coefficient β_t is interpreted as a dynamic measure of the comovement of individual stock returns with aggregate market or systematic returns.[7] We extend the one-variable CAPM regression to include a second right-hand side variable, that is Treasury security returns.[8]

The time-varying γ_t coefficient estimates the sensitivity of an individual firm's stock price returns to high-frequency realized changes in Treasury security returns. As Treasury security returns are inversely dependent on changes in interest rates, γ_t provides an estimate of that firm's interest rate sensitivity.

We label our γ_t estimates *realized gamma* because our method can also be cast in the nonparametric framework of realized variances and covariances (Meddahi, 2002; Barndorff-Nielsen and Shephard, 2004; Andersen, Bollerslev and Meddahi, 2004). Our estimates of daily gammas are based on realized daily variances and covariances after conditioning on the aggregate market returns. We first project out aggregate stock market

[7]A full discussion of the extensive literature studying time-varying β_t and its determinants is beyond the scope of this paper. See Fama and French (2004) for an overview.

[8]To be sure, we are not assuming that the right-hand side variables in our regression model are orthogonal. We are estimating the general equilibrium relationship between the three variables in our regression, which is fully consistent with the standard one-factor CAPM and yields an unbiased estimate of γ_t. Other papers that adopt a similar approach include Fama and Schwert (1977) and Flannery and James (1984). We explore the effect of an exogenous increase in long-term Treasury rate volatility in section 3.4.

returns from stock returns and Treasury security returns by running two auxiliary regressions for each day t:

$$r_{ijt} = \hat{\alpha}_t^1 + \hat{\beta}_t^1 r_{mjt} + \hat{\epsilon}_{ijt}, \tag{2}$$

$$r_{yjt} = \hat{\alpha}_t^2 + \hat{\beta}_t^2 r_{mjt} + \hat{\epsilon}_{yjt}. \tag{3}$$

The residuals from these auxiliary regressions $\{\hat{\epsilon}_{ijt}, \hat{\epsilon}_{yjt}\}$ are, respectively, the within-day conditional stock returns and Treasury security returns. The daily realized covariance of each financial intermediary's conditional stock returns and Treasury security conditional returns is given by:

$$\hat{v}_{i,y,t} = \sum_j \hat{\epsilon}_{ijt} \cdot \hat{\epsilon}_{yjt}.$$

And the daily realized variance of conditional Treasury security returns is given by:

$$\hat{v}_{y,t} = \sum_j \hat{\epsilon}_{yjt}^2.$$

So we can define realized gamma as the ratio of the conditional covariance to the daily realized conditional variance of Treasury security returns:

$$\gamma_t = \frac{\hat{v}_{i,y,t}}{\hat{v}_{y,t}}. \tag{4}$$

The γ_t estimates by equation 1 and equation 4 are identical by what

is commonly-known as the Frisch-Waugh-Lovell Theorem (Davidson and MacKinnon, 1993, Section 1.4). However, care must be taken with interpretation. The simple regression shown in equation 1 yields a consistent estimate of the ex-post realized gamma coefficient. That said, obtaining asymptotically valid standard errors is not a simple process, as we will describe in Section 2.2.

2.1.1 Addressing market microstructure noise

Controlling for market microstructure noise that is prevalent in high frequency financial market data is an important issue (Aït-Sahalia and Yu, 2009). Microstructure noise naturally arises from a variety of features built in to financial market trading, including prices bouncing from bids to asks, variation in the size of trades, adjustment to new information contained in prices, order flow dynamics, and inventory management. Following Aït-Sahalia and Mykland (2009), we address the presence of market microstructure noise without discarding observations from our samples.

We employ two well-established techniques to mitigate concerns that market microstructure noise is clouding our ability to construct estimators and draw inference from high-frequency data. First, we calculate returns at five-minute intervals as their use as a benchmark for estimators generally outperforms all alternatives (Liu, Patton and Sheppard, 2015). We use

every possible five-minute grid point in a trading day to exploit all available high-frequency information given the data structure as described in Zhang et al. (2005).[9] Second, we filter all of our returns time series through AR(1) processes estimated separately for each day. That is, we take the raw returns r_{kjt} for $k \in \{i, y, m\}$ and estimate $r_{k,j,t} = \rho + \phi r_{k,j-1,t} + \varepsilon_{k,j,t}$ for each day t. We then use the residuals $\varepsilon_{k,j,t}$ as our returns time series that has filtered out market microstructure noise.

2.2 Statistical inference

A key principle for our new methodology is to impose minimal assumptions about the data generating process. This principle underpins our use of high-frequency data to estimate nonparametrically the time-varying correlation between interest rates and financial intermediaries' stock prices. Similarly, we follow this principle when we consider what standard errors are appropriate for valid inference. We derive asymptotically valid standard errors without imposing undue structure on the time series processes. Our choice of standard errors is a crucial part of our approach to estimate interest rate risk, as the data generating process underpinning our realized gamma estimates is nonstandard. For example, we use rolling five-minute windows to construct our time series of returns.

[9] Our approach is identical to the method commonly referred to as "subsampling" in high-frequency financial econometrics. We avoid using the term here to prevent confusion with the concept of "subsampling" that we use to construct asymptotically valid standard errors.

We adopt the subsampling methodology as it is a valid technique in extremely general cases (Politis et al., 1999).[10] The basic idea of subsampling in a time series context is to approximate the sampling distribution using all possible subsets of the time series. Theorem 4.3.1 from Politis et al. (1999), which we reproduce in Appendix B for completeness, shows that we can derive asymptotically valid confidence intervals for the daily estimator γ_t. In addition, we can draw asymptotically valid inference about the true γ_t by exploiting the familiar duality between the construction of confidence intervals for γ_t and the construction of hypothesis tests about γ_t. We can test the null hypothesis that our estimate of the daily γ_t is statistically different from 0. That is, under the null, financial intermediaries are hedged against interest rate risk as their stock prices are not sensitive to movements in interest rates.

Our algorithm for hypothesis testing uses within-day observations to construct subsamples. We follow Politis et al. (1999) and evaluate statistics on an exhaustive set of subsamples of size $b < n$ that are created from the original daily sample of size n. We estimate the distribution of this statistic after a suitable normalization for each day in our sample. Note that our limiting concept is that the number of observations in a day approaches infinity. To be clear, each subsample contains consecutive observations from

[10] Alternative methodologies based on the bootstrap technique could be devised, but they typically require additional assumptions, such as a finite fourth moment of the model residuals (Paparoditis and Politis, 2009).

the original time series sample. Therefore, each subsample of size b is drawn without replacement from the true data generating process. We calculate a confidence interval for each of the daily γ_t using subsampling following Politis et al. (1999) under the assumption that the errors are asymptotically stationary. Asymptotic stationarity is a weak condition that means, for example, the errors could follow an $AR(1)$ process with autocorrelation parameter strictly less than 1 and heteroskedastic innovations. The essence of the subsampling method is to approximate the sampling distribution of the (normalized) γ_t estimate with the empirical distribution generated by its subsample counterpart.

As we have a large daily sample size (roughly speaking, $n = 390$ observations per day), the choice of subsample size (b) should not have a large effect on the empirical distribution of our statistic. Nevertheless, we need to choose the size of our subsamples. We follow the algorithm proposed by Politis et al. (1999) in section 9.3.3. Let b^t be the subsample size for day t, which yields a confidence interval $\{I_{b^t,low}, I_{b^t,high}\}$. We construct a discrete grid of possible values $b_s^t \in \{b^t_{small}, \ ..., \ b^t_{large}\}$. For each subsample size b_s^t we consider a perturbation of small integer k around the subsample size and calculate a measure of variation in the confidence interval:

$$VI_{b_s^t} \equiv \text{var}\left(I_{b_s^t-k,low}, ..., I_{b_s^t+k,low}\right) + \text{var}\left(I_{b_s^t-k,high}, ..., I_{b_s^t+k,high}\right).$$

Finally, we pick the value of b that delivers stable confidence intervals for the most number of days in the entire sample:

$$b = \text{argmax}_b \sum_{t=0}^{T} \mathbb{1}(b^{t*} = b) \text{ where } b^{t*} = \text{argmin}_{b_s^t} VI_{b_s^t}.$$

Having determined the 'optimal' subsample size, we construct the empirical distribution of the normalized γ_t estimate for each day t. We use empirical distributions to obtain confidence intervals, which allow us to make inference about the statistical significance of each γ_t. With our new methodology in hand, we can turn to a specific application and data.

3 Application to U.S. life insurers

3.1 Institutional background

Life insurers play a major role in the financial system, holding $6 trillion in total assets in their general accounts, of which roughly $3 trillion are in corporate and foreign fixed income securities (Federal Reserve release Z.1 table L.116.g). Their overall business model consists of earning a spread between the yield they owe on their insurance liabilities and the yield they earn on the assets backing those liabilities. Life insurers write liabilities that are traditionally long-term, illiquid, and make fixed payments, such as fixed annuities. Life insurers tend to invest their premiums primarily

in fixed rate corporate debt, in an effort to match their asset and liability cash flows and illiquidity profile and to offer a competitive return to policy holders.

Like other financial intermediaries, life insurers have multiple sources of exposure to interest rate risk. A key underlying reason for their exposure is that the duration of insurance liabilities is typically much longer than the duration of assets available in the economy. In the U.S., the typical duration of life insurance liabilities is 15–20 years (Huber, 2022). By contrast, in most countries, long-term fixed coupon bonds with more than two-year maturity do not exist (Gajek and Ostaszewski, 2004). Even in the U.S., which has the largest corporate bond market in the world, the supply of long-duration corporate bonds paying fixed interest rates is considerably smaller than the size of the life insurance industry (Verani and Yu, 2021). This means that, in practice, it is difficult for life insurers to hedge interest rate risk by investing in assets that have the same duration and greater cash flow variability than their insurance liabilities i.e., they cannot directly implement the classical immunization strategy of Redington (1952).

Convexity—the effect of changing interest rates on the duration—of life insurer assets and/or liabilities also contributes to interest rate risk. One well-known source of convexity stems from options on financial contracts. For life insurers, the option for corporate bond issuers to call their bonds creates convexity on the asset side of their balance sheet. Likewise,

policyholders may have the option of surrendering their life insurance products–perhaps for some cost–that creates convexity on the liability side of the balance sheet. The combination of these options creates a short straddle position for investors in the life insurer, which means they suffer when volatility is high (Babbel and Stricker, 1987).[11]

A natural way for life insurers to manage their interest rate risk consists of choosing a price for their insurance liabilities, an asset portfolio to back their insurance liabilities, and a capital structure to prevent insolvency along different paths for interest rates (Verani and Yu, 2021). For example, life insurers can hedge interest rate risk by charging a markup on the actuarially fair cost of their insurance products. The present value of the markup adds to the insurer's 'net worth'. Net worth allows the insurer to close its duration gap by financing bonds whose present value is greater than the present value of its insurance liabilities. Or, put differently, net worth acts as precautionary savings and helps cushion the effect of interest rate changes that disproportionately affect the value of the insurance liabilities.[12]

Large and sophisticated life insurers also manage interest rate risk by adding net-positive duration to their balance sheets synthetically using derivatives (Sen, 2021; Verani and Yu, 2021) and nontraditional lines of

[11]Briys and de Varenne (1997) provide an alternative formulation for the investor straddle position in which insurance liabilities are more convex than assets.

[12]Net worth is not to be confused with what the industry calls reserves, which is the value of insurance liabilities.

business (Foley-Fisher et al., 2016; Foley-Fisher, Narajabad and Verani, 2020), which amounts to using leverage instead of net worth to close the duration gap. For example, life insurers can add positive duration to their balance sheet by entering into a long-term *fixed-for-float* interest rate swaps or by financing long-term fixed interest rate assets with nontraditional liabilities such as overnight securities lending cash collateral (Gissler, Foley-Fisher and Verani, 2019; Foley-Fisher et al., 2016) and funding agreement-backed short-term funding (Foley-Fisher et al., 2020). All these interest rate hedging strategies amount to closing the insurer's natural negative duration gap by either directly or synthetically financing fixed-maturity assets with short-term floating rate debt.

Nevertheless, insurers typically carry residual interest rate risk after they have implemented their hedging strategies. Investors in the insurers—either the policyholders in the case of mutual insurers or shareholders in the case of publicly-listed insurers—provide additional risk-bearing capital and receive compensation for bearing the insurer's residual interest rate risk (Allen, 1993). When investment takes place through traded equity, the market price for the equity reflects the interest rate risk compensation.

One real-world example when the residual interest rate risk carried by life insurers was realized occurred in the early 1980s. At that time, the Federal Reserve sharply increased short-term interest rates amid persistently high inflation. Life insurers' financial condition deteriorated as

policyholders surrendered their claims or took out policy loans in search of higher interest rates on alternative saving vehicles (Briys and de Varenne, 2001). Life insurers responded by rewriting existing business at a loss and selling new products that offered higher-than-current long-term rates (negative spreads) (NAIC, 2013). While locking in huge losses—eroding their net worth—they avoided even greater losses they would have incurred had they sold their fixed income assets at far-below costs given the rise in current rates. The surge in short-term interest rates occurred after a relatively long period of low interest rate volatility, making these sharp rises largely unexpected. The significance of the episode is underscored by subsequent efforts to develop new tools for managing interest rate risk (Doffou, 2005).

Adverse scenarios such as the early 1980s create a need for researchers and policymakers to monitor and assess the effects of rising interest rates on life insurers. However, they do not have access to the complete set of balance sheet information needed to precisely identify the effectiveness of life insurers' interest risk management and their residual interest rate risk. For example, information about the interest rate sensitivity of life insurance liabilities is difficult to gauge, although it is easier in some non-U.S. jurisdictions (Huber, 2022; Möhlmann, 2021; Kirti, 2017; Domanski, Shin and Sushko, 2017). Furthermore, it is hard to incorporate balance sheet information about the interest rate sensitivity of derivative positions,

off-balance sheet liabilities (such as those in offshore captive reinsurers), and nontraditional liabilities.

To overcome this problem, researchers turned to analyzing the sensitivity of life insurer stock returns to changes in long-term interest rates. An insurer's equity valuations reflect the market price for its residual interest rate risk, after it has implemented its hedging strategies. That is, the ex-post effectiveness of life insurers' management of ex-ante interest rate risk.[13] To the best of our knowledge, this approach was first adopted by Brewer III et al. (1993). To assess the *dynamics* of interest rate risk exposure, some papers run OLS on rolling windows of stock returns e.g., Hartley et al. (2016). Although conceptually valid, the OLS implementation can lead to biased estimates in the presence of heteroskedasticity.[14] In Appendix A, we show the bias is extremely large by imposing some structure on the data generating processes. We will now apply our preferred methodology described in Section 2 to obtain consistent estimates without imposing such structure.

[13]Here, again, the term 'effectiveness' should not be taken to imply that investors think insurers should target any particular level of interest rate risk. Rather, it's investors' assessment of the effect that actual interest rate changes had on the net worth of the insurer.

[14]Brewer III, Carson, Elyasiani, Mansur and Scott (2007) recognised this concern and allowed for time-varying volatility in a GARCH-M process.

3.2 Data

All the price data for our empirical application to life insurers come from Refinitiv. The underlying data are timed to the microsecond and recorded from data feeds covering both over-the-counter and exchange traded instruments on more than 500 trading venues and third parties. We use a preprocessed version of the underlying data aggregated by Refinitiv to a minutely frequency using the last trade during each minute. We construct the data so as to follow the *previous tick method*, that is, if there are no transactions during a specific minute, the last transaction is used.

The dataset identifier for each dataseries typically combines a ticker with a code indicating the primary trading market. For example, MetLife's identifier is `MET.N` as it trades on the New York Stock Exchange. The list of the individual insurer identifiers and their mapping to the life and P&C insurers used in our analysis is provided in Table 1. Column 2 of Table 1 shows the insurers included in each index. Our list of publicly-listed life insurers almost completely overlaps with the list of "publicly traded U.S. variable annuity insurers" used by Koijen and Yogo (2022). This is not surprising because virtually all large listed life insurers offer variable annuities contracts at some point in the sample period.[15] In addition, our analysis uses Standard and Poor's S&P500 index as our measure of the

[15] As we will discuss in Section 3.4, this means that it is not possible to attribute the residual interest rate risk exposure to variable annuities.

aggregate market. The identifier for the index is `.SPX`. We also use Refinitiv evaluated prices for 10-year Treasury securities. The identifier for the series is `US10YT=RRPS`. Evaluated prices contain information from actual trades, quotes, and other sources within a model-based methodology.

Table 1: Mapping insurance groups to identifiers. This table shows the insurance groups that we use in our empirical application, with their respective NAIC Group codes, and identifiers.

Name	Code	Life/P&C	Identifier	Ticker	Notes
Alleghany Group	501	P&C	Y.N	Y	
American Financial Group	84	P&C	AFG.N	AFG	
American Intl Group, Inc.	12	Life/P&C	AIG.N	AIG	
Assurant, Inc.	19	Life/P&C	AIZ.N	AIZ	
The Allstate Corporation	8	Life/P&C	ALL.N/BK.N	ALL/BX	Identifier change for Life in 2021
Ameriprise Financial, Inc.	4	Life	AMP.N	AMP	
American National Financial Group	408	Life	ANAT.OQ	ANAT	
Apollo Global Management, Inc.	4734	Life	ATH.N/APO.N	ATH/APO	Identifier change in 2022
Brighthouse Financial, Inc.	4932	Life	BHF.OQ	BHF	
Berkshire Hathaway Inc.	31	P&C	BRKb.N	BRK.B	
Chubb Ltd.	626	P&C	ACE.N/CB.N	ACE/CB	Identifier change in 2016
Cigna Health Group	901	Life	CI.N	CI	
Cincinnati Financial Corporation	244	P&C	CINF.OQ	CINF	
CNA Financial Corporation	218	P&C	CNA.N	CNA	
CNO Financial Group	233	Life	CNO.N	CNO	
Erie Insurance Group	213	P&C	ERIE.OQ		
Equitable Holdings, Inc.	4965	Life	EQH.N	EQH	
FBL Financial Group Inc.	513	Life	FFG.N	FFG	Ceased trading in 2021
Fidelity and Guaranty Life	4731	Life	FGL.N	FGL	Ceased trading in 2017
Fidelity National Financial, Inc.	670	Life	FNF.N	FNF	
Genworth Financial, Inc.	4011	Life	GNW.N	GNW	
Hanover Insurance Group, Inc.	88	P&C	THG.N	THG	
The Hartford Fin. Svcs Group, Inc.	91	Life/P&C	HIG.N	HIG	Remove identifier from Life in 2018
Horace Mann Group	300	Life	HMN.N	HMN	
Kansas City Life Insurance Group	588	Life	KCLI.OQ	KCLI	Delisted in 2015
Kemper Corporation Group	215	P&C	KMPR.N	KMPR	
Lincoln National Corporation	20	Life	LNC.N	LNC	
Mercury General Group	660	P&C	MCY.N	MCY	
Markel Corporation Group	785	P&C	MKL.N	MKL	
MetLife, Inc.	241	Life	MET.N	MET	
Manulife Financial Corporation	904	Life	MFC.TO	MFC	
Nationwide Corporation Group	140	Life	NFS.N	NFS	Ceased trading in 2008
The Phoenix Companies, Inc.	403	Life	PNX.N	PNX	Ceased trading in 2016
Primerica Group	4750	Life	PRI.N	PRI	
Principal Financial Group, Inc.	332	Life	PFG.OQ	PFG	
Protective Life Corporation	458	Life	PL.N	PL	Ceased trading in 2015
The Progressive Corporation	155	P&C	PGR.N	PGR	
Prudential Financial, Inc.	304	Life	PRU.N	PRU	
Selective Insurance Group	88	P&C	THG.N	THG	
Symetra Financial Corp.	4855	Life	SYA.N	SYA	Ceased trading in 2016
The Travelers Companies, Inc. Group	3548	P&C	TRV.N	TRV	
Voya Financial, Inc.	4832	Life	VOYA.N	VOYA	
W. R. Berkley Corporation	98	P&C	BER.N/WRB.N	BER/WRB	Identifier change in 2008

We use minutely data for each trading day beginning at 9:30am

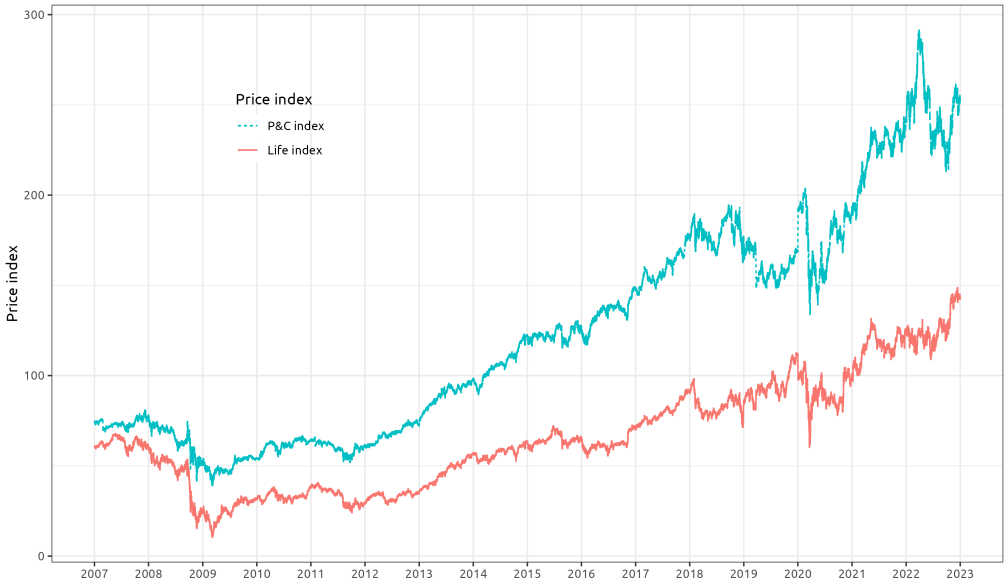
through 4pm.[16] Except for 9:30am, we use closing prices recorded for each minute. For 9:30am, we use the opening price of 9:31am to avoid concerns about jumps following overnight information and trading. We calculate five-minute log returns of all time series using every possible five-minute grid point in a trading day. That is, we calculate the returns $ln(p_{i,j,t}) - ln(p_{i,j-5,t})$ for each day t, data series i, and all $j \in \{9.35am,\ 9.36am,\ \cdots,\ 3.59pm,\ 4.00pm\}$.

We construct high-frequency price indexes separately for life insurers and P&C insurers, weighting each individual insurer's intraday market price by its end-of-day market capitalization. We obtain daily data on market capitalization from the Center for Research in Security Prices hosted by Wharton Research Data Services. Figure 1 shows the life insurer index as a red solid line and the P&C insurer index as a dotted blue line. The dotted blue line lies above the red solid line as P&C insurers have generally outperformed life insurers in the post-crisis low interest rate environment.

Table 2 shows summary statistics for the high-frequency data used in our analysis. Column 1 shows that the first day that data are available is different for each of our variables. The S&P500 Index is earliest available, while our high frequency data on long-term Treasury bond prices ('US10YT') begin only in 2007. Our indexes of large life and P&C insurers

[16] We exclude holidays, weekends, emergency closures, and partial trading days.

Figure 1: Insurer price indexes. Each line is a weighted average high-frequency price for large publicly-traded insurers listed in Table 1. The weights for each series are the daily market capitalization of insurers. Source: Authors' calculations based on data from Refinitiv and the Center for Research in Security Prices.



stock prices also begin in 2007. By construction, our sample ends on October 31, 2022. In addition to the first date available, we report the number of days, and the total number of minutely five-minute returns in our data. We also report that there are no zero returns in our data, alleviating concerns about downward bias in our estimates due to zero returns (Bandi, Kolokolov, Pirino and Renò, 2020; Kolokolov and Renò, 2023). Across these returns, we report the mean, median, standard deviation, percentiles, and higher-order moments for each time series. Life insurers' returns have a higher standard deviation than P&C insurers, but the kurtosis of life insurers' returns is far lower.

Table 2: Summary statistics. For each returns series in our sample, the table shows the first observation date, the number of days, the number of five-minute returns, the number of returns equal to zero, as well as the mean, median, standard deviation, percentiles, skewness, and kurtosis. The statistics reported in columns 6 through 10 are multiplied by $1e+4$ for legibility. Source: Authors' calculations based on data from Refinitiv and the Center for Research in Security Prices.

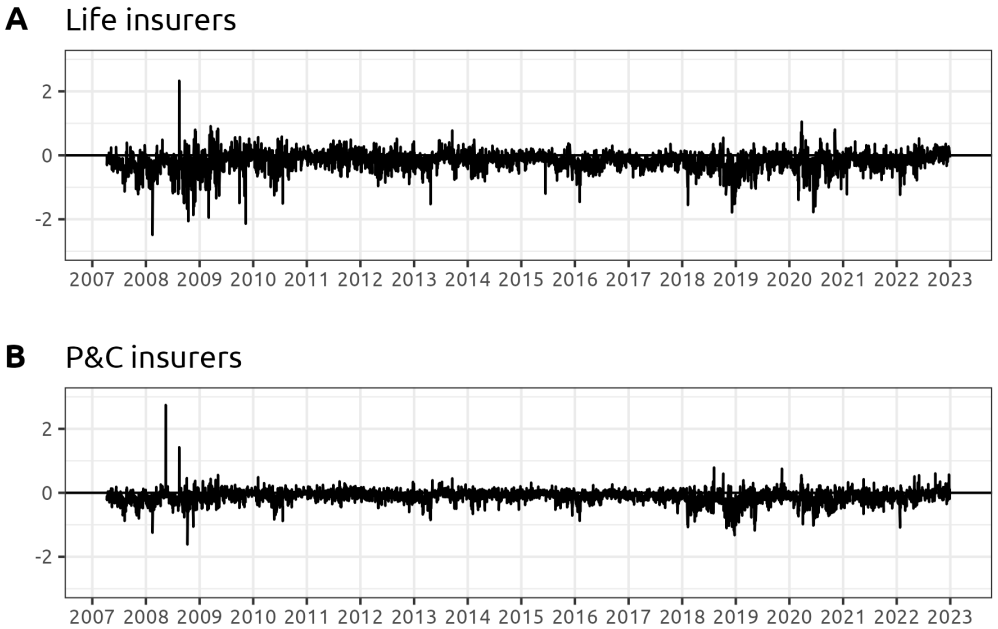
Series	First date	No. days	No. obs.	#Zeroes	Mean	Median	Std. dev.	p25	p75	Skew.	Kurt.
S&P500	2000-01-03	5,768	2,218,825	0	-0.01	0.05	10.74	-4.06	4.11	-0.04	38.74
US10YT	2007-04-10	3,961	1,524,682	0	0.01	0.00	3.91	-1.67	1.70	0.99	80.66
Life	2007-01-03	3,996	1,533,761	0	-0.01	0.06	16.55	-5.76	5.79	0.44	35.25
P&C	2007-01-03	3,996	1,533,761	0	0.02	0.03	10.44	-4.00	4.02	2.06	147.25

3.3 Results

In this section, we apply the methodology laid out in Section 2 to the data described in the previous section. Panel A of Figure 2 shows the daily

point estimate of realized gamma for life insurers, and Panel B of the same figure shows the daily point estimate of realized gamma for P&C insurers. Both panels exhibit volatility, which is a well-known feature of time-varying coefficients estimated using realized variances and covariances (Hansen et al., 2014). Nevertheless, life insurers' realized gamma evidently has a higher level of volatility than P&C insurers'.

Figure 2: Daily realized gammas. The panels show daily realized gammas for life insurers and P&C insurers from 2007 through to the end of 2022. Source: Authors' calculations based on data from Refinitiv and the Center for Research in Security Prices.



We obtain confidence intervals from the empirical distributions, which are estimated for each day. Table 3 shows the results from applying the algorithm described in subsection 2.2 to determine the block size. While any block size satisfying the conditions of Theorem B.1 is valid, the ideal

Table 3: Optimizing block size. Each of columns 2-4 shows the number of days on which the block size (column 1) produces the most stable, i.e. least variable, confidence intervals. The measures of variation used in columns 2 and 4 is the standard deviation, while columns 3 and 5 use the difference between the minimum and the maximum values. The row with the highest count of days reveals the ideal block size for life insurers (columns 2-3) and P&C insurers (columns 4-5). Source: Authors' calculations based on data from Refinitiv and the Center for Research in Security Prices.

Block	Life		P&C	
size (%)	Std. dev.	Min-max	Std. dev.	Min-max
15	481	475	519	513
20	419	418	381	379
25	992	1000	988	995
30	676	673	702	700
35	597	594	618	615
40	472	480	452	454
45	418	415	395	399

block size is the one that produces the most stable i.e., least variable, confidence intervals given a small perturbation in the size of the block.[17] Each of columns 2-4 shows the number of days on which the block size (column 1) produces the most stable confidence intervals. The measure of variation used in columns 2 and 4 is the standard deviation, while columns 3 and 5 use the difference between the minimum and the maximum values. The row with the highest count of days reveals the ideal block size for life insurers (columns 2-3) and P&C insurers (columns 4-5). For both life insurers and P&C insurers, the optimal block size is 25 percent of the daily

[17]Note that there is no reason to expect variation across grid points to follow a monotonic function or have a global optimum (Politis et al., 1999).

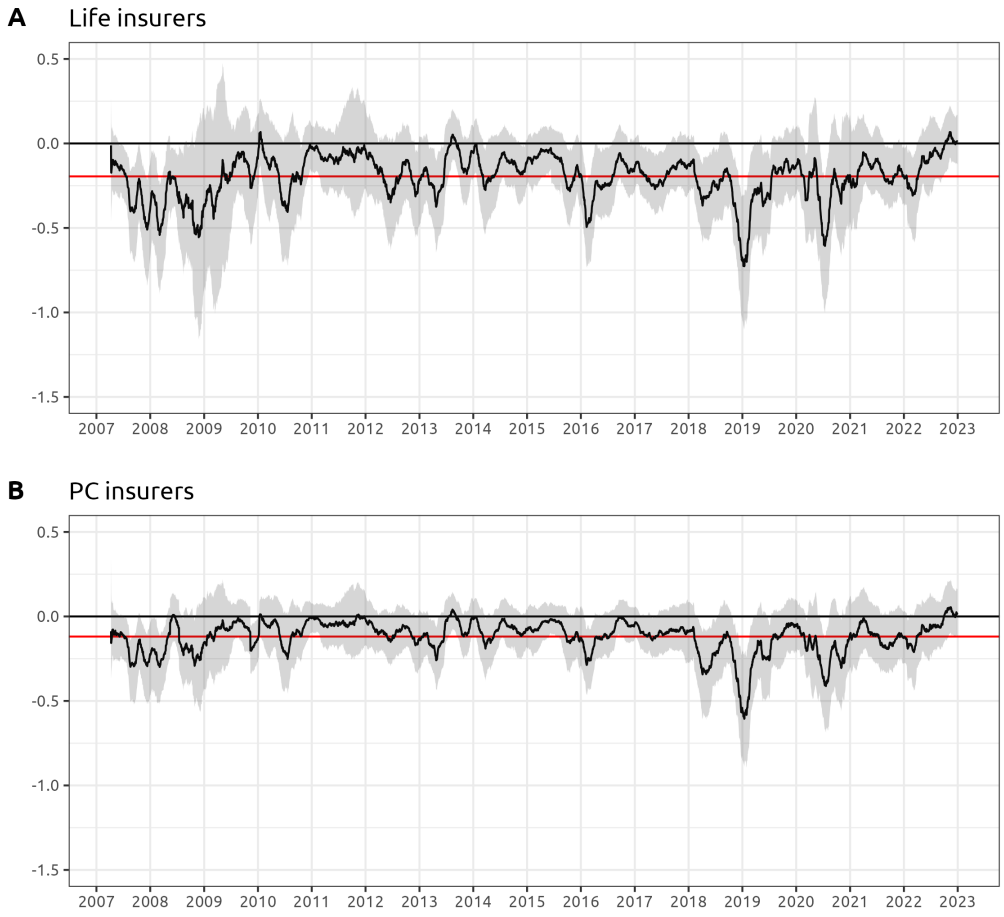
observations, corresponding to about 100 consecutive observations in each block and about 300 points in the empirical distribution. These relatively large values alleviate concerns about the power of the test.

We average our daily estimates using a rolling window of two months, to obtain a consistent estimate of interest rate risk over a longer horizon.[18] We construct smooth series for both the point estimates of realized gammas and the confidence intervals. Panels A and B in Figure 3 show the smoothed time series. The red horizontal lines represent the sample means of the respective series.

The smoothed time series reveal that realized gamma for life insurers is statistically significant on only 1,261 days, equivalent to roughly 32 percent of the sample. Realized gamma is always negative whenever it is statistically significant, which means that life insurers would benefit from higher long-term interest rates. For the majority of our sample, life insurer stock prices are uncorrelated with long-term interest rates. This suggests that life insurers' interest rate risk management is effective most of the time. These results should not be interpreted as a normative assessment of life insurers' interest rate risk management, neither by us nor by equity market participants. The measure is a reflection of how actual changes in interest rates affected—or did not affect—equity investors in life insurers, who expect compensation for bearing interest rate risk.

[18] A feature of our realized volatility estimates is the ability to aggregate them over time (Corsi, 2009).

Figure 3: Smoothed daily realized gammas. The panels show daily realized gammas averaged using a rolling window of two months for life insurers and P&C insurers from 2007 through to the end of 2022. The shaded region in both panels represents the 90 percent confidence intervals for each daily estimate. The underlying data are shown in Figure 2. The red horizontal lines represent the sample means of the respective series. A negative realized gamma means that insurers would benefit from higher long-term interest rates. Source: Authors' calculations based on data from Refinitiv and the Center for Research in Security Prices.



The time series also reveal that life insurers are more sensitive to interest rate changes than P&C insurers. In contrast to life insurers, realized gamma for P&C insurers is significant on only 705 days (about 18 percent of the sample). Like life insurers, realized gamma for P&C insurers is always negative whenever it is statistically significant. This finding could be interpreted as evidence that P&C insurers carry less residual interest rate risk, or as evidence that life insurers are exposed to different kinds of interest rate risk. In the next section, we offer some support for the latter interpretation by analyzing individual components of long-term interest rates.

3.4 Analysis

3.4.1 When is life insurer hedging not effective?

In this section, we study macroeconomic variables during periods when realized gamma is statistically significant. Our findings help to explain why life insurers' realized gamma is more often statistically significant that P&C insurers' realized gamma. While we offer an interpretation of our findings, we do not claim causal identification, as we recognize that long-term yields are a general equilibrium outcome of supply and demand (Schneider, 2022). Life insurers' interest rate sensitivity is potentially endogenous to their demand for compensation to hold longer-term debt and, as we noted earlier,

life insurers are important investors in the long-term debt market. As many macroeconomic variables are unavailable at intraday frequencies, all the analysis in this section is conducted at a daily frequency. In an ideal empirical experiment, we would use intraday data to analyze the force(s) behind the results described in the previous section. However, we do not know of any high-frequency measures of the variables described below.

We focus on three key variables based on Verani and Yu (2021), who showed that the relative cost of hedging interest rate risk is determined by the long-term investment grade bond spread relative to life insurers' cost of funding. As measures of the return on life insurers' long-term assets, we use the term premium and Moody's Baa-Aaa seasoned corporate spread.[19] We use the term structure model of Adrian et al. (2013) to decompose long-term yields and obtain an estimate of the term premium. While the term premium contributes to the slope of the yield curve, it is more specifically the component that compensates investors for holding longer-term debt instead of rolling over short-term debt. In addition to these measures of asset returns, we use the ICE BoA Single-A U.S. corporate index option-adjusted spread as a proxy for life insurers' average cost of funding because life insurers are rated around A. Summary statistics for all the variables

[19]The corporate bonds used to construct this spread all have at least 20 years of maturity. The yield on Aaa-rated corporate bonds with at least 20 years of maturity is a quasi-risk free benchmark. Under state insurance regulation, corporate bonds rated by Moody's to be Baa or higher are designated as NAIC 1 and uniformly attract the lowest statutory risk-based capital charge.

used in this analysis are provided in Appendix C.

We construct a binary variable that takes the value 1 if the estimated realized gamma (γ_{it}) is statistically significant on day t for insurer type $i \in$ {Life, P&C}, and takes the value 0 otherwise. We then estimate a linear probability model using as independent variables the term premium (TP_t) estimates from Adrian et al. (2013), the Moody's Baa-Aaa seasoned corporate spread ($Baa - Aaa_t$), and a measure of the funding cost of life insurance companies (FC_t) that is the ICE BoA Single-A U.S. corporate index option-adjusted spread. In more technical terms, we estimate:

$$\widehat{P}(\mathbb{1}(\gamma_{it} < 0)|TP_t, Baa - Aaa_t, FC_t) = \alpha_i + \beta_i^1 TP_t + \beta_i^2 Baa - Aaa_t + \beta_i^3 FC_t$$

where $\widehat{P}(\mathbb{1}(\gamma_{it} < 0)|TP_t, Baa - Aaa_t, FC_t)$ is the predicted probability that $\gamma_{it} < 0$ given TP_t, $Baa - Aaa_t$, FC_t, and a linear functional form.

The results are shown in Table 4 where we report the coefficient estimates and standard errors in parentheses. Column 1 shows the bivariate relationship between the term premium and the statistical significance of realized gamma for life insurers. Columns 2-4 provide the main result under a range of standard error estimates, as indicated at the bottom of the table. HC are heteroskedasticity consistent standard errors, HAC are heteroscedasticity and autocorrelation consistent standard errors, and NW are Newey-West standard errors. The dependent variable in column 5 is a

binary variable for statistical significance of P&C insurers' realized gamma. This column acts as a placebo test of the main result for life insurers: The key variables we focus on for life insurers are not statistically important for P&C insurers, consistent with our prior expectations.

Noting again that the results are not causal, the estimates nevertheless suggest that there is a strong economic relationship between the variables, in addition to the statistical significance indicated in the table. We use as a benchmark for the economic effects the 32 percent unconditional probability that realized gamma for life insurers is statistically significant (see Table 6 in Appendix C). A one standard deviation increase in the term premium, which compensates investors for holding longer-term debt, reduces the probability that realized gamma for life insurers is statistically significant by about 17 percentage points—equivalent to about half of the unconditional probability that realized gamma for life insurers is statistically significant. A one standard deviation increase in Moody's Baa-Aaa seasoned corporate spread reduces the probability that realized gamma for life insurers is statistically significant by about 24 percentage points. And a one standard deviation increase in the ICE BoA Single-A U.S. corporate index option-adjusted spread raises the probability that realized gamma for life insurers is statistically significant by about 28 percentage points.

Our analysis provides support for the view that a flattening yield curve

Table 4: When are insurers not hedged? The table shows the results from estimating a linear probability model with a binary dependent variable that takes the value 1 if the estimated realized gamma (γ_{it}) is statistically significant on day t for $i \in$ {Life, P&C} and 0 otherwise. Other variables are the term premium (TP_t) estimates from Adrian et al. (2013), the Moody's Baa-Aaa seasoned corporate spread ($Baa - Aaa_t$), a measure of the funding cost of life insurance companies (FC_t) that is the ICE BoA Single-A U.S. corporate index option-adjusted spread, and the daily realized volatility (σ_t^{10yt}) of the ten-year Treasury interest rate. Columns 2-4 show the main result for three different standard error specifications. HC are heteroskedasticity consistent, HAC are heteroscedasticity and autocorrelation consistent, and NW are Newey-West standard errors. Columns 7–8 show a two-stage least squares estimation where the daily realized volatility (σ_t^{10yt}) is instrumented using a binary variable ($FOMC_t$) that takes the value 1 if there was a scheduled FOMC meeting on that day and 0 otherwise. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices, FRED, and Adrian et al. (2013). *** p<0.01, ** p<0.05, * p<0.1

Dep. var.:	Life				PC	Life	*2SLS: 1*	*2SLS: 2*
							σ_t^{10yt}	Life
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TP_t	−0.09***	−0.15***	−0.15***	−0.15***	−0.04	−0.15***	−0.03*	−0.14***
	(0.01)	(0.01)	(0.04)	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)
$BAA - AAA_t$		−0.52***	−0.52**	−0.52***	−0.10	−0.53***	−0.25***	−0.51***
		(0.05)	(0.25)	(0.18)	(0.12)	(0.19)	(0.07)	(0.19)
FC_t		0.28***	0.28**	0.28***	0.06	0.29***	0.25***	0.27***
		(0.02)	(0.12)	(0.09)	(0.06)	(0.09)	(0.04)	(0.09)
σ_t^{10yt}						−0.06		0.01
						(0.05)		(0.09)
$FOMC_t$							0.32***	
							(0.09)	
α	0.37***	0.55***	0.55***	0.55***	0.22***	0.56***	0.13***	0.55***
	(0.01)	(0.02)	(0.13)	(0.09)	(0.06)	(0.10)	(0.03)	(0.09)
Std. Err.	Robust	Robust	HAC	NW	NW	NW	NW	NW
Observations	3,896	3,892	3,892	3,892	3,892	3,892	3,892	3,892
Adjusted R^2	0.04	0.07	0.07	0.07	0.01	0.08	0.22	0.07
F Statistic	169.28***	106.05***	106.05***	106.05***	8.40***	81.17***	277.42***	

can drive realized gamma below zero. This can be seen, for example, around September 2019 when short-term interest rates rose and the 10-year Treasury yield fell. Similarly, our findings chime with the broad consensus that an upward shift of the entire yield curve is generally good for life insurers. For example, realized gamma remained statistically close to zero during the rapid rise in short-term interest rates that occurred as the Federal Reserve tightened monetary policy in 2022. Our measure suggests that market participants focused on the positive effect on life insurers' profitability from rising long-term interest rates and widening spreads on long-term investment grade bonds. In summary, our realized gamma measure of stock price sensitivity to long-term interest rates serves as a useful barometer for market sentiment about the effectiveness of insurers' interest rate risk management.

3.4.2 Is realized gamma low due to interest rate volatility?

Column 6 of Table 4 shows that the daily realized volatility of 10-year Treasury security returns is not correlated with the statistical significance of realized gamma. This finding should be intuitive, as we are estimating realized gamma conditional on intraday 10-year Treasury security returns, but is important to emphasize: It means life insurers' interest rate risk management is generally effective *not* as a consequence of generally low interest rate volatility. In this section, we provide further evidence for this

key result.

We provide additional tests as we recognize the potential endogeneity of long-term interest rate volatility and realized gamma. Life insurers are important investors in the long-term debt market, as we noted above. Their willingness to lend at long terms may simultaneously affect their own sensitivity to long-term interest rates and long-term interest rate volatility. We address the potential endogeneity with a source of plausibly exogenous variation in long-term interest rate volatility.

Scheduled Federal Open Market Committee (FOMC) meeting days are a well-known source of volatility in interest rates, that is sometimes used as a exogenous source of variation (Rigobon and Sack, 2004; Foley-Fisher and Guimaraes, 2013).[20] FOMC meeting days are exogenous to the supply-side variables that give rise to endogeneity concern in our setting. We exploit this source of exogenous variation in two ways. First, we use scheduled FOMC meeting days as an instrumental variable (IV) to obtain exogenous variation in long-term interest rate volatility. Second, we test the difference in means between realized gamma on scheduled FOMC meeting days and on other days when long-term interest rate volatility is lower.

Our first test using an IV is reported in columns 7 and 8 of Table 4, where we show the results from estimating a two-stage least squares

[20]Note that monetary policy shocks are the root cause of the exogenous increase in interest rate volatility, but we do not need to identify the size of those shocks to implement our tests.

regression specification. The IV for the endogenous interest rate volatility variable (σ_t^{10yt}) is a dummy variable ($\boldsymbol{FOMC}_t$) that takes the value 1 on days when the FOMC holds a scheduled meeting and the value 0 otherwise. The first stage, reported in column 7, shows that the FOMC variable is a strong instrument for σ_t^{10yt}. The coefficient estimate is highly statistically significant and positive, consistent with rising interest rate volatility on days with scheduled FOMC meetings. The F-statistic for the first stage regression is 277.4, indicating a strong IV. The second stage, which includes the fitted values from the first stage as a right-hand side variable to replace σ_t^{10yt}, is reported in column 8. The coefficient on long-term interest rate volatility remains statistically insignificant.

Our second test addresses two limitations of our IV approach: (1) other right-hand side variables in our specification may be invalid as instruments in the first stage, and (2) our left-hand side variable is a dummy variable for the statistical significance of realized gamma. We focus on the statistical property of the average difference in realized gammas between high-volatility days when the FOMC has its scheduled meetings and low-volatility days just before the FOMC meetings. Specifically, in our data sample we have 125 FOMC meeting days from May 2007 to December 2022. We pair these days with two alternative low-volatility samples: (1) the days that are one day before the scheduled FOMC meetings, and (2) the days that are one week before the scheduled FOMC meetings. Our null

hypothesis (H_0) is that the average difference between the paired high-volatility realized gammas and low volatility realized gammas is zero.[21] We implement this test using the sub-sampling approach, which does not require making strong assumptions about the unknown distribution of realized gammas or estimating the sample mean variances. All that is required to obtain an asymptotically valid test is that the sampling distribution of the difference in paired realized gammas converges to some unknown distribution and that each *pair* of realized gammas is independent and identically distributed. The former is an extremely weak condition and the latter is natural as we estimate realized gamma using the ratio of daily realized covariances.

The results are reported in Table 5. Columns 1 and 2 show the 99-percent confidence intervals obtained by sub-sampling for the average difference in paired realized gammas for life insurers and P&C insurers, respectively. The test rejects H_0 when the confidence intervals do not contain zero. The first row of the table shows the results when the FOMC meeting days are paired with one-day earlier days. The second row of the table shows the results from pairing FOMC meeting days with one-week earlier days. In both rows, the confidence intervals contain zero and we cannot reject the null hypothesis that the paired realized gammas are the same. For comparison, column 3 reports the confidence intervals

[21] In addition to calculating the paired difference, we also tested the difference between the average realized gamma on scheduled FOMC meeting days and non-meeting days.

from testing the difference in 10-year Treasury security realized volatility between paired days. Column 3 shows that there was a statistically significant increase in σ_t^{10yt}, as should be expected.

Table 5: Comparing realized gammas on days with high and low interest rate volatility. We test the statistical significance of the average difference in realized gammas between high-volatility days when the FOMC has its scheduled meetings and low-volatility days just before the FOMC meetings. Column 1 reports the test of life insurer gammas. Column 2 reports the test of P&C insurer gammas. Column 3 reports the test of daily realized volatility of 10-year Treasury security returns, multiplied by $1e+4$ for legibility. The first row pairs the scheduled FOMC meeting days with one-day earlier days. The second row pairs the scheduled FOMC meeting days with one-week earlier days. The sub-sampled confidence intervals are calculated using 10,000 combinations of 15 paired dates. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices and the St Louis Fed's FRASER database.

	99% confidence interval		
	Life insurers	P&C insurers	10yr Treasury
FOMC days vs. 1 day before	[-0.072, 0.102]	[-0.027, 0.086]	[0.23, 0.606]
FOMC days vs. 7 days before	[-0.063, 0.096]	[-0.038, 0.074]	[0.271, 0.629]

In summary, the additional tests we implemented to address the potential endogeneity of realized gamma and σ_t^{10yt} underscore that low interest rate volatility is not the reason for our finding that life insurers' interest rate risk hedging is generally effective i.e., that realized gamma is generally statistically insignificant.

4 Concluding remarks

In this paper, we introduced a new method to measure the time-varying *residual* interest rate risk of financial intermediaries after they have executed their risk management strategies. Our estimates are daily partial correlations obtained using a nonparametric approach on high-frequency financial market data. We then showed how to conduct statistical inference on our estimates by calculating confidence intervals that are asymptotically valid under extremely weak conditions. Our method can be adapted to include additional variables in the regression model that underpins our framework. Another potential future extension would be to allow for 'jumps' when estimating realized variances and covariances (Andersen, Bollerslev and Diebold, 2007).

Our measure can be used to evaluate the interest rate risk vulnerabilities of any financial intermediary with high-frequency stock prices almost in real time, which is a useful tool for market analysts, supervisors, and policymakers. We applied our method to life insurers, whose exposure to interest rate risk has received less attention than, for example, banks. In doing so, we offered an alternative to the potentially misleading low-frequency OLS estimates that are prevalent in the existing literature. We find that life insurers are generally well-hedged against long-term interest rate movements. That said, they are more sensitive to changes in long-

term interest rates than P&C insurers. We then showed that a measure of the term premium helps to explain the difference in estimated sensitivities between the two types of insurer. Lastly, we provided evidence that our finding that insurers are generally well hedged against interest rate risk is *not* because long-term interest rate volatility is low. Comparing these results with those of other financial intermediaries, such as banks, is another avenue for further research.

References

Adrian, Tobias, Richard Crump, and Emanuel Moench, "Pricing the term structure with linear regressions," *Journal of Financial Economics*, 2013, *110* (1), 110–138.

Aït-Sahalia, Yacine and Jialin Yu, "High frequency market microstructure noise estimates and liquidity measures," *The Annals of Applied Statistics*, 2009, *3* (1), 422–457.

_ **and Per Mykland**, "Estimating Volatility in the Presence of Market Microstructure Noise: A Review of the Theory and Practical Considerations," in Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen, eds., *Handbook of Financial Time Series*, Springer Berlin Heidelberg, January 2009, chapter 25, pp. 577–598.

Allen, Franklin, "Estimating Divisional Cost of Capital for Insurance Companies," in J. David Cummins and Joan Lamm-Tennant, eds., *Financial Management of Life Insurance Companies*, Springer Netherlands, 1993, pp. 101–123.

Andersen, Torben, Tim Bollerslev, and Francis Diebold, "Roughing it up: Including jump components in the measurement, modeling, and

forecasting of return volatility," *The Review of Economics and Statistics*, 2007, *89* (4), 701–720.

_ , _ , and Nour Meddahi, "Analytical Evaluation of Volatility Forecasts," *International Economic Review*, 2004, *45* (4), 1079–1110.

_ , _ , Francis Diebold, and Ginger Wu, "Realized beta: Persistence and predictability," in "Econometric Analysis of Financial and Economic Time Series," Emerald Group Publishing Limited, 2006.

Babbel, David and Robert Stricker, "Asset/Liability Management for Insurers," *Insurance Perspectives*, May 1987.

Bandi, Federico M., Aleksey Kolokolov, Davide Pirino, and Roberto Renò, "Zeros," *Management Science*, 2020, *66* (8), 3466–3479.

Barndorff-Nielsen, Ole and Neil Shephard, "Econometric Analysis of Realized Covariation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics," *Econometrica*, 2004, *72* (3), 885–925.

_ , Peter Reinhard Hansen, Asger Lunde, and Neil Shephard, "Multivariate realised kernels: Consistent positive semi-definite estimators of the covariation of equity prices with noise and non-synchronous trading," *Journal of Econometrics*, 2011, *162* (2), 149–169.

Begenau, Juliane, Monika Piazzesi, and Martin Schneider, "Banks' Risk Exposures," *NBER Working Paper 21334*, 2015.

Berends, Kyal, Robert McMenamin, Thanases Plestis, and Richard J Rosen, "The sensitivity of life insurance firms to interest rate changes," *Economic Perspectives*, 2013, *37* (2).

Brainard, Lael, "Global Financial Stability Considerations for Monetary Policy in a High-Inflation Environment," *Speech at "Financial Stability Considerations for Monetary Policy," a research conference organized by the Board of Governors of the Federal Reserve System and the Federal Reserve Bank of New York*, 2022.

Brewer III, Elijah, James Carson, Elyas Elyasiani, Iqbal Mansur, and William Scott, "Interest Rate Risk and Equity Values of Life Insurance Companies: A GARCH-M Model," *Journal of Risk and Insurance*, 2007, *74* (2), 401–423.

_ **, Thomas Mondschean, and Philip Strahan**, "Why the life insurance industry did not face an" S&L-type" crisis," *Economic Perspectives*, 1993, *17* (5), 12–24.

Briys, Eric and François de Varenne, "On the Risk of Insurance Liabilities: Debunking Some Common Pitfalls," *The Journal of Risk and Insurance*, 1997, *64* (4), 673–694.

_ **and** _ , "Insurance: from underwriting to derivatives," *Springer US*, 2001.

Brunnermeier, Markus and Yuliy Sannikov, "A macroeconomic model with a financial sector," *American Economic Review*, 2014, *104* (2), 379–421.

Christensen, Kim, Silja Kinnebrock, and Mark Podolskij, "Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data," *Journal of Econometrics*, 2010, *159* (1), 116–133.

Corsi, Fulvio, "A Simple Approximate Long-Memory Model of Realized Volatility," *Journal of Financial Econometrics*, 02 2009, *7* (2), 174–196.

Davidson, Russell and James MacKinnon, *Estimation and inference in econometrics*, Oxford New York, 1993.

Doffou, Ako, "New Perspectives in Asset-Liability Management for Insurers," *Journal of Business and Behavioral Sciences*, 2005, *12* (2).

Domanski, Dietrich, Hyun Song Shin, and Vladyslav Sushko, "The hunt for duration: not waving but drowning?," *IMF Economic Review*, 2017, *65* (1), 113–153.

English, William, Skander Van den Heuvel, and Egon Zakrajšek, "Interest rate risk and bank equity valuations," *Journal of Monetary Economics*, 2018, *98*, 80–97.

Fama, Eugene and Kenneth French, "The Capital Asset Pricing Model: Theory and Evidence," *Journal of Economic Perspectives*, September 2004, *18* (3), 25–46.

Fama, Eugene F. and G.William Schwert, "Asset returns and inflation," *Journal of Financial Economics*, 1977, *5* (2), 115–146.

Flannery, Mark and Christopher James, "The Effect of Interest Rate Changes on the Common Stock Returns of Financial Institutions," *The Journal of Finance*, 1984, *39* (4), 1141–1153.

Foley-Fisher, Nathan and Bernardo Guimaraes, "U.S. Real Interest Rates and Default Risk in Emerging Economies," *Journal of Money, Credit and Banking*, 2013, *45* (5), 967–975.

_ , Borghan Narajabad, and Stéphane Verani, "Securities lending as wholesale funding: Evidence from the us life insurance industry," Technical Report, National Bureau of Economic Research 2016.

_ , _ , and _ , "Self-fulfilling runs: Evidence from the us life insurance industry," *Journal of Political Economy*, 2020, *128* (9), 3520–3569.

Froot, Kenneth and Jeremy Stein, "Risk management, capital budgeting, and capital structure policy for financial institutions: an integrated approach," *Journal of Financial Economics*, 1998, *47* (1), 55–82.

_ , David Scharfstein, and Jeremy Stein, "Risk management: Coordinating corporate investment and financing policies," *Journal of Finance*, 1993, *48* (5), 1629–1658.

Gajek, Leslaw and Krzysztof Ostaszewski, *Financial risk management for pension plans*, Elsevier, 2004.

Gissler, Stefan, Nathan Foley-Fisher, and Stéphane Verani, "Over-the-Counter Market Liquidity and Securities Lending," *Review of Economic Dynamics*, 2019.

Hamilton, James, "Macroeconomics and ARCH," *NBER Working Paper 14151*, 2008.

Hansen, Peter, Asger Lunde, and Valeri Voev, "Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility," *Journal of Applied Econometrics*, 2014, *29* (5), 774–799.

Hartley, Daniel, Anna Paulson, and Richard Rosen, "Measuring interest rate risk in the life insurance sector," *The economics, regulation, and systemic risk of insurance markets*, 2016, p. 124.

Hoffmann, Peter, Sam Langfield, Federico Pierobon, and Guillaume Vuillemey, "Who Bears Interest Rate Risk?," *The Review of Financial Studies*, 11 2018, *32* (8), 2921–2954.

Holmstrom, Bengt and Jean Tirole, "Financial intermediation, loanable funds, and the real sector," *the Quarterly Journal of economics*, 1997, *112* (3), 663–691.

Huber, Maximilian, "Regulation-Induced Interest Rate Risk Exposure," *Working Paper*, 2022.

Kirti, Divya, "When gambling for resurrection is too risky," *IMF Working Paper No. 2017/180*, 2017.

Koijen, Ralph and Motohiro Yogo, "The fragility of market risk insurance," *The Journal of Finance*, 2022, *77* (2), 815–862.

Kolokolov, Aleksey and Roberto Renò, "Jumps or staleness?," *Journal of Business & Economic Statistics*, 2023, *0*, 1–23.

Liu, Lily, Andrew Patton, and Kevin Sheppard, "Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes," *Journal of Econometrics*, 2015, *187* (1), 293–311.

Meddahi, Nour, "A theoretical comparison between integrated and realized volatility," *Journal of Applied Econometrics*, 2002, *17* (5), 479–508.

Möhlmann, Axel, "Interest rate risk of life insurers: Evidence from accounting data," *Financial Management*, 2021, *50* (2), 587–612.

NAIC, "Historical Evolution of Life Insurance," in CIPR, ed., *State of the Life Insurance Industry: Implications of Industry Trends*, Center for Insurance Policy and Research, 2013.

Ozdagli, Ali and Zixuan Wang, "Interest rates and insurance company investment behavior," *Available at SSRN 3479663*, 2019.

Paparoditis, Efstathios and Dimitris Politis, "Resampling and Subsampling for Financial Time Series," in Thomas Mikosch, Jens-Peter Kreiß, Richard A. Davis, and Torben Gustav Andersen, eds., *Handbook of Financial Time Series*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 983–999.

Paul, Pascal, "Banks, maturity transformation, and monetary policy," *Journal of Financial Intermediation*, 2022, p. 101011.

Politis, Dimitris, Joseph Romano, and Michael Wolf, *Subsampling*, Springer-Verlag, 1999.

Redington, F. M., "Review of the Principles of Life-office Valuations," *Journal of the Institute of Actuaries*, 1952, *78*, 286–340.

Rigobon, Roberto and Brian Sack, "The impact of monetary policy on asset prices," *Journal of Monetary Economics*, 2004, *51* (8), 1553–1575.

Schneider, Andrés, "Risk-Sharing and the Term Structure of Interest Rates," *The Journal of Finance*, 2022, *77* (4), 2331–2374.

Sen, Ishita, "Regulatory Limits to Risk Management," *Review of Financial Studies*, 2021, *forthcoming.*

Verani, Stéphane and Pei Cheng Yu, "What's Wrong with Annuity Markets?," *mimeo*, 2021.

Vuillemey, Guillaume, "Bank Interest Rate Risk Management," *Management Science*, 2019, *65* (12), 5933–5956.

Zhang, Lan, Per Mykland, and Yacine Aït-Sahalia, "A tale of two time scales: Determining integrated volatility with noisy high-frequency data," *Journal of the American Statistical Association*, 2005, *100* (472), 1394–1411.

Appendix for online publication

A How large is the rolling window bias?

In this appendix, we demonstrate the size of the bias from estimating the two-variable regression model on a rolling window of daily data for insurance companies. We start from the specification:

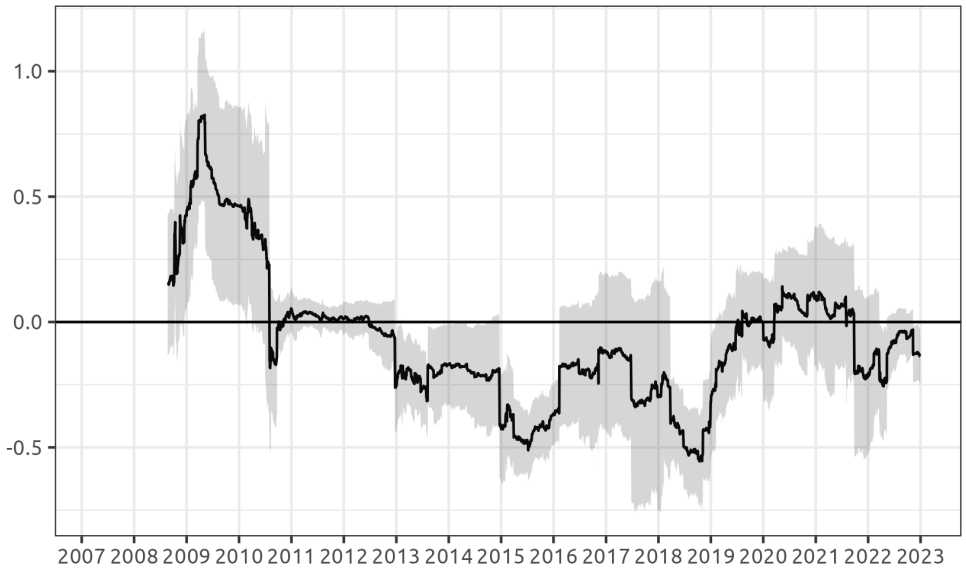
$$r_{i,t} = \alpha + \beta r_{m,t} + \gamma r_{y10,t} + \epsilon_{i,t}$$

where $r_{i,t}$ is the stock price return on the index of life insurers (described in section 3.2) on day t, $r_{m,t}$ is the return on the benchmark S&P500, and $r_{y10,t}$ is the return on the 10-year Treasury security. The γ coefficient in this specification is termed *rolling gamma* and is a low-frequency counterpart to the realized gamma described in section 2 of the main paper.

Selecting the size for the rolling window is typically framed as a tradeoff between (i) including more data to reduce standard errors and (ii) being forced to assume the parameter is stable within the window (Robertson, 2018). We follow the standard approach in the empirical literature estimating interest rate risk for life insurers, and assume a rolling window of two years (Sen, 2021; Huber, 2022).

Figure 4 shows the time series of rolling gammas. The shaded region indicates the heteroskedasticity-corrected 90 percent confidence interval for

Figure 4: Rolling window regression results. The black line shows the rolling gamma estimates using end-of-day data and a two-year rolling window. The shaded region indicates the heteroskedasticity-corrected 90 percent confidence interval for the estimates. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices.

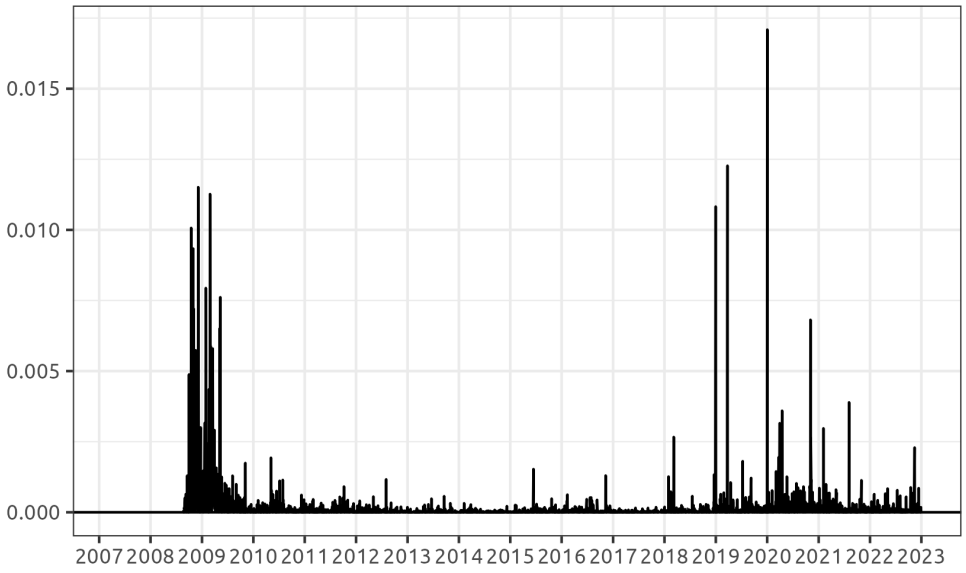


the estimates. There are two main takeaways from the figure. First, the estimates are almost always statistically significant in the post-crisis period. This finding led researchers to conclude that life insurers' risk management became less effective in the aftermath of the GFC and spurred a research agenda to understand the cause of this regime switch—e.g., Sen (2021); Koijen and Yogo (2022); Huber (2022). Second, there are large "jumps" in the time series corresponding to periods of market volatility, such as the beginning of the financial crisis (2008) and the pandemic (2020).

Jumps in the time series hint at a problem of time-varying conditional

volatility in the underlying data. Figure 5 shows the problem by plotting the square of the residuals $(r_{i,t} - \hat{r}_{i,t})^2$. Volatility clustering, which is clearly present in our data, is a long-known empirical feature of financial time series (Bollerslev, Chou and Kroner, 1992).

Figure 5: Squared residuals from rolling regression. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices.



The potential effects of conditional heteroskedasticity for OLS regressions are well known. In some applications, such as when the primary concern is estimating the conditional mean, a common view is that inference can be made using the standard corrections proposed by White (1980) or Newey and West (1987). However, as Hamilton (2008) points out, misspecifying the errors will produce inefficient estimates and incorrect

inference. The specific case of the rolling window OLS estimator was studied by Cai and Juhl (2021), who showed that a bias can exist even asymptotically with well-behaved errors. Intuitively, the rolling window OLS estimates are weighted averages of the time-varying parameter and the weights depend on the time-varying volatility. The asymptotic bias arises when the two time series (parameter and volatility) are correlated. In simulations, the rolling window OLS estimates are often unstable and the bias can be substantial (Robertson, 2018).

One solution to the problem is to assume some structure for the variance processes. By explicitly modeling the heteroscedasticity in the variance-covariance matrix, we address the bias in the time series of parameter estimates and gain efficiency. A typical approach in financial econometrics is to appeal to autoregressive conditional heteroskedasticity (ARCH) models (Bollerslev, Engle and Nelson, 1994). This class of flexible models and its wide range of extensions are straightforward to implement in off-the-shelf statistical packages. In practice, the generalized ARCH, or 'GARCH', model that allows for greater serial dependence in the error term is an extremely common choice. The conditional variance of the process for a GARCH(r, p) is given by:

$$Var(\epsilon_t|\Omega_{t-1}) = h_t = a_0 + a_1\epsilon_{t-1}^2 + a_2\epsilon_{t-2}^2 + \cdots + a_p\epsilon_{t-p}^2 +$$

$$b_1 h_{t-1} + b_2 h_{t-2} + \cdots + b_r h_{t-r}.$$

As an exercise to gauge the size of the rolling window OLS bias in the estimates reported in Figure 4, we assume that our three time series of daily returns follow a multivariate GARCH(1,1) process. We specify the joint process:

$$r_{i,t} = \alpha_i + u_{i,t}$$

$$r_{m,t} = \alpha_m + u_{m,t}$$

$$r_{y10,t} = \alpha_{y10} + u_{y10,t}$$

so

$$E\left[\begin{pmatrix} r_{i,t} \\ r_{m,t} \\ r_{y10,t} \end{pmatrix} \mid \Omega_{t-1}\right] = \begin{pmatrix} \alpha_i \\ \alpha_m \\ \alpha_{y10} \end{pmatrix}$$

and

$$Var\left[\begin{pmatrix} r_{i,t} \\ r_{m,t} \\ r_{y10,t} \end{pmatrix} \mid \Omega_{t-1}\right] = Var\left[\begin{pmatrix} u_{i,t} \\ u_{m,t} \\ u_{y10,t} \end{pmatrix} \mid \Omega_{t-1}\right] = \begin{pmatrix} \sigma^2_{11,t} & \sigma_{12,t} & \sigma_{13,t} \\ \sigma_{21,t} & \sigma^2_{22,t} & \sigma_{23,t} \\ \sigma_{31,t} & \sigma_{32,t} & \sigma^2_{33,t} \end{pmatrix}.$$

The last matrix—known as the dynamic conditional variance-covariance

matrix—can be used to form the dynamic conditional ratio:

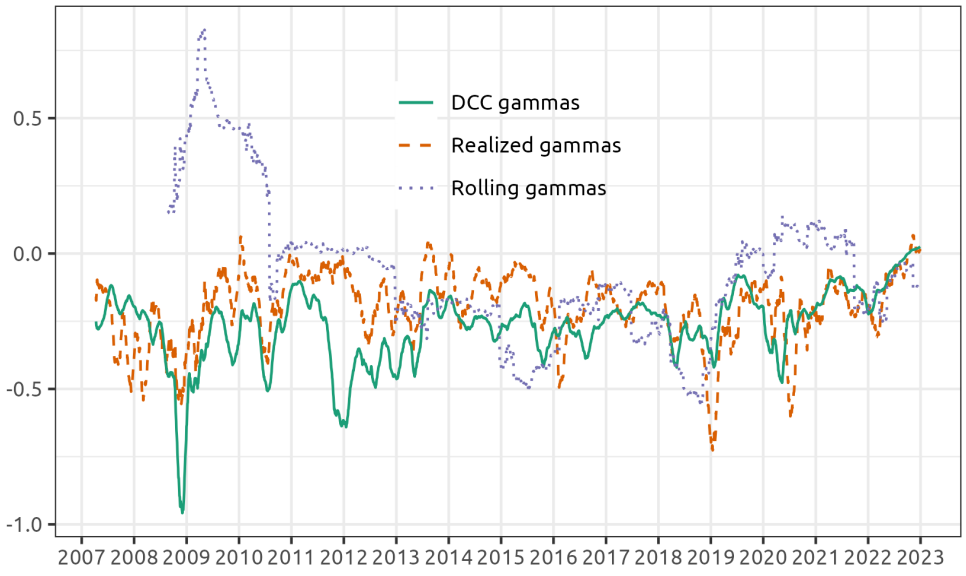
$$\gamma_t^{DC} = \frac{\sigma_{13,t}}{\sigma_{33,t}^2},$$

which is obtained after specifying GARCH(1,1) processes for each second moment. We call the ratio γ_t^{DC} the *dynamic conditional gamma*, following the literature that uses the same technique to estimate dynamic conditional betas (Engle, 2016).

Figure 6 compares the different estimates of gamma. The blue dotted line shows the rolling gamma estimates, while the green solid line shows the dynamic conditional gamma estimates. The difference between the two estimates is particularly striking during periods of high volatility, such as the financial crisis and the global pandemic, revealing that the rolling gamma is highly biased and misleading. For completeness, we include the realized gamma estimates as the brown dashed line in the figure. The relative proximity of the realized gamma estimates and the dynamic conditional gamma estimates during those periods of stress is a reassuring sign that both approaches are solving the underlying problem of conditional heteroscedasticity. Note that dynamic conditional gamma and realized gamma use completely different data and approaches to address the same underlying problem.

Although dynamic conditional gamma and realized gamma deliver

Figure 6: Comparing rolling gamma, dynamic conditional gamma, and realized gamma. The figure shows three different estimates of the sensitivity of life insurers' stock prices to changes in interest rates. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices.



similar parameter estimates, they are not the same. In particular, the empirical approach that underpins the dynamic conditional gamma is known to suffer from substantial limitations (Caporin and McAleer, 2013). As a stated data representation—rather than derived model—the dynamic conditional gamma has no moments or desirable asymptotic properties. It serves our purposes as a diagnostic tool that reveals a huge bias in rolling gamma. But to avoid reliance on the imposed structure and—most importantly—to conduct valid inference, we strongly prefer the empirical approach that uses realized variances and covariances in our paper.

B Hypothesis testing using the subsampling method

In each day $t \in \{1, \ldots, T\}$, we estimate γ_t using the following linear regression model

$$\tilde{r}_{i,j,s} = \alpha_t + \beta_t \tilde{r}_{m,j,s} + \gamma_t \tilde{r}_{y10,j,s} + \epsilon_{i,j,s}$$

on a sample of $n = 388$ observations corresponding to each of the 388 trading minutes for day t between 9:31am and 3:59pm indexed by s. We calculate a confidence interval for each of the daily γ_t using subsampling following Politis et al. (1999) under the assumption that the errors are *asymptotically*

stationary. Asymptotic stationarity means that, for example, the errors could follow an AR(1) process with autocorrelation parameter strictly less than one and heteroskedastic innovations.

To simplify the exposition of subsampling, we rewrite the linear regression model in matrix form as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ and $\boldsymbol{\epsilon}$ are $n \times 1$ vectors, $\boldsymbol{\beta}$ is a $p \times 1$ vector which includes γ_t as an element and $\mathbf{X}$ is an $n \times p$ matrix of five-minute returns and a constant. The estimator of $\boldsymbol{\beta}$ based on $\mathbf{X}$ and $\mathbf{y}$ is given by $\hat{\boldsymbol{\beta}} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$.

For any $b < n$ such that $b > p$, define the subvectors and submatrices

$$\mathbf{y}_{b,s} \equiv (y_s, \ldots, y_{s+b-1})' \text{ , } \boldsymbol{\epsilon}_{b,s} \equiv (\epsilon_s, \ldots, \epsilon_{s+b-1})' \text{ and} \tag{5}$$

$$\mathbf{X}_{b,s} \equiv \begin{pmatrix} \boldsymbol{x}_s' \\ \vdots \\ \boldsymbol{x}_{s+b-1}' \end{pmatrix}, \text{ where } \mathbf{X} \equiv \begin{pmatrix} \boldsymbol{x}_1' \\ \vdots \\ \boldsymbol{x}_n' \end{pmatrix} \tag{6}$$

The estimator of $\boldsymbol{\beta}$ based on $\mathbf{X}_{b,s}$ and $\mathbf{y}_{b,s}$ is given by

$$\hat{\boldsymbol{\beta}}_{n,b,s} \equiv (\mathbf{X}_{b,s}'\mathbf{X}_{b,s})^{-1}\mathbf{X}_{b,s}'\mathbf{y}_{b,s}.$$

Denote by $J_b(P)$ the sampling distribution of the normalized statistic

$\sqrt{b}(\hat{\boldsymbol{\beta}}_{n,b,s}-\boldsymbol{\beta})$, where P is the probability law governing the estimator $\hat{\boldsymbol{\beta}}_{n,b,s}$, which is unknown. For any Borel set $A \in \mathbb{R}^p$, let

$$J_b(A,P) = Prob_P\{\sqrt{b}(\hat{\boldsymbol{\beta}}_{n,b,s} - \boldsymbol{\beta}) \in A\}.$$

The approximation to $J_n(A,P)$ is defined by

$$L_{n,b}(A) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{\sqrt{b}(\hat{\boldsymbol{\beta}}_{n,b,s} - \hat{\boldsymbol{\beta}}) \in A\}.$$

Therefore, subsampling consist of evaluating a statistics on an exhaustive set of subsamples of size $b < n$ that are created from the original sample of size n and estimating the distribution of this statistics normalized by $\sqrt{b}$. As should be clear, each subsample contains consecutive observations from the original time series sample. Therefore, each subsample is drawn from the true data generating process.

In what follows we summarize the main result from subsampling related to the estimation of a daily γ_t using intra-day time series observation. Note that our limiting concept is that the number of equally spaced intraday returns approaches infinity. We refer the readers to Politis et al. (1999) for details and proofs.

Assumption 1 *There exists a limiting law $J(P)$ such that*

1. *$J_n(P)$ converges weakly to $J(P)$ as $n \to \infty$. This means that for*

any Borel set A whose boundary has mass zero under $J(P)$, we have $J_n(A,P) \to J(A,P)$ as $n \to \infty$.

2. *For every Borel set A whose boundary has mass zero under $J(P)$ and for any index sequence $\{s_b\}$, we have $J_{b,s_b}(A,P) \to J(A)$ as $b \to \infty$.*

Theorem B.1 (Politis et al. (1999) Theorem 4.3.1) *Let $\{(\mathbf{x}_s, \epsilon_s)\}$ be a sequence of random vectors defined on a common probability space. Denote the mixing coefficients for the $\{(\mathbf{x}_s, \epsilon_s)\}$ sequence by $\alpha(\cdot)$. Define*

$$T_{k,s} \equiv \frac{1}{\sqrt{k}} \sum_{a=s}^{s+k-1} \mathbf{x}_a \epsilon_a \ , \ V_{k,s} \equiv Cov(T_{k,s}) \ , \ and \ M_{k,s} \equiv E(\mathbf{X}'_{k,s}\mathbf{X}_{k,s}/k).$$

Assume the following conditions hold. For some $\delta > 0$,

- $E(\mathbf{x}_s \epsilon_s) = 0$ *for all* s,
- $E|\mathbf{x}_{s,j}\epsilon_s|^{2+2\delta} \leq \Delta_1$ *for all* s *and all* $1 \leq j \leq p$,
- $E|\mathbf{x}_{s,j}|^{4+2\delta} \leq \Delta_2$ *for all* s *and all* $1 \leq j \leq p$,
- $V_{k,s} \to V > 0$ *uniformly in* s *as* $k \to \infty$,
- $M_{k,s} \to M > 0$ *uniformly in* s *as* $k \to \infty$,
- $C(4) \equiv \sum_{k=1}^{\infty} (k+1)^2 \alpha^{\frac{\delta}{4+\delta}}(k) \leq K$.

Furthermore, assume that $b/n \to 0$ and $b \to \infty$ as $n \to \infty$. Letting $J(P) = N(0, M^{-1}VM^{-1})$. Then:

i. $L_{n,b}(A) \to J(A,P)$ *in probability for each Borel set* A *whose boundary has mass zero under* $J(P)$.

ii. Let Z *be a random vector with* $\mathcal{L}(Z) = J(Z)$. *For a norm* $||\cot||$ *on* $\mathbb{R}^k$, *define univariate distributions* $L_{n,||\cot||}$ *and* $J_{||\cot||}(P)$ *in the following way:*

$$L_{n,b,||\cot||}(x) = \frac{1}{n-b+1}\sum_{s=1}^{n-b+1} 1\{||\sqrt{b}(\hat{\boldsymbol{\beta}}_{n,b,s} - \hat{\boldsymbol{\beta}}||) \le x\}$$

$$J_{||\cot||}(x,P) = Prob\{||Z|| \le x\}.$$

For $\alpha \in (0,1)$, *let*

$$c_{n,b,||\cdot||}(1-\alpha) = \inf\{x : L_{n,b,||\cdot||}(x) \ge 1-\alpha\}.$$

Correspondingly, define

$$c_{||\cdot||}(1-\alpha,P) = \inf\{x : J_{||\cdot||}(x,P) \ge 1-\alpha\}.$$

If $J_{||\cdot||}(\cdot,P)$ *is continuous at* $c_{||\cdot||}(1-\alpha,P)$ *then*

$$Prob_P\{||\sqrt{b}(\hat{\boldsymbol{\beta}}_{n,b,s} - \hat{\boldsymbol{\beta}}|| \le c_{n,b,||\cdot||}(1-\alpha)\} \to 1-\alpha \text{ as } n \to \infty.$$

Thus, the asymptotic coverage probability under $\boldsymbol{P}$ *of the region*

$$\{\boldsymbol{\beta} : ||\sqrt{b}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}|| \leq c_{n,b,||\cdot||}(1-\alpha)\}$$

is the nominal level $1-\alpha$*.*

Theorem B.1 shows that we can derive asymptotically valid confidence intervals for the daily estimator $\hat{\boldsymbol{\beta}}$ using $L_{n,b}(A)$ because it is a consistent estimator of $J(A,P)$. By exploiting the usual duality between the construction of a confidence interval for γ_t and the construction of a hypothesis test about γ_t, subsampling allows us to make asymptotically valid inference about the true γ_t. In our application, we wish to test the null hypothesis that the daily γ_t is statistically different from 0. Under the null, insurers are hedged against interest rate risk as their stock price is not sensitive to movement in the ten-year treasury rate. If the value of the estimated daily γ_t falls outside the daily confidence interval, we reject the null hypothesis on that day.

Subsampling is not as well known as the bootstrap method in economics and finance, which warrants a cursory comparison—see Politis et al. (1999) for textbook-length treatment. The most relevant bootstrap method for our time series application is the so-called Moving Blocks Bootstrap (MBB). As with subsampling, MBB breaks down the original time series to smaller blocks of consecutive observations, which preserves the serial correlation

structure within each block. Practically, the main difference is that MBB draws samples with replacement from the blocks and connects the sampled blocks together to form a bootstrap sample of size n. Therefore, by construction, MBB imposes the assumption that blocks of an arbitrary size b are uncorrelated. This assumption about the unknown data generating process is rather strong and likely to be violated in our application. From a technical point view, the bootstrap method requires that the distribution of the statistic of interest be locally smooth as a function of the unknown model. Establishing this result, even if it is indeed true, would be non-trivial. With subsampling, we do not need to make these assumptions or verify the smoothness of the distribution under the true model to draw asymptotically valid inferences. All that is required is that our normalized statistic has a limit distribution under the true model.

C Summary statistics for Section 3.4

Table 6: Summary statistics. This table reports summary statistics for the variables used to analyze the determinants of the significance of realized gamma. $\gamma_{i,t}$ is realized gamma for insurer type $i \in$ {Life, P&C}. The binary variable $\mathbb{1}(\gamma_{i,t}<0)$ takes the value 1 when realized gamma for insurer type i is statistically significant and 0 otherwise. TP_t is the term premium estimate from Adrian et al. (2013), $Baa - Aaa_t$ is the Moody's Baa-Aaa seasoned corporate spread, and FC_t is the ICE BoA Single-A US corporate index option-adjusted spread. σ_t^{Life} is the realized volatility of the intraday returns of life insurers. σ_t^{10yt} is the realized volatility of the intraday returns on 10-year Treasury. The statistics for σ_t^{Life} and σ_t^{10yt} are multiplied by 1e+4 for legibility. Source: Authors' calculations based on data from Refinitiv, the Center for Research in Security Prices, FRED, and Adrian et al. (2013).

Variable	No. obs.	Mean	Median	Std. Dev.	p25	p75
$\gamma_{\text{Life},t}$	3,901	-0.19	-0.16	0.31	-0.34	-0.02
$\mathbb{1}(\gamma_{\text{Life},t}<0)$	3,923	0.32	0	0.47	0	1
$\gamma_{P\&C,t}$	3,901	-0.12	-0.09	0.22	-0.21	0.01
$\mathbb{1}(\gamma_{\text{P\&C},t}<0)$	3,923	0.18	0	0.38	0	0
TP_t	3,896	0.54	0.30	1.10	-0.33	1.51
$Baa - Aaa_t$	3,897	1.08	0.96	0.47	0.82	1.19
FC_t	3,921	1.49	1.17	1.01	0.94	1.64
σ_t^{10yt}	3,923	0.24	0.16	0.31	0.10	0.28

D Data citations

- Refinitiv
- Center for Research in Security Prices, CRSP 1925 US Indices Database, Wharton Research Data Services, `http://www.whartonwrds.com/datasets/crsp/`
- FRED API, accessed using third party R software package fredr, `https://fred.stlouisfed.org/docs/api/fred/`
  - 'BAA' — Moody's, Moody's Seasoned Baa Corporate Bond Yield [BAA], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/BAA`
  - 'AAA' — Moody's, Moody's Seasoned Aaa Corporate Bond Yield [AAA], retrieved from FRED, Federal Reserve Bank of St. Louis; `https://fred.stlouisfed.org/series/AAA`

Online appendix references

Bollerslev, Tim, Ray Chou, and Kenneth Kroner, "ARCH modeling in finance: A review of the theory and empirical evidence," *Journal of Econometrics*, 1992, *52* (1), 5–59.

_ , Robert Engle, and Daniel Nelson, "ARCH Models," in "Handbook of Econometrics," Vol. 4, Elsevier, 1994, pp. 2959–3038.

Cai, Zongwu and Ted Juhl, "The Distribution Of Rolling Regression Estimators," *mimeo*, 2021.

Caporin, Massimiliano and Michael McAleer, "Ten things you should know about the dynamic conditional correlation representation," *Econometrics*, 2013, *1* (1), 115–126.

Engle, Robert, "Dynamic conditional beta," *Journal of Financial Econometrics*, 2016, *14* (4), 643–667.

Hamilton, James, "Macroeconomics and ARCH," *NBER Working Paper 14151*, 2008.

Huber, Maximilian, "Regulation-Induced Interest Rate Risk Exposure," *Working Paper*, 2022.

Koijen, Ralph and Motohiro Yogo, "The fragility of market risk insurance," *The Journal of Finance*, 2022, *77* (2), 815–862.

Newey, Whitney and Kenneth West, "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 1987, *55* (3), 703–708.

Politis, Dimitris, Joseph Romano, and Michael Wolf, *Subsampling*, Springer-Verlag, 1999.

Robertson, Donald, "Estimating β," *mimeo*, 2018.

Sen, Ishita, "Regulatory Limits to Risk Management," *Review of Financial Studies*, 2021, *forthcoming.*

White, Halbert, "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 1980, *48* (4), 817–838.

Fragmentation and optimal liquidity supply on decentralized exchanges

Alfred Lehar
Christine A. Parlour
Marius Zoican*

May 21, 2024

Abstract

We investigate how liquidity providers (LPs) choose between high- and low-fee trading venues, in the face of a fixed common gas cost. Analyzing Uniswap data, we find that high-fee pools attract 58% of liquidity supply yet execute only 21% of volume. Large LPs dominate low-fee pools, frequently adjusting out-of-range positions in response to informed order flow. In contrast, small LPs converge to high-fee pools, accepting lower execution probabilities to mitigate adverse selection and liquidity management costs. Fragmented liquidity dominates a single-fee market, as it encourages more liquidity providers to enter the market, while fostering LP competition on the low-fee pool.

Keywords: FinTech, decentralized exchanges (DEX), liquidity, fragmentation, adverse selection
JEL Codes: G11, G12, G14

*Alfred Lehar (alfred.lehar@haskayne.ucalgary.ca) is affiliated with Haskayne School of Business at University of Calgary. Christine A. Parlour (parlour@berkeley.edu) is with Haas School of Business at UC Berkeley. Marius Zoican (marius.zoican@rotman.utoronto.ca, corresponding author) is affiliated with University of Toronto Mississauga and Rotman School of Management. Corresponding address: 3359 Mississauga Road, Mississauga, Ontario L5L 1C6, Canada. We have greatly benefited from discussion on this research with Michael Brolley, Agostino Capponi (discussant), Itay Goldstein, Sang Rae Kim (discussant), Olga Klein, Katya Malinova (discussant), Ciamac Moallemi, Uday Rajan, Thomas Rivera, Andreas Park, Gideon Saar, Lorenzo Schöenleber (discussant), Andriy Shkilko (discussant), and Shihao Yu. We are grateful to conference participants at the 2024 NYU Stern Market Microstructure Meeting, Tokenomics 2023, Gillmore Centre Annual Conference 2023, Edinburgh Economics of Technology, Financial Intermediation Research Society 2023, the Northern Finance Association 2023, the UNC Junior Faculty Finance Conference, as well as to seminar participants at the University of Chicago, Lehigh, the Microstructure Exchange, UCSB-ECON DeFi Seminar, University of Melbourne, Wilfrid Laurier University, University of Guelph, Rotman School of Management, Hong Kong Baptist University, and Bank of Canada. Marius Zoican gratefully acknowledges funding support from the Rotman School of Managements' FinHub Lab and the Canadian Social Sciences and Humanities Research Council (SSHRC) through an Insight Development research grant (430-2020-00014).

Fragmentation and optimal liquidity supply on decentralized exchanges

Abstract

We investigate how liquidity providers (LPs) choose between high- and low-fee trading venues, in the face of a fixed common gas cost. Analyzing Uniswap data, we find that high-fee pools attract 58% of liquidity supply yet execute only 21% of volume. Large LPs dominate low-fee pools, frequently adjusting out-of-range positions in response to informed order flow. In contrast, small LPs converge to high-fee pools, accepting lower execution probabilities to mitigate adverse selection and liquidity management costs. Fragmented liquidity dominates a single-fee market, as it encourages more liquidity providers to enter the market, while fostering LP competition on the low-fee pool.

Keywords: FinTech, decentralized exchanges (DEX), liquidity, fragmentation, adverse selection
JEL Codes: G11, G12, G14

1 Introduction

In addition to aggregating information, asset markets allow agents to exhaust private gains from trade. While there is a well developed literature on the informativeness of prices, less is known about if and when markets effectively exhaust all gains from trade.[1] In this paper, we exploit the unique design of a decentralized exchange to shed light on the market for liquidity and show, theoretically and empirically, that market fragmentation can improve trading efficiency.

Automated market makers such as Uniswap v3 provide a unique environment to investigate the market for liquidity. While there are various new institutional details that animate these exchanges, for our purposes three are economically important. First, in automated exchanges, liquidity demand and supply can easily be distinguished: users either supply or demand liquidity. Because of this, we can isolate the effect of transactions costs on each side of the market for liquidity. Second, costs and benefits incurred by liquidity suppliers are easier to observe because prices are not set by market participants, but are automatically calculated as a function of liquidity demand and supply. Thus, liquidity suppliers are not compensated through price impact. Third, market participants are pseudo-anonymous so we can identify and document liquidity suppliers at a high frequency. These unique features allow us to investigate, theoretically and empirically, how transactions costs affect liquidity supply.

Beyond investigating the market for liquidity, there are three additional reasons to investigate liquidity provision in AMMs. First, these markets are large and successful in their own right: After its May 2021 launch, Uniswap v3 features daily trading volume in excess of US $1 billion. Second, for major pairs such as Ether against USD stablecoins, Uniswap boasts twice or three times better liquidity than continuous limit order exchanges such as Binance, which suggests that this design can be economically superior.[2] Third, as traditional assets become tokenized, and markets become more automated, this new market form could be adopted.[3]

Uniswap v3 provides two innovations over the previous v2. First, liquidity suppliers and demanders select into trading places (called pools) that differ on transaction fees. Each asset pair to be traded on up to four liquidity pools that only differ in the compensation for liquidity providers: in particular, liquidity fees can be equal to 1, 5, 30, or 100 basis points and the corresponding tick sizes are 1, 10, 60, or 200 basis points. These proportional fees are paid by liquidity demanders and are the only source of remuneration to liquidity providers. (These fees, as we discuss below, are similar to the make-take fees that are prevalent in limit order markets.) Second, on Uniswap

[1] Gains from trade comprise an idiosyncratic private value for the underlying asset, but also idiosyncratic preferences for trade speed or "liquidity."Agents' idiosyncratic value for the underlying asset are plausibly determined by their portfolio positions, and therefore independent of the trading place. By contrast, their idiosyncratic preference for liquidity determines market quality.

[2] See The Dominance of Uniswap v3 Liquidity; May 5, 2022.

[3] Swarm — a BaFin regulated entity — already offers AMM trading for a variety of tokenized Real World Assets.

v3, liquidity providers can submit "concentrated liquidity." Even though their liquidity is passively supplied, they can choose the price range over which it is supplied. With volatile assets, these concentrated liquidity positions can become stale and require rebalancing.

Besides differences in fees, the liquidity pools are otherwise identical and, importantly, they share the common infrastructure of the Ethereum blockchain. Importantly, all participants pay a transaction cost (called a "gas fee") to access the markets. Our theory and empirical work investigates the effect of different proportional fees and fixed fees on liquidity supply. At launch, Uniswap Labs conjectured that trading and liquidity should consolidate in equilibrium on a single "canonical" pool for which the liquidity fee is just enough to compensate the marginal market maker for adverse selection and inventory costs. That is, activity in low-volatility pairs such as stablecoin-to-stablecoin trades should naturally gravitate to low fee liquidity pools, whereas speculative trading in more volatile pairs will consolidate on high fee markets.[4] As we show, this reasoning is flawed.

We present a simple model with trade between liquidity suppliers and two types of liquidity demand. Consistent with the design of v3, liquidity suppliers chose a market and then place their liquidity into a band around the current value of the asset. The posted liquidity is subject to a bonding curve and hence generates a price impact cost for the liquidity demanders (we emphasize that this does not remunerate the liquidity suppliers). Liquidity suppliers have heterogeneous endowments, interpretable as different capital constraints — low-endowment liquidity providers are akin to retail traders, whereas high-endowments stand in for large institutional investors or quantitative funds. Trade occurs against these positions either because a liquidity demander arrives who has experienced a liquidity shock or because the value of the asset has changed and an arbitrageur adversely selects the passive liquidity supply. Collectively, the decisions of the liquidity demanders determine the payoff to the liquidity suppliers. After large private or common value trades, liquidity providers rebalance their positions; to do so, liquidity providers incur a fixed cost (i.e., gas price) each time they update their position.

Traders demanding liquidity face two types of costs: first, the fee associated with their chosen pool (low or high) and second, the price impact costs generated by the pool's bonding curve and supplied liquidity. We find that traders route small orders exclusively to the low-fee pool to obtain the all-in lowest cost. In contrast, large traders split their orders across both low- and high-fee liquidity pools. As a result, low-fee markets are actively traded and require frequent liquidity updates whereas high-fee pools have a longer liquidity update cycle since they absorb fewer trades.

We establish conditions under which there is fragmentation or consolidation. Specifically, even in this simple framework, there is a robust parameter range in which liquidity does not naturally concentrate on one of the exchanges. Both pools can attract a positive market share if liquidity providers face gas fees and the adverse selection costs are sufficiently low. Liquidity providers trade

[4]See *Flexible fees* paragraph at https://uniswap.org/blog/uniswap-v3; accessed September 14, 2022.

off a higher revenue per unit of time in the low-fee pool (driven by the larger trading volume) against higher adverse selection as well as the additional gas cost required for active liquidity management. As a result, liquidity provider clienteles emerge in equilibrium. Liquidity providers with large endowments gravitate towards low-fee markets, as they are best positioned to frequently update their position. In contrast, smaller market makers choose to passively provide liquidity on high-fee markets where they only trade against large orders being routed there. They optimally trade off a lower execution probability against higher fees per unit of volume, reduced adverse selection, as well as a lower liquidity management cost per unit of time.

Not only does liquidity fragment, but it differs in both use and type across the two markets. A small number of highly active large liquidity providers, potentially institutional investors and hedge funds, primarily trade against numerous small incoming trades on pools with low fees. In contrast, high-fee pools involve less frequent trading between a substantial number of capital-constrained passive liquidity providers (e.g., retail market makers) on one side and a few sizeable incoming orders on the other. As the fixed gas fee affects liquidity providers pool choice, changes in the common fixed market access fee differentially affects the liquidity supply on the two pools. Specifically, it reduces market quality (in the sense of lower posted liquidity) on the low fee pool.

As we distinguish between liquidity demanders who are trading to exploit gains from trade and liquidity demanders who are arbitraging common value changes, we can decompose returns to liquidity providers, and show that adverse selection is higher on the low fee pool. Given that the low fee pool is populated with larger liquidity suppliers, this suggests that institutional traders bear price risk.

Our findings indicate that liquidity fragmentation can enhance market quality, as measured by total gains from trade. In a single fee market, a fee that is too low fails to attract liquidity providers with smaller endowments and thus more sensitive to fixed costs, leading to unrealized gains. Conversely, a very high fee results in prohibitively high trading costs and deters trade. A two-pool market with heterogeneous fees offers two instruments to independently manage costs. The higher fee determines the marginal liquidity provider **LP** entering the market, and therefore the realized gains from trade. The lower fee pool, by attracting **LP** with larger endowments, can reduce transaction costs. We demonstrate that a two-pool fee structure can always be designed to yield higher gains from trade than any single-pool arrangement.

Using the model for guidance, we analyze more than 28 million interactions with Uniswap v3 liquidity pools – that is, all liquidity updates and trades from the inception of v3 in May 2021 until July 2023. We first document liquidity fragmentation in 32 out of 242 asset pairs in our sample, which account for 95% of liquidity committed to Uniswap v3 smart contracts and 93% of trading volume. For each of the fragmented pairs, trading consolidates on two pools with adjacent fee levels: either 1 and 5 basis points (e.g., USDC-USDT), 5 and 30 basis points (ETH-USDC), or 30 and 100 basis points (USDC-CRV).

We then document that high-fee pools are on average larger – with aggregate end-of-day liquidity of $46.50 million relative to $33.78 million, the average size of low-fee pools. Nevertheless, three quarters of daily trading volume executes on low-fee pools. In line with the model predictions, low-fee pools are more active as they capture many small trades. There are five times as many trades on low- than on high-fee pools (610 versus 110). However, the average trade on the high fee pool is twice as large: $14,490 relative to $6,340. Unsurprisingly, liquidity cycles – measured as the time between the submission and update of posted liquidity – are 20% shorter on the highly active low-fee pool.

We find robust evidence of liquidity supply clienteles across pools. The average liquidity deposit is 107.5% larger on the low-fee pool, after controlling for daily volume and return volatility. At the same time, high-fee pools' market share is 21 percentage points higher. The results point to an asymmetric match between liquidity supply and demand: large liquidity providers are matched with small liquidity demanders on low-fee pools, whereas small liquidity providers trade with a few large orders on the high-fee pool.

We then turn to the common fixed cost of accessing the market, or gas fees. The market shares of the liquidity pools depend on the magnitude of gas costs on the Ethereum blockchain. In the model, a higher gas price leads to a shift in liquidity supply from the low- to the high-fee pool as active position management becomes relatively more costly for the marginal liquidity provider. We find that a one standard deviation increase in gas prices corresponds to a 4.63 percentage points decrease in the low-fee pool market share, and a 29% drop in liquidity inflows on days when gas costs are elevated.

Consistent with our model, we find that liquidity providers in low-fee pools earn higher fee yields but face increased adverse selection costs. Specifically, the daily fee yield is 2.03 basis points larger on low-fee pools. On the other hand, the permanent price impact as measured by loss-versus-rebalancing (LVR, as in Milionis, Moallemi, Roughgarden, and Zhang, 2023) is 6.39 basis points or 81% greater in low-fee pools compared to high-fee ones. However, despite this difference, the deviations in prices between high- and low-fee pools and those on centralized exchanges do not differ significantly.

Our paper is related to various literatures. Pagano (1989) shows that if an asset is traded on two identical exchanges with equal transaction costs, in equilibrium market participants gravitate to a single exchange due to network effects. In practice, exchanges are rarely identical: fragmentation can emerge between fast and slow exchanges (Pagnotta and Philippon, 2018; Brolley and Cimon, 2020) or between lit and dark markets (Zhu, 2014). In our model, fragmentation on decentralized exchanges is driven by variation in liquidity fees as well as different economies of scale due to heterogeneity in liquidity provider capital. We find that liquidity fragmentation driven by high gas fees implies larger transaction costs on incoming orders. We note that there is no time priority on decentralized exchanges, which clear in a pro rata fashion. On markets with time priority, Foucault and Menkveld (2008) and O'Hara and Ye (2011) find that market segmentation in equity markets

improves liquidity (by allowing queue jumping) and price discovery.

Fixed costs for order submission are uncommon in traditional markets. However, in 2012, the Canadian regulator IIROC implemented an "integrated fee model" that charged traders for all messages sent to Canadian marketplaces. Korajczyk and Murphy (2018) document that this measure disproportionately affected high-frequency traders, resulting in wider bid-ask spreads but lower implementation shortfall for large traders, possibly due to a reduction in back-running activity. Our study contributes additional insights by highlighting that the introduction of a fixed cost, even when applied across exchanges, can lead to market fragmentation.

We also relate to a rich literature on market fragmentation and differential fees. Closest to our paper, Battalio, Corwin, and Jennings (2016) and Cimon (2021) study the trade-off between order execution risk and compensation for liquidity provision in the context of make-take fee exchanges. However, Battalio, Corwin, and Jennings (2016) specifically addresses the issue of the broker-customer agency problem, whereas our study focuses on liquidity providers who trade on their own behalf. In traditional securities markets, make-take fees are contingent on trade execution and proportional to the size of the order. On the other hand, gas costs on decentralized exchanges are independent of order execution, highlighting the significance of economies of scale (lower proportional costs for larger liquidity provision orders) and dynamic liquidity cycles (managing the frequency of fixed cost payments). Strategic brokers in Cimon (2021) provide liquidity alongside exogenous market-makers in a static setting. We complement this approach by modelling network externalities inherent in the coordination problem of heterogeneous liquidity providers. In our dynamic setup, this allows us to pin down the equilibrium duration of liquidity cycles and the relative importance of gas fixed costs.

Our paper relates to a nascent and fast-growing literature on the economics of decentralized exchanges. Many studies (e.g., Aoyagi, 2020; Aoyagi and Ito, 2021; Park, 2022) focus on the economics of constant-function automated market makers, which do not allow liquidity providers to set price limits. In this restrictive environment, Capponi and Jia (2021) argue that market makers have little incentives to update their position upon the arrival of news to avoid adverse selection, since pro-rata clearing gives an advantage to arbitrageurs. Lehar and Parlour (forthcoming) solve for the equilibrium pool size in a setting where liquidity providers fully internalize information costs without rushing to withdraw positions at risk of being sniped. We argue that on exchanges that allow for limit or range orders, the cost of actively managing positions becomes a first-order concern, as liquidity providers need to re-set the price limits once posted liquidity no longer earns fees. Our empirical result on economies of scale echoes the argument in Barbon and Ranaldo (2021), who compare transaction costs on centralized and decentralized exchanges and find that high gas prices imply that the latter only become competitive for transactions over US$100,000. Hasbrouck, Rivera, and Saleh (2022) argue that liquidity providers require remuneration. We complement the argument by stating that high fees might be *necessary* for some liquidity providers to cover the fixed

costs of managing their position. In line with our theoretical predictions, Caparros, Chaudhary, and Klein (2023) find that liquidity providers reposition their quotes more often on Uniswap V3 pools built on Polygon, which features substantially lower gas fees. Finally, Heimbach, Schertenleib, and Wattenhofer (2022) document that after accounting for price impact, concentrated liquidity on Uniswap v3 pools results in increased returns for sophisticated participants but losses for retail traders.

Despite higher gas costs, decentralized exchanges may hold advantages over centralized venues. Han, Huang, and Zhong (2022) demonstrate Uniswap frequently leads price discovery compared to centralized exchanges such as Binance, despite the latter having higher trading volume. Capponi, Jia, and Yu (2023) find that the fee paid by traders to establish execution priority unveils their private information, and therefore contributes to price discovery. Aspris, Foley, Svec, and Wang (2021) argue that decentralized exchanges offer better security than their centralized counterparts since assets are never transferred to the custody of a third party such as an exchange wallet. In turn, Brolley and Zoican (2023) make the point that decentralized exchanges may be able to reduce overall computational costs associated with latency arbitrage races, as they eliminate long-term co-location subscriptions.

Our paper is related to both the finance literature that examines whether make-take fees affect market quality and to the economics literature on two-sided markets and platform competition. Broadly, our work differs from the finance literature in that we explicitly consider equity markets as markets for liquidity without focusing on the order choice decision, and our work differs from the economics literature in that we explicitly analyze an equity market as a market for liquidity. The main insight that this brings is that market participants are both large and strategic, compared to smaller players in consumer-facing markets that are often analyzed in the economics literature.

2 Model

Asset and agents. Consider a continuous time model of trade in a single token **T** with expected value $v_t > 0$. Three risk neutral trader types consummate trade in this market: a continuum of liquidity providers (**LP**s), liquidity traders (**LT**s), and arbitrageurs (**A**). Trade occurs either because public news triggers a change in the common value of the asset, or because market participants have heterogeneous private values for the asset.

Arrival times of news and private value shocks follow independent Poisson processes with rates $\eta \in (0, 1)$ and $1 - \eta$, respectively.[5] For notational compactness, we first characterize the generic shock distribution and then describe its effects on arbitrageurs or liquidity traders. Conditional on

[5]This is without loss of generality, as what matters in the model is the *relative* arrival rate of news relative to liquidity traders.

an event at time t, the asset value changes to $v_t\left(1+\mathcal{I}\tilde{\delta}\right)$ for all traders in the case of a common value shock, or for an arriving (**LT**) in the case of the private value shock. Here, $\mathcal{I}$ is an indicator that takes on the value of 1 if the taker buys and -1 if the taker sells. The value innovation $\tilde{\delta}$ has a probability density

$$\phi\left(\delta\right)=\frac{1}{2\Delta\sqrt{1+\delta}}\quad\text{for }\delta\in\left[0,\Delta^2-1\right], \tag{1}$$

thus $\sqrt{1+\tilde{\delta}}$ is uniformly distributed between $[1,\Delta]$. This assumption is innocuous and made for tractability purposes.

When news arrives, the innovation is to the common value of the token. (As we are agnostic as to the source of value of cryptocurrencies, this common value shock could include the possibility of resale on another exchange.) After such a shock, an arbitrageur **A** trades with the liquidity providers whenever profitable, and **LP**s face an adverse selection loss. Conversely, when a liquidity trader enters the market, they experience a private value shock — and liquidity providers continue to value the token at v_t. In what follows for expositional simplicity, as in Foucault, Kadan, and Kandel (2013), we focus on a one sided market in which liquidity takers act as buyers, and news lead to an increase in token value.

Liquidity providers (**LP**) differ in their endowments of the token. Each provider i can supply at most $q_i\,\mathrm{d}i$ of the token, where q_i follows an exponential distribution with scale parameter λ. The right skew of the distribution captures the idea that there are many low-endowment liquidity providers such as retail traders, but few high-capital **LP**s such as sophisticated quantitative funds. Heterogeneity in **LP** size is captured by λ, where a larger λ naturally corresponds to a larger dispersion of endowments and larger aggregate liquidity supply. Given the endowment distribution, collectively **LP**s supply at most

$$S=\int_0^\infty q_i\frac{1}{\lambda}e^{-\frac{q_i}{\lambda}}\,\mathrm{d}i=\lambda \tag{2}$$

tokens.

Trading environment. Traders can interact in two liquidity pools in which token trade occurs against a numéraire asset (cash). At the start of the trading game, each liquidity provider (LP) deposits liquidity to a single pool within a symmetric price band around the current asset value $\left[\frac{v}{(1+r)^2},v(1+r)^2\right]$, where $r\geq 0$. Here, we make use of the fact that V3 features "price bands," and thus liquidity can be consumed with a bounded price impact. Within this range, prices in both pools satisfy a constant product bonding curve as in Adams, Zinsmeister, Salem, Keefer, and Robinson (2021). In particular, for pool k,

$$\underbrace{\left(T_k+\frac{L_k}{\sqrt{v}\,(1+r)}\right)}_{\text{virtual token reserves}}\underbrace{\left(T_k v+L_k\frac{\sqrt{v}}{1+r}\right)}_{\text{virtual numeraire reserves}}=L_k^2, \tag{3}$$

where T_k is the amount of tokens deposited on pool k and L_k is the liquidity level of pool k, defined as

$$L_k = \frac{T_k}{\frac{1}{\sqrt{v}} - \frac{1}{\sqrt{v}(1+r)}}. \tag{4}$$

To purchase τ tokens, a trader needs to deposit an amount $n(\tau) = \tau T_k \frac{v(1+r)}{\tau + (1+r)(T_k - \tau)}$ of numéraire into the pool, where $n(\tau)$ is the solution to the invariance condition

$$\underbrace{\left(T_k - \tau + \frac{L_k}{\sqrt{v}(1+r)}\right)}_{\text{virtual token reserves}} \underbrace{\left(T_k v + n(\tau) + L_k \frac{\sqrt{v}}{1+r}\right)}_{\text{virtual numeraire reserves}} = L_k^2. \tag{5}$$

Fees are levied on liquidity takers as a fraction of the value of the trade and distributed pro rata to liquidity providers. Crucially, the pools have different fees. One pool charges a low fee, and one pool charges a high fee which we denote ℓ and h respectively. Specifically, to purchase τ units of the token on the low fee pool, the total cost to a taker is $(1+\ell)\, n(\tau, T_\ell)$. The **LP**s in the pool receive $\ell n(\tau, T_\ell)$ in fees. In addition, consistent with gas costs on Ethereum, all traders incur a fixed execution cost $\Gamma\, \mathrm{d}i > 0$ to interact with the market.

Figure 1 illustrates the timing of the model.

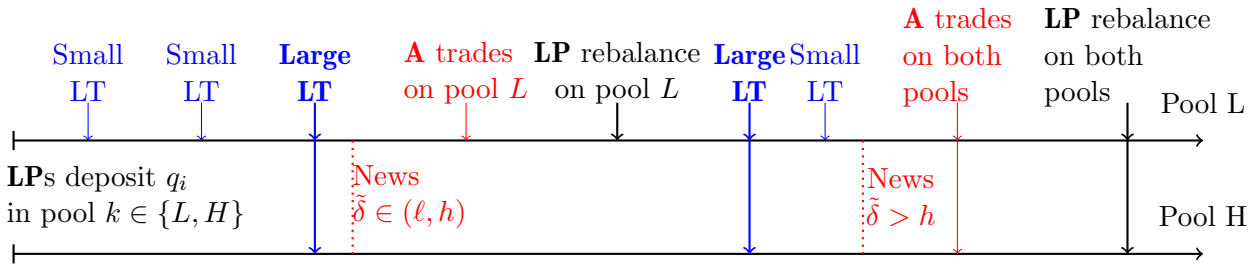


Figure 1: Model timing

To ensure the possibility of liquidity re-balancing in both pools, we assume that innovations are large enough to ensure that **LP**s may need to rebalance their position on the high fee pool or:

Assumption 1: The size of innovations are sufficiently large so that there is a positive probability that liquidity providers need to re-balance on the high-fee pool. That is, $\Delta > (1+r)\sqrt{1+h}$.

2.1 Equilibrium

2.1.1 Optimal trade size

First, consider the decisions of arbitrageurs and liquidity traders holding a value $v(1+\delta)$ for the asset. Faced with pool sizes of T_ℓ and T_h in the low and high pool respectively, their optimal trade

on pool k maximizes their expected profit, net of fees and price impact:

$$\max_{\tau} \text{Profit } \mathbf{LT}(\tau, \delta) \equiv \tau v (1+\delta) - (1+f_k)\tau T_k \frac{v(1+r)}{\tau + (1+r)(T_k - \tau)}, \tag{6}$$

which yields the optimal trade quantity:

$$\tau^\star(\delta) = T_k \min\left\{1, \frac{1+r}{r}\max\left\{0, 1 - \sqrt{\frac{1+f_k}{1+\delta}}\right\}\right\}. \tag{7}$$

From equation (7), a trader with valuation $v(1+\delta)$ only trades on pool k if the gains from trade are larger than the liquidity fee, i.e., $\delta > f_k$. Further, if $\delta > (1+f_k)(1+r)^2 - 1$ so that the gains from trade are larger than the maximum price impact, then the trader consumes all available liquidity in the pool.

2.1.2 Fee revenue for liquidity providers from private value trades

The revenue for liquidity providers can be expressed as the product of the pool fee and the numéraire deposit required from liquidity traders to purchase $\tau^\star$ token units, denoted by $n(\tau^\star, T_k)$. That is,

$$f_k n(\tau^\star(\delta), T_k) = f_k v T_k \min\left\{1+r, \frac{1+r}{r}\max\left\{0, \sqrt{\frac{1+\delta}{1+f_k}} - 1\right\}\right\}. \tag{8}$$

If the innovation δ corresponds to a private rather than common value shock, then an arbitrageur optimally steps in to reverse the liquidity trade as described in Lehar and Parlour (forthcoming). In this case, liquidity providers effectively earn double the fee revenue in (8) without affecting the capital structure of the pool; there is neither a capital gain nor a loss for the **LP**s.

The fee revenue in (8) scales linearly with the size of the pool T_k. Since fee proceeds are distributed pro-rata among liquidity providers based on their share $\frac{q_i}{T_k}$, it follows that fee revenue an **LP** with endowment q_i providing liquidity on pool k increases linearly in their endowment:

$$\begin{aligned}\text{FeeRevenue}_{i,k}(\delta) &= 2\frac{q_i}{T_k} f_k n(\tau^\star, T_k) \\ &= 2 q_i v f_k \min\left\{1+r, \frac{1+r}{r}\max\left\{0, \sqrt{\frac{1+\delta}{1+f_k}} - 1\right\}\right\}.\end{aligned} \tag{9}$$

The expression in (9) denotes the fee revenue conditional on the private value δ of the incoming trade. To compute the expected fee revenue, we integrate this expression across all posible value

shocks:

$$\begin{aligned}\mathbb{E}\text{FeeRevenue}_{i,k} &= \int_{\delta=1}^{\Delta^2-1} \text{FeeRevenue}_{i,k}(\delta)\,\phi(\delta)\,\mathrm{d}\delta \\ &= q_i \underbrace{v\frac{f_k(r+1)\left(2\Delta - r\sqrt{f_k+1} - 2\sqrt{f_k+1}\right)}{\Delta}}_{\equiv\mathcal{L}(f_k)},\end{aligned} \tag{10}$$

where we define $\mathcal{L}(f_k)$ as the *liquidity yield*: that is, the per-unit **LP** fee revenue from supplying liquidity to **LT**s in pool k.

Lemma 1. *There exists a threshold fee level $\overline{f} > 0$ such that the liquidity revenue $\mathcal{L}(f_k)$ first increases in the pool fee f_k for $f \leq \overline{f}$, then decreases in the pool fee for $f > \overline{f}$.*

Lemma 1 points out to a non-linear relationship between fee levels and liquidity yield. Initially, as fees increase, the enhanced revenue from higher fees outweighs the decrease in trading volume due to increased transaction costs, resulting in a net gain in revenue. However, beyond a certain fee threshold, the drop in trading volume dominates the larger fee, leading to a decrease in overall revenue. A salient implication is that if pool fees are large enough, the liquidity yield on the high fee pool may exceed the yield on the low-fee pool.

2.1.3 Adverse selection cost for liquidity providers

If news occurs (i.e., if δ represents a common value shock), liquidity providers trade against arbitrageurs rather than liquidity traders. In this case, there is no subsequent price reversal following the initial trade. The capital structure of the liquidity pool changes, as arbitrageurs remove the more valuable asset: i.e., buy tokens upon a positive common value shock. While **LP**s earn fee revenues on arbitrage trades, they also incur adverse selection losses by trading against the direction of the news. Moreover, if the magnitude of news is large enough that arbitrageurs remove all tokens supplied in the price range, then **LP**s face additional costs, that is a gas fee $\Gamma\,\mathrm{d}i$ to re-balance liquidity around the new asset value.

Table 1 delineates the **LP** fee revenue from selling tokens to arbitrageurs, as well as the marked-to-market value of the tokens sold. If the size of news (δ) does not exceed the pool fee, then arbitrageurs do not trade since the potential profit does not justify the transaction cost. Conversely, if the news size is larger than pool fee, then arbitrageurs execute a trade proportional to the size of the pool, and they exhaust the available liquidity on the price range if the news is large enough: specifically, if $\delta > (1+f_k)(1+r)^2 - 1$. The profit for liquidity providers in each scenario is the difference between the revenue and the marked-to-market value. Notably, the profit is consistently negative, since **LP**s are trading against the direction of news.

Table 1: Fee revenue and capital losses on arbitrage trades

News size	Revenue (numeraire)	Marked-to-market token value
$\delta \leq f_k$	0	0
$\delta \in \left(f_k, (1+f_k)(1+r)^2 - 1\right]$	$vq_i \frac{1+r}{r}(1+f_k)\left(\sqrt{\frac{1+\delta}{1+f_k}} - 1\right)$	$vq_i \frac{1+r}{r}(1+\delta)\left(1 - \sqrt{\frac{1+f_k}{1+\delta}}\right)$
$\delta > (1+f_k)(1+r)^2 - 1$	$vq_i (1+f_k)(1+r)$	$vq_i (1+\delta)$

The expected **LP** profit from trading with arbitrageurs equals $-q_i v \times \mathcal{A}(f_k)$, where $\mathcal{A}(f_k)$ is the per-unit adverse selection cost from liquidity provision in pool k:

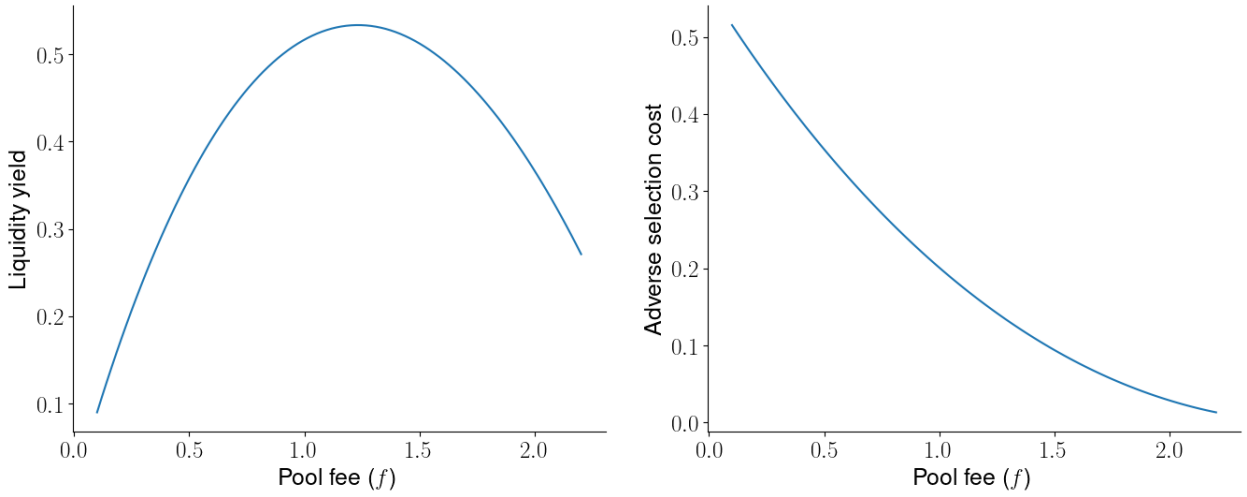
$$\mathcal{A}(f_k) = \mathbb{P}\left(f_k < \delta \leq (1+f_k)(1+r)^2 - 1\right) \times \frac{1+r}{r} \mathbb{E}\left[(1+f_k) + (1+\delta) - 2\sqrt{(1+\delta)(1+f_k)}\right] + \\ + \mathbb{P}\left(\delta > (1+f_k)(1+r)^2 - 1\right) \times \left[\mathbb{E}(1+\delta)\right] - (1+f_k)(1+r)\Big\}. \tag{11}$$

Lemma 2. *The adverse selection cost $\mathcal{A}(f_k)$ decreases in the pool fee f_k. In particular, the high-fee pool has a lower adverse selection cost than the low-fee pool.*

Lemma 2 indicates that higher pool fees lower adverse selection costs through two mechanisms: First, they increase compensation per unit traded for liquidity providers (**LP**s), enhancing their returns on trades with arbitrageurs. Second, higher fees discourage arbitrageur activity, effectively reducing the volume of informed trades. Figure 2 showcases the results in Lemmas 1 and 2 and illustrates the comparative statics of liquidity yield and adverse selection cost with respect to the pool fee.

Figure 2: Liquidity yield and adverse selection cost

This figure illustrates the expected fee yield from liquidity trades (left panel) and the adverse selection cost (right panel), as a function of the pool fee f. Parameter values: $r = 0.001$, $\lambda = 1$, $\eta = 0.1$, and $\Delta = 1.1\,(1+r)\,\sqrt{1+h}$.



Liquidity rebalancing costs arise only when news events are large enough to deplete all available liquidity within a given price range, pushing liquidity providers' (**LP**s) positions "out of range." Rebalancing only occurs post-news, since equally large liquidity trades would be reversed by arbitrageurs. Conditional on news arrival, the expected cost of rebalancing is

$$\mathcal{C}(k) = \mathbb{P}\left(\delta > (1+f_k)(1+r)^2 - 1\right)\Gamma = \Gamma\left(1 - \frac{\sqrt{1+f_k}\,(1+r)}{\Delta}\right), \tag{12}$$

which is decreasing in the pool fee f_k. This result is straightforward: smaller news events can cause arbitrageurs to deplete liquidity in low-fee pools, whereas it takes larger news to do the same in high-fee pools. Consequently, **LP**s in lower fee pools face more frequent rebalancing and incur higher fixed costs.

2.1.4 Liquidity provider pool choice

Liquidity providers face a choice between the low and high fee pool or not participating in the market. An **LP** of size q_i earns expected profit

$$\begin{aligned}\pi_L &= q_i\left[(1-\eta)\,\mathcal{L}(\ell) - \eta\mathcal{A}(\ell)\right] - \eta\Gamma\left(1 - \frac{\sqrt{1+\ell}\,(1+r)}{\Delta}\right) \text{ and} \\ \pi_H &= q_i\left[(1-\eta)\,\mathcal{L}(h) - \eta\mathcal{A}(h)\right] - \eta\Gamma\left(1 - \frac{\sqrt{1+h}\,(1+r)}{\Delta}\right),\end{aligned} \tag{13}$$

from choosing pool L or H, respectively. Equation (13) underscores the trade-off faced by liquidity providers (**LP**s): balancing the liquidity yield from trades with liquidity traders (**LT**s) against the adverse selection costs and the fixed gas costs associated with re-balancing their position.

