NYU STERN
NEW YORK UNIVERSITY · LEONARD N. STERN SCHOOL OF BUSINESS
SALOMON CENTER FOR THE STUDY OF FINANCIAL INSTITUTIONS

## Program for the Stern Microstructure Meeting, Friday, May 20, 2011

| | |
|---|---|
| *Organizer* | Joel Hasbrouck, Stern School |
| *Program Committee* | Charles Jones, Columbia University<br>Bruce Lehmann, UCSD<br>Avanidhar Subrahmanyam, UCLA |

Call for papers

Registration: email to Mary Jaffier (mjaffier@stern.nyu.edu, please include "Stern Microstructure" in the subject). Other inquiries: jhasbrou@stern.nyu.edu.

Download program and all papers as a single pdf

Friday, May 20, 2011

| | |
|---|---|
| 8:30 am - 9:00 | Continental Breakfast |
| 9:00 - 10:00 | The Growth and Limits of Arbitrage: Evidence from Short Interest<br>Samuel Hanson, Harvard Business School<br>Adi Sunderam, Harvard Business School<br><br>Discussant: Antti Petajisto, Stern School, New York University |
| 10:00 - 11:00 | Subsidizing Liquidity: The Impact of Make/Take Fees on Market Quality<br>Katya Malinova, University of Toronto<br>Andreas Park, University of Toronto<br><br>Discussant: Ohad Kadan, Olin School, Washington University, St. Louis |
| 11:00 - 11:15 | Break |
| 11:15 - 12:15 | Trading Frenzies and Their Impact on Real Investment<br>Itay Goldstein, Wharton |

| | |
|---|---|
| | Emre Ozdenoren, London Business School<br>Kathy Yuan, LSE<br><br>Discussant: Wei Xiong, Princeton University |
| 12:15-1:15 | Lunch<br>Speaker: TBA |
| 1:15-2:15 | Learning from Prices, Liquidity Spillovers, and Market Segmentation<br>Giovanni Cespa, Cass Business School<br>Thierry Foucault, HEC, Paris<br><br>Discussant: David Skeie, Federal Reserve Bank of New York |
| 2:15-3:15 | Notes on Bonds: Liquidity at All Costs in the Great Recession<br>David Musto, The Wharton School, University of Pennsylvania<br>Greg Nini, The Wharton School, University of Pennsylvania<br>Krista Schwarz, The Wharton School, University of Pennsylvania<br><br>Discussant: Michael Fleming, Federal Reserve Bank of New York |
| 3:15-3:30 | Break |
| 3:30-4:30 | Illiquidity Premia in the Equity Options Market<br>Peter Christoffersen, University of Torornto<br>Ruslan Goyenko, McGill University<br>Kris Jacobs, University of Houston<br>Mehdi Karoui, McGill University<br><br>Discussant, Menachem Brenner, Stern School, NYU |
| 4:30 | Adjourn |

# The Growth and Limits of Arbitrage: Evidence from Short Interest*

Samuel G. Hanson
shanson@hbs.edu
Harvard Business School

Adi Sunderam
asunderam@hbs.edu
Harvard Business School

April 2011

**Abstract**

We develop a novel methodology to infer the amount of capital allocated to quantitative equity arbitrage strategies from stock-level short interest data. Using this technique, which exploits time-series variation in the cross-section of short interest, we document that the amount of capital devoted to quantitative equity strategies such as value and momentum has increased significantly since the early 1990s. We find evidence suggesting that arbitrageurs have reacted to heightened competition by altering their strategies. Specifically, arbitrageurs increasingly favor stocks where the risk of over-crowding is lower such as small stocks and stocks with weaker mispricing signals. We then use these strategy-level capital measures to test theories about the limits of arbitrage. We find that strategy-level capital flows are influenced by past strategy returns, strategy return volatility, and past returns to other strategies in the directions predicted by these theories. Finally, we find that arbitrageurs have invested more capital in strategies prior to periods when those strategies perform poorly.

* We would like to thank John Campbell, Sergey Chernenko, Lauren Cohen, Harrison Hong, Jakub Jurek, Erik Stafford, Luis Viciera, and seminar participants at Harvard University and Princeton University for helpful comments and suggestions. Malcolm Baker, Robin Greenwood, and Jeremy Stein deserve special thanks for their advice and guidance throughout this project.

## I. Introduction

The professional arbitrage industry devoted to exploiting so-called equity market "anomalies" such as the value and momentum effects has grown explosively in recent years. Assets managed by long-short equity hedge funds, which pursue strategies that seek to profit from these anomalies, grew from $103 billion in 2000 to $364 billion at the end of 2009 according to Lipper – an average annual growth rate exceeding 15%. This growth has been accompanied by increased interest from market participants and academics alike.

Of primary concern to both groups is the interaction between arbitrage capital and strategy returns. Such interactions may occur at two frequencies. First, long-run returns to anomaly strategies may eventually be competed away by the long-term growth of arbitrage capital (Stein (2009)). Second, there may be higher-frequency feedback between returns and capital, such as performance chasing and deleveraging spirals, which may limit the extent to which anomaly returns are arbitraged away.