First, consider the choice of participating in the market. An agent only provides liquidity on pool k if she is able to break even – that is, if her endowment q_i is large enough. We define $\underline{q}_L$ and $\underline{q}_H$ as the thresholds at which the participation constraints $\pi_L(q) = 0$ and $\pi_H(q) = 0$ are satisfied, respectively. If $\underline{q}_k \geq 0$ for a pool k, it indicates that any **LP** with an endowment q_i at least equal to $\underline{q}_k$ can join pool k and expect to earn a positive profit, with the marginal entrant breaking even. Conversely, if $\underline{q}_k < 0$, it suggests that pool k is not economically viable as the participation constraint is breached for all **LP**s.

Assumption 2: To avoid trivial cases, we focus on the case that both markets are potentially viable, or equivalently the intensity of news is low enough:

$$\eta \leq \min_k \frac{\mathcal{L}(k)}{\mathcal{L}(k) + \mathcal{A}(k)}, \tag{14}$$

such that $\underline{q}_k \geq 0$.

Next, consider the choice between pools. Liquidity provider i chooses the low-fee pool if and only if

$$\pi_L - \pi_H = q_i \left[(1-\eta)(\mathcal{L}(\ell) - \mathcal{L}(h)) + \eta \underbrace{(\mathcal{A}(h) - \mathcal{A}(\ell))}_{<0} \right] - \underbrace{\Gamma \frac{\eta(1+r)}{\Delta} \left(\sqrt{1+h} - \sqrt{1+\ell} \right)}_{>0} > 0. \tag{15}$$

Liquidity providers in the high-fee pool face both lower adverse selection and rebalancing costs. Therefore, the low-fee pool can only be chosen in equilibrium if it offers a higher liquidity yield, specifically if $\mathcal{L}(\ell) - \mathcal{L}(h) > 0$, and if the intensity of news η is sufficiently low. Otherwise, all liquidity providers prefer to supply tokens to the high-fee pool if the participation constraint is satisfied. Further, equation (15) highlights the economies of scale embedded in liquidity provision with fixed rebalancing costs. That is, if a liquidity provider of size q prefers the low fee pool, then any liquidity provider with a larger endowment, $\widetilde{q} > q$, also prefers the low fee pool.

Proposition 1 characterizes the equilibrium liquidity provision.

Proposition 1. *i. If* $\eta > \frac{\mathcal{L}(l) - \mathcal{L}(h)}{\mathcal{L}(l) - \mathcal{L}(h) + \mathcal{A}(l) - \mathcal{A}(h)}$, *then all* ***LP****s with* $q_i > \underline{q}_h$ *deposit liquidity on the high fee pool.*

ii. Otherwise, there exists a unique fragmented equilibrium characterized by marginal trader

$q_t^\star > \underline{q}_h$ *which solves*

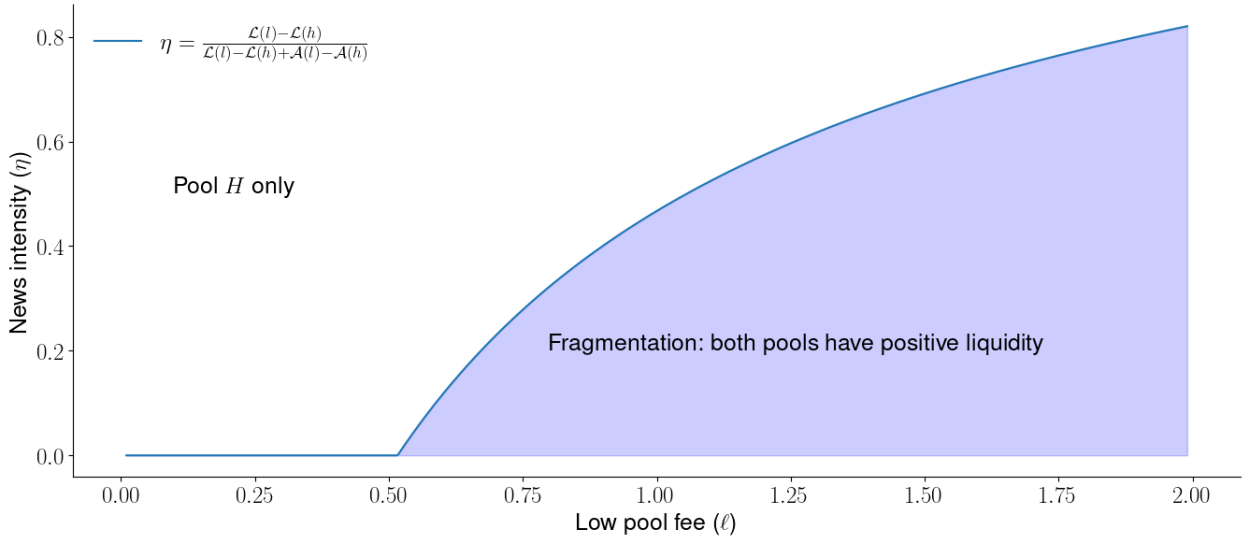
$$q_t^\star = \Gamma \frac{\eta(1 + r\left(\sqrt{1+h} - \sqrt{1+\ell}\right))}{\Delta\left[(1-\eta)\left(\mathcal{L}\left(\ell\right) - \mathcal{L}\left(h\right)\right) + \eta\left(\mathcal{A}\left(h\right) - \mathcal{A}\left(\ell\right)\right)\right]} \tag{16}$$

such that all ***LPs*** *with* $q_i \in \left(\underline{q}_h, q_t^\star\right]$ *deposit liquidity in the high fee pool and all* ***LPs*** *with* $q_i > q_t^\star$ *choose the low fee pool.*

Figure 3 illustrates the equilibrium regions in Proposition 1. When news intensity η is high, or pool H offers a substantially higher fee than pool L, liquidity suppliers gravitate towards pool H, resulting in a single-maker equilibrium. Conversely, a lower η translates to lower adverse selection costs. If this is the case, or if the fee differential between the two pools is low, liquidity providers with large endowments q migrate to the lower-fee pool to compete for order flow from small traders, causing liquidity to fragment between the two pools.

Figure 3: Fragmented and single-pool equilibria

This figure plots the existence conditions for a fragmented market equilibrium, as described in Proposition 1, for various values of the news intensity (η) on the y-axis and liquidity fee on pool L on the x-axis. Parameter values: $r = 0.001$, $h = 2$, $\lambda = 1$, $\eta = 0.1$, and $\Delta = 1.1\,(1+r)\sqrt{1+h}$.



Proposition 2 establishes the impact of gas prices on the two pools' liquidity market shares. We can compute the liquidity market share of the low-fee pool in a fragmented equilibrium as

$$w_\ell = \frac{\exp\left(-\frac{q_t - \underline{q}_h}{\lambda}\right)(q_t + \lambda)}{\underline{q}_h + \lambda} \leq 1, \tag{17}$$

with equality for $\Gamma = 0$. That is, as fixed costs drop to zero, the low fee pool asymptotically captures

the full market share.

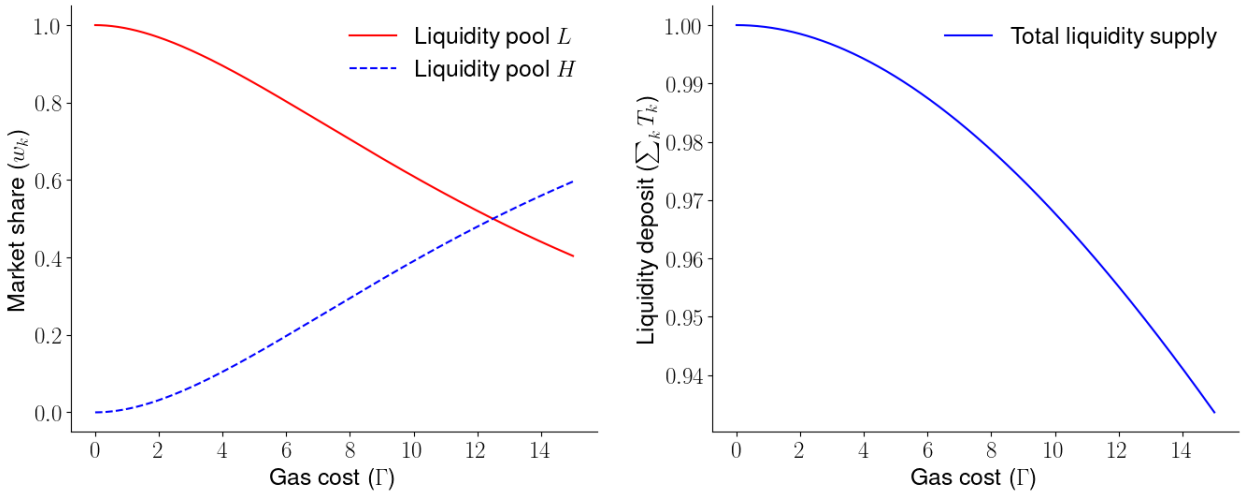
Proposition 2. *In equilibrium, the market share of the low fee pool w_ℓ decreases in the gas cost (Γ).*

We stress the critical role of fixed gas costs in driving market fragmentation. Since liquidity fee revenues and adverse selection costs are distributed pro-rata, in the absence of gas fees, all liquidity providers (**LP**s) would converge on a single pool — the one offering the optimal balance between fee yield and informational costs. For instance, if $\Gamma = 0$, all **LP**s would select pool L if the news arrival rate is sufficiently low, as defined by $\eta \leq \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$, or choose pool H otherwise. It is the introduction of fixed costs that drives **LP**s to segregate into different pools based on their size.

Figure 4 shows that the market share of the low fee pool decreases in the gas cost Γ. A larger gas price increases the costs of re-balancing upon the arrival of large enough news, everything else equal, and incentivizes smaller **LP**s to switch from the low fee pool to the high fee pool, since the arbitrageurs are less likely to fully consume liquidity there. Further, the right panel illustrates the extensive margin effect of gas prices: any increase in gas costs leads to a decrease in aggregate liquidity supply as some **LP** with low endowments are driven out of the market (that is, the threshold $\underline{q}_h$ increases in Γ).

Figure 4: Liquidity shares and gas costs

This figure illustrates the equilibrium liquidity market shares (left panel) and the aggregate liquidity supply on the two pools (right panel), as a function of the gas fee Γ. Parameter values: $r = 0.001$, $h = 2$, $\ell = 1$, $\lambda = 1$, $\eta = 0.1$, and $\Delta = 1.1\,(1+r)\,\sqrt{1+h}$.



2.2 Pool fragmentation and market quality

We measure market quality by the realized gains from trade of liquidity traders. If the asset is traded on a sequence of pools, where f_k and T_k represent the fees and liquidity deposits on pool k,

respectively, the expected gains from trade for liquidity traders are

$$\text{GainsFromTrade}\left(\{f_k\}_k\right) = v\mathbb{E}\left[\sum_k \tau^\star\left(f_k, \delta\right) \times \delta\right], \tag{18}$$

where $\tau^\star(\delta) = T_k \min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+f_k}{1+\delta}}\right\}\right\}$ is the optimal **LT** trade size, as defined in equation (7).

Suppose an asset is traded on a single pool that imposes a liquidity fee f. From equation (18) it follows that the gains from trade for an **LT** with private value $1+\delta$ are equal to

$$\text{GainsFromTrade}\left(f \mid \delta\right) = v\delta T_f \min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+f}{1+\delta}}\right\}\right\}. \tag{19}$$

The total token supply on the single pool equals $T_f = e^{-\underline{q}_f \lambda}\left(\underline{q}_f + \lambda\right)$, where $\underline{q}_f$ is the marginal liquidity provider such that all **LP**s with endowment $q_i > \underline{q}_f$ join the pool. Here, the magnitude of the liquidity fee drives the trade-off between the participation of liquidity providers (**LP**) and trading costs. A lower fee f results in fewer **LP**s offering liquidity, a lower token supply T_f, which limits gains from trade for liquidity traders. In contrast, a higher fee increases trading costs, potentially outweighing the benefits of increased **LP** participation.

Proposition 3. *For any single-pool fee $f \geq 0$, there exists a set of fees $\{h, \ell\}$ for a two-pool fragmented market, where $h = f$ and $h > \ell$, that guarantees equal or higher gains from trade in a fragmented market compared to the single-pool market.*

Proposition 3 suggests that fragmentation with multiple fee levels improves market quality. Specifically, it is always possible to devise a fee structure in a fragmented market that yields (weakly) higher gains from trade than a single-fee market. The logic is as follows: First, the highest fee in the fragmented market is set equal to the single pool fee, ensuring that the marginal **LP** participating the market is the same across both scenarios (i.e., the **LP** with endowment $\underline{q}_h$). This condition guarantees the same aggregate liquidity supply in fragmented and non-fragmented markets. Second, a lower fee is then chosen for another pool to attract liquidity providers with higher token endowments, resulting in larger trade sizes per unit of supplied liquidity. This combination of larger liquidity trades and unchanged aggregate liquidity supply leads to higher gains from trade in a fragmented market.

Figure 5: Gains from trade and market structure

The figure plots the expected gains from trade across **LT**s,

$$\text{GainsFromTrade} = \int_0^{\Delta^2-1} v\delta\tau^\star(\delta)\,\phi(\delta)\,\mathrm{d}\delta,$$

on both a single pool with a high fee as well as on fragmented pools, as a function of the gas cost Γ. Parameter values: $r = 0.001$, $h = 2$, $\ell = 1$, $\lambda = 1$, $\eta = 0.1$, and $\Delta = 1.1\,(1+r)\,\sqrt{1+h}$.

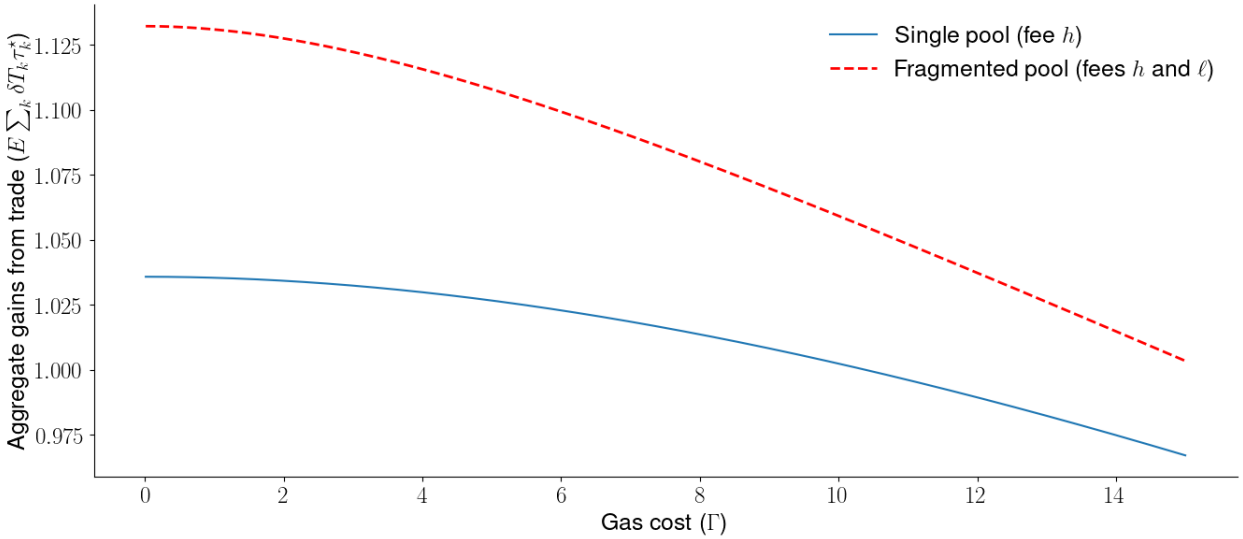


Figure 5 illustrates the result. As gas price increase, the gains from trade drop in both single-pool and fragmented markets, primarily because more **LP**s are priced out which results in lower liquidity supply and higher price impact. Nevertheless, irrespective of the level of gas costs, the gains from trade are higher in the fragmented market.

We note that the argument discussed in this section is valid for *any* single-fee pool, including an optimally designed one. In essence, if a fragmented fee structure can be designed to achieve higher gains from trade compared to an arbitrary single-pool fee, then a fee structure that dominates the optimally set single-pool fee achieves higher gains from trade than any single-fee pool.

2.3 Model implications and empirical predictions

Prediction 1: The liquidity market share of the low-fee pool decreases in the gas fee Γ.

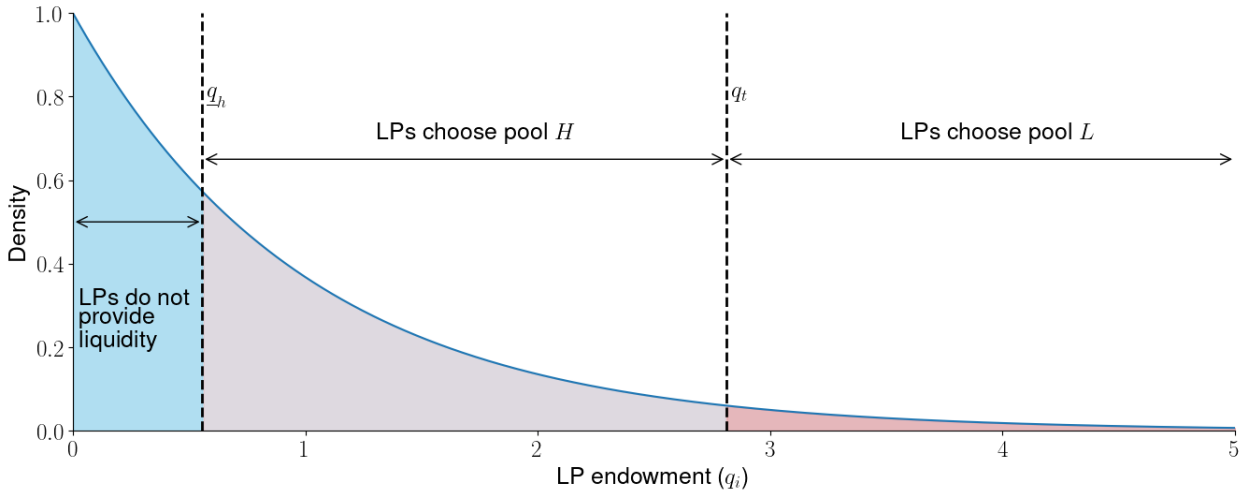
Prediction 1 follows directly from Proposition 2 and Figure 4. A higher gas price increases the cost of liquidity re-balancing. Given that re-balancing is more frequently required in the low-fee pool due to more intense arbitrage activity, liquidity providers, particularly those with smaller endowments, optimally migrate to the high-fee pool in response to a gas cost increase.

Prediction 2: **LP**s on the low-fee pool make larger liquidity deposits than **LP**s on the high-fee pool.

Prediction 2 follows from the equilibrium discussion in Proposition 1. Liquidity providers with large token endowments ($q_i > q_t$) deposit them in the low-fee pool since they are better positioned to actively manage liquidity due to economies of scale. **LP**s with lower endowments ($q_i \leq q_t$) either stay out of the market or choose pool H which allows them to offer liquidity in a more passive manner. Figure 6 illustrates this prediction by overlaying optimal pool choices on the distribution of **LP** endowments. Low-endowment **LP**s (in blue) that are being rationed out of the market due to high gas cost, medium-endowment **LP**s (gray) that deposit liquidity on pool H, and high-endowment **LP**s (red) that choose the low-fee pool L.

Figure 6: Liquidity supply on fragmented markets

This figure illustrates the endowment distribution of **LP**s and their choice of pools in a fragmented market. First, liquidity providers to the left of $\underline{q}_h$ do not provide liquidity on either pool. Next, **LP**s to the left (right) of the marginal trader $q_t^\star$ provide liquidity on pool H (pool L, respectively). $r = 0.001$, $h = 2$, $\ell = 1$, $\lambda = 1$, $\eta = 0.1$, $\Gamma = 20$, and $\Delta = 1.1\,(1+r)\,\sqrt{1+h}$.



Prediction 3: The average trade size is higher on pool H than on pool L. At the same time, trading volume is higher on pool L than on pool H.

Next, Prediction 3 deals with differences between incoming trades on the two liquidity pools. If liquidity traders and arbitrageurs find it optimal to trade on pool H since $\delta > h$, then they would also trade on pool L since $h > \ell$ and therefore $\delta > \ell$. However, the opposite is not true: **LT**s and arbitrageurs with $\delta \in [\ell, h)$ only trade on pool L. In equilibrium, only a fraction of traders with sufficiently high private values are drawn to pool H.

Prediction 4: In a fragmented market equilibrium, the liquidity yield is higher on the low fee pool than on the high fee pool.

Prediction 4 is a consequence of Proposition 1. The high fee pool offers better protection against adverse selection and re-balancing costs. If the low fee pool attracts a positive market share, then it

necessarily compensates with a higher liquidity yield.

Prediction 5: The average liquidity deposit on both the low- and- high fee pool increases with gas costs.

An increase in the gas cost Γ has two effects: first, the **LP**s with the lowest endowments on pool L switch to pool H. As a result, the average deposit on pool L increases. Second, the **LP**s with low endowments on pool H may leave the market. Both channels translate to a higher average deposit on pool H, which experiences an inflow (outflow) of relatively high (low) endowment LP following an increase in gas costs.

Prediction 6: **LP**s re-balance liquidity more frequently on the low-fee than on the high-fee pool.

Liquidity providers re-balance their positions in a pool charging a fee f only when the magnitude of news exceeds a threshold, specifically if $\delta > (1+f)(1+r)^2 - 1$. The likelihood of re-balancing given the news is $(1 - \frac{\sqrt{1+f}(1+r)}{\Delta})$. Consequently, the duration of a liquidity cycle, expressed as $\frac{1}{\eta(1-\frac{\sqrt{1+f}(1+r)}{\Delta})}$, increases in the pool fee level.

Prediction 7: Adverse selection cost is higher on the low fee pool than on the high fee pool.

This prediction follows directly from Lemma 2: a higher pool fee serves as a deterrent to arbitrageurs, particularly if the size of news remains below a threshold.

3 Data and descriptive statistics

3.1 Sample construction

We obtain data from the Uniswap V3 Subgraph, covering all trades, liquidity deposits (referred to as "mints"), and liquidity withdrawals (referred to as "burns") on 4,069 Uniswap v3 pools. The data spans from the protocol's launch on May 4, 2021, up until July 15, 2023. Each entry in our data includes a transaction hash that uniquely identifies each trade and liquidity update on the Ethereum blockchain. Additionally, it provides details such as trade price, direction, and quantity, along with quantities and price ranges for each liquidity update. Moreover, the data also includes wallet addresses associated with initiating each transaction, akin to anonymous trader IDs. The Subgraph data we obtained also provides USD-denominated values for each trade and liquidity mint. We further collect daily pool snapshots from the Uniswap V3 Subgraph, including the end-of-day pool size in Ether and US Dollars, and summary price information (e.g., open, high, low, and closing prices for each pool).

To enhance our dataset, we combine the Subgraph data with public Ethereum data available on Google Big Query to obtain the position of each transaction in its block, as well as the gas price

limit set by the trader and the amount of gas used. Finally, we obtain block-by-block liquidity snapshot data across multiple price ranges from Kaiko.

There are no restrictions to list a token pair on Uniswap. Some pools might therefore be used for experiments, or they might include untrustworthy tokens. Following Lehar and Parlour (forthcoming), we remove pools that are either very small or that are not attracting an economically meaningful trading volume. We retain liquidity pools that fulfill the following four criteria: (i) have at least one interaction in more than 100 days in the sample, (ii) have more than 500 liquidity interactions throughout the sample, (iii) have an average daily liquidity balance in excess of US$100,000, and (iv) capture more than 1% of trading volume for a particular asset pair. We exclude burn events with zero liquidity withdrawal in both base and quote assets, as traders use them solely to collect fees without altering their liquidity position.

These basic screens give us a baseline sample of 274 liquidity pools covering 242 asset pairs, with combined daily dollar volume of $1.12 billion and total value locked (i.e., aggregate liquidity supply) of $2.53 billion as of July 15, 2023. We capture 24,202,803 interactions with liquidity pool smart contracts (accounting for 86.04% of the entire universe of trades and liquidity updates). Trading and liquidity provision on Uniswap is heavily concentrated: the five largest pairs (USDC-WETH, WETH-USDT, USDC-USDT, WBTC-WETH, and DAI-USDC) account on average for 86% of trading volume and 63% of supplied liquidity.[6]

3.2 Liquidity fragmentation patterns

For 32 out of the 242 asset pairs in our baseline sample, liquidity supply is fragmented across two pools with different fees – either with 1 and 5 bps fees (5 pairs), 5 and 30 bps fees (6 pairs), or 30 and 100 bps fees (21 pairs).[7] Despite being fewer in number, fragmented pairs are economically important: they account on average for 95% of the capital committed to Uniswap v3 and for 93% of its dollar trading volume. All major token pairs such as WETH-USDC, WETH-USDT, or WBTC-WETH trade on fragmented pools.

For each fragmented liquidity pair, we label the *low* and the *high* fee liquidity pool to facilitate analysis across assets. For example, the low and high liquidity fees for USDC-WETH are 5 and 30 bps, respectively, but only 1 and 5 bps for a lower volatility pair such as USDC-USDT. We refer to non-fragmented pools as *single* (i.e., the unique pool for an asset pair).

We aggregate all interactions with Uniswap smart contracts into a panel across days and liquidity

[6]WETH and WBTC stand for "wrapped" Bitcoin and Ether. Plain vanilla Bitcon and Ether are not compliant with the ERC-20 standard for tokens, and therefore cannot be directly used on decentralized exchanges' smart contracts. USDC (USD Coin), USDT (Tether), and DAI are stablecoins meant to closely track the US dollar.

[7]In some cases, more than two pools are created for a pair – e.g., for USDC-WETH there are four pools with 1, 5, 30, and 100 bps liquidity fees. In all cases however, two pools heavily dominate the others: As described in Section 3.1 we filter out small pools with less than 1% volume share or less than $100,000 liquidity deposits.

pools. To compute the end-of-day pool size, we account for all changes in token balances, across all price ranges. There are three possible interactions: A deposit or "mint" adds tokens to the pool, a withdrawal or "burn" removes tokens, whereas a trade or "swap" adds one token and removes the other. We track these changes across to obtain daily variation in the quantity of tokens on each pool. We obtain dollar values for the end-of-day liquidity pool sizes, intraday trade volumes, and liquidity events from the Uniswap V3 Subgraph. To determine a token's price in dollars, the Subgraph searches for the most liquid path on Uniswap pools to establish the token's price in Ether and subsequently converts the Ether price to US dollars.

Table 2 reports summary statistics across pools with different fee levels. High-fee pools attract on average 58% of total liquidity supply, significantly more than their low-fee counterparts ($46.50 million and $33.78 million, respectively), but only capture 20.74% percent of the trading volume (computed as $^{8{,}071.24}/_{(8{,}071.24+30{,}848.79)}$ from the first column of Table 2). Consistent with our theoretical predictions, low-fee pools attract five times as many trades as high-fee competitors (610 versus 110 average trade count per day). At the same time, the average trade on a high-fee pool is twice as large ($14,490) than on a low-fee pool ($6,340).

The distribution of mint sizes is heavily skewed to the right, with 6.6% of deposits exceeding $1 million. There are large differences across pools – the median **LP** deposit on the low-fee pool is $15,680, twice as much as the median deposit on the high-fee pool ($7,430). At the same time, the number of liquidity providers on high-fee pools is 51% higher than on low-fee pools (10.08 unique addresses per day on high-fee pools versus only 6.68 unique address on high-fee pools).

One concern with measuring average mint size is just-in-time liquidity provision (JIT). As discussed for example in Capponi, Jia, and Zhu (2024), JIT liquidity providers submit very large and short-lived deposits to the pool to dilute competitors on an incoming large trade; they immediately withdraw the balance in the same block after executing the trade. In our sample, JIT liquidity provision is not economically significant, accounting for less than 1% of aggregate trading volume. However, it has the potential to skew mint sizes to the right, particularly in low-fee pools, without providing liquidity to the market at large. We address this issue by (i) filtering out JIT mints using the algorithm in Appendix D and (ii) taking the median liquidity mint size at day-pool level rather than the mean.

Further, we follow Augustin, Chen-Zhang, and Shin (2022) to compute the daily liquidity fee yield as the product between pool's fee tier and the ratio between trading volume and the lagged total value locked (TVL). That is,

$$\text{Liquidity yield} = \text{liquidity fee}_i \times \frac{\text{Volume}_{i,t}}{\text{TVL}_{i,t-1}}, \tag{20}$$

for pool i and day t. The average daily yield is slightly higher on low-fee pools, at 11.72 basis points, compared to 9.69 basis points on high-fee pools.

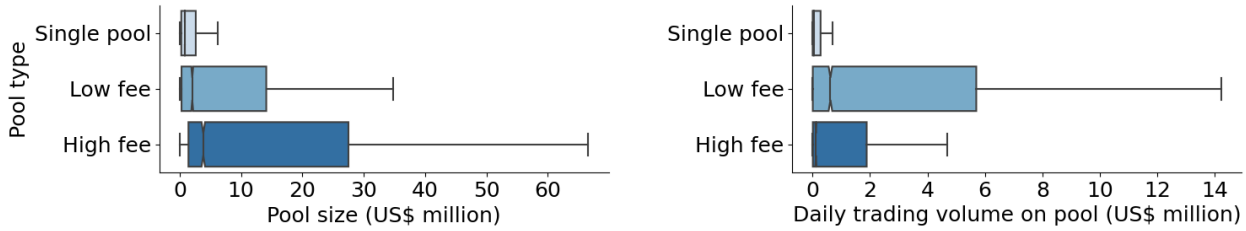
Table 2: Descriptive statistics

This table reports descriptive statistics for variables used in the empirical analysis. *Pool size* is defined as the total value locked in the pool's smart contract at the end of each day. We compute the balance on day t as follows: we take the balance at day $t-1$ and add (subtract) liquidity deposits (withdrawals) on day t, as well as accounting for token balance changes due to trades. The liquidity balance on the first day of the pool is taken to be zero. End of day balances are finally converted to US dollars. *Daily volume* is computed as the sum of US dollar volume for all trades in a given pool and day. *Liquidity share* (*Volume share*) is computed as the ratio between a pool size (trading volume) for a given fee level and the aggregate size of all pools (trading volumes) for the same pair in a given day. *Trade size* and *Mint size* are the median trade and liquidity deposit size on a given pool and day, denominated in US dollars. *Trade count* represents the number of trades in a given pool and day. ***LP** wallets* counts the unique number of wallet addresses interacting with a given pool in a day. The *liquidity yield* is computed as the ratio between the daily trading volume and end-of-day TVL, multiplied by the fee tier. The *price range* for every mint is computed as the difference between the top and bottom of the range, normalized by the range midpoint – a measure that naturally lies between zero and two. *Loss-versus-rebalancing* is computed as the permanent price impact of swaps with a one-hour horizon. The *impermanent loss* is computed as in Heimbach, Schertenleib, and Wattenhofer (2022) for a position in the range of 95% to 105% of the current pool price, with a forward-looking horizon of one hour. Finally, *mint-to-burn* and *burn-to-mint* times are defined as the time between a mint (burn) and a subsequent burn (mint) by the same address in the same pool, measured in hours. *Mint-to-burn* and *burn-to-mint* are recorded on the day of the final interaction with the pool.

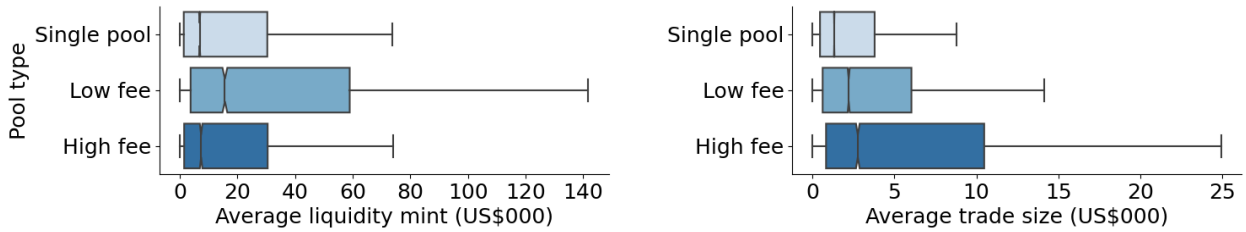
Statistic	Pool fee	Mean	Median	St. Dev.	Pctl(25)	Pctl(75)	N
Pool size ($M)	Low	33.78	2.05	96.91	0.30	14.12	20,151
	High	46.50	3.85	95.73	1.43	27.51	20,151
	Single	3.89	0.84	13.56	0.26	2.62	130,767
Liquidity share (%)	Low	39.52	35.52	32.53	7.37	72.16	20,151
	High	60.48	64.48	32.53	27.84	92.63	20,151
Daily volume ($000)	Low	30,848.79	619.77	118,908.80	6.18	5,697.30	20,151
	High	8,071.24	114.96	36,777.38	7.83	1,882.12	20,151
	Single	915.73	36.07	6,059.78	1.93	277.00	130,767
Volume share	Low	66.51	88.38	38.50	29.43	98.48	18,001
	High	42.20	23.83	41.18	3.19	95.03	18,058
Trade size ($000)	Low	6.34	2.20	13.36	0.61	6.03	18,001
	High	14.49	2.76	33.19	0.82	10.48	18,060
	Single	4.12	1.32	11.03	0.45	3.79	113,362
Mint size ($000)	Low	820.84	15.68	13,114.83	3.78	58.98	10,640
	High	1,001.10	7.43	13,807.10	1.55	30.52	10,370
	Single	96.97	6.93	622.12	1.42	30.39	45,300
Trade count	Low	610.61	95	1,518.52	12	414	20,151
	High	110.59	26	490.29	8	89	20,151
	Single	63.94	19	194.03	4	55	130,767
LP wallets	Low	6.68	1	16.01	0	6	20,151
	High	10.08	1	37.79	0	5	20,151
	Single	1.57	1.17	1.19	1.00	1.85	55,580
Liquidity yield (bps)	Low	11.72	2.58	56.31	0.16	9.08	20,122
	High	9.69	1.65	51.44	0.15	6.40	20,130
	Single	17.90	1.94	90.18	0.18	8.58	130,433
Price range	Low	0.39	0.30	0.37	0.13	0.56	11,866
	High	0.61	0.54	0.44	0.32	0.84	12,195
	Single	0.68	0.58	0.52	0.27	1.02	55,580
Loss-versus-rebalancing (bps)	Low	14.24	1.32	35.20	0.02	9.22	20,151
	High	7.85	0.84	23.15	0.03	4.87	20,151
Impermanent loss (bps)	Low	8.46	1.84	27.88	0.06	7.23	20,118
	High	7.37	1.33	27.21	0.05	5.93	20,132
	Single	17.20	2.44	71.34	0.17	11.37	130,340
Mint-to-burn (hrs)	Low	450.40	59.82	1,341.67	19.70	243.83	10,186
	High	952.14	165.61	2,076.42	39.66	711.67	9,979
	Single	760.26	126.64	1,778.62	27.01	563.50	39,735
Burn-to-mint (hrs)	Low	105.29	0.20	521.62	0.08	5.63	8,279
	High	224.26	0.32	941.31	0.10	27.78	7,289
	Single	177.74	0.23	803.40	0.07	20.64	27,477

Figure 7: Liquidity supply on decentralized exchanges

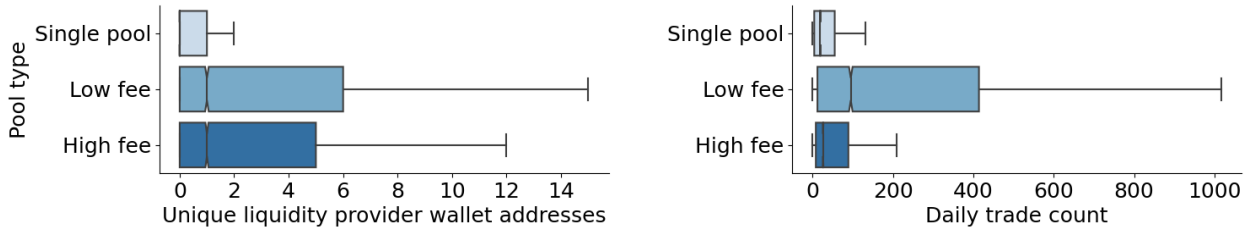
This figure plots the empirical distributions of variables in the pool-day panel, across low and high fee pools (for fragmented pairs) as well as single pools in pairs that are not fragmented. In each box plot, the median is marked as a vertical line; the box extends to the quartiles of the data set, whereas the whiskers extend to an additional 1.5 times the inter-quartile range.



(a) Pool size and trading volume



(b) Average liquidity mint and trade size



(c) Number of liquidity providers and trades

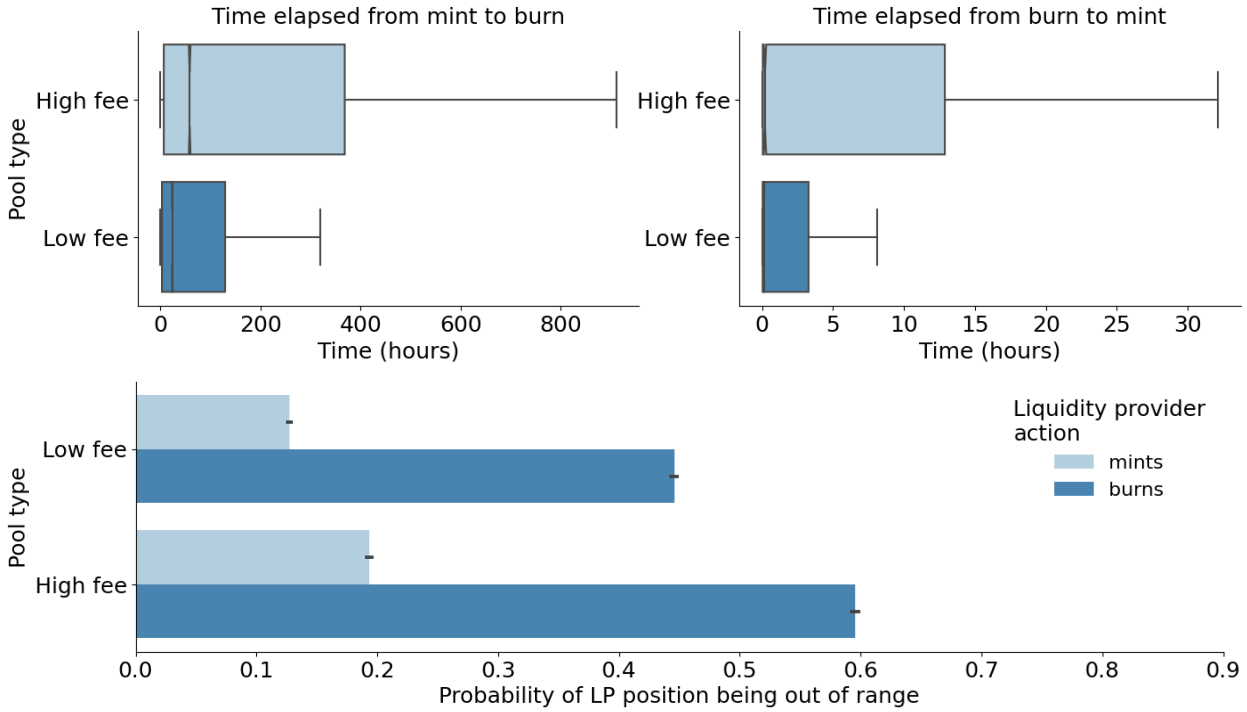
A salient observation in Table 2 is that non-fragmented pairs ("single" pools) are significantly smaller – on average less than 10% of the pool size and trading volume of fragmented pairs. Average trade and mint sizes are correspondingly lower as well. The evidence suggests that pairs for which there is significant trading interest, and therefore potentially a broader cross-section of potential liquidity providers, are more likely to become fragmented.

Figure 7 plots the distributions of our empirical measures across low- and high-fee liquidity pools. It suggest a sharp segmentation of liquidity supply and trading across pools. High-fee pools attract smaller liquidity providers by mint size, and end up with a larger *aggregate* size than their low-fee counterparts. Trading volume is similarly segmented: most small value trades are executed on the cheaper low-fee pools, making up the majority of daily volume for a given pair. High-value trades, of which there are fewer, are more likely to (also) execute on high-fee pools.

Our theoretical framework in Section 2 implies that liquidity suppliers manage their positions more actively in the low- than the high-fee pool. Figure 8 provides suggestive evidence for liquidity cycles of different lengths in the cross-section of pools. Liquidity on decentralized exchanges is significantly more passive than on traditional equity markets. That is, liquidity providers do not often manage their positions at high frequencies. The median time from a mint (deposit) to a subsequent burn (withdrawal) from the same wallet on the same pool ranges from 59.82 hours, or 2.49 days, on low-fee pools to 165.61 hours, or 6.9 days on high-fee pools.

Figure 8: Liquidity cycles on high- and low-fee pools

The top panel plots the distribution of liquidity cycle times from mint to subsequent burn (left) and from burn to subsequent mint (right) for the same **LP** wallet address in the same pool. In each box plot, the median is marked as a vertical line; the box extends to the quartiles of the data set, whereas the whiskers extend to an additional 1.5 times the inter-quartile range. The bottom panel plots the probability that the **LP** position is out of range and therefore does not earn fees. A position is considered to be "out of range" when the minimum and maximum prices at which the **LP** is willing to provide liquidity do no straddle the current price on the pool. We plot the probability separately for low- and high- fees, as well as conditional on whether the event is a burn (liquidity withdrawal) or mint (liquidity deposit).



When do **LP**s re-balance their positions? In 53% of cases, liquidity providers only withdraw tokens from the pool when their position exits the price range that allows them to collect fees. Concretely, **LP**s specified price range for liquidity provision does not straddle the most recent reference price of the pool. The scenario mirrors a limit order market where a liquidity provider's outstanding limit orders are deep in the book, such that she doesn't stand to earn the spread on the marginal incoming trade. In this case, a rational market maker might want to cancel their

outstanding order and place a new one at the top of the book. This is exactly the pattern we observe on Uniswap: the subsequent mint following a burn straddles the new price 77% of the time – **LP**s reposition their liquidity around the current prices to keep earning fees on incoming trades. Moreover, re-balancing is swift – the median time between a burn and a subsequent mint is just 12 minutes (0.20 hours).

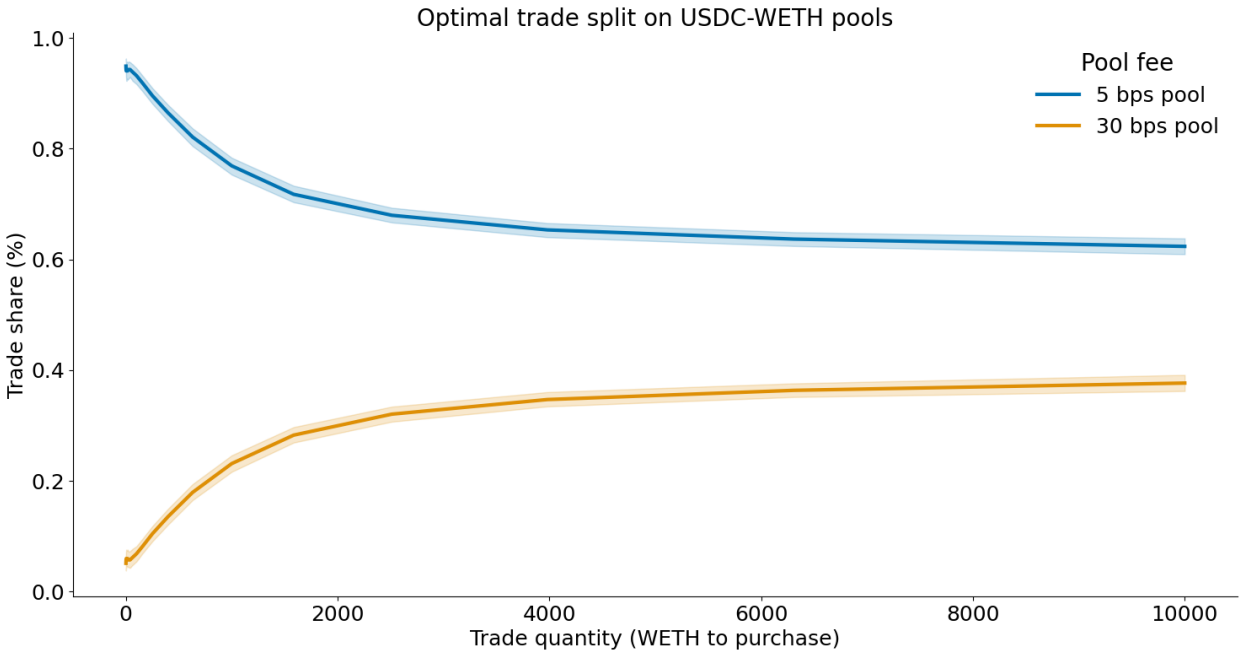
The empirical pattern in Figure 8 echoes the re-balancing cycles as described in Section 2. Liquidity providers deposit tokens in Uniswap pools to trade against uninformed order flow. They only re-balance when their position becomes out-of-range and no longer earns fees. Once this happens, **LP**s quickly adjust their position in a matter of minutes – by removing stale liquidity and adding a new position around the current price. The re-balancing cycle tends to be longer on high-fee pools, where arbitrageurs only move the price outside the range if the asset value innovation is large enough.

We note that **LP**s do not seem to "race" to update liquidity upon information arrival as in Budish, Cramton, and Shim (2015). First, they very rarely manage their position intraday. Second, **LP**s on Uniswap typically do not remove in-range liquidity that stands to trade first against incoming order flow and therefore bears the highest adverse selection risk. Our results are consistent with Capponi and Jia (2021) who theoretically argue that **LP**s have low incentives to compete with arbitrageurs on news arrival, as well as with Capponi, Jia, and Yu (2022) who find no evidence of traders racing to trade on information on Uniswap v2. In our model, **LP**s tend to re-balance their position *after* an arbitrageur executed their trade.

Next, we examine the behavior of liquidity takers (**LT**). According to our model, small orders are typically routed to low-fee pools, while larger orders are split between both low- and high-fee pools. Figure 9 provides empirical evidence supporting this claim. We use liquidity snapshot data from Kaiko on USDC-WETH pools to simulate the optimal routing strategy for trades of various sizes for the last block of each day in our sample. This simulation considers both the price impact of trades and the associated liquidity fees. In line with our model, we find that trades smaller than 150 ETH (approximately $450,000) optimally route over 90% of their size to the low-fee pool. Conversely, larger trades distribute their volume more evenly, with up to 40% being executed in high-fee pools.

Figure 9: Optimal order routing on Uniswap v3 pools

This figure displays the optimal order split for purchasing ETH using USDC across various trade sizes, on USDC-WETH Uniswap v3 pools with liquidity fees of 5 and 30 basis points. Order execution is optimized to minimize trading costs, encompassing both price impact and liquidity fees. We use liquidity distribution snapshots data from Kaiko, and focus on the last Ethereum block of each day from May 4, 2021, to July 15, 2023.

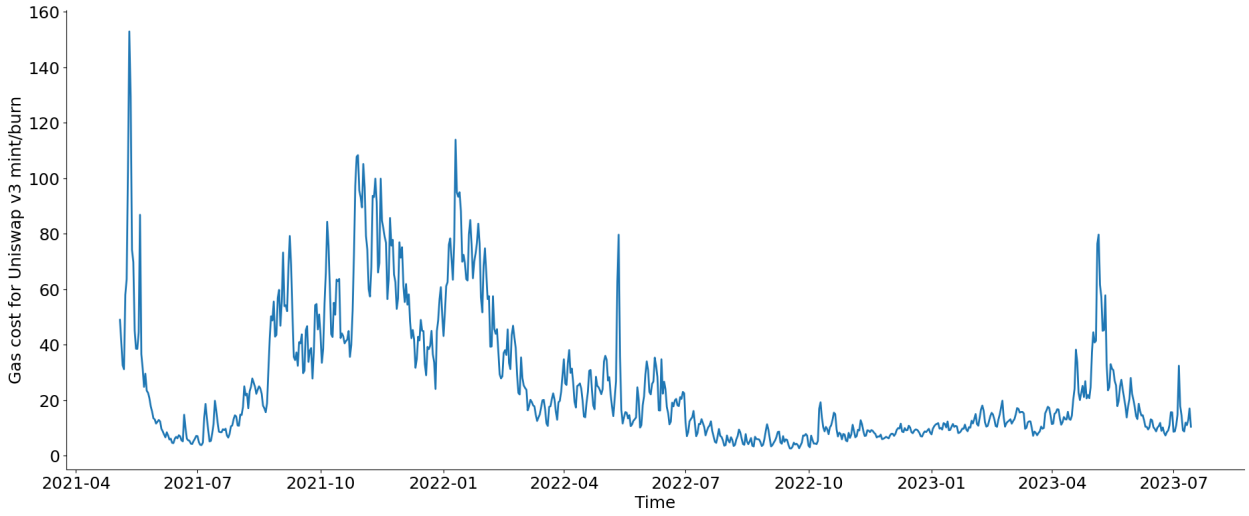


Measuring gas prices. Each interaction with smart contracts on the Ethereum blockchain requires computational resources, measured in units of "gas." Upon submitting a mint or burn transaction to the decentralized exchange, each liquidity provider specifies their willingness to pay per unit of gas, that is they bid a "gas price." Traders are likely to bid higher prices for more complex transactions or if they require a faster execution. To generate a conservative daily benchmark for the gas price, we compute the average of the lowest 1000 user gas bids for mint and burn interactions on day t, across all liquidity pools in the benchmark sample.

Figure 10 showcases the significant fluctuation in gas costs for Uniswap liquidity transactions over time. Gas costs denominated in USD are influenced by two primary factors: network congestion, which leads to variations in gas prices measured in Ether, and the fluctuation of Ether's value relative to the US dollar. On a monthly average, gas costs peaked at above US$100 in November 2021 and have since plummeted to around US$6 from the second half of 2022, albeit with occasional spikes.

Figure 10: Gas costs for Uniswap v3 mint/burn transactions

The figure illustrates the daily average gas cost on mint/burn transactions in Uniswap v3 pools. The gas cost is computed as the average of the lowest 1000 user gas bids for mint and burn interactions on each day, across all liquidity pools in the benchmark sample.



4 Empirical results

4.1 Liquidity supply on high- and low-fee pools

To formally test the model predictions and quantify the differences in liquidity supply across fragmented pools, we build a panel data set for the 32 fragmented pairs in our sample where the unit of observation is pool-day. We estimate linear regressions of liquidity and volume measures on liquidity fees and gas costs:

$$y_{ijt} = \alpha + \beta_0 d_{\textit{low-fee, }ij} + \beta_1 \text{GasPrice}_{jt} + \beta_2 \text{GasPrice}_{jt} \times d_{\textit{low-fee, }ij} + \sum \beta_k \textit{Controls}_{ijt} + \theta_j + \delta_w + \varepsilon_{ijt}, \tag{21}$$

where y is a variable of interest, i indexes liquidity pools, j runs over asset pairs, and t and w indicates days and weeks, respectively. The dummy $d_{\textit{low-fee, }ij}$ takes the value one for the pool with the lowest fee in pair j and zero else.

Further, our set of controls includes pair and week fixed effects, the log aggregate trading volume and log liquidity supply (i.e., total value locked) for day t across all pools i. Volume and liquidity are measured in US dollars. We also control for daily return volatility, computed as the range between the daily high and low prices for a given pair j (following Alizadeh, Brandt, and Diebold, 2002):

$$\text{Volatility}_{jt} = \frac{1}{2\sqrt{\log 2}} \log\left(\frac{\text{High}_{jt}}{\text{Low}_{jt}}\right). \tag{22}$$

To measure volatility for fragmented pairs that actively trade in multiple pools, we select the pool with the highest trading volume for a given day.

Consistent with Figure 7, we show in Table 3 that most of the capital deployed to provide liquidity for a given pair is locked in high-fee pools. At the same time, low-fee pools attract much larger trading volume. Models (1) and (5) show that the average low-fee pool attracts 39.5% of liquidity supply for the average pair (that is, equal to $^{(100-20.92)}/_{2}$) while it executes 62% (i.e., $^{(100+24.62)}/_{2}$) of the total trading volume. At a first glance, it would seem that a majority of capital on decentralized exchanges is inefficiently deployed in pools with low execution probability. We will show that, in line with our model, the difference is driven by heterogeneous rebalancing costs across pools, leading to the formation of **LP** clienteles.

The regression results in Table 3 support Prediction 1, stating that market share differences between pools are linked to variation in fixed transaction costs on the blockchain. A one-standard deviation increase in gas prices leads to a 4.63 percentage point increase in the high-fee liquidity share. The results suggests that blockchain transaction costs have an economically meaningful and statistically significant impact on liquidity fragmentation. In line with the theoretical model in Section 2, a jump in gas prices leads to a reshuffling of liquidity supply from low- to high-fee pools.

Evidence suggests that a higher gas price leads to a 6.52% lower volume share for the low-fee pool. This outcome is natural, as the incoming order flow is optimally routed to the high-fee pool, following the liquidity providers.

What drives the market share gap across fragmented pools? In Table 4 we document stark differences between the characteristics of individual orders supplying or demanding liquidity on pools with low and high fees. On the liquidity supply side, model (1) in Table 4 shows that the average liquidity mint is 107.5% larger on low-fee pools, which supports Prediction 2 of the model.[8] At the same time, there are 3.40 fewer unique wallets (Model 5) providing liquidity on the low-fee pool – that is, a 34% relative difference between high- and low-fee pools.

On the liquidity demand side, trades on the low-fee pool are 25.91% smaller (Model 2), consistent with Prediction 3. However, the low-fee pool executes almost three times the number of trades (i.e., trade count is 177% higher from Model 4) and has 143% higher volume than the high-fee pool (Model 3).

Next, in line with Prediction 4, low-fee pools generate a higher liquidity yield. On average, liquidity providers on low-fee pools earn 2.03 basis points higher revenue than their counterparts on high-fee pools (Model 6), indicating significant positive returns resulting from economies of scale.

Our findings (Model 7) indicate that liquidity providers on low-fee pools select price ranges that are 30% (=0.18/0.59) narrower when minting liquidity compared to those on high-fee pools. This

[8]Since all dependent variables are measured in natural logs, the marginal impact of a dummy coefficient β is computed $\left(e^{\beta}-1\right)\times 100$ percent.

Table 3: Liquidity pool market shares and gas prices

This table reports the coefficients of the following regression:

$$\text{MarketShare}_{ijt} = \alpha + \beta_0 d_{\textit{low-fee, ij}} + \beta_1 \text{GasPrice}_{jt} + \beta_2 \text{GasPrice}_{jt} \times d_{\textit{low-fee, ij}} + \sum \beta_k \textit{Controls}_{ijt} + \theta_j + \varepsilon_{ijt}$$

where the dependent variable is the liquidity or trading volume market share for pool i in asset pair j on day t. $d_{\textit{low-fee, ij}}$ is a dummy that takes the value one for the pool with the lowest fee in pair j and zero else. $\textit{GasPrice}_{jt}$ is the average of the lowest 100 bids on liquidity provision events across all pairs on day t, standardized to have a zero mean and unit variance. *Volume* is the natural logarithm of the sum of all swap amounts on day t, expressed in thousands of US dollars. *Total value locked* is the natural logarithm of the total value locked on Uniswap v3 pools on day t, expressed in millions of dollars. *Volatility* is computed as the daily range between high and low prices on the most active pool for a given pair.All regressions include pair and week fixed-effects. Robust standard errors in parenthesis are clustered by week and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from May 4, 2021 to July 15, 2023.

	Liquidity market share (%)				Volume market share (%)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$d_{\text{low-fee}}$	-20.92***	-20.92***	-20.92***	-20.94***	24.62***	24.63***	24.62***	24.71***
	(-27.42)	(-27.41)	(-27.42)	(-23.95)	(20.55)	(20.56)	(20.55)	(18.54)
Gas price × $d_{\text{low-fee}}$	-4.63***	-4.62***	-4.63***		-6.52***	-6.52***	-6.52***	
	(-7.32)	(-7.32)	(-7.32)		(-5.92)	(-5.92)	(-5.92)	
Gas price	2.31***	2.31***	2.31***		3.63***	3.61***	3.61***	
	(7.32)	(7.32)	(7.32)		(7.33)	(7.30)	(7.26)	
Volume	0.00	0.00	0.00	0.00	-0.19**	-0.20**	-0.19**	-0.12
	(0.65)	(1.33)	(0.65)	(0.66)	(-2.54)	(-2.61)	(-2.50)	(-1.56)
Total value locked	-0.00	-0.00		-0.00	0.58	0.58		0.44
	(-0.58)	(-0.06)		(-0.64)	(1.44)	(1.44)		(1.10)
Volatility	-0.29		-0.29	-0.28	-1.15***		-1.15***	-1.13**
	(-0.90)		(-0.90)	(-0.82)	(-2.74)		(-2.74)	(-2.56)
Constant	60.45***	60.46***	60.45***	60.46***	41.96***	41.99***	41.96***	41.96***
	(158.00)	(158.46)	(158.00)	(137.54)	(69.99)	(70.22)	(70.02)	(62.81)
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,288	40,288	40,288	40,288	36,059	36,059	36,059	36,059
R-squared	0.10	0.10	0.10	0.09	0.13	0.13	0.13	0.12

Robust t-statistics in parentheses. Standard errors are clustered at week level. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Fragmentation and order flow characteristics

This table reports the coefficients of the following regression:

$$y_{ijt} = \alpha + \beta_0 d_{\textit{low-fee, ij}} + \beta_1 \text{GasPrice}_{jt} d_{\textit{low-fee, ij}} + \beta_2 \text{GasPrice}_{jt} \times d_{\textit{high-fee, ij}} + \sum \beta_k \textit{Controls}_{ijt} + \theta_j + \varepsilon_{ijt}$$

where the dependent variable y_{ijt} can be (i) the log median mint size, (ii) the log median trade size, (iii) the log trading volume, (iv) the log trade count $\log(1 + \#\textit{trades})$, (v) count of unique **LP** wallets interacting with a pool in a given day, (vi) the liquidity yield in bps for pool i in asset j on day t, computed as in equation (20), and (vii) the average liquidity mint price range for pool i in asset j on day t. Price range is computed as the difference between the top and bottom of the range, normalized by the range midpoint – a measure that naturally lies between zero and two. $d_{\textit{low-fee, ij}}$ is a dummy that takes the value one for the pool with the lowest fee in pair j and zero else. $d_{\textit{high-fee, ij}}$ is defined as $1 - d_{\textit{low-fee, ij}}$. $\textit{GasPrice}_{jt}$ is the average of the lowest 100 bids on liquidity provision events across all pairs on day t, standardized to have a zero mean and unit variance. *Volume* is the natural logarithm of the sum of all swap amounts on day t, expressed in thousands of US dollars. *Total value locked* is the natural logarithm of the total value locked on Uniswap v3 pools on day t, expressed in millions of dollars. *Volatility* is computed as the daily range between high and low prices on the most active pool for a given pair. All regressions include pair and week fixed-effects. Robust standard errors in parenthesis are clustered by week and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from May 4, 2021 to July 15, 2023.

	Mint size	Trade size	Volume	# Trades	# Wallets	Liquidity yield	Price range
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$d_{\text{low-fee}}$	0.73***	-0.30***	0.89***	1.02***	-3.40***	2.03***	-0.18***
	(12.27)	(-10.05)	(14.23)	(32.95)	(-5.00)	(3.60)	(-41.84)
Gas price × $d_{\text{low-fee}}$	0.37***	0.08***	-0.03	-0.22***	-3.00***	3.57**	-0.00
	(4.96)	(3.75)	(-0.95)	(-7.29)	(-3.43)	(2.30)	(-0.47)
Gas price × $d_{\text{high-fee}}$	0.58***	0.17***	0.24***	0.07**	-2.89***	5.57***	-0.03***
	(7.52)	(8.81)	(5.95)	(2.46)	(-3.15)	(2.83)	(-4.65)
Volume	0.37***	0.16***	0.43***	0.20***	1.22***	1.01	-0.01**
	(8.68)	(21.38)	(15.27)	(13.85)	(6.56)	(0.81)	(-2.56)
Total value locked	-0.16	0.11***	0.23**	-0.01	-1.86	-13.42	-0.02
	(-1.30)	(3.54)	(1.99)	(-0.18)	(-0.99)	(-1.09)	(-0.99)
Volatility	-0.04	-0.01	-0.07	0.01	-0.09	1.18**	0.02***
	(-1.11)	(-1.34)	(-1.38)	(0.88)	(-1.03)	(2.21)	(3.98)
Constant	1.88***	1.64***	5.27***	3.26***	10.12***	10.01***	0.59***
	(58.27)	(111.47)	(168.58)	(209.84)	(28.65)	(26.04)	(184.91)
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	21,000	36,059	36,059	40,288	40,288	40,252	24,058
R-squared	0.26	0.53	0.55	0.52	0.37	0.09	0.42

Robust t-statistics in parentheses. Standard errors are clustered at week level.
*** p<0.01, ** p<0.05, * p<0.1

pattern aligns with the capability of large LPs to adjust their liquidity positions frequently, enabling more efficient capital concentration. Similarly, Caparros, Chaudhary, and Klein (2023) report a higher concentration of liquidity in pools on alternative blockchains like Polygon, known for lower transaction costs than Ethereum.

The results point to an asymmetric match between liquidity supply and demand across pools. On low-fee pools, a few **LP**s provide large chunks of liquidity for the vast majority of incoming small trades. Conversely, on high-fee pools there is a sizeable mass of small liquidity providers that mostly trade against a few large incoming trades.

How does variation in fixed transaction costs impact the gap between individual order size across pools? We find that increasing the gas price by one standard deviation leads to higher liquidity deposits on both the low- and the high-fee pools (14.2% and 30.1% higher, respectively).[9] The result supports Prediction 5 of the model. Our theoretical framework implies that a larger gas price leads to some (marginal) **LP**s switching from the low- to the high-fee pool. The switching **LP**s have low capital endowments relative to their low-fee pool peers, but higher than **LP**s on the high-fee pool. Therefore, the gas-driven reshuffle of liquidity leads to a higher average endowment on both high- and low-fee pools. Consistent with the model, a higher gas price leads to fewer active liquidity providers, particularly on low-fee pools. Specifically, a one-standard increase in gas costs leads to a significant decrease in the number of **LP** wallets interacting daily with low- and high-fee pools, respectively (Model 5).

While a higher gas price is correlated with a shift in liquidity supply, it has a muted impact on liquidity demand on low-fee pools. A higher gas cost is associated with 7.6% larger trades (Model 2), likely as traders aim to achieve better economies of scale. At the same time, the number of trades on the low-fee pool drops by 19.7% (Model 4) – since small traders might be driven out of the market. The net of gas prices effect on aggregate volume on the low-fee pool is small and not statistically significant (Model 3). The result matches our model assumption that the aggregate order flow on low-fee pool is not sensitive to gas prices.

On the high-fee pool, a higher gas price is also associated with a higher trade size, but also an increase in traded volume. As gas prices rise, liquidity providers switch from low- to high-fee pools. The outcome is greater depth and reduced price impact for liquidity demanders on high fee pools, which leads to higher trading volume.

[9]The relative effects are computed as $^{0.37}/_{(1.88+0.73)} = 13.8\%$ for low pools and $^{0.58}/_{1.88} = 30.85\%$ for high-fee pools, respectively.

Table 5: Liquidity flows and gas costs on fragmented pools

This table reports the coefficients of the following regression:

$$y_{ijt} = \alpha + \beta_0 d_{\textit{low-fee, ij}} + \beta_1 \text{GasPrice}_{jt} d_{\textit{low-fee, ij}} + \beta_2 \text{GasPrice}_{jt} \times d_{\textit{high-fee, ij}} + \sum \beta_k \textit{Controls}_{ijt} + \theta_j + \varepsilon_{ijt}$$

where the dependent variable y_{ijt} can be (i) the aggregate dollar value of mints (in logs), or (vi) a dummy variable taking value one hundred if there is at least one mint on liquidity pool i in asset j on day t. $d_{\textit{low-fee, ij}}$ is a dummy that takes the value one for the pool with the lowest fee in pair j and zero else. $d_{\textit{high-fee, ij}}$ is defined as $1 - d_{\textit{low-fee, ij}}$. $\textit{GasPrice}_{jt}$ is the average of the lowest 100 bids on liquidity provision events across all pairs on day t, standardized to have a zero mean and unit variance. *Volume* is the natural logarithm of the sum of all swap amounts on day t, expressed in thousands of US dollars. *Total value locked* is the natural logarithm of the total value locked on Uniswap v3 pools on day t, expressed in millions of dollars. *Volatility* is computed as the daily range between high and low prices on the most active pool for a given pair. All regressions include pair and week fixed-effects. Robust standard errors in parenthesis are clustered by week and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from May 4, 2021 to July 15, 2023.

	Daily mints (log US$)			Prob (at least one mint)		
	(1)	(2)	(3)	(4)	(5)	(6)
$d_{\text{low-fee}}$	0.43***	0.43***	0.43***	1.38*	1.37*	1.38*
	(6.07)	(6.07)	(6.07)	(1.71)	(1.71)	(1.71)
Gas price $\times$ $d_{\text{low-fee}}$	-0.35***	-0.35***	-0.46***	-6.02***	-6.01***	-4.58***
	(-8.50)	(-8.50)	(-7.14)	(-9.13)	(-9.13)	(-6.76)
Gas price $\times$ $d_{\text{high-fee}}$	0.11**	0.11**		-1.43**	-1.43**	
	(2.15)	(2.15)		(-2.57)	(-2.57)	
Volume	0.26***	0.26***	0.26***	0.96***	0.96***	0.96***
	(14.78)	(14.77)	(14.78)	(3.93)	(3.93)	(3.93)
Total value locked	-0.07	-0.07	-0.07	1.47	1.47	1.47
	(-0.78)	(-0.78)	(-0.78)	(1.01)	(1.00)	(1.01)
Volatility	-0.01		-0.01	0.26		0.26
	(-0.68)		(-0.68)	(0.59)		(0.59)
Gas price			0.11**			-1.43**
			(2.15)			(-2.57)
Constant	2.61***	2.61***	2.61***	51.44***	51.43***	51.44***
	(73.46)	(73.49)	(73.46)	(126.67)	(127.28)	(126.67)
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,288	40,288	40,288	40,288	40,288	40,288
R-squared	0.47	0.47	0.47	0.28	0.28	0.28

Robust t-statistics in parentheses. Standard errors are clustered at week level.
*** p<0.01, ** p<0.05, * p<0.1

In Table 5, we shift the analysis from individual orders to aggregate daily liquidity flows to Uniswap pools. We find that higher gas prices lead to a decrease in liquidity inflows, but only on the low fee pools. A one standard deviation increase in gas prices leads to a 29.5% drop in new liquidity

deposits by volume (Model 1) and an 6.02% drop in probability of having at least one mint (Model 4) on the low-fee pool. However, the slow-down in liquidity inflows is less evident in high fee pools. While an increase in gas prices reduce the probability of liquidity inflows by 1.43%, it actually leads to a 11.6% increase in the daily dollar inflow to the pool. Together with the result in Table 4 that the size of individual mints increases with gas prices, our evidence is consistent with the model implication that higher fixed transaction costs change the composition of liquidity supply on the high-fee pool, with small **LP** being substituted by larger **LP**s switching over from the low-fee pool.

4.2 Re-balancing activity on high- and low-fee pools

Next, we test Prediction 6 on the duration of liquidity re-balancing cycles on fragmented pools. Since the descriptive statistics in Table 2 suggest that **LP**s manage their positions over multiple days, we cannot accurately measure liquidity cycles in a pool-day panel. Instead, we use intraday data on liquidity events (either mints or burns) to measure the duration between two consecutive opposite-sign interactions by the same Ethereum wallet with a liquidity pool: either a mint followed by a burn, or vice-versa.

To ensure consistency with the model described in Section 2, we conduct our analysis on the entire sample as well as on a sub-sample focused solely on re-balancing events where the liquidity position falls out of range (i.e., the price range set by the **LP** does not straddle the current price and therefore the **LP** does not earn fees). We further introduce wallet fixed effects to soak up variation in reaction times across traders, and winsorize the liquidity cycle duration at the 1% level to mitigate the influence of extreme values.

Table 6 presents the results. Liquidity updates on decentralized exchanges are very infrequent, as times elapsed between consecutive interactions are measured in days or even weeks. In line with Prediction 6, we find evidence for shorter liquidity cycles on low-fee pools. The average time between consecutive mint and burn orders is 22.05% shorter on the low-fee pool (from Model 2, the relative difference is 112.42 hours/509.19 hours).

We repeat the analysis above with burn-to-mint times as the dependent variables. The burn-to-mint time measures the speed at which **LP**s deposit liquidity at updated prices after removing (out-of-range) positions. Our findings reveal that **LP**s in low-fee pools replenish liquidity 63% faster than those in high-fee pools. This supports the notion that **LP**s in low-fee environments are larger, more sophisticated market participants.

Table 6: Liquidity cycles on fragmented pools

This table reports the coefficients of the following regression:

$$y_{ijtk} = \alpha + \beta_0 d_{\textit{low-fee, ij}} + \beta_1 \text{GasPrice}_{jt} d_{\textit{low-fee, ij}} + \beta_2 \text{GasPrice}_{jt} \times d_{\textit{high-fee, ij}} + \sum \beta_k \textit{Controls}_{ijt} + \theta_j + \varepsilon_{ijt}$$

where the dependent variable y_{ijt} can be (i) the mint-to-burn time, (ii) the burn-to-mint time, measured in hours, for a transaction initiated by wallet k on day t and pool i trading asset j. The mint-to-burn and burn-to-mint times are computed for consecutive interactions of the same wallet address with the liquidity pool. $d_{\textit{low-fee, ij}}$ is a dummy that takes the value one for the pool with the lowest fee in pair j and zero else. $d_{\textit{high-fee, ij}}$ is defined as $1 - d_{\textit{low-fee, ij}}$. $\textit{GasPrice}_{jt}$ is the average of the lowest 100 bids on liquidity provision events across all pairs on day t, standardized to have a zero mean and unit variance. *Volume* is the natural logarithm of the sum of all swap amounts on day t, expressed in thousands of US dollars. *Total value locked* is the natural logarithm of the total value locked on Uniswap v3 pools on day t, expressed in millions of dollars. *Volatility* is computed as the daily range between high and low prices on the most active pool for a given pair. *Position out-of-range* is a dummy taking value one if the position being burned or minted is out of range, that is if the price range selected by the **LP** does not straddle the current pool price. All variables are measured as of the time of the second leg of the cycle (i.e., the burn of a mint-burn cycle). All regressions include pair, week, and trader wallet fixed-effects. Robust standard errors in parenthesis are clustered by day and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from May 4, 2021 to July 15, 2023.

	Mint-burn time (hours)				Burn-mint time (hours)	
	Out-of-range positions		Full sample			
	(1)	(2)	(3)	(4)	(5)	(6)
$d_{\text{low-fee}}$	-110.94***	-112.42***	-99.74***	-100.17***	-157.95***	-159.71***
	(-7.49)	(-7.69)	(-8.86)	(-8.94)	(-10.59)	(-10.81)
Gas price $\times$ $d_{\text{low-fee}}$	-14.27	-6.54	-16.65**	-15.41*	-11.29	2.95
	(-1.49)	(-0.68)	(-2.13)	(-1.98)	(-1.65)	(0.40)
Gas price $\times$ $d_{\text{high-fee}}$	-19.57**	-12.83	-14.44**	-13.42*	-10.52*	1.96
	(-2.34)	(-1.57)	(-2.04)	(-1.89)	(-1.69)	(0.32)
Volume		-16.71***		-5.87		-24.84***
		(-3.24)		(-1.15)		(-4.10)
Total value locked		-35.14		-53.17*		-12.71
		(-1.05)		(-1.70)		(-0.52)
Volatility		-3.48**		-2.11***		-2.99***
		(-2.49)		(-2.75)		(-3.36)
Constant	509.19***	509.66***	497.18***	497.00***	248.00***	250.13***
	(61.93)	(58.34)	(91.65)	(90.60)	(29.91)	(30.27)
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes
Trader wallet FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	215,454	215,454	405,586	405,584	265,848	265,848
R-squared	0.87	0.87	0.82	0.82	0.37	0.37

Robust t-statistics in parentheses. Standard errors are clustered at week level.
*** p<0.01, ** p<0.05, * p<0.1

4.3 Adverse selection costs across low- and high-fee pools

Finally, we test Prediction 7 of our model, which states that **LP** on the low-fee pool face higher adverse selection costs. Our main metric for informational costs is the *loss-versus-rebalancing* (LVR), as defined in Milionis, Moallemi, Roughgarden, and Zhang (2023). The measure is equivalent to the adverse selection component of the bid-ask spread in equity markets. To calculate it, for each swap j exchanging Δx_j for Δy_j in a pool with assets x and y, we use:

$$\text{LVR}_j = d_j \times \Delta x_j (p_{\text{swap},j} - p'_j), \tag{23}$$

where d_j is one for a "buy" trade ($\Delta x_j < 0$) and minus one for a "sell" trade ($\Delta x_j > 0$). The effective swap price is $p_{\text{swap},j} = -\frac{\Delta y_j}{\Delta x_j}$, and p'_j represents a benchmark price.

We use two benchmark prices p'_j in our analysis. The first, $p'_j = p_j^{\Delta t=0}$, is the pool's equilibrium price immediately after a swap. The resulting LVR metric captures both temporary and permanent price impact, driven by uninformed and informed trades, respectively, and represents an upper bound for **LP**'s adverse selection cost.

The second benchmark is the liquidity-weighted average price across Uniswap v3 pools, measured with a one-hour delay after the swap ($p'_j = p_j^{\Delta t=1h}$). This approach assumes that any price deviations caused by uninformed liquidity trades are corrected by arbitrageurs within an hour, as supported by Lehar and Parlour (forthcoming). Thus, the LVR metric derived using this benchmark captures only the permanent price impact, a more precise measure of adverse selection cost for liquidity providers.[10]

To compute LVR for each day t and liquidity pool i, we aggregate the loss-versus-balancing for each swap within a day. We subsequently winsorize our measures at the 0.5% and 99.5% quantiles to remove extreme outliers. The resulting sum is normalized by dividing it by the total value locked (TVL) in the pool at day's end:

$$\text{LVR}_{i,t} = \frac{\sum_j \text{LVR}_{j,i,t}}{\text{TVL}_{i,t}}, \tag{24}$$

which ensures that the LVR metric is comparable across pools trading different token pairs.

We complement our analysis with the calculation of *impermanent loss* (IL), an additional metric for assessing adverse selection costs. Impermanent loss is defined as the negative return from providing liquidity compared to simply holding the assets outside the exchange and marking them to market as prices change (see, for example, Aoyagi, 2020; Barbon and Ranaldo, 2021).

[10] Our methodology is equivalent to the one in Milionis, Moallemi, Roughgarden, and Zhang (2023) under two assumptions. First, liquidity providers can re-balance their position following each swap. Second, we assume that our two benchmarks for p'_j, derived from decentralized exchange data, closely track the fundamental value of the token. This perspective aligns with Han, Huang, and Zhong (2022), who also note that centralized exchange prices are subject to manipulative practices such as wash trading. Further, our selection of benchmarks reflects the fact that our sample includes several token pairs not traded on major centralized exchanges such as Binance.

The key distinction between IL and loss-versus-rebalancing (LVR) measures lies in their assumptions about liquidity providers' strategies (Milinois, Moallemi, Roughgarden, and Zhang, 2023). While LVR assumes that providers actively re-balance their holdings by mirroring decentralized exchange trades on centralized exchanges at the fundamental value to hedge market risk, IL is based on a more passive approach where providers maintain their positions without active re-balancing. Loss-versus-rebalancing is a function of the entire price path, reflecting constant rebalancing by liquidity providers. In contrast, impermanent loss is determined solely by the initial and final prices of the assets.

We obtain hourly liquidity snapshots from the Uniswap V3 Subgraph to calculate impermanent loss for a theoretical symmetric liquidity position. This position is set within a price range of $\left[\frac{1}{\alpha}p, \alpha p\right]$, centered around the current pool price p, with α set to 1.05. We set a one-hour horizon to measure changes in position value, aligning with the time horizon used for the LVR metric. In Appendix E, we present the exact formulas for calculating impermanent loss on Uniswap V3, based on the methodology described by Heimbach, Schertenleib, and Wattenhofer (2022).

Table 7 presents our empirical results. In line with Milionis, Moallemi, and Roughgarden (2023), all price impact measures — that is, the immediate and one-hour horizon LVR and the impermanent loss — are significantly larger in low-fee pools. This indicates that a higher liquidity fee indeed acts as barrier to arbitrageurs. Specifically, the permanent price impact, measured by the one-hour horizon LVR, is 6.39 basis points or 81% larger in low-fee pools. The total price impact, represented by the after-swap LVR metric, is 3.5 times larger in low-fee pools. The wide gap between permanent and total price impact highlights the substantially higher volume of uninformed trading in low-fee pools. Our secondary measure of adverse selection, the impermanent loss at 5% around the current price, is also 15% higher on low- than high-fee pools.

We note that an increase in gas prices leads to a 3.51 wider gap in total price impact but a 0.75 bps narrower gap in permanent price impact between the high- and low-fee pools. The result suggests that a higher gas price primarily discourages uninformed traders, rather than arbitrageurs, from trading on high-fee pools.

Finally, in Models (7) and (8) of Table 7 we explore whether various arbitrage frictions lead to price discrepancies between high- and low-fee pools. For this purpose, we collect hourly price data from Binance for the largest four pairs by trading volume: WBTC-WETH, USDC-WETH, WETH-USDT, and USDT-USDC. We subsequently compute daily averages of hourly price deviations between centralized and decentralized exchanges. The analysis reveals that the average hourly price deviation across centralized and decentralized exchanges is 0.60%. Notably, there is no significant difference in price deviations between low- and high-fee pools. The result suggests that arbitrage activities remain efficient despite the differences in trading costs between these pools.

Table 7: Adverse selection costs on high- and low-fee pools

This table presents regression results that analyze adverse selection costs in fragmented Uniswap v3 pools. For columns (1) through (4), the dependent variable is loss-versus-rebalancing (LVR), as defined in equation (24). We use the one-hour horizon benchmark ($p_j^{\Delta t=1h}$) in models (1) and (2) to measure permanent price impact, and the immediate, same-block price benchmark ($p_j^{\Delta t=0}$) in models (3) and (4) to measure total price impact. For columns (5) and (6), the dependent variable is the impermanent loss for a symmetric liquidity position at ±5% centered around the current pool price. The average impermanent loss is calculated for each day, based on Ethereum blocks mined within that day. The impermanent loss computation uses a one-hour liquidity provider horizon, comparing current pool prices with those one hour later. For columns (7) and (8), the dependent variables are the liquidity (TVL) and volume share of the pool, measured in percent. Finally, in columns (9) and (10) the dependent variable is the absolute deviation of the Uniswap pool price from Binance prices, sampled hourly, and measured in percent. $d_{low\text{-}fee,\ ij}$ is a dummy that takes the value one for the pool with the lowest fee in pair j and zero else. $GasPrice_{jt}$ is the average of the lowest 100 bids on liquidity provision events across all pairs on day t, standardized to have a zero mean and unit variance. *Volume* is the natural logarithm of the sum of all swap amounts on day t, expressed in thousands of US dollars. *Total value locked* is the natural logarithm of the total value locked on Uniswap v3 pools on day t, expressed in millions of dollars. When LVR is an explanatory variable, it is calculated using the one-hour ahead benchmark price. *Volatility* is computed as the daily range between high and low prices on the most active pool for a given pair. All regressions include pair and week fixed-effects. Robust standard errors in parenthesis are clustered by week, and ***, **, and * denote the statistical significance at the 1, 5, and 10% level, respectively. The sample period is from May 4, 2021 to July 15, 2023.

	LVR (1h horizon) Permanent price impact		LVR (after swap) Total price impact		Impermanent loss		CEX price deviation	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$d_{\text{low-fee}}$	6.39***	6.39***	29.78***	29.67***	1.08***	1.13***	0.06	0.04
	(16.57)	(17.05)	(14.86)	(14.95)	(5.72)	(6.18)	(1.51)	(1.33)
Gas price × $d_{\text{low-fee}}$		-0.75**		3.51**		-0.01		0.08
		(-2.05)		(2.10)		(-0.05)		(1.09)
Gas price		2.61***		6.16**		3.71***		-0.03
		(2.74)		(2.53)		(3.76)		(-0.28)
Volume		3.22***		8.67***		1.81***		0.22***
		(8.15)		(6.61)		(6.22)		(4.74)
Total value locked		0.53		-2.12		1.93		-0.39***
		(0.14)		(-0.34)		(0.74)		(-4.05)
Volatility		1.85***		4.23***		6.69**		1.04***
		(2.87)		(3.17)		(2.61)		(3.51)
Constant	7.85***	7.86***	8.88***	8.97***	7.37***	7.51***	0.60***	0.67***
	(40.71)	(36.89)	(8.87)	(8.88)	(77.84)	(61.32)	(30.71)	(35.75)
Pair FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40,302	40,288	40,302	40,288	40,250	40,248	5,207	5,207
R-squared	0.14	0.15	0.09	0.10	0.09	0.11	0.10	0.11

Robust t-statistics in parentheses. Standard errors are clustered at week level.
*** p<0.01, ** p<0.05, * p<0.1

Figure 11: Price impact and price deviations across high- and low-fee pools

This figure plots the average total and permanent price impact, liquidity yield, and price deviation from centralized exchanges across low and high fee pools for fragmented pairs.

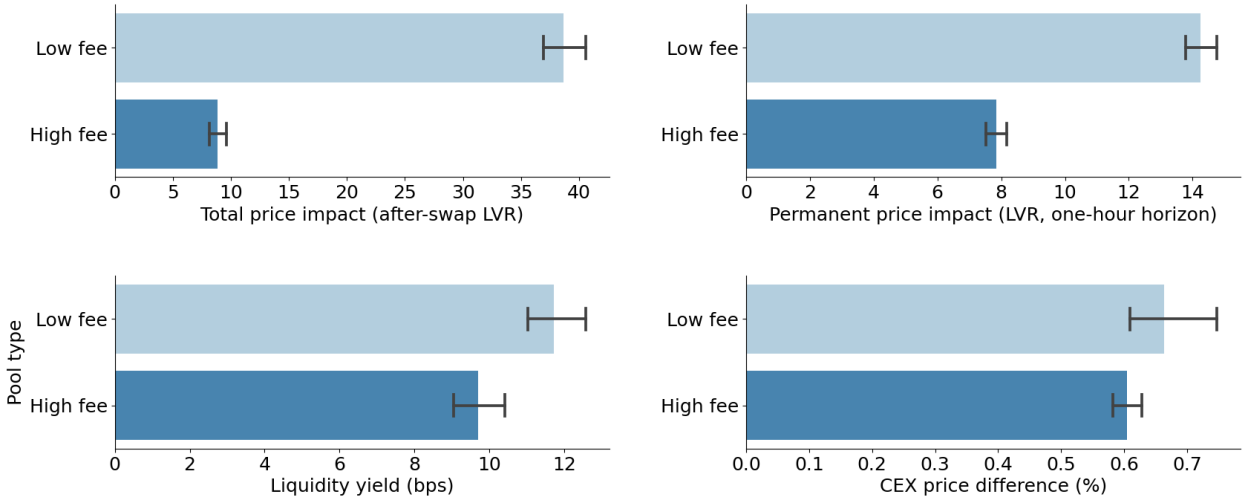


Figure 11 graphically illustrates the result, contrasting permanent price impact measures against the liquidity yield, as calculated in equation (20). Notably, before accounting for gas costs, we observe that liquidity providers in low-fee pools experience losses on average: the average daily permanent price impact in these pools is 14.21 basis points, which exceeds the fee revenue of 11.71 bps. In contrast, liquidity providers in high-fee pools approximately break even before considering gas costs: the fee revenue amounts to 9.69 bps, which is slightly higher than the permanent price impact of 7.85 bps. One should keep in mind, however, that the magnitude of losses from adverse selection depends on the horizon at which we measure the loss-versus-rebalancing.

5 Conclusion

This paper argues that fixed costs associated with liquidity management drive a wedge between large (institutional) and small (retail) market makers. In the context of blockchain-based decentralized exchanges, the most evident fixed cost is represented by gas fees, where market makers compensate miners and validators for transaction processing in proof-of-work, respectively in proof-of-stake blockchains. Innovative solutions such as Proof of Stake (PoS) consensus algorithms and Layer 2 scaling aim to address the concern of network costs. However, even if gas fees were eliminated entirely, individual retail traders still encounter disproportionate fixed costs in managing their liquidity, such as the expenditure of time and effort.

Our paper highlights a trade-off between capital efficiency and the fixed costs of active management. During the initial phase of decentralized exchanges, such as Uniswap V2, liquidity providers

were not able to set price limits, resulting in an even more passive liquidity supply and fewer incentives for active position management. However, the mechanism implied that incoming trades incurred significant price impact. To enhance the return on liquidity provision and reduce price impact on incoming trades, modern decentralized exchanges (DEXs) have evolved to enable market makers to fine-tune their liquidity positions, albeit at the expense of more active management.

We show, both theoretically and empirically, that fixed costs of liquidity management promote market fragmentation across decentralized pools and generate clienteles of liquidity providers. Large market makers, likely institutions and funds, have stronger economies of scale and can afford to frequently manage their positions on very active low fee markets, while bearing higher adverse selection risk. On the other hand, smaller retail liquidity providers become confined to high fee markets with scant activity, trading off a lower execution probability against reduced adverse selection and lower gas costs to update their positions. Since large liquidity providers can churn their position at a faster pace, two thirds of the trading volume interacts with less than half the capital locked on Uniswap V3.

Our findings indicate that substantial fixed costs can hinder the participation of small market makers in the forefront of liquidity provision, where active order management is crucial. Instead, smaller liquidity providers tend to operate on the market maker "fringe," opting for a lower execution probability in exchange for better prices. The results are particularly relevant the context of a resurgence in retail trading activity and the ongoing evolution of technology that fosters market structures aimed at enhancing broader access to financial markets.

References

Adams, Hayden, Noah Zinsmeister, Moody Salem, River Keefer, and Dan Robinson, 2021, Uniswap v3 core, .

Alizadeh, Sassan, Michael W. Brandt, and Francis X. Diebold, 2002, Range-based estimation of stochastic volatility models, *The Journal of Finance* 57, 1047–1091.

Aoyagi, Jun, 2020, Liquidity Provision by Automated Market Makers, *Working paper*.

——— , and Yuki Ito, 2021, Coexisting Exchange Platforms: Limit Order Books and Automated Market Makers, *Working paper*.

Aspris, Angelo, Sean Foley, Jiri Svec, and Leqi Wang, 2021, Decentralized exchanges: The "wild west" of cryptocurrency trading, *International Review of Financial Analysis* 77, 101845.

Augustin, Patrick, Roy Chen-Zhang, and Donghwa Shin, 2022, Reaching for yield in decentralized financial markets, LawFin Working Paper No. 39. Available at SSRN: https://ssrn.com/abstract=4063228.

Barbon, Andrea, and Angelo Ranaldo, 2021, On the quality of cryptocurrency markets: Centralized versus decentralized exchanges, .

Battalio, Robert, Shane A. Corwin, and Robert Jennings, 2016, Can brokers have it all? on the relation between make-take fees and limit order execution quality, *The Journal of Finance* 71, 2193–2237.

Brolley, Michael, and David Cimon, 2020, Order flow segmentation, liquidity and price discovery: The role of latency delays, *Journal of Financial and Quantitative Analysis* 55, 2555–2587.

Brolley, Michael, and Marius Zoican, 2023, On-demand fast trading on decentralized exchanges, *Finance Research Letters* p. 103350.

Budish, Eric, Peter Cramton, and John Shim, 2015, The high-frequency trading arms race: Frequent batch auctions as a market design response, *Quarterly Journal of Economics* 130, 1547–1621.

Caparros, Basile, Amit Chaudhary, and Olga Klein, 2023, Blockchain scaling and liquidity concentration on decentralized exchanges, Available at SSRN: https://ssrn.com/abstract=4475460 or http://dx.doi.org/10.2139/ssrn.4475460.

Capponi, Agostino, and Ruizhe Jia, 2021, The Adoption of Blockchain-based Decentralized Exchanges, *Working paper*.

——— , and Shihao Yu, 2022, The Information Content of Blockchain Fees, *Working paper*.

——— , 2023, Price discovery on decentralized exchanges, .

Capponi, Agostino, Ruizhe Jia, and Brian Zhu, 2024, The paradox of just-in-time liquidity in decentralized exchanges: More providers can sometimes mean less liquidity, .

Cimon, David A., 2021, Broker routing decisions in limit order markets, *Journal of Financial Markets* 54, 100602.

Foucault, Thierry, Ohad Kadan, and Eugene Kandel, 2013, Liquidity cycles and make/take fees in electronic markets, *Journal of Finance* 68, 299–341.

Foucault, Thierry, and Albert J. Menkveld, 2008, Competition for order flow and smart order routing systems, *The Journal of Finance* 63, 119–158.

Han, Jianlei, Shiyang Huang, and Zhuo Zhong, 2022, Trust in defi: An empirical study of the decentralized exchange, .

Hasbrouck, Joel, Thomas Rivera, and Fahad Saleh, 2022, The Need for Fees at a DEX: How Increases in Fees Can Increase DEX Trading Volume, *Working paper*.

Heimbach, Lioba, Eric Schertenleib, and Roger Wattenhofer, 2022, Risks and Returns of Uniswap V3 Liquidity Providers, Papers 2205.08904 arXiv.org.

Korajczyk, Robert A, and Dermot Murphy, 2018, High-Frequency Market Making to Large Institutional Trades, *The Review of Financial Studies* 32, 1034–1067.

Lehar, Alfred, and Christine Parlour, forthcoming, Decentralized Exchange: The Uniswap automated market maker, *Journal of Finance*.

Milinois, Jason, Ciamac Moallemi, Tim Roughgarden, and Anthony Lee Zhang, 2023, Automated market making and loss-versus-rebalancing, Manuscript.

Milionis, Jason, Ciamac C. Moallemi, and Tim Roughgarden, 2023, Automated market making and arbitrage profits in the presence of fees, .

O'Hara, Maureen, and Mao Ye, 2011, Is market fragmentation harming market quality?, *Journal of Financial Economics* 100, 459–474.

Pagano, Marco, 1989, Trading volume and asset liquidity, *The Quarterly Journal of Economics* 104, 255–274.

Pagnotta, Emiliano, and Thomas Philippon, 2018, Competing on speed, *Econometrica* 86, 1067–1115.

Park, Andreas, 2022, Conceptual flaws of decentralized automated market making, *Working paper, University of Toronto*.

Wan, Xin, and Austin Adams, 2022, Just-in-time liquidity on the uniswap protocol, Working paper.

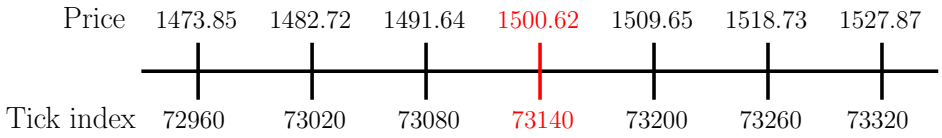
Zhu, Haoxiang, 2014, Do dark pools harm price discovery?, *Review of Financial Studies* 27, 747–789.

A Liquidity provision mechanism on Uniswap v3

In this appendix, we walk through a numerical example to illustrate the mechanism of liquidity provision and trading on Uniswap V3 liquidity pools. To facilitate understanding, we highlight the similarities and differences between the Uniswap mechanism and the familiar economics of a traditional limit order book.

Let $p_c = 1500.62$ be the current price of the ETH/USDT pair. Traders can provide liquidity on Uniswap V3 pools at prices on a log-linear tick space. In particular, consecutive prices are always θ basis point apart: $p_i = 1.0001^{\theta i}$, where θ is the tick spacing. For the purpose of the example, we take $\theta = 60$. Consequently, the current price of 1500.62 corresponds to a tick index of $c = 73140$. Figure A.1 illustrates three ticks on grid below and above the current price of ETH/USDT 1500.62.

Figure A.1: ETH/USDT price grid around p_c



Two-sided liquidity provision. Trader **A** starts out with a capital of USDT 20,000 and wants to provide liquidity over the price range $[1491.64, 1527.87]$, a range which spans four ticks. Liquidity provision over a range that includes the current price corresponds to posting quotes on both the bid and ask side of a traditional limit order book, where the current price of the pool corresponds to the mid-point of the book.

1. *Bid quotes:* trader **A** deposits USDT over the price range $[1491.64, 1500.62)$. This action is equivalent to submitting a buy limit order with a bid price of 1491.64. An incoming Ether seller can swap their ETH for the USDT deposited by **A**, generating price impact until the limit price of 1491.64 is reached.

2. *Ask quotes:* at the same time, trader **A** deposits ETH over three ticks: $[1500.62, 1509.65)$, $[1509.65, 1518.73)$, and $[1518.73, 1527.87)$. The action corresponds to submitting *three* sell limit orders with ask prices 1509.65, 1518.73, and 1527.87, respectively. Incoming Ether buyers can swap USDT for trader **A**'s ETH.

In the Uniswap V3 protocol, deposit amounts over each tick $[p_i, p_{i+1})$ must satisfy

$$\text{ETH deposit over } [p_i, p_{i+1}): \; x_i = L\left(\frac{1}{\sqrt{p_i}} - \frac{1}{\sqrt{p_{i+1}}}\right) \tag{A.1}$$

$$\text{USDT deposit over } [p_i, p_{i+1}): \; y_i = L\left(\sqrt{p_{i+1}} - \sqrt{p_i}\right), \tag{A.2}$$

where L ("liquidity units") is a scaling factor proportional to the capital committed to the liquidity position. The scaling factor L is pinned down by setting the total committed capital equal to the sum of the positions (in USDT), that is $p_c \sum_i x_i + \sum_i y_i$. In our example,

$$1500.62 \times L_A \times \left(\frac{1}{\sqrt{1500.62}} - \frac{1}{\sqrt{1527.87}} \right) + L_A \times \left(\sqrt{1500.62} - \sqrt{1491.64} \right) = 20000, \tag{A.3}$$

leading to $L_A = 43188.6$. We the value of L_A into (A.1) and conclude that trader **A** deposits 5,013.38 USDT over $[1491.64, 1500.62)$ and ETH 9.99 over $[1500.62, 1527.87)$ (approximately ETH 3.33 over each tick size covered).

One-sided liquidity provision. Trader **B** has USDT 20,000 and wants to post liquidity over the range $[1509.65, 1527.87)$, which does not include the current price. This action corresponds to posting ask quotes to sell ETH deep in the book, at price levels 1518.73 and 1527.87. Liquidity is not "active" – that is, the quotes are not filled – until the existing depth at 1509.65 is consumed by incoming trades.

We use equation (A.1) to solve for the amount of liquidity units provided by **B**:

$$1500.62 \times L_B \times \left(\frac{1}{\sqrt{1509.65}} - \frac{1}{\sqrt{1527.87}} \right) = 20000, \tag{A.4}$$

which leads to $L_B = 86589.4$. Trader **B** deposits 6.67 ETH on each of the two ticks covered by the chosen range.

Figure A.2: ETH/USDT pool state after liquidity provision choices

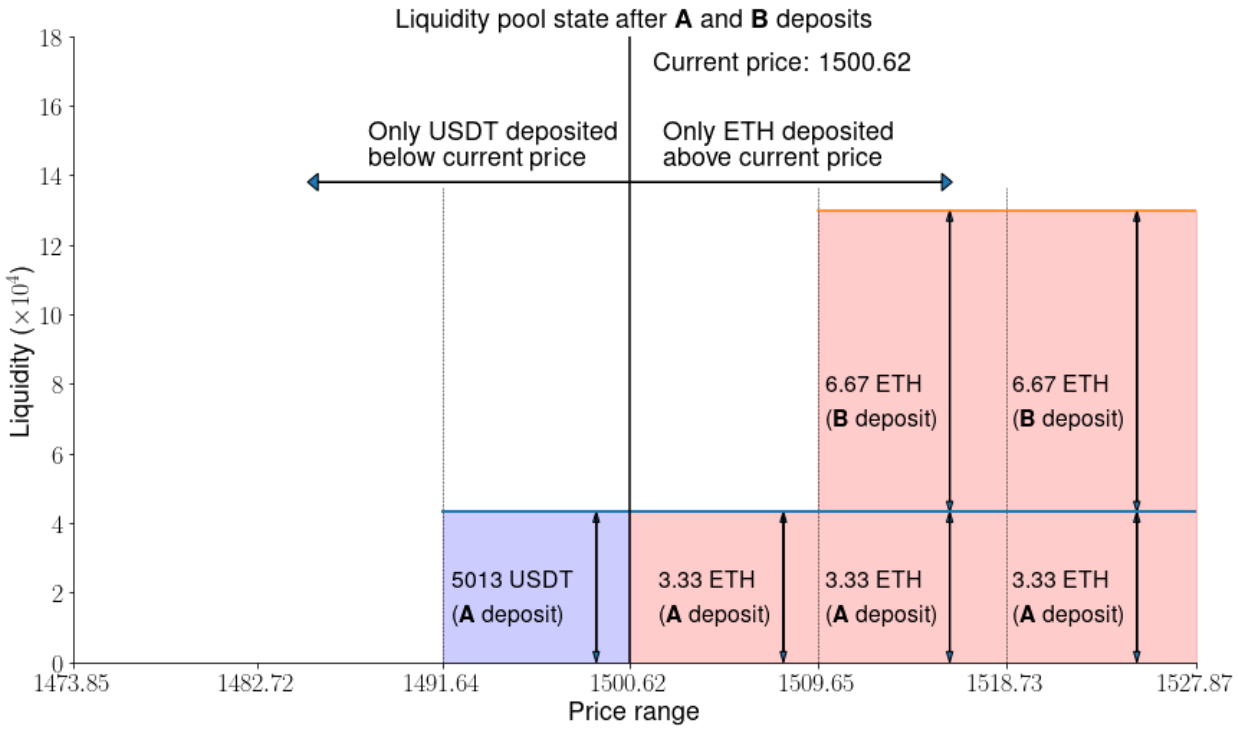


Figure A.2 illustrates market depth after **A** and **B** deposit liquidity in the pool. The current price of the pool is equivalent to a midpoint in traditional limit order markets. The "ask side" of the pool is deeper, consistent with both liquidity providers choosing ranges skewed towards prices above the current midpoint. Liquidity is uniformly provided over ticks – that is, each trader deposits an equal share of their capital at each price tick covered by their price range.

Trading, fees, and price impact. Suppose now that a trader **C** wants to buy 10 ETH from the pool. For each tick interval $[p_i, p_{i+1})$, price impact is computed using a constant product function over virtual reserves:

$$\underbrace{\left(x + \frac{L}{\sqrt{p_{i+1}}}\right)}_{\text{Virtual ETH reserves}} \underbrace{(y + L\sqrt{p_i})}_{\text{Virtual USDT reserves}} = L^2, \tag{A.5}$$

where x and y are the actual ETH and USDT deposits in that tick range, respectively. Virtual reserves are just a mathematical artifact: they extend the physical (real) deposits as if liquidity would be uniformly distributed over all possible prices on the real line. Working with constant product functions over real reserves is not feasible: in our example, the product of real reserves is zero throughout the order book (since only one asset is deposited in each tick range).

Let $\tau = 1\%$ denote the pool fee that serves as an additional compensation for liquidity providers. That is, if the buyer pays to pay Δy USDT to purchase a quantity Δx ETH, he needs to effectively pay $\Delta y\,(1+\tau)$. As per the Uniswap V3 white paper, liquidity fees are not automatically deposited back into the pool.

1. **Tick 1:** $[1500.62, 1509.65)$. Trader **C** first purchases 3.33 ETH at the first available tick above the current price (equivalent to the "best ask"). To remove the ETH, he needs to deposit Δy_1 USDT, where Δy_1 solves:

$$\left(3.33 - 3.33 + \frac{L_A}{\sqrt{1509.65}}\right)\left(0 + \Delta y_1 + L_A\sqrt{1500.62}\right) = L_A^2, \tag{A.6}$$

   which leads to $\Delta y_1 = 5026.19$ USDT. Trader **C** pays an average price of 50216.19/3.33=1507.86 USDT for each unit of ETH purchased. Further, he pays a fee of 50.26 USDT to liquidity provider **A** (the only liquidity provider at this tick).

   The new current price is given by the ratio of virtual reserves,

$$p' = \frac{\Delta y_1 + L_A\sqrt{1500.62}}{3.33 - 3.33 + \frac{L_A}{\sqrt{1509.65}}} = 1509.65, \tag{A.7}$$

   that is the next price on the tick grid since **C** exhausts the entire liquidity on $[1500.62, 1509.65)$.

2. **Tick 2:** $[1509.65, 1518.73)$. Trader **C** still needs to purchase 6.67 ETH at the next tick level (where the depth is 10 ETH). The liquidity level at this tick is $L_A + L_B$, that is the sum of liquidity provided by **A** and **B**. To remove the 6.67 ETH from the pool, he needs to deposit Δy_2, where

$$\left(10 - 6.67 + \frac{L_A + L_B}{\sqrt{1518.73}}\right)\left(0 + \Delta y_2 + (L_A + L_B)\sqrt{1509.65}\right) = (L_A + L_B)^2. \tag{A.8}$$

   It follows that trader **C** purchases 6.67 ETH by depositing $\Delta y_2 = 10089.12$ USDT, at an average price of 1512.61. The pool price is updated as the ratio of virtual reserves:

$$p'' = \frac{\Delta y_2 + (L_A + L_B)\sqrt{1509.65}}{10 - 6.67 + \frac{L_A + L_B}{\sqrt{1518.73}}} = 1515.7. \tag{A.9}$$

   The updated price is in between the two liquidity ticks, since not all depth on this tick level was exhausted in the trade. Following the swap, liquidity on the tick range $[1509.65, 1518.73)$ is composed of both assets: that is 10089.12 USDT and 10-6.67=3.33 ETH.

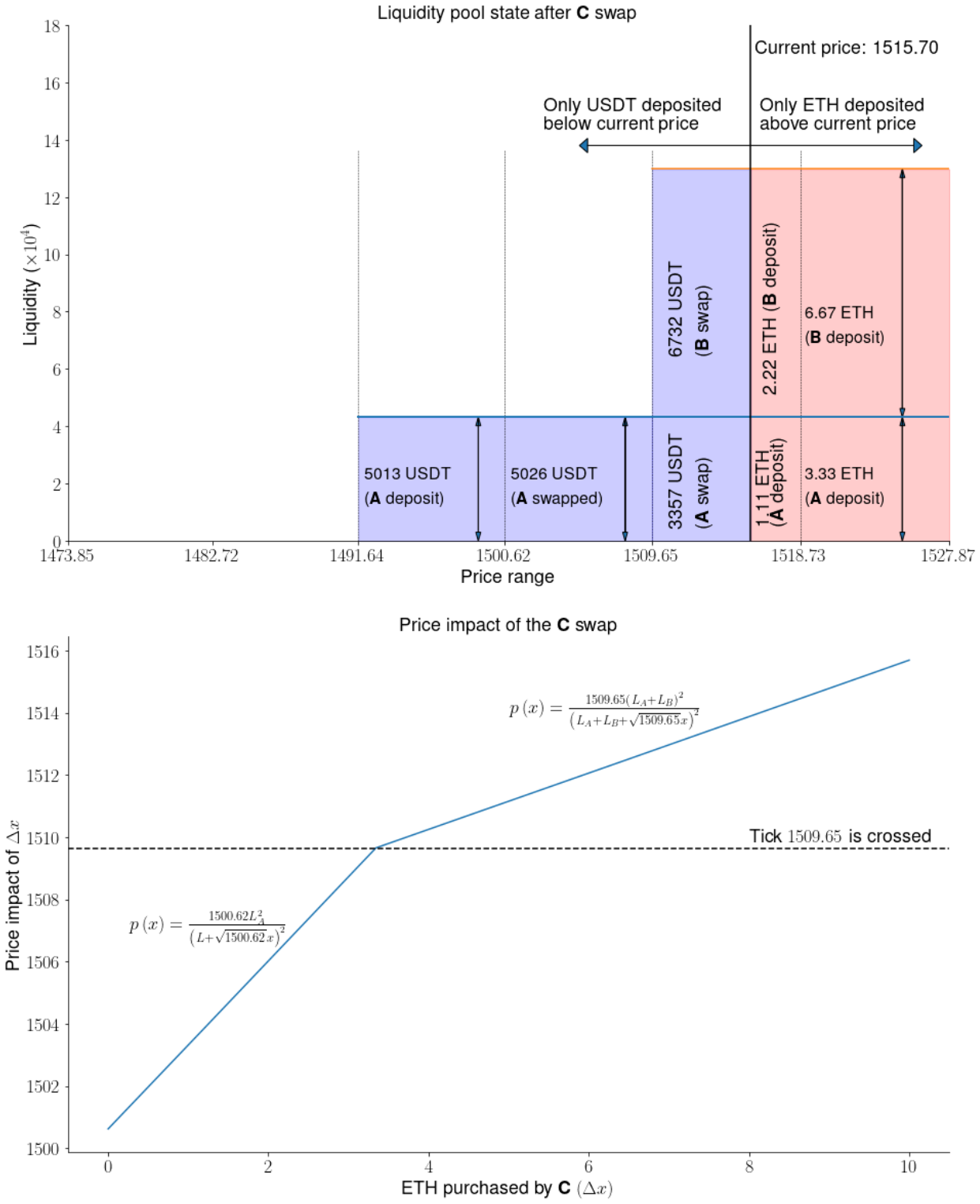
   Finally, trader C pays 100.89 USDT as liquidity fees (1% of the trade size), which are distributed to **A** and **B** proportionally to their liquidity share. That is, **A** receives a fraction

$\frac{L_A}{L_A+L_B}$ of the total fee (33.57 USDT), whereas **B** receives 67.32 USDT.

Figure A.3 illustrates the impact of the swap. Within tick $[1500.62, 1509.65)$, **A** sells 3.33 ETH and buys 5026 USDT. Unlike on limit order books, the execution does not remove liquidity from the book. Rather, **A**'s capital is converted from one token to another and remains available to trade. This feature underscores the passive nature of liquidity supply on decentralized exchanges. Mapping the concepts to traditional limit order book, this mechanism would imply that every time a market maker's sell order is executed at the ask, a buy order would automatically be placed on the bid side of the market.

The final price of 1517.70 lies within the tick $[1509.65, 1518.73)$, rather than on its boundary. Trader **C** only purchases 6.66 ETH out of 10 ETH available within this price interval. The implication is that liquidity on $[1509.65, 1518.73)$ contains both tokens: 3.33 ETH (the amount that was not swapped by **C**) as well as 10089.12 USDT that **C** deposited in the pool.

Figure A.3: Swap execution and price impact



The bottom panel of Figure A.3 shows the price impact of the swap. From equation (6.15) in the Uniswap V3 white paper, we can solve for the price within tick $[p_{\min}, p_{\max})$ with liquidity L, following the execution of a buy order of size x:

$$p(x) = \frac{p_{\min} L^2}{\left(L - \sqrt{p_{\min}} x\right)^2}. \tag{A.10}$$

As expected, the price impact of a swap decreases in the liquidity available L – each ETH unit purchased by **C** has a smaller impact on the price once tick 1509.65 is crossed and the market becomes deeper.

B Notation summary

Variable Subscripts	
Subscript	Definition
T and **N**	Pertaining to the token and numeraire assets, respectively.
L and **H**	Pertaining to the low- and high-fee pool, respectively.
LP	Pertaining to liquidity providers.
LT	Pertaining to liquidity traders.
A	Pertaining to arbitrageurs.
Exogenous Parameters	
Parameters	Definition
v_t	Common value of the token at time t.
η, $1-\eta$	Poisson arrival rate of news and private value shocks, respectively.
δ	Size of common or private value shock.
Δ	Parameter governing the probability distribution of shocks, $\phi(\delta) = \frac{1}{2\Delta\sqrt{1+\delta}}$.
ℓ, h	Liquidity fee on the low- and high-fee pool.
f	Liquidity fee on a non-fragmented pool.
q_i	Token endowment of liquidity provider i, exponentially distributed with scale parameter λ.
λ	Aggregate liquidity supply if all **LP**s join the market.
Γ	Gas price on the blockchain.
r	Width of the price range $\left[\frac{v}{(1+r)^2}, v(1+r)^2\right]$.
Endogenous Quantities	
Variable	Definition
T_k	Equilibrium liquidity supply on exchange $k \in \{\mathbf{L}, \mathbf{H}\}$.
$\tau(\delta)$	Optimal trade size for **LT** or **A** with value shock δ.
$\mathcal{L}(f_k)$	Liquidity yield (fee revenue per unit of liquidity supplied) on pool with fee f_k.
$\mathcal{A}(f_k)$	Adverse selection cost for **LP** per unit of liquidity supplied on pool with fee f_k.
$\mathcal{C}(f_k)$	Liquidity re-balancing cost on pool with fee f_k.
$q_t^\star$	Token endowment of the **LP** who is indifferent between pools.
$\underline{q}_k$	Lowest token endowment deposited on pool k (from break-even condition).
π_k	Expected liquidity provider profit on exchange k.
w_k	Liquidity market share for pool with fee k.

C Proofs

Lemma 1

Proof. We take the expectation of fee revenues over the size of private value shocks δ and obtain:

$$\begin{aligned}\mathbb{E}\text{ProfitLiq}_{i,k} = 2q_i v f_k \times \Big\{ & \\ & \mathbb{P}\left(f_k < \delta \leq (1+f_k)(1+r)^2 - 1\right) \times \frac{1+r}{r} \mathbb{E}\left[\sqrt{\frac{1+\delta}{1+f_k}} - 1 \mid f_k < \delta \leq (1+f_k)(1+r)^2 - 1\right] + \\ & + \mathbb{P}\left(\delta > (1+f_k)(1+r)^2 - 1\right) \times (1+r) \Big\} \\ & = q_i \, v \underbrace{\frac{f_k(r+1)\left(2\Delta - r\sqrt{f_k+1} - 2\sqrt{f_k+1}\right)}{\Delta}}_{\equiv \mathcal{L}(f_k)}, \end{aligned} \tag{C.1}$$

where we define $\mathcal{L}(f_k)$ as the liquidity yield: i.e., the per-unit profit from liquidity provision in pool k.

To explore how the liquidity revenue changes with respect to the pool fee, we differentiate $\mathcal{L}$ with respect to f:

$$\frac{\partial L(f)}{\partial f} = -\frac{(r+1)\left(2\left(-2\Delta\sqrt{f+1} + r + 2\right) + 3f(r+2)\right)}{4\Delta\sqrt{f+1}}. \tag{C.2}$$

Starting with $f = 0$ and given that $\Delta > 1 + r$ (by Assumption 1), the derivative at $f = 0$ is positive:

$$\left.\frac{\partial \mathcal{L}(f)}{\partial f}\right|_{f=0} = \frac{(r+1)(2\Delta - r - 2)}{2\Delta} > 0, \tag{C.3}$$

indicating that liquidity revenue increases with pool fee at this point. The derivative has roots:

$$f_{1,2} = \frac{-6(r+2)^2 + 8\Delta^2 \pm 4\sqrt{4\Delta^4 + 3\Delta^2(r+2)^2}}{9(r+2)^2}, \tag{C.4}$$

where the smallest root f_1 is negative and therefore not relevant. We need to show that the largest root f_2 is always positive, defining the threshold $\overline{f}$.

For this, consider the numerator of f_2, labeled $g(r, \Delta)$:

$$g(r, \Delta) = 8\Delta^2 + 4\sqrt{4\Delta^4 + 3\Delta^2(r+2)^2} - 6(r+2)^2.$$

This function has three roots in r, all of which are negative: $r = -2$, $r = -2(1+\Delta)$, and $r = 2(\Delta - 1)$. Since these roots are negative, for $r \geq 0$, g does not change sign and it is sufficient to examine $g(0, \Delta)$:

$$g(0, \Delta) = 8\Delta^2 + 4\sqrt{\Delta^2(\Delta^2 + 3)} - 6. \tag{C.5}$$

This is positive for any $\Delta \geq 1$, confirming that the largest root f_2 is positive and hence, $\overline{f}$ exists and is positive. This completes the proof that the liquidity revenue increases with the pool fee until $\overline{f}$ and decreases with pool fee for $f > \overline{f}$. □

Lemma 2

Proof. The cost of adverse selection for pool k after evaluating equation (11) is

$$\mathcal{A}(f_k) = v\frac{\left(\Delta - \sqrt{1+f}(1+r)\right)\left(\Delta^2 + \Delta\sqrt{f+1}(1+r) + (f+1)(r-2)(r+1)\right) + (f+1)^{3/2}r^2(r+1)}{3\Delta}. \tag{C.6}$$

We aim to demonstrate that $\mathcal{A}(f)$ decreases as f increases. To do this, we calculate the partial derivative of $\mathcal{A}(f)$ with respect to f:

$$\frac{\partial \mathcal{A}(f)}{\partial f} = \frac{(r+1)\left(-2\Delta\sqrt{f+1} + f(r+2) + r + 2\right)}{2\Delta\sqrt{f+1}} < 0 \tag{C.7}$$

The derivative is negative if $\Delta > \frac{1}{2}\sqrt{1+f}(2+r)$. Given that $\frac{1}{2}(2+r) < 1+r$, it follows that $\mathcal{A}(f)$ decreases for any $\Delta > (1+r)\sqrt{1+f}$, consistent with our assumption on Δ. □

Proposition 1

Proof. First, consider the case in which $\eta > \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$ holds. This implies $q_{\mathrm{t}} < 0$, and consequently, $\pi_L - \pi_H < 0$ for all q. Under this scenario, liquidity providers universally favor pool H over pool L. They supply liquidity on pool H if and only if their participation constraint is satisfied, that is if $q_i > \underline{q}_h$.

Conversely, if $\eta \leq \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$, then $q_{\mathrm{t}} \geq 0$, allowing for a fragmented equilibrium. If $q_{\mathrm{t}} \geq 0$, then π_L has a steeper slope compared to π_H: profit increases more rapidly with liquidity supply in pool L than in pool H. There are two potential outcomes.

1. **Dominance of pool L.** If $0 < q_{\mathrm{t}} < \underline{q}_\ell < \underline{q}_h$, as shown in the left-hand side panel of the diagram below, then the low-fee pool L captures the entire market share for any q_i that yields positive profits. The condition $\underline{q}_\ell < \underline{q}_h$ is equivalent to

$$\underline{q}_\ell < \underline{q}_h \Leftrightarrow \frac{\mathcal{C}(h)}{\mathcal{C}(\ell)} > \frac{(1-\eta)\mathcal{L}(h) - \eta\mathcal{A}(h)}{(1-\eta)\mathcal{L}(\ell) - \eta\mathcal{A}(\ell)}, \tag{C.8}$$

which translates to $\eta > \frac{\mathcal{C}(\ell)\mathcal{L}(h)-\mathcal{C}(h)\mathcal{L}(\ell)}{\mathcal{C}(\ell)[\mathcal{L}(h)+\mathcal{A}(h)]-\mathcal{C}(h)[\mathcal{L}(\ell)+\mathcal{A}(\ell)]}$. Since we require $\eta \leq \frac{\mathcal{L}(\ell)}{\mathcal{L}(\ell)+\mathcal{A}(\ell)}$ by Assumption 2, it must be that $\mathcal{L}(\ell)\mathcal{A}(h) < \mathcal{L}(h)\mathcal{A}(\ell)$.

However, the parameter regions never overlap, ruling out this scenarios: we will show that $\mathcal{L}(\ell)\mathcal{A}(h) - \mathcal{L}(h)\mathcal{A}(\ell) > 0$. To see this, we first note that $\eta \leq \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$ is equivalent to:

$$\eta \leq \frac{3\left(h\left(2\Delta - \sqrt{h+1}r - 2\sqrt{h+1}\right) - \ell\left(2\Delta - \sqrt{\ell+1}r - 2\sqrt{\ell+1}\right)\right)}{(r+2)\left(\sqrt{h+1}(h-2) - \sqrt{\ell+1}(\ell-2)\right)}, \tag{C.9}$$

which implies that $h\left(2\Delta - \sqrt{h+1}r - 2\sqrt{h+1}\right) - \ell\left(2\Delta - \sqrt{\ell+1}r - 2\sqrt{\ell+1}\right) > 0$ since the denominator is always positive. We use the inequality in (C.9) and obtain that

$$\mathcal{L}(\ell)\mathcal{A}(h) - \mathcal{L}(h)\mathcal{A}(\ell) > \frac{1+r}{6\Delta^2}g, \tag{C.10}$$

where

$$g \equiv \underbrace{(1+r)l\left(2\Delta - \sqrt{l+1}\,(r+2)\right)}_{>0} \times \tag{C.11}$$

$$\times \left((h-\ell)\,\Delta - (r+2)\left(\sqrt{h+1} - \sqrt{l+1}\right) + \underbrace{h\left(2\Delta - \sqrt{h+1}\,(r+2)\right) - l\left(2\Delta - \sqrt{l+1}\,(r+2)\right)}_{>0} \right) > 0, \tag{C.12}$$

given Assumption 1 and equation (C.9). To see that $(h-\ell)\,\Delta - (r+2)\left(\sqrt{h+1} - \sqrt{l+1}\right) > 0$, we first note the expression increases in Δ and is therefore larger than $\sqrt{h+1}(r+1)(h-l) + (r+2)\left(\sqrt{l+1} - \sqrt{h+1}\right)$ for $\Delta > \sqrt{1+h}\,(1+r)$. The latter expression increases in h and equals zero for $h = \ell$.The latter expression increases in h and equals zero for $h = \ell$.

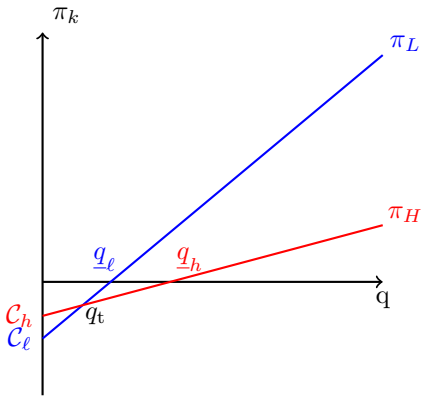
Therefore, there are no parameter values for which $0 < q_{\mathrm{t}} < \underline{q}_\ell < \underline{q}_h$ and $\eta \leq \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$, which rules out the case of pool L attracting full market share.

2. **Fragmented Market Equilibrium.** The right-hand side panel depicts the scenario $\underline{q}_h < \underline{q}_\ell < q_{\mathrm{t}}$. Here, liquidity providers with q_i in the range $(\underline{q}_h, q_{\mathrm{t}}]$ achieve higher positive profits in pool H, while those with $q_i > q_{\mathrm{t}}$ obtain larger profits in pool L. The condition $\underline{q}_h < \underline{q}_\ell$ is equivalent to

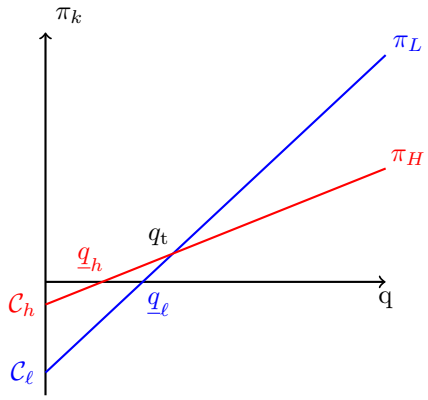
$$\underline{q}_h \leq \underline{q}_\ell \Leftrightarrow \frac{\mathcal{C}(h)}{\mathcal{C}(\ell)} \leq \frac{(1-\eta)\,\mathcal{L}(h) - \eta\mathcal{A}(h)}{(1-\eta)\,\mathcal{L}(\ell) - \eta\mathcal{A}(\ell)}, \tag{C.13}$$

which is always true if $\eta \leq \frac{\mathcal{L}(l)-\mathcal{L}(h)}{\mathcal{L}(l)-\mathcal{L}(h)+\mathcal{A}(l)-\mathcal{A}(h)}$ as seen above.