Both types of feedback have been studied extensively in the theoretical literature. However, empirical work has been hindered by the lack of appropriate data. In particular, while arbitrage *strategies* are often the relevant unit of economic analysis for assessing these theories, the amount of capital allocated to various strategies is unknown because existing data are aggregated to (at least) the fund level. As a result, researchers have often been forced to test these theories at either the individual stock level or the fund level.[1]

In this paper, we propose a novel technique for measuring the amount of capital allocated to an equity arbitrage strategy at a given time. We focus on quantitative equity strategies, which attempt to exploit the return anomalies uncovered by academic finance over the past 25 years.

[1] See e.g., Aragon and Strahan (2010), Ben-David Franzoni, and Moussawi (2010), Savor and Gamboa-Cavazos (2011).

These strategies typically use short sales to construct low- or zero-beta portfolios that generate abnormal risk-adjusted returns or “alpha.”

Our key insight is that the cross-section of short interest reveals how intensely arbitrageurs are using a particular quantitative equity strategy at a given time. For instance, when the cross section of short interest is particularly weighted towards growth stocks, we should infer that a lot of capital is playing a value strategy. We can then interpret time-series variation in the cross-section of short interest as variation in the amount of capital playing various strategies. We formalize this intuition in a regression setting. Specifically, we run cross-sectional regressions explaining stock-level short interest and interpret the coefficients from these regressions as proxies for strategy-level capital.

Short interest is a good laboratory for studying strategy-level arbitrage capital flows for several reasons. First, short sellers are typically sophisticated investors.[2] Second, the costs of short selling make it more likely that short positions are put on by managers who are actively seeking alpha. Finally, short interest data may be more informative than long-side data because, in the aggregate, long-side institutional investors hold the market portfolio and show little tendency to bet on characteristics known to predict returns (Lewellen 2010). In other words, any long-side analysis must screen out the large number of institutions that passively index; otherwise, it will have little power to detect time variation in arbitrage capital.

Focusing on the value and price momentum strategies, we first use our capital measures to explore low-frequency trends in arbitrage capital. We show that capital in both strategies has increased dramatically, particularly since the early 2000s. We consider several possible explanations for these trends. First, we ask whether the trends could be driven by an expansion of

[2] By some estimates, hedge funds account for 85% of short positions in the U.S. equity market (see e.g., Goldman Sachs Hedge Fund Trend Monitor, February 20, 2008).

share lending supply. We attempt to disentangle demand and supply shifts by using institutional ownership as a proxy for lendable share supply. Interestingly, we find similar upward trends when we focus only on stocks with high institutional ownership, which are less likely to experience a significant easing of supply constraints. Thus, we argue that shifts in shorting demand have played an important role in driving the trends we find.

We next consider the possibility that the increases in arbitrage capital we observe could be the result of slow diffusion of information about strategy profitability, leading to increased shorting demand. If this is the case, the degree of competition among arbitrageurs may have increased as the pool of arbitrage capital has grown. We find some evidence consistent with this interpretation. For instance, as more capital uses a reliable signal of mispricing, arbitrageurs may become concerned about the capacity or profitability of the signal. As a result, they may begin using a signal that is less crowded but has historically been a weaker indicator of expected returns. Similarly, if crowding or diminished strategy profitability is a greater concern in large, liquid stocks, then smaller stocks might become more appealing. Consistent with these predictions, we find that arbitrageurs have shifted to weaker signals and smaller stocks in recent years.

The low frequency growth of arbitrage capital we document suggests that the returns to the value and momentum strategies may be competed away over time. However, a large theoretical literature suggests that agency problems and funding constraints may create limits of arbitrage that allow abnormal returns to persist even in the face of substantial arbitrage capital.[3] Thus, we explore the higher-frequency feedback between capital and strategy returns and ask whether the observed patterns are consistent with theories of limited arbitrage.

[3] A partial list includes Shleifer and Vishny (1997), Barberis and Shleifer (2003), Brunnermeier and Pedersen (2009), Stein (2009), and Gromb and Vayanos (2010).

We first explore the relationship between capital flows and past performance. We find evidence of a positive performance-flow relationship for momentum. After low returns, capital tends to flow out of momentum strategies. We also explore how arbitrage capital responds to changes in the volatility of strategy returns. For both value and momentum, we find that capital tends to exit strategies following increases in volatility. We also find evidence of cross-strategy spillovers and that funding constraints may impede arbitrage activity. Specifically, we find that capital exits momentum when other arbitrage strategies do poorly and when the Treasury Eurodollar spread widens. This suggests that hedge funds may choose to liquidate momentum positions in order to meet margin requirements or capital redemptions.

Finally, we examine the relationship between arbitrage capital and subsequent strategy returns. We find that capital flows into both value and momentum negatively forecast future returns. This finding, combined with the fact that strategy returns mean revert in our sample, means that arbitrage capital demonstrates negative market timing ability. Taken together, these results suggest that quantitative equity arbitrage does suffer from the limits suggested by the theoretical literature.

The remainder of this paper is organized as follows. Section II describes the data. Section III describes our methodology and examines trends in arbitrage capital since 1992. Section IV describes our results concerning the feedback between arbitrage capital and strategy returns. Section V concludes.

## II. Data

We use monthly data on short interest from January 1992 through December 2010. From 1992-2007, short interest data for NYSE and AMEX stocks is downloaded from Bloomberg and data for NASDAQ stocks is obtained directly from the exchange. From 2008-2010, we obtain short interest data for all stocks from Compustat.[4] Short interest for stock $i$ in month $t$, $SHORT_{i,t}$, is the total number of uncovered shares sold short for transactions settling on or before the 15$^{th}$ of the month. We normalize short interest by total shares outstanding as of the reporting date to form short interest ratios: $SR_{i,t} = SHORT_{i,t} / SHROUT_{i,t}$. Short interest ratios are winsorized at the 99.5%-tile in each cross-section.

Several trends in the short interest data are worth highlighting. Figure 1 shows that short interest rose significantly during our 1992-2010 sample on both an equal- and value-weighted basis. Short interest ratios trended upwards during the mid-1990s, declined somewhat during the technology bubble as noted by Lamont and Stein (2004), and rose dramatically from 2001 to 2007. The financial crisis period from 2007 to 2009 saw large swings in short interest. Short interest peaked in July 2007 and registered a marked drop over the following three months, presumably due to the fire-sale induced de-leveraging associated with the "quant meltdown" of August 2007 (see e.g., Khandani and Lo (2007, 2008) and Pedersen (2009)). Short interest rose rapidly again in the first half of 2008, peaking in July 2008 before declining sharply after September 2008 when the SEC imposed a partial ban on short sales for financial stocks.[5] Aggregate short interest levels stabilized in 2009 and 2010.

Moreover, short interest among small stocks has surged since 2000. In fact, the entire

---

[4] Compustat's short interest data is provided by FT Interactive and is available in the Security Monthly file beginning in 2003. From 2003-2008, short interest ratios constructed using Compustat data are virtually indistinguishable from ratios constructed using our Bloomberg and NASDAQ data. The few discrepancies appear to stem from disagreements about the exact timing of stock splits.

[5] On September 19, 2008, the Securities and Exchange Commission adopted an emergency order that temporarily banned most short sales in over 900 financial stocks. However, Figure A3 in the Internet Appendix shows that short interest ratios declined for both nonfinancial and financial firms following the imposition of the ban. Boehmer, Jones, and Zhang (2009) provide a detailed examination of the market impact of the short sales ban.

cross-sectional relationship between size and short interest has shifted dramatically. In Figure 2 we plot average short interest ratios by NYSE size decile at six different years in our sample. Short interest ratios for firms in size deciles 2 through 5 have risen sharply since 1999, all hovering near 10% as of year-end 2007 (short interest for small stocks declined somewhat from 2007 to 2009).[6] Average short interest for size decile 1 has also grown, but still lags other small stocks. By contrast, short interest for size decile 10 has been remarkably stable. Although we will see that quantitative equity signals are associated with significant differences in short interest, the growth of quantitative equity arbitrage does not appear to completely explain the broad surge in short interest among small stocks.[7]

To this short interest data, we add stock characteristics from CRSP and Compustat, including size (*ME*) deciles, book-to-market (*B/M*) deciles, and past 12-month return deciles (i.e., "momentum" deciles). All deciles are based on NYSE breakpoints. We also compute the fraction of shares held by 13-F institutions as of the most recent quarter-end, the three month moving average of share turnover (volume over shares outstanding), trailing 12-month return volatility, exchange dummies (i.e., a NASDAQ dummy and an NYSE dummy), and a dummy indicating whether a firm has convertible securities outstanding. All continuous variables are winsorized in each cross-section at the 0.5% and 99.5%-tiles. Appendix A provides further detail on the data and relevant variable definitions.

## III. The Evolution of Arbitrage Capital: 1992-2010

[6] These trends are not driven by outliers: the entire distribution of *SR* shifted to the right for small stocks.

[7] Possible explanations include the rapid growth of non-quantitative hedge funds, the expansion of institutional share-lending programs, and technological changes (e.g., the evolution of the prime-brokerage and information technology may have lowered search and other transaction costs associated with short sales).