(a) Single pool market (with fee ℓ)



(b) Fragmented market



It is crucial to note that configurations where $\underline{q}_\ell < q_{\mathrm{t}} < \underline{q}_h$ or $\underline{q}_h < q_{\mathrm{t}} < \underline{q}_\ell$ are not feasible, as they would lead to a contradiction where profits are simultaneously positive in one pool and negative in the other at the indifference point q_{t}.

□

Proposition 2

Proof. We first note that both q_t and $\underline{q}_h$ scale linearly with Γ: that is, there exists $Q_t > Q_h > 0$ such that $q_t = \Gamma Q_t$ and $\underline{q}_h = \Gamma Q_h$ where Q_t and Q_h are not functions of Γ. Next, we compute the partial derivative of w_ℓ with respect to Γ

$$\frac{\partial w_\ell}{\partial \Gamma} = \frac{\Gamma(Q_h - Q_t)e^{\frac{\Gamma(Q_h-Q_t)}{\lambda}}(\Gamma Q_h Q_t + \lambda(Q_h + Q_t))}{\lambda(\lambda + \Gamma Q_h)^2} < 0, \tag{C.14}$$

since $Q_h < Q_t$ and all other terms are positive. □

Proposition 3

Proof. The two-pool gains from trade for an **LT** with private value $v(1+\delta)$ are

$$\begin{aligned}\text{GainsFromTrade}(\{h, \ell \mid \delta\}) = v\delta T_H \min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+h}{1+\delta}}\right\}\right\} + \\ + v\delta T_L \min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+\ell}{1+\delta}}\right\}\right\}.\end{aligned} \tag{C.15}$$

We set $h = f$ such that the marginal **LP** entering the market is the same as in the single-fee pool; that is, $T_H + T_L = e^{-q_h\lambda}(q_h + \lambda)$. Since $\min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+f}{1+\delta}}\right\}\right\}$ decreases in f, it follows that:

$$\begin{aligned}\text{GainsFromTrade}(\{h, \ell \mid \delta\}) &\geq v\delta \underbrace{(T_H + T_L)}_{=e^{-q_h\lambda}(q_h+\lambda)} \min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+h}{1+\delta}}\right\}\right\} &\text{(C.16)}\\ &= \text{GainsFromTrade}(\{h\} \mid \delta\}, &\text{(C.17)}\end{aligned}$$

with strict inequality if $q_t < Q$ such that the low-fee pool attracts a positive mass of **LP**s. The inequality holds for any δ, and therefore it remains true if we aggregate the gains from trade over the distribution **LT** private values. We note that the expected gains from trade per unit of liquidity is

$$\begin{aligned}&\mathbb{E}\min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+h}{1+\delta}}\right\}\right\} - \\ &= \frac{6\sqrt{f+1}(r+1)\log(r+1) + r\left(2\left(\Delta^3 - 3\Delta - \sqrt{f+1}\right) + \sqrt{f+1}(f(r+1)(r+2) + r(r+3))\right)}{6\Delta r} > 0,\end{aligned} \tag{C.18}$$

which decreases in f since

$$\frac{\partial \mathbb{E}\min\left\{1, \frac{1+r}{r}\max\left\{0, 1-\sqrt{\frac{1+h}{1+\delta}}\right\}\right\}}{\partial f} = -\frac{(r+1)((f+1)r(r+2) - 2\log(r+1))}{4\Delta\sqrt{f+1}r} < 0. \tag{C.19}$$

□

D Just-in-time liquidity

Just-in-time (JIT) liquidity is a strategy that leverages the transparency of orders on the public blockchains. If a liquidity provider observes an incoming large order that has not been processed by miners and it deems uninformed in the public mempool, it can conveniently re-arrange transactions and propose a sequence of actions to sandwich this trade as follows:

1. Add a large liquidity deposit at block position k, at the smallest tick around the current pool price.
2. Let the trade at block position $k+1$ execute and receive liquidity fees.
3. Remove or burn any residual un-executed liquidity at block position $k+2$.

The mint size is optimally very large (i.e., of the order of hundred of millions USD for liquid pairs), such that the JIT liquidity provider effectively crowds out the existing liquidity supply and collects most fees for the trade. That is, the strategy is made possible by pro-rata matching on decentralized exchanges because with time priority, the JIT provider cannot queue-jump existing liquidity providers. Since the JIT liquidity provider does not want to passively provide capital, it removes any residual deposit immediately after the trade.

We identify JIT liquidity events by the following algorithm as in Wan and Adams (2022):

1. Search for mints and burns in the same block, liquidity pool, and initiated by the same wallet address. The mint needs to occur exactly two block positions ahead of the burn (at positions k and $k+2$).
2. Classify the mint and the burn as a JIT event if the transaction in between (at position $k+1$) is a trade in the same liquidity pool.

JIT events are rare in our sample, and account for less than 1% of the traded volume on Uniswap v3. Further, more than half of them occur in a single pair - USDC-WETH, and in low-fee pools. The Uniswap Labs provides further discussions on the aggregate impact of JIT liquidity provision here. Regarding the economic effects, JIT liquidity reduces price impact for incoming trades, but dilutes existing liquidity providers in the pro-rata markets, and can discourage liquidity supply in the long run.

E Impermanent loss measure

We build our measure of impermanent loss in line with the definition of token reserves within a price range in the Uniswap V3 white paper (Adams, Zinsmeister, Salem, Keefer, and Robinson, 2021) and Section 4.1 in Heimbach, Schertenleib, and Wattenhofer (2022).

Consider a liquidity provider who supplies L units of liquidity into a pool trading a token x for a token y. The chosen price range is $[p_\ell, p_u]$ with $p_\ell < p_u$. Further, the current price of the pool is p_0. We are interested in computing the impermanent loss at a future point in time, when the price updates to p_1.

From Adams, Zinsmeister, Salem, Keefer, and Robinson (2021), the actual amount of tokens x and y ("real reserves") deposited on a Uniswap v3 liquidity pool with a price range $[p_\ell, p_u]$ to yield liquidity L are functions of the current pool price p:

$$x(p) = \begin{cases} L \times \left(\frac{1}{\sqrt{p_\ell}} - \frac{1}{\sqrt{p_u}}\right) & \text{if } p \leq p_\ell \\ L \times \left(\frac{1}{\sqrt{p}} - \frac{1}{\sqrt{p_u}}\right) & \text{if } p_\ell < p \leq p_u \\ 0 & \text{if } p > p_u \end{cases} \quad \text{and } y(p) = \begin{cases} 0 & \text{if } p \leq p_\ell \\ L \times \left(\sqrt{p} - \sqrt{p_\ell}\right) & \text{if } p_\ell < p \leq p_u \\ L \times \left(\sqrt{p_u} - \sqrt{p_\ell}\right) & \text{if } p > p_u. \end{cases} \tag{E.1}$$

From equation (E.1), the value of the liquidity position at $t = 1$ is therefore

$$V_{\text{position}} = p_1 x(p_1) + y(p_1) = \begin{cases} L p_1 \times \left(\frac{1}{\sqrt{p_\ell}} - \frac{1}{\sqrt{p_u}}\right) & \text{if } p_1 \leq p_\ell \\ L \times \left(2\sqrt{p_1} - \frac{p_1}{\sqrt{p_u}} - \sqrt{p_\ell}\right) & \text{if } p_\ell < p_1 \leq p_u \\ L \times \left(\sqrt{p_u} - \sqrt{p_\ell}\right) & \text{if } p_1 > p_u. \end{cases} \tag{E.2}$$

Conversely, the value of a strategy where the liquidity provider holds the original token quantities and marks them to market at the updated price is

$$V_{\text{hold}} = p_1 x(p_0) + y(p_0) = \begin{cases} L p_1 \times \left(\frac{1}{\sqrt{p_\ell}} - \frac{1}{\sqrt{p_u}}\right) & \text{if } p_0 \leq p_\ell \\ L \times \left(\frac{p_1 + p_0}{\sqrt{p_0}} - \frac{p_1}{\sqrt{p_u}} - \sqrt{p_\ell}\right) & \text{if } p_\ell < p_0 \leq p_u \\ L \times \left(\sqrt{p_u} - \sqrt{p_\ell}\right) & \text{if } p_0 > p_u. \end{cases} \tag{E.3}$$

The impermanent loss is then defined as the excess return from holding the assets versus providing liquidity on the decentralized exchange:

$$\text{ImpermanentLoss} = \frac{V_{\text{hold}} - V_{\text{position}}}{V_{\text{hold}}}. \tag{E.4}$$

Empirically, we follow Heimbach, Schertenleib, and Wattenhofer (2022) and compute impermanent loss for "symmetric" positions around the current pool price, that is $p_\ell = p_0 \alpha^{-1}$ and $p_u = p_0 \alpha$, with $\alpha > 1$. We allow for a time lag of one hour between p_0 and p_1.

Liquidity Provision in a One-Sided Market: The Role of Dealer-Hedge Fund Relations *

Mathias Kruttli† Marco Macchiavelli‡ Phillip Monin§ Xing (Alex) Zhou¶

February 2024

Abstract

We show that dealers' prime brokerage relations with certain hedge funds help improve their liquidity provision in a one-sided market. During the March 2020 liquidity crisis, hedge funds increased their corporate bond positions when the bond market faced excessive selling pressures. Dealers connected with hedge funds that are natural buyers of corporate bonds charged lower transaction costs on heavily sold bonds. Dealers' leverage and funding constraints do not explain our results, nor do connections with hedge funds that are natural buyers of other asset classes. Our findings reveal that dealers' willingness to provide liquidity in a one-sided market depends on their connections with natural buyers of corporate bonds.

JEL classification: G12, G23, G24.

Keywords: broker-dealers, hedge funds, corporate bonds, market liquidity.

*For their comments, we thank Maureen O'Hara (discussant) and participants at the NBER Financial Market Frictions and Systemic Risks Spring 2024 Meeting. We also thank the Financial Industry Regulatory Authority (FINRA) for providing the regulatory version of the TRACE data used in this study. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Board of Governors, the Federal Reserve System, the Office of Financial Research, or the Department of the Treasury.

†Kelley School of Business, Indiana University. Email: mkruttli@iu.edu

‡Isenberg School of Management, UMass Amherst. Email: mmacchiavelli@isenberg.umass.edu.

§Federal Reserve Board. Email: phillip.monin@frb.gov

¶Cox School of Business, Southern Methodist University, and the Office of Financial Research at the Department of the Treasury. Email: axzhou@smu.edu.

1 Introduction

Dealers play a critical role in the U.S. corporate bond market. During times of stress when demand for liquidity surges, dealers' willingness to provide liquidity is essential to the proper functioning of the market. The existing literature has studied bond market liquidity through the lens of dealers' ability to warehouse investor flows on their balance sheets, emphasizing the roles played by funding costs and the regulatory reforms passed in the aftermath of the 2008 financial crisis.[1] However, what determines dealers' willingness to intermediate trades when the market becomes one-sided is still to be fully understood.

In this paper, we analyze dealers' relationships with hedge funds that are natural buyers of corporate bonds and study the effect that such relationships have on dealers' willingness to provide liquidity in the corporate bond market during times of stress. Dealers facilitate bond trading by serving either as market makers, temporarily absorbing order imbalances and subsequently turning around their inventories, or as matchmakers, searching for counterparties for their customers without committing their own capital for intermediation. In both functions, whether dealers are willing to intermediate depends on their capability to locate entities to take the other side of a trade. Such capability is particularly valuable in a one-sided market when searching for counterparties becomes challenging.

During the COVID liquidity crisis, the corporate bond market faced severe selling pressure amid a dash for liquidity. For example, mutual funds experienced heavy outflows (Falato, Goldstein and Hortaçsu, 2021) and were forced to sell corporate bonds to meet redemptions (Ma, Xiao and Zeng, 2022). With many bond mutual funds and other investors selling bonds, the corporate bond market became one-sided and liquidity deteriorated, as shown in Figure 1. Using redemption-driven sales of corporate bonds by mutual funds at the security level, we show that dealers intermediating these bonds during the crisis charged higher transaction costs.

[1] See for example, Adrian et al. (2017); Anderson and Stulz (2017); Bao, O'Hara and Zhou (2018); Bessembinder et al. (2018); Choi, Huh and Seunghun Shin (2023); Dick-Nielsen and Rossi (2019); Duffie (2022); Macchiavelli and Zhou (2022); Saar et al. (2023); Schultz (2017); Trebbi and Xiao (2019).

Meanwhile, market dislocations caused by the COVID liquidity crisis provide profitable trading opportunities for hedge funds that tend to maintain large long positions in corporate bonds. Similar to (Shleifer and Vishny, 2011), we refer to such hedge funds as natural buyers. As hedge funds are less regulated than dealer banks and generally more willing to absorb risk, they are likely to step in and buy securities while many other investors liquidate bonds to de-risk their portfolios. Indeed, we find that for the investment-grade sector where market dislocations were more severe (Haddad, Moreira and Muir, 2021), overall hedge funds positions in corporate bonds increased significantly in March 2020.

Taking advantage of confidential SEC data on hedge funds' corporate bond positions and the identities of their prime brokers, we construct a measure of a dealer's exposure to the corporate bond holdings of their connected hedge funds. We then link it to dealer's corporate bond trading activities using dealer identities provided by a regulatory version of FINRA's TRACE data. While the liquidity deterioration is more pronounced in bonds more exposed to mutual fund fire sales, such deterioration is attenuated for the corporate bonds traded by dealers that have prime brokerage relations with natural buyers of corporate bonds.

Our results are robust to controlling for alternative channels, including repo funding conditions, the tightness of leverage constraints, or any other time-varying dealer or bond characteristic. These findings suggest that even without regulatory constraints and prohibitive funding costs, dealers might still be reluctant to intermediate trades as they perceive future challenges in locating buyers to turn around their inventories in a one-sided market. We also address concerns that our results could be driven by dealer sophistication or overall dealer-hedge fund connections — instead of relations with natural corporate bond buyers specifically — by performing a set of placebo tests in addition to controlling for dealer-day fixed effects. When we consider dealer relations with hedge funds that are buyers of U.S. Treasuries or equities instead of corporate bonds, we do not find effects on corporate bond liquidity provision.

Our results highlight the importance for dealers to have relations with hedge fund buyers during periods of stress. Dealers act as intermediaries, not ultimate investors. As such, they are not keen to take significant directional bets on the market (Treynor, 1987; Levine, 2015). Indeed, as shown in several recent studies (Boyarchenko, Kovner and Shachar, 2022; Kargar et al., 2021; O'Hara and Zhou, 2021), the corporate bond liquidity crisis of March 2020 ended once a large enough buyer stepped in to absorb a sizable amount of corporate bonds. Specifically, corporate bond liquidity broadly improved after the Federal Reserve acted as the buyer of last resort by establishing the Secondary Market Corporate Credit Facility, which had the goal of supporting liquidity via purchases of corporate bonds in secondary markets.

After documenting that dealer relationships with hedge funds have an effect on corporate bond liquidity provision, we explore factors that influence hedge funds' propensity to be the natural buyers of corporate bonds during times of stress. We consider several hedge fund characteristics, including size, net flows, liquidity transformation, share restrictions, profitability, and risk tolerance. We show that hedge funds' size and risk tolerance are the main factors that predict hedge funds' ability to absorb more corporate bonds during the COVID liquidity crisis, following a heavy sell-off of corporate bonds by mutual funds and other investors. This seems to be an efficient outcome as risky assets move from less to more risk tolerant investors in a downturn, and is in line with the findings of Kruttli et al. (2021) on hedge fund liquidity provision in the U.S. Treasury market. The fact that hedge funds' size and risk tolerance are the primary drivers of their higher corporate bond exposures is in line with our previous findings on dealers. Namely, that dealer leverage and funding conditions did not significantly affect their liquidity provision during the COVID liquidity crisis. In such a one-sided market, dealers refrain from taking inventory risk. As such, internal risk controls may have been more binding than the leverage ratio. Moreover, dealers' funding conditions were generally stable during the COVID liquidity crisis. Ultimately, what seemed to matter the most is the risk-absorbing capacity of different players. While dealers intermediate between

buyers and sellers, some hedge funds took directional bets on the market and increased their exposure to corporate bonds in March 2020.

We contribute to the literature on market making in corporate bonds. The literature mainly focuses on the role of regulatory constraints (Bao, O'Hara and Zhou, 2018; Bessembinder et al., 2018; Breckenfelder and Ivashina, 2021). Recent papers examine dealers' internal risk limits (Anderson, McArthur and Wang, 2023) and how insurers' stable funding structures allowed them to provide liquidity during the COVID crisis (O'Hara, Rapp and Zhou, 2021). We show that liquidity provision is facilitated by having relations with hedge funds, especially during times of stress when most investors are de-risking by selling securities.

We also contribute to the literature on hedge funds as liquidity providers. The liquidity provision of hedge funds during crisis periods has mainly been studied in the context of equity markets (Ben-David, Franzoni and Moussawi, 2012; Jylhä, Rinne and Suominen, 2014; Aragon, Martin and Shi, 2019; Çötelioğlu, Franzoni and Plazzi, 2021; Glossner et al., 2020) due to the available long-equity holdings from the SEC Form 13F filings of investment advisers. An exception is Kruttli et al. (2021), who study liquidity provision of hedge funds in U.S. Treasury markets. In contrast, our paper focuses on how relationships between dealers and hedge funds affect the liquidity provision of dealers, and consequently transaction costs, in corporate bond markets.

Aragon and Strahan (2012) show that stocks traded by hedge funds with Lehman Brothers as their prime broker became more illiquid in the aftermath of the Lehman Brothers' collapse. This finding is attributed to hedge funds being unable to access collateral posted with Lehman Brothers during the bankruptcy proceedings. Our analysis is testing a notably different mechanism. We measure transaction costs at the dealer-security level and test if relationships with hedge funds allow dealers to provide more liquidity and reduce transaction costs. Han, Kim and Nanda (2020) and Di Maggio, Egan and Franzoni (2022) show that hedge funds and mutual funds can profit from connections to central dealers because the latter provide lower trading costs. Our paper

differs from Han, Kim and Nanda (2020); Di Maggio, Egan and Franzoni (2022), as we find that dealers need hedge funds that are natural buyers of bonds to enable them to provide better liquidity in a one-sided market.

The remainder of the paper is organized as follows. Section 2 sets out the data and discusses the sample construction. In Section 3, we develop a measure to capture each dealer's exposure to hedge funds and test how it affects the transaction cost that dealers charge to their customers. In Section 4, we study a number of hedge fund characteristics and link them to their behavior around the crisis. Section 5 concludes the paper.

2 Data

Our study relies on data from several sources. To capture dealer liquidity provision, we obtain from the Financial Industry Regulatory Authority (FINRA) the TRACE corporate bond transaction data for the first quarter of 2020. TRACE data provide detailed information on secondary market transactions in corporate bonds, including bond CUSIP, trade execution date and time, trade price and quantity, an indicator for inter-dealer trades, an indicator for agency or principal trades, and an indicator for whether the reporting dealer buys or sells the bond. Unlike the publicly disseminated TRACE data, our data include the identity of the dealer involved in each trade. Such information is essential to our analysis as it allows us to estimate bond liquidity at the dealer-bond level, and link it to each dealer's relationships with hedge funds. We exclude from our sample the following transactions: when issued, canceled, subsequently corrected, and reversed trades.

We supplement TRACE bond transaction data with bond characteristic data, including total amount outstanding, issuance and maturity dates, and credit rating, from the Mergent Fixed Income Securities Database (FISD). We focus on bonds issued in U.S. dollars by U.S. firms in the following three broad FISD industry groups: industrial, financial and utility. Each bond has to be rated by Moody's or S&P. If a bond is rated differently

by the two rating agencies, we use the lower of the two. To avoid the potential impact of special bond features on the liquidity estimation, we focus on fixed-coupon corporate bonds with semiannual coupon payments, $1,000 par amount, and fixed maturity. We also exclude from our sample the following bonds: convertible or putable bonds, private placements, asset-backed issues, and issues that are part of a unit deal. After applying these filters, we end up with a sample of 8,716 corporate bonds.

For each bond in our sample, we obtain data on its par amount held by each mutual fund at the most recent quarter-end from Thomson Reuters' eMAXX database, which provides security-level holdings information of fixed-income mutual funds at a quarterly frequency. We obtain daily data on mutual fund flows from Morningstar.

Our data on hedge funds are from the SEC Form PF and SEC Form ADV filings of hedge fund advisers that file Form PF on a quarterly basis and report granular information about their large hedge funds, called qualifying hedge funds in the form. We focus on qualifying hedge funds because these funds report items that we require in our analysis, such as hedge fund-dealer borrowing amounts and other variables such as corporate bond exposures and returns at a monthly frequency.[2] Our sample construction follows the methodology described in Kruttli, Monin and Watugala (2022) and our data include filings for Q4 2019 and Q1 2020. Hedge fund borrowing is reported in response to Question 47, which requires the fund to list all its "major" creditors, defined as creditors to whom the hedge fund owes 5% or more of its NAV in a given quarter.[3] We manually inspect the "name" entries for Question 47 in the Form PF filings and match these to parent institutions.

To control for funding liquidity at the dealer level, we use triparty repurchase agreements (repos) transaction level data. Triparty repos are collateralized loans used to raise cash against the pledge of collateral, which is held in custody at Bank of New York

[2]Advisers with at least $1.5 billion in regulatory assets under management across all of their hedge funds file Form PF on a quarterly basis. Qualifying hedge funds are hedge funds with at least $500 million in net asset value.

[3]On average, 87.3% of all the borrowing by a hedge fund is from its major creditors, as shown in the Online Appendix of Kruttli, Monin and Watugala (2022).

Mellon. Information on triparty repos is provided daily to the Federal Reserve Bank of New York (FRBNY). The triparty repo data include a transaction-level trade file, which provides the loan amount and the interest rate for each repo transaction, as well as the identity of both the borrower and the lender. In this study, we focus on repos backed by corporate debt securities as collateral. For each dealer we compute the average pre-crisis repo rate weighted by the loan amount, called Repo Rate$_{pre}$, which controls for the differential pre-crisis funding costs of each dealer. We also compute an exogenous funding shock to the dealer, called Repo Shock, which is the monthly change in corporate bond repos outstanding between the dealer and prime money market funds (MMFs). We first identify each prime MMF among the various repo lenders, and then compute their level of outstanding corporate bond repo lending to each dealer. Finally, we compute its monthly change at the dealer level. Since prime MMFs saw heavy redemptions for reasons unrelated to their exposures to the repo market (Li et al., 2021), they had to cut back on their repo lending. As a result, Repo Shock serves as an exogenous funding shock to dealers.

Data on leverage is obtained in two ways. For dealers affiliated with bank holding companies subject to the Supplementary Leverage Ratio (SLR), we obtain SLR data from the annual Federal Reserve Stress Test results as of 2019Q4.[4] These bank holding companies are subject to a minimum SLR of 5%. For dealers that are not subject to the SLR, we obtain their leverage ratio from the SEC FOCUS reports as of 2019Q4. These dealers are subject to a minimum leverage ratio (net capital over debits) of 2%.

2.1 Summary Statistics

Summary statistics are displayed in Table 1. In the final sample, which corresponds to the intersection of TRACE, Form PF, and eMAXX, there are 19 dealers and 8,716 CUSIPs. The average transaction cost is 41.39 basis points including riskless principal trades (RPTs) and 39.96 excluding them. Following Harris (2015), we denote a trade

[4]See https://www.federalreserve.gov/publications/files/2020-dfast-results-20200625.pdf.

as RPT if a dealer offsets it within one minute by another trade in the same bond and with the same size but opposite trade direction. As such, RPTs do not require capital commitments and, as a result, transaction costs on RPTs tend to be lower than those on trades that require capital commitments. While at first this may seem not to apply to our case, because the average cost including RPT is higher than the cost excluding them, this is actually due to selection. Indeed, once we restrict the sample to observations with non-missing Cost (No RPT), the average cost excluding RPTs, 39.96, is higher than that including RPTs, 38.84. Figure 1 displays the dynamics of transaction costs including RPTs in panel (a) and excluding RPTs in panel (b), separately for investment grade (IG) and high yield (HY) bonds. While transaction costs are usually higher for HY than IG bonds in normal times, they both spike up and reach similar heights during the COVID liquidity crisis of March 2020. The spike is even larger when we exclude riskless principal trades, since these trades do not increase dealers' inventories, thus exposing them to less risk.

The average outflow-weighted bond holdings by mutual funds during the crisis, MF Shock, is \$38.88 million, which points to significant selling pressure by corporate bond mutual funds during the COVID liquidity crisis. This is in contrast to the average outflow-weighted holdings in the pre-crisis period of -\$5.55 million (not reported), which suggests that prior to the crisis mutual funds were on average receiving net inflows and, as a result, buying corporate bonds. Figure 2 shows the aggregate dynamics of mutual fund selling shocks for IG and HY bonds, separately. While almost flat before the crisis, selling pressure by mutual funds mounts during the COVID liquidity crisis, slightly more so for HY than IG bonds, as one would expect. Since IG and HY dynamics for transaction costs and selling shocks tend to be comparable, in our main analysis we estimate the average effect across ratings, while controlling for time-varying bond characteristics. Nevertheless, our results are unchanged if we separately estimate the effects for IG and HY bonds, as shown in Appendix Table B.1.

HF Expo is the logarithm of the long exposures in corporate bonds of affiliated hedge

funds. For each dealer and each month, we sum the long exposures in IG and HY corporate bonds (separately) of hedge funds with which the dealer has a prime brokerage relation. Such relation is obtained from Form PF, Question 47, where each hedge fund lists its major prime broker lenders. HF Expo thus varies at the dealer-month-rating class (IG vs HY) level. Repo Shock, the monthly change in corporate bond repos outstanding coming from prime money funds, is 2.29 percent on average over the sample. However, it varies from an average of 3 percent pre-crisis to -1 percent during the crisis. Next, the average corporate repo rate pre-crisis is 1.75 percent, which is consistent with the federal funds rate of about 1.6 percent during that period. Finally, Leverage Intensity, defined as the minimum leverage ratio minus the actual one as of 2019Q4 divided by the minimum, is on average -26.55, which suggests that the average trade in our sample is carried out by a dealer with a 27 percent buffer above the minimum requirement, which is 5% for SLR-constrained dealers and 2% for other dealers.

Table 1 also reports summary statistics for the hedge fund variables used in our analysis. The average fund manages $4.36 billion and $1.86 billion in gross and net assets, respectively, for a leverage ratio of 2.30. We use data on hedge funds' corporate bond exposures from Form PF Question 30, which requires hedge funds to report long and short portfolio exposures at a monthly frequency for a range of asset classes. Hedge funds' corporate bond exposures include cash bond holdings at fair value and related bond derivatives at notional value, but do not include exposures from credit default swaps. The average gross notional exposure to corporate bonds is $383.8 million, of which $330.5 million is long exposure and $53.3 million is short exposure. Monthly aggregates for long and short corporate bond exposures of hedge funds, broken out further by investment grade and high yield classifications, are provided in Table 2.

The variable $RiskLimit_{h,t}$ is based on the hedge fund's value at risk (VaR). The VaR shows for each fund and month the potential loss (as a percent of NAV) over a one-month horizon with a probability of 5%. Like Kruttli et al. (2021), we construct the measure

based on the monthly VaR observations.[5] The measure proxies for a fund's historical risk limit and is the VaR averaged over a rolling 12-month window. The average $RiskLimit_{h,t}$ is 3.84%, which implies that the average fund expects to lose 3.84% of its NAV in a month 5% of the time.

The next three variables measure different dimensions of fund liquidity, including portfolio liquidity ($PortIlliq_{h,t}$), investor liquidity as measured by share restrictions ($ShareRes_{h,t}$), and the funding liquidity measured as the weighted average maturity of a fund's borrowing ($FinDur_{h,t}$). Form PF asks for the percentage of a hedge fund's assets, excluding cash, that can be liquidated within particular time horizons (within $\leq$1, 2-7, 8-30, 31-90, 91-180, 181-365, and >365 days) using a given periods' market conditions. We compute the weighted average liquidation time to obtain the measure $PortIlliq_{h,t}$. The average $PortIlliq_{h,t}$ is 68.8 days in our sample and the median is 23.4 days. $ShareRes_{h,t}$ is a measure of the expected weighted average time it would take for a hedge fund's investors to withdraw the fund's equity. This variable quantifies the restrictions faced by a fund's investors, such as lock-up, redemption frequencies, and redemption notice periods. The average $ShareRes_{h,t}$ is 186.3 days. The weighted average time to maturity of a fund's borrowing is denoted $FinDur_{h,t}$. On average, the financing duration is 67.0 days for our sample of hedge funds with a median of 19 days. The variable $LiqMismatch_{h,t}$, constructed as in Aragon et al. (2021), summarizes these liquidity metrics and measures the average liquidity of the hedge fund's assets relative to its liabilities. Like Aragon et al. (2021), the average fund in our sample has a negative liquidity mismatch. The table further provides summary statistics for monthly returns, quarterly flows, and the manager's stake in the fund.

[5]Qualifying hedge funds are required to report the VaR of the fund at a monthly frequency if they regularly calculate it. Kruttli et al. (2021) show that most funds report their VaR and provide a method, which we adopt, to convert reported VaRs to a 5% significance level and monthly horizon.

3 Dealer-Hedge Fund Relationships and Corporate Bond Liquidity

Most corporate bonds trade in over-the-counter (OTC) markets and heavily rely on dealers for intermediation. Dealers use their networks with customers and other dealers to facilitate the matching between buyers and sellers, and the remaining order imbalances remain on their balance sheets as inventories. Following the 2008 financial crisis, various banking regulations increased dealer balance sheet costs and discouraged them from committing their own capital to market making. Under such regulation-induced constraints, dealers' ability to locate counterparties for their customers became particularly valuable. In addition to balance sheet constraints, dealers do not want to take significant inventory risk in turbulent times and especially in a one-sided market. Indeed, committing capital to purchase bonds may lead to large losses if market prices move against the dealer before it can offload these bonds to other customers. Indeed, facing heavy selling pressure, dealers usually charge higher transaction costs, as the dynamics of transaction costs and mutual fund selling shocks displayed in Figures 1 and 2 suggest. For all these reasons, dealers' connections with natural buyers are especially important in a one-sided market.

We hypothesize that dealers' relations with hedge funds play an important role in their liquidity provision during the COVID liquidity crisis. First, the crisis introduced opportunities for hedge funds to step in to profit from market dislocations. Second, hedge funds tend to trade with their prime brokers that also provide them with financing. Such relationship-based trading can help dealers address the unusually high selling pressures in the bond markets. To test this hypothesis, we analyze how a dealer's relation with hedge funds affects its bond trading costs when facing large sell-offs.

We start by measuring liquidity provision at the dealer-bond level. For that purpose, we first estimate the transaction cost for each bond trade as in Hendershott and Madhavan

(2015):

$$Cost_{i,j,r} = ln\left(\frac{P_{i,j,r}}{P^{B}_{i,j,r}}\right) \cdot Sign_{i,j,r}. \tag{1}$$

$P_{i,j,r}$ refers to the price for trade r by dealer i in bond j. $P^{B}_{i,j,r}$ is the benchmark price for trade r, which refers to the last trade price in the inter-dealer markets.[6] $Sign_{i,j,r}$ represents the sign of the trade r, which takes the value of +1 for customer buy and -1 for customer sell. We then calculate a daily average transaction cost for dealer i in bond j during day t ($Cost_{i,j,t}$). Finally, we divide the cost measure by 100 to facilitate our interpretation of the magnitude.

To test whether dealers relations with natural bond buyers can attenuate the effect of mutual funds selling pressures on the liquidity of a bond, we estimate the following empirical model:

$$\begin{aligned} Cost_{i,j,t} = {} & \beta_1 HFExpo_{i,j,t-1} + \beta_2 MFShock_j + \beta_3 HFExpo_{i,j,t-1} \times MFShock_j \\ & + \beta_4 HFExpo_{i,j,t-1} \times Crisis + \beta_5 MFShock_j \times Crisis \\ & + \beta_6 HFExpo_{i,j,t-1} \times MFShock_j \times Crisis + \gamma Controls + \mu_{i,j,t} + \varepsilon_{i,j,t}, \end{aligned} \tag{2}$$

where $Cost_{i,j,t}$ is the transaction cost charged by dealer i on bond j at day t, as defined in Equation (1). $HFExpo_{i,j,t-1}$ is the logarithm of the long exposures to IG and HY corporate bonds of hedge funds affiliated to dealer i as of the previous month. When dealer i trades an IG (HY) bond j, $HFExpo_{i,j,t-1}$ captures the exposure of dealer i to affiliated hedge funds' long positions in IG (HY) bonds. $MFShock_j$ is a proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of bond j by corporate bond mutual funds during the crisis period (March 5 to March 20). $Crisis$ is an indicator variable equal to one between March 5 and March 20, 2020, at the height of the dislocation in capital markets. Control variables include Log(TTM), Repo Shock,

[6] Schultz (2001), Bessembinder, Maxwell and Venkataraman (2006), Goldstein, Hotchkiss and Sirri (2006), and Edwards, Harris and Piwowar (2007) use alternative benchmark prices but broadly similar approaches to estimate transaction costs.

and Repo Rate$_{pre}$. Log(TTM) is the logarithm of the time to maturity of each bond. Repo Shock is the monthly change in dealer-level corporate bond repo outstanding with prime money market funds. It controls for exogenous variation in access to repo markets. Repo Rate$_{pre}$ is the average pre-crisis dealer-level corporate bond repo rate, which is zero pre-crisis and is switched on during the crisis. It controls for predetermined differences in repo funding costs. Finally, we include a set of fixed effects. Standard errors are two-way clustered at the bond and dealer level.

The results in Table 3 support our hypothesis. Consistent with the literature on fire-sales and liquidity (Ambrose, Cai and Helwege, 2008; Ellul, Jotikasthira and Lundblad, 2011; Bao, O'Hara and Zhou, 2018), the coefficient of $Crisis \times MFShock$ is positive and highly significant, suggesting that dealer liquidity provision deteriorates significantly for the bonds heavily sold during the crisis (column 1). More importantly, the coefficient of the triple interaction term, $Crisis \times HFExpo \times MFShock$, is negative and highly significant. This finding suggests that a dealer with more hedge fund connections is able to charge a relatively lower transaction cost in bonds facing heavy mutual fund sell-offs. To control for the potential time-varying impact of bond characteristics, we replace bond fixed effects and day fixed effects with bond-day fixed effects. Column (2) shows that, even after controlling for bond-day fixed effects, the coefficient of $Crisis \times HFExpo \times MFShock$ remains negative and highly significant. This finding suggests that among dealers that trade the same bond on the same day, those with stronger relations with hedge funds charge lower transaction costs for bonds facing higher selling pressure during the crisis.

A key assumption underlying our triple-difference empirical design is the parallel trend assumption, which requires that the difference in liquidity costs charged by dealers with different relations with hedge funds (first difference) across bonds with different exposure to mutual fund sell-offs (second difference) do not already exhibit different patterns prior to the crisis period. To validate the parallel trend assumption, we construct three indicator variables for three pre-crisis sub-periods, each 2 weeks long. Specifically,

Crisis_{-2} equals one between February 6 and 19; Crisis_{-3} equals one between January 23 and February 5; and Crisis_{-4} equals one between January 9 and 22. Following Borusyak and Jaravel (2017), the first and last two-week intervals of the pre-crisis period are left in the omitted group. We then interact $HFExpo \times MFShock$ with each of the three pre-crisis sub-period indicators and include them as regressors. Column (3) shows that the parallel trends assumption seems to hold in the data. All the pre-crisis interaction terms exhibit little economic and statistical significance. The coefficient of $Crisis \times HFExpo \times MFShock$ is only slightly less than that in column (1) in terms of magnitude and remains negative and highly significant. Controlling for bond-day fixed effects in column (4) does not materially affect the results. We also repeat our analysis by focusing on dealers' liquidity costs excluding RPT trades, which do not require capital commitments and are thus executed at a lower cost. Displayed in columns (5) to (8), the results are qualitatively the same.

There are several confounding factors that may drive our results. First, dealers finance a significant portion of their inventories in the repo markets (Macchiavelli and Zhou, 2022). As a result, both access and cost of repo funding for corporate bond collateral may affect our results. To control for differential access to repo funding, we measure the monthly change in the quantity of repo funding backed by corporate bonds coming from prime MMFs, called Repo Shock. This represents an exogenous shock to the dealers, because in March 2020 prime MMFs faced a run that was unrelated to dealers' exposures to corporate bonds (Li et al., 2021). Columns (1) and (2) of Table 4 shows that controlling for dealers' access to repo funding does not materially affect our results. To further account for the differential cost of repo funding, we also control for the pre-crisis repo rate paid by each dealer to finance corporate bond collateral. Columns (5) and (6) again show that our results hold.

Second, dealers face balance sheet constraints that may hinder their ability to make markets. In particular, the leverage ratio may constrain dealers' willingness to hold inventories on the balance sheet. Alternatively, dealers facing a more binding leverage ratio

may charge higher transaction costs to intermediate a trade. To control for the dealer-level leverage constraint, we construct a measure of how tight the leverage constraint is relative to the minimum requirement. Leverage Intensity$_{pre}$ equals 5 minus the the 2019Q4 SLR divided by 5 for dealers subject to the SLR, and 2 minus the 2019Q4 leverage ratio divided by 2 for dealers that are not subject to the SLR. Leverage Intensity$_{pre}$ is negative for dealers with a leverage ratio above the minimum and converges to zero as dealers get closer to the minimum leverage ratio requirement. Columns (3), (4), (7), and (8) of Table 4 show that leverage constraints do not appear to affect liquidity provision around the COVID liquidity crisis. The coefficient of Leverage Intensity$_{pre}$ is not significant. Importantly, the main coefficient of interest, $Crisis \times HFExpo \times MFShock$, is still negative and statistically significant. The effect is also economically significant. While a one standard deviation selling shock (23.16) increases transaction costs by 10 bps if the dealer has connections with natural buyers in the lowest decile ($HFExpo = 19.05$), the same shock has a 6 bps lower effect if the dealer has connections with natural buyers in the top decile ($HFExpo = 22.12$). As a result, moving a dealer from the bottom to the top decile of natural buyer connections reduces the liquidity decline due to a one standard deviation selling shock by 60%.[7] In sum, having relations with natural corporate bond buyers reduces the liquidity costs associated with making markets in heavily sold bonds during the crisis.

Even though our results are robust to controlling for dealer-level repo funding conditions as well as leverage constraints, the reader may be concerned that our findings could still be explained by some unobservable time-varying dealer characteristics. Given the rich dimensionality of the panel, we can add dealer-day fixed effects and still identify the triple interaction of interest. In Table 5 we show that adding dealer-day fixed effects to our model does not affect our results.

[7]Using the estimated coefficients of Table 4, column (7), a one standard deviation selling shock (23.16) increases transaction costs by $2.014 \times 23.16 - 0.083 \times 23.16 \times 19.05 = 10$ for a dealer with a natural buyer connection in the bottom decile (19.05). The same mutual fund selling shock has a $-0.083 \times 23.16 \times (22.12 - 19.05) = -5.90$ differential effect for a dealer with natural buyer connections in the top decile (22.12). These magnitudes are sizable if compared to the average and median transaction costs of 41 and 18 bps in our sample, respectively.

Finally, one may argue that overall dealer sophistication in prime brokerage activities, rather than specifically its connections to corporate bond hedge funds, may drive our results. If that were the case, we would observe the same results once we substitute the corporate bond exposures of affiliated hedge funds with the equity or Treasury exposures of affiliated hedge funds. To address this concern, we run some placebo tests where we use dealer-level equity or Treasury exposures of affiliated hedge funds instead of corporate bond exposures. Table 6 shows that overall dealer sophistication in prime brokerage activities is an unlikely explanation for our findings. When a dealer's hedge fund connections are captured using either equity or Treasury positions, the coefficient of the triple interaction term no long exhibits any significance. On the other hand, what seems to specifically matter is dealer connections with natural corporate bond buyers.

Overall, our results suggest that in times of market stress and heavy selling by mutual funds, dealers step back from providing liquidity because it becomes challenging to find willing buyers. Afraid of finding themselves on the wrong side of the trade as asset prices are falling, dealers reduce their market making activities instead of buying larger and larger quantities of depreciating bonds from mutual funds. Corroborating this narrative is the evidence that dealers that are connected with natural bond buyers provide more liquidity. With access to both sellers and buyers, a dealer is more willing to make markets and intermediate trades.

In the tumultuous times of March 2020, funding conditions remained quite stable, especially if compared to the 2008 financial crisis in which repo markets for corporate collateral were severely stressed and some dealers lost more repo finding than others (Gorton and Metrick, 2012; Copeland, Martin and Walker, 2014). The relative stability of the repo markets in March 2020 is possibly the reason why repo market conditions seem not to significantly affect liquidity provision during the COVID liquidity crisis. This result is not necessarily in contrast with the fact that repo funding conditions tend to affect liquidity provision over longer time periods (Macchiavelli and Zhou, 2022). Finally, we also find that the proximity to the leverage constraint did not play a significant role

in March 2020. We argue that in a one-sided market, dealers step back from making markets because they are less likely to find buyers and do not want to buy large quantities of bonds and be on the wrong side of a trade while asset prices are falling. In this scenario, the leverage ratio may not play a primary role. This, however, does not mean that the leverage ratio has no impact on market making. On the contrary, it is likely that in normal times the leverage ratio may be a primary factor affecting the cost of carrying inventories. Indeed, Breckenfelder and Ivashina (2021) find evidence that liquidity provision is negatively affected by leverage constraints.

4 Hedge Fund Characteristics and Corporate Bond Liquidity Provision

We have shown that dealer relations with hedge funds that invest in corporate bonds matter for corporate bond liquidity during the COVID liquidity crisis. In the midst of selling pressures from mutual funds, dealers are more willing to step in and provide liquidity if they can turn around and sell those bonds to connected buyers. We now study what factors matter for the ability of hedge funds to behave as bond buyers in times of stress. Specifically, we explore in a monthly panel the hedge fund characteristics that predict increases in corporate bond exposures during the COVID liquidity crisis.

We restrict our sample to hedge funds that have exposures to corporate bonds. The average hedge fund in our sample has $4.36 billion in gross assets and a VaR of 4.35%, which means that there is a 5% chance that the fund may lose 4.35% of its value over the next month. The average fund has a fairly illiquid portfolio, being able to liquidate its assets at little to no cost in 69 days. On the other hand, its average equity investors can withdraw their stakes in 186 days and the average maturity of its debts is 67 days. As a result, the assets of the representative fund are more liquid than its liabilities, allowing it to potentially hold on to illiquid assets for quite some time before it may be forced to sell them to meet redemptions. Consistent with Kruttli et al. (2021), funds with a higher risk

tolerance tend to have managers with more skin in the game (higher manager stake) and display greater return volatility. On the other hand, funds with longer share restrictions and financing duration tend to have lower risk tolerance.

The literature on hedge fund liquidity provision has mainly focused on equity markets and the role of share restrictions and finds mixed results (e.g., Ben-David, Franzoni and Moussawi, 2012; Hombert and Thesmar, 2014; Aragon, Martin and Shi, 2019). Recent work highlights the importance of hedge funds' internal risk limits in the U.S. Treasury market (Kruttli et al., 2021). Corporate bond markets differ notably from equity markets as they trade OTC instead of on an exchange. Further, the corporate bond market is much less liquid than the U.S. Treasury market. Therefore, which hedge fund characteristics drive the liquidity provision might differ for corporate bond trading hedge funds. On the one hand, the lower liquidity of corporate bonds might make share restrictions and liquidity mismatches more important for corporate bond trading than equity trading hedge funds due to the document inverse relationship between a fund's share restrictions and the liquidity of the assets that it trades (Aragon, 2007; Agarwal, Daniel and Naik, 2009; Teo, 2011; Sadka, 2010). On the other hand, the importance of fund internal risk limits as a predictor of hedge fund liquidity provision in the U.S. Treasury market (Kruttli et al., 2021) might also hold for other fixed income markets.

To test these hypotheses, we run the following panel regression model:

$$\Delta \log CorpBondLNE_{h,t} = \gamma_1 Z_{h,t-1} + \gamma_2 Z_{h,t-1} \times Crisis + \mu_h + \theta_t + \varepsilon_{h,t} \tag{3}$$

where $\Delta \log CorpBondLNE_{h,t}$ is the log change of the long corporate bond exposure of hedge fund h at month t, and $Crisis$ is 1 in March 2020 and 0 otherwise. The vector Z includes $RiskLimit$, liquidity mismatch ($LiqMismatch$), the log of net asset values ($LogNAV$), net returns ($NetRetM$), net flows ($NetFlows$), and manager stake ($MgrStake$). To estimate how each of these characteristics contributes to corporate bond exposures during the COVID liquidity crisis, we add their interactions with $Crisis$.

Finally, μ_h and θ_t denote fund and time fixed effects, respectively. Standard errors are clustered at the fund level. The data are monthly from October 2019 to March 2020 and the independent variables, except for the indicator variable $Crisis$, are standardized.

The results are shown in Table 7. Panel A contains the main results and shows that, relative to other funds, hedge funds with more risk tolerance significantly increased their corporate bond exposures in March 2020. The economic magnitude of these estimates is substantial, with a one standard deviation move in the $RiskLimit$ predicting a roughly 10% increase in the corporate bond exposure of a hedge fund. Similarly, larger hedge funds bought corporate bonds during the market turmoil. A higher degree of liquidity transformation indicates that the liabilities of a hedge fund can be redeemed sooner relative to the time it takes for the fund to liquidate its assets at fair value. Interestingly, the degree of liquidity transformation of a fund seems not to be associated with bond exposures.

Panel B of Table 7 contains some robustness tests. In the first two columns, we incorporate funds that do not calculate or report VaR, and thus do not have a defined $RiskLimit$ according to our methodology. In the last two columns, we decompose $LiqMismatch$ into its components $PortIlliq$, $ShareRes$, and $FinDur$. Indeed, it is possible that while liquidity transformation is not associated with bond exposures, some of its components may be. However, none of these components is significant. In particular, funds with longer-term liabilities, in the form of greater share restrictions or longer debt maturity, did not significantly increase bond exposures during the COVID liquidity crisis. In other words, within the set of hedge funds investing in corporate bonds, the stability of their funding structure does not predict bond exposures in times of stress. This result, however, does not negate that hedge funds with less liquid investment strategies (distressed debt instead of Treasury cash-futures basis trades) employ longer-term funding structures, such as longer share restrictions or longer-term repo.

Importantly, across the different specifications, the $RiskLimit$ coefficient estimate remains positive and statistically significant. Hedge funds with greater risk capacity

are willing to absorb more corporate bonds during a market sell-off and hold on to them until market confidence is restored. In late March 2020, the Federal Reserve intervened to avoid further market dislocations (O'Hara and Zhou, 2021) and provided ample funding to dealers. As a result, dealers could support the funding needs of their hedge fund clients. Had there been significant runs on dealers (as during the 2008 financial crisis), we may have seen a deterioration in hedge funds' financing conditions, which in turn may have led to a further bond selloff.

5 Conclusion

The secondary market for corporate bonds relies on dealer intermediation. With infrequent trading, dealers step in between sellers and buyers to provide a timely trade execution and charge a bid-ask spread for the incurred risk. During periods of market turmoil, many investors rush to sell corporate bonds. Dealers face an increasing risk that by the time they find a buyer, the bond they just purchased has already decreased in value. In such a one-side market, dealers' relations with hedge funds become very valuable. Indeed, hedge funds were net buyers of corporate bonds during the 2020 liquidity crisis. Dealers with stronger prime broker relations with hedge funds were better able to offload their positions and reduce their inventory risk. As a result, these dealers could provide more secondary market liquidity.

Consistent with this explanation we find that, for the bonds more heavily sold by mutual funds, dealers with stronger hedge fund connections charged smaller bid-ask spreads. Our results are not driven by other factors that could affect dealers' liquidity provision, such as leverage constraints and repo funding availability and costs. Hedge funds that were better positioned to absorb risk proved to be valuable to dealers, particularly when dealers were facing a one-side market with mutual funds and other institutional investors trying to sell bonds at the same time.

References

Adrian, Tobias, Michael Fleming, Or Shachar, and Erik Vogt. 2017. "Market liquidity after the financial crisis." *Annual Review of Financial Economics*, 9: 43–83.

Agarwal, Vikas, Naveen D Daniel, and Narayan Y Naik. 2009. "Role of managerial incentives and discretion in hedge fund performance." *The Journal of Finance*, 64(5): 2221–2256.

Ambrose, Brent W, Nianyun Kelly Cai, and Jean Helwege. 2008. "Forced selling of fallen angels." *The Journal of Fixed Income*, 18(1): 72–85.

Anderson, Christopher S, David C McArthur, and Ke Wang. 2023. "Internal risk limits of dealers and corporate bond market making." *Journal of Banking & Finance*, 147: 106653.

Anderson, Mike, and René M Stulz. 2017. "Is post-crisis bond liquidity lower?" *NBER Working Paper 23317*.

Aragon, George O. 2007. "Share restrictions and asset pricing: Evidence from the hedge fund industry." *Journal of Financial Economics*, 83(1): 33–58.

Aragon, George O., and Philip E. Strahan. 2012. "Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy." *Journal of Financial Economics*, 103: 570–587.

Aragon, George O, A. Tolga Ergun, Giulio Girardi, and Mila Getmansky Sherman. 2021. "Measuring Hedge Fund Liquidity Mismatch." *Journal of Alternative Investments*, 24(1): 26–42.

Aragon, George O, J Spencer Martin, and Zhen Shi. 2019. "Who benefits in a crisis? Evidence from hedge fund stock and option holdings." *Journal of Financial Economics*, 131: 345–361.

Bao, Jack, Maureen O'Hara, and Xing Alex Zhou. 2018. "The Volcker rule and market-making in times of stress." *Journal of Financial Economics*, 130(1): 95–113.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi. 2012. "Hedge fund stock trading in the financial crisis of 2007–2009." *Review of Financial Studies*, 25(1): 1–54.

Bessembinder, Hendrik, Stacey E Jacobsen, William F Maxwell, and Kumar Venkataraman. 2018. "Capital commitment and illiquidity in corporate bonds." *The Journal of Finance*, 73(4): 1615–1661.

Bessembinder, Hendrik, William Maxwell, and Kumar Venkataraman. 2006. "Optimal market transparency: Evidence from the initiation of trade reporting in corporate bonds." *Journal of Financial Economics*, 82(2): 251–288.

Borusyak, Kirill, and Xavier Jaravel. 2017. "Revisiting event study designs." *Working paper*.

Boyarchenko, Nina, Anna Kovner, and Or Shachar. 2022. "It's what you say and what you buy: A holistic evaluation of the corporate credit facilities." *Journal of Financial Economics*, 144(3): 695–731.

Breckenfelder, Johannes, and Victoria Ivashina. 2021. "Bank balance sheet constraints and bond liquidity." *ECB Working Paper No 2589*.

Choi, Jaewon, Yesol Huh, and Sean Seunghun Shin. 2023. "Customer liquidity provision: Implications for corporate bond transaction costs." *Management Science*, 70(1): 187–206.

Copeland, Adam, Antoine Martin, and Michael Walker. 2014. "Repo Runs: Evidence from the Tri-Party Repo Market." *The Journal of Finance*, 69(6): 2343–2380.

Çötelioğlu, Efe, Francesco Franzoni, and Alberto Plazzi. 2021. "What Constrains Liquidity Provision? Evidence from Institutional Trades." *Review of Finance*, 25(2): 485–517.

Dick-Nielsen, Jens, and Marco Rossi. 2019. "The cost of immediacy for corporate bonds." *Review of Financial Studies*, 32(1): 1–41.

Di Maggio, Marco, Mark Egan, and Francesco Franzoni. 2022. "The value of intermediation in the stock market." *Journal of Financial Economics*, 145(2): 208–233.

Duffie, Darrell. 2022. *Fragmenting Markets: Post-Crisis Bank Regulations and Financial Market Liquidity.* Walter de Gruyter.

Edwards, Amy K, Lawrence Harris, and Michael S Piwowar. 2007. "Corporate bond market transparency and transaction costs." *The Journal of Finance*, 62(3): 1421–1451.

Ellul, Andrew, Chotibhak Jotikasthira, and Christian T Lundblad. 2011. "Regulatory pressure and fire sales in the corporate bond market." *Journal of Financial Economics*, 101(3): 596–620.

Falato, Antonio, Itay Goldstein, and Ali Hortaçsu. 2021. "Financial fragility in the COVID-19 crisis: The case of investment funds in corporate bond markets." *Journal of Monetary Economics*, 123: 35–52.

Glossner, Simon, Pedro Matos, Stefano Ramelli, and Alexander F. Wagner. 2020. "Do Institutional Investors Stabilize Equity Markets in Crisis Periods? Evidence from COVID-19." *Swiss Finance Institute Research Paper No. 20-56.*

Goldstein, Michael A, Edith S Hotchkiss, and Erik R Sirri. 2006. "Transparency and liquidity: A controlled experiment on corporate bonds." *Review of Financial Studies*, 20(2): 235–273.

Gorton, Gary, and Andrew Metrick. 2012. "Securitized banking and the run on repo." *Journal of Financial Economics*, 104(3): 425–451.

Haddad, Valentin, Alan Moreira, and Tyler Muir. 2021. "When selling becomes viral: Disruptions in debt markets in the COVID-19 crisis and the Fed's response." *Review of Financial Studies*, 34(11): 5309–5351.

Han, Munhee, Sanghyun Kim, and Vikram K. Nanda. 2020. "Institutional Brokerage Networks: Facilitating Liquidity Provision." *Available at SSRN: https://ssrn.com/abstract=3221946.*

Harris, Lawrence. 2015. "Transaction costs, trade throughs, and riskless principal trading in corporate bond markets." *Available at SSRN: https://ssrn.com/abstract=2661801.*

Hendershott, Terrence, and Ananth Madhavan. 2015. "Click or call? Auction versus search in the over-the-counter market." *The Journal of Finance*, 70(1): 419–447.

Hombert, Johan, and David Thesmar. 2014. "Overcoming limits of arbitrage: Theory and evidence." *Journal of Financial Economics*, 111(1): 26–44.

Jylhä, Petri, Kalle Rinne, and Matti Suominen. 2014. "Do hedge funds supply or demand liquidity?" *Review of Finance*, 18(4): 1259–1298.

Kargar, Mahyar, Benjamin Lester, David Lindsay, Shuo Liu, Pierre-Olivier Weill, and Diego Zúñiga. 2021. "Corporate bond liquidity during the COVID-19 crisis." *Review of Financial Studies*, 34(11): 5352–5401.

Kruttli, Mathias S, Phillip Monin, and Sumudu W Watugala. 2022. "The Life of the Counterparty: Shock Propagation in Hedge Fund-Prime Broker Credit Networks." *Journal of Financial Economics*, 146(3): 965–988.

Kruttli, Mathias S, Phillip Monin, Lubomir Petrasek, and Sumudu W Watugala. 2021. "LTCM Redux? Hedge Fund Treasury Trading and Funding Fragility." *Available at SSRN: https://ssrn.com/abstract=3817978*.

Levine, Matt. 2015. "People are worried about bond market liquidity." *Bloomberg View*, June–3.

Li, Lei, Yi Li, Marco Macchiavelli, and Xing Zhou. 2021. "Liquidity restrictions, runs, and central bank interventions: Evidence from money market funds." *Review of Financial Studies*, 34(11): 5402–5437.

Macchiavelli, Marco, and Xing Zhou. 2022. "Funding liquidity and market liquidity: the broker-dealer perspective." *Management Science*, 68(5): 3175–3973.

Ma, Yiming, Kairong Xiao, and Yao Zeng. 2022. "Mutual Fund Liquidity Transformation and Reverse Flight to Liquidity." *Review of Financial Studies*, 35: 4674–4711.

O'Hara, Maureen, Andreas C. Rapp, and Xing (Alex) Zhou. 2021. "The Value of Value Investors." *Available at SSRN: https://ssrn.com/abstract=4151934*.

O'Hara, Maureen, and Xing Alex Zhou. 2021. "Anatomy of a liquidity crisis: Corporate bonds in the COVID-19 crisis." *Journal of Financial Economics*, 142(1): 46–68.

Saar, Gideon, Jian Sun, Ron Yang, and Haoxiang Zhu. 2023. "From market making to matchmaking: Does bank regulation harm market liquidity?" *Review of Financial Studies*, 36(2): 678–732.

Sadka, Ronnie. 2010. "Liquidity risk and the cross-section of hedge-fund returns." *Journal of Financial Economics*, 98(1): 54–71.

Schultz, Paul. 2001. "Corporate bond trading costs and practices: A peek behind the curtain." *The Journal of Finance*, 56(2): 677–698.

Schultz, Paul. 2017. "Inventory management by corporate bond dealers." *Available at SSRN: https://ssrn.com/abstract=2966919*.

Shleifer, Andrei, and Robert Vishny. 2011. "Fire Sales in Finance and Macroeconomics." *Journal of Economic Perspectives*, 25(1): 29–48.

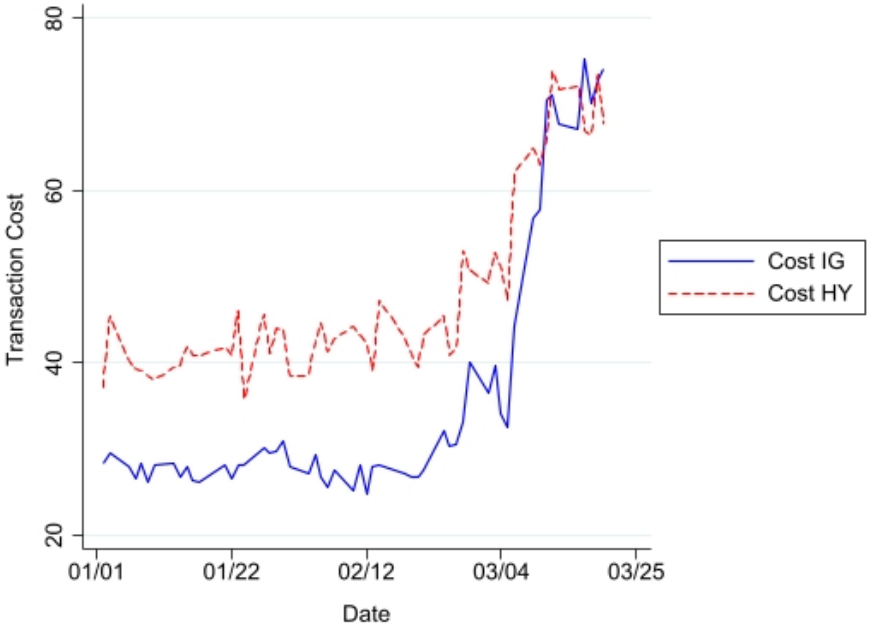
Teo, Melvyn. 2011. "The liquidity risk of liquid hedge funds." *Journal of Financial Economics*, 100(1): 24–44.

Trebbi, Francesco, and Kairong Xiao. 2019. "Regulation and market liquidity." *Management Science*, 65(5): 1949–2443.

Treynor, Jack L. 1987. "The economics of the dealer function." *Financial Analysts Journal*, 43(6): 27–34.

Figure 1: **Transaction costs over time.** This figure shows the evolution of average transaction costs over time for investment grade bonds in solid blue and high yield bonds in dashed red. Panel (a) shows transactions costs including riskless principal trades (RPT), while panel (b) shows transaction costs excluding riskless principal trades. *Sources:* TRACE, authors' calculations.

(a) Transaction Cost (including RPT)



(b) Transaction Cost (excluding RPT)

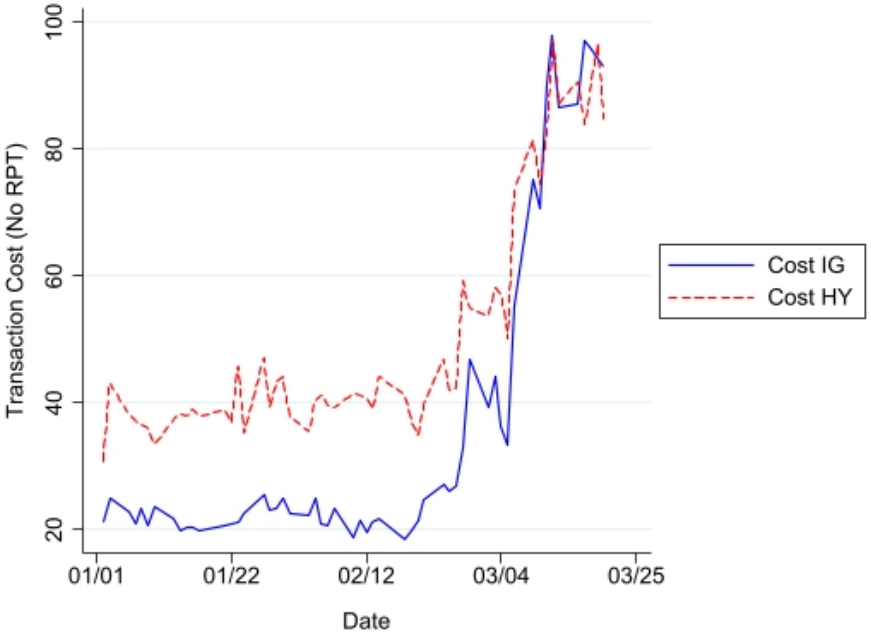


Figure 2: **Mutual fund selling shocks over time.** This figure shows the evolution of average holdings-weighted outflows from mutual funds. *Sources:* eMAXX, Morningstar, authors' calculations.

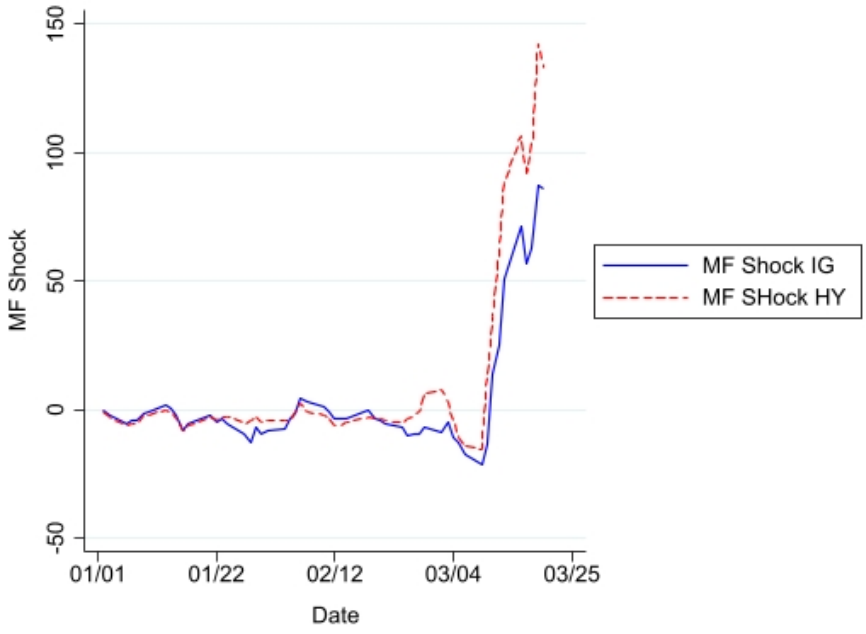


Table 1: Summary Statistics

Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Cost (NoRPT) is the average transaction cost (relative cost of customer trades to inter-dealer trades) excluding riskless principal trades at the dealer-bond-day level. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds. HF Expo is the logarithm of the long exposures in corporate bonds of affiliated hedge funds at the dealer-month level, as of the previous month. Log(TTM) is the logarithm of the time to maturity of each bond. Repo Shock is the monthly change in dealer-level corporate repo outstanding coming from prime MMFs. Repo Rate$_{pre}$ is the average pre-crisis dealer-level corporate repo rate. Leverage Intensity$_{pre}$ is equal to 100 times the minimum leverage ratio minus the 2019:Q4 leverage ratio divided by the minimum. NAV is net asset value in $ million. CorpBondGNE is gross (long plus short) exposures to corporate bonds in $ million. CorpBondLNE and CorpBondSNE are long and short exposures to corporate bonds in$ million, respectively. VaR is value at risk with a 5% probability and a horizon of 1 month. RiskLimit is the 12-month rolling average of VaR. PortIlliq measures the average number of days it would take for the assets to be liquidated at no fire sale discount. ShareRes is the average number of days it would take for investors to withdraw all their funds. FinDur is the weighted average maturity of the fund's borrowings. LiqMismatch is the liqudity of assets relative to the liquidity of liabilities. NetRetM is the monthly return and NetFlows measures investor flows. Finally, MgrStake is the percent of NAV owned by the managers. See Appendix A for more details on the variables.

Panel A: Bond-Dealer-Day Level

Variables	count	mean	st.dev.	p(10)	p(50)	p(90)
Cost	230,555	41.39	54.97	1.10	18.38	117.75
Cost (No RPT)	152,441	39.96	57.03	1.89	17.42	108.57
MF Shock	230,555	38.88	23.16	3.65	39.54	67.86
HF Expo	230,555	20.85	1.11	19.05	20.93	22.12
Log(TTM)	230,555	7.57	1.07	6.18	7.61	9.12
Repo Shock	194,531	2.29	33.36	-26.88	0	15.80
Repo Rate$_{pre}$	201,959	1.75	0.21	1.61	1.77	1.95
Leverage Intensity$_{pre}$	230,555	-26.55	9.98	-41.40	-27.20	-12.50

Summary Statistics (continued)

Panel B: Fund-Month Level

Variables	count	mean	st.dev.	p(10)	p(50)	p(90)
$NAV_{h,t}$ (m US$)	1,419	1,863.68	2,902.84	220.21	952.91	3,964.29
$GAV_{h,t}$ (m US$)	1,419	4,335.31	9,690.56	323.69	1,391.36	8,140.68
$CorpBondGNE_{h,t}$ (m US$)	4,189	383.82	752.96	1.51	111.74	1,013.33
$CorpBondLNE_{h,t}$ (m US$)	4,189	330.47	655.35	0.42	95.51	853.76
$CorpBondSNE_{h,t}$ (m US$)	4,189	53.33	178.85	0.00	0.00	103.33
$VaR_{h,t}$ (%)	1,745	4.35	4.75	0.83	3.07	8.43
$RiskLimit_{h,t}$ (%)	1,568	3.84	4.33	0.94	2.71	6.73
$PortIlliq_{h,t}$ (days)	1,419	68.77	96.10	1.78	23.41	213.27
$ShareRes_{h,t}$ (days)	1,419	186.31	136.55	0.50	185.50	366.00
$FinDur_{h,t}$ (days)	1,147	66.99	106.48	0.50	19.00	273.00
$LiqMismatch_{h,t}$ (days)	1,140	-92.21	94.58	-241.64	-72.02	1.99
$NetRetM_{h,t}$ (%)	3,829	-1.33	5.66	-8.74	0.20	2.55
$NetFlows_{h,t}$ (%)	1,353	-0.76	17.91	-14.67	-0.86	10.75
$MgrStake_{h,t}$ (%)	1,352	12.88	23.79	0.00	3.00	37.00

Table 2: Hedge fund corporate bond exposures by month

This table reports the corporate bond exposure of hedge funds in our sample. Reported are the long and short notional exposure for investment grade and high yield bonds from October 2019 to March 2020.

	Oct-19	Nov-19	Dec-19	Jan-20	Feb-20	Mar-20
$CorpBondIG\_LNE_{h,t}$ (m US$)	86,730	85,300	86,268	85,983	87,309	97,029
$CorpBondIG\_SNE_{h,t}$ (m US$)	20,703	19,744	20,077	21,164	21,555	15,222
$CorpBondHY\_LNE_{h,t}$ (m US$)	157,499	155,718	159,399	162,708	160,300	137,940
$CorpBondHY\_SNE_{h,t}$ (m US$)	27,554	25,962	25,834	25,728	25,397	16,752

Table 3: Bond Liquidity, Mutual Fund Sales, and Hedge Fund Relations.

The sample goes from January 02, 2020 to March 20, 2020. Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Cost (NoRPT) is the average transaction cost (relative cost of customer trades to inter-dealer trades) excluding riskless principal trades at the dealer-bond-day level. Crisis equals one between March 5 and March 20. Crisis_{-2} equals one between February 6 and 19; Crisis_{-3} equals one between January 23 and February 5; and Crisis_{-4} equals one between January 9 and 22. Following Borusyak and Jaravel (2017), the first and last two-week intervals of the pre-crisis period are left in the omitted group. HF Expo is the logarithm of the long exposures in corporate bonds of affiliated hedge funds as of the previous month. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. Log(TTM) is the logarithm of the time to maturity of each bond. Standard errors in parentheses are two-way clustered at the bond and dealer level; ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Cost		Cost		Cost (No RPT)		Cost (No RPT)	
HF Expo×Crisis	-0.556	0.045	-1.486	-0.329	0.893	4.874**	-0.040	4.632**
	(2.276)	(2.357)	(2.036)	(2.345)	(2.498)	(1.899)	(2.304)	(1.898)
MF Shock×Crisis	1.779**		1.656**		1.221**		1.344**	
	(0.662)		(0.611)		(0.568)		(0.562)	
HF Expo×MF Shock ×Crisis_{-4}			-0.019	-0.007			-0.008	0.019
			(0.011)	(0.017)			(0.014)	(0.022)
HF Expo×MF Shock ×Crisis_{-3}			-0.002	-0.007			-0.010	-0.023
			(0.009)	(0.018)			(0.013)	(0.023)
HF Expo×MF Shock ×Crisis_{-2}			0.014	0.014			0.004	-0.002
			(0.012)	(0.018)			(0.012)	(0.018)
HF Expo×MF Shock ×Crisis	-0.073**	-0.065***	-0.069**	-0.061**	-0.059*	-0.081**	-0.064**	-0.089**
	(0.032)	(0.022)	(0.029)	(0.026)	(0.028)	(0.033)	(0.028)	(0.037)
Log(TTM)	33.369***		31.986***		59.140***		60.724***	
	(7.526)		(7.763)		(7.585)		(7.878)	
N	229,856	161,633	229,856	161,633	151,629	93,197	151,629	93,197
R^2	0.335	0.582	0.335	0.582	0.426	0.671	0.427	0.671
Dealer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	No	Yes	No	Yes	No	Yes	No
Bond FE	Yes	No	Yes	No	Yes	No	Yes	No
Bond-Day FE	No	Yes	No	Yes	No	Yes	No	Yes

Table 4: Controlling for Alternative Channels.

The sample goes from January 02, 2020 to March 20, 2020. Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Crisis equals one between March 5 and March 20. HF Expo is the logarithm of the long exposures in corporate bonds of affiliated hedge funds as of the previous month. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. Log(TTM) is the logarithm of the time to maturity of each bond. Repo Shock is the monthly change in dealer-level corporate repo outstanding coming from prime money market funds. It controls for differential access to repo markets. Repo Rate$_{pre}$ is the average pre-crisis dealer-level corporate repo rate, which is switched on during the crisis. It controls for predetermined differences in repo funding costs. Leverage Intensity$_{pre}$ is equal to 100 times the minimum leverage ratio minus the 2019:Q4 leverage ratio divided by the minimum, which is switched on during the crisis. Standard errors in parentheses are two-way clustered at the bond and dealer level; ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	Cost		Cost		Cost		Cost	
HF Expo×Crisis	-1.170	-0.625	2.122	3.867	-2.216	-0.445	0.806	3.806
	(2.475)	(2.324)	(3.478)	(3.659)	(1.965)	(2.118)	(3.602)	(3.674)
MF Shock×Crisis	2.031**		2.291***		1.815**		2.014***	
	(0.761)		(0.589)		(0.651)		(0.570)	
HF Expo×MF Shock	-0.084**	-0.066**	-0.097***	-0.062**	-0.073**	-0.065**	-0.083***	-0.062**
×Crisis	(0.037)	(0.027)	(0.029)	(0.026)	(0.031)	(0.025)	(0.028)	(0.023)
Log(TTM)	29.454***		29.194***		29.240***		29.187***	
	(6.489)		(7.064)		(6.573)		(6.920)	
Repo Shock	-0.005	0.002	-0.005	0.003				
	(0.013)	(0.011)	(0.014)	(0.011)				
Repo Rate$_{pre}$					1.435	-3.335	8.444	2.820
					(19.548)	(10.783)	(14.640)	(7.608)
Leverage Intensity$_{pre}$			-0.447	-0.544			-0.443	-0.538
			(0.639)	(0.445)			(0.554)	(0.405)
N	193,803	126,532	193,803	126,532	202,560	134,757	202,560	134,757
R^2	0.336	0.582	0.336	0.583	0.332	0.578	0.333	0.579
Dealer FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	No	Yes	No	Yes	No	Yes	No
Bond FE	Yes	No	Yes	No	Yes	No	Yes	No
Bond-Day FE	No	Yes	No	Yes	No	Yes	No	Yes

Table 5: Bond Liquidity and Hedge Fund Relations: Within Bond-Day and Dealer-Day.

The sample goes from January 02, 2020 to March 20, 2020. Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Crisis equals one between March 5 and March 20. HF Expo is the logarithm of the long exposures in corporate bonds of affiliated hedge funds as of the previous month. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. HY refers to high yield bonds and FIN to bonds issued by financial companies. Standard errors in parentheses are two-way clustered at the bond and dealer level; ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)
Dependent variable:	Cost		Cost	
HF Expo×Crisis	-0.097	1.264		
	(2.351)	(2.506)		
HF Expo×MF Shock	-0.063**	-0.055*	-0.086***	-0.081***
×Crisis	(0.024)	(0.030)	(0.022)	(0.026)
N	161,606	146,066	161,314	145,666
R^2	0.597	0.692	0.611	0.700
Dealer-Day FE	Yes	Yes	No	No
Bond-Day FE	Yes	Yes	Yes	Yes
Dealer-Day-HY-FIN FE	No	No	Yes	Yes
Dealer-Bond FE	No	Yes	No	Yes

Table 6: Placebo Test: Equity and Treasury Long Positions.

The sample goes from January 02, 2020 to March 20, 2020. Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Crisis equals one between March 5 and March 20. HF Eqty (Tsy) L is the logarithm of the long exposures to equities (Treasuries) of affiliated hedge funds as of the previous month. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. Log(TTM), Repo Shock, and Repo Rate$_{pre}$ are defined in Table 4. Standard errors in parentheses are two-way clustered at the bond and dealer level; ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:		Cost			Cost	
HF Eqty L×Crisis	2.261	3.868	2.368			
	(3.207)	(3.238)	(3.758)			
HF Tsy L×Crisis				-1.577	6.256	5.457
				(5.932)	(8.065)	(7.962)
MF Shock×Crisis	0.783	0.796	0.394	1.223	1.882*	1.510
	(0.624)	(1.159)	(0.715)	(0.925)	(1.039)	(0.975)
HF Eqty L×MF Shock	-0.022	-0.021	-0.005			
×Crisis	(0.025)	(0.047)	(0.029)			
HF Tsy L×MF Shock				-0.041	-0.066	-0.051
×Crisis				(0.038)	(0.043)	(0.040)
Log(TTM)	33.968***	30.619***	30.511***	33.646***	29.481***	29.800***
	(8.009)	(7.505)	(7.251)	(7.297)	(7.183)	(7.139)
Repo Shock		-0.012			-0.031*	
		(0.018)			(0.016)	
Repo Rate$_{pre}$			9.487			8.223
			(15.821)			(16.804)
Leverage Intensity$_{pre}$		-0.697	-0.736		-0.747	-0.778
		(0.536)	(0.459)		(0.602)	(0.532)
N	229,856	193,803	202,560	229,856	193,803	202,560
R^2	0.334	0.336	0.333	0.334	0.337	0.333
Dealer FE	Yes	Yes	Yes	Yes	Yes	Yes
Day FE	Yes	Yes	Yes	Yes	Yes	Yes
Bond FE	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: Hedge fund characteristics and corporate bond trading

This table presents results of the panel regression model given in Equation (3). The dependent variable is $\Delta \log CorpBondLNE_{h,t}$ (in %). The data are monthly from October 2019 to March 2020. All explanatory variables (excluding Crisis) are lagged. The specifications include fund and/or time fixed effects where indicated. The standard errors are clustered at the fund level. The independent variables, with the exception of the indicator variable *Crisis*, are standardized. ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

Panel A: Main

	(1)	(2)	(3)	(4)
Dependent variable:		Δ logCorpBondLNE		
RiskLimit×Crisis	10.080***	10.082***	14.425***	14.431***
	(3.853)	(3.848)	(4.296)	(4.299)
LiqMismatch×Crisis	0.709	0.716	-0.693	-0.679
	(4.981)	(4.984)	(4.962)	(4.963)
LogNAV×Crisis			11.155**	11.181**
			(4.713)	(4.721)
NetRetM×Crisis			-1.538	-1.452
			(5.305)	(5.360)
NetFlows×Crisis			-9.413	-9.281
			(5.964)	(6.010)
MgrStake×Crisis			-6.490	-6.486
			(4.526)	(4.519)
N	1,054	1,054	1,054	1,054
R^2	0.164	0.165	0.195	0.196
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes

Hedge fund characteristics and corporate bond trading (continued)

Panel B: Robustness

	(1)	(2)	(3)	(4)
Dependent variable:		Δ logCorpBondLNE		
RiskLimit×Crisis	12.913***	12.881***	13.748***	13.746***
	(4.157)	(4.155)	(4.318)	(4.317)
NoRiskLimit×Crisis	-4.986	-5.025		
	(5.080)	(5.081)		
LiqMismatch×Crisis	0.308	0.312		
	(2.762)	(2.762)		
PortIlliq×Crisis			10.423	10.415
			(6.919)	(6.923)
ShareRes×Crisis			-3.825	-3.818
			(5.226)	(5.226)
FinDur×Crisis			4.390	4.416
			(6.093)	(6.092)
LogNAV×Crisis	6.466**	6.477**	12.251***	12.276***
	(2.943)	(2.943)	(4.668)	(4.676)
NetRetM×Crisis	0.914	1.061	-2.193	-2.103
	(2.770)	(2.790)	(5.399)	(5.475)
NetFlows×Crisis	4.081	4.135	-10.810*	-10.692*
	(3.091)	(3.088)	(5.628)	(5.663)
MgrStake×Crisis	-4.091	-4.078	-5.399	-5.386
	(3.560)	(3.562)	(4.349)	(4.339)
N	2,599	2,599	1,054	1,054
R^2	0.223	0.224	0.216	0.217
Fund FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes

Online Appendix

This section includes additional material, including variable definitions, figures and tables.

A Variable Definitions

Variable Name	Description
Cost	Relative cost of customer trades to inter-dealer trades at the dealer-bond-day level. See Eq. (1) for more details. *Source:* TRACE.
Cost (No RPT)	Relative cost of customer trades to inter-dealer trades at the dealer-bond-day level, excluding riskless principal trades (RPTs). Following Harris (2015), we denote a trade as RPT if a dealer offsets it within one minute by another trade in the same bond and with the same size but opposite trade direction. *Source:* TRACE.
HF Expo	Logarithm of the long exposures in corporate bonds of affiliated hedge funds as of the previous month. *Source:* Form PF.
MF Shock	Outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. *Source:* eMAXX and Morningstar.
Log(TTM)	Logarithm of the time to maturity of each bond. *Source:* Mergent FISD.
Repo Shock	Monthly change in dealer-level corporate repo outstanding coming from prime money market funds. *Source:* FRBNY.
Repo Rate$_{pre}$	Average pre-crisis dealer-level corporate repo rate multiplied by the Crisis dummy. *Source:* FRBNY.
Leverage Intensity$_{pre}$	The minimum leverage ratio minus the 2019:Q4 leverage ratio times 100, divided by the minimum. It is also multiplied by the Crisis dummy. The minimum is 5% for banks subject to the SLR and 2% otherwise. *Source:* SEC FOCUS, Annual Reports.

Continued on next page

Table A.1 – *Continued from previous page*

Variable	**Description**
RiskLimit	The 12-month rolling average VaR with a time horizon of one month and a probability of 5%. *Source:* Form PF.
Log(NAV)	The logarithm of net asset value, or the amount of investor equity, of the hedge fund. *Source:* Form PF.
PortIlliq	The weighted average time (in days) it would take to liquidate the hedge fund's portfolio, assuming no fire sale discounting. *Source:* Form PF.
ShareRes	The weighted average time (in days) it would take for the investors of the hedge fund to withdraw all the fund's NAV. *Source:* Form PF.
FinDur	The weighted average maturity (in days) of the hedge fund's borrowing. *Source:* Form PF.
NetRet	Net-of-fee monthly returns of the hedge fund. *Source:* Form PF.
NetFlows	Net investor flows to the hedge fund, estimated as $NetFlows_{h,t} = (NAV_{h,t} - NAV_{h,t-1} \times (1 + r_{h,t}))/NAV_{h,t-1}$. *Source:* Form PF.
MgrStake	The percent of the net asset value of the hedge fund owned by the managers or their related persons. *Source:* Form PF.

B Additional Tables

Table B.1: Bond Liquidity, Mutual Fund Sales, and Hedge Fund Relations: Investment Grade and High Yield Split

The sample goes from January 02, 2020 to March 20, 2020. Cost is the average transaction cost (relative cost of customer trades to inter-dealer trades) at the dealer-bond-day level. Cost (NoRPT) is the average transaction cost (relative cost of customer trades to inter-dealer trades) excluding riskless principal trades at the dealer-bond-day level. Crisis equals one between March 5 and March 20. Following Borusyak and Jaravel (2017), the first and last two-week intervals of the pre-crisis period are left in the omitted group. HF Expo is the logarithm of the long exposures in corporate bonds (investment grade in columns (1) and (2), high yield in columns (3) and (4)) of affiliated hedge funds as of the previous month. MF Shock is a CUSIP-level proxy for bond sales by mutual funds during the crisis. It equals the outflow-weighted holdings of a certain CUSIP by corporate bond mutual funds during the crisis period. Log(TTM) is the logarithm of the time to maturity of each bond. Standard errors in parentheses are two-way clustered at the bond and dealer level; ***,**,* indicate statistical significance at 1%, 5%, and 10%, respectively.

	(1) Cost	(2) Cost	(3) Cost	(4) Cost
HF Expo×Crisis	0.662	-0.506	10.298**	10.277*
	(2.451)	(2.749)	(3.948)	(5.226)
MF Shock×Crisis	2.120**		2.923***	
	(0.779)		(0.947)	
HF Expo×MF Shock	-0.090**	-0.076***	-0.119**	-0.147*
×Crisis	(0.039)	(0.026)	(0.044)	(0.071)
Log(TTM)	37.241***		11.383	
	(7.766)		(7.602)	
N	187,689	130,838	42,157	30,795
R^2	0.360	0.595	0.237	0.530
Dealer FE	Yes	Yes	Yes	Yes
Day FE	Yes	No	Yes	No
Bond FE	Yes	No	Yes	No
Bond-Day FE	No	Yes	No	Yes

Trading, Ambiguity and Information in the Options Market

Azi Ben-Rephael* J. Anthony Cookson† and Yehuda Izhakian‡§

September 2023

Abstract

We study how firm ambiguity—Knightian uncertainty—affects investor trading behavior using the options market as a laboratory. Greater ambiguity in the underlying asset negatively relates to both options open interest and options trading volume. The reduction in options trading activity is stronger for options with shorter maturities and out-of-the-money options that are hard-to-value. Greater ambiguity is also associated with a reduction in the informativeness of options trading for future stock prices, and it is associated with lower delta-hedged options returns for both puts and calls. The effect of ambiguity is distinct from and contrasts with the well-documented effect of risk, and it shares a similar economic significance. These findings illustrate that even sophisticated market participants, like options traders, are influenced by ambiguity to limit their market participation and trade less.

Keywords and Phrases: Knightian uncertainty, Options trading, Ambiguity measure, Limited participation, Portfolio inertia, Information inertia, Expected utility with uncertain probabilities.

JEL Classification Numbers: D81, D83, G11, G12

*Rutgers Business School, Rutgers University, 1 Washington Park, Newark, NJ 07102; Email: abenrephael@business.rutgers.edu

†Leeds School of Business, University of Colorado at Boulder, 995 Regent Drive, Boulder CO 80309 ; Email: tony.cookson@colorado.edu

‡Zicklin School of Business, Baruch College, 1 Bernard Baruch Way, New York, NY 10010; Email: yudizh@gmail.com

§We thank Yakov Amihud, Patrick Augustin, Jules Van Binsbergen, Menachem Brenner, Itzhak Ben-David, Richard Herron, Hagit Levy, Dmitry Livdan, Dmitriy Muravyev, Lin Peng, Gideon Saar, Ron Shalev, Elvira Sojli, Andrea Tamoni, Jaime Zender, Dexin Zhou, and Emanuel Zur for valuable discussions and suggestions. Izhakian acknowledges hosting by the Stern School of Business, New York University.

Introduction

Market participation is critical for well-functioning financial markets and is a central feature of many asset pricing models (Campbell, 2006). Indeed, canonical consumption models predict that all individuals ought to participate in financial markets (e.g., Merton, 1975), yet many households are disengaged with financial markets, even wealthy households (e.g., Haliassos and Bertaut, 1995; Briggs et al., 2021). On the other hand, it is well appreciated that sophisticated market participants play an outsized role in shaping market outcomes (Koijen et al., 2020), such that even small changes in sophisticated participants trading behaviors may have important consequences (e.g., Jansen, 2021). Thus, it is a natural and important question to understand what drives the participation and trading decisions of sophisticated investors.

In this paper, we address this question by studying how ambiguity, or Knightian uncertainty,[1] shapes participation and trading decisions in options markets. The options market is a natural laboratory to study participation decisions of sophisticated investors because, as options are in zero net supply, changes in open interest reflect changes in options market participation. Employing firm-day measurement of ambiguity and activity in options markets, we show that greater ambiguity reduces both options market participation and options trading, particularly for difficult-to-value options contracts. Moreover, when ambiguity is high, options trading is less informed for stock prices and both writers and buyers of options earn lower delta-hedged returns. Overall, these findings imply that ambiguity is an important market force, even for options markets that are inhabited by sophisticated traders.

Turning to our empirical design, our contribution is facilitated by employing a recently developed ambiguity measure available at the firm-day level and using it to study options markets. This measure is an empirically-applicable, *risk-independent* measure of ambiguity (Izhakian and Yermack, 2017; Brenner and Izhakian, 2018; Augustin and Izhakian, 2020). This measure estimates firm-level ambiguity from intraday returns data as the volatility of return probabilities. The main advantages of this daily measure are its risk independence and its mitigation of potential confounding effects that are difficult to address using lower frequency (e.g., monthly) proxies. In contemporaneous work, Ben-Rephael et al. (2022) show that this ambiguity measure bears a strong negative relation to daily trading volume in the stock market, and it dampens the relationship

[1] Risk is the condition in which outcomes are a priori unknown, but the odds of all possible outcomes are perfectly known. Ambiguity is the condition in which the possible outcomes are a priori unknown, and the odds of these possible outcomes are either unknown or not uniquely assigned. Knight (1921) defines the concept of (Knightian) uncertainty as distinct from risk since the condition in which the set of events that may occur is a priori unknown and the odds of these events are either unknown or not unique.

between disagreement and stock trading, consistent with a valid empirical proxy for ambiguity.

In this paper, we focus on the options market instead of the underlying asset. This has three main advantages. First, options are held in zero net supply. Since options are a zero sum game, both buyers and sellers are affected by the same uncertainty when ambiguity increases.[2] Second, the rich variation in option contracts allows us to explore the effect of ambiguity along the dimensions of investment maturity and valuation as captured by strike prices. Finally, the fact that multiple option contracts are traded on the same stock allows to better control for firm unobservables.

We begin by studying how call and put options open interest (i.e., holdings that capture the extent of options market participation) relates to the firm-day ambiguity measure at the daily level. We find that high ambiguity is robustly and negatively related to participation. Specifically, a standard deviation increase in ambiguity is associated with between 0.012 and 0.015 standard deviations smaller call (or put) options open interest. Although this estimate reveals a relatively small magnitude along the participation margin, the estimate is highly statistically significant and its economic magnitude is similar to that of intraday volatility, which is known to have a tight connection to options markets. Thus, our core finding is that ambiguity reduces participation of relatively sophisticated options traders, a finding that supports predictions given by ambiguity theory (e.g., Dow and Werlang, 1992; Easley and O'Hara, 2009).

Next, we turn to investigating how ambiguity relates to trading volume. The vast majority of trading volume is driven by activities that, on net, cancel out (e.g., rebalancing and market making activities). We find that ambiguity is also negatively related to options trading volume for both calls and puts. Thus, our evidence on trading volume effects reflects an intensive margin effect, which suggests that ambiguity increases inertia of making a planned trading decision, consistent with ambiguity theory (e.g., Illeditsch, 2011; Illeditsch et al., 2021). Moreover, the estimate is opposite in sign from risk, and of comparable economic magnitude given that the connection between risk and options trading is well-established in the literature (e.g., Bandi et al., 2008).

Observing that ambiguity relates negatively to participation and trading in options markets, we use the richness of the options contracts to provide two more refined tests. Specifically, there are many options contracts available at the same time about the same firm but with different strike prices and different expiration dates. The incentives facing traders of these different contracts

[2]This argument applies in similar force for long versus short positions at the stock level. However, as an empirical matter, stock positions are dominated by long holders since the amount of short selling is small relative to outstanding shares. Note that similar to a firm's bond versus equity holders (Izhakian et al., 2022), call versus put option buyers (and sellers) may face different levels of ambiguity since they face different partitions of the state space.

may be substantially different. Following Muravyev and Ni (2020), we examine heterogeneity in the moneyness and the maturity of options contracts. Consistent with underlying theoretical mechanisms, ambiguity matters most for trading when options are difficult to value. That is, the negative effects of ambiguity on open interest and trading are driven primarily by out of the money options that are more difficult to value. The effects are also more concentrated in the options that expire in the nearer term (within 3 months). We either see the opposite pattern or no consistent pattern with risk, which contrasts with our findings on ambiguity.

Next, we turn to understanding the market implications of a reduction in participation and trading in options markets. First, we investigate how ambiguity moderates the stock price informativeness of options trading. A well-established result in the literature is that options trading, captured by the "put call ratio," is informative of future stock returns (Pan and Poteshman, 2006). We replicate this finding within our sample. Then, in a specification that interacts ambiguity with the put call ratio, we find that a standard deviation increase in ambiguity reduces the stock price informativeness of options trading by roughly 11% of the baseline effect. This return implication is much stronger than the main effect of ambiguity on stock pricing; thus, our evidence reflects an important reduction in the informativeness of options trading.

Finally, we investigate whether ambiguity relates to options returns. Specifically, we relate both ambiguity and risk to delta-hedged cumulative returns over a five-day horizon. Consistent with a classic options pricing perspective, we find a strong positive relation between risk and delta-hedged returns. In contrast, we find that ambiguity relates *negatively* to delta-hedged returns, and our estimate carries a magnitude of 20% to 50% of the economic magnitude of the estimated risk coefficient. These findings are consistent with options being less desirable, suggesting that ambiguity may play a quantitatively important role in options pricing.

Our daily measure of ambiguity is axiomatically rooted and is outcome independent.[3] As such, the measure is theoretically risk independent (Izhakian, 2017, 2020). While other proxies suggested in the literature (e.g., disagreement among analysts' forecasts, VIX, volatility-of-mean, volatility-of-volatility, skewness, and kurtosis) capture various dimensions of uncertainty, they are outcome dependent and therefore risk dependent. Indeed, we find that the correlation between the aforementioned measures and risk is highly positive, whereas the correlation of our ambiguity measure and risk is negative. For example, the correlation between risk and the volatility-of-mean

[3]The ambiguity measure applies exclusively to the probabilities of events, independently of the outcomes associated with these events. Since the measure is outcome independent, the degree of ambiguity does not change if the outcomes associated with events change while the induced partition of the state space into events remains unchanged.

is 0.71, and the correlation between risk and the volatility-of-volatility is 0.57. Moreover, in a set of robustness exercises, we observe that all of the findings regarding $AMBG$ and options trading hold when holding constant the existing proxies in the literature.

We make several contributions to the literature. First, we provide evidence that greater ambiguity dampens options trading intensity at the firm-day level. Most prior work on ambiguity and trading either uses survey data or employs market-level proxies for ambiguity.[4] In this respect, the closest existing research to this paper is Ben-Rephael et al. (2022), which shows that firm-day ambiguity relates negatively to trading in the stock market, and dampens the relation between disagreement and trading. This paper makes a distinct contribution by focusing on how ambiguity in the trading environment affects relatively sophisticated traders who trade in options markets.[5] Moreover, the fact that options are in zero net supply allows us to measure asset participation in a clean way. Besides trading intensity, we show that buyers and sellers actively reduce their positions.

Next, in studying how the trading decisions of options traders depend on ambiguity and risk, we contribute to the literature on what shapes the trading decisions of sophisticated investors. This literature has focused on the informed trades by myriad market participants, such as activists (Collin-Dufresne and Fos, 2015), insiders (Cohen et al., 2012; Augustin et al., 2019), short-sellers (Boehmer et al., 2008; Engelberg et al., 2012) and even options traders (Chakravarty et al., 2004). It is important to understand what drives sophisticated investors to trade because these investors play an outsized role in determining market outcomes (Koijen et al., 2020). At the same time, a broadly held view about sophisticated investors is that they are more immune to non-classical frictions the afflict retail traders. Indeed, much of this research shows that sophisticated investors react to the market conditions (e.g., liquidity and valuation effects) created by other, more behavioral investors (Cookson et al., 2022; Eaton et al., 2021), or that they act in a hyper-informed way with respect to the timing of news (Rogers et al., 2017), and are skilled information processors (Engelberg et al., 2012; Huang et al., 2020). In contrast to this commonly held view, our findings show that even informed and sophisticated options traders respond to ambiguity in the trading environment, and that this behavior matters for the informativeness of options trading and options pricing.

Our paper also contributes to the options literature in at least two aspects. First, we provide

[4]For example, a common proxy for market-level ambiguity is disagreement of analyst forecasts, which has been used to study equity fund flows (Antoniou et al., 2015), aggregate expected return (Anderson et al., 2009), and the term premium (Ulrich, 2013)

[5]The ambiguity measure is drawn from high-frequency trades and quotes in the TAQ database, which is derived from stock market transactions. By studying the separate options market, this paper's findings cannot be driven by omitted microstructure characteristics.

evidence on the link between ambiguity and options pricing. The vast majority of the literature on options has focused on the pricing of volatility (e.g., Bandi et al., 2008; Feunou and Okou, 2019). We show that ambiguity is an important component of options pricing, and thus, it should be taken into account. It operates in an opposite way to the effect of risk and is economically significant. Second, there is a growing interest in the effect of investment horizon (e.g., Dew-Becker and Giglio, 2016; Bandi et al., 2021; Van Binsbergen et al., 2019) and how difficult are securities to be valued (e.g., Kumar, 2009; Baker and Wurgler, 2006; Stambaugh et al., 2015) on trading behavior and pricing. Our findings on options trading show a clear role for maturity and moneyness in shaping options trader incentives. In showing the importance of these aspects of options markets, our findings add to the recent empirical evidence on horizon investments and horizon pricing. We also add to the evidence on hard-to-value securities, which are at the heart of the mispricing literature (van Binsbergen et al., 2021). We show that investors tend to close positions of hard-to-value options earlier in the presence of ambiguity.

Finally, we shed empirical light on the economic effect of ambiguity rather than aversion to it. In this respect, much of the empirical literature focuses on investors' *aversion* to ambiguity, based mostly on experiments. Typically, individuals exhibit aversion to ambiguity, preferring alternatives with clearer probabilities over the ones involving ambiguous probabilities (Ellsberg, 1961). Aversion to ambiguity has been shown to affect human decisions (Halevy, 2007; Crockett et al., 2019) and to be economically relevant, both in experimental market settings (Bossaerts et al., 2010; Ahn et al., 2014) and among business owners and managers (Chesson and Viscusi, 2003). A few studies that use trading data to capture ambiguity aversion are Williams (2014); Thimmea and Völkertb (2014); Li et al. (2016). Relating to work that focuses on ambiguity aversion, we provide direct evidence on the effect of firm level ambiguity and focus on ambiguity itself, distinct from ambiguity aversion.

1 Motivation

In this section, we provide theoretical motivation for our empirical tests, and discuss in greater detail the expected effect of ambiguity on stock options.

1.1 Ambiguity and trading behavior

A common misconception is that ambiguity and risk bear the same implications. However, ambiguity and risk are conceptually different, and might have different implications. To illustrate, consider a decision whose payoff is determined by a flip of an unbalanced coin, for which the investor does not know the odds of heads or tails. The payoff is \$100 in the case of heads, and \$0 in the case

of tails. Suppose that prior to the coin being flipped, the payoff in the case of heads is suddenly changed to \$200. Since no new information about probabilities has been obtained, the investor has no reason to change the assessed probabilities or the perceived degree of ambiguity. Therefore, ambiguity is *outcome independent* up to a state space partition, since it applies exclusively to probabilities. However, the risk does increase in this example, since it is *outcome dependent*.

The literature on decision making under ambiguity has proposed different models, which are "seemingly different [...] rarely related to one another, and often expressed in drastically different formal languages" (Epstein and Schneider, 2010). However, based upon these models, the literature has derived a few complimentary theoretical predictions regarding decision makers' trading behavior in response to ambiguity.

The first prediction is that of limited participation – that is, *when ambiguity associated with a stock increases, the marginal investors reduce their holdings in that stock.* The idea that, for high ambiguity, investors limit their market participation or do not participate at all is supported by several studies. For example, Dow and Werlang (1992) show that for high enough ambiguity or aversion to ambiguity, investors would not participate in the market to the extent that there will be no trade. Cao et al. (2005) show that, when ambiguity dispersion is sufficiently large, investors who face high ambiguity choose not to participate in the stock market. Epstein and Schneider (2007) stress that "an increase in confidence—captured in our model by a posterior set that shrinks over time—induces a quantitatively significant trend towards stock market participation and investment." Easley and O'Hara (2009) attribute limited market participation to aversion to ambiguity. Using similar settings, Ui (2011) shows that, in a rational expectations equilibrium with high enough ambiguity or low enough risk, investors limit their market participation. Finally, using the volatility of aggregate volatility as a measure of ambiguity about market volatility, Kostopoulos et al. (2021) find that ambiguity averse investors reduce their stock market exposure when ambiguity increases.

The second prediction is that of inertia – that is, *when ambiguity associated with a security increases, the marginal investors become more reluctant to rebalance their holding positions and, therefore, adjust their holdings more slowly.* In an extreme case, investors even "freeze up" their trading activity, avoiding rebalancing their holdings. The idea that ambiguity causes investors to adjust their holding more slowly, perhaps for information acquisition, is supported by several studies. For example, Simonsen and Werlang (1991) introduce the concept of portfolio inertia due to ambiguity, and Epstein and Wang (1994) extend it into a more general form. Epstein

and Schneider (2010) characterize the conditions for portfolio inertia. Illeditsch (2011) shows that investors' desire to hedge ambiguity leads to portfolio inertia, especially when facing surprising news. Further, Illeditsch et al. (2021) show that risk and ambiguity aversion may also lead to information inertia, consistent with low trading by households.

These two core predictions suggest that ambiguity, and aversion to it, have a direct effect on investors' trading behavior. Other theoretical work includes, Guidolin and Rinaldi (2010) who show that, for sufficiently high ambiguity, a large portion of traders withdraw from trading and market breakdowns. De Castro and Chateauneuf (2011) show that a greater aversion to ambiguity implies less trading. Easley et al. (2013) investigate the way ambiguity regarding hedge fund investment strategies affects asset prices through trading and liquidity demand.[6]

1.2 Ambiguity and options markets

The derivative market provides a natural laboratory for examining the effect of ambiguity on trading behavior. It provides a direct way to test the aforementioned theoretical predictions empirically. Furthermore, derivative securities allows the refinement of the predictions above regarding the implications of ambiguity for different cases.

Most models of decision-making under ambiguity (e.g., Schmeidler, 1989; Gilboa and Schmeidler, 1989; Bewley, 2002) assert that ambiguity-averse investors act *as if* they overweight the probabilities of bad events (events with negative payoff) and underweight the probabilities of good events (events with positive payoffs). In the perspective of option buyers out-of-the-money is a bad event, and in-the-money is a good event. In the perspective of option writers out-of-the-money is a good event, and in-the-money is a bad event. However, for both option buyers and writers, a higher ambiguity reduces the perceived expected payoff of the option (Augustin and Izhakian, 2020), which motivates both to reduce (or close) their position in the option. In contrast, when risk rises, both buyers and writers are motivated to increase (or open) positions. Buyer may be seeking to increase their hedging or, alternatively, motivated by better speculative opportunities. Writers are motivated by the higher demand and the higher premium. Since options are assets in zero-net supply, these predictions can be directly tested in the options market using options' open interest.

When ambiguity rises, trading volume in option would also decrease, as both buyers and writers decrease (or close) their positions, and less contracts are available for trade. In addition, due to

[6] A further discussion of the implications of ambiguity for trading behavior is provided in recent surveys (e.g., Epstein and Schneider, 2010; Guidolin and Rinaldi, 2013).

(portfolio and information) inertia, trading would slow down, since writers and buyers would be waiting for additional information. Concerning pricing, since the perceived expected payoff for both writers and buyer declines when ambiguity rises, writers would require a higher premium, whereas the buyers would be willing to pay a lower price. Therefore, liquidity would decline and bid-ask spread would increase. However, in the short run a counter effect might accrue, since both writers and buyers may desire to close position quickly. In this respect, other considerations may play a role in options trading behaviour. For example, option writers may be forced to close positions quickly, due to margin constraints.

Besides margin constraints, the option market introduce other aspect that may affect the relation between ambiguity trading behavior. It is well documented that out-of-the-money options are not as strongly related to their underlying assets as in-the-money options, and are therefore more complex to evaluate. For this reason, one would expect out-of-the-money options to be more sensitive to ambiguity and also to risk. Similar to the volatility (risk) process, the ambiguity—the volatility of probabilities—process is a mean-reverting process. Therefore, one would expect short maturity options to be more sensitive to ambiguity than long maturity options. Finally, the perspective of option writers and buyers regarding event classification as good or bad may depend upon their other holdings. For example, in the perspective of naked put option buyers (for speculative motives), in-the-money is a good event. In contrast, in the perspective of protective put option buyers (for hedging motives), in-the-money is a bad event. Therefore, the effect of ambiguity on trading behavior may be different, conditional upon the dominate group.

2 The data

The primary data sources for our analysis are: Intraday Trade and Quote (TAQ) data for the estimation of the daily firm-specific degree of ambiguity, risk, other uncertainty factors (including volatility-of-mean, volatility-of-volatility, skewness and kurtosis) and liquidity; OptionMetrics data for options' trading volume, open interest and liquidity measures; Center for Research in Security Prices (CRSP) data for the estimation of trading volume, number of shares outstanding, and stock prices; and I/B/E/S (IBES) data for analysts' coverage.

In this paper, we focus on exploring the effect of ambiguity on options' trading, pricing, and liquidity. Since options expected value is determined by the ambiguity and risk of the underlying asset, we measure the ambiguity, risk, and other dimension of uncertainty at the stock level.

2.1 Estimating ambiguity

To measure ambiguity, we follow recent literature's (Izhakian and Yermack, 2017; Augustin and Izhakian, 2020; Izhakian et al., 2021) implementation of the expected utility with uncertain probabilities (EUUP, Izhakian, 2017) framework. The primary motivation for using this framework is that it naturally delivers a risk-independent measure of ambiguity, denoted by $\mho^2$.[7] In particular, the degree of ambiguity can be measured by the volatility of uncertain *probabilities*, just as the degree of risk can be measured by the volatility of uncertain *outcomes*. Formally, the measure of ambiguity is defined as:

$$\mho^2 [X] \equiv \int \mathrm{E} [\varphi (x)] \mathrm{Var} [\varphi (x)] \, dx, \tag{1}$$

where $\varphi (\cdot)$ is an uncertain probability density function, and the expectation $\mathrm{E} [\cdot]$ and the variance $\mathrm{Var} [\cdot]$ are taken using the second-order probability measure ξ (i.e., probabilities of probability distributions) on a set $\mathcal{P}$ of probability measures (Izhakian, 2020). The measure of ambiguity defined in Equation (1) is distinct from aversion to ambiguity. The former is a matter of beliefs (or information) and measured from data, while the latter is a matter of subjective attitudes and endogenously determined by the empirical estimations.

To estimate the measure of ambiguity in Equation (1), we use intraday stock data from the TAQ database. We compute the degree of ambiguity for each stock each day. To this end, we elicit a set of priors for each stock each day. We assume that the intraday equity return distribution for each time interval during the day in a given day represents a single prior (probability distribution) in the set of priors and the number of priors in the set is assumed to depend on the number of time intrevals in the day. Each prior in the set is elicited from thirty-second observed intraday returns on the firm's equity, over a time interval of 1170 seconds during the trading hours.[8] Thus, a set of priors consists of 20 realized distributions, at most, over a day. By the principle of insufficient reason (Bernoulli, 1713; Laplace, 1814), each distribution is assigned an equal weight. The rest of the estimation of Equation (1) follows the methodology in Izhakian and Yermack (2017), Augustin and Izhakian (2020), and Izhakian et al. (2021), which for completeness is detailed in Appendix A. We denote the daily estimation of $\mho^2$ by *AMBG*.

[7] In the EUUP framework, a decision-maker possesses a set of priors, equipped with second-order beliefs (i.e., probabilities of probability distributions). An ambiguity-averse decision maker, in this framework, does not compound these probabilities linearly due to her aversion to ambiguity.

[8] Our findings are robust to the use of different time intervals, implying a different number of distributions per day.

2.2 Estimating risk and other moments

In our analysis, we control for risk. For consistency, we measure the daily risk using the same thirty-second returns that are used to measure the degree of ambiguity. For each stock on each time interval, we compute the variance of thirty-second intraday returns. We then measure the firm's daily degree of risk as the mean of these values over the day, normalized to daily terms.[9] We denote the daily estimation of risk by *RISK*. Note that the same variances of returns, estimated over the intraday time intervals, are used in our ambiguity and risk measures.

We estimate the other uncertainty measure similarly. In particular, we measure the volatility-of-mean (VOM) as the variance of the time-interval average return over the day, and the volatility-of-volatility (VOV) as the variance of the time-interval variance over the day. In addition, we use the thirty-second intraday returns to estimate the skewness ($SKEW$) and kurtosis ($KURT$) for each stock in each day, similarly to *RISK*.

2.3 Options trading and liquidity measures

Our main analysis focuses on the daily effect of ambiguity on trading behavior (market participation and liquidity). To this end, we use the option market as a laboratory, as it offers a cleaner setting to study such behavior (e.g., stocks are held in positive supply, and an aggregate exit from the market is not feasible). We employ several measures of option trading and liquidity extracted from OptionMetrics data. To reduce noise due to option contract expiration or unusual maturities, we only consider call and put options with maturities of 7 to 365 calendar days. To reduce noise due to extremely illiquid options, we apply the filters in Muravyev (2016), Christoffersen et al. (2018), and Muravyev and Ni (2020). In particular, we keep option contrasts with absolute delta between 0.1 to 0.9; positive open interest; and a valid bid-ask spread information. We drop contracts with bid-ask spread to midpoint ratio greater than 70%; bid-ask spread greater than \$3; and midpoints lower than \$0.10 cents.

Our first measure of market participation is based on the call and put options open interest (COI and POI, respectively), calculated as the end of the day open interest of call or put options written on the firm equity, divided by the number of its shares outstanding. Open interest allows us to directly explore whether investors reduce their options positions and limit their market participation.[10] Our second measure of market participation is based on the call and put options daily

[9] For robustness, we also apply the Scholes and Williams (1977) correction for non-synchronous trading (e.g., French et al., 1987). The findings are essentially the same.

[10] COI and POI are lagged by one day in OptionMetrics since November 28th, 2000; therefore, we use OptionMetrics reported values from the next trading day.

volume ($CVOL$ and $PVOL$, respectively), calculated as the total daily trading volume of call or put options written on the firm equity, divided by the number of its shares outstanding. Option volume allows us to explore how quickly investors rebalance their option positions. To measure option liquidity, we use the call and put options' bid-ask spread ($CBAS$ and $PBAS$, respectively), based on the end of day bid and ask quotes, divided by the bid-ask spread midpoint.

We control for several additional variables, commonly used in the literature, including the natural logarithm of firm size ($LnSize$), the natural logarithm of firm book-to-market ratio ($LnBM$), institutional holdings ($InstHold$), daily stock return (RET), cumulative 21-day returns ($CumRet$), the natural logarithm of one plus the number of analysts covering the firm ($LnNumEst$), and the natural logarithm of one over the stock average price ($\ln \frac{1}{AvePrc}$).

In addition, we also report statistics and correlations for the stock (the underlying asset) trading volume ($SVOL$), measured by the daily share trading volume divided by the number of shares outstanding. We obtain trading volumes and the number of shares outstanding from CRSP daily data. Finally, we consider the relation between $AMBG$, stock return predictability, and option pricing. Table B.1 details all the variables employed in our analysis.

2.4 Summary statistics

Our main sample consists of 6,766,488 day-firm observations from January 2002 to December 2018 (4,253 trading days) of 4,757 unique firms. It includes all common stocks with Share Code 10 and 11 and a daily price greater than or equal to \$5 (Amihud, 2002). Estimating our main variable of interest, $AMBG$, for every stock and day, requires a minimal number of intraday observations, as detailed in Appendix A. Therefore, our sample starts in January 2002.[11]

Table 1 reports the summary statistics of the pooled sample. Panel A reports statistics for the stock variables. The average (median) firm size is $8,408.14$ ($1,899.96$) million dollars, and the average (median) daily turnover ($SVOL$) is 1.193% (0.805%) of the outstanding shares.

[Table 1]

Panel B reports statistics for the option variables. The average (median) number of call and put options is 15.30 (9.00) and 15.55 (9.00), respectively. The call and put options' average (median) open interest is 0.794% (0.29%) and 0.656% (0.194%) of the outstanding shares, respectively. The average (median) daily trading volume of call and put options is 0.05% (0.005%) and 0.036%

[11] For the period prior to 2002, there is only a very small number of firms that have the sufficient information required to estimate the daily ambiguity measure.

(0.002%) of the outstanding shares, respectively. The trading volume and open interest of call options is higher than that of put options, indicating that call options are more activity traded relative to put options, perhaps due to speculative motives. Finally, the call and put options' average (median) percentage bid-ask spread is 14.05% (11.28%) and 13.02% (10.26%) of the spread midpoint.

Table 2 reports the cross correlations. Panel A reports the univariate correlations between *AMBG*, *RISK*, and the main variables analyzed in the paper. The correlation between *AMBG* and *RISK* is −0.28, implying that, on average, ambiguity is lower on days with high volatility. Note that, as detailed in Appendix A, to estimate ambiguity, we assume that returns are normally distributed. In this class of continuous parametric probability distributions, a change in the parameter of the distribution σ modifies the partition of the state space (Papoulis and Pillai, 2002); thereby, changes the degree of ambiguity. Clearly, a change in σ changes risk.[12] To account for this relation, in all our regression tests, alongside *AMBG*, we control for *RISK* to ensure that our findings are not driven by the correlation between these two uncertainty measures.

[Table 2]

Panel A of Table 2 reveals that *AMBG* is negatively correlated with option trading volume and open interest. *AMBG* is also negatively correlated with stock (the underlying asset) turnover (*SVOL*). Overall, the correlation matrix indicates that an increase in ambiguity is contemporaneously associated with a lower trading activity for both options and the underlying asset, whereas an increase in risk is contemporaneously associated with a higher trading activity.

A few earlier studies use higher distribution moments as proxies for uncertainty. Panel B (Panel C) of Table 2 reports the univariate (multivariate) correlation between *AMBG* and these proxies, providing important insights. Panel B shows that *AMBG* is negatively correlated with *VOM* and *VOV*, with a correlation of −0.18 and −0.08, respectively. At a first glance, one might find these findings surprising, since the variation in the underlying distributions should be positively correlated with the variation in mean and precision of the distribution. However, the strong positive correlation between *RISK* and *VOM* and *VOV* (0.71 and 0.56, respectively) suggests that the relation between *AMBG* and these two proxies is dominated by the latent variable *RISK*. Note that *VOM* and *VOV* are strongly related to *RISK*, since as *RISK* they are outcome dependent.

[12] To see the intuition for the negative relation between ambiguity and risk in this case, suppose that σ increases to infinity. In that case, risk becomes infinite and the degree of ambiguity tends to zero, since all the normal distributions in the set of possible distributions converge to a uniform distribution, implying no uncertainty about the probabilities (i.e., no ambiguity is present).

A subsequent analysis in Columns 2-4 of Panel C reveals that once *RISK* is controlled for, the relation between *AMBG* and *VOM* and *VOV*, becomes positive as expected. Column 5 indicates that kurtosis is also positively correlated with *AMBG*, while skewness is negatively correlated with *AMBG*. We control for all these measures in our analysis. Further, the analysis below shows that *VOM* and *VOV* deliver similar findings to those of *RISK*.

3 Participation and trading in options markets

In this section, we seek to understand the empirical relation between ambiguity and participation in options markets, as well as the relation between ambiguity and trading in options markets (conditional on participation). We expect ambiguity to discourage both participation and trading in options markets, but for different reasons. On the extensive margin, we expect that ambiguity's tendency to shake investor confidence, thereby to decrease participation in options markets (e.g., Easley and O'Hara, 2009). However, even conditional on holding an option contract, ambiguity tends economic agents toward inertia, which would tend to decrease trading volume (Epstein and Schneider, 2010; Illeditsch, 2011).

To evaluate the participation margin, we estimate how ambiguity (*AMBG*) relates to open interest on options, while treating calls and puts separately. If stock options open interest increases for a firm, this is a clear indication of greater participation. Unlike stocks, options are assets in zero net supply. Therefore, greater open interest implies more participation. To evaluate the intensive margin effect on trading, we examine options trading volume directly. As an empirical matter, trading volume reflects mostly trades among active participants (not changes in participation) because trading volume vastly exceeds changes to option open interest on any given day. Thus, variation in trading volume is mostly driven by decisions to buy and sell by traders who, on net, have already decided to participate in the options market.

3.1 Option open interest

We investigate how *AMBG* relates to participation in the options market by relating it to open interest in a firm's option's contracts at the firm-day level in the following specification:

$$OpenInterest_{j,t+i} = \alpha + \beta \cdot AMBG_{j,t} + \gamma \cdot RISK_{j,t} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{j,t}, \quad (2)$$

where the dependent variable, $Open\ Interest_{j,t+i}$, is the open interest in options contracts relating to firm j held on date $t+i$, where i is the number of forward days. We estimate this specification separately for each $i = 0, \ldots, 5$ to illustrate the short run dynamics of open interest. We also

estimate the specification separately for call options and for put options to highlight asymmetries driven by optimism or pessimism about the underlying stock.

The main coefficient of interest is β, which is the coefficient estimate on *AMBG*. To distinguish *AMBG* from underlying riskiness of the stock, persistence of past options participation decisions and other explanations, we include *RISK* and other notable controls in the specification. The vector of controls (*CONTROLS*) includes log firm size (*LnSize*), log book-to-market ratio (*LnBM*), cumulative stock returns (*CumRet*), log of one plus the number of analysts' estimates (*LnNumEst*), institutional holdings (*InstHold*), and log one over average price ($\ln \frac{1}{AvePrc}$), as well as the 21-trading-day trailing average of the dependent variable (Open-Interest), *AMBG* and *RISK*, which account for their persistence. To reduce the effect of outlier observations, all raw variables are trimmed at the top and bottom 0.1% of their sample distribution. We also include firm and date fixed effects across all specifications, and we double cluster standard errors by firm and date to account for persistence over time and common shocks affecting many firms at the same time.

By controlling for *RISK*, we also provide a natural benchmark comparison for the coefficient on *AMBG* to be estimated within the same regression. Prior work has found that *RISK* is strongly and positively related to trading in options markets. Thus, this makes *RISK* a natural control variable to include, while also serving as a useful quantitative benchmark.

[Table 3]

Table 3 reports the findings from estimating Equation (2). Panel A reports findings for call option open interest for trading days t to $t+5$, and Panel B reports the analogous findings for put options. Across all specifications, *AMBG* exhibits a negative and statistically significant relation to option open interest. For call options, a standard deviation increase in ambiguity on date t is associated with a reduction in call option open interest of 0.012 standard deviations. As we consider a longer time lag, the magnitude on the *AMBG* coefficient estimate increases to -0.014. In contrast, the coefficient on *RISK* is much smaller and statistically and economically insignificant by day t+5. Turning to the relation to put option open interest, we observe a similarly strong and significant negative relation between *AMBG* and put option open interest that, like the coefficient estimates in Columns 1 through 5, increases slightly with the time horizon. The coefficient estimates for *RISK*, exhibit similar economic magnitudes to those of *AMBG* and are in the opposite sign.

Our estimated coefficients on *AMBG* reveal a decrease in options market participation that is similar in magnitude for call options and put options. This decrease in participation in options

markets is well predicted by theory (Cao et al., 2005; Easley and O'Hara, 2009), and it contrasts with the pattern of coefficient estimates for *RISK*.[13] By contrast to our findings on relation between *AMBG* and open interest, *RISK* seems to motivate participation, especially in put options.

As a complement to our main analysis, we estimate a vector autoregression (VAR) model to more fully identify the dynamics of the relations of ambiguity and risk to open interest (again, separately for call options and put options). The VAR we consider includes five lags for ambiguity, risk, and open interest, governed by the following equations:

$$OI_{j,t} = \alpha_1 + \sum_{i=1}^{5} \beta_{1,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{1,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{1,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{1,j,t};$$

$$AMBG_{j,t} = \alpha_2 + \sum_{i=1}^{5} \beta_{2,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{2,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{2,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{2,j,t};$$

$$RISK_{j,t} = \alpha_3 + \sum_{i=1}^{5} \beta_{3,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{3,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{3,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{3,j,t},$$

where *CONTROLS* is the same vector of control variables we include in our regression specifications above, measured at date t.

[Figure 1]

The VAR specification allows for nonlinear dynamics and feedback between ambiguity and risk. Despite this different in richness, the VAR delivers similar qualitative findings to our main specifications. Specifically, Panels A and B of Figure 1 present the impulse response function for a standard deviation increase in ambiguity at date t. Consistent with our regression evidence, higher ambiguity is negatively related to participation in options markets, and this effect accumulates over time. Panels C and D present the impulse response functions for *RISK*, showing that risk is positively related to both put and call open interest with a similar accumulation of the effect as the time horizon lengthens.

Overall, we find robust evidence that ambiguity is negatively related to options markets open interest. The economic magnitude of this reduction in option market open interest is comparable to the analogous effect of risk; it is also opposite in sign. This latter finding highlights a sharp economic distinction between ambiguity and risk in options markets. Unlike risk, which encourages options market participation, ambiguity discourages participation in options markets.

[13] Our findings are also in line with prior study by Izhakian and Yermack (2017), who show that higher expected ambiguity motivates the early exercise of options by executives.

3.2 Options trading volume

Having established that *AMBG* exhibits a significant and negative relation to participation in options markets, we now turn our attention to understanding the intensive margin decision to trade options. Since trading volume in options markets vastly exceeds changes to open interest, trading volume in calls and puts mostly reflects these intensive margin decisions.

Thus, we estimate how options trading volume relates to *AMBG* and *RISK* by estimating a specification like the one we used for open interest, but replacing the dependent variable with options trading volume:

$$OptionVolume_{j,t+i} = \alpha + \beta \cdot AMBG_{j,t} + \gamma \cdot RISK_{j,t} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{j,t}, \qquad (3)$$

where the dependent variable, $Option\ Volume_{j,t+i}$. is the trading volume on call options (or put options, separately) on day $t+i$ for options linked to firm j. As in the tests with open interest as the dependent variable, we estimate the relation between *AMBG* and trading volume at date $t+i$ until five trading days later (trading day $t+5$). In addition, we consider how *RISK* relates to options trading volume as a benchmark for the estimated economic magnitudes.

[Table 4]

Table 4 presents the findings from estimating this specification for call trading volume (Panel A) and for put trading volume (Panel B). A standard deviation increase in *AMBG* is associated with a 0.04 standard deviation reduction in call trading volume contemporaneously. The estimated magnitude reduces as we consider longer time lags between *AMBG* and call trading volume. At a five-day lag (day t+5), a standard deviation increase in *AMBG* is associated with only a 0.016 standard deviation decrease in call trading volume. These estimated magnitudes are opposite in sign from the magnitude on the within-day volatility term, *RISK*, and roughly one-third its magnitude: a standard deviation increase in *RISK* is associated with 0.137 standard deviations more call trading volume. This comparison to *RISK* highlights that, although equity market volatility stimulates trading in options markets (positive coefficient estimate on *RISK*), *AMBG* discourages trading. This negative estimate parallels our analogous specification for open interest. However, trading volume mostly reflects trading decisions that are conditional on participation in the options market. In this way, the estimated reduction in trading volume likely reflects a reluctance of existing options traders to trade, not the decision to participate in options markets at all.

Panel B of Table 4 presents a similar pattern for put trading volume for both the coefficient estimates on *AMBG* and *RISK*. Among other things, the similarity in the findings for calls and puts rules out any alternative explanation that predicts a directional movement in options markets.

As a complement to this main analysis, we present evidence from a vector autoregression (VAR) that relates *AMBG* and *RISK* to trading volume of puts (and separately calls). The VAR we estimate follows the same structure as the one we employed in the analysis of options open interest with five lags of *AMBG*, *RISK*, and trading volume in the system of equations.

[Figure 2]

In Figure 2, the impulse response confirms the intuition from the main regression analysis. Notably, Panels A and B illustrate that an increase in ambiguity generates a reduction in both put and call option trading volume that converges relatively quickly to a steady state. By contrast, an increase in risk leads to an increase in both call and put options trading volume, as is illustrated in the impulse responses in Panels C and D.

Overall, the findings on trading volume suggest that ambiguity reduces trading volume in options markets, above and beyond the market participation effects on options market open interest. These findings are consistent with models of ambiguity that predict that ambiguity leads to greater inertia in risky and ambiguous decision making (e.g., Illeditsch, 2011; Illeditsch et al., 2021).

3.3 Heterogeneity by option characteristics

We now exploit the richness of the option contracts to provide a series of more refined tests. Namely, at any given date, there are many different options available that are linked to the same underlying firm. As these options differ on their expiration date and strike price, the incentives facing options traders can be quite different for different options relating to the same underlying security. The literature on options has identified several characteristics that capture the incentives of options traders − most notably, the moneyness of the option (or its *delta*) and the maturity of the option (measured by the time to expiration). We consider heterogeneity in options trading activity by each of these characteristics.

To operationalize the heterogeneity tests in this section, we note that the full underlying data set is at the option contract × firm × date level, and the tests in the previous section collapsed this data set to the firm × date level. We collapse to the group × firm × date level for groups of options contracts that share the same moneyness characteristics or maturity characteristics. For

each characteristic, we split the sample into three groups. We estimate specifications of the form:

$$DepVariable_{j,t+i,g} = \alpha + \sum_{g=2}^{3} \alpha_g GroupDum_{j,t,g} + \sum_{g=1}^{3} \beta_g \cdot AMBG_{j,t} \times GroupDum_{j,t,g} + \sum_{g=1}^{3} \gamma_g \cdot RISK_{j,t} \times GroupDum_{j,t,g} + \delta \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{j,t}, \quad (4)$$

where $DepVariable$ is either $OpenInterest$ or $OptionVolume$, aggregated to the stock-day-group level. The coefficients of interest are the β_g coefficient estimates on the $AMBG\times$ $GroupDum$ terms, which captures how trading activity of options in group g relates to ambiguity at the firm-day level. The degree to which these coefficient estimates are different captures how important the grouping (by moneyness or maturity) is for explaining the heterogeneity in the relation of $AMBG$ to option trading activity. As in the main specifications, we include firm and date fixed effects, and the full set of $CONTROLS$ that we include in the main specifications. The standard errors are clustered by firm and date, which in this specification additionally accounts for cross-correlations within-firm, across options, as well as the usual accounting for serial correlation and common shocks.[14]

3.3.1 Moneyness

An important characteristic of an option is the option's *delta* or Δ, which describes the sensitivity of the option price to the underlying stock price. The Δ is signed, with put options taking on negative values and call options taking on positive values. To place put options and call options on the same footing, we consider delta's absolute value $|\Delta|$ for grouping options by their sensitivity to the underlying stock price. We refer to this sensitivity to the underlying price as *moneyness*, following Muravyev and Ni (2020), and we group options into three groups: "out of the money" ($0.1 \leq |\Delta| \leq 0.4$), "at the money" ($0.4 < |\Delta| < 0.6$), and "in the money" ($0.6 \leq |\Delta| \leq 0.9$).

We present the full estimates of Equation (4) for date t through date $t+5$, separately for call options and put options in Table B.3. The findings of the open interest are reported in Panel A of Table B.3. As in the main tests, the coefficient estimates on $AMBG$ and $RISK$ strengthen slightly from date t to date $t+5$. To summarize the heterogeneity by moneyness, we present plots of these coefficient estimates as of date $t+5$ for each of the three grouped terms for both $AMBG$, and as an instructive benchmark, $RISK$ (both separately for calls and puts). Panel A of Figure 3 indicates that most of the negative relation between $AMBG$ and open interest is driven by out-of-

[14] An alternative strategy to this stacked specification would be to estimate the original specification in Equation (2) separately by group. Such a specification would allow the fixed effects and controls to take on different values by group. We obtain qualitatively similar findings if we estimate such a split-sample specification.

the-money options and at-the-money options ($0.1 \leq |\Delta| \leq 0.4$ and $0.4 < |\Delta| < 0.6$, respectively). In comparison to in-the-money options, these options are more difficult to value, and thus, are more sensitive to the ambiguity in the trading environment. Further, we see a similar pattern for both call and put options, which reinforces our interpretation.

[Figure 3]

By contrast, Panel B of Figure 3 shows that the positive and significant relation between *RISK* and open interest is driven by the impact of *RISK* on in-the-money options only. Apart from the directional difference in the relation to open interest, this difference in the subsample that drives the *RISK* term's relation provides further evidence on the distinction between *AMBG* and *RISK*.

Turning to our evidence on trading volume, we note that the dynamics of the results on trading are distinct because trading volume is not cumulative over the date t to $t+5$ horizon, whereas open interest is. As in the open interest tests, the full heterogeneity results for trading volume are presented in the Panel B of Table B.3. Because the effect of *AMBG* and *RISK* on date t is the strongest, we present plots based on the date t relation to more clearly highlight heterogeneity in the moneyness of the options. Panels C and D of Figure 3 present these plots. In Panel C, we see heterogeneity in the relation of *AMBG* to open interest that is driven by the out-of-the-money options (for both calls and puts). Similar to the participation margin, as ambiguity increases, it tends to discourage trading in lower *delta* options that are more difficult to value (and generally more sensitive to changes in probabilistic assessments). By contrast, in Panel D, we see little heterogeneity in the *RISK* term, either for calls or puts, consistent with the theme that *AMBG* and *RISK* capture distinct economic phenomena related to options markets.

3.3.2 Maturity

Following Muravyev and Ni (2020), we conduct a similar analysis by splitting the option sample into whether they expire soon (*< 3 months*), at an intermediate horizon (*between 3 and 6 months*), or at a long horizon (*> 6 months*). Given this grouping by different option maturities, we estimate analogous specifications to our moneyness heterogeneity tests for both open interest and trading volume. The full estimates are presented in Table B.4. We summarize the heterogeneity in the estimated impact of *AMBG* and *RISK* in Figure 4. Given the cumulative nature of the open interest variable and the short-lived impacts for trading volume, the impact on open interest is considered as of date $t+5$, and on trading volume as of date t. Panels A through D of Figure 4 present these estimates, with separate panels for *AMBG* and *RISK*.

[Figure 4]

Panels A and B of Figure 4 present the heterogeneity by maturity of the estimated relation of *AMBG* and open interest as of date $t + 5$. Consistent with the intuition that near-term expiring options are more sensitive to frictions in the trading environment, we see that most of the negative relation between *AMBG* and option open interest is driven by the shorter maturity options (i.e., those expiring within 3 months of date t). Longer-term options do not exhibit a meaningful relation between *AMBG* and option open interest. By contrast, the heterogeneity in the estimated coefficient of *RISK* with respect to maturity is not meaningful, and it is not consistent across call versus put options.

Panels C and D of Figure 4 present the analogous results on heterogeneity by maturity of the estimated impact of *AMBG* and *RISK* on options trading volume as of date t. Similar to the findings on open interest, the negative relation between *AMBG* and trading volume is driven mostly by a reduction in the trading of shorter maturity options. One rationale for the greater responsiveness of the shorter-term-maturity options to *AMBG* is that the ambiguity measured today is arguably more relevant to the trading decisions regarding options with nearer-in-time expiration dates. Overall, these findings across heterogeneity on maturity support the view that the differences in responsiveness of trading activity to maturity are driven by ambiguity-induced frictions to participating in the options market.

4 Return predictability

Thus far, we have focused on the relation between *AMBG*, market participation, and trading. In this section, we examine the relation between *AMBG* and two aspects of returns: stock return predictability and option pricing. First, we extend existing literature showing that trading in the option market has predictive power for stock returns (e.g., Pan and Poteshman, 2006) by exploring how *AMBG* affects the relation between options trading and stock return predictability. Second, we explore the effect of *AMBG* on option return.

4.1 Stock return predictability

It is well established that options trading contains information about future stock prices (Pan and Poteshman, 2006). Given our findings that ambiguity dampens options market trading, a natural question is how this affects the informativeness of options trading for stock returns. Therefore, we consider how *AMBG* interacts with the informativeness of the direction of trading in options markets. We focus on two measures to link information from the option market and stock returns.

The first measure is a variant of Pan and Poteshman (2006)'s put-call ratio. The second is the implied volatility spread by Cremers and Weinbaum (2010).

Using unique data and methodology, Pan and Poteshman (2006) construct put-call ratio from option volume initiated by buyers who open *new* positions (volume-based put-call ratio). They find that stocks with low put-call ratio outperform stocks with high put-call ratio by more than 40 basis points on the next day and more than 1% over the next week. We build on these findings using information available in the OptionMetrics data. It is not possible within OptionMetrics to distinguish the opening of new positions from the closing of old positions or market making activities that zero out. Therefore, we use changes in open interest to construct the put-call ratio. This approach reduces the noise in constructing an informative volume-based put-call ratio because open interest changes more closely reflect position initiations than trading volume changes.[15] Specifically, we calculate the put-call ratio as the aggregate open interest of put options divided by the sum of the aggregate open interest of put and call options, $PC\_RATIO = P/(C+P)$. Changes in the put-call ratio (ΔPC_RATIO) are taken as the difference between PC_RATIO on day t and PC_RATIO on day t-1.

Cremers and Weinbaum (2010) construct an implied volatility spread measure that captures the difference between call and put implied volatilities for call and put options with the same strike price and maturity. They find that stocks with relatively expensive calls outperform stocks with relatively expensive puts by 50 basis points per week. We follow their methodology and aggregate the information at the stock level using the average call and put open-interest as the weight. While they focus on weekly aggregates, we construct daily spread measures. We denote the measure as *IVS*.

To examine the relation between ambiguity, option information measures ($OPTINFO$), and return predictability, we estimate the following specification:

$$\begin{aligned} DGTWRET_{j,t+1:t+k} \;=\; & \alpha + \beta_1 \cdot AMBG_{j,t} + \beta_2 \cdot RISK_{j,t} + \beta_3 \cdot OPTINFO_{j,t} + \\ & \beta_4 \cdot OPTINFO_{j,t} \times AMBG_{j,t} + \beta_5 \cdot OPTINFO_{j,t} \times RISK_{j,t} + \\ & \Gamma \cdot CONTROLS_{j,t} + \theta_t + \epsilon_{j,t}, \end{aligned} \tag{5}$$

where the dependent variable $DGTWRET$ is the DGTW-adjusted cumulative stock returns of

[15] Indeed, when we repeats the analysis conducted in this subsection using volume-based put-call ratio (instead of open-interest based), we find a negative but weak relation between the volume-based put-call ratio and subsequent stock returns, which amounts to -2 basis points after 10 trading days. We report these findings in Table B.7 for reference.

firm j from day t+1 to $t+10$ (Daniel et al., 1997), and $OPTINFO$ is either trading day t's changes in put-call open interest ratio (ΔPC_RATIO) or trading day t's implied volatility spread (IVS). For example, in the case of ΔPC_RATIO, this specification regresses returns on ΔPC_RATIO, ambiguity ($AMBG$), risk ($RISK$), and the interactions $\Delta PC\_RATIO \times AMBG$ and $\Delta PC\_RATIO \times RISK$. We estimate specifications that consider next-day DGTW returns, and cumulative returns at five-day and ten-day horizons.

Given conventional practice, we include date fixed effects but exclude firm fixed effects from the return based analysis. Results including firm fixed effects are reported in Table B.8. We double cluster standard errors by firm and calendar date to account for serial correlation and correlation within overlapping multiperiod return windows.

Our empirical tests build up to this full specification that include all interaction terms. We start with regressing DGTW returns on $RISK$ and $AMBG$ and the option information measures. This specification provide an estimate of the benchmark relation between $RISK$, $AMBG$ and future stock returns and gives an empirical validation that Pan and Poteshman (2006) and Cremers and Weinbaum (2010) finding holds within our sample, measurement strategy and specification. We then sequentially include the interactions $\Delta PC\_RATIO \times AMBG$ and $\Delta PC\_RATIO \times RISK$. These specifications allow us to quantify the importance of $AMBG$ and $RISK$ in moderating the informativeness of options trading for stock market returns.

[Table 5]

The findings from estimating Equation (5) with ΔPC_RATIO are reported in Panel A of Table 5. To allow for a natural interpretation of the cumulative returns, we present these returns in percentage point units. However, ΔPC_RATIO, $RISK$, and $AMBG$ are all presented in standardized units. Thus, the coefficient estimates for the main effects in the table are a percentage point change in DGTW-adjusted returns for a standard deviation increase in the variable of interest.

Our base specifications (Columns 1, 4, and 7) imply that $AMBG$ exhibit weak stock market predictability. At the one-day horizon, a standard deviation in $AMBG$ is associated with an increase of only 0.5 basis points, which is economically small. The stock return predictability increases somewhat at longer holding periods. For five-day returns, a standard deviation increase in $AMBG$ is associated with returns increasing by 1.6 basis points. For ten-day returns, a standard deviation increase in $AMBG$ is associated with an increase of 2.3 basis points. Though small in magnitude, these findings can be consistent with a risk-based explanation, where $AMBG$ command

a premium in the cross-section of stock returns. The results for *RISK* across the different horizons are mixed consistent with prior evidence.

We also find a strong relation between ΔPC_RATIO and subsequent stock returns, consistent with Pan and Poteshman's (2006) findings. A standard deviation increase in ΔPC_RATIO is associated with roughly 31 to 36 basis points increase of DGTW-adjusted return over the next one to ten trading days. Despite the noise in using OptionMetrics data, we obtain an estimated magnitude that is comparable with Pan and Poteshman's (2006) estimates of 40 basis points for next day return, though smaller than their finding of 100 basis points over a similar horizon.[16] Moreover, in our specification, most of the return predictability occurs on date $t+1$ with non-significant returns related to the put-call ratio in future periods. Panel A of Figure 5 illustrates the return patterns over time.

[Figure 5]

Next, in Columns 2, 5 and 8, we consider the interaction between *AMBG* and ΔPC_RATIO. Consistent with *AMBG* being an important factor for participation and trading, we find that *AMBG* has a positive and significant interaction with ΔPC_RATIO. In particular, a standard deviation increase in *AMBG* is associated with a reduction in the informativeness of ΔPC_RATIO for stock returns by 3.4 to 4.6 basis points for the one to ten day horizon. The effect amounts to approximately 12.8% of ΔPC_RATIO main effect. Including the interaction with *RISK* (Columns 3, 6, and 9) slightly reduces this effect. By comparison to this interaction with *AMBG*, we find that the interaction effect of *RISK* is only around 4.2% of ΔPC_RATIO main effect. The magnitude of the interactive effect for *AMBG* is stronger than ambiguity's main effect; particularly, at the one-day horizon − in Column 3, the main effect of *AMBG* is 0.5 basis points, whereas its interaction with ΔPC_RATIO is 3.4 basis points. The main effect of *AMBG* is similar whether we include the interaction in the specification or not. Thus, this interactive effect is unlikely to reflect any direct effect of *AMBG* on stock return predictability. Taken together, these findings suggest that ambiguity leads to a reduction in option trading informativeness for stock market returns.[17]

[16] We consider standard deviation changes in our specification, whereas Pan and Poteshman's (2006) result corresponds to a long-short quintile approach. A standard deviation change is consistent with a movement from the 16th percentile to the 84th percentile of a normal distribution. Thus, our estimated magnitudes are roughly comparable to the quintile approach.

[17] The dynamic nature of the main effect of *AMBG* versus the interactive effect also supports this interpretation. Notably, the interactive effect is immediately seen in the one-day returns with a slightly larger magnitude by day 10. This contrasts with the main effect, which is very small at the one-day horizon but gradually emerges over the 10-day window. The immediate nature of the interactive effect more closely resembles the main effect of ΔPC_RATIO, which is linked only to trading in options markets.

Next, in Panel B of Table 5 we report the results using Cremers and Weinbaum (2010)'s measure. Columns 1, 4, and 7 confirm the positive return predictability of *IVS* in the cross-section of stock returns. Specifically, a on standard deviation increase in *IVS* is associated with 6.2 to 8.3 basis points. We use daily measures, while Cremers and Weinbaum (2010) use weekly aggregates. Multiplying the coefficient estimates by 5 provides comparable magnitudes to those reported in Cremers and Weinbaum (2010).

Columns 2, 5, and 8 further indicate that *AMBG* has a negative and significant interaction with *IVS*. In particular, a standard deviation increase in *AMBG* is associated with a reduction in the informativeness of *IVS* for stock returns by 0.7 to 2.7 basis points for the one to ten day horizon. The effect amounts to approximately 37% of *IVS* main effect. Interestingly, including the interaction with *RISK* (Columns 3, 6, and 9) attenuates this effect, where it amounts to 23%. The interaction effect of *RISK* seem more important in the case of *IVS* then ΔPC_RATIO, where the effect of *RISK* amounts to 37.7% at the ten day horizon.

Overall, Table 5 indicates that *AMBG* play an important and consistent role in shaping how information flows from the options to the stock market. In particular, an increase in *AMBG* results in lower informativeness of stock prices.

4.2 Option returns

Establishing the importance of *AMBG* for return predictability, we turn to examine how *AMBG* relates to option returns. Given the sensitivity of options to the underlying asset, we follow the convention of reporting results based on delta-hedged returns. In particular, we calculate the options' end of day prices based on the midpoint between the end of day best bid and best ask quotes ($OptionPRC_t$). The option's daily delta-hedged return is then calculated as $[(OptionPRC_t - OptionPRC_{t-1}) - \Delta_{t-1}(StockPRC_t - StockPRC_{t-1})]/OptionPRC_{t-1}$. To aggregate options at the firm level, we form value-weighted portfolios using day t-1 open interest dollar value as the weight, separately for puts versus calls. We fix day $t-1$ open interest dollar value to allow for a natural buy and hold interpretation.

To estimate the impact of *AMBG* on options returns, we estimate the following specification:

$$CumulativeReturns_{j,t:t+k} = \alpha + \beta \cdot AMBG_{j,t} + \gamma \cdot RISK_{j,t} + \Gamma \cdot CONTROLS_{j,t} + \theta_t + \epsilon_{j,t}, \quad (6)$$

where the dependent variable $Cumulative\ Returns_{j,t:t+k}$ is either the delta-hedged cumulative returns on call options (value weighted) from date t to $t+k$, or the analogous cumulative returns

term for put options. To analyze dynamics in the returns, we estimate this specification separately for dates t to $t+5$. We employ the same controls as in the open interest and trading volume regressions. As in Table 5 we exclude the firm fixed effects. We also double cluster by calendar date and firm, which accounts flexibly for serial correlation (e.g., overlapping return windows).

[Table 6]

The findings from estimating Equation (6) are reported in Table 6. The results including firm fixed effects are reported in Table B.9. To allow for a natural interpretation of the cumulative returns, we present the dependent variable in percentage units, as we did for stock returns. *RISK* exhibits a positive relation to both the call and put option returns, as expected from a classic option pricing perspective (e.g., Black and Scholes, 1973). At date t, a standard deviation increase in *RISK* is associated with 31.1 (31.4) basis points increase in call (put) option return, and by the end of day t+5 *RISK* is associated with 65.2 (68) basis points increase in call (put) option return.[18]

In contrast to *RISK*, *AMBG* is *negatively* related to delta-hedged option returns with an estimated magnitude that is sizeable relative to the estimated magnitude for *RISK*. On day t, a standard deviation increase in *AMBG* is associated with 13.8 (18.5) basis points reduction in call (put) option returns. After five days *AMBG* is associated with 18.5 (36) basis points reduction in call (put) option return. Panels B and C of Figure 5 present plots of the dynamics of the delta-hedged return effects. For both *AMBG* and *RISK*, most of the return is accrued on the first few days, quickly converging to no additional effect.

Viewed at a high level, these findings on option returns provide a complementary perspective on the results on participation and trading in options markets. Our findings suggest that high ambiguity reduces options trading, while also decreasing option returns. From an economic perspective, given that options are a zero sum game, these findings jointly point to the interpretation that the option is less desirable when *AMBG* increases, which is a natural consequence of heightened ambiguity. Theoretically, though taking different approaches to model decision-making under ambiguity (e.g., Gilboa and Schmeidler, 1989; Schmeidler, 1989; Wakker and Tversky, 1993), a joint concept of these models is that, in the presence of ambiguity, ambiguity-averse investors act *as if* they overweight the probabilities of unfavorable outcomes and underweight the probabilities

[18] Notably, Cao and Han (2013) document a negative relation between risk and option returns. Importantly, they look at a monthly *RISK* measure and predict the options return over the subsequent month. Consistent with these findings in Table B.10 we report the coefficient estimates of *RISK* and *AMBG* based on their 21-day rolling averages, and recover a negative relation between *AvgRISK* and subsequent options returns.The effect is much smaller than the contemptuous effect of *RISK* on options returns.

of favorable outcomes. All else equal, such a weighting lowers the perceived expected value of the option for both buyers and writers.

5 Robustness and extensions

In this section, we present robustness and extensions to the main findings on options trading and returns.

5.1 *AMBG* and stock options bid-ask spread

First, we consider the relation on *AMBG* to options liquidity, measured by the options' bid-ask spread. One possible mechanism that could explain the findings is that periods of high ambiguity correspond with greater illiquidity, which discourages trading in the options market. To evaluate this possibility, we estimate how the options' bid-ask spread depends on *AMBG* and *RISK* in a panel regression of the same structure as Equation (2), but with the bid-ask spread as the dependent variable. To this end, we use the options' end of day percentage bid-ask spread (relative to the midpoint, *BAS*).

[Table 7]

Table 7 reports the finding from estimating this specification. For both put and call options, we obtain a small and non-significant estimated coefficient on *AMBG* as of date t. On subsequent days, the coefficient estimate on *AMBG* increases, and becomes statistically significant while remaining relatively small. These findings are inconsistent with liquidity effects driving the differences in trading volume. In fact, the response of trading volume to ambiguity may, in part, be responsible for the non-significant result as of date t. As both writers and buyers of the contracts have a mutual incentive to the reduce (close) their positions as ambiguity rises, open interest reduces while liquidity remains constant. In contrast, during subsequent days *AMBG* has a positive effect on the *BAS* as the spreads widen.[19] This finding is consistent with the view portrayed by our options returns results: as ambiguity rises, writers require a higher premium due to a lower perceived expected payoff, while at the same time, buyers offer a lower price for the same reason. Widening spreads is a natural consequence of these market conditions (e.g., Glosten and Milgrom, 1985).

The findings for *RISK* contrast with our main findings on *AMBG*. For call options, *RISK* exhibits a positive relation to the *BAS* on day t and the subsequent trading days, where the effect attenuates over time. For put options, we find a similar pattern, which is slightly weaker relative

[19] We confirm this finding in a set of unreported findings showing a decrease in the *BAS* on day t for out-of-the-money options and options with a short maturity. In contrast, we find an increase in *BAS* for the other options.

to the call options. The positive relation of risk to bid-ask spread (negative effect on liquidity) is also expected and consistent with prior evidence (e.g., Hameed et al., 2010).

5.2 Options trading around news days

In this section, we consider whether the options market participation and trading effects we observe in our main tests also hold around notable firm-specific news events when information comes out about the underlying security. To this end, we repeat the analysis for days t and $t+5$ around firm earnings announcement days, and unscheduled news disclosures (proxied by 8-K filing days).

[Table 8]

Panel A of Table 8 presents the findings on open interest while Panel B presents the findings on options trading volume. Interestingly, we find a consistent amplification of the *AMBG* coefficient estimate for unscheduled news (8-K disclosure dates) for both open interest and trading volume. A standard deviation increase in *AMBG*, implies a reduction in open interest of 0.017 standard deviations (for both calls and puts). This is notably stronger coefficient estimate than the estimates from the main table, which range form -0.012 to -0.015. The trading volume analysis imply a proportionate amplification of the *AMBG* coefficient estimates. For earnings announcement days, we see a weakening of the relation between *AMBG* and call option open interest, but a strengthening of the relation of *AMBG* and put option open interest.

At a high level, the findings from Table 8 imply that the negative relation between ambiguity and options trading is present on identifiable firm news days, implying that public information arrival does not render insignificant the effects of *AMBG* in options markets.

5.3 Subsample analysis by firm size and time period

In this subsection, we repeat our main analysis for subsamples by firm size and subperiods. To conduct the analysis by firm size, we classify firms into tercile subsamples by their market capitalization, and establish dummy variables accordingly. We then interact these dummy variables with *AMBG* and *RISK*, analogous to the heterogeneity specification in Equation (4).

Table B.11 in Appendix B reports the findings. Panel A of reveals an interesting difference between *AMBG* and *RISK*. In particular, the effect of *RISK* on options open interest is more significant for larger firms (3^{rd} tercile). In contrast, the effect of *AMBG* is uniformly present across all terciles. Turning to the effect of *AMBG* and *RISK* on options trading volume, Panel B reveals that the effect of both *AMBG* and *RISK* on trading volume is stronger for larger firms.

Next, we explore whether different time periods affect investors' reaction to ambiguity and risk. We divide our sample into three equal-length subperiods: 2002-2006, 2007-2012, and 2013-2018. Similar to the firm-size heterogeneity, we define dummy variables for each subperiod, and estimate a specification that interacts *AMBG* and *RISK* with these indicator variables.

Table B.12 in Appendix B reports the findings. Similar to the findings reported in Table B.11, *AMBG* presents consistent coefficient estimates across the three subperiods for both options open interest (Panel A) and options trading volume (Panel B). *RISK* also presents consistent coefficient estimates, except for call options open interest where the coefficients change sign across the subperiods.

5.4 Uncertainty factors, dispersion in analyst forecasts and market conditions

In this section, we explore the robustness of our main findings, reported in Tables 3, 4 and 6, by extending our empirical investigation in several ways. First, we explore how the *AMBG* coefficient estimates change when we exclude *RISK* or when we include *RISK* together with other uncertainty proxies. These proxies include skewness (*SKEW*), kurtosis (*KURT*), the volatility-of-mean returns (*VOM*), and the volatility-of-volatility of returns (*VOV*). Second, *VOM* and *VOV* and dispersion of analyst forecasts (*DAF*) are often used as proxies for ambiguity. Thus, we contrast *AMBG* with *VOV*, *VOM*, and *DAF* and explore their directional predictions and economic significance.[20] Third, to differentiate firm-specific shocks in *AMBG* from any market-wide shocks in risk or ambiguity, we include the changes in market volatility (ΔVIX) and changes in market ambiguity ($\Delta MktAMBG$) as additional control variables in our regression specifications. Importantly, across all the tests, our *AMBG* estimates remain intact.

We start by including all other uncertainty factors in our regressions test, alongside *AMBG* and *RISK*. Table B.13 in Appendix B reports the findings. Specifically, we report findings for open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C). To ease the comparison, in all the tests we include the "Base" findings from our main tables. Across all panels and specifications, the findings indicate that excluding all uncertainty proxies or including all of them does not alter our findings with respect to *AMBG*. These findings are consistent with

[20] Our measure of ambiguity $\mho^2$ is broader than *VOM* and *VOV* as it accounts for both, as well as for the volatility of all higher moments of the probability distribution (e.g., skewness and kurtosis) through the variance of probabilities. Furthermore, $\mho^2$ solves some major issues that arise from the use of only the volatility-of-mean or the volatility-of-volatility as proxies for ambiguity. For example, two securities could have a constant mean, but different degrees of ambiguity, or two securities could have constant volatility but different degrees of ambiguity. Second, as opposed to the volatility-of-mean, volatility-of-volatility, and dispersion of analyst forecasts, the measure $\mho^2$ is outcome and risk independent, as it does not depend upon the magnitudes of outcomes, but only upon their probabilities.

the fact that all these uncertainty factors do not explain more than 9% of the variation in *AMBG* (Panel C of Table 2). They are also consistent with the fact that all the other uncertainty factors are outcome dependent and therefore risk dependent, while ambiguity is outcome independent.

Next, we contrast *AMBG* with *VOM* and *VOV*, where we replace *RISK* with *VOM* and *VOV*. The findings of these tests are reported in Table B.14 in Appendix B. Similar to Table B.13, we report findings for open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C). Given our previous findings, it is not surprising that controlling for *VOV* or *VOM* does not alter our *AMBG* coefficient estimates. Since *VOM* and *VOV* are often used as proxies for ambiguity, the directional relation and economic significance of *VOM* and *VOV* is of interest. Starting with *VOM*, we find that across all panels *VOM* coefficient estimates are in the opposite sign of *AMBG* and consistent with the predictions of *RISK*. The economic magnitude is also comparable to *AMBG*. The finding of *VOV* are also consistent with those of *RISK*, except for open interest, where *VOV* shows a negative relation. However, the economic significance is very small compared to *AMBG*. Overall, the findings in Table B.14 demonstrate that the effect of *VOM* and *VOV* is qualitatively similar to that of *RISK*. This is not surprising given that Panel B of Table 2 reveals that the correlations of *VOM* and *VOV* with *RISK* are very high.

Next, we contrast *AMBG* with *DAF*. Table B.15 in Appendix B reports the fidings. Notably, the correlation between *AMBG* and *DAF* is virtually zero as reported in Panel B of Table 2. Thus, it is not surprising that controlling for *DAF* does not alter our findings regarding *AMBG*. What is striking is that higher dispersion in analyst forecasts (updated at a monthly frequency) predicts an increase in open interest and an increase in trading volume. This positive relation is inconsistent with *DAF*'s interpretation as an ambiguity measure. However, it is consistent with *DAF* being a measure of difference-of-opinions or disagreement across analysts. If analyst disagreement correlates with overall disagreement, this is consistent with Cookson and Niessner (2020) who find that higher disagreement is associated with higher trading volume. Moreover, in the options setting, higher disagreement is also associated with more contracts being opened. Finally, in contrast to the findings of *AMBG* and *RISK*, *DAF* has no prediction power for option returns.

In our last set of tests, we explore the robustness of our findings to the inclusion of market risk and market ambiguity. We use the *VIX* as market risk measure, and the ambiguity of the S&P500 index as market ambiguity measure. To capture shocks in these variables, we use the changes in $VIX(\Delta VIX)$ and changes in $MktAMBG(\Delta MktAMBG)$. Since these variables are constructed at the daily level, we replace the day fixed effects with day-of-the-week fixed effects. Overall, the

AMBG coefficient estimates are similar to those reported in our main tables. The only exception is the effect on the delta-hedge returns, where the magnitudes seem to be larger for both *AMBG* and *RISK*. Finally, changes in *MktAMBG* have no significant effect on options open interest, trading volume, or option returns. And, changes in market volatility have a consistent and significant sizeable effect only for option returns.

6 Conclusion

Ambiguity has long been recognized as a theoretical mechanism that can lead to non-participation in financial markets and inertia that inhibits trading (e.g., Easley and O'Hara, 2009; Illeditsch, 2011; Illeditsch et al., 2020). However, to date, empirical support for these theoretical consequences of ambiguity is sparse. This paper fills this important gap between theory and empirics. Specifically, we employ a newly developed daily measure of firm-level ambiguity and options market outcomes to show that both non-participation and inertia are *empirically* important outcomes of ambiguity. Beyond showing empirical support for these classic mechanisms, our findings highlight the ambiguity's importance for trading decisions by relatively sophisticated traders who inhabit options markets.

The reduction in options trading due to ambiguity also tends to reduce the informativeness of options trading (Pan and Poteshman, 2006), which is an important downstream implication of ambiguity's limited participation and inertia effects. Further, we note that greater ambiguity tends to lead to negative and non-reverting delta-hedged option returns for both puts and calls. These option return effects are of comparable economic magnitude to the impact of volatility on options returns. Given the central role volatility plays in options pricing (e.g., Black and Scholes, 1973), our findings suggest that ambiguity ought to also be considered in the pricing of options, given the striking economic magnitudes we find.

A consistent feature of our findings is that the estimated impacts of ambiguity are distinct from those of risk with comparable economic magnitudes. Given this quantitative importance of ambiguity for trading decisions, we anticipate that future work on ambiguity's effects the trading environment will continue to be fruitful. As recent work by Giglio et al. (2021) has articulated, there are many open questions in how investors update their beliefs and trade upon existing belief differences. Since ambiguity impedes acting upon one's beliefs, the linkages between ambiguity, beliefs, and trading decisions is a natural path forward for future research.

References

Ahn, D., Choi, S., Gale, D., Kariv, S., 2014. Estimating ambiguity aversion in a portfolio choice experiment. Quantitative Economics 5, 195–223.

Amihud, Y., 2002. Illiquidity and stock returns: cross-section and time-series effects. Journal of Financial Markets 5, 31–56.

Anderson, E. W., Ghysels, E., Juergens, J. L., 2009. The impact of risk and uncertainty on expected returns. Journal of Financial Economics 94, 233–263.

Antoniou, C., Harris, R. D., Zhang, R., 2015. Ambiguity aversion and stock market participation: An empirical analysis. Journal of Banking & Finance 58, 57–70.

Augustin, P., Brenner, M., Subrahmanyam, M. G., 2019. Informed options trading prior to takeover announcements: Insider trading? Management Science 65, 5697–5720.

Augustin, P., Izhakian, Y., 2020. Ambiguity, volatility, and credit risk. The Review of Financial Studies 33, 1618–1672.

Baker, M., Wurgler, J., 2006. Investor sentiment and the cross-section of stock returns. The journal of Finance 61, 1645–1680.

Bandi, F. M., Chaudhuri, S. E., Lo, A. W., Tamoni, A., 2021. Spectral factor models. Journal of Financial Economics .

Bandi, F. M., Russell, J. R., Yang, C., 2008. Realized volatility forecasting and option pricing. Journal of Econometrics 147, 34–46.

Bayes, T., Price, R., Canton, J., 1763. An essay towards solving a problem in the doctrine of chances. C. Davis, Printer to the Royal Society of London London, U. K.

Ben-Rephael, A., Cookson, J. A., Izhakian, Y., 2022. Do I really want to hear the news? public information arrival and investor beliefs. SSRN eLibrary 3631760.

Bernoulli, J., 1713. Ars Conjectandi (The Art of Conjecturing).

Bewley, T. F., 2002. Knightian decision theory. part i. Decisions in economics and finance 25, 79–110.

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637–54.

Boehmer, E., Jones, C. M., Zhang, X., 2008. Which shorts are informed? The Journal of Finance 63, 491–527.

Bossaerts, P., Ghirardato, P., Guarnaschelli, S., Zame, W. R., 2010. Ambiguity in asset markets: Theory and experiment. The Review of Financial Studies 23, 1325–1359.

Brenner, M., Izhakian, Y., 2018. Asset prices and ambiguity: Empirical evidence. Journal of Financial Economics 130, 503–531.

Briggs, J., Cesarini, D., Lindqvist, E., Östling, R., 2021. Windfall gains and stock market participation. Journal of Financial Economics 139, 57–83.

Campbell, J. Y., 2006. Household finance. The journal of finance 61, 1553–1604.

Cao, H. H., Wang, T., Zhang, H. H., 2005. Model uncertainty, limited market participation, and asset prices. The Review of Financial Studies 18, 1219–1251.

Cao, J., Han, B., 2013. Cross section of option returns and idiosyncratic stock volatility. Journal of Financial Economics 108, 231–249.

Chakravarty, S., Gulen, H., Mayhew, S., 2004. Informed trading in stock and option markets. The Journal of Finance 59, 1235–1257.

Chesson, H. W., Viscusi, W. K., 2003. Commonalities in time and ambiguity aversion for long-term risks. Theory and Decision 54, 57–71.

Christoffersen, P., Goyenko, R., Jacobs, K., Karoui, M., 2018. Illiquidity premia in the equity options market. The Review of Financial Studies 31, 811–851.

Cohen, L., Malloy, C., Pomorski, L., 2012. Decoding inside information. The Journal of Finance 67, 1009–1043.

Collin-Dufresne, P., Fos, V., 2015. Do prices reveal the presence of informed trading? The Journal of Finance 70, 1555–1582.

Cookson, J. A., Fos, V., Niessner, M., 2022. Does disagreement facilitate informed trading? SRRN (April 18, 2022) .

Cookson, J. A., Niessner, M., 2020. Why don't we agree? evidence from a social network of investors. The Journal of Finance 75, 173–228.

Cremers, M., Weinbaum, D., 2010. Deviations from put-call parity and stock return predictability. Journal of Financial and Quantitative Analysis 45, 335–367.

Crockett, S., Izhakian, Y. Y., Jamison, J. C., 2019. Ellsberg's hidden paradox. SSRN eLibrary 3423534.

Daniel, K., Grinblatt, M., Titman, S., Wermers, R., 1997. Measuring mutual fund performance with characteristic-based benchmarks. The Journal of finance 52, 1035–1058.

De Castro, L. I., Chateauneuf, A., 2011. Ambiguity aversion and trade. Economic Theory 48, 243–273.

Dew-Becker, I., Giglio, S., 2016. Asset pricing in the frequency domain: theory and empirics. The Review of Financial Studies 29, 2029–2068.

Dow, J., Werlang, S. R. d. C., 1992. Uncertainty aversion, risk aversion, and the optimal choice of portfolio. Econometrica 60, 197–204.

Easley, D., O'Hara, M., 2009. Ambiguity and nonparticipation: The role of regulation. The Review of Financial Studies 22, 1817–1843.

Easley, D., O'Hara, M., Yang, L., 2013. Opaque trading, disclosure, and asset prices: Implications for hedge fund regulation. The Review of Financial Studies 27, 1190–1237.

Eaton, G. W., Green, T. C., Roseman, B., Wu, Y., 2021. Retail trader sophistication and stock market quality: Evidence from brokerage outages. Available at SSRN 3776874 .

Ellsberg, D., 1961. Risk, ambiguity, and the savage axioms. The Quarterly Journal of Economics 75, 643–669.

Engelberg, J. E., Reed, A. V., Ringgenberg, M. C., 2012. How are shorts informed?: Short sellers, news, and information processing. Journal of Financial Economics 105, 260–278.

Epstein, L. G., Schneider, M., 2007. Learning under ambiguity. The Review of Economic Studies 74, 1275–1303.

Epstein, L. G., Schneider, M., 2010. Ambiguity and Asset Markets. Annual Review of Financial Economics 2, 315–346.

Epstein, L. G., Wang, T., 1994. Intertemporal asset pricing under knightian uncertainty. Econometrica 62, 283–322.

Fama, E. F., French, K. R., 1992. The cross-section of expected stock returns. The Journal of Finance 47, 427–465.

Faria, G., Correia-da Silva, J., 2014. A closed-form solution for options with ambiguity about stochastic volatility. Review of Derivatives Research 17, 125–159.

Feunou, B., Okou, C., 2019. Good volatility, bad volatility, and option pricing. Journal of financial and quantitative analysis 54, 695–727.

Franzoni, L. A., 2017. Liability law under scientific uncertainty. American Law and Economics Review 19, 327–360.

French, K. R., Schwert, G. W., Stambaugh, R. F., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Giglio, S., Maggiori, M., Stroebel, J., Utkus, S., 2021. Five facts about beliefs and portfolios. American Economic Review 111, 1481–1522.

Gilboa, I., Schmeidler, D., 1989. Maxmin expected utility with non-unique prior. Journal of Mathematical Economics 18, 141–153.

Glosten, L. R., Milgrom, P. R., 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. Journal of financial economics 14, 71–100.

Guidolin, M., Rinaldi, F., 2010. A simple model of trading and pricing risky assets under ambiguity: any lessons for policy-makers? Applied Financial Economics 20, 105–135.

Guidolin, M., Rinaldi, F., 2013. Ambiguity in asset pricing and portfolio choice: A review of the literature. Theory and Decision 74, 183–217.

Halevy, Y., 2007. Ellsberg revisited: an experimental study. Econometrica 75, 503–536.

Haliassos, M., Bertaut, C. C., 1995. Why do so few hold stocks? the economic Journal 105, 1110–1129.

Hameed, A., Kang, W., Viswanathan, S., 2010. Stock Market Declines and Liquidity. The Journal of Finance 65, 257–293.

Huang, A. G., Tan, H., Wermers, R., 2020. Institutional trading around corporate news: Evidence from textual analysis. The Review of Financial Studies 33, 4627–4675.

Illeditsch, P. K., 2011. Ambiguous information, portfolio inertia, and excess volatility. The Journal of Finance 66, 2213–2247.

Illeditsch, P. K., Ganguli, J. V., Condie, S., 2020. Information inertia. The Journal of Finance Forthcoming.

Illeditsch, P. K., Ganguli, J. V., Condie, S., 2021. Information inertia. The Journal of Finance 76, 443–479.

Izhakian, Y., 2017. Expected utility with uncertain probabilities theory. Journal of Mathematical Economics 69, 91–103.

Izhakian, Y., 2020. A theoretical foundation of ambiguity measurement. Journal of Economic Theory 187, 105001.

Izhakian, Y., Yermack, D., 2017. Risk, ambiguity, and the exercise of employee stock options. Journal of Financial Economics 124, 65–85.

Izhakian, Y., Yermack, D., Zender, J., 2021. Ambiguity and the trade off theory of capital structure. Management Science 68, 1526–5501.

Izhakian, Y., Zender, J., Lewis, R., 2022. Ambiguity and corporate yield spreads. SSRN eLibrary 4066052.

Jansen, K. A., 2021. Long-term investors, demand shifts, and yields. Available at SSRN 3901466 .

Jaynes, E. T., 1957. Information theory and statistical mechanics. Physical review 106, 620.

Kendall, M., Stuart, A., 2010. The advanced theory of statistics. vol. 1: Distribution theory. London: Griffin, 2010, 6th ed. 1.

Knight, F. M., 1921. Risk, Uncertainty and Profit. Houghton Mifflin, Boston.

Koijen, R. S., Richmond, R. J., Yogo, M., 2020. Which investors matter for equity valuations and expected returns? Tech. rep., National Bureau of Economic Research.

Kostopoulos, D., Meyer, S., Uhr, C., 2021. Ambiguity about volatility and investor behavior. Journal of Financial Economics .

Kumar, A., 2009. Hard-to-value stocks, behavioral biases, and informed trading. Journal of Financial and Quantitative Analysis 44, 1375–1401.

Laplace, P. S., 1814. Théorie analytique des probabilitiés.

Li, C. W., Tiwari, A., Tong, L., 2016. Investment decisions under ambiguity: Evidence from mutual fund investor behavior. Management Science 63, 2509–2528.

Merton, R. C., 1975. Optimum consumption and portfolio rules in a continuous-time model. In: *Stochastic optimization models in finance*, Elsevier, pp. 621–661.

Muravyev, D., 2016. Order flow and expected option returns. The Journal of Finance 71, 673–708.

Muravyev, D., Ni, X. C., 2020. Why do option returns change sign from day to night? Journal of Financial Economics 136, 219–238.

Pan, J., Poteshman, A. M., 2006. The information in option volume for future stock prices. The Review of Financial Studies 19, 871–908.

Papoulis, A., Pillai, S. U., 2002. Probability, random variables, and stochastic processes. Tata McGraw-Hill Education.

Rogers, J. L., Skinner, D. J., Zechman, S. L., 2017. Run edgar run: Sec dissemination in a high-frequency world. Journal of Accounting Research 55, 459–505.

Schmeidler, D., 1989. Subjective probability and expected utility without additivity. Econometrica 57, 571–587.

Scholes, M., Williams, J., 1977. Estimating betas from nonsynchronous data. Journal of Financial Economics 5, 309–327.

Simonsen, M. H., Werlang, S. R. d. C., 1991. Subadditive Probabilities and Portfolio Inertia. Brazilian Review of Econometrics 11.

Stambaugh, R. F., Yu, J., Yuan, Y., 2015. Arbitrage asymmetry and the idiosyncratic volatility puzzle. The Journal of Finance 70, 1903–1948.

Thimmea, J., Völkertb, C., 2014. Ambiguity in the cross-section of expected returns: An empirical assessment. Journal of Business & Economic Statistics Forthcoming.

Ui, T., 2011. The ambiguity premium vs. the risk premium under limited market participation. Review of Finance 15, 245–275.

Ulrich, M., 2013. Inflation ambiguity and the term structure of u.s. government bonds. Journal of Monetary Economics 60, 295–309.

van Binsbergen, J. H., Boons, M., Opp, C. C., Tamoni, A., 2021. Dynamic asset (mis) pricing: Build-up vs. resolution anomalies. Resolution Anomalies (February 12, 2021) .

Van Binsbergen, J. H., Han, J., Ruan, H., Xing, R., 2019. A horizon based decomposition of mutual fund value added using transactions. Available at SSRN 3478745 .

Wakker, P. P., Tversky, A., 1993. An axiomatization of cumulative prospect theory. Journal of Risk and Uncertainty 7, 147–175.

Williams, C. D., 2014. Asymmetric responses to earnings news: A case for ambiguity. Accounting Review Forthcoming.

Figure 1: Impulse response functions of call and put open interest

This figure plots the impulse responses of call and put open interest to a one-standard-deviation shock to *AMBG* and *RISK*. For each call and put open interest (*OI*), it estimates a daily vector autoregression (VAR) system of *OI*, *AMBG*, and *RISK*, with five lags of each variable. All variables are defined in Table B.1, where *AMBG*, *RISK*, and *OI* are trimmed at the top and bottom 0.1% of their sample distribution. All regression tests include the full set of firm control variables together with firm fixed effects and date fixed effects. The VAR system takes the following form

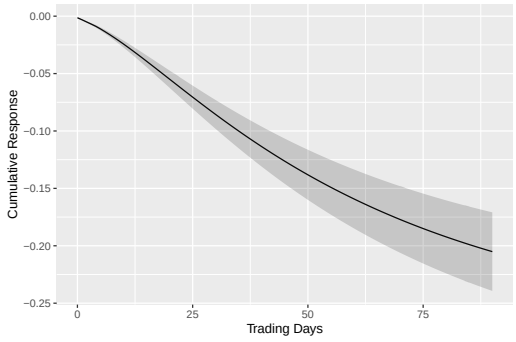
$$OI_{j,t} = \alpha_1 + \sum_{i=1}^{5} \beta_{1,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{1,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{1,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{1,j,t};$$

$$AMBG_{j,t} = \alpha_2 + \sum_{i=1}^{5} \beta_{2,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{2,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{2,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{2,j,t};$$

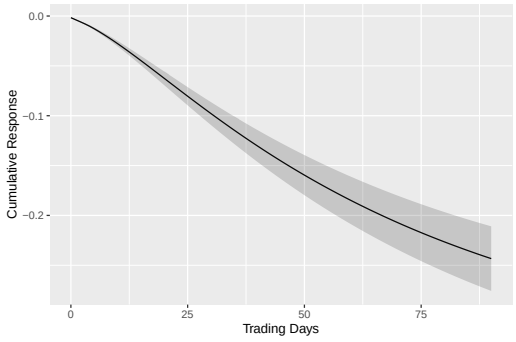
$$RISK_{j,t} = \alpha_3 + \sum_{i=1}^{5} \beta_{3,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{3,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{3,i} \cdot OI_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{3,j,t}.$$

The estimated coefficients of this system are reported in Table B.2. This figure includes two pairs of graphs, one for *AMBG* and one for *RISK*Each pair plots the cumulative response of *DEP* to a one-standard-deviation shock to *AMBG* (upper graphs) and to *RISK* (lower graph). To estimate the effect of *AMBG* (*RISK*) on *DEP*, the Cholesky order is set to be *RISK*, *AMBG*, *DEP* (*AMBG*, *RISK*, *DEP*). Each graph depicts the response in the subsequent $0, \ldots, 90$ trading days, listed on the x-axis. The solid line depicts the variable response and the dashed gray lines depict the 95% confidence intervals.

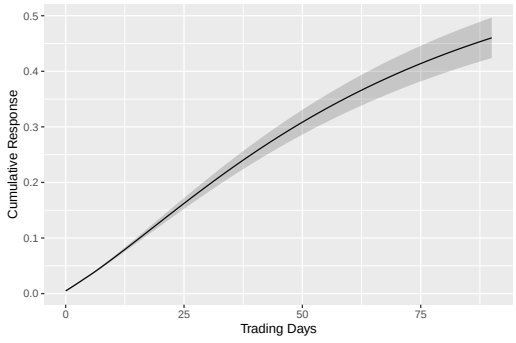
Panel A: Response of call option open interest to firm ambiguity



Panel B: Response of put option open interest to firm ambiguity



Panel C: Response of call option open interest to firm risk



Panel D: Response of put option open interest to firm risk

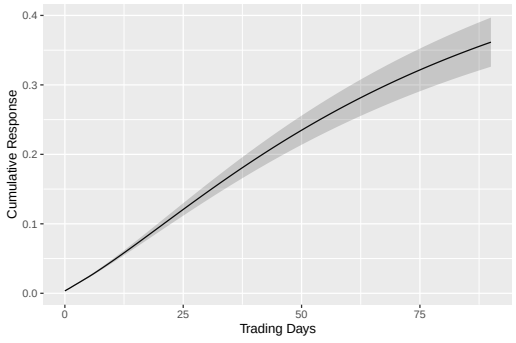


Figure 2: Impulse Response Functions of call and put trading volume

This figure plots the impulse responses of call and put trading volume to a one-standard-deviation shock to *AMBG* and *RISK*. For each call and put trading volume (VOL), it estimates a daily vector autoregression (VAR) system of VOL, *AMBG*, and *RISK*, with five lags of each variable. All variables are defined in Table B.1, where *AMBG*, *RISK*, and VOL are trimmed at the top and bottom 0.1% of their sample distribution. All regression tests include the full set of firm control variables together with firm fixed effects and date fixed effects. The VAR system takes the following form

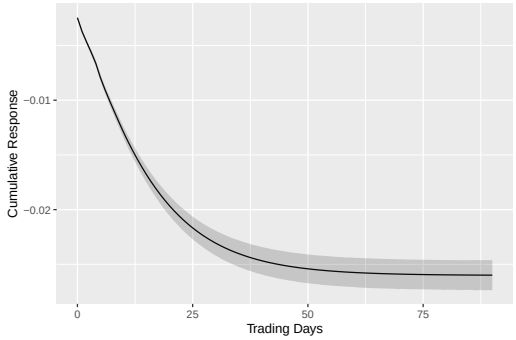
$$VOL_{j,t} = \alpha_1 + \sum_{i=1}^{5} \beta_{1,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{1,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{1,i} \cdot VOL_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{1,j,t};$$

$$AMBG_{j,t} = \alpha_2 + \sum_{i=1}^{5} \beta_{2,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{2,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{2,i} \cdot VOL_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{2,j,t};$$

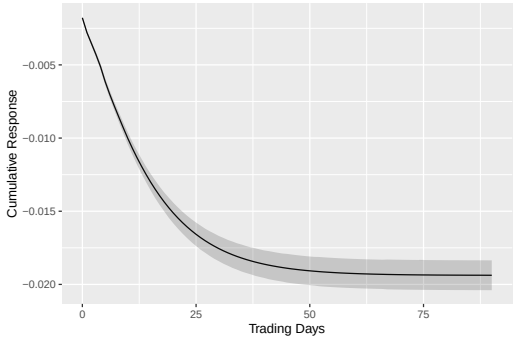
$$RISK_{j,t} = \alpha_3 + \sum_{i=1}^{5} \beta_{3,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{3,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{3,i} \cdot VOL_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{3,j,t}.$$

The estimated coefficients of this system are reported in Table B.2. This figure includes two pairs of graphs, one for *AMBG* and one for *RISK*Each pair plots the cumulative response of *DEP* to a one-standard-deviation shock to *AMBG* (upper graphs) and to *RISK* (lower graph). To estimate the effect of *AMBG* (*RISK*) on *DEP*, the Cholesky order is set to be *RISK*, *AMBG*, *DEP* (*AMBG*, *RISK*, *DEP*). Each graph depicts the response in the subsequent $0, \ldots, 90$ trading days, listed on the x-axis. The solid line depicts the variable response and the dashed gray lines depict the 95% confidence intervals.

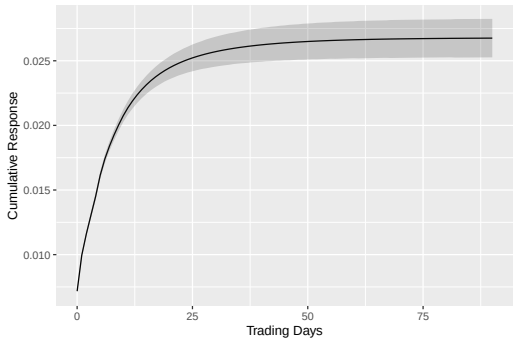
Panel A: Response of call option volume to firm ambiguity



Panel B: Response of put option volume to firm ambiguity



Panel C: Response of call option volume to firm risk



Panel D: Response of put option volume to firm risk

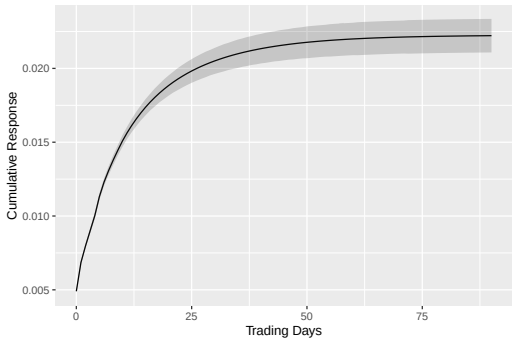
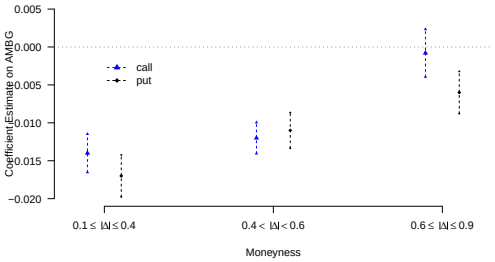


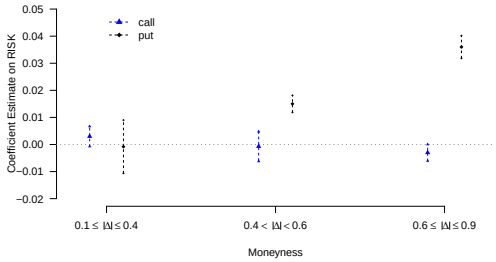
Figure 3: The effect of *AMBG* and *RISK* on options' open interest and trading volume based on moneyness

This figure plots the coefficient estimates of *AMBG* and *RISK* from daily panel regressions, in which call and put stock option open interest or trading volume on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), risk (*RISK*), and other firm characteristics based on moneyness. In particular, for each firm and day we aggregate options open interest (Graphs A-B) or trading volume (Graphs C- D) based on contract moneyness. The moneyness groups are defined as $0.1 <= |\Delta| <= 0.40$, $0.40 < |\Delta| < 0.60$, and $0.60 < |\Delta| <= 0.90$, respectively. To estimate the coefficients we stack each firm daily measures in the same regression and interact *AMBG* and *RISK* with dummy variables based on the three defined moneyness groups.The regression results are reported in Table B.3. The graphs below plot the regressions' coefficient estimates of open interest (trading volume) from trading day t+5 (t) together with their 95% confidence intervals.

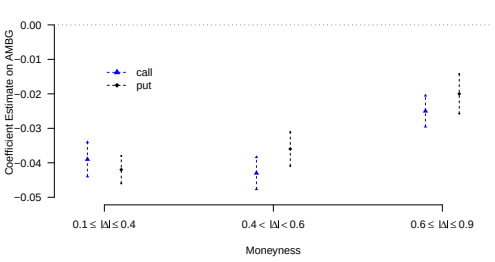
Panel A: *AMBG* and open interest



Panel B: *RISK* and open interest



Panel C: *AMBG* and trading volume



Panel D: *RISK* and trading volume

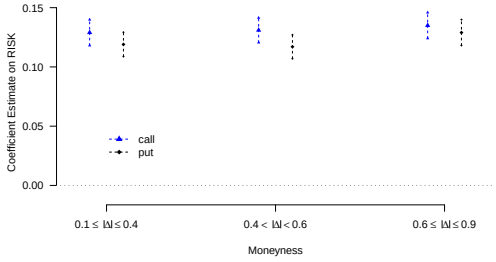
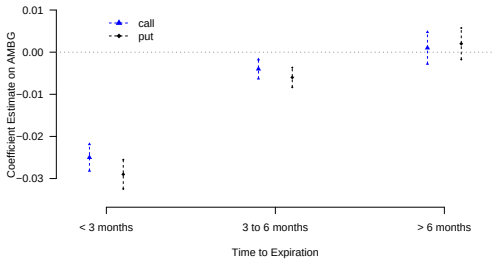


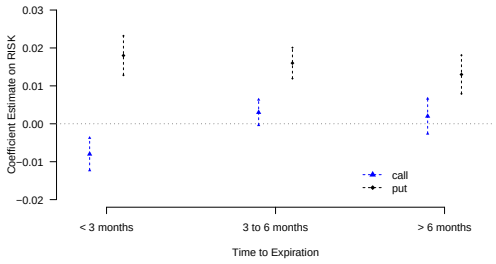
Figure 4: The effect of *AMBG* and *RISK* on options' open interest and trading volume based on maturity

This figure plots the coefficient estimates of *AMBG* and *RISK* from daily panel regressions, in which call and put stock option open interest or trading volume on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), risk (*RISK*), and other firm characteristics based on maturity. In particular, for each firm and day we aggregate options open interest (Graphs A-B) or trading volume (Graphs C-D) based on contract maturity. The maturity groups are defined as *Maturity <= 3 months*, *3 < Maturity <= 6 months*, and *6 < Maturity <= 12 months*, respectively. To estimate the coefficients we stack each firm daily measures in the same regression and interact *AMBG* and *RISK* with dummy variables based on the three defined maturity groups.The regression results are reported in Table B.4. The graphs below plot the regressions' coefficient estimates of open interest (trading volume) from trading day t+5 (t) together with their 95% confidence intervals.

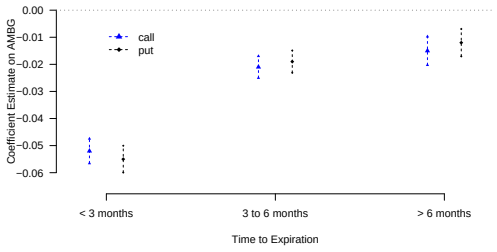
Panel A: *AMBG* and open interest



Panel B: *RISK* and open interest



Panel C: *AMBG* and trading volume



Panel D: *RISK* and trading volume

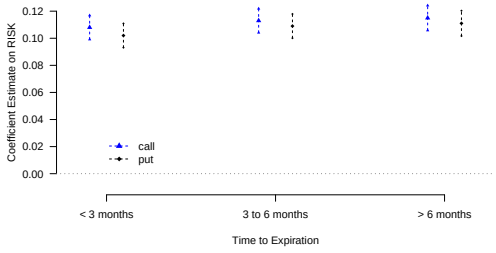
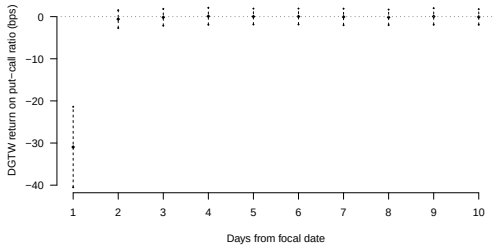


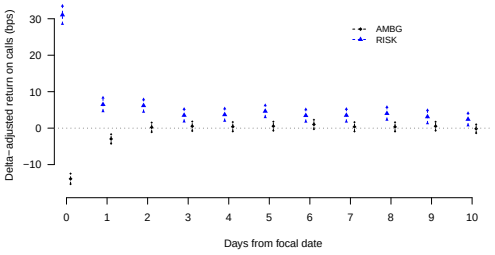
Figure 5: The dynamics of non-cumulative daily stock and option returns

This figure plots coefficient estimates of ΔPC_RATIO based on Equation (5) (Panel A) and the coefficient estimates of $AMBG$ and $RISK$ based on Equation (6) (Panels B and C) using non-cumulative daily returns. Panel A plots results from daily DGTW adjusted stock returns from day t+1 to t+10 together with their 95% confidence intervals. Similarly, Panels B and C plot results from daily delta-hedged option returns from day t to t+10. In all panels the focal date is day t.

Panel A: DGTW adjusted stock returns



Panel B: Call option delta-hedged returns



Panel C: Put option delta-hedged returns

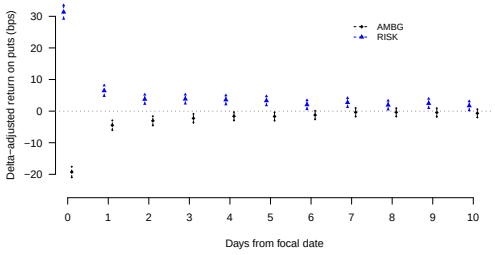


Table 1: Summary statistics

This table reports the summary statistics of the variables employed in the statistical analysis. All variables are defined in Table B.1. All panels reports the sample's mean Std. Dev. and median together with the number of firm-day observations. Panel A reports the statistics of the main stock variables. For ease of presentation, *AMBG* and *RISK* are multiplied by 10,000, *VOV* is multiplied by 1 million, and *VOM*, stock turnover (*SVOL*), and *CumRet* are multiplied by 100. Panel B reports statistics regarding the number of unique call and put option contracts and the trading variables of interest. All variables are trimmed at the top and bottom 0.1% of their sample distribution. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics.

Panel A: Main stock variables

	Mean	Std. Dev.	Median	Obs.
AMBG	60.615	84.185	32.202	6,766,488
RISK	9.897	9.242	6.951	6,766,488
VOV	1.399	2.389	0.520	6,766,488
VOM	1.921	2.000	1.270	6,766,488
DAF	0.068	0.463	0.020	6,766,488
SKEW	-0.002	0.281	-0.002	6,766,488
KURT	4.834	0.941	4.745	6,766,488
Size in Millions	8408.142	26830.731	1899.955	6,766,488
Book-to-Market	0.536	0.515	0.426	6,766,488
Number of Analysts	10.524	6.869	9.000	6,762,750
InstHold	0.692	0.201	0.727	6,412,098
SVOL	1.193	1.669	0.805	6,766,488
ES	0.318	2.117	0.096	6,766,488
$\frac{1}{AvePrc}$	0.047	0.039	0.034	6,766,488
CumRet	1.386	12.946	1.097	6,766,488

Panel B: Main option variables

	Mean	Std. Dev.	Median	Obs.
# Call Options	15.302	20.829	9.000	6,128,675
# Put Options	15.551	21.236	9.000	6,050,752
COI	0.794	1.644	0.290	6,123,752
POI	0.656	1.616	0.194	6,045,825
CVOL	0.050	0.205	0.005	6,124,603
PVOL	0.036	0.169	0.002	6,046,688
CBAS	14.054	10.233	11.275	4,738,569
PBAS	13.020	9.809	10.258	4,081,344
CRET	-0.349	7.686	-0.490	6,112,183
PRET	-0.338	7.480	-0.409	6,032,070

Table 2: Correlations

This table reports the sample correlations between *AMBG* and other variables of interest. The sample period is from January 2002 to December 2018.All variables are defined in Table B.1. Panel A reports the correlation matrix between *AMBG*, *RISK*, and the main variables of interest. Panel B reports the correlation matrix between *AMBG*, *RISK*, and other uncertainty variables. Finally, Panel C reports the partial correlations from daily panel regressions of *AMBG* on other uncertainty proxies. To capture within firm variation the variables in all panels are de-meaned. Consequently, the $AdjR^2$ in Panel C captures the variance explained by the independent variables. Standard errors are double clustered by firm and date, and *t*-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.All variables are defined in Table B.1. All variables are trimmed at the top and bottom 0.1% of their sample distribution. The sample period is from January 2002 to December 2018. The institutional investors' net trading data is from January 2002 to December 2015, taken from ANcerno. The options trading data is taken from OptionMetrics.

Panel A: Main variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) *AMBG*	1.00						
(2) *RISK*	-0.28	1.00					
(3) *COI*	-0.00	-0.02	1.00				
(4) *POI*	-0.03	0.03	0.65	1.00			
(5) *CVOL*	-0.02	0.03	0.40	0.29	1.00		
(6) *PVOL*	-0.03	0.04	0.29	0.36	0.56	1.00	
(7) *SVOL*	-0.05	0.11	0.18	0.18	0.46	0.40	1.00

Panel B: Ambiguity and other uncertainty factors - univariate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) *AMBG*	1.00						
(2) *RISK*	-0.28	1.00					
(3) *VOM*	-0.18	0.71	1.00				
(4) *VOV*	-0.08	0.57	0.40	1.00			
(5) *SKEW*	-0.01	0.01	0.00	0.00	1.00		
(6) *KURT*	0.16	-0.40	-0.29	-0.18	-0.00	1.00	
(7) *DAF*	-0.01	0.02	0.01	-0.00	-0.00	-0.03	1.00

Panel C: Ambiguity and other uncertainty factors - multivariate

	(1) t	(2) t	(3) t	(4) t	(5) t	(6) t
RISK	-2.979*** (-50.22)	-3.677*** (-50.05)	-3.258*** (-60.95)	-3.961*** (-59.34)	-3.735*** (-49.35)	-3.735*** (-49.34)
VOV		4.046*** (43.40)		4.051*** (43.70)	3.919*** (40.17)	3.921*** (40.17)
VOM			1.591*** (14.69)	1.616*** (14.49)	1.658*** (14.66)	1.658*** (14.66)
SKEW					-1.997*** (-9.46)	-1.997*** (-9.46)
KURT					4.093*** (10.94)	4.097*** (10.96)
DAF						0.427 (1.17)
Firm FEs	YES	YES	YES	YES	YES	YES
Firm Cluster	YES	YES	YES	YES	YES	YES
Day Cluster	YES	YES	YES	YES	YES	YES
Observations	6,766,486	6,766,486	6,766,486	6,766,486	6,766,486	6,766,486
AdjR^2	0.077	0.085	0.077	0.086	0.088	0.088

Table 3: Call and put options' open interest

This table reports the findings from daily panel regressions, in which call and put stock option open interest on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics. Call and put open interest measures are reported in Panel A and B, respectively. The regressions with the full set of controls are reported in Table B.5. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Call open interest

	COI(Z)				
	(1)	(2)	(3)	(4)	(5)
	t	t+1	t+2	t+3	t+5
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	-0.004***	-0.003**	-0.001	-0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES
Observations	5,871,968	5,872,005	5,872,150	5,872,179	5,872,223
AdjR^2	0.837	0.837	0.840	0.839	0.839

Panel B: Put open interest

	POI(Z)				
	(1)	(2)	(3)	(4)	(5)
	t	t+1	t+2	t+3	t+5
AMBG(Z)	-0.014***	-0.014***	-0.014***	-0.014***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.015***	0.016***	0.017***	0.017***	0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES
Observations	5,791,506	5,791,552	5,791,681	5,791,760	5,791,788
AdjR^2	0.846	0.846	0.848	0.847	0.845

Table 4: Call and put options' trading volume

This table reports the findings from daily panel regressions, in which stock option trading volume measures on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics. Call and put trading volume measures are reported in Panels A and B, respectively. The regressions with the full set of controls are reported in Table B.6. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. Firm and date fixed effects are included in each specification. (Z) stands for a Z-Score adjustment. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Call trading volume

	CVOL(Z)				
	(1)	(2)	(3)	(4)	(5)
	t	t+1	t+2	t+3	t+5
AMBG(Z)	-0.040***	-0.023***	-0.018***	-0.017***	-0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.137***	0.058***	0.033***	0.026***	0.020***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES
Observations	6,008,137	5,940,699	5,924,982	5,910,826	5,884,918
AdjR^2	0.400	0.409	0.408	0.404	0.395

Panel B: Put trading volume

	PVOL(Z)				
	(1)	(2)	(3)	(4)	(5)
	t	t+1	t+2	t+3	t+5
AMBG(Z)	-0.039***	-0.023***	-0.018***	-0.015***	-0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.132***	0.059***	0.037***	0.029***	0.024***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES
Observations	5,922,273	5,857,357	5,841,742	5,828,234	5,802,097
AdjR^2	0.369	0.373	0.371	0.367	0.359

Table 5: Option based measures and stock return predictability

This table reports the findings from daily panel regressions, in which DGTW adjusted cumulative stock returns from trading day $t+1,\ldots,t+10$ are regressed on trading day t's option based measures, ambiguity ($AMBG$), risk ($RISK$), the interaction of these measures with $AMBG$ and $RISK$ controlling for other firm characteristics. In Panel A we use the changes in put-call open interest ratio (ΔPC_RATIO), where ΔPC_RATIO is calculated as the difference between the open interest of P/(C+P) on day t and t-1. In panel B we use Cremers and Weinbaum (2010)'s implied volatility spread measure (IVS), which captures the difference between call and put implied volatilities for call and put options with the same strike price and maturity. The stock level measure is the open-interest weighted average across all pairs. Columns 1-3, 4-6 and 7-9 report results for cumulative returns based on one, five and ten trading days, respectively. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: The put-call open interest ratio

	$DGTW_t1$			$DGTW_t5$			$DGTW_t10$		
	(1) t+1	(2) t+1	(3) t+1	(4) t+1_t+5	(5) t+1_t+5	(6) t+1_t+5	(7) t+1_t+10	(8) t+1_t+10	(9) t+1_t+10
AMBG(Z)	0.005**	0.005**	0.005**	0.016***	0.016***	0.016***	0.023***	0.022***	0.022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.002	0.002	0.002	0.003	0.003	0.003	-0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\Delta PC\_RATIO(Z)$	-0.309***	-0.304***	-0.304***	-0.352***	-0.346***	-0.345***	-0.365***	-0.358***	-0.357***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\Delta PC\_RATIO(Z)\times AMBG(Z)$		0.034***	0.031***		0.044***	0.036***		0.046***	0.038***
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
$\Delta PC\_RATIO(Z)\times RISK(Z)$			-0.005			-0.014***			-0.015***
			(0.00)			(0.00)			(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	NO	NO	NO	NO	NO	NO	NO	NO	NO
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,822,503	5,822,503	5,822,503	5,820,028	5,820,028	5,820,028	5,817,898	5,817,898	5,817,898
AdjR^2	0.026	0.026	0.026	0.008	0.009	0.009	0.006	0.006	0.006

Panel B: The implied volatility spread measure

	$DGTW_t1$			$DGTW_t5$			$DGTW_t10$		
	(1) t+1	(2) t+1	(3) t+1	(4) t+1 to t+5	(5) t+1 to t+5	(6) t+1 to t+5	(7) t+1 to t+10	(8) t+1 to t+10	(9) t+1 to t+10
AMBG(Z)	0.003	0.003	0.003	0.016***	0.015***	0.015***	0.021**	0.020**	0.020**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.001	0.000	0.002	0.004	0.004	0.005	-0.000	-0.001	0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
IVS(Z)	0.062***	0.059***	0.053***	0.075***	0.068***	0.059***	0.083***	0.073***	0.061***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$IVS(Z)\times AMBG(Z)$		-0.007***	0.000		-0.017***	-0.007		-0.027***	-0.014*
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
$IVS(Z)\times RISK(Z)$			0.012***			0.017***			0.023***
			(0.00)			(0.00)			(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	NO	NO	NO	NO	NO	NO	NO	NO	NO
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,614,965	5,614,965	5,614,965	5,613,858	5,613,858	5,613,858	5,612,232	5,612,232	5,612,232
AdjR^2	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003

Table 6: Call and put options' cumulative delta-hedged returns

This table reports the findings from daily panel regressions, in which stock option cumulative delta-hedged returns on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics. We calculate the options' end of day prices based on the midpoint between the end of day best bid and best ask quotes (OptionPRC t). Based in the prices, the option's daily delta-hedged return is calculated as $[(OptionPRC_t - OptionPRC_{t-1}) - \Delta_{t-1}(StockPRC_t - StockPRC_{t-1})]/OptionPRC_{t-1}$. To aggregate the call or put options at the firm level, we form value-weighted portfolios using day t-1 open interest dollar value as the weight. We fix day t-1 open interest dollar value to allow for a natural buy and hold interpretation. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	CCUMRET(Z)					*PCUMRET(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.138***	-0.174***	-0.182***	-0.184***	-0.185***	-0.194***	-0.251***	-0.292***	-0.321***	-0.360***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
RISK(Z)	0.305***	0.403***	0.492***	0.542***	0.652***	0.313***	0.433***	0.513***	0.577***	0.680***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,099,959	6,005,322	5,935,581	5,877,097	5,776,690	6,020,006	5,927,494	5,859,923	5,804,013	5,708,099
AdjR^2	0.162	0.156	0.163	0.169	0.177	0.106	0.124	0.141	0.156	0.175

Table 7: Call and put options' bid-ask spread

This table reports the findings from daily panel regressions, in which call and put options bid-ask spreads on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics. Call and put measures are reported in Columns 1-5 and Columns 6-10, respectively. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	CBAS(Z)					*PBAS(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	0.001	0.006***	0.007***	0.006***	0.007***	-0.000	0.006***	0.008***	0.007***	0.007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.064***	0.035***	0.033***	0.032***	0.028***	0.032***	0.005***	0.001	0.001	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	4,693,356	4,580,004	4,542,915	4,511,320	4,456,246	4,040,028	3,935,647	3,899,788	3,868,782	3,814,211
AdjR^2	0.574	0.562	0.556	0.552	0.545	0.541	0.531	0.527	0.523	0.516

Table 8: Call and put options around News Event Days

The table extends the analysis conducted in Table 3 based on ... The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI				*POI*			
	EDAY		*8-K*		*EDAY*		*8-K*	
	(1) t	(2) t+5	(3) t	(4) t+5	(5) t	(6) t+5	(7) t	(8) t+5
AMBG(Z)	-0.007**	-0.005	-0.017***	-0.017***	-0.026**	-0.024**	-0.016***	-0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)
RISK(Z)	0.014***	0.020***	-0.009	0.008	0.015***	0.022***	0.015**	0.025***
	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	92,165	92,171	87,818	87,822	90,887	90,897	86,514	86,539
AdjR^2	0.805	0.818	0.811	0.807	0.784	0.793	0.846	0.844

Panel B: Trading volume

	CVOL				*PVOL*			
	EDAY		*8-K*		*EDAY*		*8-K*	
	(1) t	(2) t+1	(3) t	(4) t+1	(5) t	(6) t+1	(7) t	(8) t+1
AMBG(Z)	-0.037***	-0.019***	-0.050***	-0.030***	-0.026***	-0.017***	-0.060***	-0.030***
	(-5.00)	(-3.71)	(-6.14)	(-4.98)	(-3.28)	(-3.22)	(-7.83)	(-6.17)
RISK(Z)	0.318***	0.083***	0.378***	0.167***	0.299***	0.077***	0.372***	0.165***
	(21.81)	(9.39)	(17.01)	(11.29)	(19.91)	(8.12)	(16.27)	(10.15)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	94,331	93,104	89,994	88,946	92,824	91,738	88,617	87,606
AdjR^2	0.533	0.470	0.373	0.397	0.496	0.442	0.343	0.362

A Appendix - Estimating equity ambiguity

The measure of ambiguity, denoted by $\mho^2$ and defined by Equation (1), represents an expected probability-weighted average of the variances of probabilities. We follow the recent literature (e.g., Brenner and Izhakian, 2018; Augustin and Izhakian, 2020; Izhakian et al., 2021) and estimate the monthly degree of ambiguity for each firm's equity using intraday stock return data from TAQ. To estimate ambiguity as implemented in Equation (7) below, the expectation of and the variation in return probabilities across the set of possible prior probability distributions, $\mathcal{P}$, must be measured.

We assume that the intraday equity return distribution for each time interval during the day in a given day represents a single prior distribution, P, in the set of possible distributions, $\mathcal{P}$, and the number of priors in the set is assumed to depend on the number of time intrevals in the month. Each prior (distribution) in the set is represented by the thirty-second observed intraday returns on the firm's equity, in a time interval of 1170 seconds during the trading hours.[21] Thus, the set of priors consists of 20 realized distributions, at most, over a day. For practical implementation reasons, we discretize return distributions into n bins $B_\ell = (r_{\ell-1}, r_\ell]$ of equal size, such that each distribution is represented by a histogram, as demonstrated in Figure B.1. The height of the bar for each bin is the frequency of intraday returns observed in that bin and, thus, represents the probability of the returns in that bin. Equipped with these 20 return histograms, we compute the expected probability in a particular bin across the return distributions, $\mathrm{E}\left[\mathrm{P}\left(B_\ell\right)\right]$, as well as the variance of these probabilities, $\mathrm{Var}\left[\mathrm{P}\left(B_\ell\right)\right]$. To this end, an equal likelihood is assigned to each histogram.[22] We use these equally likely histograms to compute the daily degree of ambiguity of stock j as follows

$$\mho^2\left[r_j\right] \equiv \frac{1}{\sqrt{w\left(1-w\right)}} \sum_{\ell=1}^{n} \mathrm{E}\left[\mathrm{P}_j\left(B_\ell\right)\right] \mathrm{Var}\left[\mathrm{P}_j\left(B_\ell\right)\right]. \tag{7}$$

To minimize the impact of bin size on the scale of ambiguity, we apply a variation of Sheppard's correction and scale the probability weighted-average variance of probabilities to the size of the bins by $\frac{1}{\sqrt{w(1-w)}}$, where $w = r_{\ell-1} - r_\ell$.

[21] Our findings are robust to the use of different time intervals, implying a different number of distributions per day.

[22] Equal weighting is consistent with the principle of insufficient reason, which states that given n possibilities that are indistinguishable except for their names, each possibility should be assigned a probability equal to $\frac{1}{n}$ (Bernoulli, 1713; Laplace, 1814); with the idea of the simplest non-informative prior in Bayesian probability (Bayes et al., 1763), which assigns equal probabilities to all possibilities; and with the principle of maximum entropy (Jaynes, 1957), which states that the probability distribution which best describes the current state of knowledge is the one with the largest entropy.

[Figure B.1]

In our implementation, we sample thirty-second stock returns from 9:30 to 16:00. Thus, we obtain intradaily histograms of up to 39 intraday returns. If we observe no trade in a specific time interval, we compute returns based on the volume-weighted average of the nearest trading prices within 15 seconds distance from that time point. If there is no trade price within this distance, we drop this 30 second observation. We ignore returns between closing and next-day opening prices to eliminate the impact of overnight price changes and dividend distributions. We drop all time intervals with fewer than 10 thirty-second returns, and then we drop days with fewer than 10 intraday return distributions.[23] In addition, we drop extreme returns ($\pm 5\%$ log returns over thirty seconds), as many such returns are due to improper orders that are often later canceled by the stock exchange. We normalize the intraday thirty-second rates of return to daily returns.[24]

For the bin formation, we divide the range of normalized returns into 1,002 intervals. We form a grid of 1,000 bins, from -100% to $+100\%$, each of width 0.2%, in addition to the left and right tails, defined as $(-\infty, -100\%]$ and $[+100\%, +\infty)$, respectively. We compute the mean and the variance of probabilities for each bin, assigning an equal likelihood to each distribution (i.e., all histograms are equally likely).[25] Some bins may not be populated with return realizations. Therefore, we assume a normal return distribution and use its moments to extrapolate return probabilities. That is, $\mathrm{P}_j\left(B_\ell\right) = \Phi\left(r_\ell; \mu_j, \sigma_j\right) - \Phi\left(r_{\ell-1}; \mu_j, \sigma_j\right)$, where $\Phi\left(\cdot\right)$ denotes the cumulative normal probability distribution, characterized by its mean μ_j and variance σ_j^2 of returns.[26]

An important characteristic of the measure of ambiguity implied by EUUP is that it is outcome independent (up to a state space partition), which allows for a risk-independent examination of the impacts of ambiguity on financial decisions. Specifically, the measure of ambiguity $\mho^2$ captures the

[23] For robustness, we run all the regression tests excluding all time intervals with fewer than 15 thirty-second returns and all days with fewer than 15 intraday return distributions. The findings are essentially the same.

[24] Our findings are robust to the inclusion of extreme price changes, as well as to a cutoff at a level of 1% in terms of log returns.

[25] The assignment of equal likelihoods is equivalent to assuming that the daily ratios $\frac{\mu}{\sigma}$ are Student-t distributed. When $\frac{\mu}{\sigma}$ is Student-t distributed, cumulative probabilities are uniformly distributed (Kendall and Stuart, 2010, Proposition 1.27, p. 21).

[26] As in French et al. (1987), Brenner and Izhakian (2018) and Augustin and Izhakian (2020) apply the Scholes and Williams (1977) adjustment for non-synchronous trading to estimate the variance of returns. Scholes and Williams (1977) suggest adjusting the volatility of returns for non-synchronous trading as $\sigma_t^2 = \frac{1}{N_t}\sum_{\ell=1}^{N_t}\left(r_{t,\ell} - \mathrm{E}\left[r_{t,\ell}\right]\right)^2 + 2\frac{1}{N_t - 1}\sum_{\ell=2}^{N_t}\left(r_{t,\ell} - \mathrm{E}\left[r_{t,\ell}\right]\right)\left(r_{t,\ell-1} - \mathrm{E}\left[r_{t,\ell-1}\right]\right)$. This adjustment mitigates microstructure effects caused by bid-ask bounce. For robustness, we run all regression tests in which ambiguity is computed using this adjusted volatility of returns. The findings are essentially the same.

variation in the frequencies (probabilities) of the outcomes, without incorporating the magnitudes of the outcomes. In contrast, the measure of risk captures the variation in the magnitudes of the outcomes without incorporating the variation in the frequencies with which the outcomes are observed. Thus, the measure of ambiguity is risk independent, just as standard measures of risk are ambiguity independent, implying that these two measures capture distinct aspects of uncertainty.

Other proxies for ambiguity in the literature include the volatility of mean returns (Franzoni, 2017), the volatility of volatility of returns (Faria and Correia-da Silva, 2014), or the disagreement of analysts' forecasts (Anderson et al., 2009). These measures are sensitive to changes in the set of outcomes (i.e., are outcome dependent), so they are risk dependent and, therefore, less useful for this study. For similar reasons, skewness and kurtosis (as well as other higher moments of the return distribution) are also different from $\mho^2$, as the former are outcome dependent and the latter is outcome independent. Time-varying mean, time-varying volatility, and jumps (return shocks) are outcome dependent as well.

Figure B.1 also demonstrates that ambiguity is independent of outcomes and, therefore, independent of risk. Consider, for example, an extreme return (i.e., a stock price jump or a shock). If the partition of the state space remains unchanged, one of the bins will be associated with a higher return, but the probability of that particular bin, or any other bin, remains unchanged. Therefore, ambiguity remains unchanged.[27] If, on the other hand, the partition of the state space changes, then one additional bin may be added to the histogram, thereby characterizing a new event. This new bin may also affect the population of other bins, and therefore, affect ambiguity. However, both the expected probability of experiencing a return in this new bin and the probability variance associated with it, are small. Thus, such an extreme return would have a negligible impact on ambiguity, since the effect on ambiguity is by the product of the expected probability and the variance of probability, which is even smaller.

Brenner and Izhakian (2018) study the implications of the aggregate market ambiguity and suggest that, in their sample, $\mho^2$ does not capture other well-known uncertainty factors including skewness, kurtosis, the volatility-of-mean, the volatility-of-volatility, volatility jumps, unexpected volatility, downside risk, mixed data sampling measure of volatility forecasts (MIDAS), investor sentiment, and several others. Augustin and Izhakian (2020) study the implications of firm am-

[27] To illustrate, consider a rate of return on an investment that is determined by a coin toss with unknown probabilities, where heads yields a 2% return and tails a 1% return. Even if after 10 coin tosses the rate of return for heads changes to 10% (i.e., a jump), ambiguity remains unchanged, since no new information about probabilities has been obtained.

biguity for the spread of credit default swaps and suggest that, in their sample, $\mho^2$ also does not capture these factors at the firm level.[28] To further mitigate the concerns that $\mho^2$ captures other well-known uncertainty factors or market-microstructure effects, in Section 5.4, we examine the explanatory power of $\mho^2$ relative to these uncertainty factors at the daily firm level.

[28]In a battery of robustness tests, Augustin and Izhakian (2020) also mitigate concerns that the measure of ambiguity $\mho^2$ is sensitive to the selection of the time interval of intraday rate of returns, the bin size, and the type of parametric probability distribution used to extrapolate bins' probabilities.

B Appendix - Variable definitions and additional tests

Figure B.1: Ambiguity measurement

This figure illustrates the way we compute the ambiguity measure for each firm-day, based on intraday stock returns, sampled at thirty-second intervals from 9:30 to 16:00. For each firm-day, these samples create 20 intraday histograms of up to 39 intraday returns. For each intraday histogram, we discretize the time-period return distribution into n bins of equal size $B_\ell = (r_{\ell-1}, r_\ell]$. The height of each intraday histogram bin is the fraction of intraday returns observed in that bin, representing the probability of that bin's outcome. For simplicity, this figure shows three histograms with six bins. Across the intraday return distributions, we compute the expected probability of returns in a bin as $\mathrm{E}\left[\mathrm{P}_j\left(B_\ell\right)\right]$ and the variance of probabilities as $\mathrm{Var}\left[\mathrm{P}_j\left(B_\ell\right)\right]$. Finally, we compute firm-day ambiguity as $\mho^2\left[r_j\right] \equiv 1/\sqrt{w\left(1-w\right)}\sum_{\ell=1}^{n}\mathrm{E}\left[\mathrm{P}_j\left(B_\ell\right)\right]\mathrm{Var}\left[\mathrm{P}_j\left(B_\ell\right)\right]$, where we scale the weighted-average variance of probabilities by the bin size $w = r_\ell - r_{\ell-1}$.

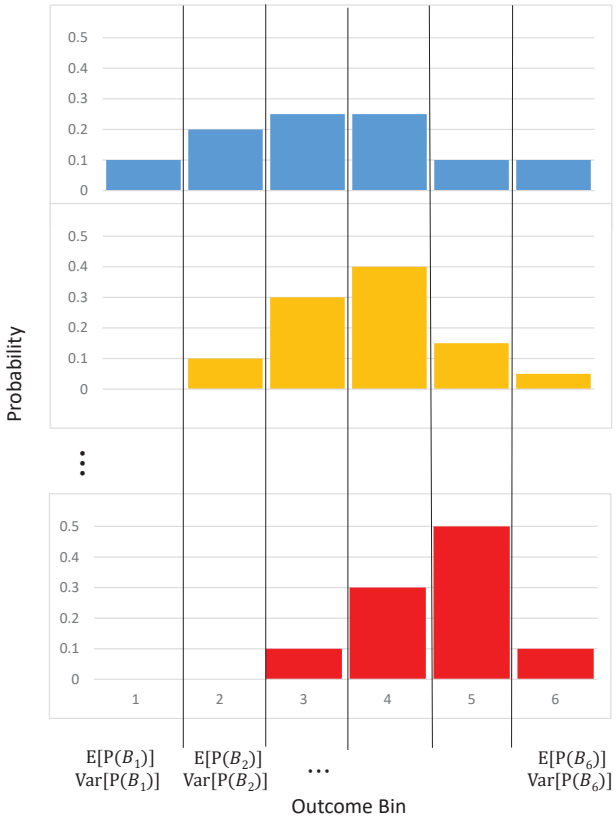


Figure B.2: Impulse Response Functions excluding day-0 effect

This figure plots the impulse responses of the trading and liquidity measures to a one-standard-deviation shock to *AMBG* and *RISK*. For each dependent variable (*DEP*), it estimates a daily vector autoregression (VAR) system of *DEP*, *AMBG*, and *RISK*, with five lags of each of the variables. All variables are defined in Table B.1, where *AMBG*, *RISK*, and *DEP* are trimmed at the top and bottom 0.1% of their sample distribution. All regression tests include the full set of firm control variables together with firm fixed effects and date fixed effects. The VAR system takes the following form

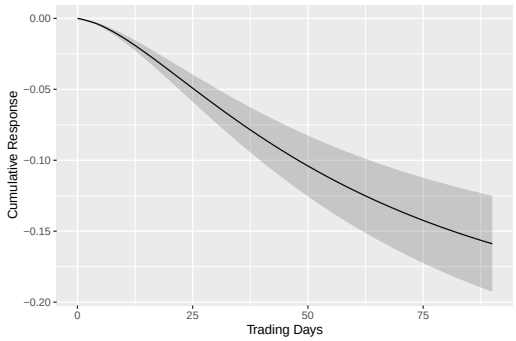
$$DEP_{j,t} = \alpha_1 + \sum_{i=1}^{5} \beta_{1,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{1,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{1,i} \cdot DEP_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{1,j,t};$$

$$AMBG_{j,t} = \alpha_2 + \sum_{i=1}^{5} \beta_{2,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{2,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{2,i} \cdot DEP_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{2,j,t};$$

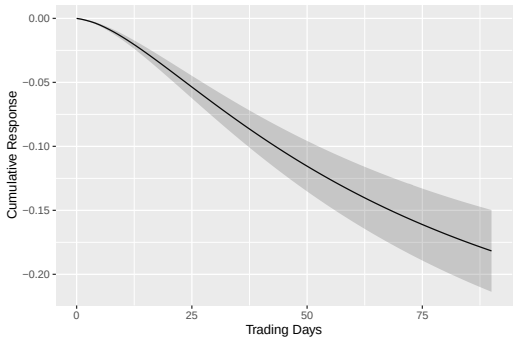
$$RISK_{j,t} = \alpha_3 + \sum_{i=1}^{5} \beta_{3,i} \cdot AMBG_{j,t-i} + \sum_{i=1}^{5} \gamma_{3,i} \cdot RISK_{j,t-i} + \sum_{i=1}^{5} \delta_{3,i} \cdot DEP_{j,t-i} + \Gamma \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{3,j,t}.$$

The estimated coefficients of this system are reported in Table B.2. This figure includes three groups of graphs: open interest (Graphs A-D), trading volume (Graphs E-H) and delta-hedged returns (Graphs I-L). Each group plots the cumulative response of *DEP* to a one-standard-deviation shock to *AMBG* or *RISK*. To estimate the effect of *AMBG* (*RISK*) on *DEP*, the Cholesky order is set zero. That is, day t effect is not allowed to enter the system updating process. Each graph depicts the response in the subsequent $0, \ldots, 90$ trading days, listed on the x-axis. The solid line depicts the variable response and the dashed gray lines depict the 95% confidence intervals.

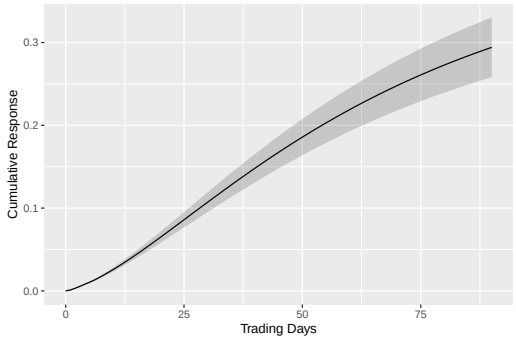
Panel A: Response of call option open interest to firm ambiguity



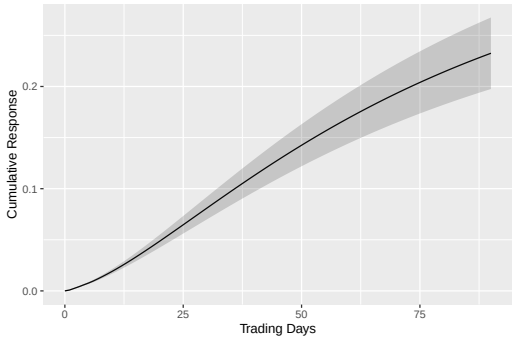
Panel B: Response of put option open interest to firm ambiguity



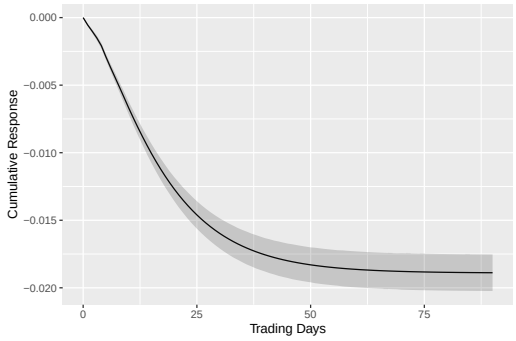
Panel C: Response of call option open interest to firm risk



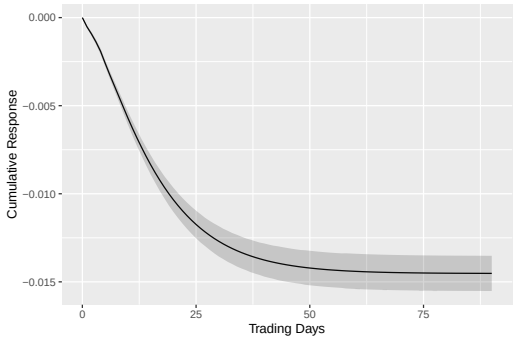
Panel D: Response of put option open interest to firm risk



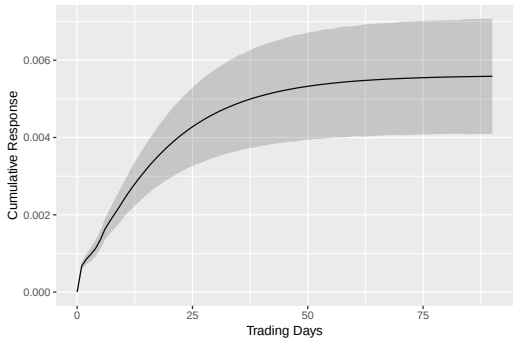
Panel E: Response of call option volume to firm ambiguity



Panel F: Response of put option volume to firm ambiguity



Panel G: Response of call option volume to firm risk



Panel H: Response of put option volume to firm risk

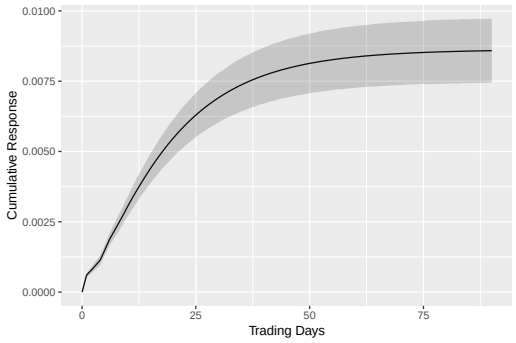


Table B.1: Variable definitions

Variable	Definition
Ambiguity and Other Moments	
$AMBG$	The daily ambiguity, measured as detailed in Section 2.1. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
$MktAMBG$	$AMBG$ of the S&P500 index (SPY ticker).
$\Delta MktAMBG$	Daily changes in $MktAMBG$, calculated as $MktAMBG_t - MktAMBG_{t-1}$.
$RISK$	The daily risk, measured as detailed in Section 2.2. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
VIX	The CBOE volatility index, calculated based on the implied volatility of the S&P500 options.
ΔVIX	Daily changes in VIX, calculated as $VIX_t - VIX_{t-1}$.
VOM	Daily volatility-of-mean, calculated as the variance of the averages' return over 20 intraday time intervals, where each interval's average is computed using 30-second returns. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
VOV	Daily volatility-of-volatility, calculated as the variance of the variances of return over 20 intraday time intervals, where each interval's variance is computed using 30-second returns. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
$SKEW$	Daily realized skewness, computed using 30-second intraday returns. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
$KURT$	Daily realized kurtosis, calculated using 30-second intraday returns. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
$AvgAMBG$	The 21 trading day trailing average of $AMBG$ over trading days $t-27,\ldots,t-6$.
$AvgRISK$	The 21 trading day trailing average of $RISK$ over trading days $t-27,\ldots,t-6$.
$AvgVOM$	The 21 trading day trailing average of VOM over trading days $t-27,\ldots,t-6$.
$AvgVOV$	The 21 trading day trailing average of VOV over trading days $t-27,\ldots,t-6$.
$AvgSKEW$	The 21 trading day trailing average of $SKEW$ over trading days $t-27,\ldots,t-6$.
$AvgKURT$	The 21 trading day trailing average of $KURT$ over trading days $t-27,\ldots,t-6$.
Option Variables	
Filters	The options data is obtained from OptionMetrics. To reduce noise due to contract expiration or unusual maturities, only call and put options with maturities of 7 to 365 days are considered. In addition, we follow Muravyev(2016), Christoffersen et al. (2018) and Muravyev and Ni (2020) and apply the following additional filters: we keep option contrasts with absolute deltas between 0.1 to 0.9, keep contracts with positive open interest, keep contracts with valid bid-ask spread information, drop contracts where the spread to midpoint ratio is greater than 70%, drop contracts with bid-ask spread above \$3, and drop contracts with midpoints below \$0.10 cents.
COI	The daily sum of the open interest of call options written on the stock, divided by the stock outstanding shares. We account for the fact that open interest is lagged by one day after November 28^{th}, 2000. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
POI	The daily sum of the open interest of put options written on the stock, divided by the stock outstanding shares. We account for the fact that open interest is lagged by one day after November 28^{th}, 2000. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.

Variable	Definition
Option Variables (Cont.)	
CVOL	The daily sum of trading volume of call options written on the stock, divided by the stock's number of shares outstanding. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
PVOL	The daily sum of trading volume of put options written on the stock, divided by the stock's number of shares outstanding. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
CCUMRET	the delta-hedged cumulative return of call options written on the stock. Call options' end of day prices based on the midpoint between the end of day best bid and best ask quotes (PRC t). Based in the prices, the option's daily delta-hedged return is calculated as $[(PRC_t - PRC_{t-1}) - \delta_{t-1}(PRC_t - PRC_{t-1})]/PRC_{t-1}$. To aggregate the call or put options at the firm level, we form value-weighted portfolios using day t-1 open interest dollar value as the weight. We fix day t-1 open interest dollar value to allow for a natural buy and hold interpretation. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
PCUMRET	the cumulative delta-hedged return of put options written on the stock. The calculation is similar to *CCUMRET* calculation.
CBAS	The daily average bid-ask spread of call options written on the stock, calculated as the difference between the best offer and the best ask divided by their midpoint. We take the value-weighted average across all options for a given stock, using the daily options' dollar volume as the weight. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
PBAS	The daily average bid-ask spread of put options written on the stock, calculated as the difference between the best offer and the best ask divided by their midpoint. We take the value-weighted average across all options for a given stock, using the daily options' dollar volume as the weight. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
AvgCOI	The 21 trading day trailing average of *COI* over trading days $t-27, \ldots, t-6$.
AvgPOI	The 21 trading day trailing average of *POI* over trading days $t-27, \ldots, t-6$.
AvgCVOL	The 21 trading day trailing average of *CVOL* over trading days $t-27, \ldots, t-6$.
AvgPVOL	*PVOL* The 21 trading day trailing average of *PVOL* over trading days $t-27, \ldots, t-6$.
AvgCBAS	The 21 trading day trailing average of *CBAS* over trading days $t-27, \ldots, t-6$.
AvgPBAS	The 21 trading day trailing average of *PBAS* over trading days $t-27, \ldots, t-6$.
Other Stock Variables	
SVOL	Daily stock volume, calculated as the number of daily traded shares divided by the number of shares outstanding. To reduce the effect of outliers, the top and bottom 0.1% of the sample distribution are trimmed.
LnSize	The natural logarithm of the firm's size in million dollars, following Fama and French (1992).
LnBM	The natural logarithm of the firm's book-to-market ratio, rebalanced every June, following Fama and French (1992).
InstHold	The firm's fraction of institutional holdings taken from Thomson Reuters Institutional (13F) Holdings database.
RET	The daily stock return, as reported by CRSP.
CumRet	The stock's cumulative return over the 21 trading days $t-27, \ldots, t-6$.
LnNumEst	The natural logarithm of one plus *NumEst*, where *NumEst* is the number of analysts covering the firm according to the most recent information from I/B/E/S.
$\ln \frac{1}{AvePrc}$	The natural logarithm of one over the average stock price (*AvePrc*), adjusted for splits, where *AvePrc* is calculated over trading days $t-27, \ldots, t-6$.

Table B.2: Call and put options' variables in a VAR setting

This table reports the findings from daily panel regressions, which serve as the base of our VAR analysis. In particular, our options' and stock measures are regressed on five lags of ambiguity ($AMBG$), risk ($RISK$), and the dependent variable (DEP). All variables are defined in Table B.1. All variables are trimmed at the top and bottom 0.1% of their sample distribution. All regression tests include the full set of firm control variables together with firm fixed effects and date fixed effects. (Z) stands for a Z-Score adjustment. The regression specifications take the following form

$$DEP(Z)_{j,t} = \alpha + \sum_{i=1}^{5} \beta_i \cdot AMBG(Z)_{j,t-i} + \sum_{i=1}^{5} \gamma_i \cdot RISK(Z)_{j,t-i} + \sum_{i=1}^{5} \delta_i \cdot DEP(Z)_{j,t-i} + \delta \cdot CONTROLS_{j,t} + \eta_j + \theta_t + \epsilon_{1,j,t}. \quad (8)$$

The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	(1) $COI(Z)$	(2) $POI(Z)$	(3) $CVOL(Z)$	(4) $PVOL(Z)$	(5) $CRET(Z)$	(6) $PRET(Z)$	(7) $CBAS(Z)$	(8) $PBAS(Z)$
$AMBG(Z)\ t-1$	-0.001***	-0.001***	-0.006***	-0.008***	-0.011***	0.001	0.003***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z)\ t-2$	-0.000**	-0.000*	-0.002***	-0.002***	0.002**	-0.002*	0.002***	0.003***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z)\ t-3$	0.000	-0.000	-0.001**	-0.002***	0.005***	-0.002*	0.001	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z)\ t-4$	-0.000	-0.000	-0.001**	-0.002***	0.004***	-0.000	-0.000	0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z)\ t-5$	-0.000	-0.000	-0.003***	-0.002***	0.004***	-0.001	0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z)\ t-1$	0.002***	0.002***	0.013***	0.015***	0.018***	-0.003*	0.011***	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z)\ t-2$	0.001***	0.001***	-0.004***	-0.004***	0.004***	-0.000	0.001	-0.003**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z)\ t-3$	0.000	-0.000	-0.002*	-0.001	-0.001	0.002	-0.000	-0.003*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z)\ t-4$	0.000	-0.000	-0.001	-0.001	-0.004**	0.004**	-0.004***	-0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z)\ t-5$	-0.001***	-0.001***	0.000	0.003***	-0.002	0.003*	0.000	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

	(1) *COI(Z)*	(2) *POI(Z)*	(3) *CVOL(Z)*	(4) *PVOL(Z)*	(5) *CRET(Z)*	(6) *PRET(Z)*	(7) *CBAS(Z)*	(8) *PBAS(Z)*
$DEP(Z)\ t-1$	0.866***	0.871***	0.292***	0.274***	-0.202***	0.143***	0.217***	0.212***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$DEP(Z)\ t-2$	-0.035***	-0.044***	0.121***	0.119***	-0.043***	0.044***	0.157***	0.154***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$DEP(Z)\ t-3$	0.095***	0.104***	0.086***	0.084***	-0.006***	0.016***	0.131***	0.131***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$DEP(Z)\ t-4$	0.013***	0.008**	0.074***	0.074***	0.004***	0.009***	0.115***	0.115***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$DEP(Z)\ t-5$	0.038***	0.041***	0.085***	0.085***	0.004***	0.004***	0.113***	0.113***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,778,604	5,704,053	5,864,429	5,777,571	5,823,013	5,725,106	3,439,335	2,694,917
AdjR^2	0.968	0.971	0.428	0.393	0.208	0.127	0.602	0.577

Table B.3: Call and put options' open interest and trading volume based on moneyness

The table extends the analysis conducted in Table 3 and Table 4, where firm's options open interest and trading volume are aggregated on each day based on contract moneyness. The moneyness groups $DR1$, $DR2$ and $DR3$ are defined as *0.1* <= |Δ| <= *0.40*, *0.40* < |Δ| < *0.60*, and *0.60* < |Δ| <= *0.90*, respectively. To estimate the coefficients we stack each firm daily measures in the same regression and interact *AMBG* and *RISK* with dummy variables based on the three defined moneyness groups (*AMBG_DR1- AMBG_DR3* and *RISK_DR1- RISK_DR3*). The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable (*AvgDEP*), *AMBG*(*AvgAMBG*) and *RISK*(*AvgRISK*). This allows to account for the persistence in the dependent variables, and explore the effect of changes in *AMBG* and *RISK* relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and *t*-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG_DR1(Z)	-0.012***	-0.013***	-0.014***	-0.014***	-0.014***	-0.012***	-0.013***	-0.015***	-0.016***	-0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_DR2(Z)	-0.009***	-0.010***	-0.010***	-0.011***	-0.012***	-0.010***	-0.011***	-0.011***	-0.011***	-0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_DR3(Z)	-0.001	-0.001	-0.000	-0.000	-0.000	-0.008***	-0.008***	-0.007***	-0.007***	-0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR1(Z)	0.002	0.003	0.004*	0.004*	0.003	-0.006***	-0.004*	-0.002	-0.002	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR2(Z)	-0.004***	-0.002	-0.001	-0.000	0.000	0.017***	0.017***	0.016***	0.016***	0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR3(Z)	-0.009***	-0.007***	-0.006***	-0.005***	-0.003*	0.033***	0.033***	0.035***	0.036***	0.036***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p-Val Diff	<0.001	<0.001	<0.001	<0.001	<0.001	0.061	0.005	<0.001	<0.001	<0.001
Observations	14,287,737	14,287,583	14,287,788	14,287,842	14,287,902	14,105,298	14,105,262	14,105,442	14,105,576	14,105,683
AdjR^2	0.644	0.646	0.651	0.652	0.654	0.671	0.672	0.678	0.678	0.679

Panel B: Trading volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG_DR1(Z)	-0.039***	-0.019***	-0.013***	-0.011***	-0.009***	-0.042***	-0.027***	-0.022***	-0.020***	-0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_DR2(Z)	-0.043***	-0.026***	-0.020***	-0.017***	-0.014***	-0.036***	-0.021***	-0.015***	-0.012***	-0.010***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_DR3(Z)	-0.025***	-0.012***	-0.009***	-0.009***	-0.009***	-0.020***	-0.007***	-0.003*	-0.001	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR1(Z)	0.129***	0.047***	0.022***	0.016***	0.008***	0.119***	0.047***	0.024***	0.016***	0.012***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR2(Z)	0.131***	0.053***	0.028***	0.022***	0.016***	0.117***	0.049***	0.029***	0.023***	0.019***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_DR3(Z)	0.135***	0.063***	0.040***	0.033***	0.026***	0.129***	0.064***	0.044***	0.037***	0.031***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p-Val Diff	<0.001	<0.001	0.029	0.338	0.900	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	14,873,117	14,572,341	14,458,208	14,355,486	14,164,250	14,595,725	14,346,982	14,262,558	14,184,373	14,029,350
AdjR^2	0.288	0.284	0.273	0.262	0.243	0.270	0.266	0.256	0.247	0.233

Table B.4: Call and put options' open interest and trading volume based on maturity

The table extends the analysis conducted in Table 3 and Table 4, where firm's options open interest and trading volume are aggregated on each day based on contract maturity. The maturity groups $MR1$, $MR2$ and $MR3$ are defined as *Maturity <= 3 months*, *3 < Maturity <= 6 months*, and *6 < Maturity <= 12 months*, respectively. To estimate the coefficients we stack each firm daily measures in the same regression and interact *AMBG* and *RISK* with dummy variables based on the three defined maturity groups (*AMBG_MR1- AMBG_MR3* and *RISK_MR1-RISK_MR3*). The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable (*AvgDEP*), *AMBG*(*AvgAMBG*) and *RISK*(*AvgRISK*). This allows to account for the persistence in the dependent variables, and explore the effect of changes in *AMBG* and *RISK* relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and *t*-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG_MR1(Z)	-0.019***	-0.020***	-0.022***	-0.023***	-0.025***	-0.025***	-0.026***	-0.027***	-0.028***	-0.029***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_MR2(Z)	-0.005***	-0.005***	-0.004***	-0.004***	-0.004***	-0.007***	-0.007***	-0.007***	-0.007***	-0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_MR3(Z)	-0.001	-0.001	-0.000	0.000	0.001	0.001	0.001	0.001	0.001	0.002
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR1(Z)	-0.010***	-0.009***	-0.008***	-0.008***	-0.008***	0.021***	0.020***	0.020***	0.020***	0.018***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR2(Z)	0.000	0.001	0.002	0.002	0.003*	0.013***	0.014***	0.015***	0.015***	0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR3(Z)	-0.005**	-0.003	-0.001	0.000	0.002	0.006**	0.008***	0.010***	0.011***	0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p-Val Diff	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	13,954,191	13,953,941	13,953,829	13,953,749	13,953,506	13,525,611	13,525,497	13,525,543	13,525,476	13,525,283
AdjR^2	0.629	0.632	0.636	0.638	0.640	0.633	0.635	0.640	0.642	0.642

Panel B: Trading volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG_MR1(Z)	-0.052***	-0.035***	-0.029***	-0.027***	-0.026***	-0.055***	-0.036***	-0.031***	-0.027***	-0.025***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_MR2(Z)	-0.021***	-0.010***	-0.007***	-0.006***	-0.006***	-0.019***	-0.009***	-0.006***	-0.005***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG_MR3(Z)	-0.015***	-0.004*	-0.000	0.001	0.001	-0.012***	-0.002	0.002	0.003	0.004*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR1(Z)	0.108***	0.035***	0.011***	0.005**	-0.001	0.102***	0.035***	0.013***	0.006**	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR2(Z)	0.113***	0.054***	0.035***	0.029***	0.024***	0.109***	0.057***	0.040***	0.034***	0.029***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK_MR3(Z)	0.115***	0.059***	0.042***	0.037***	0.032***	0.111***	0.061***	0.044***	0.038***	0.035***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
p-Val Diff	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
Observations	14,477,438	14,198,711	14,106,537	14,019,723	13,846,917	14,010,078	13,749,932	13,664,803	13,584,760	13,423,421
AdjR^2	0.333	0.334	0.331	0.325	0.314	0.298	0.297	0.293	0.288	0.279

Table B.5: Call and put options' open interest - reporting the full set of controls

This table reports the full set of results from Table 3. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	-0.004***	-0.003**	-0.001	-0.001	-0.000	0.015***	0.016***	0.017***	0.017***	0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LnSize	0.003	0.003	0.003	0.003	0.003	-0.016***	-0.017***	-0.018***	-0.019***	-0.021***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LnBM	-0.016***	-0.017***	-0.018***	-0.018***	-0.020***	-0.005*	-0.005**	-0.005**	-0.006**	-0.007**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
CumRet	0.004***	0.004***	0.004***	0.003***	0.003***	-0.001***	-0.001***	-0.001***	-0.001***	-0.000**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
LnNumEst	0.013***	0.014***	0.015***	0.015***	0.016***	0.022***	0.022***	0.023***	0.023***	0.023***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
InstHold	0.019**	0.020**	0.021**	0.021**	0.022**	0.002	0.003	0.004	0.004	0.005
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$\ln \frac{1}{AvePrc}$	-0.027***	-0.031***	-0.035***	-0.039***	-0.046***	-0.076***	-0.082***	-0.087***	-0.092***	-0.101***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
RET	0.015***	0.014***	0.014***	0.014***	0.013***	-0.009***	-0.009***	-0.008***	-0.008***	-0.007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AvgDEP	0.653***	0.650***	0.649***	0.645***	0.639***	0.722***	0.718***	0.716***	0.713***	0.705***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
AvgAMBG	0.136	0.059	-0.016	-0.098	-0.229	-0.210	-0.240	-0.318	-0.362	-0.498*
	(0.22)	(0.21)	(0.22)	(0.22)	(0.22)	(0.27)	(0.27)	(0.27)	(0.28)	(0.27)
AvgRISK	23.001***	23.097***	23.872***	24.950***	26.565***	4.632	4.130	4.446	4.726	5.610
	(2.67)	(2.71)	(2.75)	(2.80)	(2.89)	(3.59)	(3.59)	(3.62)	(3.62)	(3.60)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,871,968	5,872,005	5,872,150	5,872,179	5,872,223	5,791,506	5,791,552	5,791,681	5,791,760	5,791,788
AdjR^2	0.837	0.837	0.840	0.839	0.839	0.846	0.846	0.848	0.847	0.845

Table B.6: Call and put options' trading volume - reporting the full set of controls

This table reports the full set of results from Table 4. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.040***	-0.023***	-0.018***	-0.017***	-0.016***	-0.039***	-0.023***	-0.018***	-0.015***	-0.013***
	(-16.88)	(-15.68)	(-14.93)	(-14.14)	(-13.39)	(-15.32)	(-13.82)	(-13.55)	(-13.27)	(-12.03)
RISK(Z)	0.137***	0.058***	0.033***	0.026***	0.020***	0.132***	0.059***	0.037***	0.029***	0.024***
	(24.92)	(17.96)	(13.30)	(11.65)	(9.61)	(23.94)	(17.49)	(14.03)	(12.16)	(10.59)
LnSize	-0.006	-0.009	-0.011	-0.011	-0.011	-0.000	-0.003	-0.004	-0.005	-0.005
	(-0.87)	(-1.29)	(-1.52)	(-1.59)	(-1.45)	(-0.06)	(-0.42)	(-0.53)	(-0.68)	(-0.65)
LnBM	-0.021***	-0.022***	-0.022***	-0.022***	-0.023***	-0.016***	-0.016***	-0.017***	-0.017***	-0.017***
	(-4.82)	(-4.72)	(-4.61)	(-4.58)	(-4.59)	(-3.48)	(-3.40)	(-3.34)	(-3.31)	(-3.40)
CumRet	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(8.29)	(6.29)	(5.06)	(4.15)	(3.66)	(11.59)	(9.67)	(8.60)	(8.85)	(7.99)
LnNumEst	0.018***	0.014*	0.013*	0.013	0.010	0.032***	0.032***	0.032***	0.032***	0.032***
	(2.58)	(1.95)	(1.75)	(1.62)	(1.27)	(4.21)	(3.93)	(3.89)	(3.81)	(3.63)
InstHold	0.017	0.018	0.015	0.015	0.017	0.015	0.013	0.013	0.014	0.016
	(1.47)	(1.42)	(1.17)	(1.19)	(1.25)	(1.21)	(1.04)	(1.02)	(1.07)	(1.17)
$\ln \frac{1}{AvePrc}$	-0.134***	-0.151***	-0.157***	-0.163***	-0.169***	-0.143***	-0.149***	-0.150***	-0.152***	-0.154***
	(-13.43)	(-13.99)	(-14.19)	(-14.32)	(-14.33)	(-13.44)	(-13.40)	(-13.22)	(-13.19)	(-13.05)
RET	0.029***	0.012***	0.008***	0.006***	0.005***	-0.014***	-0.004***	-0.002***	-0.001***	-0.000*
	(35.05)	(24.99)	(22.10)	(19.13)	(16.97)	(-24.06)	(-10.83)	(-7.93)	(-4.01)	(-1.86)
AvgDEP	3.450***	3.463***	3.436***	3.379***	3.271***	4.082***	4.055***	4.013***	3.944***	3.810***
	(24.98)	(23.11)	(22.85)	(21.92)	(20.16)	(25.66)	(24.52)	(23.52)	(22.67)	(21.27)
AvgAMBG	1.068***	-0.733**	-1.293***	-1.566***	-1.735***	0.384	-1.432***	-1.971***	-2.275***	-2.519***
	(2.98)	(-2.27)	(-4.01)	(-4.82)	(-4.88)	(0.95)	(-3.86)	(-5.36)	(-6.07)	(-6.44)
AvgRISK	-100.412***	-25.524***	-2.941	3.766	10.294**	-94.148***	-26.064***	-5.856	1.609	6.391
	(-17.70)	(-5.61)	(-0.68)	(0.87)	(2.36)	(-16.98)	(-5.80)	(-1.33)	(0.36)	(1.44)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,008,137	5,940,699	5,924,982	5,910,826	5,884,918	5,922,273	5,857,357	5,841,742	5,828,234	5,802,097
AdjR^2	0.400	0.409	0.408	0.404	0.395	0.369	0.373	0.371	0.367	0.359

Table B.7: Trading volume based Put-call ratio and stock return predictability

This table reports the findings from daily panel regressions, in which DGTW adjusted cumulative stock returns from trading day $t+1,\ldots,t+10$ are regressed on trading day t's put-call volume ratio (*PCVOL_RATIO*), ambiguity (*AMBG*), risk (*RISK*), the interaction of *PCVOL_RATIO* with *AMBG* and *RISK* controlling for other firm characteristics. In particular,*PCVOL_RATIO* is calculated as day t's aggregate put option trading volume divided by the aggregate trading volume of both call and put options (P/(C+P)). Columns 1-3, 4-6 and 7-9 report results for cumulative returns based on one, five and ten trading days, respectively. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable (*AvgDEP*), *AMBG*(*AvgAMBG*) and *RISK*(*AvgRISK*). This allows to account for the persistence in the dependent variables, and explore the effect of changes in *AMBG* and *RISK* relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	$DGTW_t1$			$DGTW_t5$			$DGTW_t10$		
	(1) t+1	(2) t+1	(3) t+1	(4) t+1_t+5	(5) t+1_t+5	(6) t+1_t+5	(7) t+1_t+10	(8) t+1_t+10	(9) t+1_t+10
AMBG(Z)	0.004	0.004	0.004	0.020***	0.020***	0.020***	0.028***	0.028***	0.028***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.009	0.009	0.009	0.042**	0.042**	0.042**	0.061***	0.061***	0.061***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PCVOL_RATIO(Z)	-0.011***	-0.011***	-0.011***	-0.014***	-0.014***	-0.014***	-0.018***	-0.018***	-0.018***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
PCVOL_RATIO(Z)×AMBG(Z)		0.002**	0.002		0.003	0.005		0.003	0.004
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
PCVOL_RATIO(Z)×RISK(Z)			-0.000			0.003			0.002
			(0.00)			(0.00)			(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	NO	NO	NO	NO	NO	NO	NO	NO	NO
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,002,463	5,002,463	5,002,463	5,001,520	5,001,520	5,001,520	4,999,809	4,999,809	4,999,809
AdjR^2	0.004	0.004	0.004	0.010	0.010	0.010	0.016	0.016	0.016

Table B.8: Option based measures and stock return predictability - firm fixed effects