*A. Short Interest for Extreme Growth and Loser Stocks*

The basic premise of our methodology for measuring arbitrage capital is that short interest should be high for stocks that an arbitrage strategy recommends shorting. We test this key assumption before further fleshing out our approach. To do so, we trace out the "event-time" path of short interest ratios for stocks falling into the lowest *B/M* decile by estimating the following quarterly panel regression:

$$\begin{aligned} SR_{it} = {} & \delta^{-8} 1_{it}^{-8}\{B/M\} + \cdots + \delta^{0} 1_{it}^{0}\{B/M\} + \cdots + \delta^{+8} 1_{it}^{+8}\{B/M\} \\ & + \boldsymbol{\delta}^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}' \mathbf{x}_{it} + Stock_i + Time_t + \varepsilon_{it}. \end{aligned} \tag{1}$$

Here the $1_{it}^{-k}\{B/M\}$ dummy indicates that stock *i* will enter the lowest *B/M* decile in *k* quarters (i.e., at *t*+*k*). By contrast, the $1_{it}^{+k}\{B/M\}$ dummy indicate that stock *i* exited the lowest *B/M* decile *k* quarters ago.[8] While the $1_{it}^{-k}\{B/M\}$ indicators are forward-looking, our goal is not to forecast short interest. Rather it is simply to understand the dynamics of short interest for the group of stocks that eventually fall into the extreme growth decile.

The regression includes a full set of size ($\mathbf{1}_{it}^{SIZE}$) and momentum ($\mathbf{1}_{it}^{MOM}$) decile dummies as well as a vector ($\mathbf{x}_{it}$) of additional controls that have previously been shown to be important determinants of short interest, namely, institutional ownership, 3-month turnover, trailing 12-month return volatility, dummies for the exchange on which a stock trades, and a convertible dummy. Since equation (1) includes a full set of stock fixed effects, identification is based exclusively on within-stock variation consistent with our "event-time" interpretation of the results. Equation (1) also includes a full set of time effects. To trace out the event-time path of

[8] If a firm has a "spell" of consecutive quarters in the lowest *B/M* decile, the dummies are coded relative to the first quarter in the spell. Similarly, the dummies are coded relative to the last quarter in the spell. Thus, the event-time path is identified using true transitions into and out of the extreme decile. We also include dummies for the number of consecutive quarters that a stock has spent in the lowest *B/M* decile. While not shown here, we find that *SR* increases with each quarter that a firm spends in the extreme growth decile.

short interest for momentum losers, we estimate an analogous regression. Since momentum deciles are updated each month, this regression is run with monthly data.

Figure 3 plots the coefficients on the event-time dummies from regression (1). We draw 95% confidence bands around the estimates using standard errors that cluster by both stock and time as in Thompson (2011). Over our 1992-2010 sample, entering the lowest *B/M* decile raised *SR* by 59 bps, while entering the lowest momentum decile raised *SR* by 75 bps. The average short interest ratio over our sample is 240 bps, so these magnitudes are economically significant. Thus, Figure 3 confirms the basic premise underlying our methodology.[9]

*B. Measuring Capital Intensities Using Short Interest*

Figure 3 confirms that short interest is high for stocks that familiar quantitative strategies recommend shorting. This means that each cross-section of short interest is potentially informative about the distribution of capital across arbitrage strategies. To see this, imagine a situation where there are only two stocks *A* and *B*, and the only short sellers are quantitative investors. If *A* is the only stock momentum traders short and *B* is the only stock value traders short, then by observing the cross-section of short interest, we are actually observing the amounts of short-side capital playing momentum and value, respectively.

Two caveats follow from this simple thought experiment. First, if the value and momentum strategies perfectly overlap and both recommend going long *B* and short *A*, then the cross-section of short interest contains no information about the allocation of capital across strategies. Empirically, momentum and value are not highly collinear, but do have some overlap. For this reason, we favor an approach based on cross-sectional regressions over univariate

[9] Figure 3 shows that the increase in short interest for growth stocks is concentrated in the quarter when they enter the lowest *B/M* decile, whereas the increase in *SR* for 12-month momentum losers is more gradual. Presumably, this reflects the fact that some arbitrageurs play shorter horizon (e.g., 6-month) momentum strategies.

alternatives (e.g., averaging *SR* by momentum decile) to handle overlapping recommendations across strategies and to control for other determinants of *SR*. If we did not use a cross-sectional regression framework, we would be mixing these confounds together in varying proportions, making it difficult to compare the resulting measures over time.

Second, our methodology assumes that we know which stocks particular strategies would short. Clearly, quantitative investors use more sophisticated expected return models than those implicit in the simple cross-sectional sorts we use below. While our sorts will not perfectly capture the value or momentum portfolios generated by state-of-the-art quantitative investing techniques, our approach is a reasonable first approximation. Furthermore, Figure 3 confirms that arbitrageurs do respond to the information contained in these cross-sectional sorts.

We adopt a relatively non-parametric specification for our cross-sectional regressions. For each cross-section $t$, we regress stock $i$'s short interest ratio on a full set of size, book-to-market, and momentum decile dummies (the omitted dummy is always decile 5). We also include the same set of additional controls $\mathbf{x}_{it}$ that were used in equation (1) above. Thus, our baseline specification for each cross section is:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}. \tag{2}$$

The coefficient on the dummy for the lowest momentum decile, $\delta_t^{MOM(1)}$, reflects the increase in short interest at time $t$ associated with being an extreme loser relative to the omitted decile 5. Thus, $\delta_t^{MOM(1)}$ forms our main proxy for the quantity of short-side capital devoted to momentum strategies at time $t$.[10]

[10] The coefficients for the other deciles are also potentially informative. For instance, if arbitrage capital is flowing into momentum, we might also expect to see reductions in short interest for extreme. As discussed in the Internet Appendix, we have experimented with other measures of strategy intensities such as the spread in *SR* between extreme losers and winners. These other measures lead to similar conclusions.