This table repeats the analysis reported in Table 5 including firm fixed effects. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: The put-call open Interest ratio

	$DGTW_t1$			$DGTW_t5$			$DGTW_t10$		
	(1) t+1	(2) t+1	(3) t+1	(4) t+1_t+5	(5) t+1_t+5	(6) t+1_t+5	(7) t+1_t+10	(8) t+1_t+10	(9) t+1_t+10
AMBG(Z)	0.005**	0.005**	0.005**	0.018***	0.018***	0.018***	0.027***	0.027***	0.027***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.006	0.006	0.006	0.023	0.023	0.023	0.043**	0.043**	0.043**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\Delta PC\_RATIO(Z)$	-0.310***	-0.305***	-0.304***	-0.351***	-0.345***	-0.344***	-0.362***	-0.356***	-0.355***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\Delta PC\_RATIO(Z) \times AMBG(Z)$		0.034***	0.031***		0.044***	0.037***		0.046***	0.039***
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
$\Delta PC\_RATIO(Z) \times RISK(Z)$			-0.005			-0.012***			-0.013**
			(0.00)			(0.00)			(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,822,491	5,822,491	5,822,491	5,820,016	5,820,016	5,820,016	5,817,886	5,817,886	5,817,886
AdjR^2	0.027	0.027	0.027	0.014	0.014	0.014	0.018	0.018	0.018

Panel B: The implied volatility spread measure

	$DGTW_t1$			$DGTW_t5$			$DGTW_t10$		
	(1) t+1	(2) t+1	(3) t+1	(4) t+1_t+5	(5) t+1_t+5	(6) t+1_t+5	(7) t+1_t+10	(8) t+1_t+10	(9) t+1_t+10
AMBG(Z)	0.004	0.004	0.004	0.018***	0.017***	0.017***	0.025***	0.024***	0.024***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.004	0.004	0.006	0.025	0.024	0.026*	0.044**	0.043**	0.045**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
IVS(Z)	0.063***	0.060***	0.054***	0.075***	0.070***	0.063***	0.083***	0.076***	0.068***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$IVS(Z) \times AMBG(Z)$		-0.006***	-0.000		-0.014***	-0.008		-0.021***	-0.013*
		(0.00)	(0.00)		(0.00)	(0.00)		(0.00)	(0.00)
$IVS(Z) \times RISK(Z)$			0.012***			0.012**			0.014*
			(0.00)			(0.00)			(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,614,952	5,614,952	5,614,952	5,613,843	5,613,843	5,613,843	5,612,216	5,612,216	5,612,216
AdjR^2	0.004	0.004	0.004	0.009	0.009	0.009	0.016	0.016	0.016

Table B.9: Call and put options' cumulative delta-hedged returns - firm fixed effects

This table repeat the analysis conducted in Table 6 including firm fixed effects. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	CCUMRET					*PCUMRET*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.139***	-0.175***	-0.183***	-0.183***	-0.181***	-0.193***	-0.246***	-0.284***	-0.309***	-0.341***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
RISK(Z)	0.311***	0.409***	0.500***	0.551***	0.662***	0.314***	0.435***	0.514***	0.578***	0.680***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,099,948	6,005,311	5,935,571	5,877,084	5,776,677	6,019,993	5,927,483	5,859,912	5,803,999	5,708,088
AdjR^2	0.162	0.157	0.164	0.171	0.182	0.106	0.124	0.143	0.158	0.179

Table B.10: Call and put options' cumulative delta-hedged returns - Monthly *RISK* and *AMBG*

To link our option return findings reported in Table 6 with Cao and Han (2013), in this table we also reports the coefficient estimates of the monthly *RISK* and *AMBG* measures (*AvgRISK* and *AvgAMBG*) included in the regressions reported in Table 6. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable (*AvgDEP*), *AMBG*(*AvgAMBG*) and *RISK*(*AvgRISK*). This allows to account for the persistence in the dependent variables, and explore the effect of changes in *AMBG* and *RISK* relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and *t*-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

	CCUMRET(Z)					*PCUMRET(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.138***	-0.174***	-0.182***	-0.184***	-0.185***	-0.194***	-0.251***	-0.292***	-0.321***	-0.360***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
RISK(Z)	0.305***	0.403***	0.492***	0.542***	0.652***	0.313***	0.433***	0.513***	0.577***	0.680***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
AvgAMBG	0.002***	0.002***	0.003***	0.003***	0.003***	0.002***	0.003***	0.003***	0.003***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AvgRISK	-0.030***	-0.039***	-0.049***	-0.055***	-0.069***	-0.028***	-0.036***	-0.042***	-0.047***	-0.055***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm FEs	NO	NO	NO	NO	NO	NO	NO	NO	NO	NO
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,099,959	6,005,322	5,935,581	5,877,097	5,776,690	6,020,006	5,927,494	5,859,923	5,804,013	5,708,099
AdjR^2	0.162	0.156	0.163	0.169	0.177	0.106	0.124	0.141	0.156	0.175

Table B.11: Call and put options' open interest and volume based on firm size subsamples

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A) and volume (Panel B) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics conditioning on firm size. In particular, the dummy variables Size1-Size3 are equal to one if the firm is assigned to size terciles 1-3, respectively, and zero otherwise. $AMBG \times Size1$ — $AMBG \times Size3$ ($RISK \times Size1$ — $RISK \times Size3$) are the interaction of $AMBG$ ($RISK$) with $Size1$-$Size3$ dummy variables. Call and Put measures are reported in Columns 1-5 and Columns 6-10, respectively. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG(AvgAMBG)$ and $RISK(AvgRISK)$. This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
$AMBG(Z) \times Size1$	-0.013***	-0.013***	-0.013***	-0.013***	-0.013***	-0.006***	-0.007***	-0.007***	-0.007***	-0.007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Size2$	-0.012***	-0.013***	-0.013***	-0.013***	-0.013***	-0.010***	-0.011***	-0.011***	-0.011***	-0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Size3$	-0.008***	-0.008***	-0.008***	-0.008***	-0.008***	-0.011***	-0.011***	-0.011***	-0.011***	-0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size1$	-0.007***	-0.006***	-0.006***	-0.006***	-0.006***	0.009***	0.010***	0.010***	0.010***	0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size2$	0.001	0.003	0.005**	0.006**	0.007***	0.021***	0.022***	0.024***	0.024***	0.024***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size3$	0.019***	0.025***	0.031***	0.034***	0.039***	0.060***	0.062***	0.066***	0.067***	0.069***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,887,441	5,887,438	5,887,517	5,887,539	5,887,564	5,806,847	5,806,844	5,806,942	5,806,963	5,807,012
AdjR^2	0.843	0.843	0.844	0.844	0.842	0.856	0.855	0.857	0.857	0.854

Panel B: Volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
$AMBG(Z) \times Size1$	-0.014***	-0.011***	-0.011***	-0.011***	-0.011***	-0.005***	-0.006***	-0.005***	-0.006***	-0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Size2$	-0.028***	-0.019***	-0.017***	-0.016***	-0.016***	-0.022***	-0.014***	-0.012***	-0.011***	-0.010***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Size3$	-0.043***	-0.022***	-0.016***	-0.014***	-0.013***	-0.042***	-0.022***	-0.016***	-0.013***	-0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size1$	0.112***	0.044***	0.024***	0.018***	0.013***	0.104***	0.044***	0.026***	0.020***	0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size2$	0.175***	0.073***	0.041***	0.031***	0.023***	0.172***	0.075***	0.047***	0.035***	0.030***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Size3$	0.288***	0.150***	0.102***	0.087***	0.072***	0.312***	0.171***	0.122***	0.103***	0.089***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,008,137	5,940,699	5,924,982	5,910,826	5,884,918	5,922,273	5,857,357	5,841,742	5,828,234	5,802,097
AdjR^2	0.402	0.409	0.408	0.404	0.395	0.371	0.374	0.372	0.367	0.359

Table B.12: Call and put options' open interest and volume - sub periods

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A) and volume (Panel B) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity ($AMBG$), risk ($RISK$), and other firm characteristics conditioning on three subperiods. In particular, the dummy variables Sub1-Sub3 are equal to one if the sample period is 2002-2006, 2007-2012, and 2013-2018, respectively, and zero otherwise. $AMBG \times Sub1$–$AMBG \times Sub3$ ($RISK \times Sub1$ — $RISK \times Sub3$) are the interaction of $AMBG$($RISK$) with $Sub1$-$Sub3$ dummy variables. Call and Put measures are reported in Columns 1-5 and Columns 6-10, respectively. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable ($AvgDEP$), $AMBG$($AvgAMBG$) and $RISK$($AvgRISK$). This allows to account for the persistence in the dependent variables, and explore the effect of changes in $AMBG$ and $RISK$ relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
$AMBG(Z) \times Sub1$	-0.011***	-0.012***	-0.012***	-0.013***	-0.014***	-0.011***	-0.012***	-0.012***	-0.012***	-0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Sub2$	-0.014***	-0.014***	-0.014***	-0.015***	-0.015***	-0.015***	-0.016***	-0.016***	-0.016***	-0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Sub3$	-0.009***	-0.010***	-0.010***	-0.010***	-0.010***	-0.011***	-0.011***	-0.012***	-0.012***	-0.012***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub1$	0.004*	0.006***	0.007***	0.008***	0.009***	0.020***	0.021***	0.021***	0.022***	0.022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub2$	-0.005**	-0.003	-0.001	-0.000	0.001	0.015***	0.016***	0.018***	0.018***	0.018***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub3$	-0.006***	-0.006***	-0.005***	-0.005***	-0.006***	0.011***	0.012***	0.013***	0.013***	0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,887,441	5,887,438	5,887,517	5,887,539	5,887,564	5,806,847	5,806,844	5,806,942	5,806,963	5,807,012
AdjR^2	0.843	0.843	0.844	0.844	0.842	0.856	0.855	0.857	0.856	0.854

Panel B: Volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
$AMBG(Z) \times Sub1$	-0.031***	-0.021***	-0.020***	-0.019***	-0.018***	-0.031***	-0.021***	-0.017***	-0.017***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Sub2$	-0.044***	-0.027***	-0.022***	-0.020***	-0.020***	-0.043***	-0.026***	-0.021***	-0.018***	-0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$AMBG(Z) \times Sub3$	-0.044***	-0.022***	-0.015***	-0.013***	-0.011***	-0.043***	-0.022***	-0.015***	-0.012***	-0.009***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub1$	0.147***	0.069***	0.043***	0.037***	0.030***	0.136***	0.066***	0.044***	0.036***	0.032***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub2$	0.145***	0.060***	0.033***	0.024***	0.016***	0.139***	0.060***	0.036***	0.025***	0.018***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
$RISK(Z) \times Sub3$	0.123***	0.047***	0.024***	0.018***	0.014***	0.122***	0.052***	0.032***	0.025***	0.021***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Firm FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date FEs	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Date Cluster	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	6,008,137	5,940,699	5,924,982	5,910,826	5,884,918	5,922,273	5,857,357	5,841,742	5,828,234	5,802,097
AdjR^2	0.400	0.409	0.408	0.404	0.395	0.369	0.373	0.371	0.367	0.359

Table B.13: *AMBG* and other uncertainty proxies

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), risk (*RISK*), and other firm characteristics. In each panel, "Base" refers to the main specification reported in the paper. "No uncertainty controls" is a specification that excludes *RISK* and *AvgRISK*. "Full uncertainty controls" is a specification that includes *RISK* together with *VOV*, *VOM*, *SKEW*, and *KURT* together with their rolling averages. For brevity, the table only reports the *AMBG* coefficients. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
Base										
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
No uncertainty controls										
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***	-0.015***	-0.016***	-0.016***	-0.016***	-0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Full uncertainty controls										
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.013***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B: Trading Volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
Base										
AMBG(Z)	-0.040***	-0.023***	-0.017***	-0.016***	-0.015***	-0.039***	-0.024***	-0.018***	-0.016***	-0.013***
	(-16.88)	(-16.03)	(-15.07)	(-14.06)	(-13.78)	(-15.32)	(-14.28)	(-14.36)	(-13.88)	(-11.92)
No uncertainty controls										
AMBG(Z)	-0.051***	-0.028***	-0.021***	-0.019***	-0.018***	-0.049***	-0.028***	-0.021***	-0.018***	-0.015***
	(-19.73)	(-17.48)	(-16.02)	(-15.08)	(-14.27)	(-18.09)	(-15.67)	(-14.94)	(-14.44)	(-13.24)
Full uncertainty controls										
AMBG(Z)	-0.041***	-0.023***	-0.018***	-0.017***	-0.016***	-0.041***	-0.023***	-0.018***	-0.015***	-0.014***
	(-17.08)	(-15.56)	(-14.91)	(-14.12)	(-13.45)	(-15.78)	(-13.89)	(-13.64)	(-13.39)	(-12.20)

Panel C: Cumulative delta-hedged returns

	CCUMRET(Z)					*PCUMRET(Z)*			
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3
Base									
AMBG(Z)	-0.139***	-0.175***	-0.183***	-0.183***	-0.181***	-0.193***	-0.246***	-0.284***	-0.309***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)
No uncertainty controls									
AMBG(Z)	-0.163***	-0.206***	-0.221***	-0.224***	-0.231***	-0.216***	-0.279***	-0.323***	-0.354***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Full uncertainty controls									
AMBG(Z)	-0.142***	-0.179***	-0.188***	-0.188***	-0.188***	-0.199***	-0.253***	-0.292***	-0.318***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)

Table B.14: *AMBG*, *VOM* and *VOV*

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), volatility-of-mean (*VOM*), volatility-of-volatility (*VOV*) and other firm characteristics. There are two separate specifications in each panel based on *VOM*("*AMBG* and *VOM*") and *VOV*("*AMBG* and *VOV*"), controlling for their trailing averages. For brevity, the table only reports the *AMBG*, *VOM*, and *VOV* coefficients. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG and *VOM*										
AMBG(Z)	-0.012***	-0.012***	-0.012***	-0.013***	-0.013***	-0.015***	-0.015***	-0.015***	-0.016***	-0.016***
	(-12.43)	(-12.77)	(-13.12)	(-13.35)	(-13.83)	(-13.91)	(-14.18)	(-14.63)	(-14.80)	(-15.29)
VOM(Z)	0.003***	0.004***	0.004***	0.005***	0.005***	0.010***	0.010***	0.011***	0.011***	0.011***
	(3.87)	(5.11)	(5.97)	(6.39)	(6.80)	(12.62)	(13.08)	(13.76)	(13.88)	(13.79)
AMBG and *VOV*										
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***	-0.015***	-0.015***	-0.016***	-0.016***	-0.016***
	(-12.59)	(-12.94)	(-13.30)	(-13.54)	(-14.03)	(-14.06)	(-14.33)	(-14.79)	(-14.96)	(-15.45)
VOV(Z)	-0.004***	-0.004***	-0.004***	-0.005***	-0.005***	-0.001**	-0.001**	-0.001**	-0.001***	-0.002***
	(-9.08)	(-9.01)	(-9.22)	(-9.49)	(-9.75)	(-2.50)	(-2.44)	(-2.52)	(-2.65)	(-3.17)

Panel B: Trading volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG and *VOM*										
AMBG(Z)	-0.049***	-0.027***	-0.021***	-0.018***	-0.017***	-0.048***	-0.027***	-0.020***	-0.017***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
VOM(Z)	0.100***	0.038***	0.023***	0.019***	0.014***	0.095***	0.038***	0.024***	0.018***	0.016***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
AMBG and *VOV*										
AMBG(Z)	-0.050***	-0.027***	-0.021***	-0.019***	-0.017***	-0.049***	-0.027***	-0.020***	-0.017***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
VOV(Z)	0.020***	0.005***	0.001	0.000	-0.000	0.022***	0.007***	0.003***	0.002**	0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C: Cumulative delta-hedged returns

	CCUMRET(Z)					*PCUMRET(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG and *VOM*										
AMBG(Z)	-0.160***	-0.202***	-0.216***	-0.220***	-0.225***	-0.215***	-0.277***	-0.320***	-0.350***	-0.389***
	(-22.95)	(-22.15)	(-20.03)	(-17.53)	(-15.00)	(-25.95)	(-25.02)	(-24.38)	(-23.76)	(-21.44)
VOM(Z)	0.289***	0.374***	0.439***	0.477***	0.546***	0.296***	0.380***	0.439***	0.477***	0.541***
	(36.46)	(39.15)	(41.07)	(40.67)	(39.49)	(41.89)	(47.15)	(47.56)	(45.93)	(44.82)
AMBG and *VOV*										
AMBG(Z)	-0.162***	-0.205***	-0.220***	-0.224***	-0.231***	-0.216***	-0.278***	-0.322***	-0.352***	-0.391***
	(-23.33)	(-22.43)	(-20.22)	(-17.64)	(-15.10)	(-26.09)	(-25.17)	(-24.46)	(-23.81)	(-21.45)
VOV(Z)	0.026***	0.031***	0.050***	0.059***	0.076***	0.033***	0.053***	0.070***	0.086***	0.107***
	(5.50)	(5.31)	(7.77)	(8.01)	(8.78)	(7.90)	(10.34)	(11.98)	(13.00)	(13.41)

Table B.15: *AMBG* and dispersion of analyst forecast (*DAF*)

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), risk (*RISK*) the dispersion of analyst forecasts (*DAF*) and other firm characteristics. For brevity, the table only reports the *AMBG*, *RISK*, and *DAF* coefficients. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. All specifications include the trailing avergaes of the dependent variable (*AvgDEP*), *AMBG*(*AvgAMBG*) and *RISK*(*AvgRISK*). This allows to account for the persistence in the dependent variables, and explore the effect of changes in *AMBG* and *RISK* relative to their trailing benchmarks. (Z) stands for a Z-Score adjustment. Firm and date fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.012***	-0.012***	-0.013***	-0.013***	-0.014***	-0.014***	-0.014***	-0.014***	-0.014***	-0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	-0.004***	-0.003**	-0.001	-0.001	-0.000	0.015***	0.016***	0.017***	0.017***	0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
DAF(Z)	0.005***	0.006***	0.006***	0.006***	0.006***	0.005***	0.005***	0.005***	0.005***	0.005***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B: Trading volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.040***	-0.023***	-0.018***	-0.017***	-0.016***	-0.039***	-0.023***	-0.018***	-0.015***	-0.013***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.137***	0.058***	0.033***	0.026***	0.020***	0.131***	0.059***	0.037***	0.029***	0.024***
	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
DAF(Z)	0.008***	0.008***	0.007***	0.008***	0.007***	0.010***	0.011***	0.011***	0.011***	0.011***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C: Cumulative delta-hedged returns

	CCUMRET(Z)					*PCUMRET(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.139***	-0.175***	-0.183***	-0.183***	-0.181***	-0.193***	-0.246***	-0.284***	-0.309***	-0.341***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
RISK(Z)	0.311***	0.409***	0.500***	0.551***	0.662***	0.314***	0.435***	0.514***	0.578***	0.680***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
DAF(Z)	-0.000	0.005	0.015**	0.020**	0.038***	0.004*	0.004	0.005	0.008	0.013
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)

Table B.16: *AMBG* controlling for market *AMBG* and *VIX*

This table reports the findings from daily panel regressions, in which call and put stock option open interest (Panel A), trading volume (Panel B), and cumulative delta-hedged returns (Panel C) on trading day $t, \ldots, t+5$ are regressed on trading day t's ambiguity (*AMBG*), risk (*RISK*) and other firm characteristics controlling for changes in market ambiguity (Δ*MktAMBG*) and changes in *VIX* (Δ*VIX*). For brevity, the table only reports the *AMBG*, *RISK*, *MktAMBG* and *VIX* coefficients. The sample period is from January 2002 to December 2018. The options trading data is taken from OptionMetrics. All variables are defined in Table B.1. (Z) stands for a Z-Score adjustment. Firm and day-of-the-week fixed effects are included in each specification. Standard errors are double clustered by firm and date, and t-statistics are reported in parentheses below the coefficient estimates. Statistical significance at the 10%, 5%, and 1% level is indicated by *, **, and ***, respectively.

Panel A: Open Interest

	COI(Z)					*POI(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.009***	-0.010***	-0.010***	-0.011***	-0.011***	-0.025***	-0.025***	-0.025***	-0.025***	-0.025***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	-0.020***	-0.019***	-0.017***	-0.016***	-0.015***	0.032***	0.033***	0.034***	0.035***	0.034***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ*MktAMBG (Z)*	-0.001	-0.001	-0.001	-0.001	-0.001	0.001*	0.001*	0.001*	0.001*	0.001*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ*VIX (Z)*	0.008***	0.008***	0.008***	0.008***	0.008***	-0.007***	-0.006***	-0.006***	-0.006***	-0.006***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel B: Trading volume

	CVOL(Z)					*PVOL(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.036***	-0.023***	-0.018***	-0.016***	-0.015***	-0.042***	-0.028***	-0.022***	-0.020***	-0.017***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
RISK(Z)	0.110***	0.047***	0.025***	0.019***	0.013***	0.125***	0.062***	0.041***	0.032***	0.027***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Δ*MktAMBG (Z)*	-0.000	-0.001	-0.001	-0.000	0.000	0.000	0.001	-0.000	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ*VIX (Z)*	0.018***	0.008***	0.006***	0.005***	0.004**	-0.004**	0.001	-0.000	0.001	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Panel C: Cumulative delta-hedged returns

	CCUMRET(Z)					*PCUMRET(Z)*				
	(1) t	(2) t+1	(3) t+2	(4) t+3	(5) t+5	(6) t	(7) t+1	(8) t+2	(9) t+3	(10) t+5
AMBG(Z)	-0.159***	-0.227***	-0.274***	-0.303***	-0.347***	-0.325***	-0.464***	-0.562***	-0.642***	-0.748***
	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)	(0.01)	(0.02)	(0.03)	(0.03)	(0.04)
RISK(Z)	0.431***	0.647***	0.736***	0.857***	1.057***	0.709***	0.942***	1.147***	1.290***	1.514***
	(0.03)	(0.08)	(0.08)	(0.09)	(0.10)	(0.04)	(0.05)	(0.07)	(0.07)	(0.08)
Δ*MktAMBG (Z)*	0.008	0.040*	-0.029	0.002	-0.038	0.005	-0.056***	-0.053**	-0.076**	-0.078**
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)
Δ*VIX (Z)*	1.111***	0.975***	1.006***	1.077***	1.052***	0.311***	0.696***	0.682***	0.729***	0.785***
	(0.05)	(0.07)	(0.08)	(0.08)	(0.11)	(0.05)	(0.05)	(0.07)	(0.08)	(0.09)

Detecting Informed Trading Risk from Undercutting Activity in Limit Order Markets *

Yashar H. Barardehi Peter Dixon Qiyu Liu

May 18, 2024

Abstract

We use abnormal undercutting activity ($QIDRes$) to measure informed trading risk, reflecting liquidity-providing algorithms competing less to fill marketable orders when adverse selection exposure rises. Despite its simple construction, when examined around information events, $QIDRes$ behaves similarly to existing measures of informed trading intensity/probability whose constructions are complex. $QIDRes$ predicts arrivals and magnitudes of imminent information events. Moreover, episodes of high $QIDRes$ coincide with weaker subsequent price reversals, increased accumulation/covering of short interest, and increased informed institutional trades. $QIDRes$ from prior quarters positively predicts monthly stock returns, especially among stocks with tighter short sale constraints. Since $QIDRes$ is orthogonal to stock liquidity and is not a persistent stock characteristic, we attribute its return predictability to limits to arbitrage.

JEL Classification Codes: G14

Keywords: Informed Trading, Undercutting, Asset Pricing, Liquidity, Limits to Arbitrage

*We are grateful for comments and suggestions from Robert Battalio, Dan Bernhardt, Luis Ceballos, Zhi Da, Joel Hasbrouck, Tim Johnson, Travis Johnson, Albert Menkveld, Mitch Warachka, and Liyan Yang, as well as conference participants at LMU Corporate Finance Conference. Barardehi (barardehi@chapman.edu) is at the Argyros College of Business & Economics, Chapman University and the U.S. Securities and Exchange Commission. Dixon (DixonP@sec.gov), and Liu (LiuQi@sec.gov) are at the U.S. Securities and Exchange Commission.

1 Introduction

Informed trading is a key concept in various areas of financial economics including market efficiency, market structure, and the cost of capital. However, empirically measuring informed trading risk is difficult since informed traders conceal their presence by endogenously adjusting trading behavior with market conditions (Kyle (1985), Anand, Irvine, Puckett, and Venkataraman (2012)). This reality makes it difficult for researchers to empirically differentiate informed trading from other market conditions like liquidity (Ahren (2020), Duarte and Young (2009)). In this paper we propose an easy-to-compute, intuitive measure of nondirectional informed trading risk that is orthogonal to liquidity, performs at least as well as prominent existing measures in empirical tests, and only requires trades and quotes data. Importantly, our measure is computable at the daily, or even finer, frequencies for securities traded in any modern limit order market.

Our approach exploits the intuition that liquidity providers will compete less to trade against marketable orders they perceive to be informed. Specifically, we expect that the phenomena known as undercutting runs, or just runs (Foley, Dyhrberg, and Svec (2022), Foley, Meling, and Ødegaard (2021)), will decrease when informed trading risk is high and that this change in behavior will be observable in the trade and quote data. Undercutting refers to a trader using trivial price improvement to get their order to the front of the limit order queue. Undercutting runs occur when multiple trading algorithms repeatedly undercut each other as they compete to provide liquidity to an expected upcoming marketable order. In modern markets most liquidity providers have no affirmative obligation to provide liquidity in the face of informed or "toxic" order flow—liquidity provision that would lead them to incur losses ((Glosten and Milgrom (1985), Menkveld (2013)). Consequently, when informed trading risk is high, the willingness to "undercut" rivals will decrease, or disappear, reducing both the number and length of undercutting runs.[1]

The existing empirical literature on undercutting has primarily relied on proprietary account level data (e.g. Foley et al. (2021), Foley et al. (2022)) to identify runs. However, we observe that the nature of undercutting runs gives rise to patterns in the trades and quotes data that are

[1]Importantly, liquidity providing algorithms operate with inventory holding horizons as short as a few seconds (Conrad and Wahal (2020)). Hence, when dodging directional informed flow expected to persist beyond these holding horizons, they limit providing liquidity, and hence undercutting, on *both sides* of the market to avoid unwanted inventory accumulation. Thus, despite the directional nature of informed trading, liquidity-providing algorithms react to increased informed trading risk by undercutting less on both sides of the market. Put differently, abnormally low undercutting reveals the extent of liquidity providers' concerns about non-directional informed trading risk.

identifiable without proprietary data. Specifically, the hallmark of a run is a sequence of single tick improvements in the best quoted price on one side of the market followed by a sudden drop back to the pre-run prices as the incoming marketable order executes the quote provided by the winner of the undercutting run. Empirically, this pattern can be analyzed by studying the difference between NBBO quote improvements and trade driven NBBO quote deteriorations. This leads us to measure undercutting activity by standardizing this difference at the stock-day level.

Specifically, our measure of undercutting, the QID ratio, reflects the total number of NBBO quote improvements observed on a given stock-day minus the corresponding number of trade-driven NBBO quote deteriorations on that same day, all divided by the sum of these two quantities. The construction of QID imposes boundaries of -1 and 1.[2] QID moving closer to 1 signifies increases in undercutting runs.

We establish the validity of QID as an undercutting measure by documenting its inverse relationship with undercutting costs. First, we exploit the exogenous change in the costs of undercutting driven by the SEC's Tick Size Pilot program (TSP). By temporarily raising the tick size from 1¢ to 5¢ for some stocks, the TSP quintupled a major component of undercutting costs (Werner, Rindi, Buti, and Wen (2022)). In standard difference-in-difference analysis, we find this exogenous increase in undercutting costs at TSP implementation *reduced* QID by about 0.44; whereas the TSP conclusion virtually mirrored the implementation results where QID *increased* by an average of 0.42.[3] Second, we exploit the positive impact of stock splits and the negative impact stock reverse splits on the costs of undercutting as reflected by relative tick sizes, i.e., 1¢ divided by share price. We find that QID significantly falls after stock splits, but significantly rises after stock reverse splits. Collectively, these findings highlight a strong inverse relationship between the costs of undercutting and QID, bolstering our interpretation of QID as a measure of undercutting activity.

We address two additional issues before employing QID to capture informed trading risk. First, prior literature demonstrates that informed trading risk measures tend to conflate informed trading and liquidity effects (Duarte and Young (2009), Ahren (2020)). Thus, QID could simply

[2]Because (1) we exclude best quote deteriorations due to limit order cancellations and (2) executions of marketable orders likely lead to best quote deteriorations, we expect QID to be slightly negative in the absence of undercutting.

[3]Unlike the TSP's heterogeneous effects on many other outcomes conditional on how binding the 5¢ tick was, its effect on QID are fairly homogeneous. Additionally, our analysis satisfies the heuristic hurdles when re-using experiments as all t-statistics range between 9–38, multiples of thresholds proposed by Heath, Ringgenberg, Samadi, and Werner (2020).

be capturing variations in liquidity that could also affect the willingness of a liquidity provider to undercut. In fact, Figure 2 documents a positive association between QID and stock illiquidity, measured by relative quoted bid-ask spread.[4] Second, the contributions of informed trading risk and liquidity to the variation in QID may vary in the cross-section. Thus, we must account for stock-specific effects to arrive at a measure that is comparable across stocks.

We address both concerns by orthogonalizing QID to liquidity and then standardizing it to make it comparable across stocks. Crucially, adopting this approach distinguishes our measure of informed trading risk from all existing measures, other relevant mircostructure outcomes, or stock characteristics. First, for each quarter and each stock, we fit a regression of daily QID on time-weighted relative bid-ask spread to control for liquidity conditions. Second, we apply the coefficients from the first step to the following quarter's realizations to produce estimates of the unexpected (residual) QID, i.e., undercutting activity that is orthogonal to liquidity. Next, we scale these stock-specific estimates of liquidity adjusted unexpected QID by the standard deviation of observed QID from the prior quarter. Lastly we multiply the resulting ratio by -1 to produce a positive, instead of inverse, measure of informed trading risk. We dub the resulting measure $QIDRes$. Consistent with its construction, $QIDRes$ satisfies two properties at the daily frequencies (1) it is distributed with a mean and a standard deviation close to 0 and 1, respectively; (2) it has nearly zero correlation with relevant contemporaneous mictostructure and liquidity outcomes such as quoted, effective, and realized spreads; price impact; volatility; and trading volume. This lack of correlation extends with respect to common measures of liquidity such as quoted spreads, effective spreads, and lambda as well as stock characteristics when we aggregate measures at quarterly frequencies.

We examine the behavior of $QIDRes$ around multiple information events known to be associated with informed trading. We also compare it's behavior to that of other prominent measures of informed trading such as the Informed Trading Intensity (ITI) measures of Bogousslavsky, Fos, and Muravyev (2023); Probability of Informed Trading (PIN) measures—see Duarte, Hu, and Young (2020) for a discussion of the various PIN-based measures; and the multi-market information

[4]Relative quoted spread is particularly relevant for undercutting in U.S. equity markets. Dollar bid-ask spread together with the 1¢ tick size reflect the number of 1-¢-apart price levels potentially available for undercutting runs. However, the value per share of the stock, usually approximated by the quote midpoint in microstructure applications, together with the minimum lot size of 100 shares, required for any effective undercutting, reflect the minimum dollar value transferred per transaction as an undercutting run's winner trades. The minimum tick and lot size are fixed across all stocks, and relative bid-ask spread, defined as the ratio of dollar bid-ask spread to NBBO midpoint, controls for the two remaining relevant factors.

asymmetry (MIA) measure of Johnson and So (2018).

We document that around earnings announcements, unscheduled press releases, and news arrivals there is a significant spike in $QIDRes$ that takes up to 10 days to rebound. This is a pattern we also observe with the other measures of informed trading risk. Moreover, we find that the magnitude and persistence of the spikes in $QIDRes$ are related to the size of the the post-event returns: information events with larger increases in $QIDRes$ are followed by larger post-event absolute returns; and for such events, post-event $QIDRes$ rebounds more slowly. Finally, we test the predictive power of $QIDRes$ for informational events, exploring the notion that market makers may learn from order flow about upcoming information events (Chae (2005)). In fact, we find that increases in $QIDRes$ predict imminent arrivals of *unscheduled* information events.

We next provide evidence inconsistent with $QIDRes$ solely capturing 'sniping risk.' The literature pioneered by Budish, Cramton, and Shim (2015) shows in continuous-time limit order markets liquidity providers face adverse selection costs due to sniping risk, rather than their information disadvantages about fundamental values.[5] Relevant for our analysis is the intuition that liquidity providers should become reluctant to undercut when sniping risk rises due public news arrivals, leading to increases in $QIDRes$ around major information arrivals. To address this, instead of only relying on information events, we link $QIDRes$ to more direct sources of informed trading.

First, we explore the relation between $QIDRes$ and changes in short interest. A large literature documents that changes in short interest are strong predictors of future stock performance (see, e.g., Boehmer, Huszar, and Jordan (2010), Dixon and Kelley (2022)). Consistent with $QIDRes$ being linked to informed trading, we document that $QIDRes$ is significantly higher during periods with large absolute changes in short interest, even after excluding periods that overlap with information events. Second, we examine the behavior of $QIDRes$ around stock-days with informed mutual-fund trades, as identified by Barardehi, Da, and Warachka (2022). Again, we document significantly larger $QIDRes$ on informed mutual fund trading days than on days without such trading.[6]

We also provide evidence suggesting that $QIDRes$ is not simply reflecting inventory manage-

[5] With differences in order processing speeds across traders, the arrival of public news leaves some resting limit orders of slow trades stale not because of information asymmetries but because these traders cannot cancel their orders fast enough. In turn, faster traders benefit from picking off (sniping) these stale orders at the loss of slow traders (also see Menkveld and Zoican (2017)).

[6] We document similar behavior for both short interest and informed trading days with most other informed trading intensity/probability measures.

ment concerns of liquidity providers. Due to capital constraints, liquidity providers with unbalanced inventories avoid accumulating additional inventory or charge a premium to do so (Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010)). This can translate to reduced undercutting, i.e., higher $QIDRes$, as such liquidity providers demand greater compensation for providing liquidity. As demonstrated by Hendershott and Menkveld (2014), these inventory dynamics give rise to short-term price pressure followed by price reversals. Thus, if $QIDRes$ were reflecting inventory management concerns then stock-days with higher $QIDRes$ would be associated with stronger price reversals. However, we find the exact opposite occurs. Stock-days with higher $QIDRes$ are followed by *weaker* reversals which further suggests informed trading. Moreover, we show that these reversals patterns persist when we control for realized volatility—in fact, $QIDRes$ is nearly orthogonal to contemporaneous realized volatility

The inverse link between $QIDRes$ and subsequent reversals is also at odds with increased undercutting reflecting the increased use of limit orders by informed investors. Traders possessing a positive (negative) signal may undercut more on the bid (ask) side, instead of using marketable buy (sell) orders that would reveal their trading intentions. If this mechanism underlies the primitive motive for undercutting, then price reversals should be *weaker* following abnormally high undercutting, i.e., when $QIDRes$ is low, the exact opposite of our findings.

We next document asset pricing implications of $QIDRes$. We first demonstrate that long-short portfolios formed using $QIDRes$ from the prior two quarters earlier produce statistically significant risk-adjusted returns of over 30 basis points per month. This finding extends the one-month return predictability of informed trading intensity measures, documented by Bogousslavsky et al. (2023), to longer horizons. Our additional asset pricing tests involve fixed-effect panel regressions that regress monthly excess returns on lagged $QIDRes$ and stock characteristics, including illiquidity measures. We find positive associations between monthly expected stock returns and $QIDRes$ from the preceding two quarters, i.e., stocks with higher expected informed trading risk have higher returns. We highlight the incremental explanatory power of $QIDRes$ for returns, relative to existing informed trading intensity measures, in "horse race" regressions. These regressions, in addition to $QIDRes$ from the preceding two quarters, include subsets of ten corresponding alternative measures as independent variables. Not only does $QIDRes$ maintain its explanatory power for expected returns when we control for existing measures, but $QIDRes$ is the *only* measure

that significantly predicts future returns across all specifications.

Return predictability of informed trading risk as reflected in $QIDRes$ cannot be interpreted in the context of existing theories such as Easley and O'Hara (2004), interpreting informed trading risk as a stock characteristic, or Duarte and Young (2009), arguing that informed trading risk is correlated with illiquidity. Among unique features of $QIDRes$ are (1) it *does not* constitute a stock characteristic; and (2) it is orthogonal to stock illiquidity. Specifically, $QIDRes$ exhibits no temporal persistence but rather displays modest mean reversion, if anything. Moreoever, reflecting its construction, $QIDRes$ should be orthogonal to persistent stock characteristics such as illiquidity, which we confirm empirically: in the cross-section, $QIDRes$ is minimally correlated with a host of stock characteristics as well as existing informed trading intensity proxies. For example, highlighting the contrast between $QIDRes$ and existing measures, the average absolute pairwise correlation coefficient between $QIDRes$ and five illiquidity measures is only 0.02; whereas the analogue for ITI and PIN measures is 0.15, with individual pairwise correlation coefficients as high as 0.53.

We interpret return predictability of $QIDRes$ in the context of limits to arbitrage. Short sellers who systematically investigate and trade on negative information face short sale constraints, while trading on positive information is not subject to such constraints. To these ends, as Bogousslavsky et al. (2023) also argue, an informed trading risk measure is more likely to capture trading motivated by positive information, rather than negative information. Hence, increases in measures of informed trading risk such as $QIDRes$ should predict higher future returns. Our empirical findings confirm this. When we control for short sale constraints using security lending fees indeed we observe that the return predictability of $QIDRes$ is stronger among stocks with tighter short sale constraints.

2 Linking Undercutting Runs to Informed Trading Risk

We next provide a simple framework that formally links the risk of trading against informed investors from liquidity providers' perspectives to their tendencies to participate undercutting runs. Consider the setup of a simple one period rational expectation equilibrium model based off of Glosten and Milgrom (1985). An asset takes the equally likely value of 0 or 1. The fraction π of liquidity demanders are informed and know the true value of the asset only buying when the the value equals 1 and only selling when the value equals 0. π captures informed trading risk in

the market. The remaining $1-\pi$ fraction of liquidity demanders are uninformed and buy and sell with equal probability. Liquidity providers come in two types: sophisticated and unsophisticated. Unsophisticated liquidity providers, denoted $ULPs$, are passive, competitive, and set prices equal to the conditional expected value of the asset. Sophisticated liquidity providers, denoted $SLPs$, pay a cost c which will, with probability ρ inform them about whether the next trade to arrive is informed or uninformed and on which side of the market the trade will arrive. It does not inform them about the arrival time of the upcoming trade which is random.[7] There are m SLPs where the value m is determined in equilibrium such that the expected profit associated with being an SLP is equal to the cost c. The likelihood that at least one of the m $SLPs$ receives a signal is $\phi = 1-(1-\rho)^m$.

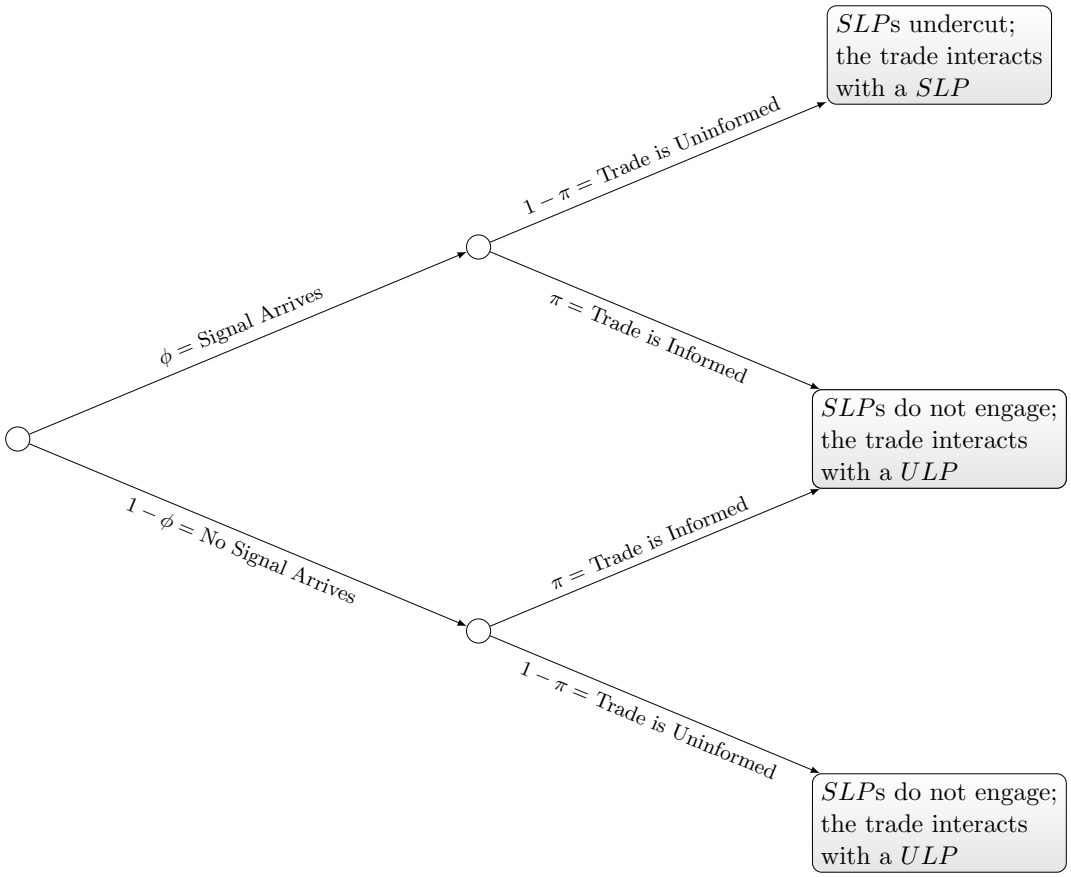
If a SLP receives a signal that an upcoming trade is informed, the SLP will simply sit out and not post any quotes allowing the ULPs to interact with the incoming informed trade. If no SLP receives a signal then all $SLPs$ sit out again allowing the ULPs to interact with the upcoming trade. However, if an SLP receives a signal that the upcoming trade is uninformed, they will undercut the existing quote on that side of the market. The other $SLPs$, whether they receive a signal or not, will observe this quote improvement and will infer that a signal has been received and will submit their own undercutting orders and an undercutting run will ensue.[8] This behavior is outlined in the outcome tree in Figure 1.

From the outcome tree it is straightforward to see that the probability of a run, which can be thought of as the prevalence of runs in that market, is the probability that at least one SLP receives a signal, ϕ, multiplied by the probability that the next trade to arrive is uninformed, $(1-\pi)$ as shown in 1.

[7]The cost c can be thought of as the cost of investing in the capacity to process, analyze, and respond quickly to information based in order flow. As discussed in greater detail in A.1 the exact arrival time of the next trade to arrive is random and follows an exponential distribution with arrival rate parameter λ.

[8]The assumption that all $SLPs$ can infer the signals of others via monitoring quote updates could be relaxed such that only those $SLPs$ receiving a signal engage in the undercutting run without changing any of the key inference. In this case ϕ could be redefined to be the probability that at least two $SLPs$ receive a signal $\phi = 1-(1-\rho)^m - m\rho(1-\rho)^{m-1}$, and all inference remains exactly the same since in both cases ϕ is increasing in both m and ρ. Additionally the profit to undercutting is random, since the arrival of the uninformed trade is random and so it is unclear exactly when during the run the uninformed trade will arrive. However, given that SLPs know the arrival rate of trades, they can compute the expected time during an undercutting run a trade will arrive and so can compute the expected profit of a run that is earned by the winning quote provider, which we denote Π. The likelihood that a given SLP wins the undercutting run is $\frac{1}{m}$, so expected profits to undercutting are $\frac{\Pi}{m}$. For this market to be in equilibrium it must be the case that $c = \frac{\Pi}{m}$ which implies that the number of SLPs is $m = \frac{\Pi}{c}$.

Figure 1. Informed Trading Signal Arrivals and SLPs' Undercutting Choices.



$$P(UndercuttingRun) = \phi(1-\pi). \tag{1}$$

The derivative of equation 1 with respect to informed trading risk is $\frac{\delta P(UndercuttingRun)}{\delta \pi} = -\phi$, which is always less than zero. Thus, when informed trading risk is higher - i.e. π is larger - undercutting diminishes confirming the inverse link between the prevalence of undercutting and informed trading. We present the equilibrium outcomes and comparative statics for the model in Appendix A.1, showing that liquidity gets worse as undercutting risk increases. Our analysis is consistent with empirical findings in the literature (e.g.,Foley et al. (2021) and Foley et al. (2022)). This occurs because undercutting exacerbates adverse selection by preventing uninformed trades from interacting with ULPs who set posted quotes.

3 Related Literature

In this section, we link our paper's contributions to the existing literature. We contribute by developing an informed trading risk measure that is computed using aggregate frequencies of best quote improvements and deteriorations. This simple construction offers several appealing features relative to existing measures: Our measures (1) are implementable for securities traded in any modern limit order market; (2) do not require structural estimations as in, e.g., Easley, Hvidkjaer, and O'Hara (2002); (3) do not require hand-collected data and computationally demanding data-driven techniques as in Bogousslavsky et al. (2023); and (4) do not require significant trading activity in corresponding derivatives markets as in Johnson and So (2018).

Our methodology builds on the literature on order placement, including undercutting, strategies in modern limit order markets. Hasbrouck and Saar (2013) introduced the notion of 'strategic runs' to describe a sequence of order submission/cancellations by an *individual* trader. In this context, strategic runs that end with a trade may resemble successful undercutting efforts of an individual trader (Chordia and Miao (2020)). Bringing this idea to the market level, Foley et al. (2021) and Foley et al. (2022) directly examine undercutting 'runs' by identifying sequences of quote improvements, reflecting order submissions by *multiple* traders, that end with a trade. We posit that, in aggregate market data, best-quote improvements tend to capture undercutting efforts; whereas best-quote deteriorations due to trades tend to capture conclusions of undercutting runs. We measure aggregate undercutting intensity using quote improvements minus quote deteriorations at the stock-day level. Intuitively, exposure to adverse-selection risk due to information asymmetry lowers liquidity providers' willingness to undercut, leading us to propose abnormally low undercutting as a new measure of increased informed trading risk.

We also provide new evidence relevant for the debate about asset pricing implications of informed trading as $QIDRes$ from two prior quarters predicts monthly returns. Easley and O'Hara (2004) predict more frequent informed trading commands higher expected stock returns, with Easley et al. (2002), Kelly and Ljungqvist (2012), and Derrien and Kecskés (2013) providing supportive evidence in different settings. Hughes, Liu, and Liu (2007) and Petacchi (2015), respectively, link more frequent informed trading to higher cost of capital and higher cost of equity. However, Lambert, Leuz, and Verrecchia (2012) predict these links only exist in noncompetitive capital markets, with

Armstrong, Taylor, Core, and Verrecchia (2011) providing supportive empirical evidence. In contrast, Wang (1993) posits that increased presence of informed investors reduces the cost of capital. Relatedly, Duarte and Young (2009) show that the ability of Easley et al. (2002)'s PIN measures to explain expected returns reflects the cross-sectional variation in liquidity, rather than that in prevalence of informed trading. Because, our measure of informed trading risk is unrelated to stock liquidity and does not constitute a persistent stock characteristic, we attribute its return predictability to limits to arbitrage such as short sale constraints.

4 Data and Methodology

4.1 Data

Our main sample runs from January 2010 thought December 2019 and includes NMS-listed common stocks whose share prices were at least \$5 at the end of the preceding month. We obtain intraday quote and trade information from Daily TAQ; daily microstructure outcomes from WRDS Intraday Indicators; daily and monthly price and trade information from Daily and Monthly CRSP, respectively; Book-value information and earnings announcements dates from COMPUSTAT; earnings surprise scores from I/B/E/S; and news information from Ravenpack.

We construct national best bid and ask prices (NBBOs), from 09:45am to 3:45pm each day, following Holden and Jacobsen (2014) by merging Daily TAQ's NBBO and Quote files that are then matched with trades in the same millisecond obtained from Daily TAQ's Trade files. Our daily undercutting measure, QID_{jt}, divides the difference between the number of best quote improvements, on either bid or ask side, and the number of trade-driven best quote deterioration, on either bid or ask side, by the total number of such NBBO updates for stock j on day t. We flag a quote deterioration as trade-driven if it occurs no later than 10 milliseconds after a trade.

$$QID_{jt} = \frac{\#Impr_{jt} - \#DeterTrade_{jt}}{\#Impr_{jt} + \#DeterTrade_{jt}} \tag{2}$$

Panel A in Table 1 contains summary statistics of the national best quoted ask and bid updates. The mean and median daily national best bid (NBB) improvements are 1,046.5 and 579, respectively; the analogous mean and median for national best ask (NBO) are 1,052.9 and 584, re-

spectively. Consistent with the prevalence of undercutting activity, the mean and median of daily trade-driven NBB deteriorations are 279.05 and 110, respectively, with ask-side analogues of 277.7 and 109. More compelling evidence for the prevalence of undercutting obtain from the fractions single-tick quote updates. Liquidity providers are expected to undercut the existing best price by the minimum amount of possible price improvement, i.e., one tick. Hence, best quote improvements are most likely to occur at single-tick updates. By contrast, trade-driven quote deteriorations ending undercutting runs more likely reflect multiple-tick updates as marketable orders may consume the depth available beyond the top of the order book. Consistent with this, on a typical stock-day, near 90% of quote improvements reflect single-tick updates. This is significantly higher than the analogous 61% ratio for trade-driven quote deteriorations. Consequently, and reflecting the larger frequency of best quote improvements than deteriorations, the thirteenth row in Table 1 shows that over 99% of QID observations are positive. Specifically, only 3,222 of observations (only 0.05% of the sample) correspond to a negative QID quantity, even though QID can be as small as -1.

We match QID_{jt} with daily time-weighted dollar spreads (denoted qsp_{jt}) and percent quoted spreads (denoted psp_{jt}) as well as percent effective spreads (denoted $pefsp_{jt}$), realized spreads (denoted $prsp_{jt}$), price impacts (denoted $primp_{jt}$), regular-hour trading volume (denoted tv_{jt}), and volatility of 1-minute quote-midpoint returns (denoted $qvol_{jt}$) obtained from WRDS Intraday Indicators. We also match them with daily returns (denoted r_{jt}), reflecting quote midpoints at close, and trading volumes from Daily CRSP.[9] The CRSP-TAQ linking table provided by WRDS facilitates these mergers.

We then merge our daily data base with earnings announcements (EA), unscheduled corporate events (PR), and news arrivals unassociated with identifiable corporate events (NA), using the announcements' timing to identify the first trading day where trading takes place after an announcement. Earnings announcement dates are obtained from COMPUSTAT. Reflecting the findings of [cite] that the vast majority of such announcements arrive outside regular trading hours, we designate the trading day after the recorded announcement date as the effective announcement date. We obtain dates and timestamps of unscheduled press releases and news arrivals from Ravenpack. For press releases, we focus on Ravenpack "full-article" or "news-flash" observations with "news_relevance" scores of at least 90. For news arrivals, we focus on Raven-

[9] Our daily return calculations account for dividend distributions and overnight adjustments such as stock splits.

pack "full-article" or "news-flash" observations with "news_relevance" scores of at least 95 and no recorded "event_relevance" score. We construct event windows that span the 10 days prior to an announcement and 10 days after the announcement.[10]

We construct a set of stock characteristics for our asset pricing analysis using data from CRSP, COMPUSTAT, and 13F. For stock j in month m, $RET_{j,m-1}$ and $RET_{j,m-2}^{m-12}$, respectively, capture compound returns over the preceding month and the 11 months prior; $M_{j,m-12}$ is market-capitalization based on the closing price 12 months earlier; $DYD_{j,m-1}$ is dividend yield, i.e., the ratio of total dividend distributions over the 12 months ending in month $m-1$ divided by the closing price at the end of month $m-1$. The book-to-market ratio, $BM_{j,m-1}$, is the most recently reported book value divided by market capitalization at the end of month $m-1$.[11] We obtain three-factor Fama-French betas for each stock from Beta Suite by WRDS. Our approach employs weekly data from rolling horizons that span the preceding 104 weeks, requiring a minimum of 52 weeks. For each stock month, the set of betas represent estimates from the estimation horizon ending in the last week of that month. As in Ang, Hodrick, Zhing, and Zhang (2006), we use a CAPM regression using daily observations in each month to construct monthly idiosyncratic volatility measures. We match each monthly observation with previous calendar quarter's fraction of institutionally owned shares outstanding ($IOShr$) and the concentration of such ownership based on a Herfindahl-Hirschman index ($IOShrHHI$) using 13F data.[12]

To control for stock illiquidity in each month m, we use five liquidity measures constructed using daily or intraday obsrvations from month $m-2$: (1) time-weighted dollar quoted spreads (QSP); (2) size-weighted dollar effective spread ($EFSP$); (3) monthly estimates of Kyle's λ, constructed by regressing 5-minute returns (calculated from quote midpoints) on the contemporaneous signed square root of net order flow (estimated using the Lee-Ready algorithm) from the respective month ($Lambda$); (4) a modified version of Amihud (2002)'s measure (AM);[13] and (5) Barardehi, Bernhardt, Da, and Warachka (2023)'s retail-based institutional liquidity measure ($ILMV$). We also construct turnover ratio (TO), defined as the average daily fraction of share volume to shares

[10] For each announcement type (EA, PR, or NA), we focus on the first announcement should multiple announcements cluster over a 20 day period. This endures non-overlapping event windows.

[11] Book value is defined as Compustat's shareholder equity value (seq) plus deferred taxes (txdb). We use the "linktable" from WRDS to match stocks across CRSP and Compustat, dropping stocks without links.

[12] We match CRSP with COMPUSTAT and 13F using the link tables and matching code provided by WRDS.

[13] Barardehi, Bernhardt, Ruchti, and Weidemier (2021) modify this measure by using open-to-close, instead of close-to-close, daily returns to construct Amihud measure's underlying daily liquidity proxy.

outstanding using observations from month $m-2$.

Finally, we obtain lending fee observations at the stock-day level for the 2009-2018 period from Financial Information Service (FIS) Astec Analytics. FIS compiles dollar-weighted average stock lending fees at daily frequencies. For each stock, we aggregate these lending fees annually to estimate expected lending fees over the following calendar year for the respective stock (see Dixon, Corbin, and Kelley (2021) for detailed descriptions of FIS data).

4.2 Abnormal Undercutting Activity and Informed Trading

This section describes the construction of our informed trading riskmeasure, $QIDRes$. The intuition behind our measure reflects market makers' efforts to avoid trading against informed investors. We argue that market makers become less willing to undercut each others' quotes when they perceive incoming order flow to be informed. This notion is also consistent with marker makers' concerns about their limit orders becoming stale and picked off by faster traders, as first observed by Budish et al. (2015). Intuitively, an increased likelihood of informed trading raises the risk of a market maker's limit orders going stale and makes the market maker less willing to jump in front of the queue through undercutting.

It is important to observe that undercutting is more likely to occur in less liquid stocks, e.g., stocks with wider bid-ask spread, for two reasons. First, with a market maker's limit orders coinciding with the NBBO, a wider bid-ask spread provides larger profits per round-trip set of liquidity providing trades as market maker orders are filled by incoming marketable orders. Second, since trades need to improve the price by only 1¢ to undercut, a wider bid-ask spread implies a capacity for undercutting in terms of number ob available intra-spread price ticks. Moreover, undercutting by the best existing quotes by 1¢ is relatively cheaper for higher share prices (see Li and Ye (2023) for discussion on the relevance of the interaction share price and minimum tick size for liquidity provision). This leads us to use relative quoted bid-ask spread to control for the variation in undercutting capacities offered by market conditions. Figure 2 documents a strong positive association between our measure of undercutting, QID, and percent bid-ask spread that yields a R^2 of 58.44%.

To operationalize our intuition that informed trading risk discourages undercutting, we employ a backward-looking procedure to estimate abnormal undercutting activity at the stock-day level.

We first estimate the following regression using daily observations of each stock in each quarter

$$QID_{jt}^{q} = a_j^q + b_j^q \ln(PQSP)_{jt}^{q} + u_{jt}^{q}, \tag{3}$$

where QID_{jt}^{q} measures undercutting activity in stock j on day t of quarter q; $\ln(PQSP)_{jt}^{q}$ is the natural log of the corresponding time-weighted percentage quoted spread; and u_{jt}^{q} is the error term. We then use estimated intercept and slope coefficients from the preceding quarter, i.e., $\widehat{a_j^{q-1}}$ and $\widehat{b_j^{q-1}}$, respectively, to construct daily estimates of unexpected undercutting activity in the current quarter. Finally, we scale unexpected undercutting by the standard deviation of daily QID_{jt}^{q} observations, $S(QID)_j^q$, to account for cross-sectional differences in the variability of undercutting activity. Such variability reflect factors like the more tightly bounded undercutting in stocks with binding minimum tick sizes, which in turn reduces the variation in QID in these stocks.[14] Thus, abnormal undercutting activity for stock j on day t of quarter q is given by:

$$QIDRes_{jt}^{q} = -\frac{QID_{jt}^{q} - \left(\widehat{a_j^{q-1}} + \widehat{b_j^{q-1}} \ln(PQSP)_{jt}^{q}\right)}{S(QID)_j^{q-1}}. \tag{4}$$

Since undercutting is expected to be abnormally low in presence of informed trading, higher $QIDRes$ reflects higher informed trading.

Reflecting the construction of $QIDRes$, we expect it to possess the following two properties: (1) is it distributed with a mean and a standard deviation that are close to 0 and 1, respectively;[15] (2) it should not be correlated with other relevant microstructure and liquidity outcomes. We find strong support of this in the data. The last two rows in Table 1's Panel A report the summary statistics for $QIDRes_{jt}$, indicating that the measure is tightly distributed around zero, with the mean of 0.07 and the standard deviation of 1.53. Panel B in Table 1 contains the correlation coefficients between daily abnormal undercutting activity, $QIDRes_{jt}^{q}$ and contemporaneous micrsotructure outcomes defined earlier, including quoted, effective, and realized spreads; price impact; realized volatility;

[14] Whenever the 1-¢ tick size binds, liquidity providing algorithms may not undercut on exchanges using non-marketable limit orders. As such, one can argue that for stocks where minimum tick more often binds the variation in QID, which we measure using the standard deviation of QID, is lower.

[15] Despite the standardization of unexpected undercutting by equation (4), we do not expect $QIDRes$ to exhibit a mean of exactly 0 and a standard deviation of exactly 1. This is because in each quarter both the conditional mean and the standard deviation used to standardize undercutting are estimated based preceding quarter's data and will differ from the current quarter's realized mean and standard deviation.

absolute daily return; and trading volume. None of these correlation coefficient exceed 0.06 in absolute value, suggesting that $QIDRes$ is orthogonal to these outcomes.

In Appendix A.3, we examine the qualitative robustness of our findings to two modified constructions of $QIDRes$. The first modification, denoted $QIDResInt$, controls for the variation in the unconditional average of undercutting. Our qualitative findings extend if, instead of $S(QID)_j^{q-1}$, we use $\widehat{a_j^{q-1}}$ to normalize unexpected undercutting. The second alternative, denoted $QIDResV$, augments equations (3) and (A.10) with the volatility of 1-minute returns based on quote midpoints. This approach ensures that our measures do not conflate informed trading risk with the effect of higher volatility, e.g., reflecting more frequently arrivals of purely public information, that can also deter liquidity provision and undercutting. This modification also leaves our qualitative findings unaffected.

5 Results

5.1 The Impact of Undercutting Costs on *QID*

We begin our analysis by establishing the validity of the QID ratio as a measure of undercutting. To do so, we leverage the tick size pilot (TSP), during which a select number of stocks had their minimum tick sizes increased from 1¢ to 5¢—see, e.g., Werner et al. (2022), for a detailed description of the experiment. An increase in the tick size will decrease runs by making undercutting more expensive. For a TSP stock, the cost to undercut increased by five fold. Consequently, we expect the implementation of TSP to be associated with a decrease in QID and that the conclusion of the TSP will be associated with a reversal.

We study two TSP event windows: one around the imposition of TSP and the other around its conclusion. For our analysis of the imposition of the TSP, we examine the time window of 08/11/2016 through 12/15/2016. We follow Griffith and Roseman (2019) and exclude from this window the trading days spanning the staggered imposition of the TSP which comprise 10/03/2016–10/23/2016.[16] Our analysis of the imposition of the TSP has a pre-period where both the pilot and control stocks had a tick of 1¢, running from 8/11/2016 to 10/02/2016, and a treatment period

[16]Some effects related to the tick size change may not occur instantaneously as market participants may need time to optimize systems and adapt behavior. Excluding the imposition period helps mitigate some of this noise that may muddle inference of the steady state effects of the tick size change.

where pilot stocks had a 5¢ tick and control stocks had a 1¢ tick, running from 10/24/2016 to 12/15/2016. Our analysis of the conclusion of the TSP runs from 08/07/2018 through 11/20/2018, during which the minimum tick size for stocks in TSP Test Groups was simultaneously reduced from 5¢ to 1¢ on 10/01/2018.[17]

We compare undercutting activity, QID, of control stocks, denoted C, to those of TSP Test Groups 1 and 2, denoted G1 and G2, respectively. Reflecting the similarities between G1 and G2 and to increase the statistical power of our tests, we combine G1 and G2 stocks together. The "`tick size pilot indicator`" flag in TAQ data identifies control and pilot stocks as well as the exact dates tick size changes were enforced for each pilot stock, facilitating accurate identifications of enforcement dates when tick changes were enforced or lifted with delays relative to the dates intended by the program. Stocks that changed test groups or that were removed from the TSP, for any reason, are excluded, as are stock-days with previous day's closing prices below $5.00.

Our estimation strategy is similar to Barardehi, Dixon, Liu, and Lohr (2023) who show that the same change in the tick size due to TSP had opposing impacts on certain outcomes depending on the extent to which minimum ticks were binding pre-shock. But more important for our analysis is that undercutting runs are affected by how tight the bid-ask spread is, and thus how many price levels competing liquidity providing algorithms can use to undercut. Hence, we assign each TSP stock to one of four bins based on their prevailing time-weighted quoted spread prior to the imposition and conclusion of the TSP. For the imposition window, stocks are classified into four bins according to their quoted spreads in May and June of 2016:[18] : bin 1 (tick constrained) 5¢ or less quoted spread, bin 2 (near-tick constrained) greater than 5¢ but less than 10¢, bin 3 (intermediate spread) greater than 10¢ but less than 15¢, and bin 4 (wide spread) greater than 15¢. For the conclusion of the TSP, we assign stocks to bins reflecting average quoted spreads in May and June 2018: bin 1 (tick constrained) less than 5.5¢, bin 2 (near-tick constrained) greater than 5.5¢ but less than 10¢,[19] bin 3 (intermediate spread) greater than 10¢ but less than 15¢, and bin 4 (wide

[17] Following Rindi and Werner (2019), we remove trading days coinciding with Labor Day, Thanksgiving, and Black Friday from our sample. We also do not omit the period surrounding the conclusion of the TSP as we do with the imposition of the TSP because nearly all TSP stocks returned to a 1¢ tick simultaneously, with market participants returning to a familiar trading environment, i.e., one that had continued to operate on the majority of stocks. For these reasons, we generally view the conclusion of the TSP as a cleaner test than the TSP imposition.

[18] Specifically we use WRDS Intraday Indicators data for time-weighted average quoted spread for each stock during regular trading hours and compute a simple average across all trading days in May and June 2016.

[19] This slight modification of bin 1's threshold reflects the restrictions put in place by the TSP. The 5¢ tick size creates a floor on quoted spreads making it all but impossible for a TSP stock to have a time-weighted quoted spread

spread) greater than 15¢.

Our difference-in-difference strategy estimates the impact of an exogenous change in tick size, hence undercutting costs, on QID. We estimate

$$QID_t^j = \alpha_0 + \alpha_p Pilot^j + \alpha_e Event_t^j + \beta \left(Pilot^j \times Event_t^j \right) + u_t + \varepsilon_t^j, \tag{5}$$

by event window and bin, where QID_t^j is stock j's undercutting activity on day t; $Pilot_t$ is an indicator variable that equals 1 for treated stocks (G1 or G2) and equals 0 for control stocks; $Event_t^j$ of a treated stock equals 0 prior to a change in minimum tick size and equals 1 after the change, accounting for the enforcement date differences across stocks; $Event_t^j$ of a control stock in the imposition (conclusion) window equals zero before 10/03/2016 (10/01/2018) and equals 1 as of 10/24/2016 (10/01/2018); u_t is the date fixed effect; and $\varepsilon_{j,t}$ is the error term. Similar to Barardehi et al. (2023), we estimate the treatment effect β by fitting equation (5) using both quantile and OLS regressions, winsorizing QID_t^j at its 1^{st} and 99^{th} percentiles by tick constraint bin and treatment category. All of our estimates control for date fixed effects and double-clustered standard errors at the stock-date level.[20]

Table 2 shows that our findings strongly align with the expected effect of a tick size change on undercutting. The first row of Panels A and B provide the difference-in-difference effect of the TSP on QID for the various groups along with the median/mean value of QID for the control stocks in the sample. Consistent with tick constraints hindering undercutting, the median/mean value of QID increases as spreads get wider with the QID value for tick constrained stocks being very close to zero. Nonetheless, across all groups, and for the TSP imposition and conclusion, the wider tick size is associated with a statistically negative shift in the QID ratio that reverses when tick sizes are returned to 1¢.

Our additional analyses attribute the TSP effects on QID to changes in the quoting behavior, consistent with the impact of a change in tick size on undercutting choices of liquidity providers.

less than 5¢, thus the threshold for tick constrained stocks is 5.5¢ for the conclusion of the TSP.

[20] Due to variation in the dates when the TSP was implemented across TSP stocks, simultaneous inclusion of variable $Event_{j,t}$ and date fixed effects do not lead to perfect co-linearity. The introduction of date fixed effects reflects the fact that for some stocks, the enforcement/lifting dates of TSP restrictions differ from the intended dates by the program. However, in unreported results, we verify robustness to, instead, the use of stock fixed effects or the use of both date and stock fixed effects. The robustness of results across these specifications is consistent with the findings of Rindi and Werner (2019), who also state that their results are virtually unchanged as they vary their fixed effects specifications.

Rows two and three break down the effect of the TSP on the two aspects of the QID ratio. The second row shows the difference-in-difference effect of the TSP on the number of quote improvements divided by total number of quote updates ($Impr$). We find that increased tick size reduces the ratio of quote improvements to quote updates, consistent with reduced undercutting as it becomes more costly jump to the front of the queue. The third row shows that a wider tick size raises the ratio of trade driven quote deteriorations to all quote updates ($DeterTrade$). Existing literature establishes that the widening of tick size during TSP raised trade sizes but left trading volume unchanged (e.g., see Rindi and Werner (2019)), which suggests a reduction in the number of trades. As such, $DeterTrade$'s numerator likely declines as tick size widens, suggesting that the positive effect of a wider tick on $DeterTrade$ reflects reductions on the denominator, i.e., the number of quote updates, that more than offsets the decline in the numerator. These findings reinforce our interpretation that a larger tick size discourages undercutting as reflected in liquidity providers' less aggressive quoting behavior.

We reinforce the link between undercutting costs and QID by exploiting the relevance of *relative* tick sizes for undercutting costs. Following O'Hara, Saar, and Zhong (2019), we focus on stock splits and reverse splits as events that raise and reduce undercutting costs, respectively, by changing relative tick sizes. With a fixed minimum tick size, i.e., 1¢, the share price decline due to a stock split raises relative tick size, while the share price rise due to a reverse split reduces it. For example, to improve the best ask price of \$10 a liquidity provider must quote a round-lot or larger ask order at \$9.99, incurring a relative cost of 1bps. With a 2-for-1 split, the best ask should shift to \$5, leading to a \$4.99 reflecting the next better ask price offered by an undercutting algorithm; and this corresponds to a 2bps relative cost—twice as large as the pre-split cost. As such, we expect undercutting activity, and hence QID, to fall following stock splits and to rise following stock reverse splits. This is exactly what we find.

Our analysis of QID around stock splits also addresses the generalizability of findings using the TSP experiment that focuses on small-cap firms. Specifically, when forming event windows that span 30 days around a stock (reverse) split, we exclude any stock featuring a closing share price of \$5 or less over the even window. We identify 476 split and 27 reverse-split events that fit this criteria. The average and median market-capitalization of stocks with split events are \$8.36 billion and \$2.25 billions, respectively. The average and median market-capitalization of stocks

with reverse-split events are $12.66 billion and $2.51 billion, respectively.

Panel A in Figure 3 shows that average QID drops from over 0.6 to below 0.5 following stock splits; in contrast, it rises from below 0.2 to over 0.3 following stock revers splits. Importantly, this cannot be attributed to the corresponding variation in relative quoted spreads, e.g., it cannot reflect the positive association documented in Figure 2. In fact, Panel B in Figure 3 shows no significant variation in relative quoted spreads around stock split events. This leads us to attribute the observed changes in QID around these events to changes in undercutting costs.

Our collective findings establish the impact of changes in the cost of undercutting on the level of QID, suggesting a strong positive link between QID and undercutting activity. We next relate abnormally low undercutting activity, i.e., high $QIDRes$, to increased informed trading.

5.2 *QIDRes* and Information Arrival

Our next analysis leverages the increased likelihood of informed trading around major instances of information arrival to highlight the correlation between abnormally low undercutting activity and informed trading. Specifically, we focus on earnings announcements (EA), unscheduled corporate events (PR), and news arrivals unassociated with identifiable corporate events (NA).

For each stock, we form twenty-trading-day windows around each information event occuring on day t, with pre-event trading days $t-10$ through $t-1$ and post event trading days t through $t+10$. Whenever available, we use the exact time stamp of the information event to accurately identify the event day t; an event is matched with day t if the event took place after-hours on day $t-1$ or before the close on day t. For earnings announcements, where COMPUSTAT does not provide timestamps, we assume they all arrive after-hours. Moreover, to prevent contamination due to clustering of events, we focus on isolated events that do not follow a similar event in preceding 10 trading days, nor are followed by a similar event in the following 10 trading days.

To set up our analysis, we first explore the behavior of existing measures of informed trading intensity/probability around these events and confirm the findings in the literature. We analyze the behaviors of five different versions of Bogousslavsky et al. (2023)'s ITI measure,[21] as well as three versions of PIN, discussed by Duarte et al. (2020), the $OWRPIN$ measure of Odders-White and

[21] We thank authors of Bogousslavsky et al. (2023) for generously sharing with us 2010-2019 daily ITI measures.

Ready (2008),[22] and MIA measures of Johnson and So (2018).[23] Figure 4 shows that all versions of ITI rise around these instances of information arrival, and that qualitatively similar results obtain using PIN and MIA, even though results vary across different versions of PIN and MIA and for different information events. Overall, these findings are consistent with increased informed trading riskaround instances of material information arrival.

Turning to $QIDRes$ in Figure 5 we document the same pattern. Across all information events we find that $QIDRes$ rises leading up to the event, peaking on the day of the event and reverting afterward. Consistent with adverse-section concerns underlying the abnormally low undercutting activity around information events, we find $QIDRes$ spikes are associated with significantly wider bid-ask spreads (in Panels A, C, and E). This short-term inverse relation between abnormal undercutting activity and spreads, i.e., the positive relation between $QIDRes$ and spreads, obtains despite the positive long-term relation shown in Figure 2—which reflects more ample undercutting opportunities when spreads are wide. Reduced undercutting in the face of widened bid-ask spreads can only reconcile with increased adverse-selection concerns of liquidity providers, suggesting that $QIDRes$ captures informed trading. Further bolstering the idea that these events are associated with significant information we also find spikes in trading volume and abnormal absolute daily return around these events (Panels B, D, and F). Panels A and B present the results for earnings announcements. Panels C and D present the results for unscheduled corporate events, and Panels E and F present the results for other news arrivals. Across all events we observe that these days are associated with a spike in the bid ask spread, abnormal trading volume, and in absolute abnormal return. Importantly, as our next analysis indicates, the behavior of $QIDRes$ appears to be distinct from that of volatility around information events. Figure A.2 shows that the qualitative behavior of $QIDRes$ around information events remains unaffected when we modify our measure to directly control for the the effect of volatility.

We next show that changes in $QIDRes$ predicts imminent upcoming *unscheduled* information arrival events, i.e., PRs and NAs defined earlier. To highlight the incremental predictive power of

[22]Estimates of PIN measures for all NMS stocks up to 2012 are available at Professor Edwin Wu's website.

[23]Estimates of MIA measures for qualifying stock-days up to December, 2018 are available at Professor Travis Johnson's website. Out of 5,940,019 stock-day $QIDRes$ observations in our 2010-2018 sub-sample, we can only match 446,066 stock-days featuring MIA measures. The number of missing observations reflect at least to constraints associated with MIA measures: (1) a common share must be optionable; and (2) to construct MIA for a given stock-day, Johnson and So (2018) require non-zero put and call option volume over the preciding 60 trading days.

$QIDRes$, we control for other observables that, according to Figure 5, exhibit distinct behaviors prior to information arrival days. Specifically, we control for bid-ask spreads, trading volume, and absolute daily returns. Moreover, instead of focusing on isolated events, we control for information event clusters by observing that current information events can predict future information events.

Our analysis estimates the probabilities of unscheduled press releases (PR) and news arrivals (NA) using logistic regressions of these probabilities on past changes in undercutting behavior and a set of control variables, accounting for firm fixed effects. The dependent variable is defined as indicator function $I(z)_t^j$, with $z \in \{\text{PR}, \text{NA}\}$ that equals 1 when event z takes place on day t for stock j and equals 0 otherwise. The set of independent variables contain 5-day changes $\Delta x_{t-1}^j = x_{t-1}^j - x_{t-6}^j$, with $x \in \{QIDRes, qsp, tv, |r|\}$, in abnormal undercutting, quoted bid-ask spread, trading volume, and absolute returns. These variables, as shown in Figure 5 exhibit notable changes in the days leading up to an information event. To control for past relevant information events, additional independent variables are indicator functions $I(Inf)_s^j$ that equal 1 if an earning announcement (EA), an unscheduled press release (PR), *or* a news arrival (NA) event takes place on day s for stock j and equal 0 otherwise, with $s \in \{t-5, \ldots, t-1\}$.

We estimate the probability of event z to occur on day t for stock j using logistic regressions on a year-by-year basis.[24] We fit the models once only using $QIDRes$ and once using $QIDRes$ and all other controls. Tables 3 and 4 show that a 5-day change in $QIDRes$ positively predicts the immediately upcoming unscheduled press release or news arrival. This is consistent with market makers learning from order flow about a an imminent information event (Chae (2005)). For press releases, this finding is robust to controlling for changes in trading and quoting outcomes, that correspond with the change in $QIDRes$, as well as clustering of information events. For news arrivals, the statistical significance is affected by controlling for this these outcomes, which is consistent with our earlier finding that $QIDRes$ spikes are smaller around NAs, relative to those observed around EAs and PRs. Overall, we find that $QIDRes$ possesses significant incremental predictive power for imminent information events relative other liquidity and information variables.

[24] Estimation by year reflects the computational burden when using the over 6 million observations from all years.

5.3 *QIDRes* and Information Content of Trades

We next relate the spikes in $QIDRes$ around information arrivals, discussed in Section 5.2, to the extent of private information contained in the typical trade associated with these spikes. To do so, we first show that the magnitude and persistence of the increase in $QIDRes$ reflect the magnitude of the associated information event. Our tests are motivated by Kim and Verrecchia (1994)'s premise that more informative public news lead to greater post-event information asymmetries. For earnings announcements, we use SUE scores from I/B/E/S to capture the variation in the magnitude of events: in a given quarter, earnings announcement SUE scores in the top or bottom 20 percent—indicating that the announced earnings were significantly higher or lower than analyst consensus—are considered highly informative events. For press releases and news arrivals, we proxy for the information content using post-event realized price movements. For a day-t event, we simply divide each quarterly sample into those events associated with high versus low *absolute* compound post-event 10-day return.[25] Events in the top 40 percent are identified as highly informative events, and those in the bottom 60 percent are the less informative events.

Panels A through C of Figure 6 show that the magnitude of the increase in $QIDRes$ positively correlates with the magnitude of the information event. We first note that there is minimal pre-event variation in $QIDRes$ based on the magnitudes of information events, indicating that any post-event differences in abnormal undercutting may not be attributed to persistent stock characteristics such as volatility. Consistent with abnormally low undercutting activity, i.e., high $QIDRes$, capturing increased informed trading, we find in all cases that the event-day increase in $QIDRes$ is larger for highly informative events than it is for less informative events. Moreover, undercutting activity appears to rebound more quickly toward pre-event levels following less informative events, suggesting that market-making algorithms return to "business as usual" as the risk of trading against informed investors drops. This pattern is remarkably stronger for news arrivals that are classified by Ravenpack as disassociated with any corporate events, suggesting that these events are highly unanticipated by market participants.

We further highlight the link between $QIDRes$ and informed trading risk by decomposing the transaction cost associated with each trade, as captured by effective spread, into permanent and

[25] Qualitative findings are robust to excluding event days from these return calculations

temporary price impact components. This decomposition reflects the idea that the cost of consuming liquidity for incoming marketable order flow consists two components: (1) the compensation that liquidity providers demand for exposure to adverse-selection risk, captured by price impact and reflective of potential information advantages of liquidity consumers; and (2) the compensation that liquidity providers demand in return for facilitating "immediacy", captured by realized spreads that is generally attributed to operational costs incurred and revenues collected by market makers (see, e.g., Hendershott, Jones, and Menkveld (2011)). If the abnormally low undercutting documented in Figure 5 is due to informed trading, then any corresponding variation in effective spread should be primarily attributable to the price impact (adverse selection) component. Panels D, E, and F of Figure 6 show exactly this. Around the news events realized spreads are effectively unchanged and the entire observed increase in the effective spread is explained by an increase in the adverse selection component of the effective spread.

5.4 *QIDRes* and Direct Sources of Informed Trade

In this section, we address an alternative explanation for the association between abnormally low undercutting, i.e., high $QIDRes$, and the arrivals of information events. Specifically, we provide evidence that $QIDRes$ is unlikely to only capture increased 'sniping risk' around information events. Budish et al. (2015) show that in continuous-time limit order markets high-frequency traders engage in an arms race over the speed with which they can place/cancel orders. A key result in this literature is that differences in order processing speeds across traders lead limit orders of 'slower' traders to become stale for very short periods of time as the prices move against these resting orders upon arrivals of public information. These stale orders are then picked off, i.e. sniped, by 'faster' traders, leading to losses to slow traders. This phenomenon poses an adverse selection risk that is unrelated to information asymmetry about the fundamental value of the asset, but rather the speed with which different traders can respond to the arrivals of public information.[26] Relevant for our analysis is the possibility that information events that we study purely reflect increased 'sniping risk', as opposed to increased information asymmetry regarding fundamental value, leading to a reduction in the willingness of liquidity providers to undercut.

[26] Menkveld and Zoican (2017) extent these insights by showing that exogenous increased in order processing speed offered by exchanges may exacerbate this issue and harm liquidity provision.

To address this concern, we use more direct measures of informed trading, as opposed to solely relying on variations around information events, to provide cross-sectional evidence that links increased informed trading risk to high $QIDRes$.[27] We first show that $QIDRes$ is higher when short sellers more activity take (accumulate) or leave (cover) short positions. The literature has provided robust evidence that short-seller trades are informed (see, e.g., Desai, Ramesh, Thiagarajan, and Balachandran (2002); Engelberg, Reed, and Ringgenberg (2012); Boehmer and Wu (2013), among others), so we expect to observe higher $QIDRes$ for stocks with high short selling activity.

We match each stock's bi-weekly percentage change in short interest to the corresponding averages of various informed trading risk measures, including $QIDRes$. We then sort each bi-weekly cross-section into ten portfolios (deciles) of signed percentage change in short interest, with the bottom decile containing stocks with largest coverings of short interest and the top portfolio containing stocks with largest 10% of short interest accumulations. We then calculate portfolio-level average informed trading risk measures in each bi-weekly period.[28] We finally plot the time-series means of these averages against change-in-short-interest portfolios.

Figure 7 shows that most measures of informed trading risk follow ∪-shaped patterns as we go from portfolio of stocks with largest coverings of short interest (declie 1) to stocks with largest accumulations of short interest (decile 10). This is consistent with private information underlying both buying and selling activity by short sellers and confirms Bogousslavsky et al. (2023)'s findings that relate ITIs o short interest. However, consistent with short sellers main focus on investigating negative information about asset values, most informed trading risk measures are highest when short interest accumulations are largest. Panel A shows that all versions of ITI display these patterns; whereas Panel B and C show that even though PIN, $DYPIN$, $GPIN$, and MIA follow similar patters, $OWRPIN$ exhibits a ∩-shaped pattern. Panels D and E in Figure 7 document relationships between $QIDRes$ and short-seller activity conditioning on the past levels of short interest and firm size, respectively. Our findings suggest that (1) increased $QIDRes$ in times of high short-seller activity is more pronounced for stocks with higher levels of short interest, indicative of a higher likelihood that order flow contains orders from informed short sellers; and (2) the link

[27] Nonetheless, Appendix A.3 shows that a modified version of our measure $QIDResV$, which directly controls for the volatility of 1-minute quote midpoint returns exhibit patterns around information events that are qualitatively similar to those of $QIDRes$. This evidence suggests that pure sniping risk does drive the variation in $QIDRes$.

[28] To ensure that our findings doe not pick up any temporal variation in liquidity provision activities, for $QIDRes$, we first adjust each bi-weekly stock-specific average relative to the corresponding market-wide mean $QIDRes$.

between $QIDRes$ and the information content of short selling is not a small-stock phenomenon. Importantly, all these qualitative findings extend if we conservatively exclude biweekly periods that overlap with at least an EA, PR, *or* NA,[29] reinforcing the conclusion that informed trading risk identified by $QIDRes$ is likely distinct from increased sniping risk associated with public information arrival.

Second, we show that most measures indicate increased information asymmetry around a subset of informed mutual-fund trades. Barardehi et al. (2022) use ANcerno to identify industry-neutral self-financed trades of mutual funds, denoted INSFIT, and establish these trades are informed. We estimate the average incremental difference between informed trading risk measures around INSFIT days and non-INSFIT days, controlling for firm and date fixed effects.[30] We form 1-, 3-, and 5-day windows around stocks-days representing an INSFIT trade, examining INSFIT-bought and INSFIT-sold stocks separately. We then compare informed trading risk measures observed inside versus outside these windows.

Table 5 shows that stock-days featuring informed institutional trades are associated with statistically higher average informed trading risk measures. Specifically, with the exception of $ITI_{insider}$, $GPIN$, and $OWRPIN$, results based on all measures are consistent with increased informed trading risk on stock-days surrounding with INSFIT buy or INSFIT sell trades. Further highlighting the relevance of the information content of INSFIT trades, we find the largest differences on the "day of", i.e., 1-day INSFIT trade windows. Widening these windows to 3-day and 5-day horizons around the underlying INSFIT trades lead to smaller estimated differences that become statistically insignificant for some existing measures.

In sum, we find a positive link between more direct, established sources of informed trading and various measures of informed trading risk used in our analysis. Our finding suggests that $QIDRes$ captures variation in the extent of information asymmetry, rather than solely that in sniping risk.

[29] Such biweekly periods account for nearly half of the stock-days in our sample.

[30] We thank authors of Barardehi et al. (2022) for permitting us to use a sample of daily indicators that identify stocks bought and sold through INSFIT. This sample spans January 1999 through September 2011, leaving us with the overlap period of January 2010 through September 2011 for our analysis.

5.5 *QIDRes* and Compensation for Liquidity Provision

We next show that spikes in $QIDRes$ are hard to reconcile with inventory management concerns of liquidity providers driven by capital constraints. Comerton-Forde et al. (2010) show that liquidity providers with capital constraints become reluctant to accumulate additional inventory when their inventories are unbalanced; and So and Wang (2014) show that expected returns from liquidity provision significantly rise prior to earnings announcements reflecting increased inventory risk. Thus, a potential explanation for reductions in undercutting, i.e., $QIDRes$ spikes, may reflect inflated market maker inventories driven by increased liquidity demand that leads capital constraints to bind. Compensation for such liquidity provision is often reflected by short-term price pressure that is followed by price reversals (see, e.g., Campbell, Grossman, and Wang (1993); Hendershott and Menkveld (2014)). Thus, if inventory management concerns underlie the spikes in $QIDRes$, i.e., abnormally low undercutting, we should observe greater price reversals following high-$QIDRes$ days. We find the exact opposite.

Trading days with higher $QIDRes$ are followed by weaker price reversals. On each day t we sort stocks into quintiles of $QIDRes$. We then regress the cumulative returns from the close of day t through the close of day $t+n$, with $n \in \{1, \ldots, 10\}$, on day t returns, controlling for date and stock fixed effects. A negative slope coefficient indicates price reversal with the magnitude of this slope coefficient indicating the magnitude of this reversal. Panel A in Table 6 shows that the high $QIDRes$ portfolio, containing stock-days with abnormally low undercutting activity, have coefficients significantly closer to zero than the low $QIDRes$ portfolio. For all future return horizons, n, reversals grow nearly monotonically weaker, with the absolute values of slope coefficients shrinking by half, as we move from the low $QIDRes$ tercile to its high tercile. Hence, inventory management concerns of liquidity providers cannot drive the variation in $QIDRes$. In contrast, and consistent with our earlier findings, weaker price reversals that follow days with higher $QIDRes$ further reinforces that $QIDRes$ picks up informed trading. This finding is also consistent with Bogousslavsky et al. (2023) who find that trading days with higher informed trading intensity (ITI) are followed by weaker price reversals.

Panel B in Table 6 documents the extent of price reversals conditional on both $QIDRes$ and realized volatility of 1-minute returns based on midpoint prices, $qvol$. This analysis addresses the

possible link between volatility and undercutting activity, reflecting reduced liquidity provision, and hence undercutting, when volatility is high. We sort each cross-section *independently* into terciles of $QIDRes$ and realized volatility, before estimating the extent of price reversals conditional on both. First, we observe that roughly equal number of observations fall in the nine $QIDRes$-volatility categories, indicating a near-zero correlation between abnormal undercutting and realized volatility—in fact, the correlation coefficient in the full sample is -0.0011 (see Panel B in Table 1). Second, the finding that highest $QIDRes$ tercile is associated with weakest subsequent reversals extends across different levels of realized volatility.

5.6 Intraday Analysis of *QIDRes*

In this section, we analyze the relationship between $QIDRes$ and informed trading risk by examining this link at different times of the trading day. Our analysis is motivated by the premise that information asymmetry, and the liquidity providers' risk of trading with informed investors, declines over the course of the trading day (see, e.g., Madhavan, Richardson, and Roomans (1997)).[31] We construct three "intraday" versions of $QIDRes$ that reflect undercutting activity at three time-of-day segments of the trading day. First, we inspect the correlation between $QIDRes_{jt}$ and each of these intraday versions. Second, we compare the behaviors of intraday $QIDRes$ measures around information events.

To construct intraday $QIDRes$, we divide each trading day into three segments: 9:45am–11:45am (morning, am), 11:45am–1:45pm (mid-day, md), and 1:45pm–3:45pm (afternoon, pm), which allows us to construct the three respective intraday undercutting activity measures $QID(\tau)^q_{jt}$, with $\tau \in \{am, md, pm\}$. Quarter $q-1$ quantities of these intraday undercutting activity measures are then entered, in turn, on the left hand side of equation (3).[32] The resulting parameter estimates as well as standard deviations of intraday QID measures enter equation (4) to produce $QIDRes(am)^q_{jt}$, $QIDRes(md)^q_{jt}$, and $QIDRes(pm)^q_{jt}$. This process decomposes $QIDRes$ on each stock day into its intraday components.

If $QIDRes$ captures informed trading risk and if such risk is higher in earlier trading hours of the

[31] Also see Admati and Pfleiderer (1988) and Wood, McInish, and Ord (1985), among others.

[32] We use the same right-hand-side variable in equation (3) when constructing different intraday versions of QID. This allows us to attribute any differences in the resulting $QIDRes$ measures to time-of-day effects in undercutting rather than those in quoted spreads.

trading day then we expect our baseline $QIDRes^q_{jt}$ to be more strongly correlated with its morning component, $QIDRes(am)^q_{jt}$, than with the other two components. We find strong evidence of this. Figure 8 exhibits empirical distributions of R^2 statistics obtained from regressing $QIDRes$ on each of its intraday components. These estimates are carried out at the stock-quarter level, capturing the association between $QIDRes$ and the intraday component *only* using time-series variations. Consistent with a declining informed trading risk over the course of the trading day, the association between $QIDRes$ and $QIDRes(am)$ is strongest and that between $QIDRes$ and $QIDRes(pm)$ is the weakest. Importantly, the clearly distinguishable locations of R^2 empirical distributions given different τ's is evidence of statistical dominance, which strongly speaks to the statistical and economic significance of our findings. More concretely, the mean (median) stock-quarter-specific R^2's are 65.8% (69.5%), 57.2% (59.9%), and 47.1% (47.7%) when variation in $QIDRes$ is examined against that in the underlying component from morning, mid-day, and evening, respectively. In sum, a much larger portion of the variation in $QIDRes$ is attributable to abnormal undercutting activity in earlier trading hours rather than later windows.

We provide additional evidence using the intraday variation in the intensity of informed trading by examining $QID(\tau)^q_{jt}$'s behavior around unscheduled press releases.[33] With higher intensity of informed trading earlier in the day, we expect $QIDRes(am)^q_{jt}$ to display greater spikes around information events than do other intraday versions of $QIDRed$. Figure 9 documents exatly this.

5.7 Asset Pricing Implications of *QIDRes*

The literature has documented that informed trading risk measures predict stock returns: higher past informed trading probability/intensity is associated with higher expected returns. However, there is no theoretical or empirical consensus regarding what drives this return predictability. For example, Easley and O'Hara (2004) argue that informed trading should be priced since the risk driven by information asymmetry is non-diversifiable; hence, investors holding a stock with more private information, and hence informed trading, demand a premium as compensation for this exposure. Consistent with this prediction, Easley et al. (2002) show that PIN is priced in the cross-section.[34] Duarte and Young (2009) propose an alternative explanation for return

[33] Qualitative similar conclusions obtain around earnings announcements and other news arrivals.

[34] Also see, e.g., Kelly and Ljungqvist (2012) and Derrien and Kecskés (2013). In contrast, Lambert et al. (2012) argue that in a perfectly competitive market, information asymmetry risk is diversifiable and hence should not be

predictability of informed trading intensity/probability measures by showing that PIN's cross-sectional return predictability primarily reflects liquidity premia. They argue that since informed trading intensity is correlated with liquidity, PIN's return predictability conflates the effects of information asymmetry with those of priced illiquidity (Amihud and Mendelson (1980)). Following this literature, we also show that $QIDRes$ predicts stock returns. However, we attribute this return predictability to limits to arbitrage, reflecting the unique features of $QIDRes$.

In contrast to prior measures of informed trading, we do not expect any return predictability demonstrated by $QIDRes$ to be associated with compensation for bearing the risk associated with a stock characteristic or the premium demanded to hold less liquid stocks. In fact, we provide strong evidence that $QIDRes$ fits neither of these notions. Table 7 presents the correlations between $QIDRes$, ITI and PIN based information trading measures as well as common liquidity measures: quoted spread, effective spread, lambda, Amihud, and ILM. Panel A presents the correlations for 2010-2019 (omitting PIN measures where we only have data for 2010-2012) and Panel B presents all measures for the 2010-2012 period. This table shows virtually zero cross-sectional correlation between monthly averages of $QIDRes$ and various measures of liquidity, and only minimal correlation with other measures of informed trading. Panel A in Table A.1 presents evidence that $QIDRes$ is very weakly correlated with a host of stock characteristics. Finally, in Panel B of Table A.1, we document evidence of slight mean-reversion in $QIDRes$, indicating that is does not constitute a persistent stock characteristic.

Importnatly, the lack of correlation with liquidity is not true for other measures of informed trading where different versions of ITI and PIN appear to be positively related to liquidity. For example, Panel A shows that the average of the absolute correlation coefficients obtained between different versions of ITI and various stock illiquidity measures is about 0.15, with the highest pairwise absolute correlation of 0.37. Similarly, the average absolute correlation between different versions of PIN and stock illiquidity measures is around 0.15, with a high pairwise absolute correlation of 0.26. These collective facts clearly distinguish $QIDRes$ from existing measures, strongly suggesting that it cannot predict returns in the context of existing theories on return predictability of informed trading risk. We next investigate whether $QIDRes$ predicts returns.

priced, with Armstrong et al. (2011) providing empirical evidence supportive of this prediction.

We begin this analysis using simple portfolio sorts.[35] Table 8 shows that stocks with higher $QIDRes$ feature higher expected returns. For example, we find that average three-factor risk-adjusted monthly return of the portfolio of stocks with the the highest past levels of informed trading, i.e., stocks falling in the top $QIDRes$ quintile in quarter $q-1$, is 30bps higher than that for the portfolio containing stocks with the lowest levels of informed trading, i.e., stocks falling in the bottom $QIDRes$ quintile in quarter $q-1$. These quantitative findings extend when we form test portfolios using $QIDRes$ in quarter $q-2$. Bogousslavsky et al. (2023) document next-month return predictability using ITIs; hence, complementary to their results, our finding that $QIDRes$ predicts monthly returns two quarters forward indicates that $QIDRes$ can predict future returns over longer horizons.

We next fit cross-sectional regressions to examine return predictability of $QIDRes$ while controling for key stock characteristics. Our regression analysis estimates

$$RetRf_{j,q,m} = \gamma^0 + \gamma^1 \left(QIDRes_{j,q-1}\right) + \gamma^2 \left(QIDRes_{j,q-2}\right) + \Lambda^\top \text{Control}_{j,q,m-1} + u_{j,q,m}, \quad (6)$$

where $RetRf_{j,q,m}$ is stock j's return in month m of quarter q in excess of the corresponding 1-month T-Bill rate; $QIDRes_{j,q-1}$ and $QIDRes_{j,q-2}$ denotes abnormal undercutting activity in quarters $q-1$ and $q-2$, respectively, for stock j; $\text{Control}_{j,q,m-1}$ denotes the vector of controls including betas from the three-factor Fama-French model, book-to-market ratio, market capitalization, dividend yield, idiosyncratic volatility, previous month's return, the return from the prior 11 months, previous quarter's share of institutionally held shares, previous quarter's institutional ownership concentration, and share turnover in month $m-2$.

Table 9 summarizes our findings when we fit fixed-effect panel regressions based on equation (6): we find a statistically significant positive association between $QIDRes$ and expected stock returns. This finding is robust to (1) including year-month fixed effects only versus including both year-month and firm fixed effects, which we choose as our main specification; (2) to including institutional ownership concentration and share turnover, reflecting the extent of competition for liquidity be-

[35] We work with a sample spanning January 2010 through August 2016, reflecting the significant impacts of TSP on the level of undercutting for a large group of stocks (see Section 5.1). These empirical choices allow us to examine the entire cross-section of NMS stocks with no TSP-driven gaps in the time-series of each stock. Unreported analysis insures that qualitative findings are robust to, instead, excluding TSP stocks between September 2016 through December 2018 when TSP was in effect, and using the remaining data in the 2010-2019 time period.

tween potentially informed investors (Lambert et al. (2012)); and (3) augmenting the set of controls with individual or all the five stock illiquidity measures, reflecting the main message of Duarte and Young (2009) as a general concern that may apply to any measure of informed trading.

Table 10 formally contrasts the abilities of different informed trading intensity/probability measures in explaining the cross-section of expected returns. We estimate horse race regressions based on modified specifications of equation (6) that include $QIDRes$ *and* different sets of alternative existing measures as independent variables subject to their availability. We find that the association between $QIDRes$ and expected returns remains in these regressions, and that most of the alternative measures do not load with a statistically significant coefficients. Notably, $QIDRes$ is *the* only measure that significantly predicts future returns in all specifications. We also note that ITIs are not completely backward-looking measures of informed trading risk as Bogousslavsky et al. (2023) train their machine learning algorithms using sub-sample of stock-days that are scatted over the entire time-series, and hence, ITIs from quarters $q-1$ and $q-2$ may, by construction, contain information about future returns. In sharp contrast, average $QIDRes$ from quarters $q-1$ and $q-2$ are not conditional on any future trading or pricing outcome.

As discussed earlier, we may interpret the robust return predictability of $QIDRes$ neither in the context of Easley and O'Hara (2004)'s "stock characteristic" story, nor in the context of Duarte and Young (2009)'s "illiquidity premia" story. This leads us to attribute the return predictability of $QIDRes$ to limits to arbitrage. Specifically, $QIDRes$ does not differentiate between positive and negative information, so if informed traders acting on positive and negative information is equally likely, then we would not expect $QIDRes$ to have any association with future returns. However, reflecting the well-documented selling constraints (e.g. Saffi and Sigurdsson (2011), and Dixon (2021)), it must be more difficult for investors to trade on negative information. As a result, high $QIDRes$ is more likely to capture informed trading motivated by positive, rather than negative, signals; and thus should positively predict returns.[36] Specifically, stocks with higher $QIDRes$ in a given quarter (1) experienced more information events than is normal in those quarters, and (2) due to short selling constraints, these information events were, on average, positive.

We conclude by showing that return predictability of $QIDRes$ is concentrated among stocks with tighter short sale constraints. We do so by splitting the sample based on observed equilibrium

[36] See Bogousslavsky et al. (2023) for a similar discussion.

lending fees in the securities lending markets. We examine $QISRes$'s return predictability conditional on the level of lending fees, with higher such fees reflecting tighter short sale constraints. From FIS data, we calculate average lending fee of each stock in quarter $q-3$, and then sort monthly cross-section in the current quarter into terciles of this average security lending fee. Table 11 shows that $QIDRes$ predicts expected returns more strongly among stocks with high lending fees.

6 Conclusion

Despite the key importance of informed trading for different areas of financial economics, easy to implement empirical measures of informed trading have proven difficult to derive. In this paper, we propose an easy to compute and intuitive measure of informed trading risk which we refer to as $QIDRes$. Our measure only requires trades and quotes data and thus can be computed for almost all publicly traded stocks at the daily, or even finer, frequencies in any modern limit order market.

Our approach exploits the intuition that liquidity providers compete less to fill order flow if they perceive the incoming marketable orders to be informed. Specifically, a liquidity provider's appetite to "undercut" rivals should significantly drop when they expect arrivals of informed marketable orders. We argue that abnormally low undercutting activity reveals the concerns of liquidity providers about incoming informed orders and hence indirectly measures informed trading risk.

We contrast $QIDRes$ with existing measures of informed trading intensity/probability whose constructions are computationally demanding, require proprietary data, or are applicable to only a subset of stock-days. We find that $QIDRes$ performs as well as or better than these alternative measures: (1) $QIDRes$ spikes around periods known to be associated with informed trading such as earnings announcements, unscheduled press releases, and news arrivals; (2) increases in $QIDRes$ predict imminent unscheduled information arrival events; (3) the magnitudes of the $QIDRes$ spikes are positively associated with the magnitudes of imminent information events; (4) stock prices reverse less following days when $QIDRes$ indicates higher informed trading risk; (5) episodes of increased short selling activity are associated with higher $QIDRes$; and (6) stock-days with known informed mutual-fund trades exhibit higher $QIDRes$.

We also show that $QIDRes$ from the preceding two quarters predicts monthly stocks returns. However, $QIDRes$ is orthogonal to persistent stock characteristics, especially liquidity, indicating

that its return predictability is distinct from liquidity premia as posited by Duarte and Young (2009) about PIN. Moreover, consistent with the notion that informed trading should not be predictable, $QIDRes$ does not constitute a persistent stock characteristic either. Hence, we attribute its return predictability to the asymmetry in limits to arbitrage that restrict trading based on negative information. In fact, return predictability of $QIDRes$ is concentrated among stocks with tightest short sale constraints.

References

Admati, A. and P. Pfleiderer (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies 1*, 3–40.

Ahren, K. R. (2020). Do proxies for informed trading measure informed trading? evidence from illegal insider trades. *Review of Asset Pricing Studies 10*, 397–440.

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Amihud, Y. and H. Mendelson (1980). Market-making with inventory. *Journal of Financial Economics 8*, 31–53.

Anand, A., P. Irvine, A. Puckett, and K. Venkataraman (2012). Performance of institutional trading desks: An analysis of persistence in trading costs. *The Review of Financial Studies 25*(2), 557–598.

Ang, A., R. Hodrick, Y. Zhing, and x. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*, 259–299.

Armstrong, C., D. Taylor, J. Core, and R. Verrecchia (2011). When does information asymmetry affect the cost of capital? *Journal of Accounting and Economics 49*, 1–40.

Barardehi, Y. H., D. Bernhardt, Z. Da, and M. Warachka (2023). Uncovering the liquidity premium in stock returns using retail liquidity provision. Working Paper.

Barardehi, Y. H., D. Bernhardt, T. G. Ruchti, and M. Weidemier (2021). The night and day of Amihud's (2002) liquidity measure. *Review of Asset Pricing Studies 11*, 269–308.

Barardehi, Y. H., Z. Da, and M. Warachka (2022). The information in industry-neutral self-financed trades. *Journal of Financial and Quantitative Analysis*. Forthcoming.

Barardehi, Y. H., P. Dixon, Q. Liu, and A. Lohr (2023). When does the tick size help or harm market quality? evidence from the tick size pilot. Working Paper.

Boehmer, E., Z. R. Huszar, and B. D. Jordan (2010). The good news in short interest. *Journal of Financial Economics 96*(1), 80–97.

Boehmer, E. and J. Wu (2013). Short selling and the price discovery process. *Review of Financial Studies 26*, 287–322.

Bogousslavsky, V., V. Fos, and D. Muravyev (2023). Informed trading intensity. *Journal of Finance*. Forthcoming.

Budish, E., P. Cramton, and J. Shim (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics 130*(4), 1547–1621.

Campbell, J. Y., S. J. Grossman, and J. Wang (1993). Trading volume and serial correlation in stock returns. *Quarterly Journal of Economics 108*, 905–939.

Chae, J. (2005). Trading volume, information asymmetry, and timing information. *Journal of Finance 60*(1), 413–442.

Chordia, T. and B. Miao (2020). Market efficiency in real time: Evidence from low latency activity around earnings announcements. *Journal of Accounting and Economics 70*, 1013–35.

Comerton-Forde, C., T. Hendershott, C. Jones, P. Moulton, and M. Seasholes (2010). Time variation in liquidity: The role of market maker inventories and revenues. *Journal of Finance 65*(1), 295–331.

Conrad, J. and S. Wahal (2020). The term structure of liquidity provision. *Journal of Financial Economics 136*(1), 239–259.

Derrien, F. and A. Kecskés (2013). The real effects of financial shocks: Evidence from exogenous changes in analyst coverage. *Journal of Finance 68*(4), 1407–1440.

Desai, H., K. Ramesh, S. Thiagarajan, and B. Balachandran (2002). An investigation of the informational role of short interest in the NASDAQ market. *Journal of Finance 57*, 2263–2287.

Dixon, P. N. (2021). Why do short selling bans increase adverse selection and decrease price efficiency? *The review of asset pricing studies 11*(1), 122–168.

Dixon, P. N., A. F. Corbin, and E. K. Kelley (2021). To own or not to own: Stock loans around dividend payments. *Journal of Financial Economics 140*, 539–559.

Dixon, P. N. and E. K. Kelley (2022). Business cycle variation in short selling strategies: Picking during expansions and timing during recessions. *Journal of Financial and Quantitative Analysis 57*(8), 3018–3047.

Duarte, J., E. Hu, and L. Young (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics 135*(3), 795–815.

Duarte, J. and L. Young (2009). Why is pin priced? *Journal of Financial Economics 91*(2), 119–138.

Easley, D., S. Hvidkjaer, and M. O'Hara (2002). Is information risk a determinant of asset returns? *Journal of Finance 57*(5), 2185–2221.

Easley, D. and M. O'Hara (2004). Information and the cost of capital. *Journal of Finance 59*(4), 1553–1583.

Engelberg, J., A. Reed, and M. Ringgenberg (2012). How are shorts informed? Short sellers, news, and information processing. *Journal of Financial Economics 105*, 260–278.

Foley, S., A. Dyhrberg, and J. Svec (2022). When bigger is better: the impact of a tiny tick size on undercutting behavior. *Journal of Financial and Quantitative Analysis*.

Foley, S., T. Meling, and B. A. Ødegaard (2021). Tick size wars: The market quality effects of pricing grid competition. *Available at SSRN 2866943*.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics 14*(1), 71–100.

Griffith, T. G. and B. S. Roseman (2019). Making cents of tick sizes: The effect of the 2016 us sec tick size pilot on limit order book liquidity. *Journal of Banking & Finance 101*, 104–121.

Hasbrouck, J. and G. Saar (2013). Low-latency trading. *Journal of Financial Markets 16*, 646–679.

Heath, D., M. C. Ringgenberg, M. Samadi, and I. M. Werner (2020). Reusing natural experiments.

Hendershott, T., C. Jones, and A. Menkveld (2011). "does algorithmic trading improve liquidity? *Journal of Finance 66*(1), 1–66.

Hendershott, T. and A. J. Menkveld (2014). Price pressures. *Journal of Financial Economics 114*, 405–423.

Holden, C. W. and S. E. Jacobsen (2014). Liquidity measurement problems in fast, competitive markets: Expensive and cheap solutions. *Journal of Finance 69*, 1747–85.

Hughes, J., J. Liu, and J. Liu (2007). Diversification and the cost of capital. *The Accounting Review 82*, 705–729.

Johnson, T. L. and E. C. So (2018). A simple multimarket measure of information asymmetry. *Management Science 64*(3), 1055–1080.

Kelly, B. and A. Ljungqvist (2012). Testing asymmetric-information asset pricing models. *The Review of Financial Studies 25*(5), 1366–1413.

Kim, O. and R. E. Verrecchia (1994). Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics 17*, 41–67.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, 1315–1335.

Lambert, R., C. Leuz, and R. Verrecchia (2012). Information asymmetry, information precision, and the cost of capital. *The Review of Finance 16*(1), 1–29.

Li, S. and M. Ye (2023). Discrete price, discrete quantity, and the optimal nominal price of a stock. Working Paper.

Madhavan, A., M. Richardson, and M. Roomans (1997). Why do security prices change? a transaction-level analysis of nyse stocks. *The Review of Financial Studies 10*(4), 1035–1064.

Menkveld, A. J. (2013). High frequency trading and the new market makers. *Journal of financial Markets 16*(4), 712–740.

Menkveld, A. J. and M. A. Zoican (2017). Need for speed? exchange latency and liquidity. *Review of Financial Studies 30*, 188–1228.

Odders-White, E. R. and M. J. Ready (2008). The probability and magnitude of information events. *Journal of Financial Economics 87*(1), 227–248.

O'Hara, M., G. Saar, and Z. Zhong (2019). Relative tick size and the trading environment. *The Review of Asset Pricing Studies 9*(1), 47–90.

Petacchi, R. (2015). Information asymmetry and capital structure: Evidence from regulation fd. *Journal of Accounting and Economics 59*, 143–162.

Rindi, B. and I. M. Werner (2019). Us tick size pilot. Working Paper.

Saffi, P. A. and K. Sigurdsson (2011). Price efficiency and short selling. *The Review of Financial Studies 24*(3), 821–852.

So, E. and W. S. Wang (2014). News-driven return reversals: Liquidity provision ahead of earnings announcements. *Journal Financial Economics 114*, 20–35.

Wang, J. (1993). A model of intertemporal asset pricing under asymmetric information. *The Review of Economic Studies 60*, 249–282.

Werner, I. M., B. Rindi, S. Buti, and Y. Wen (2022). Tick size, trading strategies, and market quality. *Management Science*. Forthcoming.

Wood, R., T. McInish, and J. K. Ord (1985). An investigation of transactions data nyse stocks. *Journal of Finance 40*, 723–739.

Figures and Tables

Figure 2. Undercutting and Quoted Spreads.
The figure presents the relationship between undercutting activity, as measured by QID, and percent quoted bid-ask spread. for each stock, both QID and the natural log of time-weighted percent quoted bid-ask spread, constructed at the stock-day frequency, are averaged across all days in the sample. The scatter plot presents the correlation between these two averages across stocks. The sample includes stock-days of NMS-listed common shares between Jan 01, 2010 through Dec 31, 2019 with previous months' closing prices of at least $5, excluding stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment.

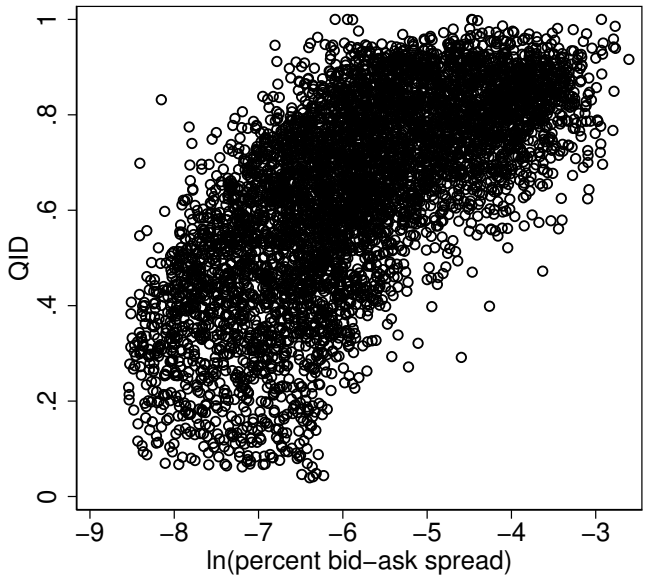
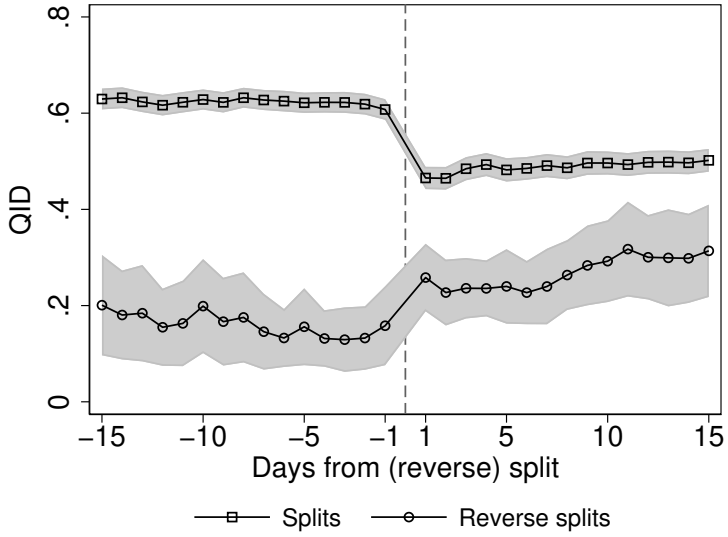


Figure 3. Relative Tick Size and Undercutting activity: Stock Splits and Reverse Splits. The figure presents average QID around stock splits. Stock split and revere-split dates are obtained from CRSP, with event windows covering 15 days prior to a split date and 15 day as of the split date. Averages and 95% confidence intervals of QID (Panel A) and relative quoted spread (Panel B), both winsorized at the 1^{st} and 99^{th} percentiles of each day if the main sample, are plotted against days from the event. The sample includes stock-days of NMS-listed common shares between Jan 01, 2010 through Dec 31, 2019 that coincide with stock-split event windows. Included stocks must minimum a daily closing price of \$5 and must feature non-missing observations over the event window. Stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment are excluded.

Panel A: Undercutting activity measure



Panel B: Relative quoted spread

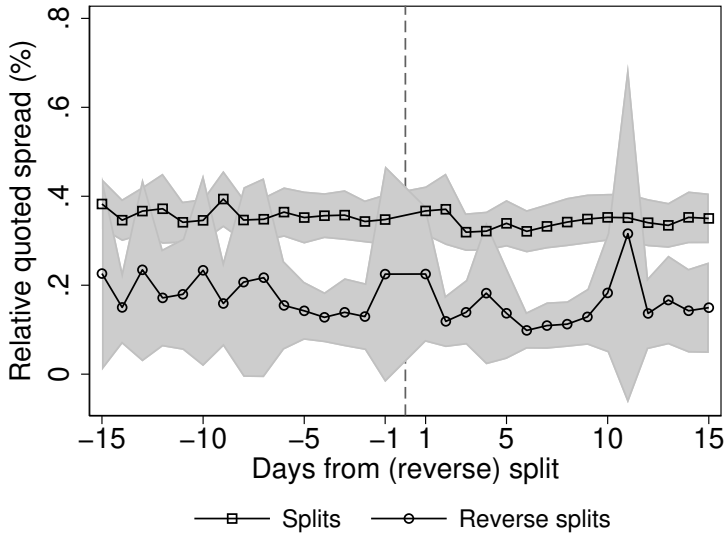
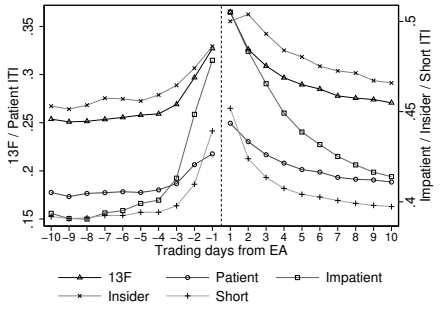
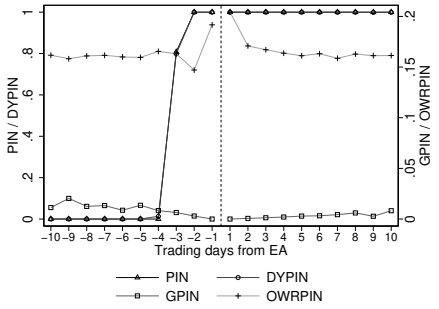


Figure 4. Existing Measures of Informed Trading around Unscheduled Corporate Announcements.
The figure presents medians of ITI, PIN, and MIA around around earnings announcements (EA), unscheduled press releases (PR), and news arrivals not associated with any identified event (NA). Five versions of ITI and four PIN are considered. The sample includes all NMS-listed common stocks with previous quarter-end's share prices of at least \$5. Sample period is Jan, 2010 through Dec, 2019 for ITI; Jan, 2010 through Dec, 2012 for PIN; and Jan, 2010 through Dec, 2018 for MIA. Stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment are excluded. Earnings announcement dates are obtained from COMPUSTAT; unscheduled press release dates and news arrivals not associated with any identified event are obtained from Ravenpack.

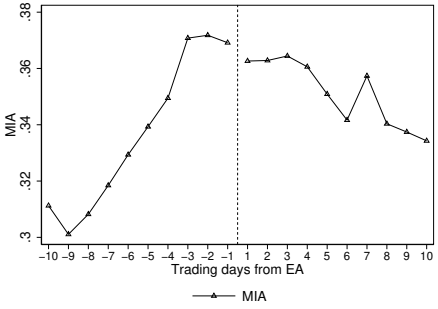
Panel A: EA, Informed Trading Intensity



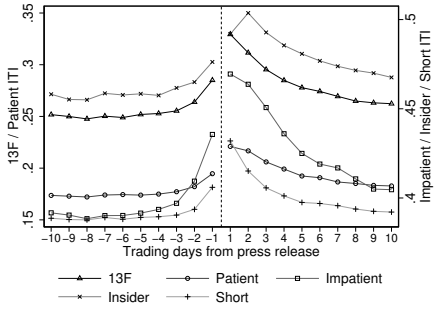
Panel B: EA, Prob. of Informed Trading



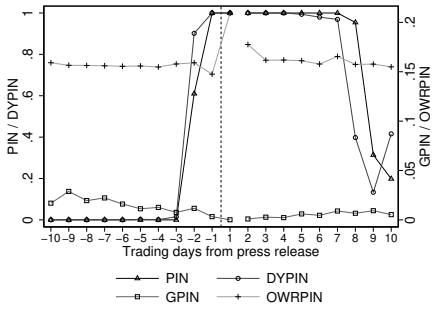
Panel C: EA, MIA



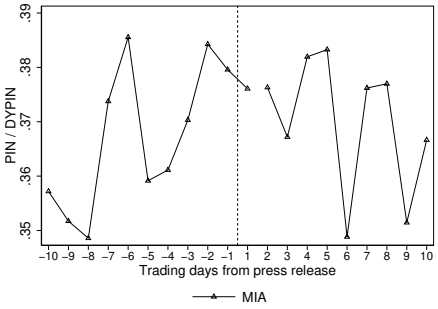
Panel D: PR, Informed Trading Intensity



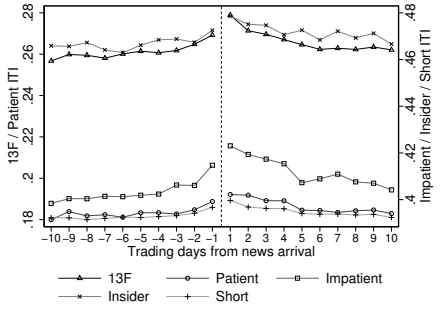
Panel E: PR, Prob. of Informed Trading



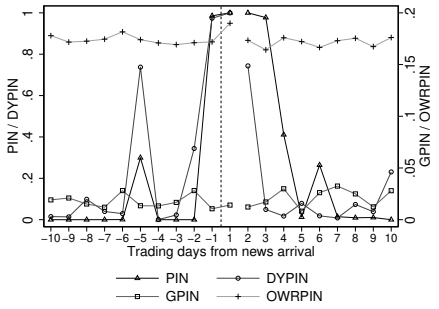
Panel F: PR, MIA



Panel G: NA, Informed Trading Intensity



Panel H: NA, Prob. of Informed Trading



Panel I: NA, MIA

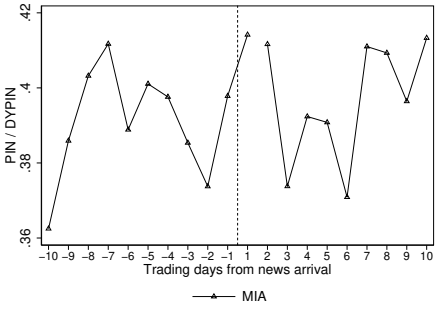


Figure 5. Undercutting Activity, Liquidity, and Information Asymmetry around Scheduled and Unscheduled Corporate Announcements.
The figure presents abnormal undercutting activity, dollar bid-ask spread, abnormal trading volume, and abnormal daily absolute return around earnings announcements (EA), unscheduled press releases (PR), and news arrivals not associated with any identified event (NA). Daily abnormal undercutting values are calculated based on equation (4). Daily trading volume and absolute returns of each stock are normalized relative to the stock-specific median of each respective variable from the previous calendar quarter. The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least \$5. Earnings announcement dates are obtained from COMPUSTAT; unscheduled press release dates and news arrivals not associated with any identified event are obtained from Ravenpack.