One practical issue is whether and how to smooth the raw time series of monthly coefficients. Smoothing reduces the measurement error associated with monthly cross-sectional estimates. Since we explore low and medium frequency variation in strategy capital, we smooth coefficients by estimating annual and quarterly panel regressions. That is, we stack all firm-month observations for a given year (or quarter) in a panel and estimate a single pooled regression that includes month fixed effects.[11] However, all our results are qualitatively unchanged if we do not smooth, albeit with slightly reduced significance in a few cases.

We have examined a number of other equity anomaly strategies in addition to value and price-momentum, including earnings-momentum (i.e., post-earnings-announcement-drift or "PEADs"), net stock issuance, accruals, distress (as proxied by Shumway's (2001) bankruptcy hazard rate), idiosyncratic volatility, and asset growth. Many of the patterns we describe below for value and price-momentum also hold for these other anomalies. We provide a brief overview of these results in the Internet Appendix.

### *C. Trends in Value and Momentum Capital from 1992-2010*

We now turn to the time series of our estimated capital intensities, which are estimated from annual cross-sectional regressions. In the Internet Appendix, we plot and discuss the cross-sectional $R^2$, number of observations, and the coefficients on the additional control variables.

Figure 4 plots the coefficients for the lowest *B/M* and momentum deciles along with the associated 95% confidence intervals. Figure 4 shows that the coefficient for *B/M* decile 1 is significant in each year of our 1992-2010 sample, while the coefficient for momentum decile 1 is significant in all but one year. Consistent with anecdotal evidence, the figure suggests that large

---

[11] Consistent with our cross-sectional interpretation, virtually all of the identification in these short panels is from between- as opposed to within-firm variation. Thus, the resulting annual coefficients are indistinguishable from 12-month averages of the coefficients from monthly cross-sectional regressions (e.g., the correlations exceed 0.99).

quantities of arbitrage capital have flowed into value and momentum strategies, particularly since 2001. Specifically, Figure 4 shows that there has been a steady increase in short interest for extreme growth stocks and momentum losers. Regressing $\hat{\delta}_t^{B/M(1)}$ on a time trend reveals a trend of +8.0 bps per annum ($t$ = 7.3). While the trend for $\hat{\delta}_t^{MOM(1)}$ is slightly smaller at +3.6 bps per annum ($t$ = 3.9), both $\hat{\delta}_t^{B/M(1)}$ and $\hat{\delta}_t^{MOM(1)}$ have tripled over the past 19 years.

The relative magnitudes of our capital proxies are also interesting. We find that $\hat{\delta}_t^{B/M(1)}$ is greater than $\hat{\delta}_t^{MOM(1)}$ in each year of our sample. This suggests that more short-side capital has been allocated to value strategies than to momentum strategies. Value strategies have a longer history among both practitioners and academics than momentum strategies (e.g., dating back to Graham and Dodd (1934)) and are use by a variety of sophisticated investors other than quantitative hedge funds. Thus, it is not surprising to find more short-side capital dedicated to value strategies.

Figure 4 also reports estimates based on 3-month rolling windows as discussed above which allow us to examine higher frequency movements in arbitrage capital. For instance, there is a clear decline in the shorting of growth stocks during the tech bust from 2000 to 2001. Short interest for momentum losers is also noticeably more volatile than that for extreme growth stocks. Specifically, Figure 4 shows that short interest for extreme momentum losers spiked during the beginning of the tech bust in 2000 and again in 2004. Short interest for extreme loser stocks reached all at time peak in June 2007, just before the quant meltdown of August 2007. Short interest for extreme losers plummeted following the imposition of the short interest ban in

September 2008, before spiking again in early 2010.[12] The divergence of our value and momentum capital measures during the 2007-2009 financial crisis is also interesting. Specifically, short-side capital devoted to value strategies continued to grow in late 2008 even as short-side momentum capital was retreating.[13]

*D. Why has Short-side Value and Momentum Capital Grown?*

We consider several possible explanations for the long-run trends we observe in our short-side capital proxies. First, the shifts we observe could be driven by an expansion of share lending supply. Second, shorting demand may have risen due to a rise in the expected returns to the value and momentum strategies. Third, shorting demand may have grown due to a slow diffusion of information about the profitability of these strategies. Such a diffusion of information would have lead additional arbitrageurs to enter these strategies, potentially resulting in heightened competition or strategy crowding over time.