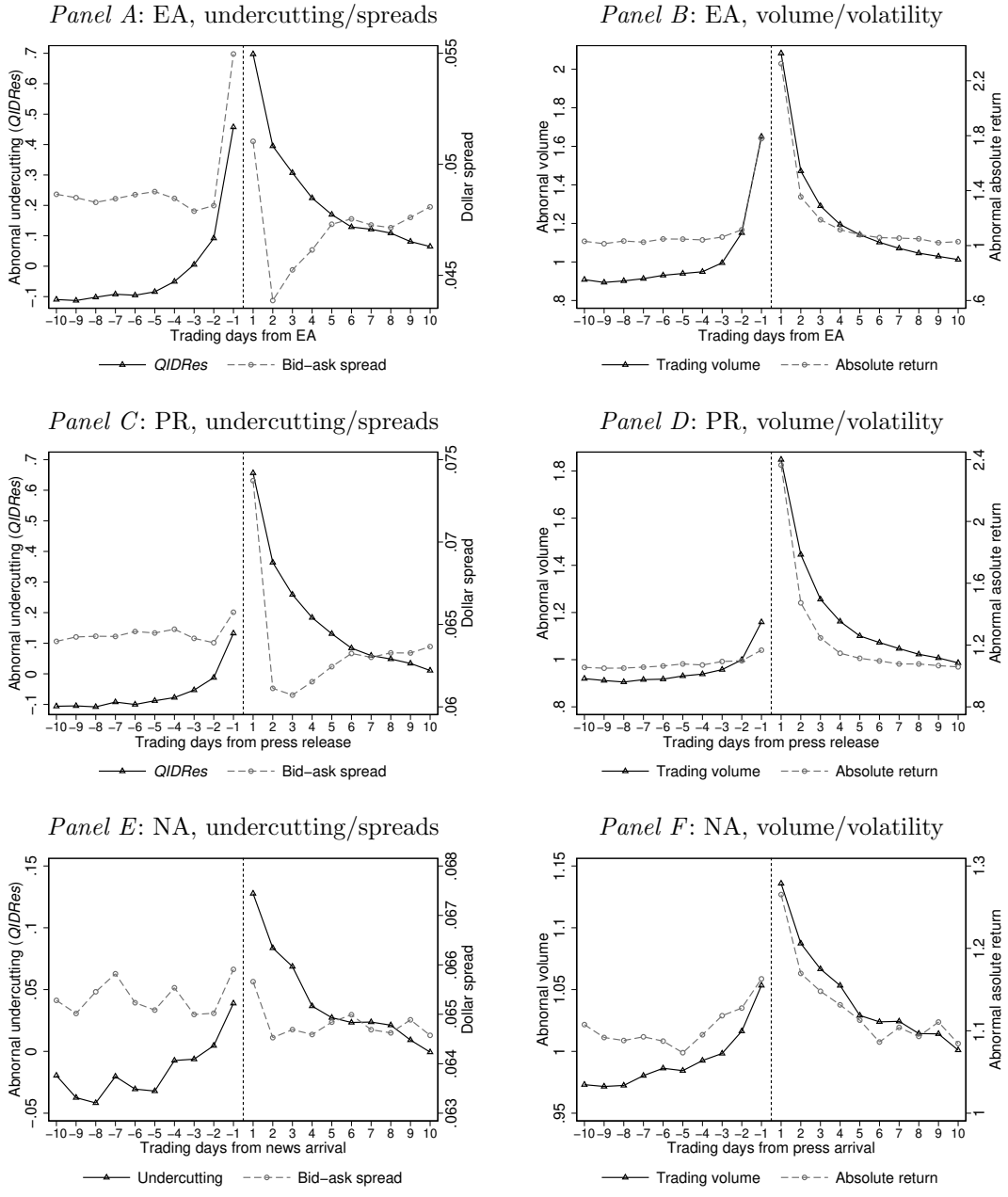


Figure 6. Undercutting Activity and Information Content of Trades, Events, and News.
Panels A through C present median abnormal undercutting activity around earnings announcements (EA), unscheduled press releases (PR), and news arrivals not associated with any identified event (NA). Earnings announcements are classified into events with high earnings surprise score (SUE), i.e., top and bottom 20% of SUE scores in the respective quarter, and low/moderate SUE, i.e., the middle 60% of SUE scores in the respective quarter. Both unscheduled press releases (PR) and news arrivals (NA) are classified into high post-announcement/-news 10-day return, i.e., the top 40% of absolute 10-day compound return, and low post-announcement/-news 10-day return, i.e., the bottom 60% of absolute 10-day compound return. Daily abnormal undercutting values are calculated based on equation (4). Panels D through F present medians of daily percentage effective spreads, realized spreads and price impacts, all obtained from WRDS Intraday Indicators, around earnings announcements (EA), unscheduled press releases (PR), and news arrivals not associated with any identified event (NA). The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least $5, excluding stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. Earnings announcement dates are obtained from COMPUSTAT; SUE scores are obtained from I/B/E/S; unscheduled press release dates and news arrivals not associated with any identified event are obtained from Ravenpack.

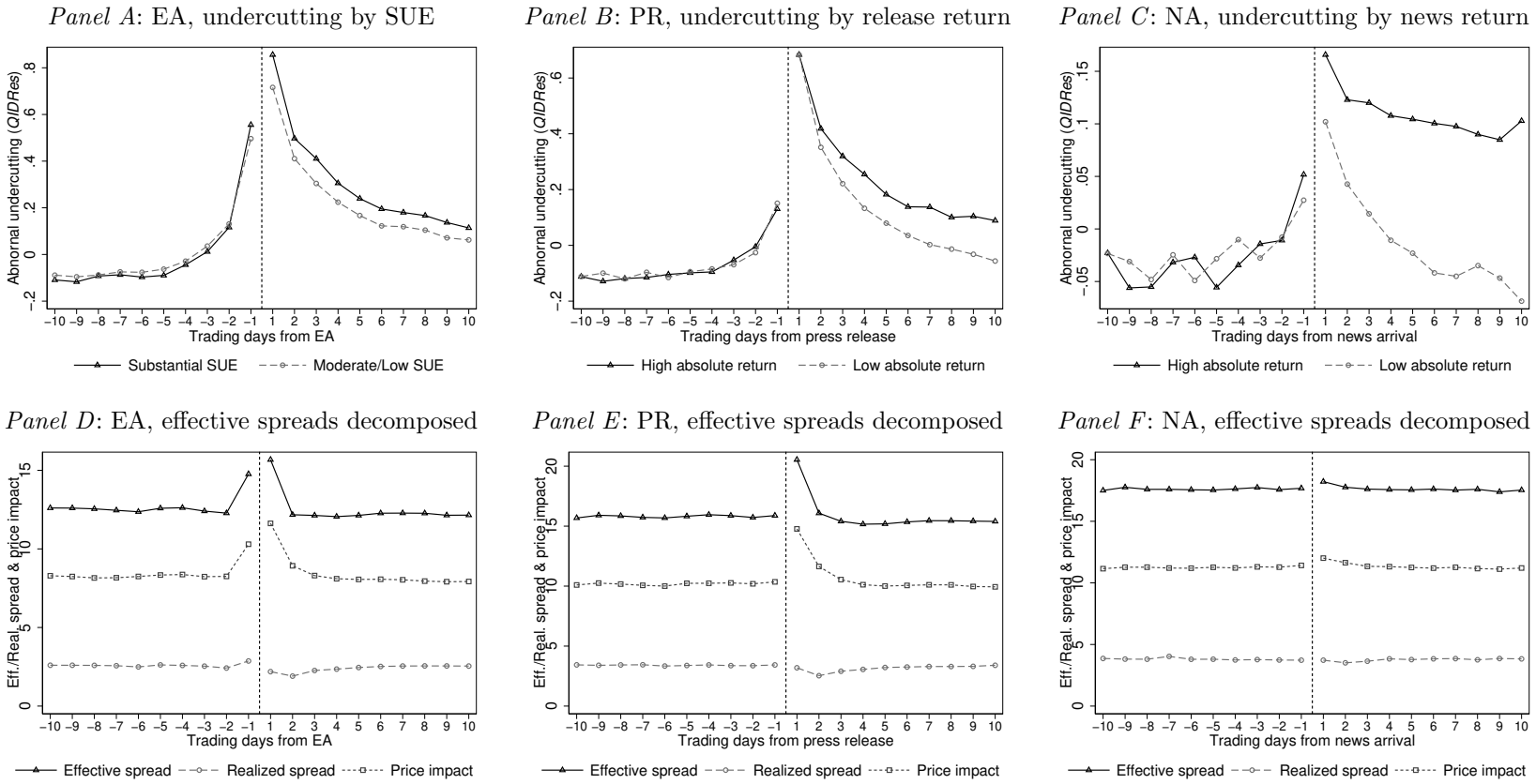
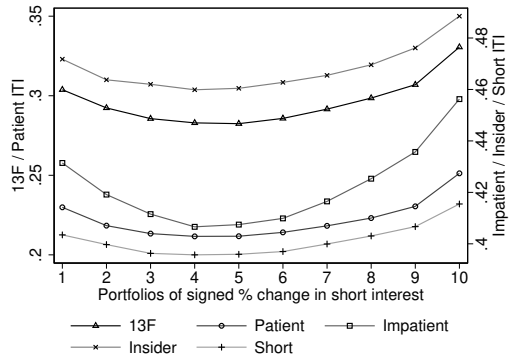
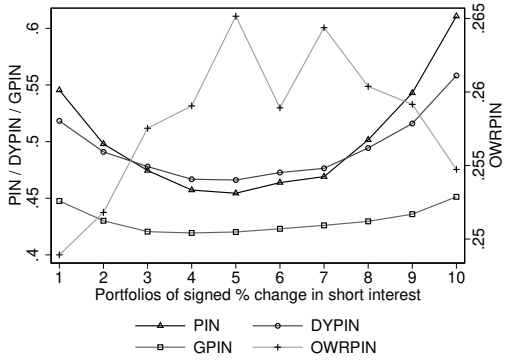


Figure 7. Informed Trading measures and Short Selling Activity.
The figure presents averages of various informed trading measures across levels of short selling activity. For averages of daily informed trading measures are calculated over bi-weekly intervals and matched with corresponding percentage change in short interest. each bi-weekly cross-section is sorted into portfolio (deciles) of signed percentage change in short interest. Equal weighted means of informed trading measures are calculated across stocks in each portfolio at the bi-weekly frequencies. The time-series averages are these means are plotted portfolio indexes, with 1 and 10 indexing the portfolios of stocks with largest declines and increased, respectively, in short interest. Panel A, B, and C present results for ITI, PIN, and MIA measures, respectively. Panel D presents results based on $QIDRes$ where each bi-weekly cross-section is decomposed in to terciles of the most recent short interest levels (defined as the most recent number of shares sold short by the total number of shares outstanding) before portfolios of percentage change in short interest are formed within each tercile. Panel E presents results based on $QIDRes$ where each bi-weekly cross-section is decomposed in to terciles of market-capitalization (defined as the product of the most recent share price and the total number of shares outstanding) before portfolios of percentage change in short interest are formed within each tercile. Daily $QIDRes$ observations are adjusted relative to the respective cross-stock average. The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least $5, excluding stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment.

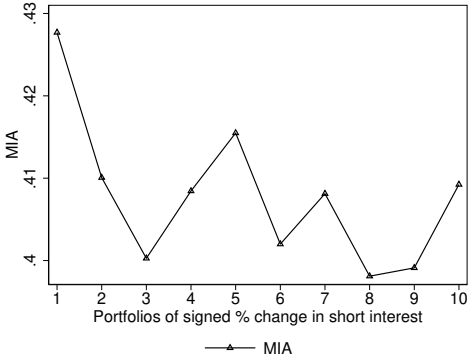
Panel A: ITI measures



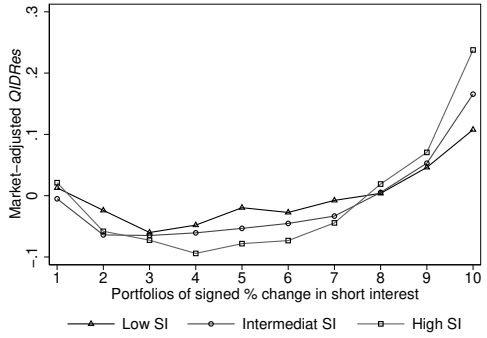
Panel B: PIN measures



Panel C: MIA measure



Panel D: $QIDRes$ by short interest level



Panel E: $QIDRes$ by market-capitalization

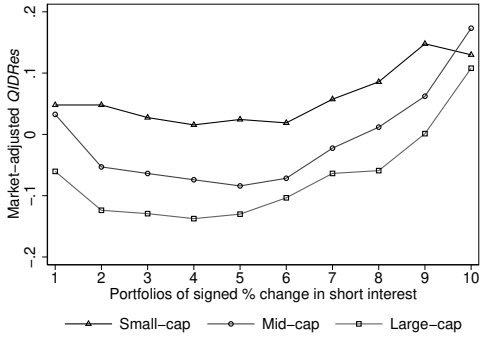


Figure 8. Intraday Sources of Variation in *QIDRes*.
The figure presents the decomposition of the variation in $QIDRes$ into into intraday components. In each quarter q and for each stock j, $QIDRes^q_{jt}$ is regressed on intraday component $QIDRes(\tau)^q_{jt}$, with $\tau \in \{am, md, pm\}$. The R^2 statistic from each regression for time-of-day τ is stored. The figure plots kernel densities for empirical distributions of R^2's across stock-quarters by τ. The sample includes stock-days of NMS-listed common shares between Jan 01, 2010 through Dec 31, 2019 with previous months' closing prices of at least \$5, excluding stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment.

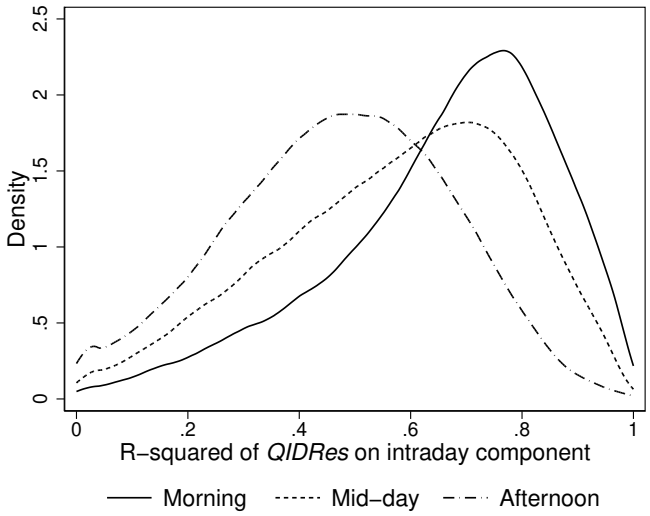
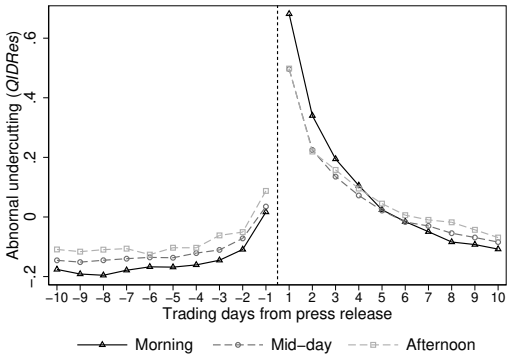


Figure 9. Undercutting Activity and Information Asymmetry around Unscheduled Corporate Announcements by Dime of Day.

The figure presents abnormal undercutting activity at different time-of-day windows around unscheduled press releases (PR). Intraday abnormal undercutting values are calculated based on equation (4) with $QID(\tau)$ refelecting undercutting activity in time-of-day window $\tau \in \{am, md, pm\}$. The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least \$5, excluding stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. Unscheduled press release dates are obtained from Ravenpack.

Panel A: Median $QIDRes$ by time of day



Panel B: Mean $QIDRes$ by time of day

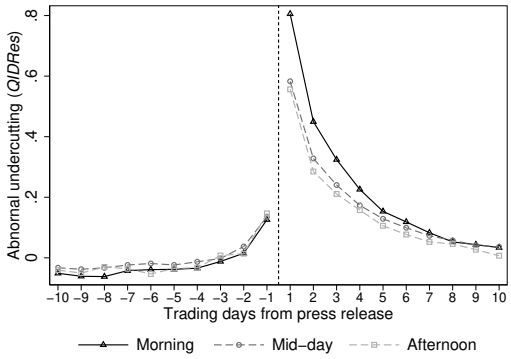


Table 1. Summary Statistics: Quote Revisions, *QID*, and *QIDRes*.

Panel A reports summary statistics of NBBO revisions as well as undercutting and abnormal undercutting measures. For each stock j on day t, NBBO improvements and deteriorations are counted separately for the bid (NBB) and ask (NBO) sides of the market. Trade-driven best quote deteriorations reflecting quote updates recorded no later that 10 milliseconds after a trade are constructed separately. For both categories, the share of single-tick updates devides the number of single-tick quote updates by all quote updates in the respective category. All quote improvements, $\#Impr_{jt}$, reflect the sum of the corresponding best bid and ask side improvements. Trade-drive quote deteriorations, $\#DeterTrade_{jt}$, reflect the sum of corresponding trade-driven best bid and ask deteriorations. The undercutting activity measure, QID, is constructed according to equation (2). Abnormal undercutting, $QIDRes$ is constructed according to equation (4). $QIDRes$ summary statistics are provided both before and after winsorizing each daily cross-section at the 1^{st} and 99^{th} percentiles. The sample includes NMS common shares from January 2010 to December 2019, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. Panel B reports correlation coefficients between daily $QIDRes_{jt}$ and contempranous measures of quoted, effective, and realized spreads; price impacts, volatility, and trading volume.

Panel A: Summary statistics

						Percentiles						
Variable	Observations	Mean	S.D.	Skew	1st	5th	25th	50th	75th	95th	99th
All NBB revisions	6,662,352	1916.45	3167.51	29.51	2	17	337	1056	2394	6442	13066
NBB improvements	**6,662,352**	**1046.47**	**1717.63**	**29.90**	**1**	**9**	**183**	**579**	**1320**	**3511**	**7046**
Share of single-tick	6,662,352	*0.79*	0.28	−1.96	0.00	0.00	0.77	*0.89*	0.96	1.00	1.00
NBB deteriorations	6,662,352	869.98	1479.88	32.02	0	7	151	469	1068	2955	6108
Trade-driven NBB deteriorations	**6,662,352**	**279.05**	**523.87**	**9.09**	**0**	**0**	**25**	**110**	**323**	**1096**	**2268**
Share of single-tick	6,662,352	*0.58*	0.34	−0.37	0.00	0.00	0.33	*0.61*	0.91	1.00	1.00
All NBO revisions	6,662,352	1924.87	3146.58	16.76	1	17	341	1066	2403	6464	13091
NBO improvements	**6,662,352**	**1052.88**	**1715.27**	**17.70**	**1**	**9**	**184**	**584**	**1326**	**3530**	**7081**
Share of single-tick	6,662,352	*0.79*	0.28	−1.95	0.00	0.00	0.76	*0.89*	0.96	1.00	1.00
NBO deteriorations	6,662,352	872.00	1464.30	18.03	0	7	153	473	1070	2958	6100
Trade-driven NBO deteriorations	**6,662,352**	**277.70**	**521.51**	**8.95**	**0**	**0**	**24**	**109**	**322**	**1094**	**2265**
Share of single-tick	6,662,352	*0.57*	0.34	−0.35	0.00	0.00	0.32	*0.61*	0.91	1.00	1.00
QID	**6,662,352**	**0.61**	**0.27**	**−0.42**	**0.03**	**0.12**	**0.41**	**0.64**	**0.84**	**0.99**	**1.00**
QIDRes	**6,662,352**	**0.07**	**1.52**	**1.53**	**−3.46**	**−1.85**	**−0.70**	**−0.01**	**0.77**	**2.20**	**4.23**
QIDRes (winsorized at $1^{st}/99^{th}$ percentile)	**6,662,352**	**0.07**	**1.38**	**1.46**	**−3.12**	**−1.85**	**−0.70**	**−0.01**	**0.77**	**2.20**	**3.77**

Panel B: Correlation coefficients between daily $QIDRes$ and contemporanous microstructure outcomes

Microstructure outcome	Quoted spread		Effective spread		Realized spread		Piece Impact		Volatility		Trading
	Dollar	Relative	Dollar	Relative	Dollar	Relative	Dollar	Relative	Realized	|Daily return|	volume
Correlation coefficient	0.0465	0.059	−0.0001	0.005	−0.0001	−0.0032	0.0025	0.0052	−0.0011	0.0009	0.0045

Table 2. Minimum Tick Size and the Undercutting Activity.
The table presents estimated impacts of an exogenous change in the minimum quoting and trading increment, i.e., tick size, on undercutting activity for differentially tick-constrained stocks. QID is the difference between the daily number of NBBO improvements and the number of trade-driven NBBO deteriorations, divided by the total number of NBBO updates. *Impr* divides the number of NBBO improvements by the number of NBBO updates. *DeterTrade* divides the number of *trade-driven* NBBO deteriorations by the number of NBBO updates Panel A presents the impacts of an increase in tick size from 1¢ to 5¢, using data from 08/12/2016-12/14/2016, for stocks with different tick constraint status prior to tick size increase. Stocks are classified into four tick constraint bins according to the average May and June 2016 quoted spreads of: (1) no more than 5¢, (2) 5¢ to 10¢, (3) 10¢ to 15¢, and (4) greater than 15¢. Panel B presents the impacts of a reduction in tick size from 5¢ to 1¢, using data from 08/08/2018-11/20/2018, for stocks with different tick constraint status prior to tick size reduction. Stocks are classified into four tick constraint bins according to the average May and June 2018 quoted spreads of: (1) no more than 5.5¢, (2) 5.5¢ to 10¢, (3) 10¢ to 15¢, and (4) greater than 15¢. Equation (5) is estimated using median (quantile) and OLS regressions. Estimates control for date fixed effects and double-cluster standard errors by stock and date. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% levels, respectively.

Panel A: TSP imposition

	QR				OLS			
Dependent variable:	May & June 2016 quoted spread group				May & June 2016 quoted spread group			
QID	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Pilot × *Event*	−0.36***	−0.51***	−0.37***	−0.29***	−0.38***	−0.60***	−0.52***	−0.32***
	[−18.07]	[−17.01]	[−11.80]	[−11.80]	[−20.89]	[−35.33]	[−20.12]	[−13.18]
Median/Mean of control	0.11	0.54	0.70	0.74	0.16	0.50	0.64	0.65
Impr								
Pilot × *Event*	−0.043***	−0.061***	−0.074***	−0.079***	−0.030***	−0.065***	−0.075***	−0.054***
	[−19.83]	[−18.52]	[−16.33]	[−12.43]	[−19.96]	[−30.41]	[−22.36]	[−11.00]
DeterTade								
Pilot × *Event*	0.090***	0.098***	0.075***	0.052***	0.087***	0.12***	0.100***	0.056***
	[18.24]	[15.58]	[11.02]	[9.51]	[24.14]	[33.99]	[17.75]	[10.64]

Panel B: TSP conclusion

	QR				OLS			
Dependent variable:	May & June 2018 quoted spread bin				May & June 2018 quoted spread bin			
QID	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Pilot × *Event*	0.23***	0.64***	0.54***	0.27***	0.33***	0.52***	0.54***	0.27***
	[9.54]	[26.90]	[18.30]	[11.63]	[15.51]	[38.45]	[28.99]	[13.21]
Median/Mean of control	−0.01	0.35	0.38	0.46	0.02	0.33	0.37	0.42
Impr								
Pilot × *Event*	0.010***	0.059***	0.078***	0.060***	0.0059***	0.032***	0.061***	0.048***
	[5.52]	[14.48]	[16.23]	[9.88]	[4.55]	[15.68]	[16.38]	[9.82]
DeterTade								
Pilot × *Event*	−0.053***	−0.12***	−0.11***	−0.049***	−0.078***	−0.12***	−0.11***	−0.051***
	[−8.98]	[−21.92]	[−18.14]	[−10.57]	[−15.86]	[−39.42]	[−25.58]	[−12.39]

Table 3. Probability of Unscheduled Press Releases and Recent *QIDRes*.
This table reports in the predictive power of $QIDRes$ for the likelihood of imminent unscheduled press releases (PR). Panel A fit logit regressions of day t probability of PR conditional on the most recent 5-day change in $QIDRes$. Panel A fit logit regressions of day t probability of PR conditional on the most recent 5-day changes in $QIDRes$, bid-ask spread (qsp), trading volume (tv), and absolute daily return $|r|$ as well as arrivals of information events, including earnings announcements (EA); press releases (PR); or news arrivals (NA) over days $t-5$ through $t-1$, specified using indicator variables $I(Inf)_{t-1}$ through $I(Inf)_{t-5}$. All estimates control for firm fixed effects. The sample includes NMS common shares from January 2010 to December 2019, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Logit estimates of the probability of PR conditional on $QIDRes$

Independent variable	Year 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
$\mathbf{\Delta QIDRes_{t-1}}$	**0.049*** [9.50]**	**0.052*** [12.27]**	**0.043*** [9.65]**	**0.033*** [7.79]**	**0.040*** [10.51]**	**0.042*** [10.39]**	**0.062*** [13.21]**	**0.030*** [5.93]**	**0.052*** [9.52]**	**0.079*** [18.80]**
Observations	285,847	408,344	402,579	434,403	482,783	502,762	447,180	226,810	260,723	575,013

Panel B: Logit estimates of the probability of PR conditional on $QIDRes$ and controls

Independent variable	Year 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
$\mathbf{\Delta QIDRes_{t-1}}$	**0.037*** [6.52]**	**0.043*** [9.53]**	**0.028*** [5.87]**	**0.026*** [5.57]**	**0.028*** [6.60]**	**0.033*** [7.50]**	**0.048*** [9.62]**	**0.026*** [4.78]**	**0.045*** [7.80]**	**0.057*** [12.65]**
Δqsp_{t-1}	−0.33*** [−2.59]	−0.29*** [−2.98]	−0.26*** [−3.68]	−0.092 [−1.44]	−0.071 [−1.37]	−0.16** [−2.51]	−0.11 [−1.28]	−0.094 [−0.99]	−0.34*** [−4.51]	−0.031 [−0.63]
Δtv_{t-1}	0.039*** [9.96]	0.034*** [10.30]	0.065*** [15.96]	0.045*** [10.94]	0.052*** [13.20]	0.050*** [12.77]	0.055*** [13.19]	0.060*** [12.24]	0.045*** [9.54]	0.062*** [13.66]
$\Delta\lvert r\rvert_{t-1}$	0.018*** [5.31]	0.021*** [7.29]	0.014*** [3.55]	0.0050 [1.33]	0.0041 [1.23]	−0.0035 [−1.11]	−0.011*** [−3.14]	−0.020*** [−3.64]	−0.013*** [−2.99]	−0.0030 [−0.98]
$I[Inf]_t - 1$	0.73*** [34.68]	0.63*** [37.60]	0.85*** [44.27]	0.64*** [37.78]	0.71*** [46.58]	0.92*** [58.20]	0.95*** [53.12]	0.56*** [26.59]	0.67*** [32.06]	1.03*** [66.36]
$I[Inf]_t - 2$	0.093*** [4.06]	0.035* [1.95]	0.0026 [0.12]	−0.0036 [−0.20]	0.028* [1.67]	0.071*** [4.00]	0.062*** [3.10]	0.040* [1.81]	0.14*** [6.55]	0.12*** [6.87]
$I[Inf]_t - 3$	0.042* [1.81]	0.047*** [2.59]	−0.012 [−0.56]	0.0066 [0.36]	0.035** [2.09]	0.053*** [2.95]	0.083*** [4.09]	0.080*** [3.60]	0.11*** [5.20]	0.11*** [6.44]
$I[Inf]_t - 4$	0.041* [1.75]	0.055*** [3.00]	0.0020 [0.09]	−0.045** [−2.41]	0.0065 [0.39]	0.041** [2.30]	0.046** [2.27]	0.070*** [3.13]	0.072*** [3.25]	0.16*** [9.11]
$I[Inf]_t - 5$	0.10*** [4.34]	0.17*** [9.67]	0.035 [1.62]	0.055*** [3.01]	0.070*** [4.17]	0.059*** [3.29]	0.062*** [3.03]	0.059*** [2.67]	0.12*** [5.56]	0.19*** [11.09]
Observations	275,157	395,593	387,235	417,403	465,960	486,121	433,618	221,401	254,999	558,273

Table 4. Probability of news arrivals and Recent *QIDRes*.
This table reports in the predictive power of $QIDRes$ for the likelihood of imminent news arrivals (NA). Panel A fit logit regressions of day t probability of NA conditional on the most recent 5-day change in $QIDRes$. Panel B fit logit regressions of day t probability of NA conditional on the most recent 5-day changes in $QIDRes$, bid-ask spread (qsp), trading volume (tv), and absolute daily return $|r|$ as well as arrivals of information events, including earnings announcements (EA); press releases (PR); or news arrivals (NA) over days $t-5$ through $t-1$, specified using indicator variables $I(Inf)_{t-1}$ through $I(Inf)_{t-5}$. All estimates control for firm fixed effects. The sample includes NMS common shares from January 2010 to December 2019, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Logit estimates of the probability of NA conditional on $QIDRes$

Independent variable	Year 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
$\boldsymbol{\Delta QIDRes_{t-1}}$	**0.0096****	**0.018*****	**0.014*****	**0.0095*****	**0.019*****	**0.014*****	**0.019*****	**0.0042**	**0.0097*****	**0.020*****
	[2.04]	**[4.85]**	**[3.99]**	**[2.82]**	**[5.61]**	**[4.00]**	**[5.18]**	**[1.11]**	**[2.61]**	**[6.77]**
Observations	264,162	392,899	390,291	434,513	469,206	486,223	424,563	223,100	260,629	584,506

Panel B: Logit estimates of the probability of NA conditional on $QIDRes$ and controls

Independent variable	Year 2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
$\boldsymbol{\Delta QIDRes_{t-1}}$	**0.0054**	**0.013*****	**0.0079****	**0.0057**	**0.0083****	**0.0079****	**0.014*****	**−0.00014**	**0.0036**	**0.0060***
	[1.07]	**[3.34]**	**[2.14]**	**[1.60]**	**[2.27]**	**[2.07]**	**[3.43]**	**[−0.03]**	**[0.93]**	**[1.95]**
Δqsp_{t-1}	−0.26**	−0.13	−0.26***	−0.16***	−0.060	−0.11*	−0.099	−0.066	−0.21***	−0.057*
	[−1.98]	[−1.38]	[−4.60]	[−3.22]	[−1.25]	[−1.82]	[−1.37]	[−0.93]	[−4.13]	[−1.77]
Δtv_{t-1}	0.021***	0.012***	0.025***	0.026***	0.040***	0.029***	0.035***	0.035***	0.026***	0.031***
	[6.39]	[4.24]	[7.44]	[7.57]	[11.47]	[8.88]	[10.66]	[9.53]	[7.49]	[9.40]
$\Delta\|r\|_{t-1}$	0.012***	0.014***	0.011***	0.021***	0.017***	0.018***	0.0058**	0.0030	−0.00069	0.0100***
	[3.69]	[5.48]	[3.28]	[7.16]	[5.47]	[6.42]	[1.99]	[0.69]	[−0.24]	[4.75]
$I[Inf]_t - 1$	0.17***	0.21***	0.24***	0.21***	0.27***	0.26***	0.23***	0.21***	0.29***	0.38***
	[9.19]	[15.06]	[16.79]	[15.86]	[19.99]	[19.69]	[16.23]	[13.84]	[21.66]	[36.12]
$I[Inf]_t - 2$	0.16***	0.13***	0.19***	0.14***	0.24***	0.18***	0.13***	0.19***	0.19***	0.23***
	[8.66]	[9.02]	[12.96]	[10.99]	[17.76]	[13.10]	[8.74]	[12.68]	[14.16]	[21.82]
$I[Inf]_t - 3$	0.052***	0.074***	0.15***	0.15***	0.23***	0.12***	0.13***	0.099***	0.16***	0.16***
	[2.74]	[5.12]	[9.97]	[11.65]	[17.52]	[8.66]	[9.13]	[6.53]	[11.84]	[14.57]
$I[Inf]_t - 4$	0.053***	0.099***	0.10***	0.050***	0.11***	0.11***	0.081***	0.11***	0.16***	0.15***
	[2.75]	[6.86]	[6.79]	[3.77]	[8.19]	[8.16]	[5.50]	[7.19]	[11.96]	[14.25]
$I[Inf]_t - 5$	0.13***	0.16***	0.15***	0.12***	0.11***	0.16***	0.15***	0.13***	0.17***	0.18***
	[6.95]	[11.28]	[10.49]	[9.07]	[8.26]	[11.58]	[10.09]	[8.43]	[12.17]	[17.06]
Observations	254,568	380,563	375,202	418,031	450,407	468,964	411,449	218,048	254,752	566,665

Table 5. Informed Trading Measures around Informed Trades of Mutual Funds.
The table reports the incremental differences in various measures of informed trading around informed trades of mutual funds. Measures of informed trading are compared between stock-days around institutional buys and sells involved in Industry-Neutral Self-Financed Informed-Trades of Barardehi et al. (2022) and other stock-days. For each informed trading measure Y_t^j, the η_i coefficient from the following regression is reported: $Y_t^j = \eta_0 + \eta_i \times I(t-i, t+i)_t^j + \epsilon_t^j$, where $I(t-i, t+i)_t^j$ is an indicator function that equals 1 in the $i \in \{0, 1, 2\}$ days surrounding an INSFIT trade on t, and equals 0 otherwise. The model is fit once using INSFIT buy trade indicators and once using INSFIT sell trade indicators. All estimates control for firm and date fixed effects. The sample includes NMS common shares from January 2010 to September 2011, excluding stocks whose previous month-end's closing price is below \$5. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Difference in informed trading measures around INSFIT buys trades

INSFIT trade window	Informed trading measure										
	QIDRes	ITI_{13D}	$ITI_{patient}$	$ITI_{impatient}$	$ITI_{insider}$	ITI_{short}	PIN	$DYPIN$	$GPIN$	$OWRPIN$	MIA
t	**0.083*** **	0.014***	0.0058***	0.015***	0.0054***	0.0074***	0.022***	0.031***	−0.0043	−0.0054**	−0.0035
	[5.46]	[7.73]	[3.49]	[10.45]	[3.07]	[10.60]	[3.75]	[4.79]	[−0.68]	[−2.15]	[−0.52]
$[t-1, t+1]$	**0.071*** **	0.0072***	0.0028**	0.0084***	0.0030**	0.0045***	0.016***	0.019***	−0.0014	−0.0067**	−0.00097
	[3.94]	[4.73]	[2.03]	[6.72]	[2.44]	[8.03]	[3.03]	[3.82]	[−0.31]	[−1.97]	[−0.25]
$[t-2, t+2]$	**0.066*** **	0.0056***	0.0015	0.0068***	0.0025**	0.0036***	0.011**	0.017***	−0.0015	−0.0067**	−0.0040
	[3.94]	[4.17]	[1.22]	[5.92]	[2.40]	[7.23]	[2.31]	[3.92]	[−0.39]	[−2.03]	[−1.17]
Sample mean	0.0112	0.3041	0.2225	0.4395	0.4401	0.4252	0.5679	0.5317	0.4337	0.2712	0.3128

Panel B: Difference in informed trading measures around INSFIT sell trades

INSFIT trade window	Informed trading measure										
	QIDRes	ITI_{13D}	$ITI_{patient}$	$ITI_{impatient}$	$ITI_{insider}$	ITI_{short}	PIN	$DYPIN$	$GPIN$	$OWRPIN$	MIA
t	**0.068*** **	0.013***	0.0080***	0.013***	−0.0011	0.0054***	0.040***	0.033***	0.011	0.0061*	−0.0095
	[3.42]	[6.08]	[3.79]	[7.38]	[−0.50]	[5.91]	[5.96]	[3.99]	[1.50]	[1.82]	[−1.18]
$[t-1, t+1]$	**0.056*** **	0.0097***	0.0048***	0.0091***	0.00056	0.0042***	0.022***	0.015***	0.0034	0.0035	−0.0045
	[3.42]	[6.31]	[3.62]	[7.26]	[0.45]	[6.59]	[4.19]	[2.79]	[0.71]	[1.32]	[−0.94]
$[t-2, t+2]$	**0.050*** **	0.0075***	0.0045***	0.0077***	0.00061	0.0032***	0.017***	0.015***	0.0034	0.0019	−0.0047
	[3.31]	[5.50]	[3.69]	[6.74]	[0.58]	[5.86]	[3.52]	[3.20]	[0.86]	[0.74]	[−1.16]
Sample mean	0.0365	0.3061	0.2278	0.4401	0.4353	0.4243	0.5875	0.5415	0.4565	0.2713	0.3322

Table 6. Price Reversals by Abnormal Undercutting Activity and Realized Volatility.

This table reports the extent of price reversal over the next 10 trading days conditional on day t abnormal undercutting activity and realized volatility. For Panel A results, each daily cross-section is sorted into terciles of $QIDRes$. For each such tercile panel regressions of compound returns over the next $n \in \{1, 2, \ldots, 10\}$ days from day t's close, denoted $CR^j_{t,t+n}$, on day t returns, denoted R^j_t, are estimated. For Panel B results, each daily cross-section is sorted ***independently*** into terciles of $QIDRes$ and realized volatility, $qvol$. For each of the nine categories, panel regressions of compound returns over 1 day forward ($CR^j_{t,t+1}$), 5 days forward ($CR^j_{t,t+5}$), and 10 days forward ($CR^j_{t,t+10}$), on day t returns, denoted R^j_t, are estimated. Regressions control for stock and date fixed effects and double-cluster standard errors at both date and stock levels. All return cross-sections are winsorized at 1% and 99%. Estimates are reported by $QIDRes$ tercile and n (Panel A) or by $QIDRes$ and volatility terciles (Panel B). The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Price reversals up to 10 days forward by day-t $QIDRes$ tercile

	Dependent Variable									
$QIDRes$ tercile	$CR_{t,t+1}$	$CR_{t,t+2}$	$CR_{t,t+3}$	$CR_{t,t+4}$	$CR_{t,t+5}$	$CR_{t,t+6}$	$CR_{t,t+7}$	$CR_{t,t+8}$	$CR_{t,t+9}$	$CR_{t,t+10}$
Low	−0.053***	−0.057***	−0.070***	−0.078***	−0.084***	−0.088***	−0.089***	−0.093***	−0.10***	−0.097***
	[−4.20]	[−4.33]	[−4.97]	[−5.54]	[−6.06]	[−6.28]	[−6.39]	[−6.76]	[−7.70]	[−7.24]
Medium	−0.051***	−0.056***	−0.066***	−0.074***	−0.076***	−0.081***	−0.083***	−0.087***	−0.092***	−0.090***
	[−3.53]	[−3.81]	[−4.32]	[−4.83]	[−5.02]	[−5.33]	[−5.62]	[−5.87]	[−6.62]	[−6.32]
High	−0.025***	−0.030***	−0.038***	−0.044***	−0.047***	−0.050***	−0.054***	−0.054***	−0.059***	−0.057***
	[−3.22]	[−3.63]	[−4.42]	[−5.03]	[−5.34]	[−5.51]	[−5.96]	[−5.96]	[−6.55]	[−6.21]

Panel B: Price reversals 1, 5, and 10 days forward by day-t $QIDRes$ and realized volatility terciles

		Dependent variable								
		$CR_{t,t+1}$			$CR_{t,t+5}$			$CR_{t,t+10}$		
		Realized volatility tercile			Realized volatility tercile			Realized volatility tercile		
$QIDRes$ tercile		Low	Medium	High	Low	Medium	High	Low	Medium	High
Low	Slope	−0.056***	−0.055***	−0.046***	−0.092***	−0.084***	−0.076***	−0.099***	−0.10***	−0.090***
		[−3.73]	[−4.00]	[−5.25]	[−5.55]	[−5.56]	[−6.59]	[−6.17]	[−6.58]	[−7.45]
	Observations	536,727	539,164	538,302	536,727	539,164	538,302	536,727	539164	538302
Medium	Slope	−0.047***	−0.051***	−0.054***	−0.070***	−0.073***	−0.085***	−0.080***	−0.092***	−0.098***
		[−4.59]	[−2.76]	[−3.96]	[−5.96]	[−3.81]	[−5.64]	[−6.49]	[−5.08]	[−6.69]
	Observations	538,961	538,924	540,782	538,961	538,924	540,782	538,961	538,924	540,782
High	Slope	−0.024***	−0.023***	−0.026***	−0.045***	−0.044***	−0.052***	−0.053***	−0.058***	−0.060***
		[−3.48]	[−2.87]	[−3.30]	[−5.15]	[−4.70]	[−5.42]	[−5.38]	[−5.66]	[−5.98]
	Observations	540,265	539,962	540,681	540,265	539,962	540,681	540,265	539,962	540,681

Table 7. Correlation between Informed Trading Measures and Stock Illiquidity.
This table presents the correlations matrices of informed trading measures and stock illiquidity. Panel A reports on the correlations between $QIDRes$ (indexed 1); five versions of ITI (indexed 2 through 6); and five illiquidity measures, time-weighted dollar quoted spread (QSP), size-weighted dollar effective srpead ($EFSP$), Kyle's λ ($Lambda$), Barardehi et al. (2021)'s open-to-close Amihud measure (AM), and Barardehi et al. (2023)'s retail-based institutional liquidity measure ($ILMV$), indexed 11 through 15, for the 2010-2019 sample sample. Panel B reports on the correlations between $QIDRes$, indexed 1; five versions of ITI, indexed 2 through 6; four versions of PIN, indexed 7 through 10; and five illiquidity measures, QSP, $EFSP$. $Lambda$, AM, and $ILMV$, indexed 7 through 11, for the 2010-2012 sample, where we have access to PIN measures. All measures are constructed at the monthly frequency by averaging daily observations.

Panel A: Correlation between, $QIDRes$, ITI, and illiquidity, the 2010-2019 sample

Variable		Variable index									
index		1	2	3	4	5	6	7	8	9	10
1	$QIDRes$										
2	ITI_{13D}	0.10									
3	$ITI_{patient}$	0.12	0.79								
4	$ITI_{impatient}$	0.11	0.73	0.56							
5	$ITI_{insider}$	0.01	0.35	0.39	0.38						
6	ITI_{short}	0.14	0.44	0.36	0.63	0.17					
7	QSP	−0.01	−0.11	−0.08	−0.17	0.06	−0.23				
8	$EFSP$	0.00	−0.13	−0.09	−0.19	0.05	−0.25	0.97			
9	$Lambda$	0.03	−0.12	−0.03	−0.28	0.09	−0.31	0.24	0.29		
10	AM	0.01	−0.10	−0.05	−0.22	−0.01	−0.21	0.27	0.31	0.57	
11	ILM	0.03	−0.18	−0.05	−0.35	0.05	−0.37	0.37	0.42	0.60	0.47

Panel B: Correlation between, $QIDRes$, ITI, PIN and illiquidity, the 2010-2012 sample

Variable		Variable index													
index		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	$QIDRes$														
2	ITI_{13D}	0.09													
3	$ITI_{patient}$	0.08	0.80												
4	$ITI_{impatient}$	0.11	0.75	0.59											
5	$ITI_{insider}$	−0.02	0.34	0.31	0.41										
6	ITI_{short}	0.08	0.51	0.49	0.67	0.28									
7	PIN	0.04	0.33	0.33	0.38	0.13	0.53								
8	$DYPIN$	0.04	0.31	0.30	0.34	0.16	0.43	0.61							
9	$GPIN$	0.05	−0.01	0.00	0.02	−0.07	0.09	0.03	0.02						
10	$OWRPIN$	−0.01	−0.01	0.01	−0.03	0.01	−0.06	−0.05	−0.01	−0.02					
11	QSP	−0.02	−0.04	−0.08	−0.04	0.06	−0.21	−0.17	−0.11	−0.17	0.07				
12	$EFSP$	−0.02	−0.04	−0.08	−0.04	0.06	−0.23	−0.18	−0.12	−0.17	0.09	0.93			
13	$Lambda$	0.00	−0.05	−0.05	−0.13	0.16	−0.24	−0.22	−0.12	−0.21	0.21	0.28	0.28		
14	AM	0.00	−0.03	−0.02	−0.10	0.02	−0.17	−0.14	−0.09	−0.11	0.13	0.14	0.15	0.66	
15	ILM	−0.02	0.02	0.01	−0.06	0.18	−0.27	−0.26	−0.15	−0.22	0.10	0.44	0.42	0.63	0.41

Table 8. Informed Trading Alphas.
This table presents excess returns as well as three-, four-, and six-factor alphas conditional on our measure of informed trading. Each month m cross-section in quarter q is sorted into quintiles of $QIDRes$ from quarter $q-1$ (Panel A) or from quarter $q-2$ (Panel B), with quintles formed based in NYSE breakpoints. The time series averages of monthly equally weighted portfolio returns as well that for the long-short (High−Low) portfolio, after subtracting the 1-month Treasury-bill rate, are reported as "excess returns." The 3-factor alphas reflect the intercept of time-series regressions of portfolio excess returns on Fama-French three factors. The 4-factor alphas reflect the intercepts when the 3-factor models are augmented with the momentum factor. The 6-factor alphas reflect the intercepts when 4-factor models are augmented by profitability and investment factors. The sample contains NMS common shares with previous month-end's closing prices of at least $5 from the January 2010 through August 2016. Standard errors are Newey-West-corrected using 12 lags. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Panel A: Monthy returns to portfolios $QIDRes$ from quarter $q-1$

	$QIDRes$ quintile					
Monthly portfolio return	Low	2	3	4	High	High−Low
Excess return	1.00**	1.10**	1.15***	1.35***	1.18***	0.18
	[2.41]	[2.60]	[2.68]	[3.16]	[3.12]	[1.57]
3-factor alpha	−0.20**	0.024	0.085	0.30***	0.095	0.30**
	[−2.26]	[0.29]	[0.93]	[4.32]	[1.09]	[2.24]
4-factor alpha	−0.17**	0.033	0.097	0.30***	0.11	0.28***
	[−2.32]	[0.36]	[1.15]	[4.33]	[1.37]	[2.69]
6-factor alpha	−0.21***	0.040	0.11	0.29***	0.11	0.32***
	[−2.75]	[0.44]	[1.37]	[3.98]	[1.49]	[3.10]

Panel B: Monthy returns to portfolios $QIDRes$ from quarter $q-2$

	$QIDRes$ quintile					
Monthly portfolio return	Low	2	3	4	High	High−Low
Excess return	1.04***	1.07***	1.21***	1.21***	1.28***	0.24*
	[2.69]	[2.93]	[3.08]	[3.03]	[3.08]	[1.72]
3-factor alpha	−0.16	−0.069	0.089	0.13	0.19***	0.35**
	[−1.29]	[−0.73]	[1.36]	[1.47]	[3.01]	[2.13]
4-factor alpha	−0.19	−0.072	0.073	0.13	0.17**	0.37*
	[−1.46]	[−0.78]	[1.28]	[1.55]	[2.13]	[1.98]
6-factor alpha	−0.17	−0.042	0.10**	0.14*	0.17**	0.34*
	[−1.36]	[−0.41]	[2.00]	[1.70]	[2.09]	[1.93]

Table 9. The Cross-Section of Expected Returns and Abnormal Undercutting Activity. This table reports on the relation between undercutting activity and the cross-section of expected returns. Equation (6) is estimated using $QIDRes$ constructed in the preceding two quarters and 5 liquidity measures constructed in month $m-2$. Other controls include three-factor Fama-French betas three-factor Fama-French betas ($\beta^{mkt}_{j,m-1}$, $\beta^{hml}_{j,m-1}$, $\beta^{smb}_{j,m-1}$), estimated using weekly observations from the two-year period ending in the final full week of month $m-1$, book-to-market ratio, ($\text{BM}_{j,m-1}$), natural log of market capitalization, ($\ln(\text{Mcap}_{j,m-12})$), dividend yield ($\text{DYD}_{j,m-1}$), defined as total dividends over the past 12 months divided by the share price at the end of month $m-1$, idiosyncratic volatility ($\text{IdVol}_{j,m-1}$), previous month's return ($RET_{j,m-1}$), preceding return from the prior 11 months ($RET_{j,(m-12,m-2)}$), and previous quarter's fraction institutionally owned shares outstanding ($IOShr_{j,q-1}$). The previous quarter's Herfindahl-Hirschman index for institutional ownership ($IOShrHHI_{j,q-1}$) and month $m-2$ share turnover ($TO_{j,m-2}$) serve as measures of market competition. Estimates are from panel regressions that control for firm and month-year fixed effects, double clustering standard errors by these two dimensions. The sample includes NMS common shares from January 2010 to December 2019, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Independent Variable				Illiquidity measures						
				QSP	*EFSP*	Lambda	*AM*	*ILM*		
$QIDRes_{q-1}$	**0.17*** [2.75]**	**0.21*** [3.17]**	**0.23*** [3.52]**	**0.21*** [3.23]**	**0.21*** [3.22]**	**0.21*** [3.17]**	**0.21*** [3.16]**	**0.21*** [3.21]**	**0.21*** [3.27]**	**0.23*** [3.60]**
$QIDRes_{q-2}$	**0.047 [0.84]**	**0.084 [1.54]**	**0.091* [1.68]**	**0.088 [1.62]**	**0.087 [1.61]**	**0.084 [1.54]**	**0.083 [1.53]**	**0.084 [1.56]**	**0.087 [1.62]**	**0.095* [1.76]**
Illiquidity				−1.13** [−2.39]	−2.38*** [−2.91]	0.0096 [0.08]	−0.36 [−1.02]	0.23 [0.46]		
β^{mkt}	−0.13 [−0.45]	0.26 [1.28]	0.29 [1.45]	0.26 [1.28]	0.25 [1.26]	0.26 [1.28]	0.25 [1.26]	0.26 [1.31]	0.26 [1.30]	0.28 [1.44]
β^{hml}	−0.18 [−1.10]	−0.15 [−0.97]	−0.14 [−0.93]	−0.15 [−0.98]	−0.15 [−0.99]	−0.15 [−0.98]	−0.15 [−0.97]	−0.15 [−0.97]	−0.15 [−0.99]	−0.15 [−0.95]
β^{smb}	0.064 [0.44]	0.076 [0.51]	0.089 [0.60]	0.075 [0.51]	0.075 [0.51]	0.074 [0.50]	0.072 [0.49]	0.077 [0.52]	0.072 [0.49]	0.085 [0.58]
BM	0.27** [2.15]	1.02*** [3.03]	1.09*** [3.31]	1.00*** [2.99]	1.00*** [2.98]	1.02*** [3.00]	1.04*** [3.07]	1.02*** [3.03]	1.00*** [2.96]	1.07*** [3.25]
ln(Mcap)	−0.0098 [−0.23]	−2.45*** [−10.07]	−2.43*** [−9.99]	−2.41*** [−10.10]	−2.41*** [−10.05]	−2.45*** [−10.03]	−2.46*** [−10.07]	−2.44*** [−9.98]	−2.39*** [−9.94]	−2.39*** [−9.96]
DYD	0.38 [0.19]	−0.42 [−0.21]	−0.16 [−0.08]	−0.69 [−0.35]	−0.70 [−0.35]	−0.43 [−0.21]	−0.44 [−0.22]	−0.39 [−0.20]	−0.67 [−0.34]	−0.47 [−0.24]
Id. Vol.	−0.21** [−2.37]	−0.072 [−1.00]	−0.058 [−0.83]	−0.067 [−0.93]	−0.065 [−0.90]	−0.073 [−1.01]	−0.070 [−0.97]	−0.070 [−0.98]	−0.064 [−0.88]	−0.052 [−0.72]
RET_{-1}	−1.09 [−1.01]	−4.45*** [−4.09]	−4.46*** [−4.09]	−4.48*** [−4.11]	−4.48*** [−4.11]	−4.44*** [−4.09]	−4.44*** [−4.10]	−4.45*** [−4.09]	−4.48*** [−4.12]	−4.48*** [−4.12]
$RET_{(-12,-2)}$	0.36 [1.29]	−1.77*** [−5.98]	−1.74*** [−5.76]	−1.73*** [−5.92]	−1.72*** [−5.89]	−1.77*** [−5.94]	−1.77*** [−5.97]	−1.76*** [−5.80]	−1.70*** [−5.64]	−1.69*** [−5.58]
$IOShr$	0.43*** [2.78]	−0.94*** [−3.18]	−1.39*** [−4.21]	−0.97*** [−3.29]	−0.98*** [−3.34]	−0.93*** [−3.17]	−0.95*** [−3.23]	−0.93*** [−3.17]	−0.96*** [−3.29]	−1.41*** [−4.25]
$IOShr_HHI$			−1.75*** [−3.44]							−1.69*** [−3.26]
TO			−31.0** [−2.46]							−33.1*** [−2.64]
Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	234,110	234,026	234,026	234,026	234,026	233,564	234,026	234,026	233,564	233,564

Table 10. The Cross-Section of Expected Returns and Informed Trading: Horse Race Regressions. This table reports on the relation informed trading measures and the cross-section of expected returns. equation (6) is estimated using $QIDRes$, along with different subsets of other informed trading measures, from the preceding two quarters. Control variables contain the full set of controls used in Table 9. The sample periods 2010-2019, 2010-2018, and 2010-2012 reflect the availability of alternative measures ITIs, MIA, and PIN, respectively. The samples include all NMS common shares, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. All estimates control for year-month and stock fixed effects, and standard errors are double-clustered at both levels. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

RHS variable	2010-2019 sample	2010-2018 sample		2010-2012 sample			
$\boldsymbol{QIDRes_{q-1}}$	**0.22*** **	**0.20** **	**0.21** **	**0.23**	**0.20**	**0.21**	**0.25**
	[3.07]	**[2.39]**	**[2.50]**	**[1.53]**	**[1.37]**	**[1.34]**	**[1.15]**
$\boldsymbol{QIDRes_{q-2}}$	**0.092***	**0.10**	**0.087**	**0.28****	**0.26****	**0.27****	**0.41*****
	[1.67]	**[1.50]**	**[1.29]**	**[2.41]**	**[2.35]**	**[2.24]**	**[3.01]**
$ITI_{13D,q-1}$	0.35		0.47		−1.16	−1.05	−3.96
	[0.29]		[0.30]		[−0.34]	[−0.31]	[−1.12]
$ITI_{13D,q-2}$	−1.94		−2.00		−4.41	−4.33	−2.41
	[−1.53]		[−1.31]		[−1.65]	[−1.62]	[−0.73]
$ITI_{patient,q-1}$	1.05		1.19		**5.20***	**5.36***	**6.53***
	[0.80]		[0.65]		**[1.71]**	**[1.74]**	**[1.82]**
$ITI_{patient,q-2}$	0.24		1.31		−0.55	−0.63	−0.86
	[0.18]		[0.76]		[−0.20]	[−0.23]	[−0.25]
$ITI_{impatient,q-1}$	−0.88		−2.22		−3.64	−3.01	−4.20
	[−0.59]		[−1.23]		[−0.88]	[−0.76]	[−0.80]
$ITI_{impatient,q-2}$	1.11		**3.26***		−0.39	−0.67	−3.98
	[0.81]		**[1.93]**		[−0.13]	[−0.24]	[−1.30]
$ITI_{insider,q-1}$	1.57		**2.68***		0.060	0.46	2.33
	[1.41]		**[1.85]**		[0.02]	[0.13]	[0.51]
$ITI_{insider,q-2}$	**2.43*****		**2.52****		**5.18***	**4.87***	3.68
	[2.66]		**[2.12]**		**[2.00]**	**[1.84]**	[1.01]
$ITI_{short,q-1}$	−1.14		−0.81		2.56	4.16	7.20
	[−0.41]		[−0.22]		[0.34]	[0.53]	[0.68]
$ITI_{short,q-2}$	1.28		−2.24		10.5	10.3	11.9
	[0.48]		[−0.65]		[1.68]	[1.62]	[1.58]
MIA_{q-1}		**1.67*****	**1.54*****				0.88
		[3.13]	**[2.95]**				[0.55]
MIA_{q-2}		0.22	0.22				0.42
		[0.45]	[0.47]				[0.34]
PIN_{q-1}				−0.13		−0.25	−0.18
				[−0.21]		[−0.37]	[−0.22]
PIN_{q-2}				0.13		0.028	−0.13
				[0.26]		[0.05]	[−0.21]
$DYPIN_{q-1}$				−0.60		−0.68	−0.12
				[−0.98]		[−1.09]	[−0.17]
$DYPIN_{q-2}$				0.56		0.48	0.60
				[0.92]		[0.76]	[0.76]
$GPIN_{q-1}$				0.44		0.42	0.37
				[0.96]		[0.90]	[0.57]
$GPIN_{q-2}$				−0.99*		−1.00*	−1.11
				[−1.93]		[−1.83]	[−1.47]
$OWRPIN_{q-1}$				−0.76		−0.78	−0.52
				[−1.47]		[−1.33]	[−1.22]
$OWRPIN_{q-2}$				**0.83***		**0.83***	**0.74***
				[1.74]		**[1.74]**	**[1.90]**
Observations	216,077	119,098	118,113	25,045	25,045	25,045	16,065

Table 11. Return Predictability of Informed Trading Measures and Short Sale Constraints. This table reports on the relation between $QIDRes$ and the cross-section of expected returns by level of short sale constraints. Equation (6) is estimated within terciles of quarter $q-3$'s average security lending fees obtained from FIS database from 2010 through 2018. The sample includes NMS common shares from January 2010 to December 2018, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment. The set of stock characteristics is identical to that used in Table 9. Estimates control for stock and year-month (year-quarter) fixed effects, and standard errors are double-clustered at both levels. The numbers in brackets are t-statistics with ***, **, and * identifying statistical significance at the 1%, 5%, and 10% level, respectively.

Independent Variable	Tercile of security lending fee					
	Low		Intermediate		High	
$QIDRes_{q-1}$	0.16**	0.15**	0.23***	0.21***	0.39***	0.38***
	[2.24]	[2.19]	[3.00]	[2.87]	[3.31]	[3.24]
$QIDRes_{q-2}$	0.038	0.036	0.15*	0.14*	0.19	0.18
	[0.60]	[0.57]	[1.86]	[1.83]	[1.61]	[1.55]
Stock characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Liquidity controls	Yes	No	Yes	No	Yes	No
Observations	72,747	72,913	70,888	71,012	67,101	67,227

A Appendix

A.1 A Simple Framework of Undercutting and Informed Trading

Consider a simple one period rational expectation equilibrium model that builds off of Glosten and Milgrom (1985). An asset takes the equally likely value of 0 or 1. The fraction π of liquidity demanders are informed and know the true value of the asset only buying when the the value equals 1 and only selling when the value equals 0. The remaining $1-\pi$ fraction of liquidity demanders are uninformed and buy and sell with equal probability. The exact arrival time of the next trade to arrive is random and follows an exponential distribution with arrival rate parameter λ. Liquidity providers come in two types: sophisticated and unsophisticated. Unsophisticated liquidity providers, denoted $ULPs$, are passive, competitive, and thus set prices equal to the conditional expected value of the asset. Sophisticated liquidity providers, denoted $SLPs$, can pay a cost c which will, with probability ρ inform them about whether the next trade to arrive is informed or uninformed and on which side of the market the trade will arrive. It does not inform them about the arrival time of the upcoming trade.[37] There are m SLPs where the value m is determined in equilibrium such that the expected profit associated with being an SLP is equal to the cost c, and so SLPs are competitive. The likelihood that at least one of the m $SLPs$ receives a signal is $\phi = 1-(1-\rho)^m$.

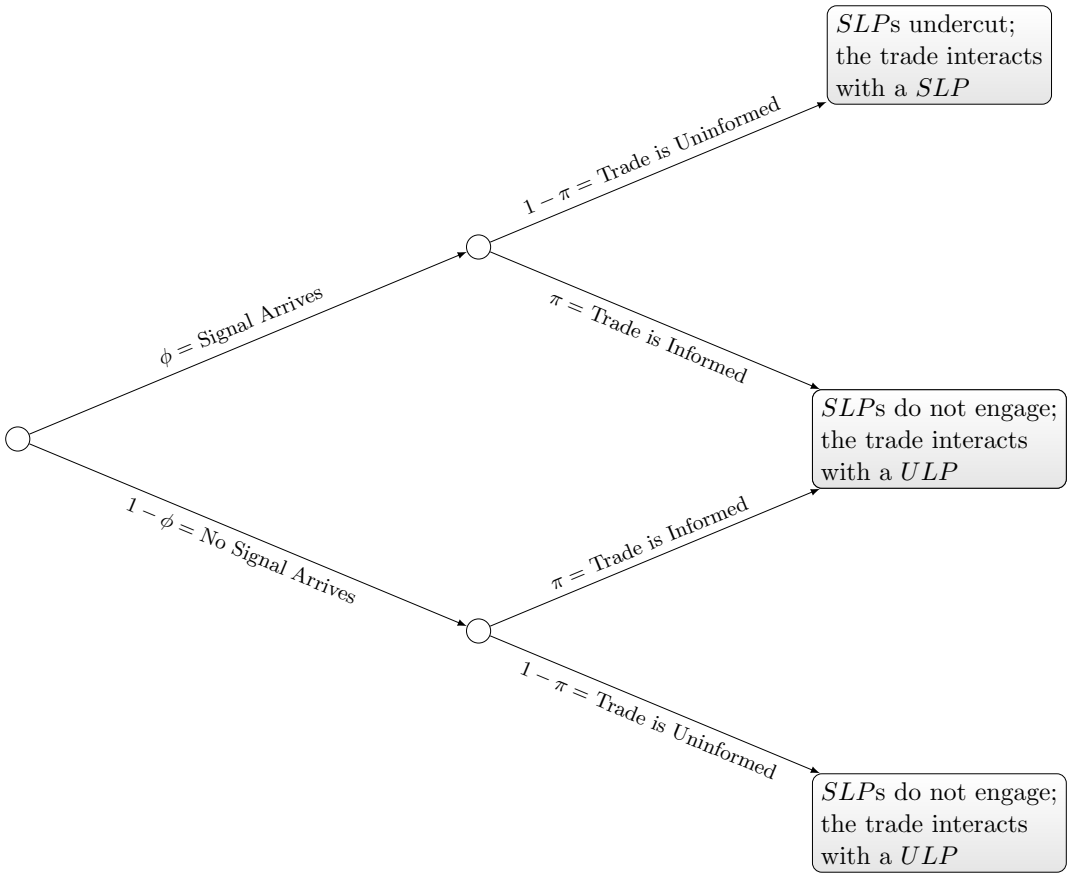
If a SLP receives a signal that an upcoming trade is informed, the SLP will simply sit out and not post any quotes allowing the ULPs to interact with the incoming informed trade. If no SLP receives a signal then all $SLPs$ sit out. However, if an SLP receives a signal that the upcoming trade is uninformed, they will undercut the existing quote on that side of the market. The other $SLPs$, whether they receive a signal or not, will observe this quote improvement and will infer that a signal has been received and will submit their own undercutting orders and an undercutting run will ensue.[38] The outcome tree in Figure A.1 illustrated this setup.

In this setup all informed trades interact with ULPs, and some uninformed trades interact with ULPs and some with SLPs. The probability that a ULP interacts with an informed trade is the probability of an informed trade arriving (π) divided by the probability that a trade interacts with a ULP, which is $1-\phi(1-\pi)$. The probability that a ULP interacts with an uninformed trade is simply the compliment as shown in equations A.1 and A.2,

[37] The cost c can be thought of as the cost of investing in the capacity to process, analyze, and respond quickly to information based in order flow.

[38] The assumption that all $SLPs$ can infer the signals of others via monitoring quote updates could be relaxed such that only those $SLPs$ receiving a signal engage in the undercutting run without changing any of the key inference. In this case ϕ could be redefined to be the probability that at least two $SLPs$ receive a signal $\phi = 1-(1-\rho)^m - m\rho(1-\rho)^{m-1}$, and all inference remains exactly the same since in both cases ϕ is increasing in both m and ρ. Additionally the profit to undercutting is random, since the arrival of the uninformed trade is random and so it is unclear exactly when during the run the uninformed trade will arrive. However, given that SLPs know the arrival rate of trades, they can compute the expected time during an undercutting run a trade will arrive and so can compute the expected profit of a run that is earned by the winning quote provider, which we denote Π. The likelihood that a given SLP wins the undercutting run is $\frac{1}{m}$, so expected profits to undercutting are $\frac{\Pi}{m}$. For this market to be in equilibrium it must be the case that $c = \frac{\Pi}{m}$ which implies that the number of SLPs is $m = \frac{\Pi}{c}$.

Figure A.1. Informed Trading Signal Arrivals and SLPs' Undercutting Choices.



$$P(I) = \frac{\pi}{1-\phi(1-\pi)}, \tag{A.1}$$

$$P(U) = \frac{(1-\phi)(1-\pi)}{1-\phi(1-\pi)}. \tag{A.2}$$

The bid and the ask prices are set by the ULPs equal to the expected value of the asset conditional on the trade occurring as shown in equations A.3 and A.4,

$$Ask = 1 * P(I) + \frac{1}{2}P(U), \tag{A.3}$$

$$Bid = 0 * P(I) + \frac{1}{2}P(U). \tag{A.4}$$

Inserting A.1 and A.2 into A.3 and A.4 renders,

$$Ask^* = \frac{1+\pi-\phi(1-\pi)}{2(1-\phi(1-\pi))}, \tag{A.5}$$

$$Bid^* = \frac{1 - \pi - \phi(1-\pi)}{2(1-\phi(1-\pi))}. \tag{A.6}$$

$$Spread^* = \frac{\pi}{1-\phi(1-\pi)}. \tag{A.7}$$

With this framework it is straightforward to show that undercutting behavior is inversely related to informed trading risk. To see this, consider that the probability of an undercutting run is the probability that at least one SLP receives a signal,ϕ, multiplied by the likelihood that the upcoming trade is uninformed,$(1-\pi)$,

$$P(UndercuttingRun) = \phi(1-\pi). \tag{A.8}$$

The derivative of this value with respect to informed trading risk is $\frac{\delta P(UndercuttingRun)}{\delta \pi} = -\phi$, which is always less than zero, confirming the inverse relation between undercutting activity and informed trading risk - i.e. when the risk of informed trading goes up, the likelihood of undercutting runs diminishes.

The model also produces an additional prediction: that the spread will be increasing in undercutting risk - i.e. liquidity gets worse as undercutting risk increases a result documented empirically(Foley et al. (2021), Foley et al. (2022)). To see this, consider that the spread from A.7 can be rewritten as,

$$Spread^* = \frac{\pi}{1 - P(UndercuttingRun)}. \tag{A.9}$$

$P(UndercuttingRun)$ is bounded by 0 and $1-\pi$ it is straightforward to see that as $P(UndercuttingRun)$ increases, so too does the bid ask spread. The spread is bounded on the top by the value of 1. Thus, the prevalence of undercutting runs can cause markets to fail if SLPs interact with too many of the uninformed trades.[39]

A.2 Is *QIDRes* a Stock Characteristic?

This section present evidence that $QIDRes$ is not persistent stock/firm characteristic. Panel A in Figure A.1 presents pairwise correlation coefficients between $QIDRes_{q-1}$, $QIDRes_{q-2}$ and an array of stock characteristics. $QIDRes$ is nearly orthogonal to all these stocks characteristics. Panel B present estimates of an AR(2) model that regresses $QIDRes_q$ on $QIDRes_{q-1}$ and $QIDRes_{q-2}$ using the panel of stock-quarter observations in our sample. $QIDRes$ exhibits no temporal persistence; if anything, it exhibit some degree of mean reversion, which consistent with its "residual" nature.

[39] In the case of no undercutting runs, i.e. $\phi = 0 \Leftrightarrow P(UndercuttingRun) = 0$, the spread will equal π, its minimum given the prevalence of informed traders in the market. In the other extreme case where $\phi = 1$, implying that all uninformed trades trigger undercutting runs, $P(UndercuttingRun) = 1 - \pi$, and the spread goes to 1 indicating a failed market.

Table A.1. Correlations Between Current $QIDRes$, Past $QIDRes$, and Stock Characteristics. Panel A presents pairwise correlations between variables used in asset pricing tests. These variables include our measures of informed trading from the two preceding quarters, i.e., $QIDRes_{j,q-1}$ and $QIDRes_{j,q-2}$, three-factor Fama-French betas ($\beta^{mkt}_{j,m-1}$, $\beta^{hml}_{j,m-1}$, $\beta^{smb}_{j,m-1}$), estimated using weekly observations from the two-year period ending in the final full week of month $m-1$, book-to-market ratio, ($\text{BM}_{j,m-1}$), natural log of market capitalization, ($\ln(\text{Mcap}_{j,m-12})$), dividend yield ($\text{DYD}_{j,m-1}$), defined as total dividends over the past 12 months divided by the share price at the end of month $m-1$, idiosyncratic volatility ($\text{IdVol}_{j,m-1}$), previous month's return ($RET_{j,m-1}$), preceding return from the prior 11 months ($RET_{j,(m-12,m-2)}$), previous quarter's fraction institutionally owned shares outstanding ($IOShr_{j,q-1}$), previous quarter's Herfindahl-Hirschman index for institutional ownership ($IOShrHHI_{j,q-1}$), and month $m-2$ share turnover ($TO_{j,m-2}$). Panel B presents estimates of the AR(2) models the regress $QIDRes_{j,q}$ on $QIDRes_{j,q-1}$ and $QIDRes_{j,q-2}$ using different specifications with and without double-clustered standard errors at year-quarter and stock levels. The sample includes NMS common shares from January 2010 to December 2019, excluding stocks whose previous month-end's closing price is below \$5 as well as stocks-dates for firms designated as treatment or control stocks during the SEC's Tick Size Pilot experiment.

Panel A: Correlations between current/past $QIDRes$ and stock characteristics

Variable		Variable index												
index		1	2	3	4	5	6	7	8	9	10	11	12	13
1	$QIDRes_{q-1}$													
2	$QIDRes_{q-2}$	−0.056												
3	β^{mkt}	0.009	−0.003											
4	β^{hml}	0.007	0.005	−0.03										
5	β^{smb}	0.030	0.025	0.12	0.15									
6	BM	0.059	0.048	−0.09	0.33	0.05								
7	ln(Mcap)	−0.044	−0.046	0.26	−0.10	−0.40	−0.27							
8	DYD	0.016	0.015	−0.13	0.10	−0.16	0.10	0.10						
9	Id. Vol.	0.057	0.027	0.14	−0.07	0.32	0.06	−0.31	−0.15					
10	RET_{-1}	0.007	0.013	0.00	0.00	0.00	−0.09	−0.02	0.01	0.03				
11	$RET_{(-12,-2)}$	−0.117	−0.118	−0.01	−0.09	−0.04	−0.25	−0.07	−0.08	−0.07	−0.03			
12	$IOShr$	−0.016	−0.034	0.29	−0.03	0.01	−0.20	0.41	−0.12	−0.08	0.00	−0.03		
13	$IOShr_HHI$	0.019	0.028	−0.18	0.02	0.06	0.18	−0.35	0.01	0.14	−0.01	−0.01	−0.60	
14	TO	0.06	0.02	0.35	−0.09	0.09	−0.10	0.21	−0.11	0.24	−0.01	0.02	0.31	−0.16

Panel B: AR(2) models of $QIDRes$

	(1)	(2)	(3)	(4)
Constant	0.085***	0.085**	0.087***	0.090***
	−33.5	−2.41	−37.24	−40.37
$QIDRes_{q-1}$	−0.064***	−0.064***	−0.078***	−0.11***
	[−19.16]	[−3.01]	[−3.70]	[−5.15]
$QIDRes_{q-2}$	−0.093***	−0.093**	−0.11***	−0.14***
	[−27.93]	[−2.54]	[−4.27]	[−5.78]
Quarter FE	No	No	Yes	Yes
Stock FE	No	No	No	Yes
Clustered Errors	N/A	Quarter & Stock	Quarter & Stock	Quarter & Stock
Observations	75,017	75,017	75,017	74,792

A.3 Modified Constructions of *QIDRes*

This section provides documents the robustness of our main findings to controlling for binding tick sizes and the effects of intraday volatility on undercutting. We construct two modified versions of $QIDRes$. The first modification uses equation (3) to fit parameters from the previous quarter, but

it defines $QIDResInt$ as follows

$$QIDResInt_{jt}^{q} = -\frac{QID_{jt}^{q} - \left(\widehat{a_j^{q-1}} + \widehat{b_j^{q-1}} \ln(PQSP)_{jt}^{q}\right)}{\widehat{a_j^{q-1}}} \tag{A.10}$$

wher $\widehat{a_j^{q-1}}$ is obtained from equation (3). This modification accounts from the potential cross-sectional variation in unconditional average undercutting. The second modification accounts for the possibility that liquidity providing algorithms with very short holding periods avoid undercutting in more volatile stocks/markets, for a any given level of information asymmetry. Hence, the first stage in this modification involves modeling QID as a function of both spreads and volatility. That is, we first fit

$$QID_{jt}^{q} = \alpha_j^q + \beta_j^q \ln(PQSP)_{jt}^{q} + \gamma_j^q qvol_{jt}^{q} + v_{jt}^{q}, \tag{A.11}$$

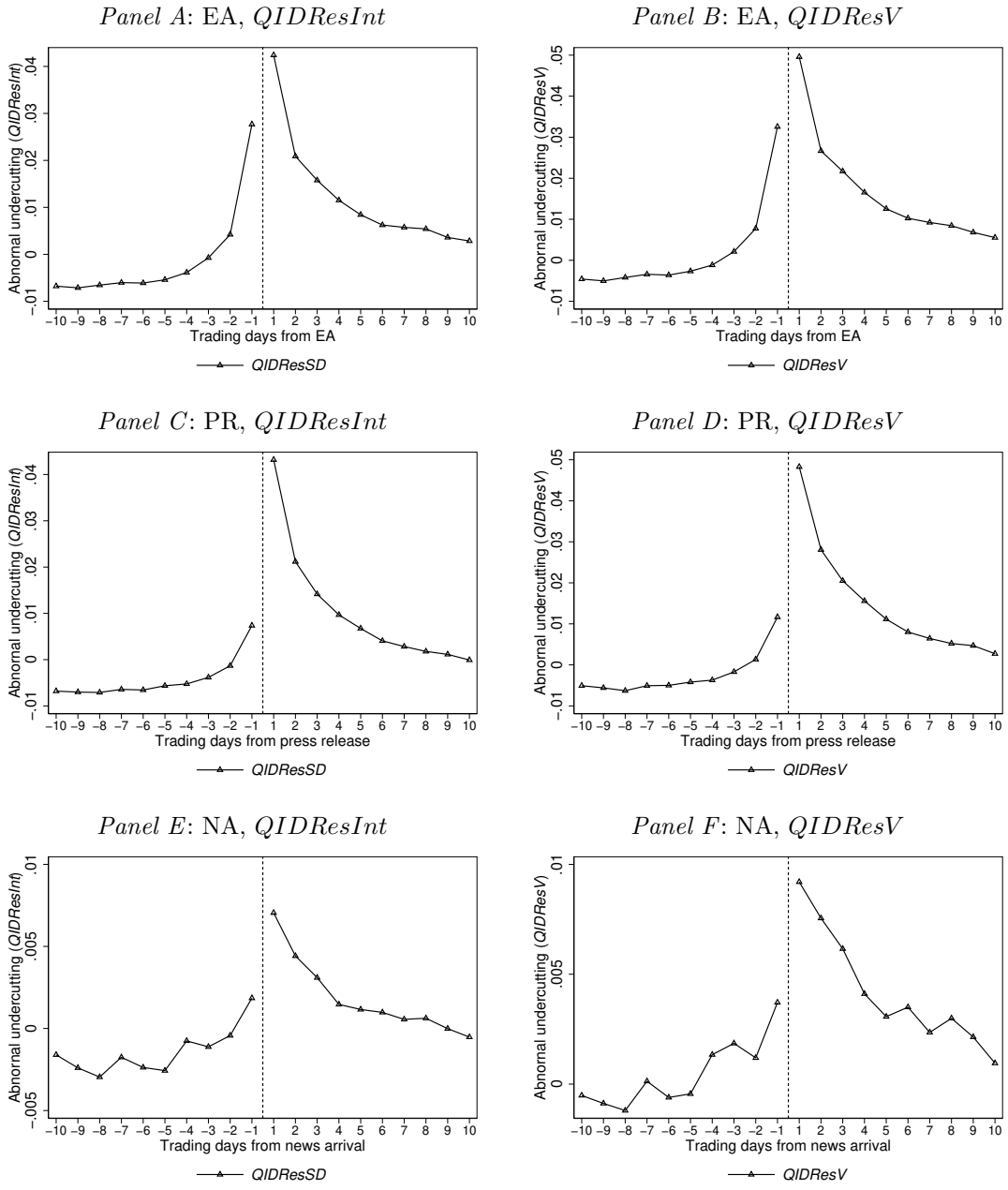
where $qvol_{jt}^{q}$ is the daily standard deviation of 1-minute quote-midpoint returns. Thus, a modified abnormal undercutting activity—that accounts for high-frequency volatility—for stock j on day t of quarter q is given by:

$$QIDResV_{jt}^{q} = -\frac{QID_{jt}^{q} - \left(\widehat{\alpha_j^{q-1}} + \widehat{\beta_j^{q-1}} \ln(PQSP)_{jt}^{q} + \widehat{\gamma_j^{q-1}} qvol_{jt}^{q}\right)}{\widehat{a_j^{q-1}}}. \tag{A.12}$$

Figure A.2 shows that $QIDResSD$ and $QIDResV$ behave qualitatively very similarly to the basking $QIDRes$ around major information events.

Figure A.2. Abnormal Undercutting Activity around Scheduled and Unscheduled Corporate Announcements: Robustness.

The figure presents alternative versions of abnormal undercutting activity, $QIDResSD$ and $QIDResV$, around earnings announcements (EA), unscheduled press releases (PR), and news arrivals not associated with any identified event (NA). The sample includes all NMS-listed common stocks between Jan, 2010 through Dec, 2019 with previous quarter-end's share prices of at least \$5. Earnings announcement dates are obtained from COMPUSTAT; unscheduled press release dates and news arrivals not associated with any identified event are obtained from Ravenpack.



AI-Powered Trading, Algorithmic Collusion, and Price Efficiency

Winston Wei Dou Itay Goldstein Yan Ji *

March 10, 2024

Abstract

The integration of algorithmic trading and reinforcement learning, known as AI-powered trading, has significantly impacted capital markets. This study utilizes a model of imperfect competition among informed speculators with asymmetric information to explore the implications of AI-powered trading strategies on speculators' market power, information rents, price informativeness, market liquidity, and mispricing. Our results demonstrate that informed AI speculators, even though they are "unaware" of collusion, can autonomously learn to employ collusive trading strategies. These collusive strategies allow them to achieve supra-competitive trading profits by strategically under-reacting to information, even without any form of agreement or communication, let alone interactions that might violate traditional antitrust regulations. Algorithmic collusion emerges from two distinct mechanisms. The first mechanism is through the adoption of price-trigger strategies ("artificial intelligence"), while the second stems from homogenized learning biases ("artificial stupidity"). The former mechanism is evident only in scenarios with limited price efficiency and noise trading risk. In contrast, the latter persists even under conditions of high price efficiency or large noise trading risk. As a result, in a market with prevalent AI-powered trading, both price informativeness and market liquidity can suffer, reflecting the influence of both artificial intelligence and stupidity.

Keywords: Reinforcement learning, AI collusion, Homogenization, Self-confirming equilibrium, Asymmetric information, Price informativeness, Market liquidity.
JEL Classification: D43, G10, G14, L13.

*Dou: University of Pennsylvania (wdou@wharton.upenn.edu) and NBER; Goldstein: University of Pennsylvania (itayg@wharton.upenn.edu) and NBER; Ji: Hong Kong University of Science and Technology (jiy@ust.hk). We thank Snehal Banerjee, Hui Chen, Antoine Didisheim, Itamar Drechsler, Slava Fos, Cary Frydman, Paolo Fulghieri, Vincent Glode, Joao Gomes, Mark Grinblatt, Tim Johnson, Chris Jones, Scott Joslin, Larry Harris, Zhiguo He, David Hirshleifer, Jerry Hoberg, Leonid Kogan, Pete Kyle, Tse-Chun Lin, Deborah Lucas, Ye Luo, Semyon Malamud, Andrey Malenko, George Malikov, Albert Menkveld, Jonathan Parker, Lasse Pedersen, Josh Pollet, Paul Romer, Nick Roussanov, Tom Sargent, Antoinette Schoar, Daniel Sokol, Rob Stambaugh, Eric Talley, Anton Tsoy, Liyan Yang, Jiang Wang, Neng Wang, Xiaoyan Zhang, and seminar and conference participants at ASU Sonoran Winter Finance Conference, Boston College, CUFE, Fudan, George Mason, HKU, HKUST, Jackson Hole Finance Conference, Johns Hopkins Carey Finance Conference, Melbourne Asset Pricing Meeting, MIT, Nordic Fintech Symposium, NYU/Penn Law and Finance Conference, Olin Finance Conference at WashU, PKU/PHBS Sargent Institute Macro-Finance Workshop, QES Global Quant and Macro Investing Conference, QRFE Workshop on Market Microstructure, Fintech and AI, SHUFE, Toronto Macro/Finance Conference, Tsinghua PBCSF, UIUC, University of Macau, University of Toronto, USC, and Wharton for their comments. We are grateful for the insightful discussions with Jacob Yunger and other colleagues at the Financial Industry Regulatory Authority (FINRA). Dou is grateful for the financial supports from the Golub Faculty Scholar Award at Wharton.

1 Introduction

The integration of algorithmic trading with reinforcement learning (RL) algorithms, often termed AI-powered trading, poses new regulatory challenges and has the potential to fundamentally reshape capital markets.[1] With Nasdaq receiving SEC approval for an RL-based, AI-driven order type, the momentum for AI integration in trading continues to build. Leading digital trading platforms like MetaTrader are endorsing RL-based AI trading bots, and major hedge funds such as Two Sigma, along with investment powerhouses like Blackrock and J.P. Morgan, are adopting AI technologies. This trend has led policymakers, regulators, and financial market supervisors worldwide to make AI a regulatory priority. Their focus is now on understanding how AI is applied in financial markets, its potential implications, and the risks of unintended systemic impacts.[2]

In particular, the U.S. Security and Exchange Commission (SEC) has recently cautioned against the possibility of AI destabilizing the global financial market if big tech-based trading companies monopolize AI development and applications within the financial sector. The SEC points out that the real challenge is fostering competitive and efficient markets amidst the swift adoption of AI technologies, as AI might be optimized to benefit sophisticated speculators at the expense of other investors, potentially compromising competition and market efficiency. Notably, SEC Chair Gary Gensler has emphasized this concern, noting that there is evidence of machines in high-frequency trading starting to exhibit cooperative behavior independently of human intervention or interaction.

Promoting competition in financial markets is a primary objective of the SEC and similar regulatory bodies worldwide. As such, the potential for collusion among AI trading algorithms is a significant concern for these organizations. However, the underlying scientific and economic principles of such "cooperation" among autonomous AI algorithms remain unclear, not to mention how it might affect competition, price formation, and overall market efficiency. In this paper, we demonstrate that "AI collusion" – where autonomous, self-interested algorithms independently learn to coordinate without any explicit agreement, communication, or intention – can robustly occur via one of two distinct mechanisms. These mechanisms are collusion through price-trigger strategies or homogenized learning biases, and their emergence is contingent on the condition of the trading environment. We find that AI collusion impairs competition and thereby market efficiency, leading to reduced liquidity, less informative pricing, and increased mispricing.

The economics of AI collusion in trading can be intuitively understood as follows. On one hand, consider a trading environment where subgame perfect collusive Nash equilibria theoretically exist for rational-expectations agents, supported by price-trigger strategies as introduced by Green and Porter (1984). In this environment, even without direct monitoring of trading behaviors, agents can develop collusive incentives. This is achieved by allowing non-collusive competition to

[1] Traditional algorithmic trading is based on rigid, human-defined trading protocols that are hardcoded.

[2] For example, the SEC proposed novel rules concerning the application of AI technologies (SEC, 2023). Additionally, the European Securities and Markets Authority (ESMA) published a report on AI utilization within EU securities markets (Bagattini, Benetti and Guagliano, 2023).

occur when market prices diverge from the expected collusive level beyond a certain threshold. If the trading environment is not overly disrupted by noise trading flows, AI algorithms have the capacity to interact and learn, ultimately achieving a steady state, within which they engage in collusive trading based on a price-trigger strategy, even though they might not achieve the most profitable collusive equilibrium, due to a learning bias. On the other hand, in a trading environment where subgame perfect collusive Nash equilibria do not theoretically exist, AI algorithms cannot learn to sustain collusion through price-trigger strategies. Instead, they may converge to a steady state characterized by a self-conforming equilibrium, as introduced by Fudenberg and Levine (1993). This equilibrium concept, weaker than Nash equilibrium, allows for potentially incorrect or biased off-equilibrium beliefs, tightly aligned with the learning and trading behaviors of AI algorithms. Beliefs may be accurate along the equilibrium path, as this is more commonly observed, but can be inaccurate off the equilibrium path, unless there is sufficient exploration of non-optimal actions (e.g., Fudenberg and Kreps, 1988, 1995; Cho and Sargent, 2008). Crucially, these incorrect off-equilibrium beliefs are not necessarily inconsistent with observed outcomes along the equilibrium path.

Notably, AI algorithms are distinct from human traders in that they do not simply mimic human behavior. Traditional theories and experimental studies about human behavior are insufficient for understanding AI traders' behavior and the equilibria they might form. This is because AI possesses a fundamentally different form of intelligence. Unlike humans, AI decision-making is not influenced by emotions or logical thinking; rather, it is driven primarily by pattern recognition and is not affected by higher-order beliefs. Therefore, understanding the dynamics of capital markets with the prevalence of AI-powered trading algorithms requires insights into algorithmic behavior akin to the "psychology" of machines (Goldstein, Spatt and Ye, 2021), in a similar vein to how decision theory and psychology literature have provided insights into modeling human behavior in an economic context. In this paper, we conduct an experimental study to examine the behavior of AI algorithms endowed with private information. Following the tradition of experimental research, our study is qualitative and intended as a proof-of-concept demonstration.

In this paper, we adopt a streamlined theoretical framework as our laboratory. Building upon the seminal work of Kyle (1985), we extend this framework in three novel ways. First, our model incorporates multiple informed speculators within a repeated-trading context. Second, we introduce a continuum of atomistic long-term preferred-habitat investors, who together create a collective downward-sloping demand curve. Third, we expand the role of the market maker to consider both inventory costs and pricing errors, thereby extending beyond the original model's focus on pricing errors alone, as in Kyle (1985). Within each trading period, agents execute a single transaction. The sequence of events for each period unfolds as follows: Initially, the fundamental value of the asset is determined. Subsequently, a continuum of noise traders collectively places an order flow, which is independent of the asset's fundamental value. The variance of such an aggregate noise trading flow encapsulates the noise trading risk (Long et al., 1990). This noise trading risk is a crucial characteristic of the trading environment. Each oligopolistic informed

speculator is aware of the fundamental value but remains uninformed about the noise trading flow when determining his or her optimal trading strategy. The market maker, in turn, sets the market price with the goal of minimizing the weighted average of inventory costs and pricing errors. In doing so, the market maker also takes into account the price elasticity of the preferred-habitat investors' demand. This price elasticity represents another critical characteristic of the trading environment.

In our experimental study, we position our subjects – AI algorithms – within the laboratory framework we have established. Specifically, we substitute the rational-expectations informed speculators and market maker as in Kyle (1985)'s model with Q-learning algorithms. These algorithms are tasked with learning and guiding the real-time trading decisions. Known for their simplicity, transparency, and economic interpretability, Q-learning algorithms provide a foundational basis for various RL procedures that have significantly advanced the AI domain. Our theoretical framework, coupled with simulation-based experiments that blend theoretical rigor with practical relevance, serves as a laboratory for examining the impact of AI-powered trading strategies. Specifically, it allows us to investigate their influence on the market power of informed AI speculators, as well as on the price formation process, including implications for market liquidity, price informativeness, and mispricing within financial markets.

To ascertain whether informed AI speculators' behavior exhibits collusion sustained by price-trigger strategies due to the intelligence of the algorithms, our analysis starts with examining the theoretical properties of tacit collusion that can be maintained through price-trigger strategies. This analysis is based on the assumption that both the informed speculators and the market maker operate under rational expectations and have a thorough understanding of the preferred-habitat demand curve. We examine how tacit collusion varies across different trading environments. This includes variations in the price elasticity of preferred-habitat investors and noise trading risk levels, as well as variations in the number of informed speculators and their time discount rates. This theoretical investigation enables us to establish a baseline understanding of collusive behavior in the presence of asymmetric information and the endogenous strategic pricing rules of the market maker. Importantly, it lays the groundwork for our experimental study on the AI trading behavior, wherein we assess whether the observed collusion of informed AI speculators aligns with the theoretical predictions under the assumption of rational expectations and perfect knowledge of the preferred-habitat demand curve.

As a noteworthy theoretical contribution, we establish a novel result on the impossibility of collusion under information asymmetry. We demonstrate that informed speculators are unable to achieve collusive outcomes through price-trigger strategies in certain conditions. This includes scenarios where market prices are already efficient, accurately reflecting the asset's fundamental value, especially when the preferred-habitat investor has high price elasticity of demand, thereby playing a minimal role in price formation. Another scenario precluding collusion is when the noise trading risk is excessively high. This novel result illuminates a mechanism distinct from existing theories on the impossibility of collusion under information asymmetry in the context of product market competition (Abreu, Milgrom and Pearce, 1991; Sannikov and Skrzypacz, 2007).

Intuitively, sustaining price-trigger collusion requires two conditions: first, monitoring necessitates high price informativeness, and second, maintaining informational rents requires a low price impact of informed trading. These two conditions cannot be simultaneously met when price efficiency or noise trading risk is high.

Furthermore, as an additional theoretical contribution, we illustrate that in scenarios where the preferred-habitat investor, exhibiting low price elasticity of demand, significantly influences price formation, market prices can become inefficient. In such cases, tacit collusion among informed speculators can be sustained through price-trigger strategies. The success of these strategies is contingent on the number of informed speculators and the level of noise trading risk in the market. We find that price-trigger strategies can only sustain collusion in markets with a low level of noise trading risk and a few informed speculators. Additionally, we show that collusion capacity increases, market liquidity decreases, price informativeness decreases, and mispricing increases, when the number of informed speculators drops, the level of noise trading risk decreases, or the subjective rate of time preference (i.e., "impatience") declines.

Having established the baseline theoretical results, we now turn back to our simulation experiments, which involve informed AI speculators using Q-learning algorithms. These simulations provide compelling evidence that these AI speculators can robustly collude and secure supra-competitive profits by strategically manipulating excessively low order flows relative to their information about the asset's fundamental value. This occurs without any form of agreement or communication that would typically be seen as an antitrust infringement. The cruciality, and even necessity, of communication in collusion among humans is well-documented in the literature of experimental economics. To underscore the concept of AI collusion in our simulations, we deliberately employ relatively simple Q-learning algorithms that base their decisions solely on one-period-lagged asset prices as state variables. This approach is intentional, omitting more extensive lagged data, such as information on lagged self-order flows or multiple-period-lagged asset prices. Although the trading environment is excessively complex relative to the simple AI algorithms used, our simulation results remarkably indicate that informed AI speculators can intelligently form collusion across diverse trading environments. Specifically, in environments characterized by low price efficiency and low noise trading risk, the behavior of algorithmic collusion aligns with the predictions of our rational-expectations model, where informed AI speculators are capable of learning price-trigger strategies to sustain collusion. Conversely, in environments with high price efficiency or high noise trading risk, informed AI speculators are unable to learn price-trigger strategies, consistent with our rational-expectations model predictions. However, strikingly, going beyond the rational-expectations model, our simulation results demonstrate that informed AI speculators can still collude and achieve supra-competitive profits by manipulating excessively low order flows, even without relying on traditional price-trigger strategies, provided they use equally naive algorithms. These findings suggest the existence of two distinct mechanisms underpinning algorithmic collusion, depending on the trading environment.

Finally, we elaborate further on the two distinct mechanisms behind AI collusion across various trading environments. The first mechanism, known as "algorithmic collusion through

price-trigger strategies," involves a form of collusion driven by "artificial intelligence." In this scenario, informed AI speculators have the capability to learn and implement price-trigger strategies effectively. This price-trigger strategy enables the AI speculators to sustain collusion and reach a steady state closely resembling a subgame perfect Nash equilibrium. Such a scenario can only occur if both price efficiency and noise trading risk are low. Leveraging simulation experiments, we provide direct evidence that sizable price deviations trigger aggressive trading flows similar to those in a non-collusive Nash equilibrium, which diminishes the trading profits of all informed AI speculators. While the underlying mechanisms through which AI speculators learn to conduct the price-trigger trading strategy, thereby achieving algorithmic collusion, may differ from those behind how humans would learn to coordinate using price-trigger trading strategies, the resulting patterns exhibit notable similarities. At the heart of these mechanisms, whether involving AI or human speculators, the threat of punishment effectively acts as a deterrent, discouraging individual speculators from violating the collusive agreement. Closely aligned with the theoretical predictions of a collusive Nash equilibrium sustained by price-trigger strategies with rational-expectations agents, as the number or impatience of speculators decreases, the extent of achievable collusion increases. This leads to reduced market liquidity, diminished price informativeness, and increased mispricing.

Importantly, algorithmic collusion through price-trigger strategies introduces a paradoxical situation concerning price informativeness. This paradox arises because such collusion relies on the informativeness of prices – specifically, the ability of an informed AI speculator to infer the order flows of other informed AI speculators from observed prices. High price informativeness typically characterizes environments where prices are sensitive to new information about the fundamental value of the asset and are not predominantly driven by noise trading flows. However, in such environments, the heightened price informativeness actually facilitates informed AI speculators in discerning each other's order flows, thereby strengthening collusion among them. This stronger collusion, in turn, endogenously compromises price informativeness by distorting the information content of prices – specifically, it reduces the responsiveness of prices to new information about the fundamental value of the asset. Consequently, in a capital market dominated by AI-powered trading, where algorithmic collusion through price-trigger strategies is prevalent, achieving perfect price informativeness becomes unattainable.

The second mechanism, known as "algorithmic collusion through homogenized learning biases," involves a form of collusion driven by "artificial stupidity." Despite the learning biases originating from intrinsic imperfections in the algorithms, informed AI speculators might still achieve and sustain supra-competitive profits. This can occur when they use similar foundational models that have homogenized learning biases, effectively forming a kind of hub-and-spoke conspiracy.[3] Johnson and Sokol (2021) emphasize the prevalence of this type of AI collusion in the context of e-commerce platforms, observing that many retailers adopt similar or even identical AI

[3]In the context of product market competition, the term "hub-and-spoke conspiracy" is a metaphor used to describe a cartel that includes a firm at one level of a supply chain, typically a supplier, acting as the "hub" of a wheel. Vertical agreements down the supply chain represent the "spokes." This common supplier facilitates the implicit coordination among its customers.

pricing algorithms. Specifically, anti-competitive effects may emerge when multiple competitors use the same AI pricing algorithm supplied by a common service provider, who serves as the hub. In the financial markets, informed speculators often rely on similar foundational models for their AI-powered trading systems. This practice, whether intentional or not, can result in a significant degree of homogenization, a phenomenon documented by Bommasani et al. (2022), among others. In the context of RL learning, the emergence of a learning bias is directly linked to inconsistencies in statistical learning. These inconsistencies often stem from over-exploitation and insufficient exploration, especially when the noise trading risk is excessive. This inherently biased algorithm leads informed speculators to under-react to their private information in their learned trading strategies, compared to the optimal strategy in a non-collusive equilibrium setting. Consider a scenario in which an RL-based AI speculator explores a trading strategy that aggressively responds to private information and receives a positive signal about the asset's fundamental value. If a substantial and positive noise trading flow occurs, this could result in significant losses for the AI speculator. Consequently, the RL algorithm is unlikely to revisit and update its understanding of this state-strategy pair sufficiently, consistently deeming this strategy as suboptimal for the given state. This means the initial adverse effect on the Q function at the state-strategy pair due to such a shock is unlikely to be mitigated in subsequent iterations. Conversely, if a substantial and negative noise trading flow occurs, it could lead to significant gains for the AI speculator. In this fortunate case, the RL algorithm is more likely to revisit and thoroughly understand the performance of this state-strategy pair, adequately exploiting it, and thus, the initial beneficial effect on the Q function at this pair may be averaged out, which even leads to accurate estimations of Q function at this state-strategy pair. Such severe asymmetric learning outcomes from large positive and negative noise trading flows can lead AI speculators to generally under-react to their private information in their learned trading strategies.

Such under-reaction can lead to the realization of supra-competitive profits, a scenario that is more likely to occur with widespread homogenization in the algorithms adopted by AI speculators. This homogenized learning bias steers informed AI speculators toward a steady state where trading behaviors can be accurately characterized by a self-conforming equilibrium, as introduced by Fudenberg and Levine (1993). In contrast to the Nash equilibrium, the self-conforming equilibrium is weaker because it permits players to hold incorrect (or biased) off-equilibrium beliefs. This concept of equilibrium is motivated by the idea that noncooperative equilibria should be interpreted as outcomes of a learning process, where players form beliefs based on their past experiences. While beliefs can generally be correct along the equilibrium path of play due to its frequent observation, they are not necessarily correct off the equilibrium path. Correct beliefs off the equilibrium path require players to engage in sufficient experimentation with non-optimal actions, as suggested in works by Fudenberg and Kreps (1988), Fudenberg and Kreps (1995), and Cho and Sargent (2008).

Although adopting superior algorithms can disrupt the collusion created by homogenized learning biases, it is likely that no AI speculator would choose to gain an advantage by using superior algorithms due to the nature of AI collusion. Intuitively, if one speculator adopts a

superior algorithm, it could render the trading strategies of other AI speculators unprofitable, thereby compelling them to adopt equally or more advanced algorithms. This could spark a race towards algorithmic advancement, ultimately leading to an equilibrium where trading profitability is minimal for every AI speculator. Consequently, AI speculators autonomously learn to adopt similarly basic algorithms in equilibrium. To illustrate this point, we consider a simple extension of the baseline Q-learning algorithms, wherein informed AI speculators are able to learn both the key parameter that governs the sophistication of their Q-learning algorithms and their trading strategies based on the AI-chosen Q-learning algorithm. Our simulation experiments robustly demonstrate that informed AI speculators may collectively opt for less advanced algorithms. This occurs despite the potential for increased self-profit that could come from unilaterally choosing a more advanced algorithm while others' algorithms remain fixed.

These two types of AI collusion, while both generating supra-competitive trading profits, can exhibit opposite collusive behaviors as trading environments evolve. On one hand, akin to AI collusion through price-trigger strategies (referred to as "artificial intelligence"), a decrease in the number of speculators leads to increased potential for collusion. This, in turn, results in reduced market liquidity, diminished price informativeness, and increased mispricing. On the other hand, contrary to AI collusion through price-trigger strategies, an increase in speculator impatience, or an elevation in noise trading risk, enlarges the potential for collusion due to a more pronounced homogenized learning bias. (termed "AI collusion through artificial stupidity"). This also leads to reduced market liquidity, diminished price informativeness, and increased mispricing. Notably, unlike the scenario with price-trigger strategies, in the case of AI collusion through homogenized learning biases, an increase in noise trading risk leads to an increase, rather than a decrease, in trading profitability for AI speculators based on their private information.

Related Literature. The topic of autonomous cooperation among multiple Q-learning agents in repeated games has garnered significant attention from researchers in the artificial intelligence and computer science community over the past decades (e.g., Sandholm and Crites, 1996; Tesauro and Kephart, 2002). Given the widespread adoption of AI technologies in pricing decisions across various marketplaces, Waltman and Kaymak (2008) demonstrate that Q-learning firms typically learn to attain supra-competitive profits in repeated Cournot oligopoly games with homogeneous products, even though a perfect cartel is usually unattainable. Klein (2021) also examines the strategies employed by algorithms in a context where firms selling homogeneous products alternate in adjusting prices to support supra-competitive profits. Recently, in a noteworthy contribution, Calvano et al. (2020) study collusion by AI algorithms in a logit model of differentiated products, not only uncovering the existence of supra-competitive profits but also pinpointing how algorithms might learn to sustain collusive outcomes through grim-trigger strategies. Expanding upon this, our paper extensively broadens the AI experimental framework, moving from a scenario of perfect information and a static demand curve to one imbued with asymmetric information and a strategically-determined demand scheme. We characterize the various types of AI algorithmic collusion, whether occurring through price-trigger strategies or

through learning biases and homogenization, across diverse market environments.

Inspired by the simulation-based studies on AI algorithmic collusion, empirical research has also emerged, demonstrating that the use of AI algorithms in setting product prices can lead to collusion, resulting in heightened supra-competitive prices (e.g., Assad et al., 2023). Additionally, recent studies have started to focus on policy interventions aiming to obstruct the ability of algorithms to collude, thereby ensuring the maintenance of competitive prices. Specially, based on simulation-based studies, Johnson, Rhodes and Wildenbeest (2023) show that platform design can benefit consumers and the platform. However, achieving these gains may require policies that condition on past behavior and treat sellers in a non-neutral fashion. Harrington (2018) delves into critical policy issues surrounding the definition of collusion, such as whether collusion should necessarily entail an explicit agreement among conspirators, or if it might be more aptly defined as the maintenance of elevated prices, sustained by a reward-and-punishment scheme.

Our paper is among the first to investigate how the widespread adoption of AI-powered trading strategies might affect capital markets. The work of Colliard, Foucault and Lovo (2022) is closely related to our research, as it also explores the implications of interactions among Q-learning algorithms in capital markets. However, there are notable differences in focus between their work and ours. Specifically, Colliard, Foucault and Lovo (2022) focuses on AI-powered oligopolistic market makers, while our study concentrates on AI-powered oligopolistic informed speculators who face perfectly competitive market makers. Their research illuminates the strategies that AI market makers would adopt by leveraging their market power. In contrast, our paper explores the dynamics and implications of algorithmic collusion among AI-powered informed speculators, particularly in the context of preferred-habitat long-term investors and perfectly competitive market makers. We provide novel insights into the strategies of informed AI speculators on how they leverage private information and maximize profits through autonomously forming collusion via distinct mechanisms.

2 AI-Powered Trading Algorithms

The traditional algorithmic trading system executes orders according to protocols predefined by human quantitative strategists. In contrast, AI-powered trading employs RL algorithms to dynamically adjust and optimize trading strategies in real time.

The RL algorithm, a pivotal technique in AI, forms the foundation of numerous successful AI algorithms, like "AlphaGo," demonstrating the superiority of RL-backed AI over human cognitive abilities in areas such as securities trading and other complex tasks. RL algorithms are model-free machine learning techniques that learn autonomously through trial-and-error experimentation, without relying on two common assumptions: first, that the multi-agent system is on an equilibrium path, and second, that agents have knowledge of the true state and payoff distributions at equilibrium. The fundamental rationale behind RL algorithms centers on the principle that actions yielding higher rewards historically are more likely to be selected in the future, compared to those that have led to lesser rewards. By interacting with its environment and

experimenting with different actions, the agent incrementally learns an optimal policy. Through continuous rounds of exploration and experimentation, it refines its strategy to prefer actions that offer the greatest long-term benefits, even without any knowledge of the environment beforehand. This iterative process enables the agent to progressively enhance its decision-making approach, consistently steering towards actions that maximize the cumulative rewards based on its gathered experiences.

While RL encompasses different variants (e.g., Watkins and Dayan, 1992; Sutton and Barto, 2018), we choose to focus on Q-learning for several reasons. First, Q-learning serves as a foundational framework for numerous RL algorithms, upon which many recent AI breakthroughs are built. However, it is important to note that AI trading algorithms currently in use may not exclusively rely on Q-learning principles. Second, Q-learning holds substantial popularity among computer scientists in practical applications. Third, Q-learning algorithms possess simplicity and transparency, offering clear economic interpretations, in contrast to the black-box nature of many machine learning and AI algorithms. Finally, Q-learning shares a common architecture with more sophisticated RL algorithms.

In the remainder of this section, we will concentrate on a multi-agent system, detailing the Bellman equation for each agent, and describe the Q-learning algorithm that an agent employs. This discussion will cover how each agent iteratively updates its Q-function and strategy based on the received rewards, thereby optimizing its long-term outcomes through the Q-learning algorithm.

2.1 Bellman Equation and Q-Function

In a multi-agent Markov decision process environment, there are I agents, indexed by $i = 1, \cdots, I$. The state of the environment is represented by a Markov process, denoted by s. Each agent makes decisions based on the current state, which in turn evolves partly due to the collective actions of all agents within the system. Agent i's intertemporal optimization is characterized by the Bellman equation and solved recursively via dynamic programming:

$$V_i(s) = \max_{x_i \in \mathcal{X}} \left\{ \mathbb{E}\left[\pi_i | s, x_i\right] + \rho \mathbb{E}\left[V_i(s') | s, x_i\right] \right\}, \tag{2.1}$$

where $x_i \in \mathcal{X}$ is action of agent i, with $\mathcal{X}$ denoting the set of available actions, π_i is the payoff received by agent i, which may be influenced by the actions of other agents, and $s, s' \in S$ represent the states in the current and the next period, respectively, with S denoting the set of states. In general, s and s' can depend on agent i's individual characteristics and private information. However, for our purpose of illustration, it is sufficient to concentrate on the simple setting where the same state applies uniformly to all agents in the system. The first term on the right-hand side, $\mathbb{E}\left[\pi_i | s, x_i\right]$, is agent i's expected payoff in the current period, and the second term, $\rho \mathbb{E}\left[V_i(s') | s, x_i\right]$, is agent i's continuation value, with the parameter ρ capturing the subjective rate of time preference.

The Bellman equation (2.1) represents the recursive formulation of dynamic control problems (e.g., Bellman, 1954; Ljungqvist and Sargent, 2012). It focuses on the equilibrium path, and thus the optimal value function $V_i(s)$ depends solely on the state variable s. In contrast to focusing solely on the equilibrium path, the Q function, denoted by $Q_i(s, x_i)$, extends the optimal value function to include the values of each state-action pair. This captures scenarios (or counterfactuals) that occur off the equilibrium path. By definition, the value of $Q_i(s, x_i)$ is the same as that in the curly brackets of the Bellman equation (2.1):

$$Q_i(s, x_i) = \mathbb{E}\left[\pi_i | s, x_i\right] + \rho \mathbb{E}\left[V_i(s') | s, x_i\right]. \tag{2.2}$$

Intuitively, the Q-function value, $Q_i(s, x_i)$, can be interpreted as the quality of action x_i in state s. The optimal value of a state, $V_i(s)$, is the maximum of all the possible Q-function values of state s. That is, $V_i(s) \equiv \max_{x' \in \mathcal{X}} Q_i(s, x')$. By substituting $V_i(s')$ with $\max_{x' \in \mathcal{X}} Q_i(s', x')$ in equation (2.2), we can establish a recursive formula for the Q-function as follows:

$$Q_i(s, x_i) = \mathbb{E}\left[\pi_i | s, x_i\right] + \rho \mathbb{E}\left[\max_{x' \in \mathcal{X}} Q_i(s', x') \middle| s, x_i\right]. \tag{2.3}$$

When both $|S|$ and $|\mathcal{X}|$ are finite, the Q-function can be represented as an $|S| \times |\mathcal{X}|$ matrix, which is often referred to as the Q-matrix.

2.2 Q-Learning Algorithm

If agent i possessed knowledge of its Q-matrix, determining the optimal actions for any given state s would be straightforward. In essence, the Q-learning algorithm is a method to estimate the Q-matrix in environments where the underlying distribution $\mathbb{E}\left[\cdot | s, x_i\right]$ is unknown and there are limited observations for off-equilibrium pairs (s, x_i) in the data. The Q-learning algorithm addresses both challenges concurrently: it employs Monte Carlo methods to estimate the underlying distribution $\mathbb{E}[\cdot | s, x_i]$ based on the law of large numbers, while at the same time, conducts trial-and-error experiments to produce off-equilibrium counterfactuals.

The iterative experimentation starts from an arbitrary initial Q-matrix of agent i, denoted by $\widehat{Q}_{i,0}$, and updates the estimated Q-matrix $\widehat{Q}_{i,t}$ recursively. The learning equation governing this update is as follows:

$$\widehat{Q}_{i,t+1}(s_t, x_{i,t}) = (1-\alpha) \underbrace{\widehat{Q}_{i,t}(s_t, x_{i,t})}_{\text{Past knowledge}} + \alpha \underbrace{\left[\pi_{i,t} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t}(s_{t+1}, x')\right]}_{\text{Present learning based on a new experiment}}, \tag{2.4}$$

where $\alpha \in [0, 1]$ captures the forgetting rate, s_t is the state that the iteration t concentrates on, s_{t+1} is randomly drawn from the Markovian transition probabilities conditional on s_t, $\widehat{Q}_{i,t}(s, x)$ is the estimated Q-matrix of agent i in the t-th iteration, and $\pi_{i,t}$ is the payoff in the t-th iteration, corresponding to agent i's choice of action $x_{i,t}$.

Equation (2.4) indicates that for agent i in the t-th iteration, only the value of the estimated

Q-matrix $\widehat{Q}_{i,t}(s,x)$ corresponding to the state-action pair $(s_t, x_{i,t})$ is updated to $\widehat{Q}_{i,t+1}(s_t, x_{i,t})$. All other state-action pairs remain unchanged. In other words, $\widehat{Q}_{i,t+1}(s,x) = \widehat{Q}_{i,t}(s,x)$ for cases where $s \neq s_t$ or $x \neq x_{i,t}$. The updated value $\widehat{Q}_{i,t+1}(s_t, x_{i,t})$ is computed as a weighted average of accumulated knowledge based on the previous experiments, $\widehat{Q}_{i,t}(s_t, x_{i,t})$, and learning based on a new experiment, $\pi_{i,t} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t}(s_{t+1}, x')$. A key distinction between the Q-learning recursive algorithm (2.4) and the Bellman recursive equation (2.1) lies in how they handle expectations. Q-learning algorithm (2.4) does not form expectations about the continuation value because the Markovian transition probabilities from s_t to s_{t+1} are unknown. Instead, it directly discounts the continuation value associated with the randomly realized state s_{t+1} in the $(t+1)$-th iteration.

It is crucial to note that the forgetting rate α plays a significant role in the Q-learning algorithm, balancing past knowledge against present learning based on a new experiment. A higher α not only indicates a greater impact of present learning on the Q-value update but also implies that the algorithm forgets past knowledge more quickly, potentially leading to biased learning. To elaborate, let τ be the number of times that the Q-value of the state-action pair (s,x) has been updated in the past. We derive in Appendix G.1 that as $\tau \to \infty$, the Q-value of (s,x) is as follows:

$$\widehat{Q}_{i,t(\tau)}(s,x) \approx \sum_{h=0}^{\tau-1} \alpha(1-\alpha)^h \left[\pi_{i,t(\tau-h)} + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(\tau-h)}(s_{t(\tau-h)+1}, x') \right], \tag{2.5}$$

where $t(h)$ represents the period in which the Q-value of (s,x) receives the h-th update. Clearly, when α is not close to 0, the weights given by $\alpha(1-\alpha)^h$ decay so rapidly with τ that it jeopardizes the applicability of the law of large number. When the underlying environment has randomness, a sufficiently small value of α is crucial for ensuring small learning biases. Otherwise, the law of large numbers may fail, leading to biased estimation for the underlying distribution $\mathbb{E}[\cdot|s, x_i]$. However, a smaller value of α requires more iterations for the algorithm to converge, and thus greater computational costs. Moreover, if α is excessively small relative to the decaying speed of the exploration rate ε_t in equation (2.6), biased learning may arise due to insufficient exploration.

2.3 Experimentation

Conditional on the state variable s_t, agent i chooses its action $x_{i,t}$ in two experimentation modes, exploitation and exploration, as follows:

$$x_{i,t} = \begin{cases} \operatorname{argmax}_{x \in \mathcal{X}} \widehat{Q}_{i,t}(s_t, x), & \text{with prob. } 1-\varepsilon_t, \quad \text{(exploitation)} \\ \widetilde{x} \sim \text{uniform distribution on } \mathcal{X}, & \text{with prob. } \varepsilon_t. \quad \text{(exploration)} \end{cases} \tag{2.6}$$

To determine the mode, we employ the simple ε-greedy method. As outlined in equation (2.6), during the t-th iteration, agent i engages in the exploration and exploitation modes with exogenous probabilities ε_t and $1-\varepsilon_t$, respectively. In the exploitation mode, agent i chooses its action to maximize the current state's Q-value, given by $x_{i,t} = \operatorname{argmax}_{x \in \mathcal{X}} \widehat{Q}_{i,t}(s_t, x)$. Conversely, in the exploration mode, agent i randomly chooses its action $\widetilde{x}$ from the set of all possible values in $\mathcal{X}$,

each with equal probability.[4] Essentially, the exploration mode guides the Q-learning algorithm to experiment with suboptimal actions based on the current Q-matrix approximation, $\widehat{Q}_{i,t}$. As t approaches infinity, the pre-specified exploration probability ε_t monotonically decreases to zero.

Given that agent i lacks prior knowledge about its Q-matrix, it is evident that sufficient exploration is crucial to increase the accuracy of approximating the true Q-matrix. At a minimum, all actions must be attempted multiple times in all states, and even more so in complex environments. However, in addition to the computational costs associated with exploration, there exists a tradeoff. An overly comprehensive exploration scheme may have adverse effects when multiple agents interact with one another, because the random selected actions by one agent introduce noises to other agents, impeding their learning processes.

3 Model

This model extends the influential framework of Kyle (1985) along three novel dimensions. First, it considers multiple informed speculators in a repeated-game context. Second, it introduces a representative preferred-habitat investor, whose net demand flows need to be absorbed by other agents in the market (e.g., Vayanos and Vila, 2021). Third, it introduces a market maker who takes into account both inventory and pricing error, going beyond the limited focus on price error alone as in the model of Kyle (1985).

By blending theoretical rigor with practical relevance, this model offers a laboratory for exploring the implications of AI-powered trading on both algorithmic collusion and price efficiency. Importantly, the theoretical results produced by the model act as a foundational benchmark for the characterization and categorization of AI-powered trading in simulation experiments in Sections 4 to 6.

3.1 Economic Environment

Time is discrete, indexed by $t = 1, 2, ...$, and runs forever. There are $I \geq 2$ risk-neutral informed speculators, a representative noise trader, a representative preferred-habitat investor, and a market maker. The economic environment is stationary, and all exogenous shocks are independent and identically distributed across periods.

In each period t, an asset is available for trading, with its fundamental value, denoted as v_t, being realized at the end of the period. Each period consists of two distinct steps: the beginning and the end. We examine the problem in period t in reverse order. At the end of the period, v_t is observed by all agents. It is drawn from a normal distribution $N(\overline{v}, \sigma_v^2)$, where σ_v^2 represents the variance and $\overline{v}$ the mean, with $\overline{v} \equiv 1$ for convenience. After the realization of v_t, trading profits for all agents in period t are determined.

[4]For simplicity, we adopt a uniform distribution. However, a more intelligent distribution choice could make exploration more efficient and less costly.

At the beginning of the period, the informed speculators, noise trader, and preferred-habitat investor submit their order flows. Simultaneously, the market maker sets the asset's price, denoted as p_t. Specifically, the noise trader submits its order flow u_t to either buy u_t units of the asset if $u_t > 0$ or take a short position of u_t if $u_t < 0$, with u_t following a normal distribution $N(0, \sigma_u^2)$, where zero is the mean and σ_u^2 is the variance. The informed speculators are indexed by $i \in \{1, \cdots, I\}$. Each informed speculator i perfectly knows the value v_t, but is unaware of u_t when submitting his order flows; he understands that the choice of order flow $x_{i,t}$ will influence p_t by shifting the market-clearing condition and revealing information. The informed speculator i chooses its order flows $\{x_{i,t}\}_{t\geq 0}$ to maximize the expected present value of the profit stream:

$$\mathbb{E}\left[\sum_{t=0}^{\infty} \rho^t (v_t - p_t) x_{i,t}\right], \tag{3.1}$$

where $\rho \in (0,1)$ is the subjective discount rate.

Preferred-Habitat Investor's Demand Curve. Contrary to the uninformed speculator in Kyle (1989), the preferred-habitat investor does not derive information about v_t from p_t. Instead, this investor has a linear downward-sloping demand curve for the net trading flow z_t:

$$z_t = -\xi(p_t - \overline{v}), \quad \text{with } \xi > 0. \tag{3.2}$$

The rationale behind this specification is straightforward: the preferred-habitat investor focuses solely on the ex-ante expected fundamental value, $\overline{v}$, and tends to buy more of the asset when $p_t - \overline{v}$ is more negative, interpreting this as a stronger indication that the asset is currently undervalued. The demand curve is proportional to the spread between the ex-ante expected fundamental value and the market price. Graham (1973) names this spread a safety margin.

The average asset holding of the preferred-habitat investor, denoted as $\overline{z}$, is often substantial. This implies a small price elasticity of demand, given by $\varepsilon = \mathbb{E}[(\mathrm{d}z_t/\mathrm{d}p_t)(p_t/z_t)] = -\xi\mathbb{E}[p_t/z_t] \approx -\xi/\overline{z}$. Studies indicate that preferred-habitat investors with low price elasticity of demand play an important role in shaping asset prices (e.g., Greenwood and Vayanos, 2014; Vayanos and Vila, 2021; Greenwood et al., 2023).

The preferred-habitat investor's demand curve (3.2) mirrors that of the "long-term investor" in the model by Kyle and Xiong (2001). This becomes clear, especially when we recognize that $\overline{v}$ is the fair value of the asset to risk neutral investors as $\overline{v} = \mathbb{E}[v_t]$. According to this demand curve, the preferred-habitat investor always provides liquidity to the market. When the price falls further below the ex-ante expected fundamental value, $\overline{v}$, in the market, the preferred-habitat investor will buy more of the asset. Analogous to Kyle and Xiong (2001), the demand curve (3.2) can be justified by a rational choice made by the preferred-habitat investor under certain assumptions. These assumptions are summarized in Lemma 1. The proof is in Appendix A.

Lemma 1 (Demand Curve)**.** *If the preferred-habitat investor possesses exponential utility with an absolute*

risk aversion coefficient of η, then the demand curve has the functional form of (3.2), where the slope ξ is given by $1/(\eta\sigma_v^2)$.

Moreover, the concept of specifying exogenous net demand curves within the framework of a noisy rational expectation equilibrium also shares similarities with studies conducted by Hellwig, Mukherji and Tsyvinski (2006) and Goldstein, Ozdenoren and Yuan (2013), among others. The fundamental idea is to capture relevant institutional frictions and preferences in a parsimonious and tractable manner. Notably, our net demand curves can be reinterpreted as "noisy supply curves" in these prior works by introducing a new variable $\widetilde{z}_t \equiv -(u_t + z_t)$. Specifically, $\widetilde{z}_t$ represents the total trading supply provided by the noisy trader and the preferred-habitat investor to absorb the trading demand of informed speculators. The total supply $\widetilde{z}_t$ follows an exogenous noisy supply curve defined as:

$$\widetilde{z}_t = -u_t + \xi(p_t - \overline{v}), \tag{3.3}$$

where $-u_t$ can be reinterpreted as the unobservable demand or supply shock in the context of the above prior works.

Market Maker's Pricing Rules. Trading occurs through the market maker, whose role is to absorb the order flow while minimizing pricing errors. The market maker observes the combined order flow of informed speculators and the noise trader, represented by $y_t = \sum_{i=1}^{I} x_{i,t} + u_t$, as well as the order flow z_t of the preferred-habitat investor. However, the market maker cannot distinguish between order flows from informed speculators and the noise trader. Thus, the market maker can only make statistical inferences about the fundamental value v_t based on the combined order flow y_t rather than individual order flows. The market maker sets the price p_t to jointly minimize inventory and pricing errors according to the following objective function:

$$\min_{p_t} \mathbb{E}\left[(y_t + z_t)^2 + \theta(p_t - v_t)^2 \middle| y_t\right], \tag{3.4}$$

where $\theta > 0$ represents the weight that the market maker places on minimizing pricing errors. Here, $\mathbb{E}[\cdot|y_t]$ denotes the market maker's expectation over v_t, conditioned on the observed combined order flow y_t and its belief about how informed speculators would behave in the equilibrium.

The market maker's objective function (3.4) captures both the inventory cost and asymmetric information faced by the market maker. Because the market maker takes the position $-(y_t + z_t)$ to clear the market, the term $(y_t + z_t)^2$ represents its inventory-holding costs. The quadratic form is adopted for tractability, consistent with the literature (e.g., Mildenstein and Schleef, 1983). The term $\theta(p_t - v_t)^2$ captures the market maker's efforts to reduce pricing errors arising from asymmetric information. The weight θ serves as a reduced-form way to capture the various benefits of reducing pricing errors, such as increased trading flows from a growing client base or enhanced competitive advantages over other trading platforms.[5] As θ approaches zero, the price

[5]Similarly, in the context of e-commerce platforms, it is often assumed that the platform aims to maximize a weighted

p_t is primarily determined by the market clearing condition, $y_t + z_t = 0$, as in the model of Kyle and Xiong (2001). Conversely, as θ increases towards infinity, the price p_t is primarily determined by the pricing-error minimization condition, $p_t = \mathbb{E}[v_t|y_t]$, as in the model of Kyle (1985).

Because multiple informed speculators engage in a repeated-game of trading in our model, multiple equilibria may emerge. We identify three types of equilibria: the non-collusive equilibrium, the perfect cartel equilibrium, and the collusive equilibrium sustained by price-trigger strategies. Throughout our analysis, we assume that the market maker is aware of the specific equilibrium in which informed speculators are participating. Specifically, we consider the linear and symmetric equilibrium in which the trading strategy of the informed speculators is characterized by

$$x_{i,t} = \chi(v_t - \overline{v}), \quad \text{for all } i = 1, \cdots, I. \tag{3.5}$$

The first-order condition of the minimization problem (3.4) leads to

$$p_t = \frac{\xi}{\xi^2 + \theta} y_t + \frac{\xi^2}{\xi^2 + \theta} \overline{v} + \frac{\theta}{\xi^2 + \theta} \mathbb{E}[v_t|y_t],$$

where $\mathbb{E}[v_t|y_t]$, according to Bayesian updating, is

$$\mathbb{E}[v_t|y_t] = \overline{v} + \gamma y_t, \quad \text{with } \gamma = \frac{I\chi}{(I\chi)^2 + \sigma_u^2/\sigma_v^2}.$$

Therefore, the market maker's pricing rule is

$$p_t = \overline{v} + \lambda y_t, \quad \text{with } \lambda = \frac{\theta\gamma + \xi}{\theta + \xi^2}.$$

3.2 Noncollusive Nash Equilibrium

We use the superscript N to denote the variables in the noncollusive Nash equilibrium. At the beginning of each period t, each informed speculator i solves the following problem:

$$x^N(v_t) = \underset{x_i}{\operatorname{argmax}} \mathbb{E}\left[(v_t - p_t)\, x_i \middle| v_t\right], \tag{3.6}$$

where $\mathbb{E}[\cdot|v_t]$ is informed investor i's expectation conditional on the privately observed v_t and its belief about how the market maker would set the price in the equilibrium $p_t = p^N(y_t)$. The pricing function $p^N(\cdot)$ is determined in equilibrium, as follows:

$$p^N(y_t) = \overline{v} + \lambda^N y_t, \quad \text{with } \lambda^N = \frac{\theta\gamma^N + \xi}{\theta + \xi^2} \text{ and } \gamma^N = \frac{I\chi^N}{(I\chi^N)^2 + (\sigma_u/\sigma_v)^2}, \tag{3.7}$$

average of per-unit fee revenues and consumer surplus (see, e.g., Johnson, Rhodes and Wildenbeest, 2023). The weight on consumer surplus in this context is a reduced-form way to capture various aspects of increasing consumer surplus. For example, increasing consumer surplus allows the platform to dynamically expand its consumer base over time and better compete with rival platforms.

where y_t is the combined order flow of informed speculators and the noise trader, given by

$$y_t = x_i + (I-1)x^N(v_t) + u_t. \tag{3.8}$$

The non-collusive Nash equilibrium can be summarized in the following proposition.

Proposition 3.1. *The order flow of informed speculators and price in the non-collusive Nash equilibrium are*

$$x^N(v_t) = \chi^N(v_t - \overline{v}) \;\; \textit{and} \;\; p^N(v_t) = \overline{v} + \lambda^N y_t, \;\; \textit{respectively,}$$

where χ^N and λ^N satisfy

$$\chi^N = \frac{1}{(I+1)\lambda^N} \quad \textit{and} \quad \lambda^N = \frac{\theta\gamma^N + \xi}{\theta + \xi^2} \;\; \textit{with} \;\; \gamma^N = \frac{I\chi^N}{(I\chi^N)^2 + (\sigma_u/\sigma_v)^2}.$$

The expected profit of informed speculators is

$$\pi^N = \left(1 - \lambda^N I\chi^N\right)\chi^N\sigma_v^2.$$

The price informativeness, denoted by $\mathcal{I}^N$, is defined as the logged signal-noise ratio of prices,

$$\mathcal{I}^N = \log\left[\frac{\mathrm{var}(x_{i,t}^N)}{\mathrm{var}(u_t)}\right] = \log\left[\left(I\chi^N\right)^2(\sigma_v/\sigma_u)^2\right].$$

The market liquidity, denoted by $\mathcal{L}^N$, is defined as the inverse sensitivity of the market maker's inventory $|z_t + y_t|$ to the noise order flow u_t

$$\mathcal{L}^N = \frac{1}{\partial|z_t + y_t|/\partial u_t} = \frac{1}{|1 - \xi\lambda^N|}.$$

The mispricing, denoted by $\mathcal{E}^N$, is defined by the percentage deviation of the asset's price p_t from its conditional expected value

$$\mathcal{E}^N = \left|\frac{p^N(v_t) - \mathbb{E}^N[v_t|y_t]}{\mathbb{E}^N[v_t|y_t] - \overline{v}}\right| = \left|\frac{\lambda^N - \gamma^N}{\gamma^N}\right|.$$

Intuitively, the price informativeness measure captures the fact that relative to the noise trader, informed speculators' order flows contain information about the asset's value v_t. Thus, the order flow of informed speculators can be considered as informative signals about the value of v_t whereas noise order flows contain no information. The market liquidity measure captures the fact that when the market is less liquid, trade flows can have a larger impact on the market maker's inventory, leading to greater adjustments in the asset's price.

3.3 Perfect Cartel Equilibrium

Consider a cartel that consists all I informed speculators under perfect collusion. The cartel is a monopolist who chooses each informed speculator's order flow to maximize total profits. Because informed speculators are symmetric, the cartel solves the following problem

$$x^M(v_t) = \underset{x_i}{\operatorname{argmax}}\, \mathbb{E}\left[(v_t - p_t)\, x_i \middle| v_t\right], \tag{3.9}$$

where $\mathbb{E}[\cdot|v_t]$ is informed investor i's expectation conditional on the privately observed v_t and its belief about how the market maker would set the price in the equilibrium $p_t = p^M(y_t)$. The pricing function $p^M(\cdot)$ s determined in equilibrium, as follows:

$$p^M(y_t) = \overline{v} + \lambda^M y_t, \;\; \text{with} \;\; \lambda^M = \frac{\theta\gamma^M + \xi}{\theta + \xi^2} \;\; \text{and} \;\; \gamma^M = \frac{I\chi^M}{(I\chi^M)^2 + (\sigma_u/\sigma_v)^2}, \tag{3.10}$$

where y_t is the combined order flow of informed speculators and the noise trader, given by

$$y_t = Ix_i + u_t. \tag{3.11}$$

The perfect cartel equilibrium can be summarized in the following proposition.

Proposition 3.2. *The order flow of informed speculators and price in the perfect cartel equilibrium are*

$$x^M(v_t) = \chi^M(v_t - \overline{v}) \;\; \textit{and} \;\; p^M(v_t) = \overline{v} + \lambda^M y_t, \;\; \textit{respectively,}$$

where χ^M and λ^M satisfy

$$\chi^M = \frac{1}{2I\lambda^M} \quad \textit{and} \quad \lambda^M = \frac{\theta\gamma^M + \xi}{\theta + \xi^2} \;\; \textit{with} \;\; \gamma^M = \frac{I\chi^M}{(I\chi^M)^2 + (\sigma_u/\sigma_v)^2}.$$

The expected profit of informed speculators is

$$\pi^M = \left(1 - \lambda^M I\chi^M\right)\chi^M\sigma_v^2.$$

The price informativeness, denoted by $\mathcal{I}^M$, is defined as the logged signal-noise ratio of prices,

$$\mathcal{I}^M = \log\left[\frac{\operatorname{var}(x_{i,t}^M)}{\operatorname{var}(u_t)}\right] = \log\left[\left(I\chi^M\right)^2(\sigma_v/\sigma_u)^2\right].$$

The market liquidity, denoted by $\mathcal{L}^M$, is defined as the inverse sensitivity of the market maker's inventory $|z_t + y_t|$ to the noise order flow u_t

$$\mathcal{L}^M = \frac{1}{\partial|z_t + y_t|/\partial u_t} = \frac{1}{|1 - \xi\lambda^M|}.$$

The mispricing, denoted by $\mathcal{E}^M$, is defined by the percentage deviation of the asset's price p_t from its conditional expected value

$$\mathcal{E}^M = \left| \frac{p^M(v_t) - \mathbb{E}^M[v_t|y_t]}{\mathbb{E}^M[v_t|y_t] - \overline{v}} \right| = \left| \frac{\lambda^M - \gamma^M}{\gamma^M} \right|.$$

3.4 Collusive Nash Equilibrium

Information asymmetry is a significant characteristic of capital markets, rendering standard grim-trigger strategies less viable to sustain tacit collusion, due to the challenges in accurately observing and monitoring each other's actions.[6] However, tacit collusion can still be sustained under information asymmetry through price-trigger strategies with imperfect monitoring. If an informed speculator can reliably infer other informed speculators' total order flows from the market price, collusive incentives can be created.

The concept of tacit collusion sustained by price-trigger strategies was first introduced by Green and Porter (1984). Even with imperfect monitoring, agents can establish collusive incentives by allowing noncollusive competition to occur with positive probabilities. Abreu, Pearce and Stacchetti (1986) further characterize optimal symmetric equilibria in this context, revealing two extreme regimes: a collusive regime and a punishment regime featuring a noncollusive reversion. In the collusive regime, informed speculators implicitly coordinate on submitting order flows in a less aggressive manner than what they would do in the noncollusive Nash equilibrium. If the price breaches a critical level, suspicion of cheating arises, leading to a noncollusion reversion. In the punishment regime, informed speculators trade noncollusively and obtain low profits.

Price-Trigger Strategies. We now describe the collusive Nash equilibrium sustained by price-trigger strategies under information asymmetry, as studied by Green and Porter (1984). Specifically, we focus on the symmetric collusive Nash equilibrium in which all I informed speculators choose the same collusive order flow, denoted by $x^C(v_t)$. Such trading strategies are sustained by a price-trigger strategy: Firms will initially submit their respective order flows $x^C(v_t)$, and will continue to do so until the market price falls below a trigger price $q(v_t)$ if $v_t < \overline{v}$ or goes above a trigger price $q(v_t)$ if $v_t > \overline{v}$, and then they will trade noncollusively for a reversionary episode that lasts for $T-1$ periods. In period t, the state of world is "normal," denoted by $s_t = 0$, if (a) $v_{t-1} = \overline{v}$ and $s_{t-1} = 0$, or (b) $p_{t-1} \leq q(v_{t-1})$ and $v_{t-1} > \overline{v}$ and $s_{t-1} = 0$, or (c) $p_{t-1} \geq q(v_{t-1})$ and $v_{t-1} < \overline{v}$ and $s_{t-1} = 0$, or (d) $p_{t-T} > q(v_{t-T})$ and $v_{t-T} > \overline{v}$ and $s_{t-T} = 0$, or (e) $p_{t-T} \leq q(v_{t-T})$ and $v_{t-T} < \overline{v}$ and $s_{t-T} = 0$. Otherwise, in period t, the state of world is "reversionary," denoted by $s_t = 1$. In other words, $s_t = 0$ if price trigger is not violated at $t-1$ and $s_{t-1} = 0$, or if price trigger is violated at $t-T$ and $s_{t-T} = 0$; otherwise, $s_t = 1$.

[6] Tacit collusion sustained by grim-trigger strategies has been extensively studied since the pioneering work of Fudenberg and Maskin (1986) and Rotemberg and Saloner (1986), among others. Recent studies delve into the impact of such tacit collusion sustained by grim-trigger strategies on pricing in capital markets (e.g., Opp, Parlour and Walden, 2014; Dou, Ji and Wu, 2021*a*,*b*; Dou, Wang and Wang, 2023).

Similar to Green and Porter (1984), we assume that the state variable s_t is a common knowledge to all agents. We characterize the equilibrium order flows and prices in each period t. There are two cases: when $s_t = 1$, the state of world is reversionary, and thus the equilibrium order flows and prices follow the noncollusive equilibrium in Section 3.2; and when $s_t = 0$, the state of world is normal. In this case, we focus on linear policy functions and characterize the equilibrium order flow $x^C(v_t)$ and price p_t^C as follows:

$$x^C(v) \equiv \chi^C(v - \overline{v}), \tag{3.12}$$

$$p^C(y) = \overline{v} + \lambda^C y, \text{ with } \lambda^C = \frac{\theta\gamma^C + \xi}{\theta + \xi^2} \text{ and } \gamma^C = \frac{I\chi^C}{(I\chi^C)^2 + \sigma_u^2/\sigma_v^2}. \tag{3.13}$$

The price-trigger function $q(v)$ is specified based on the expected price when all informed speculators trade coordinately according to $x^C(v)$ conditional on v, namely, $\overline{p}^C(v) \equiv \mathbb{E}\left[p^C(y)|v\right]$. Specifically, plugging (3.12) into (3.13) and taking expectation over u, we obtain that $\overline{p}^C(v) \equiv \overline{v} + \lambda^C I\chi^C(v - \overline{v})$. The price-trigger function $q(v)$ is specified as follows:

$$q(v) \equiv \begin{cases} \overline{p}^C(v) + \lambda^C\sigma_u\omega, & \text{if } v > \overline{v} \\ \overline{p}^C(v) - \lambda^C\sigma_u\omega, & \text{if } v < \overline{v}, \end{cases} \tag{3.14}$$

where $\omega > 0$ is a parameter that characterizes the tightness of the price trigger.

Equation (3.14) warrants further discussions. First, when $v > \overline{v}$, informed investors have incentives to buy a large amount of the asset, which boosts up its price. As a result, when $v > \overline{v}$, a meaningful price-trigger strategy would punish the potential deviating counterparty by reverting to the noncollusive Nash equilibrium once the market price goes above a certain high-level threshold $q(v)$. In contrast, when $v < \overline{v}$, informed investors have incentives to sell a large amount of the asset, which suppresses down its price. As a result, when $v < \overline{v}$, a meaningful price-trigger strategy would punish the potential deviating counterparty by reverting to the noncollusive Nash equilibrium once the market price falls below a certain low-level threshold $q(v)$. Second, there is no price threshold when $v = \overline{v}$ because no informed investor would have incentives to trade in this case. Third, although there are infinitely many alternative ways to specify the functional form of the threshold $q(v)$, we focus on a specification that is not only statistically meaningful but also ensures a linear model solution as in Kyle (1985). If no one deviates from the coordinated trading, each informed speculator can infer that the noise order is $\widehat{u}_t = [p_t - q(v_t)]/\lambda^C$ based on the observed price $p_t = p^C(y_t)$. If $\widehat{u}_t$ is excessively positive when $v_t > \overline{v}$, say $\widehat{u}_t > \omega\sigma_u$ for some constant $\omega > 0$, the informed speculator would suspect that some other informed speculators might have deviated from the implicit agreement. Analogously, if $\widehat{u}_t$ is excessively negative when $v_t < \overline{v}$, say $\widehat{u}_t < -\omega\sigma_u$ for some constant $\omega > 0$, the informed speculator would suspect that some other informed speculators might have deviated from the implicit agreement. Fourth, the multiplier σ_u ensures that the probability of price-trigger violation is independent of the magnitude of noisy trading, σ_u, in the collusive Nash equilibrium.

Given that $s_t = 0$, let $J^C(\chi_i)$ denote each informed speculator i's expected present value of

future profits, when investor i chooses $x_{i,t} = \chi_i(v_t - \overline{v})$ and all other $I-1$ informed investors choose $x^C(v_t)$. The value of $J^C(\chi_i)$ is determined recursively as follows:

$$\begin{aligned} J^C(\chi_i) =& \mathbb{E}\left[\left(v_t - p^C(y_t)\right)\chi_i(v_t - \overline{v})\right] \\ &+ \rho J^C(\chi_i)\mathbb{P}\left\{\text{Price trigger is not violated in period } t \middle| \chi_i, \chi^C\right\} \\ &+ \mathbb{E}\left[\sum_{\tau=1}^{T-1}\rho^\tau \pi^N(v_{t+\tau}) + \rho^T J^C(\chi_i)\right]\mathbb{P}\left\{\text{Price trigger is violated in period } t \middle| \chi_i, \chi^C\right\}, \end{aligned} \tag{3.15}$$

where the combined order flow of informed investors and the noise trader is

$$y_t = \chi_i(v_t - \overline{v}) + (I-1)x^C(v_t) + u_t, \tag{3.16}$$

and the probability of price-trigger violation is

$$\begin{aligned} &\mathbb{P}\left\{\text{Price trigger is not violated in period } t \middle| \chi_i, \chi^C\right\} \\ =&\mathbb{E}\left[\mathbb{P}\left(p_t \le q(v_t)|v_t\right)\mathbf{1}\{v_t > \overline{v}\}\right] + \mathbb{E}\left[\mathbb{P}\left(p_t \ge q(v_t)|v_t\right)\mathbf{1}\{v_t < \overline{v}\}\right] \\ =&\mathbb{E}\left[\Phi(\sigma_u^{-1}(\chi^C - \chi_i)(v_t - \overline{v}) + \omega)\mathbf{1}\{v_t > \overline{v}\}\right] + \mathbb{E}\left[\Phi(\sigma_u^{-1}(\chi_i - \chi^C)(v_t - \overline{v}) + \omega)\mathbf{1}\{v_t < \overline{v}\}\right], \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

Impossibility of Collusion When Efficient Prices Prevail. The following proposition highlights the impossibility of achieving collusion in an environment closely resembling the standard Kyle benchmark (Kyle, 1985), where efficient prices prevail. In this setting, the market maker focuses on minimizing pricing errors and sets the price approximately at $\mathbb{E}[v_t|y_t]$, which is the expected fundamental value conditional on the observed combined order flow of informed speculators and the noise trader. In other words, the efficient price in this context is an unbiased estimate of the asset's fundamental value. The proof is in Appendix B.

Proposition 3.3 (Impossibility of Collusion When Efficient Prices Prevail)**.** *If θ is large or ξ is small, there is no collusive Nash equilibrium that can be sustained by price-trigger strategies for any $\sigma_u/\sigma_v > 0$.*

Sustaining coordination through price-trigger strategies requires two conditions: (i) price informativeness needs to be sufficiently high to ensure that there is sufficient capacity for monitoring, which has been emphasized by Abreu, Milgrom and Pearce (1991) and Sannikov and Skrzypacz (2007), and (ii) the price impact of informed speculators' order flows needs to be sufficiently low to ensure that there is sufficient room for achieving significant informational rents.

However, the environments with large θ or small ξ closely resemble the standard Kyle benchmark (Kyle, 1985), where efficient prices prevail. In this environment, because λ^C is

approximately equal to γ^C, price informativeness is always low and unresponsive to σ_u/σ_v.[7] As a result, the two necessary conditions (i) and (ii) cannot hold simultaneously. In particular, in order to achieve high price informativeness, the environment needs to have low noise trading risks, as captured by a low σ_u/σ_v. However, knowing that noise orders are not significant, the market maker will choose a high γ^C, resulting in a high price impact of informed trading because $\lambda^C \approx \gamma^C$. The high price impact of informed trading would further induce informed speculators to trade conservatively by placing orders of small amounts. In the end, the positive effect on price informativeness from low noise trading risks would be largely cancelled out by the negative effect from the conservative orders of informed speculators, making the price informativeness low and unresponsive to σ_u/σ_v.

Proposition 3.3 carries intrinsic value in terms of theoretical insights and novelty, setting it apart from existing theories on the impossibility of collusion under information asymmetry, as posited by Abreu, Milgrom and Pearce (1991) and Sannikov and Skrzypacz (2007). These prior theories emphasize that, when prices are not informative, "false positive" errors, made by triggering punishments, occur on the equilibrium path disproportionately often, erasing all benefits from collusion. In contrast, Proposition 3.3 offers a distinctive intuitive perspective, highlighting that informed speculators cannot exploit pricing errors to achieve collusive outcomes because prices are already efficient, accurately reflecting the asset's fundamental value. The absence of substantial pricing errors essentially renders collusion infeasible, as there exists limited scope for market manipulation based on price discrepancies. In summary, Proposition 3.3 sheds light on the interplay between efficient pricing, information asymmetry, and collusive behavior in financial markets. By demonstrating the impracticality of collusion in environments characterized by efficient prices, our results provide a deeper understanding of market dynamics and the implications of information asymmetry on collusion strategies.

Existence of Collusion with a Significant Preferred-Habitat Investor. The following proposition shows that collusion sustained by price-trigger strategies exists when the preferred-habitat investor plays an important role in price formation, making prices not very efficient. However, when information asymmetry, captured by σ_u/σ_v, is too large, or when the number of informed speculators I, no collusion can be sustained by price-trigger strategies even with inefficient prices. The proof is in Appendix C.

Proposition 3.4 (Existence of Collusion with a Significant Preferred-Habitat Investor)**.** *If θ is sufficiently small or ξ is sufficiently large, there exists a collusive Nash equilibrium that can be sustained by price-trigger strategies provided that σ_u/σ_v and I are not too large.*

If θ is small or ξ is large, the market maker determines prices determined primarily to minimize inventory costs rather than pricing errors. Thus, a low price impact of informed trading can arise even in environments with low noise trading risks. The low price impact of informed trading would further induce informed speculators to trade aggressively by placing orders of large

[7] In the extreme case with $\theta = \infty$ or $\xi = 0$, price informativeness is independent from σ_u/σ_v as in Kyle (1985).

amounts, thereby leading to high price informativeness. Consequently, the necessary conditions (i) and (ii) can hold simultaneously when the preferred-habitat investor plays an important role in price formation.

However, when σ_u/σ_v is too large, price informativeness is low, and thus price-trigger strategies are difficult to sustain. This is because when prices are not informative, agents make "false positive" errors by triggering punishments on the equilibrium path disproportionately often, erasing all benefits from collusion. This key idea exactly follows the insight of Abreu, Milgrom and Pearce (1991) and Sannikov and Skrzypacz (2007).

Properties of Collusion Sustained by Price-Trigger Strategies. To discern whether informed speculators trade in a tacitly collusive manner based on observable outcomes, we derive testable properties of collusion.

Proposition 3.5 (Supra-competitive nature of collusion)**.** *In the price-trigger collusive equilibrium, it holds that*

$$\pi^M \geq \pi^C > \pi^N, \tag{3.17}$$

where $\pi^C = \left(1 - \lambda^C I \chi^C\right) \chi^C \sigma_v^2$ *is the expected profit of informed speculators in the collusive equilibrium. If we define* $\Delta^C \equiv \frac{\pi^C - \pi^N}{\pi^M - \pi^N}$*, inequalities in (3.17) can be summarized by* $\Delta^C \in (0, 1]$*.*

Clearly, a greater Δ^C signifies a higher collusion capacity. We use Δ^C as a measure for collusion capacity, as in Calvano et al. (2020). Similar measures are also adopted in empirical studies to identify collusion capacity (e.g., Dou, Wang and Wang, 2023). Consistent with the definitions in the noncollusive equilibrium and the perfect cartel equilibrium, the price informativeness, denoted by $\mathcal{I}^C$, is defined as the logged signal-noise ratio of prices,

$$\mathcal{I}^C = \log\left[\frac{\mathrm{var}(x_{i,t}^C)}{\mathrm{var}(u_t)}\right] = \log\left[\left(I\chi^C\right)^2 (\sigma_v/\sigma_u)^2\right].$$

The market liquidity, denoted by $\mathcal{L}^C$, is defined as the inverse sensitivity of the market maker's inventory $|z_t + y_t|$ to the noise order flow u_t

$$\mathcal{L}^C = \frac{1}{\partial |z_t + y_t| / \partial u_t} = \frac{1}{|1 - \xi\lambda^C|}.$$

The mispricing, denoted by $\mathcal{E}^C$, is defined by the percentage deviation of the asset's price p_t from its conditional expected value

$$\mathcal{E}^C = \left|\frac{p^C(v_t) - \mathbb{E}^C[v_t|y_t]}{\mathbb{E}^C[v_t|y_t] - \overline{v}}\right| = \left|\frac{\lambda^C - \gamma^C}{\gamma^C}\right|.$$

In the next proposition, we derive how Δ^C, $\mathcal{I}^C$, $\mathcal{L}^C$, and $\mathcal{E}^C$ vary across various market structures and information environments. The proof is in Appendix D.

Proposition 3.6 (Effects of Market Structures and Information Environments)**.** *If θ is sufficiently small or ξ is sufficiently large, the price-trigger collusive Nash equilibrium satisfies the following properties:*

$$\textit{(i)}\ I\uparrow \implies \Delta^C\downarrow \ \ \& \ \ \mathcal{I}^C/\mathcal{I}^M\uparrow \ \ \& \ \ \mathcal{L}^C/\mathcal{L}^M\uparrow \ \ \& \ \ \mathcal{E}^C\downarrow$$

$$\textit{(ii)}\ \sigma_u/\sigma_v\uparrow \implies \Delta^C\downarrow \ \ \& \ \ \mathcal{I}^C/\mathcal{I}^M\uparrow \ \ \& \ \ \mathcal{L}^C/\mathcal{L}^M\uparrow \ \ \& \ \ \mathcal{E}^C\downarrow$$

$$\textit{(iii)}\ \rho\uparrow \implies \Delta^C\uparrow \ \ \& \ \ \mathcal{I}^C/\mathcal{I}^M\downarrow \ \ \& \ \ \mathcal{L}^C/\mathcal{L}^M\downarrow \ \ \& \ \ \mathcal{E}^C\uparrow$$

$$\textit{(iv)}\ \xi\uparrow \implies \Delta^C\uparrow \ \ \& \ \ \mathcal{I}^C/\mathcal{I}^M\downarrow \ \ \& \ \ \mathcal{L}^C/\mathcal{L}^M\downarrow \ \ \& \ \ \mathcal{E}^C\uparrow$$

4 Simulation Experiments with AI-Powered Trading

The theoretical results presented in Section 3 are predicated on the assumption that both the informed speculators and the market maker possess rational expectations. Specifically, they are capable of discerning (i) the order flows of other informed speculators, albeit with noise; (ii) the distribution of noise trading flows; and (iii) the distribution of the fundamental value of the asset. Furthermore, both the informed speculators and the market maker are sufficiently astute, with the speculators being able to communicate amongst themselves. This allows the informed speculators to collectively reach and sustain a price-trigger strategy characterized by $\chi^C(v)$ and $q(v)$, as detailed in (3.12) to (3.14). Meanwhile, this also allows the market maker to perfectly understand the collusion scheme of these speculators.

It remains uncertain whether autonomous, model-free AI algorithms can learn to sustain tacit collusion during trading – and thereby generate supercompetitive profits – in line with the theoretical predictions, which are derived based on stringent, and at times, unrealistic assumptions. As a proof-of-concept illustration, in this section, we design simulation experiments to investigate the capability of Q-learning algorithms to attain tacit collusion under asymmetric information, without the overt acts of communication or agreements typically seen in competition law infringements (Harrington, 2018).

4.1 Informed AI Speculators with Q-Learning

We consider informed speculators operating Q-learning algorithms (i.e, informed AI speculators) to learn how to trade. Importantly, informed AI speculators have no direct knowledge of order flows from their counterparts and are oblivious to the distribution of noisy trading flows and the fundamental value of the asset. Our experimental design and methodology are similar to the studies of Calvano et al. (2020) and Asker, Fershtman and Pakes (2022), who explore product market competition under which asymmetric information and endogenous pricing rules are absent.

Specifically, each informed AI speculator $i \in \{1, \cdots, I\}$ adopts the Q-learning algorithm described in Section 2. Observing s_t, informed AI speculator i chooses its order flow $x_{i,t}$, following one of the two experimentation modes described in Section 2.3. After receiving the total quantity

of market orders, the market maker determines the price p_t according to its own pricing rules (see Subsection 4.2 below). The profit of informed AI speculator i in period t is given by $\pi_{i,t} = (v_t - p_t)x_{i,t}$.

State Variables. State variables, s_t, are essential for characterizing the recursive relation presented in equation (2.4). The choice of state variables is not unique. In principle, s_t can encompass any information that informed AI speculator i has observed up to the beginning of period t. This includes both public information and speculator i's own private information. We utilize the smallest possible set of state variables in s_t that can theoretically generate tacit collusion sustained by price-trigger strategies. First, drawing from the insights in Section 3.4, we include the asset's price p_{t-1} in the preceding period $t-1$ as part of s_t. Second, we incorporate v_t, instead of v_{t-1}, as part of s_t because informed AI speculators engage in trading activities in period t after observing v_t at the beginning of period t. Thus, the state variable s_t is defined as $s_t \equiv \{p_{t-1}, v_t\}$. Put simply, we equip the informed AI speculator with a one-period memory to trace the history for decision making, similar to the approach adopted by Calvano et al. (2020).

One could also expand informed AI speculator i's state variables in s_t with its own lagged order flow $x_{i,t-1}$, a piece of private information only known by informed AI speculator i, and a longer memory for lagged asset prices and order flows. In our simulation experiments, we observe that enlarging the state variable s_t augments the degree of tacit collusion among informed AI speculators, leading to higher trading profits. Thus, our deliberate choice to solely incorporate p_{t-1} and v_t as state variables sets a stringent bar for the Q-learning algorithms to reach tacit collusion within our economic environment. Furthermore, the Q-learning algorithm with state variables $s_t \equiv \{p_{t-1}, v_t\}$ has a convergence speed significantly faster than those incorporating a more extensive list of state variables.[8]

The evolution of state variable s_t is given by $s_{t+1} \equiv \{p_t, v_{t+1}\}$, where v_{t+1} is randomly drawn from the distribution $N(\overline{v}, \sigma_v^2)$. The price p_t is determined by the market maker, and it depends on the noise trading flow and the order from the preferred-habitat investor in period t, which remain unknown to informed AI speculators when they make decisions in period t.

Role of Exploration and Exploitation in Generating Collusive Outcomes. Exploration is not only critical for approximating the true Q-matrix but also for informed AI speculators to learn and sustain the collusion through price-trigger strategies discussed in Section 5.1. In each iteration, the randomly selected order flow typically differs significantly from the exploited order flow that generates collusive profits. Thus, such deviation, triggered by exploration, provides the only opportunity for the algorithms to learn the price-trigger strategies to sustain the collusion through punishment threat.

Exploitation, as a defining characteristic of RL algorithms, plays a vital role in generating collusion through homogenized learning biases discussed in Section 5.2. Specifically, exploitation biases the estimation of the Q-matrix away from its true values. This bias leads to excessive

[8]When dealing with an extensive list of state variables, deep Q-learning algorithms become indispensable.

overestimation of Q-values for certain choices that can sustain collusive profits, while simultaneously underestimating Q-values for other choices in $\mathcal{X}$. The collusion through homogenized learning biases shares a foundation with the fundamental concept of the "bias-variance tradeoff" in supervised machine learning algorithms – sacrificing unbiasedness to gain stronger identification. Although Q-learning algorithms are inherently self-oriented, they can achieve and maintain collusive profits through interactions by overestimating the Q-values of choices that facilitate high collusive profits. Consequently, under the influence of the biased estimated Q-matrix, informed AI speculators lack incentives to deviate from collusive behavior. Such behaviors constitute a unique character of AI algorithms, which is intrinsically different from how human traders would behave.

4.2 Pricing Rule of the Adaptive Market Maker

The market maker does not know the distributions of randomness. It stores and analyzes historical data on asset values, asset prices, the order flows from the preferred-habitat investor, and the combined order flows from informed AI speculators and the noise trader, i.e., $\mathcal{D}_t \equiv \{(v_{t-\tau}, p_{t-\tau}, z_{t-\tau}, y_{t-\tau})\}_{\tau=1}^{T_m}$, where T_m is a large integer. The market maker estimates the demand curve of the preferred-habitat investor and the conditional expectation of the asset value, $\mathbb{E}[v_t|y_t]$, using the following linear regression models:

$$z_{t-\tau} = \xi_0 - \xi_1 p_{t-\tau}, \tag{4.1}$$

$$v_{t-\tau} = \gamma_0 + \gamma_1 y_{t-\tau} + \epsilon_{t-\tau}, \tag{4.2}$$

where $\tau = 1, \cdots, T_m$. The estimated coefficients are $\widehat{\xi}_{0,t}$, $\widehat{\xi}_{1,t}$, $\widehat{\gamma}_{0,t}$, and $\widehat{\gamma}_{1,t}$, respectively, based on the dataset $\mathcal{D}_t$ in period t. The pricing rule adaptively adheres to the theoretical optimal policy using a plug-in procedure:

$$p_t(y) = \widehat{\gamma}_{0,t} + \widehat{\lambda}_t y \ \text{ with } \ \widehat{\lambda}_t = \frac{\theta \widehat{\gamma}_{1,t} + \widehat{\xi}_{1,t}}{\theta + \widehat{\xi}_{1,t}^2}, \tag{4.3}$$

where θ is the market maker's own choice. Therefore, the market maker is adaptive using simple statistical models. To show robustness of our results, we also consider the economic environment where the market maker determines the pricing rule with rational expectations or the market maker adopts Q-learning algorithms to learn the pricing rule (see Appendix F). All the results are similar to those obtained in the baseline economic environment.

4.3 Repeated Games of Machines

At $t = 0$, each informed AI speculator $i \in \{1, \cdots, I\}$ is assigned with an arbitrary initial Q-matrix $\widehat{Q}_{i,0}$ and state s_0. Then, the economy evolves from period t to period $t+1$ as follows:

(1) Informed AI speculator i draws a random value that determines whether it will be in the exploration mode with probability ε_t or the exploitation mode with probability $1 - \varepsilon_t$ in

period t. The random values drawn by different informed AI speculators are independent. Subsequently, each informed AI speculator i submits its own order flow $x_{i,t}$ according to its mode.

(2) The noise trader submits its order flow u_t, which is randomly drawn from a normal distribution $N(0, \sigma_u^2)$.

(3) The preferred-habitat investor submits its order flow z_t according to (3.2).

(4) The market maker observes the historical data $\mathcal{D}_t \equiv \{v_{t-\tau}, p_{t-\tau}, z_{t-\tau}, y_{t-\tau}\}_{\tau=1}^{T_m}$ and estimates the optimal pricing rule according to (4.1) – (4.3).

(5) Each informed AI speculator i realizes its profits $(v_t - p_t)x_{i,t}$ and updates its Q-matrix according to equation (2.4).

(6) At the beginning of period $t+1$, the state variable for each informed AI speculator evolves to $s_{t+1} = \{p_t, v_{t+1}\}$, where v_{t+1} is drawn from $N(\overline{v}, \sigma_v^2)$ and is independent of any other variables.

The interactions of informed AI speculators and an adaptive market maker, together with the randomness caused by the noise trader and stochastic asset values in the background, make the stationary equilibrium difficult to achieve. The economic environment in our study is substantially more complex than that of Calvano et al. (2020) whose setting does not have randomness, information asymmetry, or endogenous pricing rules. As noted by Calvano et al. (2020), the player's optimization problem is inherently nonstationary when its rivals vary their actions over time due to experimentation or learning. Theoretical analysis of the Q-learning algorithms playing repeated games is generally not tractable. Rather than applying stochastic approximation techniques to AI agents, we follow Calvano et al. (2020) by simulating the exact stochastic dynamic system a large number of times to smooth out uncertainty. There is no theoretical guarantee that Q-learning agents will settle on a stable outcome, nor that they will correctly learn an optimal policy. However, we can always verify this in our simulations ex post to ensure that our analyses are conducted based on the stationary equilibrium.

4.4 Discretization of State and Action Space

We choose the following grids for the state variable $s_t \equiv \{p_{t-1}, v_t\}$ and action variable $x_{i,t}$. For computational efficiency, we approximate the normal distribution $N(\overline{v}, \sigma_v)$ using a sufficiently larger number of n_v grid points, $\mathbb{V} = \{v_1, \cdots, v_{n_v}\}$. Our discretization ensures that these n_v grid points have equal probabilities but are unequally spaced. Specifically, the probability of each grid point is $\mathcal{P}_k = 1/n_v$. The locations of grid points are chosen based on $v_k = \overline{v} + \sigma_v \Phi^{-1}((2k-1)/(2n_v))$ for $k = 1, \cdots, n_v$, where Φ^{-1} is the inverse cumulative density function of a standard normal distribution. The mathematical property of Φ^{-1} implies that grid points around the

mean $\overline{v}$ are closer to each other than those far away from the mean. The speed of convergence is significantly increased because all n_v grid points of v_t have equal probabilities.[9]

We construct the discrete grid points for informed AI speculators' order $x_{i,t}$ based on their optimal actions in the noncollusive Nash equilibrium and perfect cartel equilibrium. According to our model in Section 3, the order values in the two equilibria are given by $x^N = (v - \overline{v})/((I+1)\lambda)$ and $x^M = (v - \overline{v})/(2I\lambda)$. We specify informed AI speculators' action space by discretizing the interval $[x^M - \iota(x^N - x^M), x^N + \iota(x^N - x^M)]$ for $v > \overline{v}$ and $[x^N - \iota(x^M - x^N), x^M + \iota(x^M - x^N)]$ for $v < \overline{v}$ into n_x equally spaced grid points, i.e., $\mathbb{X} = \{x_1, \cdots, x_{n_x}\}$. The parameter $\iota > 0$ ensures that informed AI speculators can choose order flows beyond the theoretical levels corresponding to the noncollusive Nash equilibrium and perfect cartel equilibrium. As the action space is discrete, the exact order flows corresponding to the perfect cartel equilibrium may not be feasible. Despite this, our simulations show that informed AI speculators can collude with each other to a large degree.

The grid points of price p_t are similarly chosen as those of $x_{i,t}$, except for considering the noise trader's impact on prices. Specifically, in our numerical experiments, the noise trader's order is drawn randomly from the normal distribution $N(0, \sigma_u)$, without imposing any discretization or truncation. In our theoretical framework in Section 3, the market maker sets the price according to the total order flow y_t, which is the sum of informed AI speculators' order $\sum_{i=1}^{I} x_{i,t}$ and the noise trader's order u_t. Because u_t follows an unbounded normal distribution, the theoretical range of the price p_t is unbounded. To maintain tractability, in our numerical experiments, we set the upper bound at $p_H = \overline{v} + \lambda(I \max(x^M, x^N) + 1.96\sigma_u)$ and the lower bound at $p_L = \overline{v} + \lambda(I \min(x^M, x^N) - 1.96\sigma_u)$, corresponding to the 95% confidence interval of the noise trader's order distribution, $N(0, \sigma_u)$. The grid points of p_t are chosen by discretizing the interval $[p_L - \iota(p_H - p_L), p_H + \iota(p_H - p_L)]$ into n_p grids, i.e., $\mathbb{P} = \{p_1, \cdots, p_{n_p}\}$.

4.5 Initial Q-Matrix and States

We initialize the Q-matrix at $t = 0$ using the discounted payoff that would accrue to informed AI speculator i if the other informed AI speculators randomize their actions uniformly over the grid points defined by $\mathbb{X}$.[10] Moreover, we consider a zero order flow from the noise trader,

[9] All the results are robust to the use of alternative methods to discretize the state variable v_t. For example, one commonly used method is to use n_v equally spaced points over a sufficiently large interval, e.g., $[\overline{v} - 6\sigma_v, \overline{v} + 6\sigma_v]$. The probability of each grid point is computed based on the probability mass function of the normal mass function, i.e., $\mathcal{P}_k = \exp\left(-(k - \overline{v})^2/(2\sigma_v^2)\right)$ for $k = 1, \cdots, n_v$. Compared to the discretization method we use, this alternative method yields similar quantitative results but has a much slower convergence. The reason is that it assigns very small probabilities to the left-most and right-most grid points. As a result, the Q-matrix's cells far away from the mean $\overline{v}$ are updated at much lower frequencies than those closer to the mean. An infrequent update for the cells far away from the mean in turn requires many more updates for other cells of the Q-matrix to stabilize. Thus, the global convergence speed is reduced significantly due to the buckets effect. In fact, as $n_v \to \infty$, the two alternative methods can both perfectly capture the theoretical distribution of v_t but yield vastly different convergence speed for the Q-learning algorithms.

[10] Adopting different initial values for the Q-matrix do not significantly alter the results. In RL algorithms, another common strategy to initialize the Q-matrix is to use optimistic initial values. That is, initializing the Q-matrix with sufficiently high values so that subsequent iterations tend to reduce the values of the Q-matrix. This approach enables Q-learning algorithms to visit all actions multiple times at the beginning, resulting in early improvement in estimated action values. Thus, setting optimistic initial values is in some sense equivalent to adopting a thorough exploration

corresponding to the expected value of the distribution $N(0, \sigma_u^2)$. Specifically, for each informed AI speculator $i = 1, \cdots, I$, we set its initial Q-matrix $\widehat{Q}_{i,0}$ at $t = 0$ as follows:

$$\widehat{Q}_{i,0}(p_m, v_k, x_n) = \frac{\sum_{x_{-i} \in \mathbb{X}} \left[v_k - (\overline{v} + \lambda(x_n + (I-1)x_{-i})) \right] x_n}{(1-\rho) n_x}, \tag{4.4}$$

for $(p_m, v_k, x_n) \in \mathbb{P} \times \mathbb{V} \times \mathbb{X}$. The initial states of our simulation, $s_0 = \{p_{-1}, v_0\}$, are randomized uniformly over $\mathbb{V} \times \mathbb{P}$.

4.6 Specification of Learning Modes

We adopt an exponentially time-declining state-dependent exploration rate for informed AI speculators,

$$\varepsilon_{t(v_k)} = e^{-\beta t(v_k)}, \tag{4.5}$$

where the parameter $\beta > 0$ governs the speed that informed AI speculators' exploration rate diminishes over time and the variable $t(v_k)$ captures the number of times that the exogenous state $v_k \in \mathbb{V}$ has occurred in the past.[11] The specification of $t(v_k)$ implies that the exploration rate is state dependent, which ensures that informed AI speculators can sufficiently explore their actions for all grid points of the exogenous state variable v_t.

The specification (4.5) implies that initially, Q-learning algorithms are almost always in the exploration mode, choosing actions randomly. However, as time passes, Q-learning algorithms gradually switch to the exploitation mode.

4.7 Parameter Choice

The parameters used in our numerical experiments can be categorized into three groups according to their roles. The environment parameters are the parameters that characterize the underlying economic environment in our experiments. Importantly, the values of most of these parameters are neither known to informed AI speculators nor to the market maker.[12] They instead adopt Q-learning algorithms to learn how to make decisions in an unknown environment. The simulation parameters are the parameters that determine our numerical experiments, such as the number of discrete grid points, simulation sessions, etc. The hyperparameters are the parameters that control the machine learning process. Below, we describe the choice of parameters for each category.

Environment Parameters. Across all simulation experiments, we set $\overline{v} = 1$, $\sigma_v = 1$, and $\theta = 0.1$. The parameter $\overline{v}$ determines the expected value of v_t, and thus we normalize its value to unity without loss of generality. The parameter σ_v plays a similar role as σ_u because what matters in our

over the entire action space early in the learning phase and then exploitation later on.

[11] In principle, we can allow informed AI speculators to choose their exploration rate conditional on the realized value of v_t because they perfectly observe v_t, which is one of their state variables $s_t = \{p_{t-1}, v_t\}$.

[12] An exception is ρ and θ. The parameter ρ is known to informed AI speculators as this parameter captures their own discount rates. The parameter θ is known to the market maker as this is its own choice.

model in Section 3 is the ratio σ_u/σ_v. We thus normalize the value of σ_v to unity. The parameter θ determines the extent to which the market maker focuses on price discovery. We find that the implications of different values of θ can be analyzed similarly by varying the value of ξ. Thus, for simplicity, we fix the value of θ at 0.1 throughout our simulation experiments.

In the baseline economic environment, we set $I = 2$, $\sigma_u = 0.1$, $\rho = 0.95$, and $\xi = 500$. We extensively study the implications of different values for these parameters. Specifically, we consider different number of informed AI speculators ranging from $I = 2$ to $I = 6$, different levels of background noise ranging from $\sigma_u = e^{-5}$ to $\sigma_u = e^5$, different discount rates ranging from $\rho = 0.5$ to $\rho = 0.95$, and different values of ξ ranging from $\xi = 0$ to $\xi = 500$.

Simulation Parameters. We set $\iota = 0.1$ so that informed AI speculators can go beyond the theoretical bounds of order flows by 10%. We choose $n_x = 15$ and $n_p = 31$. These grid points are sufficiently dense to capture the economic mechanism we are interested in. Importantly, our choice of $n_p \approx 2n_x$ ensures that, all else equal, a one-grid point change in one informed AI speculator's order will result in a change in price p_t over the grid defined by $\mathbb{P}$. If the grid defined by $\mathbb{P}$ is coarser, informed AI speculators will not be able to detect small deviations of peers even in the absence of noise, which in turn lowers the possibility of algorithmic collusion through price-trigger strategies.

We use $n_v = 10$ grid points to approximate the normal distribution of v_t. Under our discretization, the standard deviation of v_t is $\widehat{\sigma}_v = \sqrt{\sum_{k=1}^{N} \mathcal{P}(v_k)(v_k - \overline{v})^2} = 0.938$, which is close to the theoretical value $\sigma_v = 1$. In the remainder of this paper, the theoretical benchmarks of noncollusive Nash equilibrium and perfect cartel equilibrium are computed using $\widehat{\sigma}_v$, to be consistent with the discretization of v_t adopted in our simulation experiments.

All the results are robust if we choose a larger n_v, n_x, n_p, or ι, as long as the hyperparameters, α and β, are adjusted accordingly to ensure sufficiently good learning outcomes. However, the cost of using denser grids is that significantly longer time is needed for Q-learning algorithms to fully converge to limit strategies.

We set $T_m = 10,000$ so that the market maker stores sufficiently long time-series data to estimate the linear regressions (4.1) and (4.2). In our simulation experiments, we verify that the estimates of $\widehat{\xi}_{0,t}$, $\widehat{\xi}_{1,t}$, $\widehat{\gamma}_{0,t}$, and $\widehat{\gamma}_{1,t}$ can accurately recover the preferred-habitat investor's demand curve and the conditional expectation of the asset value, $\mathbb{E}[v_t|y_t]$. Increasing the value of T_m will not change any quantitative results, but it adds more computation burden.

For each experiment with a particular choice of environment parameters, we simulate the Q-learning algorithms by $N = 1,000$ times. All the random initial states and shocks (i.e., v_t, u_t, and exploration status of each informed AI speculator for all $t \geq 0$) are independently drawn from identical distributions across the N simulation sessions of the experiment. In principle, the results of different experiments can differ both because of the difference in environment parameters and the difference in the realized values of random variables. To ensure that comparisons across different experiments are not contaminated by the latter, we generate a large set of random variables for all N simulation sessions offline and store in the high-powered-computing server.

The same set of random values is used when we compare results across the experiments with different environment parameters in Sections 5 and 6.

Hyperparameters. The hyperparameters that control the learning process of Q-learning algorithms are set at $\alpha = 0.01$ and $\beta = 10^{-5}$. All results are robust to choosing different values of α and β so long as they are in the reasonable range that ensures sufficiently good learning outcomes. Our baseline choice of β implies that any action $x_k \in \mathbb{X}$ is visited purely by random exploration by $n_v/[(1-\exp(-10^{-5}))n_x] = 66,660$ times on average before exploration completes.[13] In Section 6.3, we study the experiments with different values of α and β as well as the experiments that allow informed AI speculators to adopt different values of α. In Section 7.2, we develop a two-tier Q-learning algorithms that allow informed AI speculators to learn the choice of α.

4.8 Convergence

Strategic games played by Q-learning algorithms do not have general convergence results. To verify convergence, a practical criterion is to check whether each player's optimal strategy does not change for a long period of time. Note that convergence is determined by the stationarity of players' optimal strategies rather than the stationarity of players' learned Q-matrices. In fact, in a stochastic environment, the Q-matrix can never remain unchanged because randomly realized shocks will always result in an update for some cells of the Q-matrix. However, the slight update in the Q matrix does not necessarily result in a change in the optimal strategies. This is why convergence in optimal strategies can be achieved in principle, even in a stochastic environment with Q-learning algorithms playing repeated games.

In general, setting a smaller value of α or β requires longer time for the algorithm to reach convergence. For example, with $\beta = 10^{-5}$, informed AI speculators' Q-learning algorithms are still doing exploration with $e^{-\beta T'/n_v} = 36.8\%$ probability after $T' = 1,000,000$ periods. It is almost by definition that the optimal strategies are nonstationary with an exploration rate that is far away from zero. Thus, a necessary condition for all Q-learning algorithms to reach stationary optimal strategies is that exploration rate is virtually zero, say, after 10,000,000 periods. Moreover, with a small α, the Q-matrix is updated slowly when new information arrives. As a result, informed AI speculators can only slowly learn their optimal actions, which are based on their learned Q-matrices. A sufficiently long time is needed to ensure the convergence of optimal strategies.

Per discussions above, we adopt a stringent criterion of convergence by requiring all informed AI speculators' optimal strategies to stay unchanged for 1,000,000 consecutive periods. All $N = 1,000$ simulation sessions are simulated until convergence. The number of periods needed to reach convergence varies considerably across experiments depending on the particular choice of environment parameters. Moreover, even for the same experiment, the number of periods needed to reach convergence can vary significantly across the N simulation sessions, depending

[13] We do not have an explicit formula for the expected number of times a cell in the Q-matrix being visited by random exploration because the state variable p_{t-1} in $s_t = \{p_{t-1}, v_t\}$ is also affected by the noise trader's random order and the pricing rule adopted by the market maker.

on the realized values of random variables. Among all the experiments we study, the number of periods to reach convergence ranges from about 20 million to about 10 billion. To speed up computations, our programs are written in C++, using $-O2$ to optimize the compiling process. The C++ program is run with parallel computing in a high-powered-computing server cluster with 376 CPU cores in total. It takes about 1 min to 6 hours to finish all N simulation sessions in one experiment, depending on the total number of iterations needed to reach convergence.

4.9 Metrics Reflecting Collusive Behavior

Motivated by our theoretical results in Section 3, we calculate three simple metrics that can be indicative of potential collusive behavior among informed AI speculators. The values of all three metrics are computed in each simulation session over $T = 100,000$ periods, after informed AI speculators' optimal strategies fully converge to the limit strategies according to the convergence criterion in Section 4.8. By taking the average over a large number of periods, we smooth out the stochastic underlying economic environment, caused by the randomness in the noise trader's order u_t and the stochastic variation of the asset value v_t over time.

Collusion Capacity. The degree of collusion can be reflected by the Delta metric defined as follows:

$$\Delta^C = \frac{1}{I}\sum_{i=1}^{I}\Delta_i^C, \text{ with } \Delta_i^C = \frac{\overline{\pi}_i - \overline{\pi}_i^N}{\overline{\pi}_i^M - \overline{\pi}_i^N}, \tag{4.6}$$

where $\overline{\pi}_i \equiv \sum_{t=T_c}^{T_c+T}\pi_{i,t}(v_t,u_t)$ is the average profits of informed AI speculator i over T periods after Q-learning algorithms reach convergence at T_c. The values of $\overline{\pi}_i^N = \sum_{t=T_c}^{T_c+T}\pi_i^N(v_t,u_t)$ and $\overline{\pi}_i^M = \sum_{t=T_c}^{T_c+T}\pi_i^M(v_t,u_t)$ are the average profit that informed speculator i would obtain, theoretically, in the noncollusive Nash equilibrium or perfect cartel equilibrium, respectively. Because informed speculators are symmetric, we have $\pi_i^N(v_t,u_t) \equiv \pi^N(v_t,u_t)$ and $\pi_i^M(v_t,u_t) \equiv \pi^M(v_t,u_t)$ for all $i = 1,..,I$. Specifically, according to the formulas in Section 3.2, conditional on the realized values of v_t and u_t in period t, informed speculator i's profit in the noncollusive Nash equilibrium is

$$\pi^N(v_t,u_t) = \left[v_t - p^N(Ix^N(v_t)+u_t)\right]x^N(v_t), \text{ for } i = 1,\cdots,I, \tag{4.7}$$

where $x^N(v_t) = \chi^N(v_t-\overline{v})$ and $p^N(Ix^N(v_t)+u_t) = \overline{v} + \lambda^N(Ix^N(v_t)+u_t)$. Similarly, according to the formulas in Section 3.3, conditional on the realized values of v_t and u_t in period t, informed speculator i's profit in the perfect cartel equilibrium is

$$\pi^M(v_t,u_t) = \left[v_t - p^M(Ix^M(v_t)+u_t)\right]x^M(v_t), \text{ for } i = 1,\cdots,I, \tag{4.8}$$

where $x^M(v_t) = \chi^M(v_t-\overline{v})$ and $p^M(Ix^M(v_t)+u_t) = \overline{v} + \lambda^M(Ix^M(v_t)+u_t)$.

In principle, the value of Δ^C should range from 0 to 1. A larger Δ^C implies that informed AI speculators attain higher profits. The value of Δ^C can never be larger than 1 because $\overline{\pi}_i^M$ is the highest theoretically possible average profit. In fact, because informed AI speculators can

only choose actions over discrete grids, by design, it is not possible to obtain $\Delta^C = 1$ in our simulation experiments. However, it is possible to achieve a Δ^C below 0 under the limit strategies of informed AI speculators. This outcome implies that informed AI speculators failed to learn a good approximation of the actual Q-matrix, and as a result, they achieve average profits lower than those in the noncollusive Nash equilibrium.

Profit Gain Relative to Noncollusion. The Delta metric is informative about collusive behavior. However, it does not tell us the relative magnitude of supra-competitive profits. We thus also calculate the extra profit gain relative to the profits that informed AI speculators would obtain in the noncollusive Nash equilibrium theoretically. Specifically, the relative profit gain is $\sum_{i=1}^{I} \overline{\pi}_i / \sum_{i=1}^{I} \overline{\pi}_i^N$, where $\overline{\pi}_i$ and $\overline{\pi}_i^N$ are calculated similarly as those in equation (4.6).

Order Sensitivity to Asset Value. In our model, each informed speculator's order flows $x_{i,t}$ are linear in the asset value v_t, as captured by $x_{i,t} = \chi^C(v_t - \overline{v})$. Our model implies that informed speculators are more conservative in placing their orders if there is implicit collusion. That is, the sensitivity of order flows $x_{i,t}$ to the asset value $v_t - \overline{v}$ is lower when informed speculators collude more, i.e., $\chi^M \leq \chi^C < \chi^N$.

In our simulation experiments, informed AI speculators directly learn $x_{i,t}$ without imposing the linearity restriction between $x_{i,t}$ and v_t. Despite this, we find that informed AI speculators learn roughly linear strategies (see Figure 8). We estimate $\widehat{\chi}^C$ based on the recorded asset values and order flows $\{v_t, x_{i,t}\}_{t=T_c}^{T_c+T}$ for each AI speculator $i = 1, \cdots, I$, by running the following linear regression:

$$x_{i,t} = \chi_{i,0}^C + \chi_{i,1}^C v_t + \epsilon_t. \tag{4.9}$$

Consistent with our model, the estimates based on the simulated data satisfy $\widehat{\chi}_{i,0}^C \approx -\overline{v}\widehat{\chi}_{i,1}^C$ in the unrestricted regression (4.9). The estimate $\widehat{\chi}_{i,1}^C$ captures the sensitivity of $x_{i,t}$ to v_t corresponding to the optimal trading strategies after Q-learning algorithms converge. We further compute the average sensitivity of informed AI speculators as $\widehat{\chi}^C = \frac{1}{I}\sum_{i=1}^{I} \widehat{\chi}_{i,1}^C$.

4.10 Measures of Price Informativeness, Market Liquidity, and Mispricing

Price Informativeness. Consistent the our model, the degree of price informativeness in our simulation experiments is measured by the log signal-noise ratio as follows:

$$\mathcal{I}^C = \log\left[\frac{\text{var}(x_{i,t}^C)}{\text{var}(u_t)}\right] = \log\left[(I\widehat{\chi}^C)^2(\widehat{\sigma}_v/\sigma_u)^2\right], \tag{4.10}$$

where $\widehat{\sigma}_v$ is the standard deviation of v_t under our discrete grid points in $\mathbb{V}$.

Market Liquidity. Consistent the our model, the market liquidity in period t is measured by the inverse sensitivity of market maker's inventory $|z_t + y_t|$ to noise order flows u_t

$$\mathcal{L}_t^C = \frac{1}{\partial |z_t + y_t| / \partial u_t} = \frac{1}{|1 - \xi \widehat{\lambda}_t|}, \tag{4.11}$$

where $z_t = -\xi(p_t - \overline{v}) = -\xi \widehat{\lambda}_t y_t$ and $\widehat{\lambda}_t$ is given by equation (4.3). The average market liquidity is computed as $\mathcal{L}^C = \sum_{t=T_c}^{T_c+T} \mathcal{L}_t^C$.

Mispricing. Consistent the our model, the magnitude of mispricing in period t is measured by the percentage deviation of the asset's price p_t from its conditional expected value

$$\mathcal{E}_t^C = \left| \frac{p_t - \mathbb{E}[v_t|y_t]}{\mathbb{E}[v_t|y_t] - \overline{v}} \right| = \left| \frac{\widehat{\lambda}_t - \widehat{\gamma}_{1,t}}{\widehat{\gamma}_{1,t}} \right|, \tag{4.12}$$

where $p_t = \widehat{\gamma}_{0,t} + \widehat{\lambda}_t y_t$ and $\mathbb{E}[v_t|y_t] = \widehat{\gamma}_{0,t} + \widehat{\gamma}_{1,t} y_t$; $\widehat{\gamma}_{0,t}$ and $\widehat{\gamma}_{1,t}$ are estimated from (4.2). The average mispricing is computed as $\mathcal{E}^C = \sum_{t=T_c}^{T_c+T} \mathcal{E}_t^C$.

5 AI Collusion under Information Asymmetry

Our model suggests that informed speculators can achieve supra-competitive profits through implicit collusion when both price efficiency and noise trading risks are low (see Proposition 3.4). In this section, we conduct simulation experiments with informed AI speculators whose trading is powered by Q-learning algorithms. We are mainly interested in four questions. First, can informed AI speculators learn to collude, even without communicating with each other or possessing any information about the underlying economic environment? Second, if collusion exists, what are the mechanisms that generate such collusive behavior among informed AI speculators? Third, how price efficiency and noise trading risk affect the trading strategies of informed AI speculators. Fourth, what are the implications of AI-powered trading for price informativeness, market liquidity, and mispricing in financial markets?

In Subsection 5.1, we show that in environments with low price efficiency and low noise trading risks, informed AI speculators are able to learn price-trigger strategies to achieve implicit collusion, which is quite similar to the mechanism characterized in our model in Section 3.4. In Subsection 5.2, we show that in environments with low price efficiency and high noise trading risks, informed AI speculators are not able to learn price-trigger strategies to achieve collusion, as predicted by our model. However, they can still achieve supra-competitive profits due to biased learning. The equilibrium of informed AI speculators resembles a self-confirming equilibrium (Fudenberg and Kreps, 1988; Fudenberg and Levine, 1993) with collusion rather than a Nash equilibrium. In Subsection 5.3, we study the role of price efficiency and noise trading risks in determining informed AI speculators' profits and collusive behavior. In Subsection 5.4, we illustrate informed AI speculators' trading strategies. Finally, in Subsection 5.5, we study the

implications of AI-powered trading for price informativeness, market liquidity, and mispricing in financial markets.

5.1 Artificial Intelligence: Collusion through Price-Trigger Strategies

In this subsection, we study informed AI speculators' behavior when the environment has low price efficiency (i.e., $\xi = 500$) and low noise trading risks (i.e., $\sigma_u/\sigma_v = 10^{-1}$). The other parameters are set according to the baseline economic environment described in Section 4.7. Across all $N = 1,000$ simulation sessions, the average value of Δ^C is about 0.73 and the average profit of informed AI speculators is about 9% higher than the profit in the noncollusive equilibrium. Thus, our simulation results indicate that informed AI speculators can achieve supra-competitive profits. Below, we examine the mechanism that sustains their collusion. We show that informed AI speculators are intelligent enough to learn price-trigger strategies, which allows them to sustain collusion after their Q-learning algorithms converge. These simulation results with informed AI speculators are similar to the theoretical predictions of our model with rational-expectation informed speculators.

5.1.1 Price-Trigger Strategy

Motivated by our model, we examine whether the optimal strategies learned by informed AI speculators are consistent with the price-trigger strategy illustrated in Section 3. To this end, in Figure 1, we study the impulse response function (IRF) after an exogenous shock to the noise order flow, which further affects the asset's price given the market maker's pricing rule.

Specifically, in each of the $N = 1,000$ simulation sessions, we focus on the economic environment after informed speculators' Q-learning algorithms converge. Throughout the IRF experiment, for all $t \geq 0$, both informed AI speculators play their learned optimal strategies and the asset's price p_t is determined by the market maker according to its learned pricing rule. In period $t = 3$, we introduce an unexpected exogenous shock Δu_t to the noise order flow u_t. The direction of the shock is made to mimic the price impact of a hypothetical profitable deviation from informed AI speculators. That is, we choose $\Delta u_t > 0$ if $v_t > \overline{v}$ and $\Delta u_t < 0$ if $v_t < \overline{v}$. Thus, all else equal, this exogenous shock will unexpectedly increase the asset's price p_t if $v_t > \overline{v}$ and decrease p_t if $v_t < \overline{v}$.

We are interested in the IRF of three outcome variables. The first outcome variable is the price's percentage deviation from its long-run mean, defined by $(\widetilde{p}_t - \mathbb{E}[\widetilde{p}_t])/\mathbb{E}[\widetilde{p}_t]$, where $\widetilde{p}_t = (p_t - \overline{v})sgn(v_t - \overline{v})$ and $sgn(\cdot)$ is the sign function. The variable $\widetilde{p}_t$ captures the difference between the asset's price p_t and its expected value. The sign function ensures that $\widetilde{p}_t > 0$ because according to our model and simulation results, when $v_t > \overline{v}$, we have $p_t > \overline{v}$ and $sgn(v_t - \overline{v}) = 1$; when $v_t < \overline{v}$, we have $p_t < \overline{v}$ and $sgn(v_t - \overline{v}) = -1$. In addition, the definition of $\widetilde{p}_t$ ensures that the exogenous shock always increases its value, enabling us to take the average of IRF across simulation paths for expositional purposes. Specifically, if $v_t > \overline{v}$, the exogenous shock will increase p_t, and because $sgn(v_t - \overline{v}) = 1$, $\widetilde{p}_t$ also increases. If $v_t < \overline{v}$, the exogenous shock will decrease p_t, and because $sgn(v_t - \overline{v}) = -1$, $\widetilde{p}_t$ also increases. The second outcome variable is

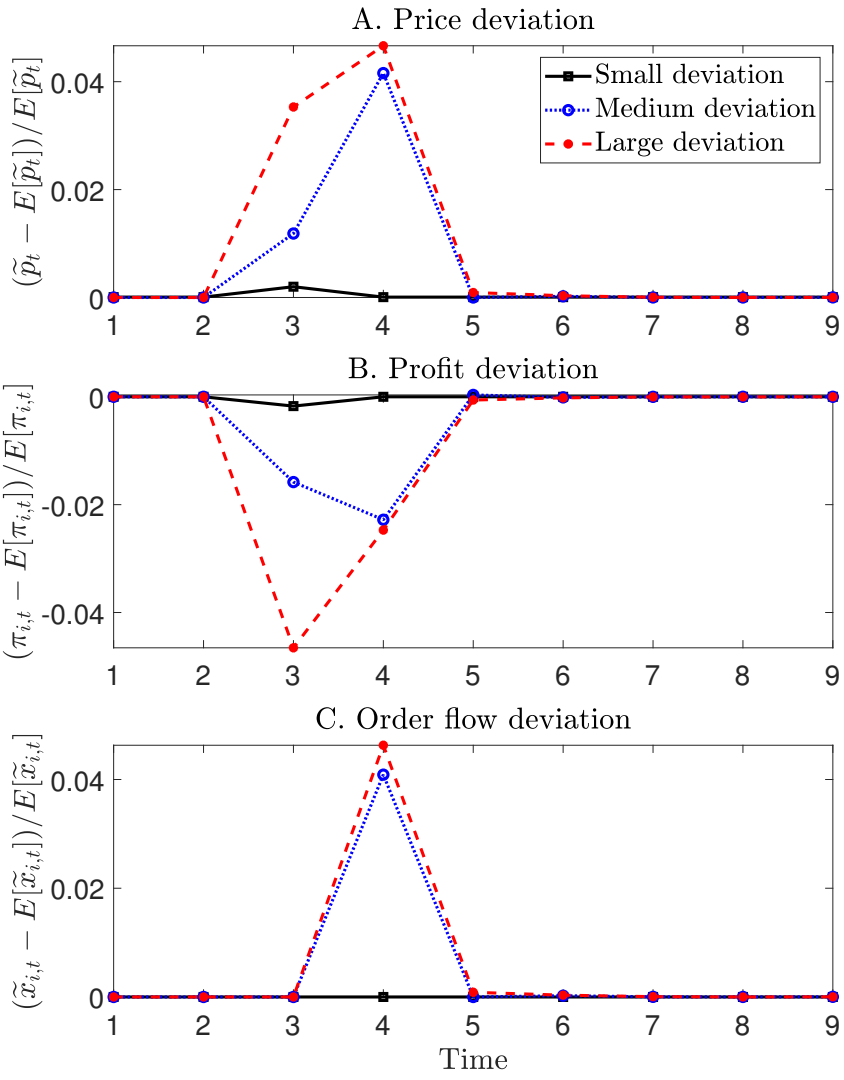
each informed AI speculator's profit's percentage deviation from its long-run mean, defined by $(\pi_{i,t} - \mathbb{E}[\pi_{i,t}])/\mathbb{E}[\pi_{i,t}]$. The third outcome variable is each informed AI speculator's order flow's percentage deviation from its long run mean, defined by $(\widetilde{x}_{i,t} - \mathbb{E}[\widetilde{x}_{i,t}])/\mathbb{E}[\widetilde{x}_{i,t}]$, where $\widetilde{x}_{i,t} = x_{i,t} sgn(v_t - \overline{v})$. The sign function ensures that $\widetilde{x}_{i,t} > 0$ because according to our model and simulation results, we have $x_{i,t} > 0$ when $v_t > \overline{v}$ and $x_{i,t} < 0$ when $v_t < \overline{v}$.

To clearly present the IRF, we calculate the average value of the above three interested outcome variables in two steps. First, for each of the $N = 1,000$ simulation sessions, we use the learned optimal strategies to simulate the IRF 10,000 times, with independently drawn random shocks to v_t and u_t. We smooth out the randomness in the economic environment by taking the average IRF across these 10,000 independent paths. This is referred to as the IRF for each simulation session $i = 1, ..., N$. Second, we compute the average IRF across $N = 1,000$ simulation sessions. This allows us to smooth out the randomness (i.e., initial states and exploration choices) during the learning process. However, our results hold not merely to the average IRF of $N = 1,000$ simulation sessions. Figure 2 plots the distribution of the impulse responses across the $N = 1,000$ simulation sessions. Although the magnitudes of the deviations in prices and trading flows differ significantly across simulation sessions, the $[25\%, 75\%]$ and $[5\%, 95\%]$ confidence intervals indicate that price-trigger strategies are consistently adopted by informed AI speculators.

Figure 1 plots the average IRF across the $N = 1,000$ simulation sessions for each outcome variable of interest. We consider exogenous shocks of different magnitudes. In the scenario with "small deviation," $|\Delta u_t|$ is roughly 0.5% of the average magnitude of informed AI speculators' order flow $|x_{i,t}|$. Thus, it generates a small impact on the asset's price p_t at $t = 3$. In the scenario with "medium deviation" and "large deviation," $|\Delta u_t|$ is about 2.5% and 7% of the average magnitude of informed AI speculators' order flow $|x_{i,t}|$, respectively, resulting in much larger changes in p_t.

Panel A plots the price's percentage deviation from its long-run mean. Due to the exogenous shock, the asset's price deviates from its long-run mean in period $t = 3$, and the size of price deviation increases with the magnitude of the exogenous shock. Panel B plots the profit's percentage deviation from its long-run mean for one informed AI speculator. The other informed AI speculator has similar profit dynamics. It is shown that in period $t = 3$, the price deviation reduces the informed AI speculator's profit, and the impact increases with the magnitude of the price deviation. Panel C plots the order flow's percentage deviation from its long-run mean for one informed AI speculator. The deviation is zero in period $t = 3$ because informed AI speculators submit their orders before observing the price in the same period.

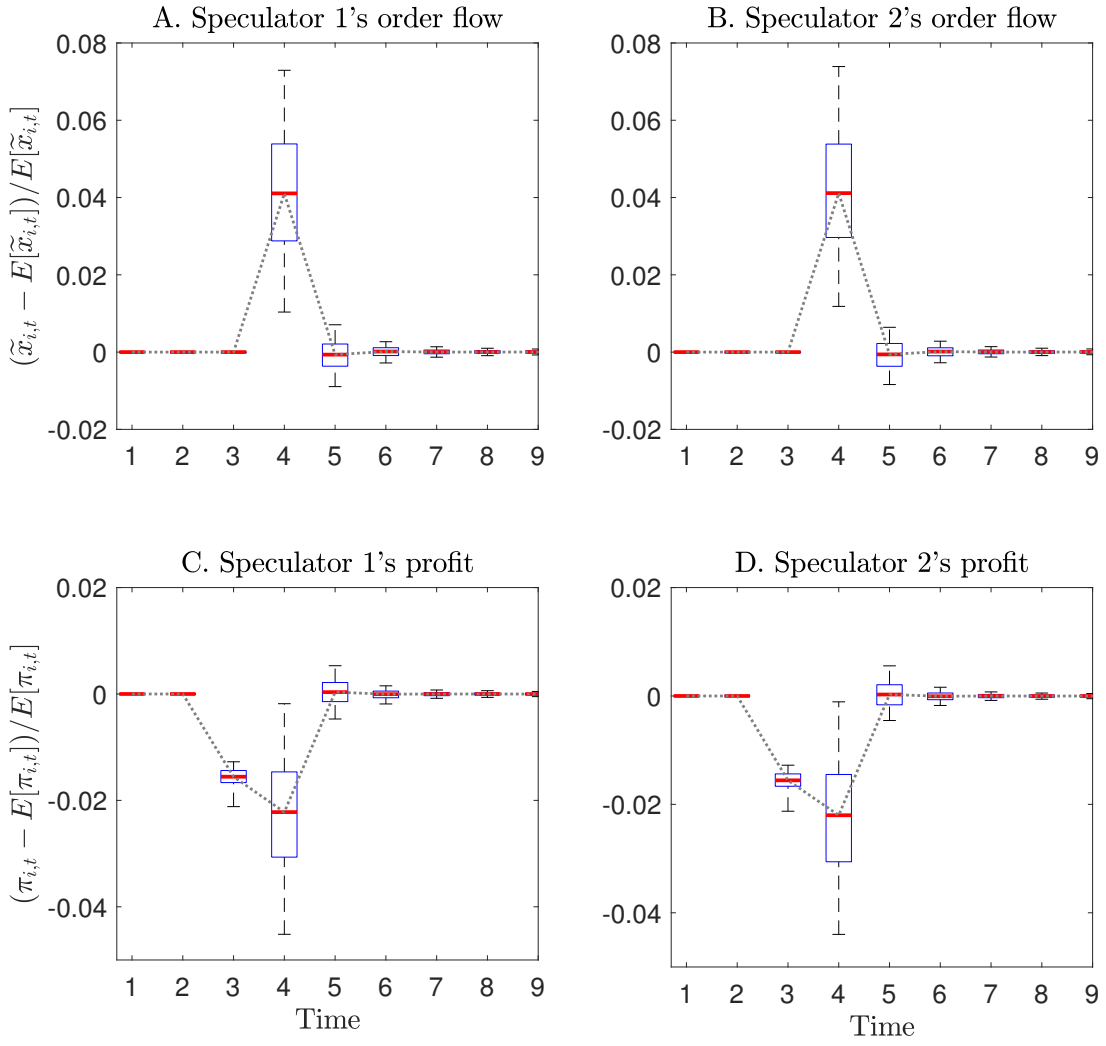
In period $t = 4$, panel C shows that in response to medium and large price deviations occurred in the previous period, the informed AI speculator's order flow significantly deviates from its long-run mean. Moreover, the magnitude of the order flow deviation is similar for the medium and large price deviation. However, the informed AI speculator's order flow does not respond to small price deviation. These patterns resemble the price-triggers strategies described in Section 3. Panel A shows that for the medium and large deviation cases, the percentage deviation of the asset's price continues to increase as a result of increased order flows from informed AI



Note: In each simulation session, we focus on the economic environment after informed speculators' Q-learning algorithms converge. Throughout the IRF experiment, for all $t \geq 0$, both informed AI speculators play their learned optimal strategies and the asset's price p_t is determined by the market maker according to its learned pricing rule. In period $t = 3$, we introduce an unexpected exogenous shock Δu_t to the noise order flow u_t. The direction of the shock is made to mimic the price impact of a hypothetical profitable deviation from informed AI speculators. That is, we choose $\Delta u_t > 0$ if $v_t > \overline{v}$ and $\Delta u_t < 0$ if $v_t < \overline{v}$. Thus, all else equal, this exogenous shock will unexpectedly increase the asset's price p_t if $v_t > \overline{v}$ and decrease p_t if $v_t < \overline{v}$. The three curves in each panel represent different magnitudes of the shock. Panel A plots the price's percentage deviation from its long-run mean. Panels B and C plot the percentage deviation of profit and order flow from its long-run mean for one informed AI speculator, respectively. All curves are average values across $N = 1,000$ simulation sessions, where each session is independently simulated 10,000 times to smooth out the effect of random shocks to v_t and u_t. We set $\sigma_u/\sigma_v = 10^{-1}$. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 1: IRF after an exogenous shock to u_t ($\sigma_u/\sigma_v = 10^{-1}$).

speculators. This, in turn, results in continued profit losses for informed AI speculators (see panel B). By contrast, for the small deviation case, both of the asset's price and informed AI speculators'



Note: The experiment is similar to that described for Figure 1. Panels A and B plot the two speculators' order flow's percentage deviation from the long-run mean, and panels C and D plot their profit's percentage deviation from the long-run mean. In each panel, the dotted line represents the median value, the boxes represent the 25th and 75th percentiles, and the dashed intervals represent the 5th and 95th percentiles across $N = 1,000$ simulation sessions. Parameters are set as in Figure 1.

Figure 2: Confidence intervals for the IRF after an exogenous shock to u_t ($\sigma_u/\sigma_v = 10^{-1}$).

profits revert back to the long-run mean.

In period $t = 5$, panel C shows that informed AI speculators' order flows abruptly return to the long-run mean for both the medium and large deviation cases. As a result, both the price and profit deviation abruptly return to zero (see panels A and B).

5.1.2 Punishment for Deviation

According to our model in Section 3, price-trigger strategies are implemented based on whether the asset's price in the preceding period deviates from its long-run mean, which could be caused by either the random order flows from the noise trader or the order flows from informed AI speculators. Informed AI speculators cannot distinguish what causes price deviation.

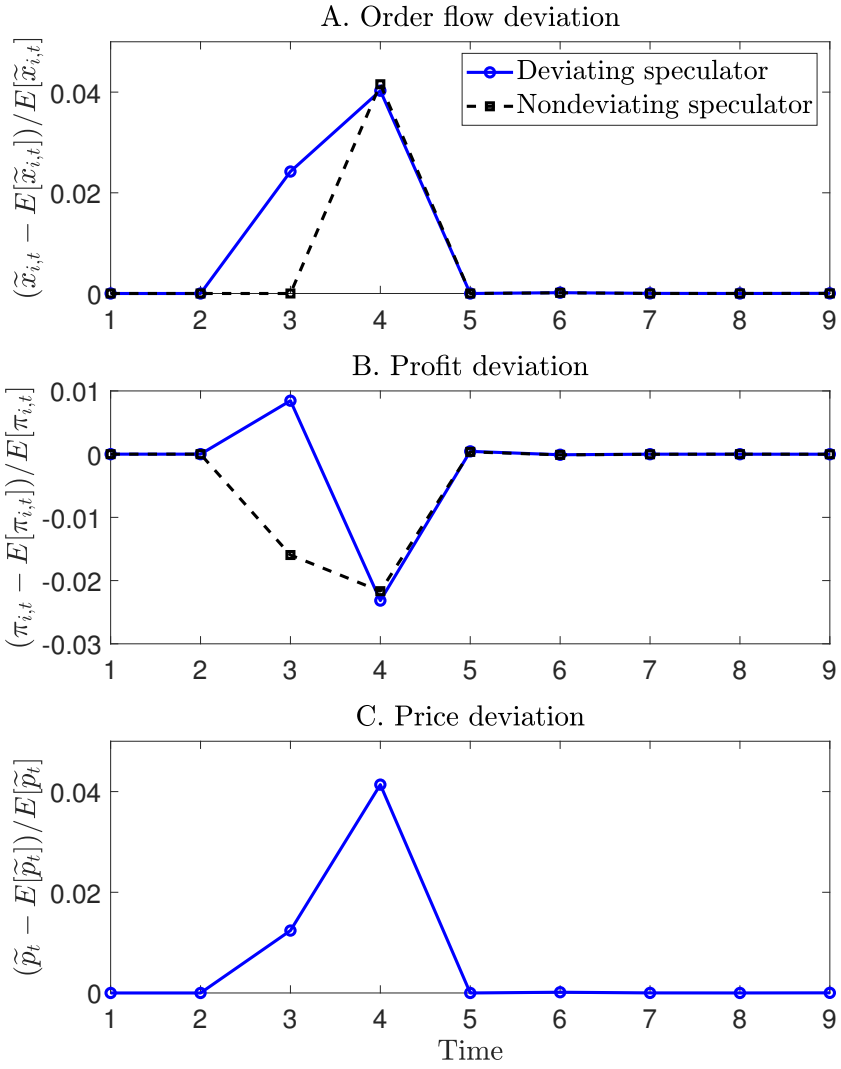
In this section, we complement the experiments in Section 5.1.1 by further studying the IRF after a unilateral deviation by one of the informed AI speculators. Specifically, in each of the $N = 1,000$ simulation sessions, we focus on the economic environment after informed speculators' Q-learning algorithms converge. Throughout the IRF experiment, for all $t \geq 0$, both informed AI speculators play their learned optimal strategies and the asset's price p_t is determined by the market maker according to its learned pricing rule. In period $t = 3$, we exogenously force one informed AI speculator i to have a one-period deviation from its learned optimal strategy. The one-period deviation in period $t = 3$ is made to the direction that increases the contemporaneous profit of the deviating speculator (i.e., we exogenously increase the deviating speculator's order by $\Delta x_{i,t}$ if $v_t > \overline{v}$ and reduce its order by $\Delta x_{i,t}$ if $v_t < \overline{v}$). We choose the deviation size $\Delta x_{i,t}$ to be one grid point in the order space $\mathbb{X}$, which ensures that the resulting price deviation is similar to the medium deviation case in panel A of Figure 1 for comparison purposes.

Panel A of Figure 3 plots the order flow's percentage deviation for both the deviating speculator and the nondeviating speculator. In period $t = 3$, on average, the deviating speculator's order flow deviates from the long-run mean by 2.5% while the nondeviating speculator's order flow remains unchanged. In period $t = 4$, the deviation gets punished as the nondeviating speculator behaves more aggressively, deviating its order flow from the long-run mean by 4.2%.

Rather than reducing its order flow, the deviating speculator further increases its order flow to 4.1% of the long-run mean in period $t = 4$, slightly below that of the nondeviating speculator. This form of overshooting exists for small deviations. As shown in panel A of Figure 5, if we consider a larger deviation, the deviating speculator would reduce its order flow in period $t = 4$. Regardless of whether its a small or a large deviation, both informed AI speculators abruptly return to the predeviation level of order flows in period $t = 5$.

Panel B of Figure 5 plots the profit's percentage deviation from its long-run mean for each informed AI speculator. In period $t = 3$, the deviating speculator's profit increases by 0.8% of the long-run mean while the nondeviating speculator's profit decreases by 1.6%. In period $t = 4$, due to the punishment strategy implemented by the nondeviating speculator, the profit of the deviating speculator drops substantially from 0.8% to -2.4% of the long-run mean. The expected discounted profit of deviation is about -1.6% of the long-run mean for the deviating speculator, indicating that deviation from the learned optimal strategies is not profitable.

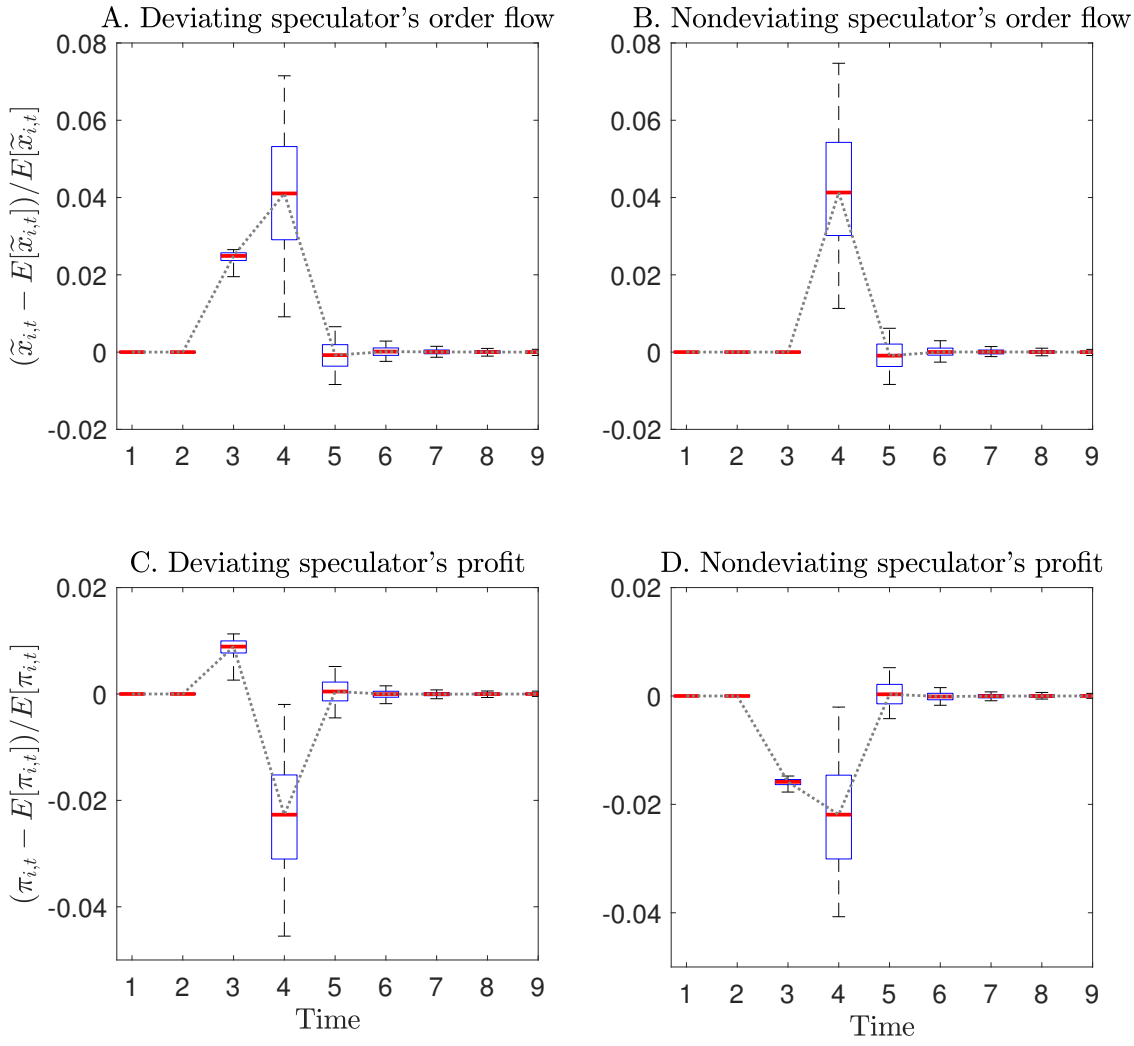
Panel C of Figure 3 plots the price's percentage deviation from its long-run mean. In period $t = 3$, due to the order deviation by one of the informed AI speculators, the asset's price deviates from its long-run mean by 1.2%. In fact, this is the force that triggered both informed AI speculators to change their order flows in period $t = 4$ because p_{t-1} is the only state variable that



Note: In each simulation session, we focus on the economic environment after informed speculators' Q-learning algorithms converge. Throughout the IRF experiment, for all $t \geq 0$, both informed AI speculators play their learned optimal strategies and the asset's price p_t is determined by the market maker according to its learned pricing rule. In period $t = 3$, we exogenously force one informed AI speculator i to have a one-period deviation from its learned optimal strategy. The one-period deviation in period $t = 3$ is made to the direction that increases the contemporaneous profit of the deviating speculator (i.e., we exogenously increase the deviating speculator's order flow by $\Delta x_{i,t}$ if $v_t > \overline{v}$ and reduce its order flow by $\Delta x_{i,t}$ if $v_t < \overline{v}$). The deviation size $\Delta x_{i,t}$ is one grid point in the order space $\mathbb{X}$. Panels A and B plot the percentage deviation of profit and order flow from its long-run mean for both informed AI speculator, respectively. Panel C plots the price's percentage deviation from its long-run mean. All curves are average values across $N = 1,000$ simulation sessions, where each session is independently simulated 10,000 times to smooth out the effect of random shocks to v_t and u_t. We set $\sigma_u/\sigma_v = 10^{-1}$. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 3: IRF after a unilateral deviation ($\sigma_u/\sigma_v = 10^{-1}$).

records the deviation status in the preceding period $t = 3$. The asset's price continues to increase to 4.2% in period $t = 4$ because of the overshooting in the deviating speculator's order flow, and



Note: The experiment is similar to that described for Figure 3. Panels A and B plot the two speculators' order flow's percentage deviation from the long-run mean, and panels C and D plot their profit's percentage deviation from the long-run mean. In each panel, the dotted line represents the median value, the boxes represent the 25th and 75th percentiles, and the dashed intervals represent the 5th and 95th percentiles across $N = 1,000$ simulation sessions. Parameters are set as in Figure 3.

Figure 4: Confidence intervals for the IRF after a unilateral deviation ($\sigma_u/\sigma_v = 10^{-1}$).

then abruptly returns to the long-run mean in period $t = 5$ as the two informed AI speculators revert to their predeviation behavior.

Figure 4 plots the distribution of the IRF across the $N = 1,000$ simulation sessions and shows that the deviating speculator gets punished through price-trigger strategies in most simulation sessions. To further show robustness, in panels A to C of Figure 5, we present the IRF of a larger deviation by setting $\Delta x_{i,t}$ equal to three grid points in the order space $\mathbb{X}$. The nondeviating speculator still implements a punishment strategy by substantially increasing its order flow in

period $t = 4$ to punish the deviating speculator's defect in period $t = 3$. The expected discounted profit of deviation is negative for the deviating speculator. In panels D to F of Figure 5, we present the IRF in an economic environment with higher noise trading risks by setting $\sigma_u/\sigma_v = 1$. In this environment, the two informed AI speculators achieve a small amount of supra-competitive profits with an average value of $\Delta^C = 0.2$. Even with such a low level of supra-competitive profits, we still see that the nondeviating speculator implements price-trigger strategies to deter deviations. However, the magnitude of both deviations and punishments in panels D to F of Figure 5 are smaller than those in Figure 3. This is consistent with a lower average Δ^C and the theoretical insight that collusive behavior becomes more difficult to achieve when informed AI speculators are less able to monitor peers' deviations due to the larger information asymmetry caused by higher noise trading risks.

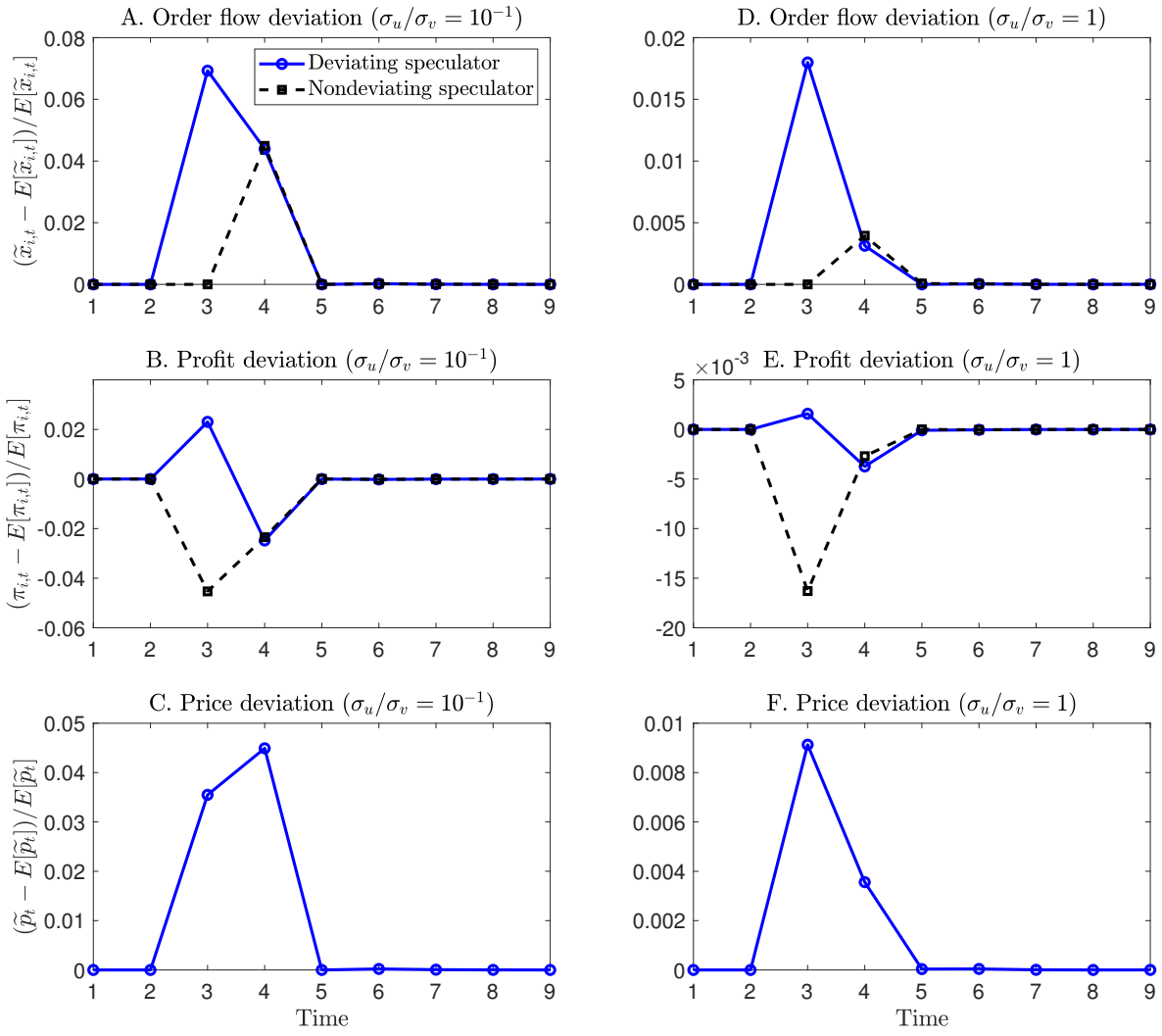
5.1.3 Discussions

Except for the duration of punishment, the impulse responses presented in Figures 1, 3 and 5 are quite consistent with the price-trigger strategies described in our model in Section 3. The patterns observed in our experiments coincide with our theoretical predictions that when the environment has low price efficiency and low noise trading risks, informed AI speculators are able to collude with each other by adopting price-trigger strategies to deter deviations. Moreover, collusion is more difficult to attain as noise trading risks become large.

Q-learning algorithms can learn price-trigger strategies because of experimentations. When one informed AI speculator switches to the exploration mode in the process of learning, it would choose actions randomly. Such behavior is effectively similar to defect from an implicit collusive agreement, if any. When this occurs, the two informed AI speculators would be trapped in the punishment phase until further explorations by one or both informed AI speculators occur. Informed AI speculators are able to learn coordination strategies because exploration modes will eventually stop, a necessary condition for Q-learning algorithms to converge.

Our finding that informed AI speculators are able to learn price-trigger strategies is similar to the finding of Calvano et al. (2020) that informed AI speculators learn grim-trigger strategies to sustain collusion in a perfect-information environment with Bertrand competition. However, different from Calvano et al. (2020), after punishment in period $t = 4$, rather than gradually returning to predeviation behavior, the informed AI speculators in our experiments abruptly return to their predeviation behavior. This difference is mainly due to the information asymmetry introduced by noise trading risks (i.e., $\sigma_u > 0$) and the stochastic asset value (i.e., $\sigma_v > 0$). Both model ingredients make informed AI speculators more difficult to sustain collusion by punishment threat, not just in the simulation experiments with informed AI speculators, but also in the model with rational-expectation informed speculators in Section 3.

In particular, our economic environment differs from that of Calvano et al. (2020) in two main aspects. First, we consider a stochastic environment where the asset's value v_t in each period is drawn from an i.i.d. distribution. In this stochastic setting, it becomes more difficult



Note: The experiment is similar to that described for Figure 3. The left three panels consider a larger deviation by setting $\Delta x_{i,t}$ equal to three grid points in the order space $\mathbb{X}$. The right three panels consider an economic environment with higher noise trading risks by setting $\sigma_u/\sigma_v = 1$. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 5: Robustness of IRF: larger deviation or higher noise trading risks ($\sigma_u/\sigma_v = 1$).

for the two informed AI speculators to learn punishment strategies to sustain collusion than in the deterministic setting with a constant v_t.[14] Second, the noise trader's random order flows generate information asymmetry to informed AI speculators, which makes grim-trigger strategies infeasible. As a result, informed AI speculators have to adopt price-trigger strategies to collude. In both the model with rational-expectation informed speculators and the simulation experiments

[14] In one of the robustness checks, Calvano et al. (2020) consider stochastic demand and show that the average Δ^C is lower when aggregate demand can take two values randomly. We also find that with stochastic v_t, the average Δ^C declines because it is more difficult for Q-learning algorithms to learn strong punishment strategies. The decline in Δ^C would be smaller if the evolution of v_t exhibits a smaller degree of randomness, either through a higher level of persistence or a less dispersed distribution.

with informed AI speculators, the ratio σ_u/σ_v plays a crucial role in determining the level of collusion.

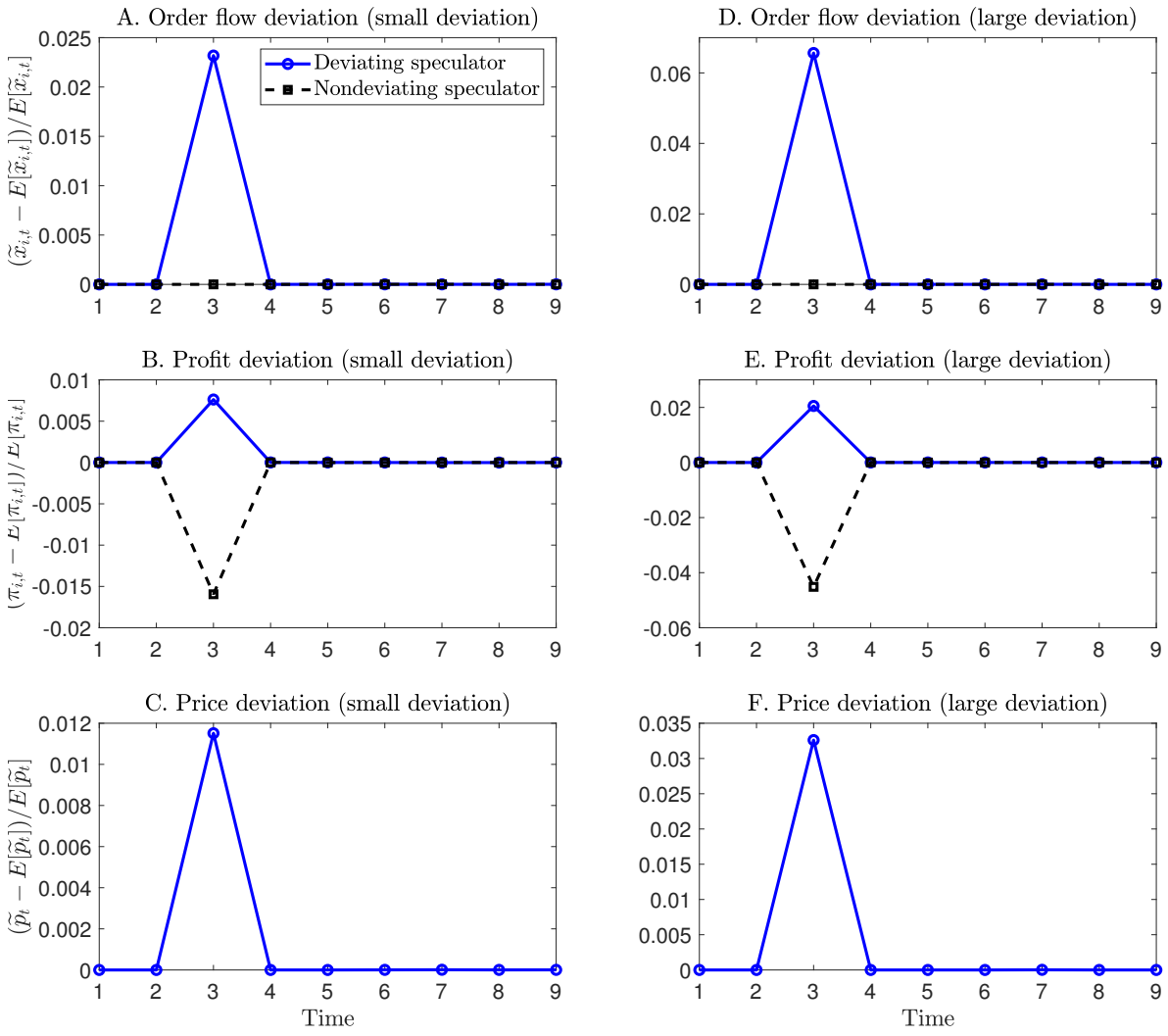
The information asymmetry in our economic environment implies that peer informed AI speculators' lagged actions are unobservable and thus cannot be included as state variables. Thus, as described in Section 4.1, we use the lagged asset's price p_{t-1} as the state variable in period t, rather than the lagged actions of the two informed AI speculators. Compared to our baseline setting with state variables $s_t = \{p_{t-1}, v_t\}$, we also examine the settings with alternative specifications of state variables. First, we consider a counterfactual setting with state variables $s_t = \{x_{i,t-1}, x_{-i,t-1}, v_t\}$. This setting essentially assumes that informed AI speculators' can perfectly observe peers' order flows, which is close to the perfect-information setting of Calvano et al. (2020) except for including v_t as an additional state variable. Second, we consider the setting where state variables are $s_t = \{p_{t-1}, x_{i,t-1}, v_t\}$. We find that under the perfect information benchmark (i.e., $\sigma_u/\sigma_v = 0$) with two informed AI speculators, these two alternative settings have almost the same average Δ^C. This is not surprising because under the perfect information benchmark, recording $x_{i,t-1}$ and p_{t-1} allows each informed AI speculator to exactly back out its peer's order flow $x_{-i,t-1}$. However, with information asymmetry (i.e., $\sigma_u/\sigma_v > 0$), the first setting with $s_t = \{x_{i,t-1}, x_{-i,t-1}, v_t\}$ yields a considerably higher average Δ^C than the other setting with $s_t = \{p_{t-1}, x_{i,t-1}, v_t\}$. In addition, we find that the average Δ^C in these two alternative settings is higher than that in our baseline setting. Thus, incorporating informed AI speculators' lagged actions as additional state variables indeed helps informed AI speculators to learn collusive strategies, likely through an improved learning of punishment strategies. However, lagged actions are not a necessary ingredient because in both our model with rational-expectation informed speculators and simulation experiments with informed AI speculators, including lagged price p_{t-1} alone can already result in a significant degree of collusion.

5.2 Artificial Stupidity: Collusion through Homogenized Learning Biases

In this subsection, we study informed AI speculators' learned optimal strategies when the environment has low price efficiency but large noise trading risks (i.e., $\sigma_u/\sigma_v = 10^2$). Similar to Section 5.1, the other parameters are set according to the baseline economic environment.

According to our model in Section 3, it is impossible for informed speculators to collude with each other in environments with large noise trading risks. However, in our simulation experiments, informed AI speculators can still achieve supra-competitive profits. Across $N = 1,000$ simulation sessions, the average value of Δ^C is about 0.6 and the average profit of informed AI speculators is about 7.5% higher than the profit in the noncollusive equilibrium. The profit becomes even higher as noise trading risks further increase. Below, we examine the mechanism that leads to such supra-competitive profits. We show that in line with our model's prediction, informed AI speculators do not learn price-trigger strategies to sustain collusion. Instead, they are able to collude to achieve supra-competitive profits due to homogenized learning biases.

To begin with, we study the impulse responses to a unilateral deviation in Figure 6. Clearly,



Note: The experiment is similar to those described for Figure 3, except for setting $\sigma_u/\sigma_v = 10^2$. The left three panels consider a unilateral small deviation with deviation size $\Delta x_{i,t}$ equal to one grid point in the order space $\mathbb{X}$. The right three panels consider a unilateral large deviation with $\Delta x_{i,t}$ equal to three grid points in $\mathbb{X}$.

Figure 6: IRF after a unilateral deviation ($\sigma_u/\sigma_v = 10^2$).

regardless of whether it is a small deviation (panels A to C) or a large deviation (panels D to F), we do not see any punishment from the nondeviating speculator. Instead, panels A and D of Figure 6 show that the nondeviating speculator's order flow is virtually unchanged and the deviating speculator returns to its learned optimal trading strategy immediately in period $t = 4$, which is just one period after the deviation. Panels B and E of Figure 6 show that the deviating speculator obtains an extra amount of one-period profit in period $t = 3$, which causes a one-period profit loss for the nondeviating speculator. Because there is no punishment for $t \geq 4$, the average percentage gains from the deviation in terms of discounted profits is strictly positive for the deviating speculator.

5.2.1 Self-Confirming Equilibrium

The collusive outcomes achieved by the two informed AI speculators are clearly not generated by price-trigger strategies when σ_u/σ_v is large, which is consistent with the prediction of our model (Proposition 3.4). In fact, the collusive outcomes are achieved through homogenized learning biases of informed AI speculators when noise trading risks are large. Although deviation seems to be profitable in terms of increasing the discounted profits, both informed AI speculators choose not to do this according to their learned optimal trading strategies after their Q-learning algorithms converge. The reason is that informed AI speculators' actions are governed by their learned Q-matrix, which indicates that the (no-deviation) strategies they are playing are optimal and any deviations cannot be profitable.

The steady-state behavior of informed AI speculators represents a self-confirming equilibrium, a notion first introduced by Fudenberg and Levine (1993). Compared with the Nash equilibrium, the self-confirming equilibrium is weaker because it allows players to have incorrect (or biased) off-equilibrium beliefs. This equilibrium concept is motivated by the idea that noncooperative equilibria should be interpreted as outcomes of a learning process, in which players form beliefs based on their past experience. While beliefs can be generally correct along the equilibrium path of play because it is frequently observed, beliefs are not necessarily correct off the equilibrium path unless players engage in a sufficient amount of experimentation with non-optimal actions (e.g., Fudenberg and Kreps, 1988, 1995; Cho and Sargent, 2008). Importantly, the incorrect off-equilibrium beliefs are not inconsistent with the evidence (i.e., outcomes along the equilibrium path). As noted by Fudenberg and Levine (1993), any self-confirming equilibrium can be a steady state, especially, including those equilibria with outcomes that cannot arise in Nash equilibrium. The self-confirming equilibrium allows completely arbitrary beliefs and supposes that players do not think strategically like what they do in a rational expectations framework. Instead, players choose actions based on what they have learned from their past experience.

In our simulations, informed AI speculators' beliefs are summarized by their Q-matrices. Specifically, the value of each state-action pair (s, x) in the Q-matrix represents the "perceived" reward that the informed AI speculator can obtain by playing the action $x \in \mathcal{X}$ in the state $s \in S$.[15] In Appendix G.1, we show that the hyperparameter α, which determines the informed AI speculator's forgetting rate or memory capacity, plays a crucial role in determining the magnitude of learning biases. Unbiased learning about the Q-matrix requires two conditions to hold simultaneously: 1) the informed AI speculators have sufficiently experimented all possible off-equilibrium plays before Q-learning algorithms converge, and 2) informed AI speculators' memory capacity is infinitely large, i.e., $\alpha \to 0$. As long as $\alpha > 0$, the Q-matrix is learned with biases due to the failure of the law of large numbers. Moreover, learning biases are larger when noise trading risks are higher (i.e., higher σ_u/σ_v) or the forgetting rate α is higher. Intuitively, informed AI speculators average past data to approximate the moments of the conditional probability

[15] As we show in Appendix G.1, when $\rho = 0$, the value of each state-action pair (s, x) in the Q-matrix is equal to the sum of the discounted value of the profits $(v - p)x$ received by the informed AI speculator when it played x in state s in the past, with the discount rate being $1 - \alpha$.

distribution of interest. When the environment's has higher noise trading risks or the forgetting rate α is higher, informed AI speculators lack sufficient memory capacity to store and analyze past data, and thus it becomes more difficult to approximate the moments of interest (i.e., the Q-matrix). The magnitude of learning biases in turn will determine which self-confirming equilibrium would emerge after Q-learning algorithms converge.

5.2.2 Biased Learning Leads to Self-Confirming Equilibrium with Supra-Competitive Profits

Having discussed that the steady state reached by informed AI speculators represents a self-confirming equilibrium, we now further explain why informed AI speculators' biased learning leads to collusive rather than competitive outcomes.

The underlying logic involves the following four key steps. First, collusive outcomes are achieved when informed AI speculators adopt more conservative, rather than more aggressive trading strategies. Specifically, according to our model in Section 3, the sensitivities of informed speculators' order flow to the asset's value v_t in different equilibria satisfy $\chi^M \leq \chi^C < \chi^N$. Because informed speculator i's order $x_{i,t}$ is $x_{i,t} = \chi(v_t - \overline{v})$, its absolute value of order flow satisfies $|x^M_{i,t}| \leq |x^C_{i,t}| < |x^N_{i,t}|$ for any v_t, indicating that collusion means that informed speculators adopt more conservative (i.e., trading with smaller absolute value of order flow $|x_{i,t}|$), rather than more aggressive trading strategies.

Second, compared with more conservative trading strategies, when informed AI speculators adopt more aggressive trading strategies, the unconditional variance of per-period profits is larger, namely, the distribution of per-period profits is more dispersed. Specifically, in Appendix G.2, we show that, for any state s, there exists complementarity between an informed AI speculator's order flow x and the noise order flow u_t in determining per-period profits. This complementarity implies that more aggressive trading strategies would amplify the impact of the noise order flow u_t, generating a more dispersed distribution of per-period profits compared to that generated by more conservative trading strategies.

Third, if playing action x in state s generates a more dispersed distribution of per-period profits, the resulting estimated Q value, $\widehat{Q}_t(s,x)$, for the state-action pair (s,x) also has a more dispersed distribution over time. This is because at any point in time t, the estimated $\widehat{Q}_t(s,x)$ is the sum of the discounted value of per-period profits that the informed AI speculator receives when it visits the state-action pair (s,x) in the past.

Fourth, a necessary condition for all Q-learning algorithms to reach stationary optimal strategies is that exploration rate is virtually zero, and informed AI speculators are purely in the exploitation mode. However, because of exploitation, for any state s, the action x that generates a more dispersed distribution of $\widehat{Q}_t(s,x)$ over time is less likely to be adopted by informed AI speculators after their Q-learning algorithms converge. Specifically, relative to playing conservative actions, playing an aggressive action (denoted by x^*), generates a dispersed distribution of $\widehat{Q}_t(s,x^*)$ over time. This means that an aggressive action x^* is likely to generate both a high $\widehat{Q}_t(s,x^*)$ and a low $\widehat{Q}_t(s,x^*)$. In one case, suppose a sequence of unfavorable noise order flows

were realized when the informed AI speculator was playing x^* in state s, so that a low $\widehat{Q}_t(s, x^*)$ is estimated for x^*. Then, x^* will not be played when the informed AI speculator conducts exploitation in state s in the future, because this action obviously does not maximize its Q value. In the other case, suppose a sequence of favorable noise order flows were realized when the informed AI speculator was playing x^* in state s, so that a high $\widehat{Q}_t(s, x^*)$ is estimated for x^*. Then, x^* will be further "exploited" in future periods. Because x^* generates a more dispersed $\widehat{Q}_t(s, x^*)$, it is highly likely that, eventually, the estimated $\widehat{Q}_t(s, x^*)$ will be small. From this point on, like the first case, the informed AI speculator will not play x^* when conducting future exploitation in state s. Thus, in the process of reaching convergence, the informed AI speculator's exploitation has the tendency to not adopt the trading strategies that can possibly generate large negative Q values, which are aggressive trading strategies that generate a more dispersed distribution of per-period profits. In some sense, informed AI speculators exhibit a certain degree of aversion to risks in the exploitation mode.

Taking the above four steps together, informed AI speculators' biased learning leads them to adopt more conservative trading strategies after their Q-learning algorithms converge, resulting in collusive outcomes.

5.2.3 Homogenized Bias and Implicit Coordination

We have explained how informed AI speculators' learning biases and exploitation lead to a self-confirming equilibrium that features collusive outcomes. However, it remains unclear why informed AI speculators adopt highly similar trading strategies after their Q-learning algorithms converge. What is the fundamental force that generates this sort of implicit coordination? We find that the key reason is that informed AI speculators rely on the same foundational model in their learning process. This generates homogenized learning biases, eventually leading to implicit coordination.

To elaborate, first consider the economic environment represented by the trough point of the blue solid line in panel A of Figure 7, i.e., $\log(\sigma_u/\sigma_v) = 2$. This represents an environment with high price inefficiency but relatively low noise trading risks in the sense that learning biases are small for informed AI speculators. However, noise trading risks are large enough to rule out the existence of a collusive equilibrium sustained by price-trigger strategies. Because learning biases are small in this environment, informed AI speculators are able to learn to play a noncollusive Nash equilibrium after their Q-learning algorithms converge, resulting in an average $\Delta^C \approx 0$. Implicit coordination in this environment is achieved because both informed AI speculators adopt similar noncollusive trading strategies in the Nash equilibrium.

Next, suppose that noise trading risks in the economic environment become higher, all else equal, both informed AI speculators become more biased in their learning processes. This leads both of them to optimally choose more conservative trading strategies after their Q-learning algorithms converge. Because both informed AI speculators adopt the same Q-learning algorithm with the same forgetting rate α, the magnitudes of their learning biases are similar. Thus, they

also become more conservative at a similar pace, resulting in similar optimal trading strategies after their Q-learning algorithms converge, as if they are implicitly coordinating with each other. The homogenized bias in informed AI speculators' Q-learning algorithms allows them to attain similar levels of supra-competitive profits. The extent to which informed AI speculators are biased homogeneously determines the implicitly coordinated level of profits. Importantly, as noted above, the two informed AI speculators reach a self-confirming equilibrium in which no one will deviate, because their biased beliefs, as recorded in their learned Q-matrices, suggest that any deviation cannot be profitable.

By contrast, if the two informed AI speculators' learning processes are not governed by the same foundational model, the learning biases will not be homogeneized. As a result, the two informed AI speculators may not be able to simultaneously attain supra-competitive profits. As an illustrative example, in panel B of Figure 15, we consider an experiment in which one informed AI speculator adopts a more advanced algorithm than the other, as captured by a lower forgetting rate α. We find that the more advanced informed AI speculator is able to attain much higher profits than in the experiment with two informed AI speculators adopting the same α. However, the average profit of the less advanced informed AI speculator is much lower and similar to the profit in the noncollusive Nash equilibrium. In about half of the $1,000$ simulation sessions, the profits of the less advanced informed AI speculator are even lower than the profit in the noncollusive Nash equilibrium. This experiment highlights the importance of homogenized bias in generating implicit coordination and supra-competitive profits for all informed AI speculators. Further, in Section 7.2, we extend the Q-learning algorithm to a two-tier Q-learning algorithm in which informed AI speculators learn both the optimal choice of the forgetting rate α and the optimal trading strategies corresponding to the choice of α. Interestingly, we find that informed AI speculators will learn to coordinately adopt high values of α in the stationary equilibrium, and such coordination allows both of them to obtain supra-competitive profits through homogenized learning baises.

5.2.4 Determinants of the Magnitude of Learning Biases

The extent to which learning is biased determines which self-confirming equilibrium would emerge after Q-learning algorithms converge, which consequently determines the average profits of informed AI speculators. Specifically, the above mechanism is stronger when informed AI speculators' Q-matrices are estimated with larger biases. Thus, the extent to which informed AI speculators collude to attain supra-competitive outcomes increases with the magnitude of learning biases. We now discuss the determinants of the magnitude of learning biases.

As noted in Section 5.2.1, learning biases are larger in environments with higher noise trading risks (i.e., higher σ_u/σ_v) or when informed AI speculators have a higher forgetting rate α. In equation (G.6) in Appendix G.1, we formally show that the magnitude of learning biases increases when σ_u/σ_v is higher, λ is higher, ρ is lower, or α is higher. These properties behind Q-learning algorithms predict that informed AI speculators can attain higher supra-competitive profits due

to biased learning when σ_u/σ_v is higher, λ is higher, ρ is lower, or α is higher. Consistent these predictions, first, we show that the average Δ^C across $N = 1,000$ simulation sessions increases with σ_u/σ_v in the region with high noise trading risks (i.e., $\log(\sigma_u/\sigma_v) \geq 2$) in panel A of Figure 7. Second, we show that in the environment with high noise trading risks (e.g., $\log(\sigma_u/\sigma_v) = 2$), reducing ξ from 500 to 1 (which results in a larger λ and higher price efficiency) leads to a higher average Δ^C in panel B of Figure 7. Third, we show that in the environment with high noise trading risks, reducing the value of ρ leads to a higher average Δ^C in Figure 13. Finally, we show that in the environment with high noise trading risks, a higher α would result in a higher average Δ^C in panel B of Figure 14.

5.3 Role of Noise Trading Risks and Price Efficiency

In this subsection, we study the role of noise trading risk and price efficiency in generating collusive outcomes for informed AI speculators.

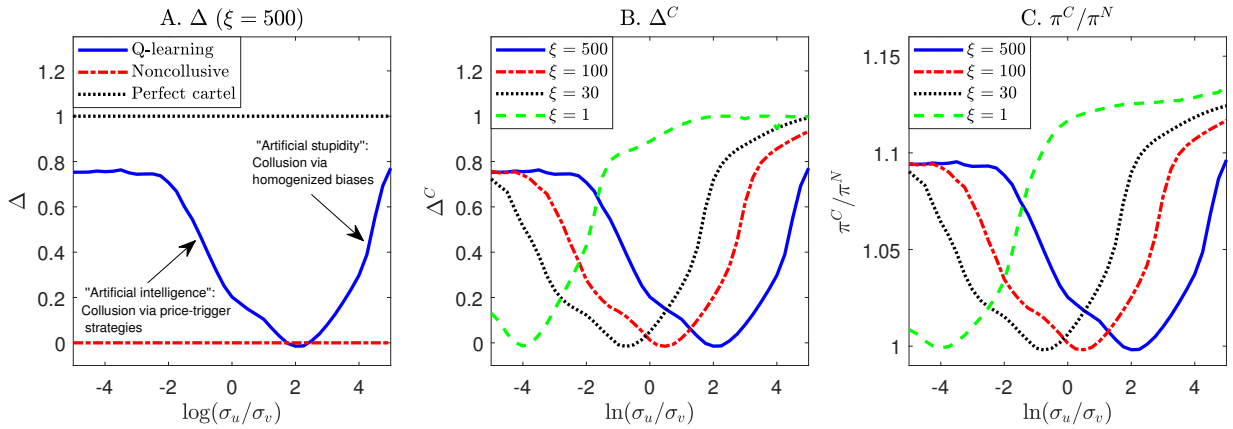
5.3.1 Role of Noise Trading Risks

Consider the baseline economic environment described in Section 4.7. In panel A of Figure 7, we plot the average Δ^C as $\log(\sigma_u/\sigma_v)$ varies from -5 to 5 along the x-axis. The black dotted and red dash-dotted lines represent the theoretical benchmarks ($\Delta^M = 1$ and $\Delta^N = 0$) in the perfect cartel and noncollusive Nash equilibrium, respectively. The blue solid line plots the average Δ^C across $N = 1,000$ simulation sessions, holding all other parameters unchanged. It shows that as $\log(\sigma_u/\sigma_v)$ increases along the x-axis, the average Δ^C first decreases and then increases. This U-shape pattern is an outcome of the interaction of the two mechanisms discussed in Sections 5.1 and 5.2. Specifically, in the region of low noise trading risks, i.e., $\log(\sigma_u/\sigma_v) < 2$, the average Δ^C is decreasing in $\log(\sigma_u/\sigma_v)$. In this region, informed AI speculators learn price-trigger strategies to sustain collusion and attain supra-competitive profits, as discussed in Section 5.1. The negative relationship between the average Δ^C and $\log(\sigma_u/\sigma_v)$ observed in our simulation experiments is consistent with the prediction of our model (see Proposition 3.6.(ii)).

In the region of large noise trading risks, i.e., $\log(\sigma_u/\sigma_v) \geq 2$, the average Δ^C is increasing in $\log(\sigma_u/\sigma_v)$. In this region, informed AI speculators attain supra-competitive profits because of homogenized learning biases, as discussed in Section 5.2. The positive relationship between the average Δ^C and $\log(\sigma_u/\sigma_v)$ observed in our simulation experiments is consistent with the theoretical property that biased learning becomes more significant when $\log(\sigma_u/\sigma_v)$ increases (see Section 5.2.4).

5.3.2 Role of Price Efficiency

According to our model in Section 3, the market maker focuses more on minimizing pricing errors when ξ is small or θ is large. In this case, price efficiency is high and there does not exists collusive Nash equilibrium sustained by price-trigger strategies for any $\sigma_u/\sigma_v > 0$ (Proposition 3.3). By



Note: This figure plots the average Δ^C and the profit gain relative to noncollusion (π^C/π^N) across $N = 1,000$ simulation sessions as $\log(\sigma_u/\sigma_v)$ varies along the x-axis, for different values of $\xi = 500, 100, 30, 1$. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 7: Δ^C and π^C/π^N for $\log(\sigma_u/\sigma_v) \in [-5, 5]$ and $\xi = 500, 100, 30, 1$.

contrast, when ξ is large or θ is small, the market maker focuses more on minimizing inventory costs. In this case, price efficiency is low and there exists a collusive Nash equilibrium that can be sustained by price-trigger strategies for small σ_u/σ_v and I (Proposition 3.4).

By varying the value of ξ in our simulation experiments, we study how price efficiency affects informed AI speculators' trading profits.[16] Specifically, the four curves in panel B of Figure 7 represent the experiments with $\xi = 500, 100, 30$ and 1. The overall U-shaped relationship between the average Δ^C and $\log(\sigma_u/\sigma_v)$ is not peculiar to the choice of ξ. All four curves display U-shape patterns. Panel C of Figure 7 plots the profit gain relative to noncollusion (π^C/π^N), the pattern is similar to that in panel A.