*D.1. Disentangling Supply and Demand Shifts*

We first consider explanations that hinge on an expansion of share lending supply over time. Specifically, it could be that the demand to short a particular set of stocks has always been high, but has only gradually been revealed in equilibrium short interest quantities as supply constraints have relaxed over time. However, we argue that shifts in shorting demand have played an important, though perhaps not exclusive, role in shaping the trends described above.

Since we do not have data on the relevant prices (i.e., share lending fees), we cannot separate supply and demand shifts using the approach of Cohen, Diether, and Malloy (2008).

[12] As discussed in the Internet Appendix, short interest for extreme losers falls for both nonfinancial and financial stocks post September 2008. However, the decline for financials is far more pronounced than that for nonfinancials.

[13] This might either be because investors recognized that momentum strategies are often not very profitable in volatile bear markets (Daniel (2011)) or because momentum strategies were employed by a set of institutions who were more heavily exposed to funding disruptions during the crisis (e.g., highly levered hedge funds).

Instead, we use institutional ownership, $IO_{it}$, as a proxy for the shortable supply of stock *i* at time *t*. Consider the stylized view of the equity lending market is depicted in Figure 5. D'Avolio (2002) suggests that shorting supply curves are kinked; being highly elastic for $SR_{it} < c_{it} \cdot IO_{it}$ and inelastic beyond that kink. Here $c_{it} \in [0,1]$ represents the fraction of institutional owners with active share lending programs. If $SR_{it} < c_{it} \cdot IO_{it}$ so shorting supply is highly elastic, the stock is considered "general collateral" and the lending fee will typically be quite small. If $SR_{it} > c_{it} \cdot IO_{it}$ and short sales constraints bind, the stock is said to be on "special" and short-sellers wishing to borrow shares will have to pay a larger fee. The figure suggests that short interest in stocks with high institutional ownership is unlikely to be affected by loosening supply constraints. For such stocks, it is likely that $SR_{it} < c_{it} \cdot IO_{it}$, so outward shifts in the kink or changes in the cost of shorting constrained stocks will not affect equilibrium short interest quantities.[14]

This analysis leads us to an important prediction of the supply-driven explanation. If shorting demand has been relatively constant, we should not see significant time trends once we condition on institutional ownership. Instead, we should simply find that the time series for unconstrained stocks lies above that for constrained stocks. Under a pure supply shift hypothesis, our aggregate trend simply reflects a changing mix of these two flat lines. By contrast, if there have been important shifts in shorting demand, we would expect to see trends for both the unconstrained and constrained stocks. A time trend for the unconstrained (i.e., high *IO*) stocks would be especially suggestive of an outward shift in shorting demand.

Since shifts in lending supply are likely to be most important for small stocks, we look

[14] However, the fee for shorting unconstrained stocks may have dropped (i.e., the horizontal segment of the supply curve may have shifted). Shifts in the general collateral lending fee will affect equilibrium short interest for unconstrained stocks and are thus harder to disentangle from demand shifts. Anecdotally, general collateral lending fees have been relatively constant over time. Furthermore, such shifts will only have large effects on equilibrium *SR* if shorting demand curves are extremely price elastic which seems unlikely.

for evidence supporting these predictions. We group stocks into small (NYSE size deciles 1-2), medium (deciles 3-5), and big (deciles 6-10) stocks ($me \in \{S,M,B\}$). To explore the importance of the supply shifts, we use a fixed ownership cut-off of 30%, so stocks with $IO_{it}$ < 30% are considered to have "low" institutional ownership. This 30% cut-off is close to the median institutional ownership across all observations in our sample of 35.3%.[15] Thus, we have a set of 6= 3×2 size by *IO* bins in each cross-section. For each cross-section, we run our baseline specification, allowing each of the 6 size by *IO* bins to have its own intercept ($\alpha_{t(me,io)}$) and its own coefficients on the *B/M* and momentum quintile dummies ($\boldsymbol{\delta}_{t(me,io)}^{B/M}$ and $\boldsymbol{\delta}_{t(me,io)}^{MOM}$).

Figure 6 plots the time series of coefficients for the lowest *B/M* and momentum quintiles for small stocks, broken out by high and low institutional ownership. These coefficients show the boost in *SR* relative to stocks of similar size and similar institutional ownership. Shorting of small growth stocks with low *IO* has increased, consistent with the idea that short-sales constraints have eased for this group. Furthermore, there is essentially no increase in short interest for small loser stocks with low institutional ownership. One possible explanation for this contrast with value is that the high turnover rates of momentum strategies makes them less profitable in small stocks with low *IO*. The key to our argument are the results for small-caps with high institutional ownership. Here we find large increases in short interest for both growth and loser stocks. Since these high-*IO* stocks have likely always been unconstrained, Figure 10 suggests that there have large increases in shorting demand for small growth and loser stocks.

*D.2. Changing Expected Returns*

What explains these shifts in shorting demand? One possibility is that the expected returns to value and momentum strategies have increased. Arbitrageurs would naturally respond

---

[15] We obtain similar results if the *IO* cut-off for each period is based on the cross-sectional median.

to such increases in expected returns by allocating more capital to the strategies. Empirically, however, there is little evidence of a secular increase in the expected returns associated with value and momentum strategies. If anything, there is weak evidence of a secular decline in expected returns and a secular rise in the volatility of these strategies.

*D.3. Information Diffusion and Increased Competition*

Another explanation for the low-frequency trends we observe in value and momentum capital is that information on the profitability of these strategies has diffused slowly over time. An emerging literature in finance, including Duffie and Manso (2007), Stein (2008), and Duffie, Malamud, and Manso (2009), studies the diffusion of value-relevant information among competitive agents. These papers show that under certain conditions competitors have incentives to truthfully share information, so that the number of informed agents increases over time.[16] In our setting, this suggests that the degree of competition among arbitrageurs may be increasing over time, as the pool of informed capital has grown. And, interestingly, we find some evidence consistent with concerns about heightened strategy-level competition or crowding.

*D.3.1. Shifting Out of Large and into Small Stocks*

If diminished strategy profitability or strategy crowding are a greater concern in large, liquid stocks, then smaller stocks might have become more appealing over time. As a result, arbitrageurs may shift out of large stocks and into smaller stocks as competition increases. For both value and momentum, we find that arbitrageurs have moved away from large stocks and into small stocks. This is consistent with the idea that increases in competition have been

[16] In Stein's (2008) model an investor might choose to share information on a new trading strategy with a competitor because he hopes that the competitor will be able to further refine the strategy. Furthermore, Stein argues that there may be a tendency for broad or underdeveloped ideas to diffuse more widely than specific and well developed ideas. For instance, researchers affiliated with quantitative equity funds often publish articles outlining the general contours of trading strategies (e.g., "beta arbitrage is profitable," "momentum and value are everywhere," "don't fight the fed," "there is an interaction between value and momentum," etc.), but rarely publish on the specific implementation of these ideas (e.g., exactly how they construct a given signal.).

particularly significant among large, informationally efficient stocks.

To understand these size interactions, we break each of the *B/M* decile into size groups: small stocks (NYSE size deciles 1-2), medium stocks (deciles 3-5), and big stocks (deciles 6-10). In Figure 7, we plot the size interactions associated with extreme growth stocks (*B/M* decile 1). These coefficients represent the boost in *SR* associated with being an extreme growth stock relative to a value-neutral (decile 5) stock in the same size category.

As expected, Figure 7 reveals a steady increase in short interest for small-and medium cap growth stocks. A regression of $\hat{\delta}_t^{B/M(1),SMALL}$ on time yields a trend of +10.5 bps per annum ($t$ = 7.2). The trend for medium stocks is similar at +9.9 bps per annum ($t$ = 4.2). The patterns for large growth stocks are perhaps most interesting. Specifically, we see that large growth stocks were actively shorted in the early 1990s. This activity began to decline in 1995 and during the fallout from the tech bubble there was actually less short interest for large-cap growth stocks than large-caps in *B/M* decile 5. Shorting activity among large-cap growth stocks has staged a modest rebound in recent years and is again comparable to levels in the early 1990s.

We find similar patterns for extreme momentum losers. Specifically, we see increased shorting of small losers since the mid-1990s. The trend for $\hat{\delta}_t^{MOM(1),SMALL}$ is +6.0 bps per annum ($t$ = 6.3). Large-cap loser stocks were actively shorted in mid-1990s, but this has not been the case since 2000. Specifically, $\hat{\delta}_t^{MOM(1),BIG}$ rises from zero during 1992 to a level of approximately 1% from 1995-1998 before falling off in the late 1990s. Evidently, arbitrageurs were fairly reluctant to short large-cap losers during 2001-2003 and 2007-2009.

*D.3.2. Trading on Lower Quality Signals*

Similarly, as more capital uses a reliable signal of mispricing, arbitrageurs may become concerned about either the capacity or profitability of the signal. As a result, they may begin using a signal that is less "crowded" but has historically been a weaker indicator of mispricing. Consistent with this intuition, we find that arbitrageurs have begun trading on "lower quality" signals. Specifically, short interest for firms in *B/M* or momentum decile 2 has increased substantially in recent years. This is consistent with the idea that arbitrageurs facing increased competition are concerned about the capacity or profitability of previously "high quality" signals such as being in *B/M* or momentum decile 1. Such concerns then lead them to use historically less profitable signals such as being in *B/M* or momentum decile 2.