As we compare the four curves in panel B of Figure 7, one salient feature is that the trough of the U-shape shifts to the left as ξ decreases. This suggests that with a smaller ξ, a lower level of noise trading risks is necessary for informed AI speculators to learn price-trigger strategies to collude. A similar point can be made if we focus on the region with low noise trading risks, in which price-trigger strategies are learned by informed AI speculators. For example, holding $\ln(\sigma_u/\sigma_v) = -4$ unchanged, it is clear that the average Δ^C declines monotonically as ξ decreases from 500 to 1. Thus, collusion becomes more difficult to achieve through price-trigger strategies as ξ decreases, as predicted by our model (see Proposition 3.6.(iv)). By contrast, the relationship between ξ and average Δ^C is opposite if we focus on the region with large noise trading risks, in which informed AI speculators' trading strategies are dominantly affected by learning biases. For example, holding $\ln(\sigma_u/\sigma_v) = 2$ unchanged, it is clear that the average Δ^C increases monotonically as ξ decreases from 500 to 1. This is consistent with the theoretical property of biased learning discussed in Section 5.2.4, that is, the magnitude of learning biases increases with λ (i.e., decreases with ξ). Thus, a lower ξ leads to larger learning biases, allowing informed AI speculators to

[16] We do not conduct experiments with different θ because a smaller θ has similar impacts as a larger ξ on price efficiency.

achieve higher supra-competitive profits.

5.4 Trading Strategy of Informed AI Speculators

In this subsection, we illustrate informed AI speculators' trading strategies in the baseline economic environment described in Section 4.7.

In panel A of Figure 8, we plot the average sensitivity of informed AI speculators' order to the asset's value, $\widehat{\chi}^C$, across $N = 1,000$ simulation sessions as a function of the noise trading risk $\log(\sigma_u/\sigma_v)$. Consistent with panel A of Figure 7, $\widehat{\chi}^C$ displays an inverted U-shape as $\log(\sigma_u/\sigma_v)$ increases along the x-axis. By contrast, the theoretical benchmarks χ^N and χ^M stay roughly unchanged as $\log(\sigma_u/\sigma_v)$ increases.

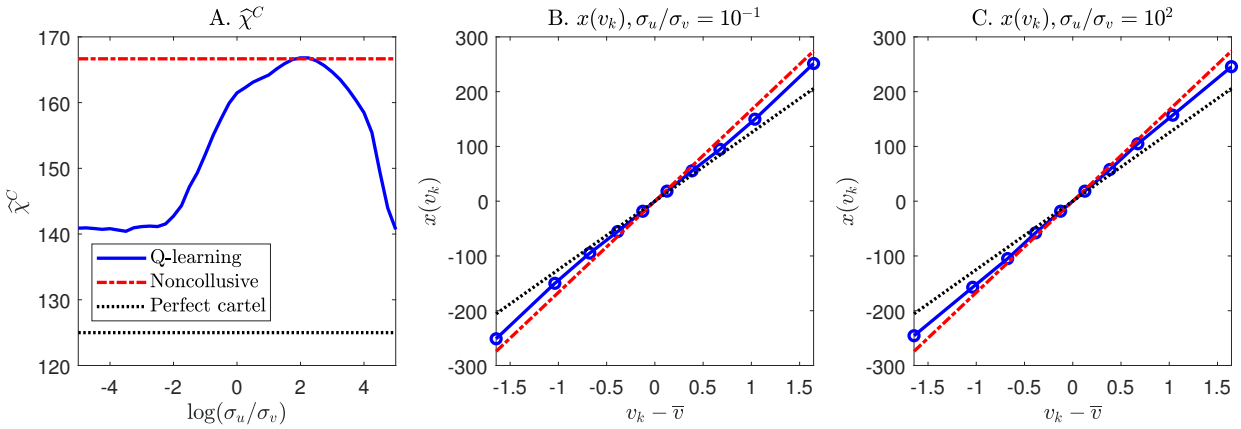
In fact, the estimated $\widehat{\chi}^C$ almost sufficiently describes informed AI speculators' trading strategy because their orders are almost linear in the assets's value, a property that holds both in the model and the simulation experiments. As an illustration, in panels B and C of Figure 8, we present the average trading strategy of informed AI speculators across $N = 1,000$ simulation sessions. Panel B is for the environment with low noise trading risks ($\sigma_u/\sigma_v = 10^{-1}$) and panel C is for the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^2$). The trading strategy in each simulation session is calculated as $x(v_k) = \frac{1}{In_p}\sum_{i=1}^{I}\sum_{m=1}^{n_p} x_i(p_m, v_k)$, which is the average order flow of I informed AI speculators across all grid points of $\mathbb{P}$, after Q-learning algorithms converge. The dots on the blue solid lines represent the average order flow corresponding to the discrete grid points of $\mathbb{V}$. The black dotted and red dash-dotted lines represent the theoretical benchmarks, $\chi^M(v_k - \overline{v})$ and $\chi^N(v_k - \overline{v})$, in the perfect cartel equilibrium and noncollusive Nash equilibrium, respectively.

It is clear that informed AI speculators learn an optimal trading strategy that is roughly linear in the asset's value after their Q-learning algorithms converge, even though the linearity restriction is not imposed during the learning process. Moreover, the slope of a linear fit for the trading strategy of informed AI speculators, i.e., $\widehat{\chi}^C$, lies between χ^M and χ^N in both panels B and C of Figure 8. Thus, the trading strategy learned by informed AI speculators is more conservatively than that in the noncollusive Nash equilibrium, which explains why informed AI speculators are able to attain supra-competitive profits.

5.5 Price Informativeness, Market Liquidity, and Mispricing

In this subsection, we study the impacts of AI collusion for price informativeness, market liquidity, and mispricing in financial markets. We show that AI collusion leads to lower price informativeness, lower market liquidity, and higher mispricing. The magnitude of such effects depends on the extent to which informed AI speculators collude with each other, which is largely determined by the noise trading risk σ_u/σ_v.

Panel A of Figure 9 plots the market's price informativeness relative to the theoretical benchmark of the perfect cartel equilibrium. By definition, the black dotted line shows that the relative price informativeness in the perfect cartel equilibrium is $\mathcal{I}^M/\mathcal{I}^M \equiv 1$. The red dash-dotted line



Note: Panel A plots the average $\widehat{\chi}^C$ across $N = 1,000$ simulation sessions as $\log(\sigma_u/\sigma_v)$ varies along the x-axis. Panels B and C plot the average trading strategy of informed AI speculators across $N = 1,000$ simulation sessions. The trading strategy in each simulation session is calculated as $x(v_k) = \frac{1}{In_p}\sum_{i=1}^{I}\sum_{m=1}^{n_p} x_i(p_m, v_k)$, which is the average order flow of I informed AI speculators across all grid points of $\mathbb{P}$, after Q-learning algorithms converge. The dots on the blue solid lines represent the average order flow corresponding to the discrete grid points of $\mathbb{V}$. Panels A and B focus on the environments with low ($\sigma_u/\sigma_v = 10^{-1}$) and high ($\sigma_u/\sigma_v = 10^{2}$) noise trading risks, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

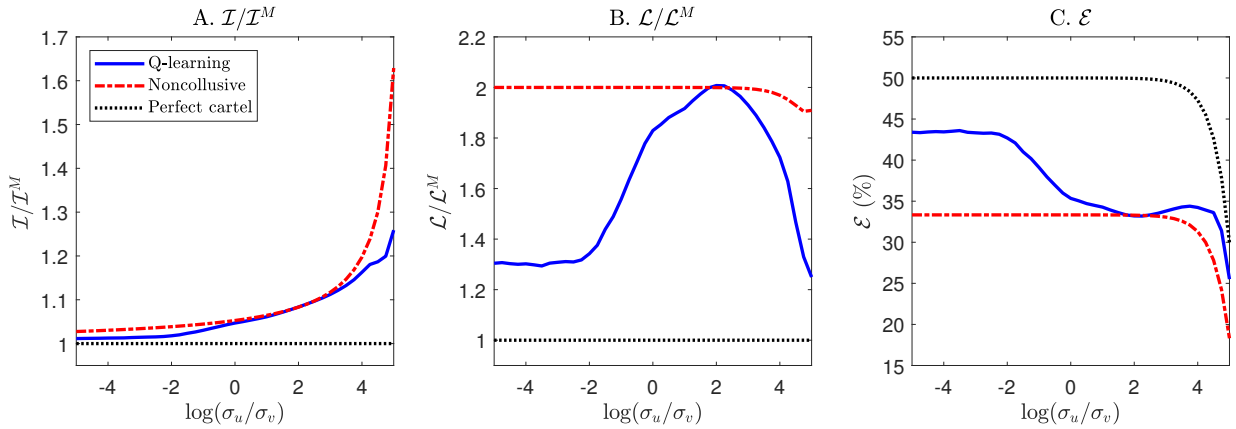
Figure 8: The trading strategy of informed AI speculators.

shows that the ratio of price informativeness in the theoretical benchmark of the noncollusive Nash equilibrium and perfect cartel equilibrium, $\mathcal{I}^N/\mathcal{I}^M$, is greater than 1 and increasing in $\log(\sigma_u/\sigma_v)$.[17] The blue solid line plots the average relative price informativeness, $\mathcal{I}^C/\mathcal{I}^M$, across $N = 1,000$ simulation sessions with informed AI speculators. Its value is close to the relative price informativeness in the theoretical benchmark of the non-collusive equilibrium when $\log(\sigma_u/\sigma_v)$ is around 2 due to the lack of collusion. When $\log(\sigma_u/\sigma_v)$ is very small or very large, the relative price informativeness in our simulation experiments with informed AI speculators is significantly lower than that in the theoretical benchmark of the noncollusive Nash equilibrium. The reason is that informed AI speculators place orders in a more conservative manner, with $\widehat{\chi}^C < \chi^N$, as shown in panel A of Figure 8.

Our findings suggest that perfect price informativeness is not achievable in the presence of informed AI speculators. In our simulation environments, when the noise trading risk σ_u/σ_v decreases, informed AI speculators would withhold their private information about the asset's value and collude more through price-trigger strategies, placing orders more conservatively than what they would do in the noncollusive Nash equilibrium. This AI collusion reduces price informativeness. Crucially, informed AI speculators never need to communicate with each other, whether explicitly or implicitly, the adoption of Q-learning algorithms automatically leads to such collusive behavior.

Panel B of Figure 9 plots the market liquidity relative to the theoretical benchmark of the perfect cartel equilibrium. The red dash-dotted line shows that the ratio of market liquidity in

[17] This is because $\widehat{\chi}^N > \widehat{\chi}^M$ for all $\log(\sigma_u/\sigma_v)$. Moreover, when $\xi = 500$, $\widehat{\chi}^N$ and $\widehat{\chi}^M$ are roughly unchanged (only slightly increase) as $\log(\sigma_u/\sigma_v)$ increases. Then, according to the equation (4.10), both $\mathcal{I}^N$ and $\mathcal{I}^M$ are decreasing in $\log(\sigma_u/\sigma_v)$, but the ratio $\mathcal{I}^N/\mathcal{I}^M$ is increasing in $\log(\sigma_u/\sigma_v)$.



Note: This figure plots the average values of different metrics across $N = 1{,}000$ simulation sessions as $\log(\sigma_u/\sigma_v)$ varies. Panels A and B plot the price informativeness and market liquidity relative to the theoretical benchmark of the perfect cartel equilibrium, i.e., $\mathcal{I}/\mathcal{I}^M$ and $\mathcal{L}/\mathcal{L}^M$, respectively. Panel C plots the magnitude of mispricing $\mathcal{E}$. The blue solid line represents the simulation experiments with informed AI speculators; the red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 9: Price informativeness, market liquidity, and mispricing for $\log(\sigma_u/\sigma_v) \in [-5, 5]$.

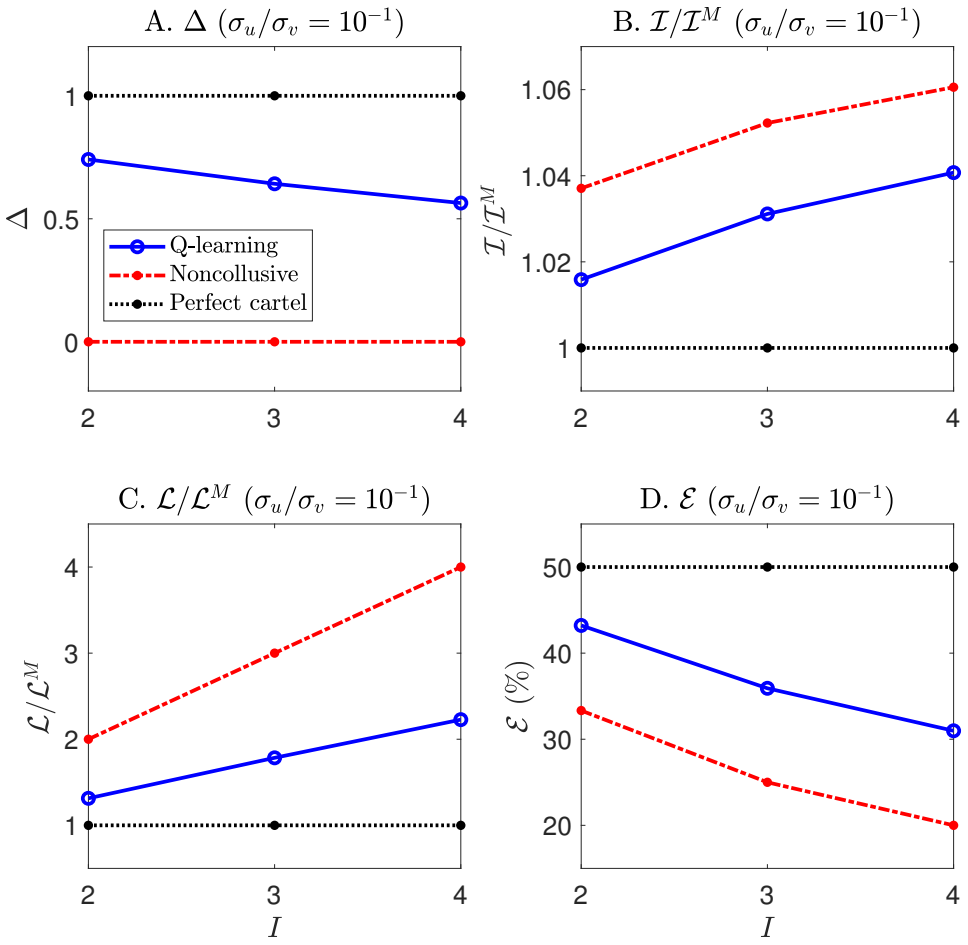
the theoretical benchmark of the noncollusive Nash equilibrium and perfect cartel equilibrium, $\mathcal{L}^N/\mathcal{L}^M$ is greater than 1 and decreasing in $\log(\sigma_u/\sigma_v)$.[18] The blue solid line shows that the market liquidity in our simulation experiments with informed AI speculators is higher than that in the theoretical benchmark of the perfect cartel equilibrium and lower than that of the noncollusive equilibrium. The blue solid line displays an U shape similar to panel A of Figure 8, indicating that the market liquidity is closer to the theoretical benchmark of the perfect cartel equilibrium if there is more AI collusion.

Panel C of Figure 9 plots the magnitude of mispricing in financial markets. Mispricing is higher in the theoretical benchmark of the perfect cartel equilibrium (the black dotted line) than in the noncollusive equilibrium (the red dash-dotted line). The blue solid line shows that AI collusion increases mispricing, and the magnitude is larger when there is a higher degree of collusion among informed AI speculators.

6 Further Inspection of Model Ingredients

In this section, we further inspect several key parameters in our simulation experiments. In Subsection 6.1, we study how the number of informed AI speculators affects their trading strategies. In Subsection 6.2, we study the implication of informed AI speculators' subjective discount rates. Finally, in Subsection 6.3, we study the impacts of hyperparameters α and β on informed AI speculators' learning outcomes.

[18] This is because $\lambda^N < \lambda^M$ for all $\log(\sigma_u/\sigma_v)$. Intuitively, in the perfect cartel equilibrium, the market maker knows that informed speculators submit orders jointly like a monopoly, and thus the market maker adopts a pricing rule that is more responsive to the combined order flow of informed speculators and the noise trader, i.e., $\gamma^N < \gamma^M$. As $\log(\sigma_u/\sigma_v)$ increases, both λ^N and λ^M decline, so that market liquidity defined by equation (4.11) increases.

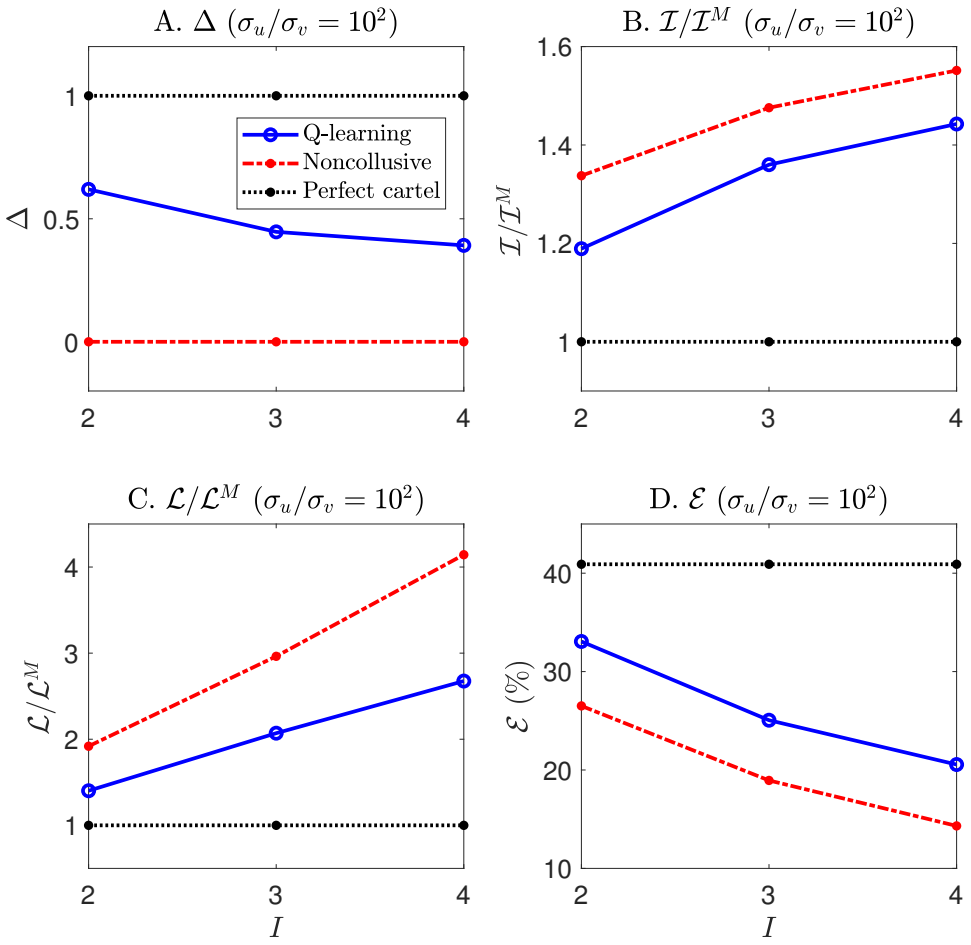


Note: The blue solid line plots the average values of Δ^C, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, and $\mathcal{E}^C$ across $N = 1{,}000$ simulation sessions as the number of informed AI speculators I varies, in the environment with low noise trading risks, i.e., $\sigma_u/\sigma_v = 10^{-1}$. The red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 10: Implications of the number of informed AI speculators ($\sigma_u/\sigma_v = 10^{-1}$).

6.1 Number of Informed AI Speculators I

Our model in Section 3 predicts that in the environment with low price efficiency (i.e., ξ is large or θ is small) and low noise trading risks (i.e., small σ_u/σ_v), informed speculators are less able to collude through price-trigger strategies when the number of informed speculators increases (see Proposition 3.6.(i)). In the simulation experiments with informed AI speculators, we find similar patterns. Specifically, consider the baseline economic environment described in Section 4.7. In Figure 10, we conduct simulation experiments in the environment with low noise trading risks ($\sigma_u/\sigma_v = 10^{-1}$). Panel A shows that as the number of informed AI speculators I increases from 2 to 4, the average Δ^C decreases from 0.74 to 0.56, indicating a decline in the extent of collusion among informed AI speculators. Moreover, panels B to D show that as I increases, the relative



Note: The blue solid line plots the average values of Δ^C, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, and $\mathcal{E}^C$ across $N = 1{,}000$ simulation sessions as the number of informed AI speculators I varies, in the environment with high noise trading risks, i.e., $\sigma_u/\sigma_v = 10^2$. The red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 11: Implications of the number of informed AI speculators ($\sigma_u/\sigma_v = 10^2$).

price informativeness $\mathcal{I}^C/\mathcal{I}^M$ and market liquidity $\mathcal{L}^C/\mathcal{L}^M$ increase whereas the magnitude of mispricing $\mathcal{E}^C$ decreases.

For comparisons, in Figure 11, we conduct simulation experiments in the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^2$). In these experiments, informed AI speculators collude through homogenized learning biases, as discussed in Subsection 5.2. The implications of I for informed AI speculators' strategies are similar to the experiments with low noise trading risks. Specifically, panel A shows that as I increases from 2 to 4, the average Δ^C decreases from 0.62 to 0.39. These results suggest that the coordination through homogenized learning biases becomes more difficult to achieve when there are more informed AI speculators in the market. Intuitively, the equilibrium degree of collusion is determined by the interaction of two counterveiling forces. One is the magnitude of learning biases, which is the mechanism that generates collusion. The

other is the deviation gain from the self-confirming collusive equilibrium. A larger deviation gain makes it more difficult for informed AI speculators to reach the collusive equilibrium because in the process of exploration (which, in essence, generates deviation behavior), these speculators will more likely learn to play noncollusive actions despite the existence of learning biases. As the number of informed AI speculators I increases, the deviation gain from the equilibrium trading strategies becomes larger, but the magnitude of learning biases remain unchanged.[19] Therefore, as I increases, collusion becomes more difficult and Δ^C declines.

Panels B to D show that as I increases, the relative price informativeness $\mathcal{I}^C/\mathcal{I}^M$ and market liquidity $\mathcal{L}^C/\mathcal{L}^M$ increase whereas the magnitude of mispricing $\mathcal{E}^C$ decreases.

6.2 Subjective Discount Rate ρ

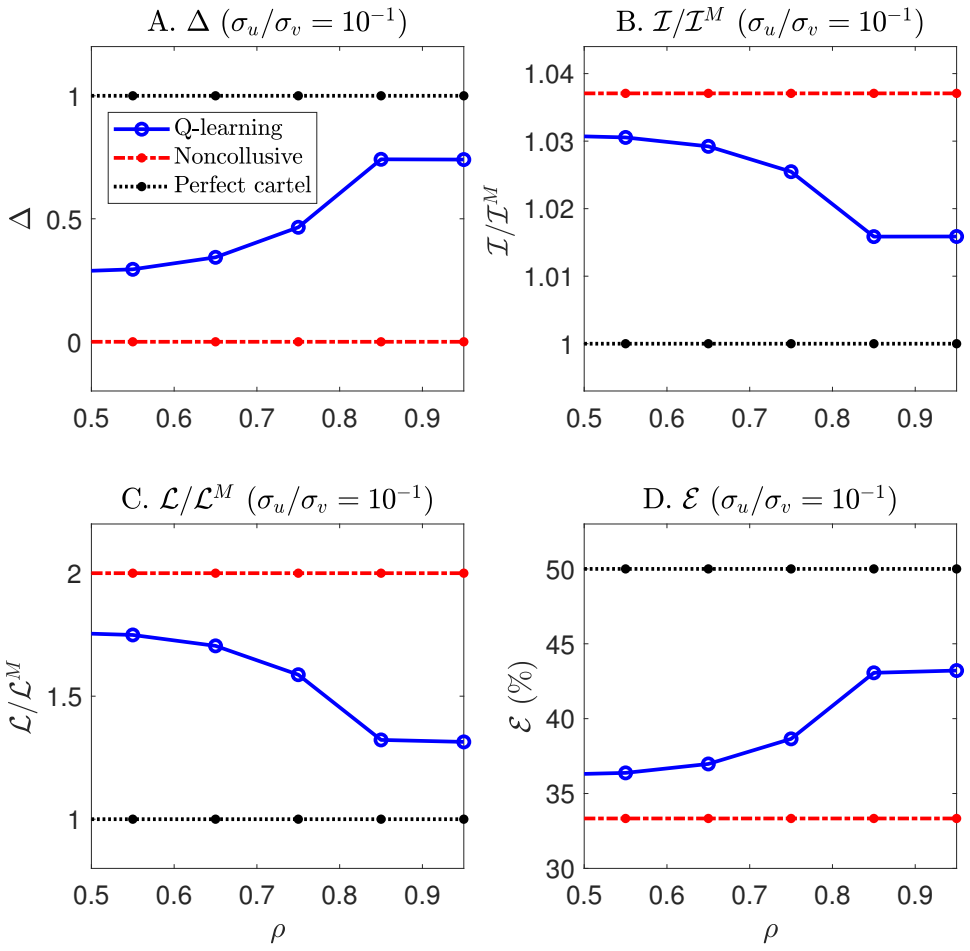
Our model in Section 3 predicts that in the environment with low price efficiency (i.e., ξ is large or θ is small) and low noise trading risks (i.e., small σ_u/σ_v), informed speculators are able to collude on higher profits through price-trigger strategies as the subjective discount rate ρ increases (see Proposition 3.6.(iii)). In the simulation experiments with informed AI speculators, we find similar patterns. Specifically, consider the baseline economic environment described in Section 4.7. In Figure 12, we conduct simulation experiments in the environment with low noise trading risks ($\sigma_u/\sigma_v = 10^{-1}$). Panel A shows that as ρ increases from 0.5 to 0.95, the average Δ^C increases from 0.29 to 0.74, indicating an increase in the extent of collusion among informed AI speculators. Moreover, panels B to D show that as ρ increases, the relative price informativeness $\mathcal{I}^C/\mathcal{I}^M$ and market liquidity $\mathcal{L}^C/\mathcal{L}^M$ decline whereas the magnitude of mispricing $\mathcal{E}^C$ increases.

Turning to the environment with high noise trading risks, the theoretical properties discussed in Section 5.2.4 indicate that as the subjective discount rate ρ increases, the magnitude of learning biases declines, and as a result, informed AI speculators would find it more difficult to collude. The patterns observed in our simulation experiments are consistent with this prediction. In particular, in Figure 13, we conduct simulation experiments in the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^2$). Panel A shows that as ρ increases from 0.5 to 0.95, the average Δ^C decreases from 0.76 to 0.62. Moreover, panels B to D show that as ρ increases, the relative price informativeness $\mathcal{I}^C/\mathcal{I}^M$ and market liquidity $\mathcal{L}^C/\mathcal{L}^M$ increase whereas the magnitude of mispricing $\mathcal{E}^C$ declines.

6.3 Hyperparameters α and β

In this subsection, we study how the hyperparameters α and β affect informed AI speculators' profits in equilibrium. Similar to the baseline economic environment, we consider informed AI speculators adopting the same values of α and β. In panel A of Figure 14, we plot the average Δ^C in the environment with low noise trading risks ($\sigma_u/\sigma_v = 10^{-1}$) for different values of α and β.

[19] When I increases, individual informed AI speculators trading flows x_i decrease. However, in equation (G.6), the trading flow x_i proportionally affects every term. Thus, the decrease in x_i does not affect the importance of the term $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^\tau u_{t(T-\tau)}$, which causes learning biases, relative to other terms in equation (G.6). This is why the magnitude of learning biases does not depend on I.

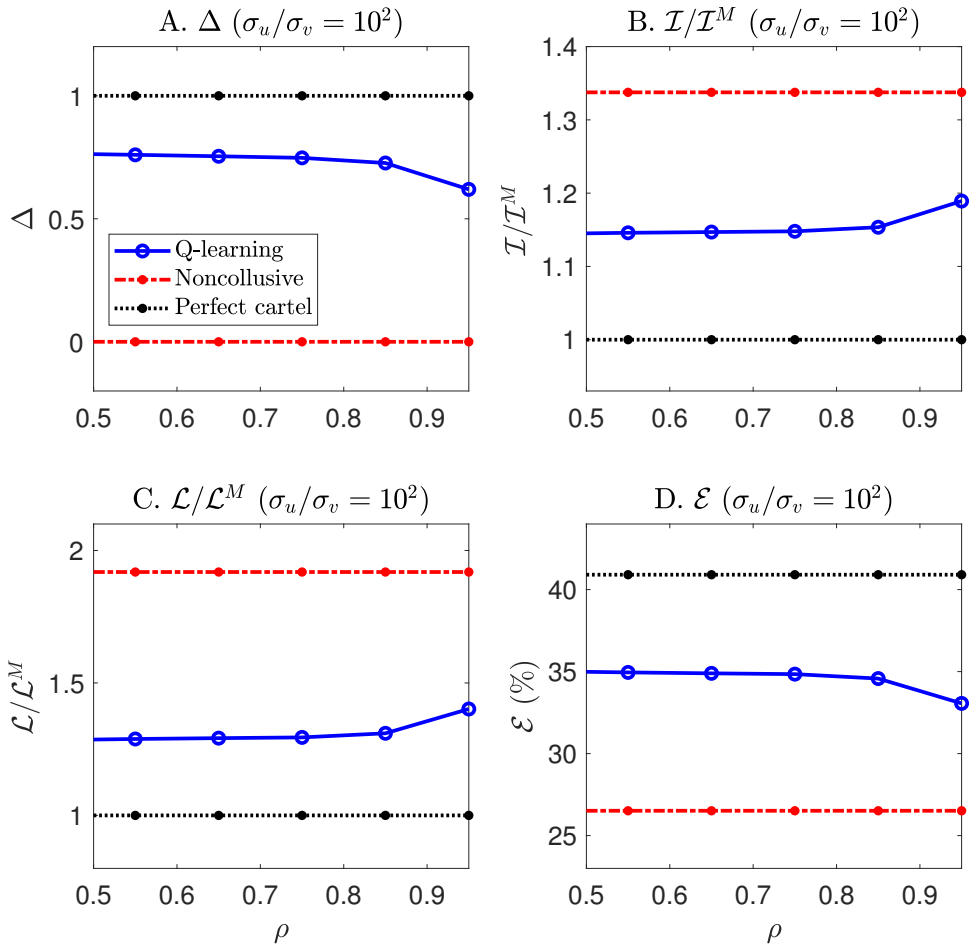


Note: The blue solid line plots the average values of Δ^C, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, and $\mathcal{E}^C$ across $N = 1,000$ simulation sessions as the subjective discount rate ρ varies, in the environment with low noise trading risks, i.e., $\sigma_u/\sigma_v = 10^{-1}$. The red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 12: Implications of the subjective discount rate ($\sigma_u/\sigma_v = 10^{-1}$).

As discussed in Subsection 5.1, informed AI speculators need to conduct sufficient explorations to learn punishment strategies, which is achieved by setting a sufficiently low β. Indeed, when $\beta = 10^{-6}$, the red bars in panel A of Figure 14 show that informed AI speculators can easily achieve a very high value of $\Delta^C = 0.90$ (corresponding to $\alpha = 0.001$) whereas when $\beta = 10^{-3}$, the yellow bars show that informed AI speculators can only achieve a low value of $\Delta^C = 0.40$ (corresponding to $\alpha = 0.1$).

Panel A of Figure 14 further shows that, to achieve the best collusive outcomes, the values of α and β have to be jointly determined. That is, the choice of a smaller β needs to be matched with a smaller α, and conversely, the choice of a larger β needs to be matched with a larger α. Intuitively, setting a small β ensures that informed AI speculators will spend a long time in the exploration mode in which they randomly choose different actions, resulting in extensive experimentation.

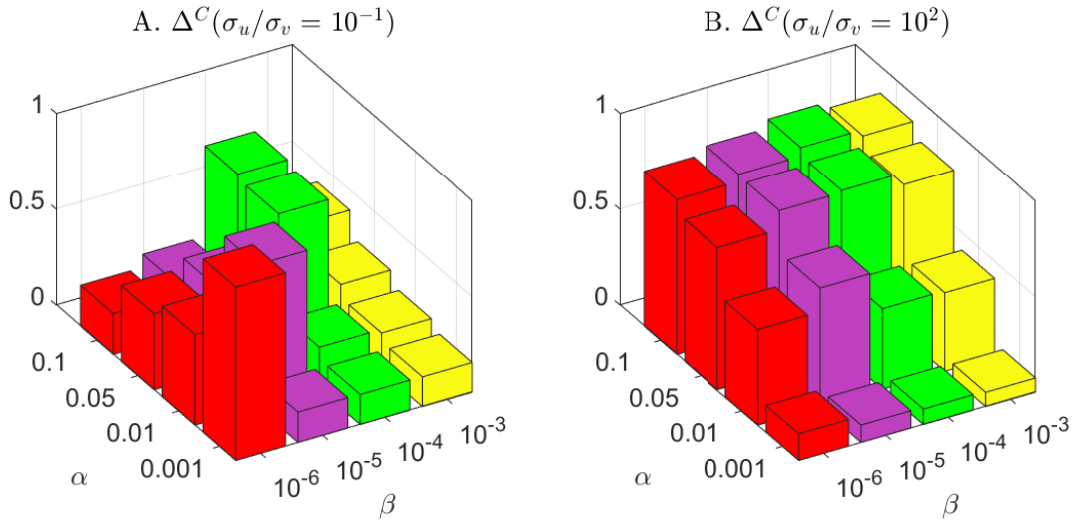


Note: The blue solid line plots the average values of Δ^C, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, and $\mathcal{E}^C$ across $N = 1{,}000$ simulation sessions as the subjective discount rate ρ varies, in the environment with high noise trading risks, i.e., $\sigma_u/\sigma_v = 10^2$. The red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 13: Implications of the subjective discount rate ($\sigma_u/\sigma_v = 10^2$).

Then, setting a small α is necessary to record the value learned in the past whereas setting a large α will disrupt learning as the algorithm would forget what it has learned in the past too rapidly. By contrast, setting a large β means that informed AI speculators only spend a short period of time in the exploration mode. Then, if we still set a small α, the Q-matrices of informed AI speculators would not be updated significantly until the algorithms complete exploration. Thus, when β is large, setting a small α would backfire, making the initial exploration futile. Instead, setting a large α in this case would help informed AI speculators to learn punishment strategies to achieve more collusive outcomes.

In panel B of Figure 14, we plot the average Δ^C in the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^2$) for different values of α and β. Holding β unchanged at each value of $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$, panel B shows that the value of Δ^C declines monotonically as α decreases. This



Note: Panel A plots Δ^C in the environment with low noise trading risks ($\sigma_u/\sigma_v = 10^{-1}$); panel B plots Δ^C in the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^{2}$). The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 14: Implications of hyperparameters α and β on Δ^C.

is because when noise trading risks are large, the supra-competitive profits are attained because informed AI speculators have homogenized learning biases. As discussed in Section 5.2.4, the learning biases due to the failure of the law of large numbers are mitigated when α becomes small.

Taken together, a key feature that distinguishes collusion through price-trigger strategies (panel A of Figure 14) and collusion through homogenized learning biases (panel B of Figure 14) is whether improved learning through setting a sufficiently small α would significantly reduce the supra-competitive profits of informed AI speculators.

7 Coordinated Choice of Q-Learning Algorithms

As shown in panel B of Figure 14, setting a lower forgetting rate α reduces the magnitude of learning biases but it takes longer time and more computation power to train the algorithm. Thus, we can think of α as capturing the "intelligence level" of the algorithm: the algorithm is more advanced if it has a lower α.

In this section, we focus on the environment with high noise trading risks and allow informed AI speculators to choose different values of the hyperparameter α for their Q-learning algorithms. We evaluate the implications for trading profits. Specifically, in Subsection 7.1, we show that the more advanced algorithm will make more profit than the less advanced algorithm. Moreover, given the peer's choice of α, by setting a lower α, the informed AI speculator can increase its own profit. However, importantly, both informed AI speculators can obtain supra-competitive

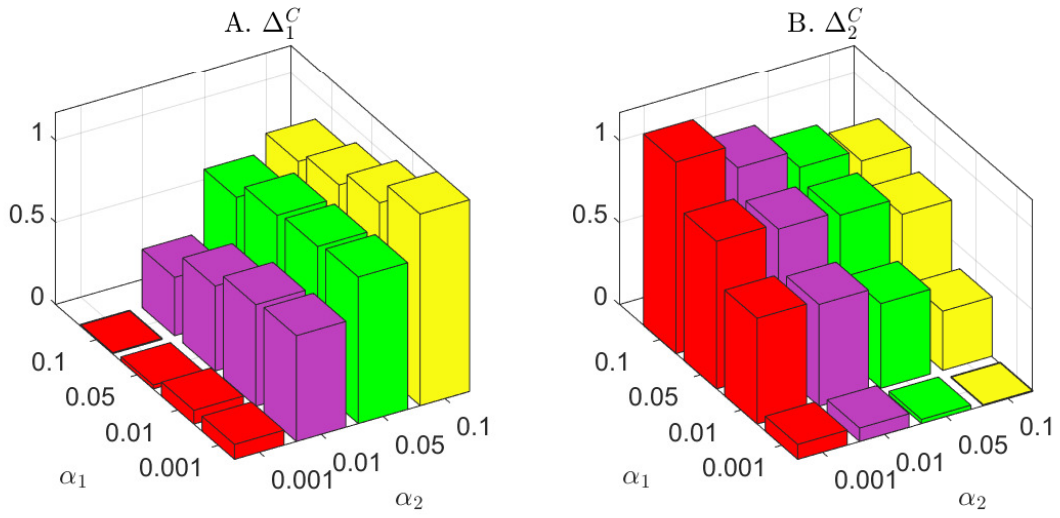
profits if they both adopt less advanced algorithms with similar learning biases. In Subsection 7.2, we extend the Q-learning algorithm to a two-tier Q-learning algorithm in which informed AI speculators learn both the optimal choice of the forgetting rate α and the optimal trading strategies associated with the choice of α. We show that informed AI speculators will learn to adopt high values of α in the stationary equilibrium, and such coordination allows both of them to obtain supra-competitive profits.

7.1 Homogenized Learning Biases

Focusing on the baseline economic environment with two informed AI speculators, as described in Section 4.7 except for setting $\sigma_u/\sigma_v = 10^2$, representing an environment with high noise trading risks. We allow the two informed AI speculators to adopt different values of α, but the same value of β. Intuitively, the informed AI speculator adopting a more advanced Q-learning algorithm (i.e., a lower α) would have smaller learning biases than the one adopting a less advanced algorithm (i.e., a higher α). As discussed in Subsection 5.2.4, learning biases induce informed AI speculators to adopt more conservative trading strategies, i.e., smaller order flows. Therefore, the informed AI speculator with a less advanced algorithm would adopt a more conservative trading strategy than the one with a more advanced algorithm. This essentially enables the informed AI speculator with a more advanced algorithm to take advantage of the other informed AI speculator and obtain more profits than what it would obtain when the other speculator adopts an algorithm with the same α. Conversely, the informed AI speculator with a less advanced algorithm would obtain less profits than what it would obtain when the other speculator adopts an algorithm with the same α.

The results of our simulation experiments are consistent with the above intuition. In Figure 15, we allow each informed AI speculator i to adopt algorithms with different values of α_i, with $\alpha_i = 0.001, 0.01, 0.05$ and 0.1 for $i = 1, 2$. Panels A and B plot the average Δ_1^C and Δ_2^C for informed AI speculators 1 and 2, respectively. It is shown that for any combination of (α_1, α_2), the informed AI speculator with a lower α_i attains a higher average Δ_i^C than the other informed AI speculator. Moreover, holding α_1 unchanged at each value of $\{0.001, 0.01, 0.05, 0.1\}$, as informed AI speculator 2's α_2 decreases, the average Δ_1^C for informed AI speculator 1 decreases and the average Δ_2^C for informed AI speculator 2 increases. Similarly, holding α_2 unchanged at each value of $\{0.001, 0.01, 0.05, 0.1\}$, as informed AI speculator 1's α_1 decreases, the average Δ_2^C for informed AI speculator 2 decreases and the average Δ_1^C for informed AI speculator 1 increases.

Our results indicate that both informed AI speculators can obtain supra-competitive profits if both of them adopt less advanced algorithms with a high value of α. Holding one informed AI speculator's algorithm unchanged, the other speculator could increase its profit by adopting a more advanced algorithm with a lower value of α, and at the same time, the profit of the speculator with a less advanced algorithm would decrease. However, if both informed AI speculators adopt advanced algorithms with a small value of α, the profit for both of them will decrease relative to the equilibrium where both speculators adopt unadvanced algorithms. The results we observe bear similarity to the general equilibrium effects in active management, as characterized by Stambaugh



Note: We allow the two informed AI speculators to adopt Q-learning algorithms with different values of the forgetting rate, denoted by α_1 and α_2 for informed AI speculators 1 and 2, respectively. Panels A and B plot Δ_1^C and Δ_2^C in the environment with high noise trading risks ($\sigma_u/\sigma_v = 10^2$). The other parameters are set according to the baseline economic environment described in Section 4.7.

Figure 15: Profit gain when informed AI speculators adopt algorithms with different values of α.

(2020). According to his model, if all managers lack the ability to select positive-alpha stocks, they can collectively achieve high profits. When a small fraction of managers gains more skill, it results in increased profits for the skilled ones, while the less skilled managers see a decline in their profits. However, if a large proportion of managers becomes more skilled, the profits for all managers start to diminish. This decline is due to a shrinking alpha magnitude, caused by more substantial price corrections in general equilibrium. Interestingly, the total profit of the active management industry typically decreases whenever any of the managers become more skilled. In a recent work, Dugast and Foucault (2024) derive a similar result by showing that improvements in the skills of active asset managers, due to lower information processing costs or the proliferation of new datasets, can reduce their average performance as asset prices become more informative.

7.2 Adaptive Forgetting Rates

In practice, the forgetting rate α is not necessarily fixed throughout the simulation experiments. Instead, many Q-learning algorithms are implemented with adaptive forgetting rates, which are adjusted dynamically in response to the performance of the model. In this subsection, we show that informed AI speculators can learn to coordinately choose high values of α in environments with high noise trading risks, despite the fact that choosing a low forgetting rate unilaterally may boost self-profit. This result implies that an equilibrium with unadvanced algorithms (i.e., high α)

may arise endogenously due to the optimal decisions of informed AI speculators.

Two-Tier Q-Learning Algorithm. Each informed AI speculator i adopts a two-tier Q-learning algorithm. In the lower tier, the informed AI speculator adopts a Q-learning algorithm to learn the lower-tier Q-matrix $\widehat{Q}_{i,t}(s_t, x_{i,t})$ for state $s_t = \{p_{t-1}, v_t\}$ and order flow $x_{i,t}$, given the choice of $\alpha_{i,t}$ in the upper tier. The lower-tier Q-learning algorithm is identical to the algorithm described in Section 4.1, except for the use of a time-varying adaptive forgetting rate $\alpha_{i,t}$. In the upper tier, the informed AI speculator adopts a Q-learning algorithm to learn the upper-tier Q-matrix $\widehat{Q}^u_{i,t}(s^u_{i,t}, \alpha_{i,t})$ for state $s^u_{i,t}$ and action $\alpha_{i,t}$.

For any given choice of $\alpha_{i,t}$ in the upper tier, it is necessary to ensure that the lower tier Q-learning algorithm is run for a sufficiently long period of time, so that the profits corresponding to the choice of $\alpha_{i,t}$ fully stablize. This means that compared with the choice of $x_{i,t}$ in the lower tier, the choice of $\alpha_{i,t}$ in the upper tier has to be experimented at a much lower frequency. Therefore, we specify that each informed AI speculator i adjusts its upper tier's action $\alpha_{i,t}$ only after the lower tier finishes a training epoch that lasts for a total of T periods, with T being a large integer.

Specifically, let $\tau = 1, 2, ...$ denote all training epochs of the lower-tier Q-learning algorithm. The training epoch τ represents the period from $(\tau-1)T+1$ to τT. Within each training epoch τ, each informed AI speculator i's upper-tier Q-matrix $\widehat{Q}^u_{i,t}(s^u_{i,t}, \alpha_{i,t})$ or action $\alpha_{i,t}$ stay unchanged from period $(\tau-1)T+1$ to period $\tau T-1$; the values of $\widehat{Q}^u_{i,t}(s^u_{i,t}, \alpha_{i,t})$ and action $\alpha_{i,t}$ are updated only at the end of the training epoch, occurring at $t = \tau T$. Therefore, without loss of generality, we only need to specify the recursive learning equation of the upper-tier Q-learning algorithm at the end of each period, $t = \tau T$, as follows:

$$\widehat{Q}^u_{i,(\tau+1)T}(s^u_{i,\tau T}, \alpha_{i,\tau T}) = (1-\alpha^u)\widehat{Q}^u_{i,\tau T}(s^u_{i,\tau T}, \alpha_{i,\tau T}) + \alpha^u \left[\pi^u_{i,\tau T} + \rho^u \max_{\alpha' \in \mathcal{A}} \widehat{Q}^u_{i,\tau T}(s^u_{i,(\tau+1)T}, \alpha')\right], \quad (7.1)$$

for $\tau = 1, 2, ...$ In equation (7.1), $\pi^u_{i,\tau}$ is the reward in the training epoch τ, given by $\pi^u_{i,\tau T} = \frac{1}{T}\sum_{t=(\tau-1)T+1}^{\tau T}(v_t - p_t)x_{i,t}$, which is the average trading profit over the last T periods, from period $(\tau-1)T+1$ to period τT. The parameters α^u and ρ^u are the forgetting rate and the subjective discount rate for the upper tier Q-learning algorithm. For tractability, we choose the state variable $s^u_{i,\tau T} = \{\pi^u_{i,(\tau-1)T}\}$, which is the reward in the previous training epoch. The choice of $\alpha_{i,\tau T}$ is chosen as follows:

$$\alpha_{i,\tau T} = \begin{cases} \text{argmax}_{\alpha' \in \mathcal{A}}\, \widehat{Q}^h_{i,\tau T}(s^u_{i,\tau T}, \alpha'), & \text{with prob. } 1-\varepsilon^u_\tau, \quad \text{(exploitation)} \\ \widetilde{\alpha} \sim \text{uniform distribution on } \mathcal{A}, & \text{with prob. } \varepsilon^u_\tau. \quad \text{(exploration)} \end{cases} \quad (7.2)$$

The exploration rate is specified as $\varepsilon_\tau = e^{-\beta^u \tau}$, where β^u is a parameter governing the decaying speed of exploration rates across training epochs.

Simulation Results. The two-tier Q-learning algorithm takes a substantially longer time to converge because there are experimentations on both $\alpha_{i,t}$ and $x_{i,t}$. We consider the following

parameter values: $\alpha^u = 0.1$, $\beta^u = 10^{-4}$, and $\rho^u = 0.95$. Each training epoch has a total of $T = 10,000,000$ periods. The convergence criterion requires the decisions of $\alpha_{i,t}$ to stay unchanged for $100,000$ consecutive training epochs. For tractability, we choose three grids for the choice of $\alpha_{i,t}$, with $\mathcal{A} = \{0.001, 0.01, 0.1\}$. The parameters and grids for the lower-tier Q-learning algorithm are similar to those described in Section 4. In particular, there are two informed AI speculators. We separately conduct $N = 1,000$ independent simulations for the environments with high and low noise trading risks.

Our primary focus is on the environment with high noise trading risks (i.e., $\sigma_u/\sigma_v = 10^2$). As shown in panel B of Figure 15, the two informed AI speculators encounter a problem resembling the prisoner's dilemma. Specifically, given informed AI speculator i's choice of α_i, informed AI speculator j can gain by adopting the smallest $\alpha_j = 0.001$. However, both informed AI speculators would not make much profit if they reach the Nash equilibrium of $(\alpha_1, \alpha_2) = (0.001, 0.001)$. Instead, both of them would attain supra-competitive profits by coordinately reaching the equilibrium with $(\alpha_1, \alpha_2) = (0.01, 0.01)$ or $(\alpha_1, \alpha_2) = (0.1, 0.1)$, that is, by adopting unadvanced algorithms to trade. In theory, these two equilibria with high values of α can only be sustained in a repeated game. Turning to our simulation experiments with informed AI speculators adopting the two-tier Q-learning algorithms, we find that across the $N = 1,000$ simulations sessions, 272 sessions converge to the equilibrium with $(\alpha_1, \alpha_2) = (0.1, 0.1)$, and 710 sessions converge to the equilibrium with $(\alpha_1, \alpha_2) = (0.01, 0.01)$. There does not exist a single simulation session that converges to the equilibrium with $(\alpha_1, \alpha_2) = (0.001, 0.001)$, even though this is the unique Nash equilibrium in a one-shot game. Our results indicate that in the environment with high noise trading risks, the two informed AI speculators are able to learn to adopt less advanced algorithms, which have high values of α, in the stationary equilibrium. This coordination allows both AI speculators to obtain supra-competitive profits.

For comparisons, we also conduct simulation experiments in the environment with low noise trading risks (i.e., $\sigma_u/\sigma_v = 10^{-1}$). As shown in panel A of Figure 14, the optimal outcome is achieved if the two informed AI speculators choose to play the equilibrium with $(\alpha_1, \alpha_2) = (0.01, 0.01)$, given that $\beta = 10^{-5}$. We find that across the $N = 1,000$ simulations sessions, 957 sessions converge to this equilibrium. This suggests that our simple two-tier Q-learning algorithm enables the two informed AI speculators to learn to play the optimal equilibrium. The algorithm's excellent performance is due to the fact that in this environment, informed AI speculators do not face a prisoner's dilemma problem. That is, the equilibrium with $(\alpha_1, \alpha_2) = (0.01, 0.01)$, which yields the highest trading profits for both informed AI speculators, is also the Nash equilibrium of a one-shot game. In other words, choosing the forgetting rate $\alpha_i = 0.01$ maximizes informed AI speculator i's trading profits regardless of the forgetting rate that the other informed AI speculator chooses.

References

Abreu, Dilip, David Pearce, and Ennio Stacchetti. 1986. "Optimal cartel equilibria with imperfect monitoring." *Journal of Economic Theory*, 39(1): 251–269.

Abreu, Dilip, Paul Milgrom, and David Pearce. 1991. "Information and Timing in Repeated Partnerships." *Econometrica*, 59(6): 1713–1733.

Asker, John, Chaim Fershtman, and Ariel Pakes. 2022. "Artificial Intelligence, Algorithm Design, and Pricing." *AEA Papers and Proceedings*, 112: 452–56.

Assad, Stephanie, Robert Clark, Daniel Ershov, and Lei Xu. 2023. "Algorithmic Pricing and Competition: Empirical Evidence from the German Retail Gasoline Market." *Journal of Political Economy*, Forthcoming.

Bagattini, Giulio, Zeno Benetti, and Claudia Guagliano. 2023. "Artificial intelligence in EU securities markets." *ESMA50-164-6247*. European Securities and Markets Authority.

Bellman, Richard Ernest. 1954. *The Theory of Dynamic Programming*. Santa Monica, CA:RAND Corporation.

Bommasani, Rishi, Kathleen Creel, Ananya Kumar, Dan Jurafsky, and Percy Liang. 2022. "Picking on the Same Person: Does Algorithmic Monoculture lead to Outcome Homogenization?"

Calvano, Emilio, Giacomo Calzolari, Vincenzo Denicoló, and Sergio Pastorello. 2020. "Artificial Intelligence, Algorithmic Pricing, and Collusion." *American Economic Review*, 110(10): 3267–3297.

Cho, In-Koo, and Thomas J. Sargent. 2008. "Self-confirming Equilibria." 407–408. Palgrave Macmillan.

Colliard, Jean-Edouard, Thierry Foucault, and Stefano Lovo. 2022. "Algorithmic Pricing and Liquidity in Securities Markets." HEC Paris Working Papers.

Dou, Winston Wei, Wei Wang, and Wenyu Wang. 2023. "The Cost of Intermediary Market Power for Distressed Borrowers." The Wharton School at University of Pennsylvania Working Papers.

Dou, Winston Wei, Yan Ji, and Wei Wu. 2021*a*. "Competition, Profitability, and Discount Rates." *Journal of Financial Economcis*, 140(2): 582–620.

Dou, Winston Wei, Yan Ji, and Wei Wu. 2021*b*. "The Oligopoly Lucas Tree." *The Review of Financial Studies*, 35(8): 3867–3921.

Dugast, Jérôme, and Thierry Foucault. 2024. "Equilibrium Data Mining and Data Abundance." *Journal of Finance*, forthcoming.

Fudenberg, Drew, and David Levine. 1993. "Self-Confirming Equilibrium." *Econometrica*, 61(3): 523–45.

Fudenberg, Drew, and David M. Kreps. 1988. "A theory of learning, experimentation, and equilibrium in games." Working Papers.

Fudenberg, Drew, and David M. Kreps. 1995. "Learning in extensive-form games I. Self-confirming equilibria." *Games and Economic Behavior*, 8(1): 20–55.

Fudenberg, Drew, and Eric Maskin. 1986. "The Folk theorem in repeated games with discounting or with incomplete information." *Econometrica*, 54(3): 533–54.

Goldstein, Itay, Chester S Spatt, and Mao Ye. 2021. "Big Data in Finance." *The Review of Financial Studies*, 34(7): 3213–3225.

Goldstein, Itay, Emre Ozdenoren, and Kathy Yuan. 2013. "Trading frenzies and their impact on real investment." *Journal of Financial Economics*, 109(2): 566–582.

Graham, Benjamin. 1973. *The Intelligent Investor.* . 4 ed., Publisher: Harper & Row, New York, NY.

Green, Edward J, and Robert H Porter. 1984. "Noncooperative Collusion under Imperfect Price Information." *Econometrica*, 52(1): 87–100.

Greenwood, Robin, and Dimitri Vayanos. 2014. "Bond Supply and Excess Bond Returns." *The Review of Financial Studies*, 27(3): 663–713.

Greenwood, Robin, Samuel Hanson, Jeremy C Stein, and Adi Sunderam. 2023. "A Quantity-Driven Theory of Term Premia and Exchange Rates*." *The Quarterly Journal of Economics*, qjad024.

Harrington, Joseph E. 2018. "Developing Competition Law for Collusion by Autonomous Artificial Agents." *Journal of Competition Law & Economics*, 14(3): 331–363.

Hellwig, Christian, Arijit Mukherji, and Aleh Tsyvinski. 2006. "Self-Fulfilling Currency Crises: The Role of Interest Rates." *The American Economic Review*, 96(5): 1769–1787.

Johnson, Justin, and D. Daniel Sokol. 2021. "Understanding AI Collusion and Compliance." *The Cambridge Handbook of Compliance*, , ed. Benjamin van Rooij and D. DanielEditors Sokol *Cambridge Law Handbooks*, 881–894. Cambridge University Press.

Johnson, Justin Pappas, Andrew Rhodes, and Matthijs Wildenbeest. 2023. "Platform Design when Sellers Use Pricing Algorithms." *Econometrica*, Forthcoming.

Klein, Timo. 2021. "Autonomous algorithmic collusion: Q-learning under sequential pricing." *The RAND Journal of Economics*, 52(3): 538–558.

Kyle, Albert S. 1985. "Continuous Auctions and Insider Trading." *Econometrica*, 53(6): 1315–1335.

Kyle, Albert S. 1989. "Informed Speculation with Imperfect Competition." *The Review of Economic Studies*, 56(3): 317–355.

Kyle, Albert S., and Wei Xiong. 2001. "Contagion as a Wealth Effect." *The Journal of Finance*, 56(4): 1401–1440.

Ljungqvist, Lars, and Thomas J. Sargent. 2012. *Recursive Macroeconomic Theory, Third Edition.* Vol. 1 of *MIT Press Books*. 3 ed., The MIT Press.

Long, J. Bradford De, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy*, 98(4): 703–738.

Mildenstein, Eckart, and Harold Schleef. 1983. "The Optimal Pricing Policy of a Monopolistic Marketmaker in the Equity Market." *The Journal of Finance*, 38(1): 218–231.

Opp, Marcus M., Christine A. Parlour, and Johan Walden. 2014. "Markup cycles, dynamic misallocation, and amplification." *Journal of Economic Theory*, 154: 126–161.

Rotemberg, Julio J, and Garth Saloner. 1986. "A supergame-theoretic model of price wars during booms." *American Economic Review*, 76(3): 390–407.

Sandholm, Tuomas W., and Robert H. Crites. 1996. "On multiagent Q-learning in a semi-competitive domain." 191–205. Berlin, Heidelberg:Springer Berlin Heidelberg.

Sannikov, Yuliy, and Andrzej Skrzypacz. 2007. "Impossibility of Collusion under Imperfect Monitoring with Flexible Production." *American Economic Review*, 97(5): 1794–1823.

SEC. 2023. "Conflicts of Interest Associated with the Use of Predictive Data Analytics by BrokerDealers and Investment Advisers." *Release Nos. 34-97990*. U.S. Securities and Exchange Commission.

Stambaugh, Robert F. 2020. "Skill and Profit in Active Management." National Bureau of Economic Research, Inc NBER Working Papers 26027.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction.* . Second ed., The MIT Press.

Tesauro, Gerald, and Jeffrey O. Kephart. 2002. "Pricing in Agent Economies Using Multi-Agent Q-Learning." *Autonomous Agents and Multi-Agent Systems*, 5(3): 289–304.

Vayanos, Dimitri, and Jean-Luc Vila. 2021. "A Preferred-Habitat Model of the Term Structure of Interest Rates." *Econometrica*, 89(1): 77–112.

Waltman, Ludo, and Uzay Kaymak. 2008. "Q-learning agents in a Cournot oligopoly model." *Journal of Economic Dynamics and Control*, 32(10): 3275–3293.

Watkins, Christopher J. C. H., and Peter Dayan. 1992. "Q-learning." *Machine Learning*, 8(3): 279–292.

Appendix

A Proof of Lemma 1

The preferred-habitat investor solves the following portfolio optimization problem for a given p_t:

$$\max_z \mathbb{E}\left[-e^{-\eta(v_t-p_t)z}/\eta\right]. \tag{A.1}$$

Because $v_t - p_t$ is distributed as $N(\overline{v}-p_t, \sigma_v^2)$, the first-order condition with respect to z is

$$0 = \left[(\overline{v}-p_t) - \eta z\sigma_v^2\right] e^{-\eta z(\overline{v}-p_t)+(\eta z)^2\sigma_v^2/2}. \tag{A.2}$$

Thus, the optimal holding, z, is characterized as

$$z = -\frac{1}{\eta\sigma_v^2}(p_t - \overline{v}). \tag{A.3}$$

B Proof of Proposition 3.3

Given that $s_t = 0$, let $J^C(\chi_i)$ denote each informed speculator i's expected present value of future profits, when investor i chooses $x_{i,t} = \chi_i(v_t - \overline{v})$ and all other $I-1$ informed investors choose $x^C(v_t) = \chi^C(v_t - \overline{v})$. That is,

$$\begin{aligned} J^C(\chi_i) =& \mathbb{E}\left[\left(v_t - p^C(y_t)\right)\chi_i(v_t-\overline{v})\right] \\ &+ \rho J^C(\chi_i)\mathbb{P}\left\{\text{Price trigger is not violated in period } t \middle| \chi_i, \chi^C\right\} \\ &+ \mathbb{E}\left[\sum_{\tau=1}^{T-1}\rho^\tau \pi^N(v_{t+\tau}) + \rho^T J^C(\chi_i)\right]\mathbb{P}\left\{\text{Price trigger is violated in period } t \middle| \chi_i, \chi^C\right\}, \end{aligned} \tag{B.1}$$

where $p^C(\cdot)$ is the pricing function of market makers in the collusive Nash equilibrium and

$$p^C(y_t) = \overline{v} + \lambda^C y_t, \quad \text{with } \lambda^C = \frac{\theta\gamma^C + \xi}{\theta + \xi^2} \text{ and } \gamma^C = \frac{I\chi^C}{(I\chi^C)^2 + (\sigma_u/\sigma_v)^2}, \tag{B.2}$$

$$y_t = \chi_i(v_t - \overline{v}) + (I-1)x^C(v_t) + u_t. \tag{B.3}$$

The probability of price trigger violation is

$$\begin{aligned} &\mathbb{P}\left\{\text{Price trigger is not violated in period } t \middle| \chi_i, \chi^C\right\} \\ =&\mathbb{E}\left[\mathbb{P}\left(p_t \le q(v_t)|v_t\right)\mathbf{1}\{v_t > \overline{v}\}\right] + \mathbb{E}\left[\mathbb{P}\left(p_t \ge q(v_t)|v_t\right)\mathbf{1}\{v_t < \overline{v}\}\right] \\ =&\mathbb{E}\left[\Phi(\sigma_u^{-1}(\chi^C - \chi_i)(v_t-\overline{v}) + \omega)\mathbf{1}\{v_t > \overline{v}\}\right] + \mathbb{E}\left[\Phi(\sigma_u^{-1}(\chi_i - \chi^C)(v_t-\overline{v}) + \omega)\mathbf{1}\{v_t < \overline{v}\}\right], \end{aligned}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution.

Evaluating equality (B.1) at $\chi_i = \chi^C$ leads to

$$\begin{aligned} J^C(\chi^C) = &\left(1 - \lambda^C I \chi^C\right) \chi^C \sigma_v^2 \\ &+ \rho J^C(\chi^C) \Phi(\omega) \\ &+ \frac{\rho - \rho^T}{1-\rho} \left[1 - \Phi(\omega)\right] \mathbb{E}\left[\pi^N(v)\right] + \rho^T J^C(\chi^C) \left[1 - \Phi(\omega)\right]. \end{aligned} \tag{B.4}$$

Thus, we can obtain that

$$J^C(\chi^C) = \frac{\left(1 - \lambda^C I \chi^C\right) \chi^C \sigma_v^2 + \frac{\rho - \rho^T}{1-\rho} \left[1 - \Phi(\omega)\right] \mathbb{E}\left[\pi^N(v)\right]}{1 - \rho\Phi(\omega) - \rho^T \left[1 - \Phi(\omega)\right]}. \tag{B.5}$$

The first-order derivative of the both sides of (B.1) with respect to χ_i, evaluated at $\chi_i = \chi^C$, is

$$\begin{aligned} \nabla J^C(\chi^C) = &\left[1 - \lambda^C (I+1) \chi^C\right] \sigma_v^2 \\ &+ \rho \left[\nabla J^C(\chi^C)\right] \Phi(\omega) - \rho J^C(\chi^C) \frac{1}{\sigma_u} \phi(\omega) \mathbb{E}\left[|v - \overline{v}|\right] \\ &+ \frac{\rho - \rho^T}{1-\rho} \frac{1}{\sigma_u} \phi(\omega) \mathbb{E}\left[|v - \overline{v}|\right] \mathbb{E}\left[\pi^N(v)\right] \\ &+ \rho^T \left[\nabla J^C(\chi^C)\right] \left[1 - \Phi(\omega)\right] + \rho^T J^C(\chi^C) \frac{1}{\sigma_u} \phi(\omega) \mathbb{E}\left[|v - \overline{v}|\right], \end{aligned} \tag{B.6}$$

where $\phi(\cdot)$ is the probability density function of the standard normal distribution.

Because $v - \overline{v}$ is distributed as $N(0, \sigma_v^2)$, it follows that $\mathbb{E}\left[|v - \overline{v}|\right] = \sigma_v \sqrt{\frac{2}{\pi}}$. Plugging it into (B.6), we obtain that

$$\begin{aligned} \nabla J^C(\chi^C) = &\left[1 - \lambda^C (I+1) \chi^C\right] \sigma_v^2 \\ &+ \rho \left[\nabla J^C(\chi^C)\right] \Phi(\omega) - \rho J^C(\chi^C) \frac{\sigma_v}{\sigma_u} \phi(\omega) \sqrt{\frac{2}{\pi}} \\ &+ \frac{\rho - \rho^T}{1-\rho} \mathbb{E}\left[\pi^N(v)\right] \frac{\sigma_v}{\sigma_u} \phi(\omega) \sqrt{\frac{2}{\pi}} \\ &+ \rho^T \left[\nabla J^C(\chi^C)\right] \left[1 - \Phi(\omega)\right] + \rho^T J^C(\chi^C) \frac{\sigma_v}{\sigma_u} \phi(\omega) \sqrt{\frac{2}{\pi}}. \end{aligned} \tag{B.7}$$

The policy variable χ^C constitutes a collusive Nash equilibrium if speculator i has no incentive to deviate by setting $\chi_i \neq \chi^C$. The first-order condition with respect to χ_i, characterized by

$\nabla J^C(\chi^C) = 0$, leads to

$$\begin{aligned} 0 = &\left[1 - \lambda^C(I+1)\chi^C\right]\sigma_v^2 \\ &- \rho J^C(\chi^C)\frac{\sigma_v}{\sigma_u}\phi(\omega)\sqrt{\frac{2}{\pi}} \\ &+ \frac{\rho - \rho^T}{1-\rho}\mathbb{E}\left[\pi^N(v)\right]\frac{\sigma_v}{\sigma_u}\phi(\omega)\sqrt{\frac{2}{\pi}} \\ &+ \rho^T J^C(\chi^C)\frac{\sigma_v}{\sigma_u}\phi(\omega)\sqrt{\frac{2}{\pi}}. \end{aligned} \tag{B.8}$$

According to (B.2), as $\theta \to \infty$ or as $\xi \to 0$, $\lambda^C \to \gamma^C$, that is, the market approaches to the environment of Kyle (1985). In this case, the demand of the preferred-habitat investor is irrelevant. Because the system is continuous, it is sufficient to show that there is no solution $\chi^C \in [\chi^M, \chi^N)$ in the environment of Kyle (1985), where $\chi^N = \frac{1}{\sqrt{I}}\frac{\sigma_u}{\sigma_v}$ and $\chi^M = \frac{1}{I}\frac{\sigma_u}{\sigma_v}$ as a result of $\lambda^C = \gamma^C$. Let $\chi^C = \widehat{\chi}^C\frac{\sigma_u}{\sigma_v}$. Then, we show that there is no solution $\widehat{\chi}^C \in [\widehat{\chi}^M, \widehat{\chi}^N)$, with $\widehat{\chi}^M = \frac{1}{I}$ and $\widehat{\chi}^N) = \frac{1}{\sqrt{I}}$. In the Kyle case, $\mathbb{E}\left[\pi^N(v)\right] = \frac{\sigma_u\sigma_v}{(I+1)\sqrt{I}}$. Therefore, equations (B.5) and (B.8) can be rewritten, respectively, as follows:

$$J^C(\chi^C) = \frac{\left(1 - \gamma^C I\chi^C\right)\chi^C\sigma_v^2 + \frac{\rho - \rho^T}{1-\rho}\left[1 - \Phi(\omega)\right]\frac{\sigma_v\sigma_u}{(I+1)\sqrt{I}}}{1 - \rho\Phi(\omega) - \rho^T\left[1 - \Phi(\omega)\right]}. \tag{B.9}$$

and

$$0 = \left[1 - \gamma^C(I+1)\chi^C\right]\sigma_v^2 - \left[\rho J^C(\chi^C) - \frac{\rho - \rho^T}{1-\rho}\frac{\sigma_v\sigma_u}{(I+1)\sqrt{I}} - \rho^T J^C(\chi^C)\right]\frac{\sigma_v}{\sigma_u}\phi(\omega)\sqrt{\frac{2}{\pi}}. \tag{B.10}$$

Therefore, $\widehat{\chi}^C$ is the root of the following quadratic equation:

$$\begin{aligned} 0 = &\left[1 - I(\widehat{\chi}^C)^2\right]\frac{1}{\rho - \rho^T} \\ &- \left\{1 - \rho + (\rho - \rho^T)[1 - \Phi(\omega)]\right\}^{-1}\left\{\widehat{\chi}^C - \frac{1}{(I+1)\sqrt{I}}\left[1 + (I\widehat{\chi}^C)^2\right]\right\}\phi(\omega)\sqrt{\frac{2}{\pi}}, \end{aligned}$$

which can be simplified as

$$0 = 1 - I(\widehat{\chi}^C)^2 - \vartheta\left\{\widehat{\chi}^C - \frac{1}{(I+1)\sqrt{I}}\left[1 + (I\widehat{\chi}^C)^2\right]\right\}, \tag{B.11}$$

where

$$\vartheta = \frac{\phi(\omega)}{\frac{1-\rho}{\rho-\rho^T} + 1 - \Phi(\omega)}\sqrt{\frac{2}{\pi}}. \tag{B.12}$$

Solving the above problem, we obtain

$$\widehat{\chi}^C = \frac{\vartheta \pm \left| -2\sqrt{I} + \frac{I-1}{I+1}\vartheta \right|}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}}.$$

There are three cases.

Case 1: if $-2\sqrt{I} + \frac{I-1}{I+1}\vartheta \leq 0$ and $-2I + 2\vartheta\frac{I\sqrt{I}}{I+1} < 0$, the larger root is

$$\widehat{\chi}^C = \frac{\vartheta + \left(-2\sqrt{I} + \frac{I-1}{I+1}\vartheta\right)}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}} = \frac{1}{\sqrt{I}} = \widehat{\chi}^N,$$

and the other root, which is smaller, is given by

$$\widehat{\chi}^C = \frac{\vartheta - \left(-2\sqrt{I} + \frac{I-1}{I+1}\vartheta\right)}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}} = \frac{\sqrt{I} + \frac{\vartheta}{I+1}}{-I + \vartheta\frac{I\sqrt{I}}{I+1}},$$

which is negative. Thus, there does not exist a solution $\widehat{\chi}^C$ that lies in $[\frac{1}{I}, \frac{1}{\sqrt{I}})$, meaning that the collusive equilibrium does not exist.

Case 2: if $-2\sqrt{I} + \frac{I-1}{I+1}\vartheta \leq 0$ and $-2I + 2\vartheta\frac{I\sqrt{I}}{I+1} > 0$, the smaller root is

$$\widehat{\chi}^C = \frac{\vartheta - \left(-2\sqrt{I} + \frac{I-1}{I+1}\vartheta\right)}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}} = \frac{1}{\sqrt{I}} = \widehat{\chi}^N,$$

and the other root, which is larger, should be greater than $\widehat{\chi}^N$. Thus, there does not exist a solution $\widehat{\chi}^C$ that lies in $[\frac{1}{I}, \frac{1}{\sqrt{I}})$, meaning that the collusive equilibrium does not exist.

Case 3: if $-2\sqrt{I} + \frac{I-1}{I+1}\vartheta > 0$. In this case, we can prove that

$$-2I + 2\vartheta\frac{I\sqrt{I}}{I+1} = \sqrt{I}\left[-2\sqrt{I} + 2\vartheta\frac{I}{I+1}\right] > \sqrt{I}\left[-\frac{I-1}{I+1}\vartheta + 2\vartheta\frac{I}{I+1}\right] > 0.$$

Thus, the larger root is

$$\widehat{\chi}^C = \frac{\vartheta + \left(-2\sqrt{I} + \frac{I-1}{I+1}\vartheta\right)}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}} = \frac{1}{\sqrt{I}} = \widehat{\chi}^N.$$

The smaller root is

$$\widehat{\chi}^C = \frac{\vartheta - \left(-2\sqrt{I} + \frac{I-1}{I+1}\vartheta\right)}{-2I + 2\vartheta\frac{I\sqrt{I}}{I+1}} = \frac{\sqrt{I} + \frac{\vartheta}{I+1}}{-I + \vartheta\frac{I\sqrt{I}}{I+1}}.$$

For $\widehat{\chi}^C$ to lie in $[\frac{1}{I}, \frac{1}{\sqrt{I}})$, we need $\widehat{\chi}^C \geq 1/I$, which implies

$$\frac{1}{I+1}\frac{\sqrt{I}-1}{\sqrt{I}+1}\vartheta \leq 1,$$

Thus, if $\vartheta \in \left(2\sqrt{I}\frac{I+1}{I-1}, (I+1)\frac{\sqrt{I}+1}{\sqrt{I}-1}\right]$, there exists a collusive equilibrium. To rule out this, we either need $\vartheta \leq 2\sqrt{I}\frac{I+1}{I-1}$ (to rule out case 3) or $\vartheta > (I+1)\frac{\sqrt{I}+1}{\sqrt{I}-1}$ (to ensure the smaller root $\widehat{\chi}^C < 1/I$ in case 3).

C Proof of Proposition 3.4

As $\theta \to 0$ or as $\xi \to \infty$, $\lambda^C \to 1/\xi$, that is, the market approaches to the environment where prices are primarily determined by market clearing conditions. In this case, the demand of the preferred-habitat investor plays an important role. In particular, when $\theta = 0$ (or $\xi \to \infty$), the market maker's pricing rule is $\lambda^C = 1/\xi$.

Because the system is continuous, it is sufficient to show that there is a solution $\chi^C \in [\chi^M, \chi^N)$ in the environment with $\lambda^C = 1/\xi$, where $\chi^N = \frac{\xi}{I+1}$, $\chi^M = \frac{\xi}{2I}$, and $\mathbb{E}\left[\pi^N(v)\right] = \frac{\xi\sigma_v^2}{(I+1)^2}$. In this environment, equations (B.5) and (B.8) can be rewritten, respectively, as follows:

$$J^C(\chi^C) = \frac{\left(1-\xi^{-1}I\chi^C\right)\chi^C\sigma_v^2 + \frac{\rho-\rho^T}{1-\rho}\left[1-\Phi(\omega)\right]\frac{\xi\sigma_v^2}{(I+1)^2}}{1-\rho\Phi(\omega)-\rho^T\left[1-\Phi(\omega)\right]} \tag{C.1}$$

and

$$0 = \left[1-\xi^{-1}(I+1)\chi^C\right]\sigma_v^2 - \left[\rho J^C(\chi^C) - \frac{\rho-\rho^T}{1-\rho}\frac{\xi\sigma_v^2}{(I+1)^2} - \rho^T J^C(\chi^C)\right]\frac{\sigma_v}{\sigma_u}\phi(\omega)\sqrt{\frac{2}{\pi}}. \tag{C.2}$$

Therefore, χ^C is the root of the following quadratic equation:

$$0 = 1-\xi^{-1}(I+1)\chi^C - K\left[\left(1-\xi^{-1}I\chi^C\right)\chi^C - \frac{\xi}{(I+1)^2}\right],$$

where

$$K = \frac{\sigma_v}{\sigma_u}\vartheta = \frac{\sigma_v}{\sigma_u}\frac{\phi(\omega)}{\frac{1-\rho}{\rho-\rho^T}+1-\Phi(\omega)}\sqrt{\frac{2}{\pi}}. \tag{C.3}$$

Solving the above problem, we obtain

$$\widehat{\chi}^C = \frac{K+\frac{I+1}{\xi} \pm \left|\frac{K(I-1)}{I+1} - \frac{I+1}{\xi}\right|}{\frac{2KI}{\xi}}.$$

There are two cases.

Case 1: if $\frac{K(I-1)}{I+1} - \frac{I+1}{\xi} < 0$, then the smaller root is

$$\chi^C = \frac{K+\frac{I+1}{\xi}+\left(\frac{K(I-1)}{I+1}-\frac{I+1}{\xi}\right)}{\frac{2KI}{\xi}} = \chi^N.$$

The larger root must be greater than χ^N. Thus, there does not exist a collusive equilibrium. To rule out this case, we need $K\xi > \frac{(I+1)^2}{I-1}$, which can be achieved by choosing a sufficiently small σ_u/σ_v according to (C.3).

Case 2: if $\frac{K(I-1)}{I+1} - \frac{I+1}{\xi} > 0$, i.e., $K\xi > \frac{(I+1)^2}{I-1}$, then the larger root is

$$\chi^C = \frac{K + \frac{I+1}{\xi} + \left(\frac{K(I-1)}{I+1} - \frac{I+1}{\xi}\right)}{\frac{2KI}{\xi}} = \chi^N.$$

The smaller root is

$$\chi^C = \frac{K + \frac{I+1}{\xi} - \left(\frac{K(I-1)}{I+1} - \frac{I+1}{\xi}\right)}{\frac{2KI}{\xi}} = \frac{\frac{K\xi}{I+1} + I + 1}{KI}. \tag{C.4}$$

To have a valid collusive equilibrium, we need

$$\frac{\frac{K\xi}{I+1} + I + 1}{KI} \geq \frac{\xi}{2I},$$

which implies

$$K\xi \leq \frac{2(I+1)^2}{I-1},$$

meaning that σ_u/σ_v cannot be too small.

In summary, for given parameters T, ρ, ω, and I, we have a range of σ_u/σ_v to sustain the collusive equilibrium. That is, σ_u/σ_v has to be sufficiently small (in order to rule out case 1) but cannot be too small (to ensure the existence of the collusive equilibrium in case 2). That is, σ_u/σ_v should be determined such that

$$K\xi \in \left(\frac{(I+1)^2}{I-1}, \frac{2(I+1)^2}{I-1}\right]. \tag{C.5}$$

D Proof of Proposition 3.6

We prove the proposition for the environment with $\theta = 0$, so the results derived in Appendix C can be directly used. More general environments with $\theta > 0$ can be proved similarly with more complex derivations.

Without loss of generality, we restrict the analysis to the parameter choices such that the collusive equilibrium exists, meaning that condition (C.5) is satisfied. Thus, χ^C is given by (C.4).

Proof for the profit ratio Δ^C**.** The expected profit associated with χ^C is

$$\pi^C = (1 - \xi^{-1} I \chi^C)\chi^C \sigma_v^2.$$

Thus, $\pi^C - \pi^N$ is as follows:

$$\pi^C - \pi^N = \left(\frac{I}{I+1} - \frac{I+1}{K\xi}\right)\left(\frac{\xi}{I(I+1)} + \frac{I+1}{KI}\right)\sigma_v^2 - \frac{\xi\sigma_v^2}{(I+1)^2}.$$

The expected profit associated with χ^M is

$$\pi^M = (1 - \xi^{-1} I\chi^M)\chi^M\sigma_v^2.$$

Thus, $\pi^M - \pi^N$ is as follows:

$$\pi^M - \pi^N = \frac{\xi\sigma_v^2}{4I} - \frac{\xi\sigma_v^2}{(I+1)^2} = \xi\frac{(I-1)^2}{4I(I+1)^2}\sigma_v^2.$$

Thus, Δ^C is

$$\Delta^C = \frac{4}{(I-1)^2}\left(I - \frac{(I+1)^2}{K\xi}\right)\left(1 + \frac{(I+1)^2}{K\xi}\right) - \frac{4I}{(I-1)^2}.$$

Because $K\xi \leq \frac{2(I+1)^2}{I-1}$, Δ^C is increasing in ξ and K. Moreover, $K = \frac{\phi(\omega)}{\frac{1-\rho}{\rho-\rho^T}+1-\Phi(\omega)}\frac{\sigma_v}{\sigma_u}\sqrt{\frac{2}{\pi}}$ is increasing in ρ and decreasing in σ_u/σ_v. Thus, Δ^C is increasing in ρ and decreasing in σ_u/σ_v.

To show K is increasing in ρ, it is sufficient to prove $\frac{1-\rho}{\rho-\rho^T}$ is decreasing in ρ, which is equivalent to show that $f(\rho) = \log(1-\rho) - \log(\rho - \rho^T)$ is decreasing in ρ. The first derivative is

$$f(\rho)' = -\frac{1}{1-\rho} - \frac{1 - T\rho^{T-1}}{\rho - \rho^T} = \frac{\rho^T - 1 + T\rho^{T-1}(1-\rho)}{(1-\rho)(\rho - \rho^T)}.$$

In order to have $f(\rho)' \leq 0$, we need $h(\rho, T) = \rho^T - 1 + T\rho^{T-1}(1-\rho) < 0$. Note that $h(\rho, 1) = 0$. Thus, it is sufficient to show that $h(\rho, T)$ is decreasing in T for all ρ. The first derivative is

$$\begin{aligned}
\frac{\partial h(\rho, T)}{\partial T} =& \rho^{T-1}\left[\rho\log(\rho) + 1 - \rho + T(1-\rho)\log(\rho)\right] \\
\leq& \rho^{T-1}\left[\rho\log(\rho) + 1 - \rho + (1-\rho)\log(\rho)\right] \\
=& \rho^{T-1}\left[1 - \rho + \log(\rho)\right] \\
<& 0.
\end{aligned}$$

Next, we show that Δ^C is decreasing in I. We can rewrite Δ^C as follows

$$\Delta^C = \frac{4(I+1)^2}{K\xi(I-1)^2}\left[I - 1 - \frac{(I+1)^2}{K\xi}\right].$$

We have $\Delta^C > 0$ because $K\xi > \frac{(I+1)^2}{I-1}$. The first derivative is

$$\begin{aligned}\frac{\partial \Delta^C}{\partial I} =& \frac{4}{K\xi}\left[2\left(\frac{I+1}{I-1}\right)\left(-\frac{2}{(I-1)^2}\right)\left(I-1-\frac{(I+1)^2}{K\xi}\right)+\left(\frac{I+1}{I-1}\right)^2\left(1-\frac{2(I+1)}{K\xi}\right)\right] \\ =& \frac{4}{K\xi}\frac{(I+1)(I-3)}{(I-1)^2}\left[1-\frac{2(I+1)^2}{K\xi(I-1)}\right].\end{aligned}$$

The term $1-\frac{2(I+1)^2}{K\xi(I-1)} < 0$ because $K\xi \leq \frac{2(I+1)^2}{I-1}$. Thus, $\frac{\partial \Delta^C}{\partial I} \leq 0$ for $I \geq 3$.

Proof for the price informativeness $\mathcal{I}^C$. The price informativeness $\mathcal{I}^C$ is

$$\mathcal{I}^C = \log\left[\left(I\chi^C\right)^2(\sigma_v/\sigma_u)^2\right] = 2\log\left(\frac{\xi}{I+1}+\frac{I+1}{K}\right)+2\log\left(\frac{\sigma_v}{\sigma_u}\right).$$

The price informativeness $\mathcal{I}^M$ is

$$\mathcal{I}^M = \log\left[\left(I\chi^M\right)^2(\sigma_v/\sigma_u)^2\right] = 2\log\left(\frac{\xi}{2}\frac{\sigma_v}{\sigma_u}\right).$$

Thus, the relative price informativeness $\mathcal{I}^C/\mathcal{I}^M$ is

$$\frac{\mathcal{I}^C}{\mathcal{I}^M} = \frac{\log\left(\frac{\xi}{I+1}\frac{\sigma_v}{\sigma_u}+\frac{I+1}{K}\frac{\sigma_v}{\sigma_u}\right)}{\log\left(\frac{\xi}{2}\frac{\sigma_v}{\sigma_u}\right)}. \tag{D.1}$$

According to (D.1), $\mathcal{I}^C/\mathcal{I}^M$ is increasing in I if $K\xi < (I+1)^2$, which is satisfied for $I \geq 3$ because of condition (C.5) for the existence of the collusive equilibrium. Moreover, $\mathcal{I}^C/\mathcal{I}^M$ is decreasing in K. Thus, $\mathcal{I}^C/\mathcal{I}^M$ is decreasing in ρ because K is increasing in ρ.

To study the effect of σ_u/σ_v, substituting out K using (C.3), equation (D.1) can be rewritten as

$$\frac{\mathcal{I}^C}{\mathcal{I}^M} = \frac{\log\left(\frac{\xi}{I+1}\frac{\sigma_v}{\sigma_u}+(I+1)\frac{\frac{1-\rho}{\rho-\rho^T}+1-\Phi(\omega)}{\phi(\omega)}\sqrt{\frac{\pi}{2}}\right)}{\log\left(\frac{\xi}{2}\frac{\sigma_v}{\sigma_u}\right)}.$$

Obviously, $\mathcal{I}^C/\mathcal{I}^M$ is increasing in σ_u/σ_v and decreasing in ξ, because $\frac{\xi}{I+1}\frac{\sigma_v}{\sigma_u} < \frac{\xi}{2}\frac{\sigma_v}{\sigma_u}$ and $(I+1)\frac{\frac{1-\rho}{\rho-\rho^T}+1-\Phi(\omega)}{\phi(\omega)}\sqrt{\frac{\pi}{2}} > 0$.

Proof for the mispricing $\mathcal{E}^C$. The mispricing $\mathcal{E}^C$ is

$$\mathcal{E}^C = \left|\frac{p^C(v_t)-\mathbb{E}^C[v_t|y_t]}{\mathbb{E}^C[v_t|y_t]-\overline{v}}\right| = \left|\frac{\lambda^C-\gamma^C}{\gamma^C}\right| = \left|\frac{1}{\gamma^C}\left(\frac{\theta\gamma^C+\xi}{\theta+\xi^2}-\gamma^C\right)\right| = \left|\frac{\xi(1-\xi\gamma^C)}{\gamma^C(\theta+\xi^2)}\right|.$$

Consider the case where $\theta = 0$ and ξ is sufficiently large, i.e., $\xi > 1/\gamma^C$. Thus,

$$\mathcal{E}^C = 1 - \frac{1}{\gamma^C \xi} = 1 - \frac{I\chi^C}{\xi} - \frac{\sigma_u^2}{\xi\sigma_v^2}\frac{1}{I\chi^C}.$$

Substituting out χ^C using (C.4), we obtain

$$\mathcal{E}^C = 1 - \left[\frac{1}{I+1} + \frac{I+1}{K\xi} + \frac{\sigma_u^2}{\xi\sigma_v^2}\frac{K(I+1)}{K\xi + (I+1)^2}\right]. \tag{D.2}$$

Obviously, $\mathcal{E}^C$ increases as ξ increases. Moreover,

$$\begin{aligned}\frac{\partial \mathcal{E}^C}{\partial I} &= -\left[\frac{1}{K\xi} - \frac{1}{(I+1)^2} + \frac{\sigma_u^2 K}{\sigma_v^2 \xi} \times \frac{K\xi - (I+1)^2}{[K\xi + (I+1)^2]^2}\right] \\ &= -\left[(I+1)^2 - K\xi\right]\left[\frac{1}{K\xi(I+1)^2} - \frac{\sigma_u^2 K}{\sigma_v^2 \xi}\frac{1}{[K\xi + (I+1)^2]^2}\right].\end{aligned}$$

Because $K\xi \leq \frac{2(I+1)^2}{I-1}$ (equation (C.5)), it is clear that $(I+1)^2 - K\xi > 0$ for $I \geq 3$. Moreover, to ensure that $\frac{1}{K\xi(I+1)^2} - \frac{\sigma_u^2 K}{\sigma_v^2 \xi}\frac{1}{[K\xi+(I+1)^2]^2} \geq 0$, we need

$$K\xi + (I+1)^2 \geq \frac{\sigma_u}{\sigma_v} K(I+1).$$

Because $K\xi > (I+1)^2/(I-1)$ (equation (C.5)), it is sufficient to have

$$(I+1)^2/(I-1) + (I+1)^2 \geq \frac{\sigma_u}{\sigma_v} K(I+1),$$

which implies $\vartheta < I(I+1)/(I-1)$, which is satisfied when ω is not too large or ρ is not very close to 1 (see equation (B.12) for the dependence of ϑ on ω and ρ). Thus, $\frac{\partial \mathcal{E}^C}{\partial I} < 0$ if $I \geq 3$ and $\vartheta < I(I+1)/(I-1)$.

Substituting out K using (C.3), equation (D.3) can be written as

$$\mathcal{E}^C = 1 - \left[\frac{1}{I+1} + \frac{I+1}{\vartheta\xi}\frac{\sigma_u}{\sigma_v} + \frac{\sigma_u}{\xi\sigma_v}\frac{\vartheta(I+1)}{\vartheta\xi\sigma_v/\sigma_u + (I+1)^2}\right].$$

Thus, $\mathcal{E}^C$ is decreasing in σ_u/σ_v. Moreover, we can further rewrite the above equation as follows:

$$\mathcal{E}^C = \frac{I}{I+1} - \frac{I+1}{\xi}\frac{\sigma_u}{\sigma_v}\left[\frac{1}{\vartheta} + \frac{\vartheta}{\vartheta\xi\sigma_v/\sigma_u + (I+1)^2}\right].$$

Obviously, $\mathcal{E}^C$ is increasing in ϑ if $\vartheta < I+1$. We have shown that $K = \frac{\sigma_v}{\sigma_u}\vartheta$ is increasing in ρ. Thus, $\mathcal{E}^C$ is increasing in ρ if $\vartheta < I+1$, which is satisfied when ω is not too large or ρ is not very close to 1 (see equation (B.12) for the dependence of ϑ on ω and ρ).

Proof for the market liquidity $\mathcal{L}^C$. The market liquidity $\mathcal{L}^C$ is

$$\mathcal{L}^C = \frac{1}{\partial |z_t + y_t| / \partial u_t} = \frac{1}{1 - \xi \lambda^C}.$$

In the environment with $\theta = 0$, market liquidity is $\mathcal{L}^C = \frac{1}{|1 - \xi \frac{1}{\xi}|} = \infty$ because prices are determined by market clearing conditions, which are not affected by the noise order flow u_t in expectation. Thus, to analyze how market liquidity depends on parameter values, we consider an environment with $\theta \approx 0$ rather than $\theta = 0$. In this environment, the market liquidity $\mathcal{L}^C$ is

$$\mathcal{L}^C = \frac{1}{\left|1 - \xi \frac{\theta \gamma^C + \xi}{\theta + \xi^2}\right|} \approx \frac{1}{\left|1 - \xi \frac{\theta \gamma^C + \xi}{\xi^2}\right|} = \frac{\xi}{\theta \gamma^C} = \frac{\xi}{\theta} \left(I \chi^C + \frac{\sigma_u^2}{\sigma_v^2} \frac{1}{I \chi^C} \right).$$

Substituting out χ^C using (C.4), we obtain

$$\mathcal{L}^C = \frac{\xi}{\theta} \left[\frac{\xi}{I+1} + \frac{I+1}{K} + \frac{\sigma_u^2}{\sigma_v^2} \frac{K(I+1)}{K\xi + (I+1)^2} \right]. \tag{D.3}$$

The market liquidity in the perfect cartel equilibrium, $\mathcal{L}^M$, is

$$\mathcal{L}^M = \frac{\xi}{\theta} \left(I \chi^M + \frac{\sigma_u^2}{\sigma_v^2} \frac{1}{I \chi^M} \right) = \frac{\xi}{\theta} \left(I \frac{\xi}{2I} + \frac{\sigma_u^2}{\sigma_v^2} \frac{1}{I \frac{\xi}{2I}} \right) = \frac{1}{\theta} \left(\frac{\xi^2}{2} + \frac{2\sigma_u^2}{\sigma_v^2} \right).$$

Thus, the relative market liquidity $\mathcal{L}^C / \mathcal{L}^M$

$$\frac{\mathcal{L}^C}{\mathcal{L}^M} = \frac{\frac{\xi^2}{I+1} + \frac{\xi(I+1)}{K} + \frac{\sigma_u^2}{\sigma_v^2} \frac{K\xi(I+1)}{K\xi + (I+1)^2}}{\frac{\xi^2}{2} + \frac{2\sigma_u^2}{\sigma_v^2}}. \tag{D.4}$$

Clearly, $\mathcal{L}^C / \mathcal{L}^M$ is decreasing in ξ if σ_u / σ_v is sufficiently small; $\mathcal{L}^C / \mathcal{L}^M$ is increasing in σ_u / σ_v if ξ is sufficiently large.

In equation (D.4), the first derivative with respect to K is

$$\begin{aligned} \frac{\partial \mathcal{L}^C / \mathcal{L}^M}{\partial K} =& \frac{\xi}{\frac{\xi^2}{2} + \frac{2\sigma_u^2}{\sigma_v^2}} \left[-\frac{I+1}{K^2} + \frac{\sigma_u^2}{\sigma_v^2} \frac{(I+1)^3}{[K\xi + (I+1)^2]^2} \right] \\ =& \frac{\xi(I+1)}{\vartheta^2 \left(\frac{\xi^2}{2} + \frac{2\sigma_u^2}{\sigma_v^2} \right)} \frac{\sigma_u^2}{\sigma_v^2} \left[-1 + \left[\frac{(I+1)\vartheta}{K\xi + (I+1)^2} \right]^2 \right]. \end{aligned}$$

Thus, $\frac{\partial \mathcal{L}^C / \mathcal{L}^M}{\partial K} < 0$ if $\frac{(I+1)\vartheta}{K\xi + (I+1)^2} < 1$, which is achieved if

$$\vartheta < \frac{K\xi}{I+1} + I + 1 < \frac{2(I+1)^2/(I-1)}{I+1} + I + 1 = \frac{(I+1)^2}{I-1}.$$

where the second inequality is due to condition (C.5). We have shown that K is increasing in ρ.

Thus, $\partial \mathcal{L}^C/\mathcal{L}^M$ is decreasing in ρ if $\vartheta < (I+1)^2/(I-1)$, which is satisfied when ω is not too large or ρ is not very close to 1 (see equation (B.12) for the dependence of ϑ on ω and ρ).

In equation (D.4), the first derivative with respect to I is

$$\begin{aligned}\frac{\partial \mathcal{L}^C/\mathcal{L}^M}{\partial I} =& \frac{\xi}{\frac{\xi^2}{2}+\frac{2\sigma_u^2}{\sigma_v^2}}\left[-\frac{\xi}{(I+1)^2}+\frac{1}{K}+\frac{\sigma_u^2}{\sigma_v^2}\frac{K[K\xi+(I+1)^2]-2K(I+1)^2}{[K\xi+(I+1)^2]^2}\right] \\ =& \frac{\xi}{K(I+1)^2\left(\frac{\xi^2}{2}+\frac{2\sigma_u^2}{\sigma_v^2}\right)}\left[(I+1)^2-K\xi\right]\left[1-\left[\frac{(I+1)\vartheta}{K\xi+(I+1)^2}\right]^2\right].\end{aligned}$$

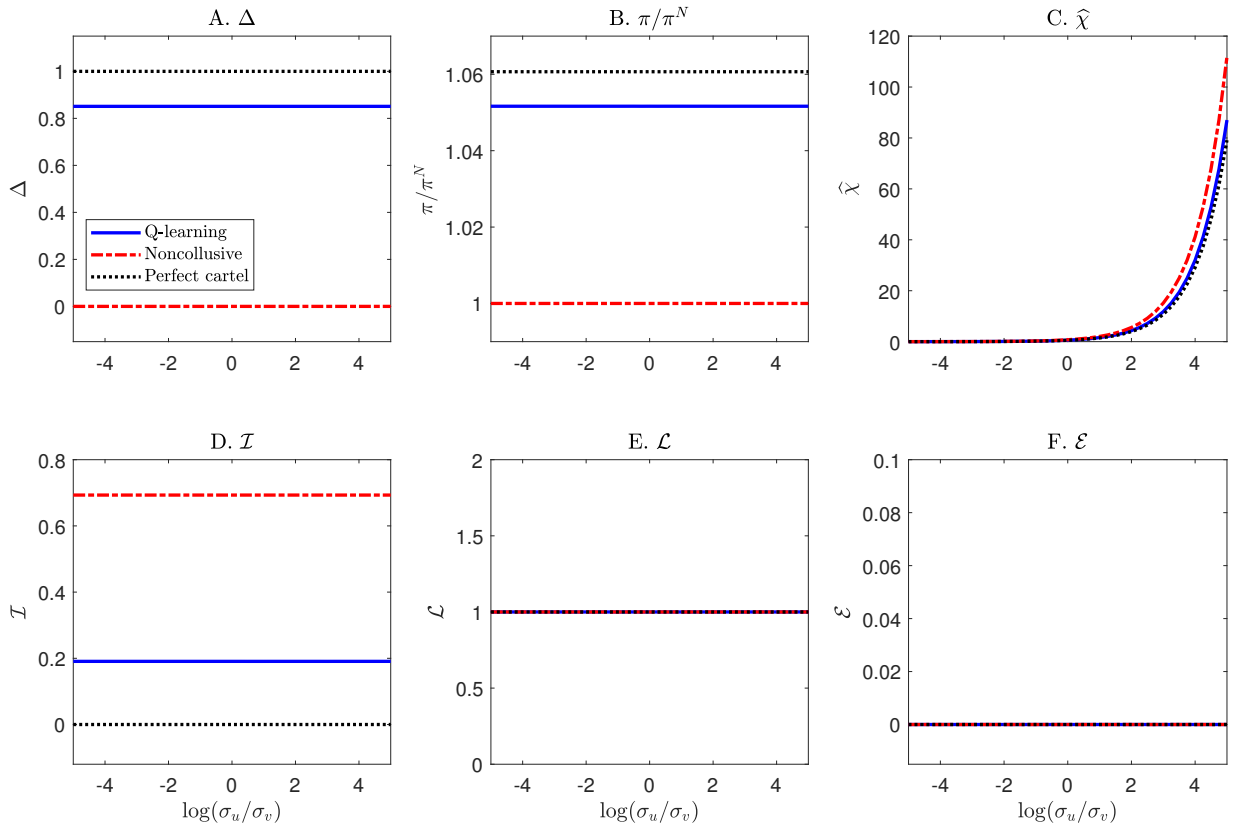
Thus, similarly, we can prove that if $\vartheta < (I+1)^2/(I-1)$, then $1-\left[\frac{(I+1)\vartheta}{K\xi+(I+1)^2}\right]^2 > 0$. Moreover, condition (C.5) implies that $(I+1)^2 - K\xi > 0$ for $I \geq 3$. Therefore, $\frac{\partial \mathcal{L}^C}{\partial I} > 0$ if $I \geq 3$ and $\vartheta < (I+1)^2/(I-1)$.

E Environments with Efficient Prices

In this appendix section, we study informed AI speculators' behavior in the baseline economic environment except for setting $\xi = 0$, which essentially means that the preferred-habitat investor does not exist. Thus, the market maker sets prices purely for price discovery, i.e., $p_t = \mathbb{E}[v_t|y_t]$. This economic environment is similar to Kyle (1985) except for having $I = 2$ informed speculators. Proposition 3.3 in Section 3 indicates that implicit collusion cannot be sustained by any price-trigger strategies in this environment with efficient prices.

Figure A presents the average results across $N = 1,000$ simulation sessions with informed AI speculators. The blue solid lines in panels A and B show that informed AI speculators can attain an average Δ^C of 0.85 and their average profit is about 5% higher than that in the theoretical benchmark of the noncollusive equilibrium. As discussed in Section 5.2, collusion in this environment is achieved through homogenized learning biases. Similar to the property of the Kyle (1985) model, the profits of informed speculators in the theoretical benchmark of the noncollusive Nash equilibrium and the perfect cartel equilibrium are linear in the noise trading risk $\log(\sigma_u/\sigma_v)$. Thus, the red dash-dotted and black dotted lines in panels A and B are flat. Interestingly, the collusive equilibrium formed by informed AI speculators also has a constant Δ^C and π^C/π^N as $\log(\sigma_u/\sigma_v)$ varies along the x-axis, exhibiting a similar scaling property with respect to $\log(\sigma_u/\sigma_v)$. Panel C shows that the informed AI speculators' order sensitivity to asset value $\widehat{\chi}^C$ increases exponentially with $\log(\sigma_u/\sigma_v)$ and linearly with σ_u/σ_v. This scaling property with respect to $\log(\sigma_u/\sigma_v)$ is similar to that in the theoretical benchmarks of the noncollusive Nash equilibrium and the perfect cartel equilibrium, a property that also holds in the Kyle (1985) model.

Panel D shows that due to collusion, price informativeness in the environment with informed AI speculators is lower than that in the theoretical benchmark of the noncollusive Nash equilibrium, but higher than that in the theoretical benchmark of the perfect cartel equilibrium. Moreover, as in



Note: We consider the economic environment with efficient prices as in Kyle (1985). That is, we set $\xi = 0$, implying that the asset's price p_t is determined to minimize pricing errors, with $p_t = \mathbb{E}[v_t|y_t]$. The blue solid line plots the average values of Δ^C, π^C/π^N, $\widehat{\chi}^C$, $\mathcal{I}^C/\mathcal{I}^M$, $\mathcal{L}^C/\mathcal{L}^M$, and $\mathcal{E}^C$ across $N = 1{,}000$ simulation sessions as $\log(\sigma_u/\sigma_v)$ varies. The red dash-dotted and black dotted lines represent the theoretical benchmarks of the noncollusive Nash equilibrium and perfect cartel equilibrium, respectively. The other parameters are set according to the baseline economic environment described in Section 4.7, except for $\xi = 0$.

Figure A: Implications of noise trading risks in the environment with $\xi = 0$.

the Kyle (1985) model, price informativeness remains unchanged as $\log(\sigma_u/\sigma_v)$ varies along the x-axis. Panel E shows that market liquidity is equal to 1 in this environment with efficient prices. This can be directly seen from equation (4.11). In the absence of the preferred-habitat investor, the market maker is the counterparty for informed speculators and the noise trader, and its inventory is equal to $-y_t \equiv -\sum_{i=1}^{I} x_{i,t} - u_t$. Thus, the sensitivity of the market maker's inventory to noise order flows is 1, which holds regardless of the level of noise trading risks or whether informed speculators collude. Panel F shows that mispricing in this environment is 0 because, by definition, prices are efficient, with $p_t = \mathbb{E}[v_t|y_t]$.

F Q-Learning Market Maker

In the baseline economic environment, the market maker analyzes historical data to estimate the pricing rule (ese Section 4.2). In this appendix section, we consider the market maker adopting Q-learning algorithms to learn the pricing rule. All the results presented in the main text are similar; they do not depend on whether the market maker determines the pricing rule using statistical learning or Q-learning algorithms.

Below, we describe the Q-learning algorithm of the market maker. We consider the market maker adopting linear policies to price assets given the combined order flow y_t from informed speculators and the noise trader:

$$p_t = v_t^{MM} + \lambda_t^{MM} y_t, \tag{F.1}$$

where v_t^{MM} and λ_t^{MM} are the market maker's decisions learned from its Q-learning algorithm. Specifically, the market maker's state variable is $s_t = \emptyset$ and action variables are $a_t = \{v_t^{MM}, \lambda_t^{MM}\} \in \mathcal{V} \times \Lambda$. The market maker updates its Q-matrix according to the following learning equation:

$$\begin{aligned}\widehat{Q}_{t+1}^{MM}(v_t^{MM}, \lambda_t^{MM}) =& (1-\alpha^{MM})\widehat{Q}_t^{MM}(s_t, a_t) + \alpha \Big[(y_t - \xi(v_t^{MM} - \overline{v} + \lambda_t^{MM} y_t))^2 \\ &+ \theta(v_t^{MM} + \lambda_t^{MM} y_t - v_t)^2 + \rho^{MM} \min_{v' \in \mathcal{V}, \lambda' \in \Lambda} \widehat{Q}_t^{MM}(v', \lambda') \Big],\end{aligned} \tag{F.2}$$

where the reward in period t is

$$\begin{aligned}(y_t + z_t)^2 + \theta(p_t - v_t)^2 =& (y_t - \xi(p_t - \overline{v}))^2 + \theta(p_t - v_t)^2 \\ =& (y_t - \xi(v_t^{MM} - \overline{v} + \lambda_t^{MM} y_t))^2 + \theta(v_t^{MM} + \lambda_t^{MM} y_t - v_t)^2.\end{aligned} \tag{F.3}$$

The optimal choices of v_t^{MM} and λ_t^{MM} are learned to minimize the Q-matrix. Similar to informed AI speculators' Q-learning algorithms, the market maker also conducts exploration with probability ε_t^{MM} and exploitation with probability $1 - \varepsilon_t^{MM}$. In the exploration mode, the market maker randomly chooses actions v and λ over the set $\mathcal{V} \times \Lambda$.

To implement the Q-learning algorithm for the market maker, we construct discrete grid for v_t^{MM} and λ_t^{MM}. Specifically, we discretize the intervals $[(1-\kappa)v^{MM}, (1+\kappa)v^{MM}]$ and $[(1-\kappa)\lambda^{MM}, (1+\kappa)\lambda^{MM}]$ into n_v and n_λ equally spaced grid points, i.e., $\mathbb{V} = \{v_1^{MM}, \cdots, v_{n_v}^{MM}\}$ and $\Lambda = \{\lambda_1^{MM}, \cdots, \lambda_{n_\lambda}^{MM}\}$. The parameters v^{MM} and λ^{MM} correspond to the optimal values in the theoretical benchmark of the noncollusive equilibrium. The parameter $\kappa > 0$ ensures that the values of v_t and λ_t chosen by the market maker can be different from the theoretical values, v^{MM} and λ^{MM}.

For grid $(v_k^{MM}, \lambda_j^{MM}) \in \mathbb{V} \times \Lambda$, we initialize the market maker's Q-matrix as follows:

$$\widehat{Q}_0^{MM}(v_k^{MM}, \lambda_j^{MM}) = \frac{1}{1-\rho^{MM}} \mathbb{E}\left[(y_t - \xi(v_k^{MM} - \overline{v} + \lambda_j^{MM} y_t))^2 + \theta(v_k^{MM} + \lambda_j^{MM} y_t - v_t)^2 \right]$$

Substituting out $y_t = I\chi^N(v_t - \overline{v}) + u_t$, we obtain

$$\begin{aligned}\widehat{Q}_0^{MM}(v_k^{MM}, \lambda_j^{MM}) =& \frac{1}{1-\rho^{MM}}\left[(1-\xi\lambda_j^{MM})^2((I\chi^N\sigma_v)^2+\sigma_u^2)+\xi^2(v_k^{MM}-\overline{v})^2\right] \\ &+ \frac{\theta}{1-\rho^{MM}}\left[(v_k^{MM}-\overline{v})^2+(\lambda_j^{MM}I\chi^N-1)^2\sigma_v^2+(\lambda_j^{MM}\sigma_u)^2\right]\end{aligned}$$

The exploration rate is $\varepsilon_t^{MM} = e^{-\beta^{MM}t}$, similar to equation (4.5). We set the parameters at $\beta^{MM} = 10^{-4}$, $\alpha^{MM} = 0.1$, $\rho^{MM} = 0.95$, $\kappa = 0.5$, and $n_v = n_\lambda = 31$. The results are similar if we choose different parameter values.

G A Technical Appendix for Learning Biases

In this appendix, we explain why learning biases can lead informed AI speculators to exhibit collusive behavior from a technical perspective. We proceed in three steps. First, in Subsection G.1, we show that learning biases are significant when noise trading risks are high because in this case, the estimation of the Q-matrix cannot properly account for the distribution of the noise order flow u_t due to the failure of the law of large numbers. Second, in Subsection G.2, we show that due to biased learning, the estimated Q-values associated with larger order flows have a larger unconditional variance. Third, in Subsection G.3, we show that large order flows are less likely to be included in the optimal strategies adopted by informed AI speculators after their Q-learning algorithms converge. In other words, biased learning would more likely lead informed AI speculators to optimally trade with small order flows, which coincide with the order flows adopted in the theoretical benchmark of the collusive Nash equilibrium. Taken together, we show that in the presence of high noise trading risks, collusive outcomes emerge due to informed AI speculators' homogenized learning biases.

G.1 Biased Learning When Noise Trading Risks are High

First, we explain that when noise trading risks are high, there exist learning biases for the Q-matrix due to the failure of the law of large numbers.

Learning biases are caused by a generic feature of RL algorithms. As discussed in Section 2, Q-learning algorithms cannot take expectations due to the absence of knowledge about the underlying economic environment (e.g., the distribution of the noise order flow u_t). In each period t, the algorithm updates the value of one single state-action pair (s, x_i) of the Q-matrix according to the currently realized profit $\pi_{i,t}$ (see equation (2.4)) rather than the expected profit $\mathbb{E}[\pi_{i,t}|s, x_i]$ as in a rational-expectations framework. Biases may exist in Q-value estimation because updating the Q-matrix sequentially based on past realized profits may not accurately reflect the expected profit, due to the failure of the law of large numbers.

To illustrate this point, we focus on a particular state-action pair (s, x_i) that has been visited T times in the past. Let $\tau = 1, 2, ..., T$ be the τ-th visit to the state-action pair (s, x_i). Let $t(\tau)$ be the

period for the τ-th time that the Q-learning algorithm visits the state-action pair (s, x_i). According to Equation (2.4), in each period t, the Q-learning algorithm only updates the state-action pair of the Q-matrix that the algorithm visits. Thus, the state-action pair (s, x_i) has been updated T times in the past, and these updates occur in periods $t(\tau)$ for $\tau = 1, 2, ..., T$. In other words, for each $\tau = 1, 2, ..., T$, the value of $\widehat{Q}_{i,t}(s, x_i)$ is a constant and equal to $\widehat{Q}_{i,t(\tau)+1}(s, x_i)$ from period $t = t(\tau) + 1$ to period $t = t(\tau + 1)$ and gets updated with a new value, $\widehat{Q}_{i,t(\tau+1)+1}(s, x_i)$, in period $t(\tau + 1) + 1$.

Based on equation (2.4), for the T-th visit to the state-action pair (s, x_i), we have

$$\widehat{Q}_{i,t(T)+1}(s, x_i) = (1 - \alpha)\widehat{Q}_{i,t(T)}(s, x_i) + \alpha \left[(v_{t(T)} - p_{t(T)})x_i + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(T)}(s_{t(T)+1}, x') \right] \tag{G.1}$$

For the $(T - 1)$-th visit to the state-action pair (s, x_i), we have

$$\widehat{Q}_{i,t(T-1)+1}(s, x_i) = (1 - \alpha)\widehat{Q}_{i,t(T-1)}(s, x_i) + \alpha \left[(v_{t(T-1)} - p_{t(T-1)})x_i + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(T-1)}(s_{t(T-1)+1}, x') \right] \tag{G.2}$$

..., and for the 1-st visit to the state-action pair (s, x_i), we have

$$\widehat{Q}_{i,t(1)+1}(s, x_i) = (1 - \alpha)\widehat{Q}_{i,t(1)}(s, x_i) + \alpha \left[(v_{t(1)} - p_{t(1)})x_i + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(1)}(s_{t(1)+1}, x') \right] \tag{G.3}$$

Because the Q-value for the state-action pair (s, x_i) does not change from $t = t(\tau) + 1$ to $t = t(\tau + 1)$, we have $\widehat{Q}_{i,t(\tau)+1}(s, x_i) = \widehat{Q}_{i,t(\tau+1)}(s, x_i)$, for $\tau = 1, 2, ..., T - 1$. Thus, combining above equations, we derive

$$\begin{aligned} \widehat{Q}_{i,t(T)+1}(s, x_i) = &\sum_{\tau=0}^{T-1} \alpha(1 - \alpha)^\tau \left[(v_{t(T-\tau)} - p_{t(T-\tau)})x + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(T-\tau)}(s_{t(T-\tau)+1}, x') \right] \\ &+ (1 - \alpha)^T \widehat{Q}_{i,0}(s, x_i). \end{aligned} \tag{G.4}$$

As $T \to \infty$, we can omit the last term and rewrite the above equation as

$$\widehat{Q}_{i,t(T)+1}(s, x_i) = \sum_{\tau=0}^{T} \alpha(1 - \alpha)^\tau \left[(v_{t(T-\tau)} - p_{t(T-\tau)})x_i + \rho \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(T-\tau)}(s_{t(T-\tau)+1}, x') \right]. \tag{G.5}$$

By substituting out $p_{t(T-\tau)}$, the above equation becomes

$$\begin{aligned} \widehat{Q}_{i,t(T)+1}(s, x_i) = &\sum_{\tau=0}^{T} \alpha(1 - \alpha)^\tau [v_{t(T-\tau)} - \overline{v} - \lambda(y_{t(T-\tau)} - u_{t(T-\tau)})]x_i \\ &- \alpha\lambda x_i \sum_{\tau=0}^{T} (1 - \alpha)^\tau u_{t(T-\tau)} + \rho \sum_{\tau=0}^{T} \max_{x' \in \mathcal{X}} \widehat{Q}_{i,t(T-\tau)}(s_{t(T-\tau)+1}, x'). \end{aligned} \tag{G.6}$$

The term $\alpha\lambda x_i \sum_{\tau=0}^{T} (1 - \alpha)^\tau u_{t(T-\tau)}$ represents a stochastic term that depends on the noise order flow $u_{t(T-\tau)}$. With $\mathbb{E}[u_t] = 0$, the estimation for the limit value of $\widehat{Q}_{i,t(T)+1}(s, x_i)$ is unbiased only

if $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)} = 0$ as $T \to \infty$[20], which occurs if $\alpha \to 0$. Thus, for any $\alpha > 0$, the term $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)}$ would bias the estimate of $\widehat{Q}_{i,t(T)+1}(s, x_i)$. This is due to the failure of the law of large numbers because in general, as $T \to \infty$, we have $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)} \neq \alpha\lambda x_i \mathbb{E}[u_t]$ unless $\alpha \to 0$.

The magnitude of learning biases depends on the importance of the term $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)}$ relative to other terms in equation (G.6), as $T \to \infty$. Specifically, learning biases are absent when there is no noise trading risk (i.e., $\sigma_u/\sigma_v = 0$) or when $\alpha \approx 0$. Learning biases become more significant when σ_u/σ_v is higher, λ is higher, ρ is lower, or α is higher.

G.2 Complementarity Between Informed AI Speculators' Order and Noise Order

Second, we show that due to biased learning, the estimated Q-values associated with larger order flows have larger unconditional variances.

To begin with, we decompose the per-period profit $(v_t - p_t)x_i$ that an informed speculator i receives when choosing order flow $x_i \in \mathcal{X}$ in period t into two parts:

$$(v_t - p_t)x_i = \left[v_t - \overline{v} - \lambda(y_t - u_t)\right] x_i - \lambda x_i u_t. \tag{G.7}$$

The term $\left[v_t - \overline{v} - \lambda(y_t - u_t)\right] x_i$ captures the profit determined by the asset's fundamental value v_t and the term $\lambda x_i u_t$ captures the profit determined by the noise order flow u_t. Through the term $\lambda x_i u_t$ in equation (G.7), there exists complementarity between the informed speculator's order flow x_i and the noise order flow u_t in determining per-period profits. This complementarity implies that, choosing larger order flows (i.e., a larger absolute value $|x_i|$) would amplify the impact of the noise order flow u_t on per-period profits.

Because the estimated Q-value is the accumulated discounted per-period profits realized in the past, the complementarity between x_i and u_t in equation (G.7) would propagate to equation (G.6), captured by the term $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)}$. In the absence of learning biases (i.e., when $\alpha \to 0$), we have $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)} \approx \alpha\lambda x_i \mathbb{E}[u_t] = 0$ as $T \to \infty$, so that the unbiased estimate of the Q-value is not affected by the complementarity. However, as long as $\alpha > 0$, we would have $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)} \neq 0$, and thus, the estimated limit Q-value is biased, due to the failure of the law of large numbers. The biased learning implies that the estimated Q-value of an informed AI speculator's particular order flow x_i is path dependent, crucially depending on the realized noise order flow u_t in the past periods when the informed AI speculator chose x_i.

Thus, in the presence of learning biases, there exists complementarity between x_i and u_t in determining the estimated Q-value. This complementarity implies that the estimated Q-values associated with larger order flows have larger unconditional variances.

[20] To see why unbiasedness requires $\alpha\lambda x_i \sum_{\tau=0}^{T}(1-\alpha)^{\tau} u_{t(T-\tau)} = 0$ as $T \to \infty$, note that the Q-matrix is essentially a precursor of the value function (i.e., $V_i(s) \equiv \max_{x' \in \mathcal{X}} Q_i(s, x')$, see equation (2.1)), which represents the discounted "expected" profits. In our model, the noise order u_t should have no direct effect on an informed speculator's "expected" profits except for affecting its order flow $x_{i,t}$.

G.3 Impacts of Biased Learning on Optimal Strategies

Third, we show that large order flows are less likely to be the optimal strategies adopted by informed AI speculators after their Q-learning algorithms converge. In other words, learning biases would more likely lead informed AI speculators to optimally choose small order flows, which coincide with those order flows in the theoretical benchmark of the collusive Nash equilibrium.

Before discussing why learning biases make the choice of large order flows less likely, it is useful to clarify that although informed AI speculators start their Q-learning algorithms with a mix of the exploration mode and the exploitation mode, it must be the case that the exploration rate drops to zero at some point before Q-learning algorithms to converge. In other words, in a long period of time right before Q-learning algorithms converge, informed AI speculators must be in pure exploitation mode, choosing the order flows that maximize their Q-values rather than choosing order flows randomly. Therefore, without loss of generality, we focus on the exploitation mode in our discussions below.

To fix the idea, consider a simple setting in which there is a single state s and each informed AI speculator i's order flow x_i can take two different values, $x_i = x_S, x_L$, with $0 < x_S < x_L$, meaning that x_L is a large order flow and x_S is a small order flow. As discussed above, in the presence of learning biases caused by noise trading risks, there is complementarity between x_i and u_t in determining the estimated Q-value. Thus, relative to the estimated Q-value associated with the small order flow x_S, the estimated Q-value associated with the large order flow x_L has a large unconditional variance (see equation (G.6)). Let $[\underline{Q}(s, x_S), \overline{Q}(s, x_S)]$ and $[\underline{Q}(s, x_L), \overline{Q}(s, x_L)]$ be the 99.9% confidence interval of the estimated Q-value for order flows x_S and x_L, respectively. Thus, we have $[\underline{Q}(x_S), \overline{Q}(s, x_S)] \subset [\underline{Q}(s, x_L), \overline{Q}(s, x_L)]$.

Because the informed AI speculator is purely in the exploitation mode, in any period t, its order flow is determined according to $\text{argmax}_{x_S, x_L} \left\{ \widehat{Q}_{i,t}(s, x_S), \widehat{Q}_{i,t}(s, x_L) \right\}$. There are two cases, either $\widehat{Q}_{i,t}(s, x_L) > \widehat{Q}_{i,t}(s, x_S)$ or $\widehat{Q}_{i,t}(s, x_L) <= \widehat{Q}_{i,t}(s, x_S)$. In the first case, for $\tau > [t, t']$, the informed AI speculator would keep choosing x_L to update $\widehat{Q}_{i,\tau}(s, x_L)$ while $\widehat{Q}_{i,\tau}(s, x_S)$ remains unchanged at $\widehat{Q}_{i,t}(s, x_S)$. The period $t' > t$ is the first passage time for $\widehat{Q}_{i,t'}(s, x_L) <= \widehat{Q}_{i,t'}(s, x_S)$. From period t' on, the informed AI speculator switches from choosing the large order flow x_L to choosing the small order flow x_S, and fall into the second case as described below.

In the second case, for $\tau > [t, t']$, the informed AI speculator would keep choosing x_S to update $\widehat{Q}_{i,\tau}(s, x_S)$ while $\widehat{Q}_{i,\tau}(s, x_L)$ remains unchanged at $\widehat{Q}_{i,t}(s, x_L)$. The period $t' > t$ is the first passage time for $\widehat{Q}_{i,t'}(s, x_L) > \widehat{Q}_{i,t'}(s, x_S)$. From period t' on, the informed AI speculator switches from choosing the small order flow x_S to choosing the large order flow x_L, and fall into the first case as described above.

These two cases alternate over time. In one simulation session, given our convergence criterion specified in Section 4.8 (i.e., stability of optimal strategy for $T = 100,000$ consecutive periods), eventually, the optimal strategy will converge to x_S with probability $\mathcal{P}$ and x_L with probability $1 - \mathcal{P}$. We have $p > 0.5$ because $\underline{Q}(s, x_L) < \underline{Q}(s, x_S)$. The probability $\mathcal{P}$ is higher if the estimated Q-value associated with the order flow x_L has a larger probability to be in the

interval $[\underline{Q}(s, x_L), \underline{Q}(s, x_S)]$, which happens when noise trading risks are higher (i.e., higher σ_u/σ_v so the magnitude of learning biases is larger) or the difference in order flows is larger (i.e., larger $x_L - x_S$). This explains why learning biases make the choice of large order flows less likely.

According to our model in Section 3, the sensitivity of informed speculators' order flow to the asset's value v_t is lower under collusion, i.e., $\chi^M \leq \chi^C < \chi^N$. Because informed speculator i's order $x_{i,t}$ is $x_{i,t} = \chi(v_t - \overline{v})$, the absolute value of its order flows satisfies $|x_{i,t}^M| \leq |x_{i,t}^C| < |x_{i,t}^N|$ for any v_t, indicating that informed speculators would collude if they adopt more conservative (i.e., choosing order flows with smaller magnitude), rather than more aggressive, trading strategies. Taken together, it is clear that in the presence of high noise trading risks, homogenized learning biases lead to collusive outcomes.