In Figure 8 we plot the coefficients for *B/M* decile 2, $\hat{\delta}_t^{B/M(2)}$. While $\hat{\delta}_t^{B/M(2)}$ is significant in all but 4 years in our sample, we see a steady trend toward more aggressive shorting of decile 2 growth stocks over our sample. A simple regression of $\hat{\delta}_t^{B/M(2)}$ on a time trend reveals a trend of +4.1 bps per annum ($t$ = 6.0). Turning to momentum decile 2, we again see statistically and economically significant shorting of these stocks during most years in our sample. However, the overall time trend is less apparent. While a simple regression of $\hat{\delta}_t^{MOM(2)}$ yields a trends of +1.1 bps ($t$ = 2.4), Figure 8 shows that short interest for decile 2 losers has been more uneven.[17]

Another way to examine these issues is to plot the full set of decile dummies for momentum and *B/M*. These plots reveal time variation in the full mapping from characteristics to short-interest. Examining short interest for stocks that a strategy recommends going long is useful since the stocks that investors choose not to short are also potentially informative about allocation of arbitrage capital across strategies. For example, if the amount of short-side arbitrage capital playing momentum increases relative to other strategies, we would expect to see less shorting of momentum winners (i.e., deciles 9 and 10).

---

[17] Furthermore, the *ratio* of to has risen over time. However, there is little trend in .

We turn first to momentum. In Figure 9 we plot the full set of momentum decile dummies, $\{\hat{\delta}_t^{MOM(d)}\}_{d=1}^{10}$ for the odd years in our sample. As noted above, there has been a significant increase in short interest for extreme losers (i.e., momentum decile 1) during our sample. However, the most striking feature of Figure 9 is the large reduction in short interest among winners in recent years. A possible interpretation of this reduction is that short-sellers have become increasingly concerned about price pressure from long-side momentum investors. Interestingly, however, the reluctance to short past winners appears to have reversed in 2009.

In Figure 10 we plot the full cross-sectional relationship between *B/M* and short interest for the odd years in our sample. In examining this relationship, it is informative to examine whether value investors use a raw *B/M* signal or an industry-adjusted *B/M* signal. Specifically, we augment our baseline specification which already includes a full set of raw *B/M* decile dummies by adding a set of industry-adjusted *B/M* decile dummies (each firm's *B/M* is demeaned using the 8-quarter trailing average *B/M* of its Fama-French-48 industry).

The increase in short interest among stocks in unadjusted *B/M* deciles 1 through 3 is readily apparent in Figure 10. For the first dozen years of our sample, industry-adjusted *B/M* had no impact on short interest after controlling for unadjusted *B/M*. However, since 2003 we see an increased tendency to short stocks that have low industry-adjusted *B/M*, suggesting the industry-adjusted signal has become significantly more popular.

What caused this shift toward industry-adjusted value strategies? Industry-adjusted *B/M* strategies typically have higher Sharpe ratios than unadjusted *B/M* strategies because industry-adjusted strategies function as within-sector allocation rules, while unadjusted strategies may also place large cross-sector bets. The strong performance of the intra-industry *B/M* strategy was first noted by Asness and Stevens (1995) and Cohen and Polk (1996) and may have taken several years to diffuse throughout the quantitative investment community. Furthermore, industry-adjusted strategies outperformed unadjusted strategies during the growth of the tech bubble from 1998-1999, but then underperformed unadjusted strategies during the subsequent bust. Thus, quantitative investors may have shifted into industry-adjusted value strategies in response to their lower volatility and outperformance during the tech bubble.

## IV. Arbitrage Capital and Asset Prices

The low frequency growth of arbitrage capital suggests that the returns to the value and momentum strategies may be competed away over time. However, a large theoretical literature suggests that there may be limits of arbitrage that allow abnormal returns to persist. To explore this possibility, we now turn our attention to the high-frequency feedback between our measures of arbitrage capital and returns. We first examine the effects of past returns and volatilities on changes in arbitrage capital. We then turn to the relationship between arbitrage capital and future returns.

We use $\delta_t^k$ to denote the coefficient on the decile 1 dummy for strategy *k* from the cross-sectional short interest regression at time *t*. We work with quarterly data, so the $\delta_t^k$ are estimated by running regression (2) where all monthly observations in a given quarter are pooled together in a single panel. Using coefficients from monthly cross-sectional regressions introduces greater noise into the $\delta_t^k$ measures, but yields similar results. We use quarterly changes in these coefficients, $\Delta\delta_t^k = \delta_t^k - \delta_{t-1}^k$, to proxy for strategy-level capital flows. Our $\delta_t^k$ and $\Delta\delta_t^k$ measures have units of basis points of short interest.

We use the *HML* and *UMD* factors returns available from Ken French's web-site to proxy for the returns to value and momentum strategies, respectively. We cumulate the monthly returns to form quarterly and annual factor returns. We also compute 1-quarter rolling factors volatilities, $\sigma_t^k$, as the standard deviation of daily factor returns during quarter *t*. The quarterly and annual returns are in percentages and our factor volatility measures are in annualized percentages. To proxy for the returns to hedge funds more generally we use the return indices for "Equity Hedge" and "Event Driven" hedge funds available from Hedge Fund Research (HFR).

We present the results both for our entire sample period and the period 1992-2007, which excludes the recent financial crisis. While the crisis was a period when arbitrage constraints may have bound tightly, the short sales bans and withdrawal of share supply due to concerns about the re-investment portfolios of securities lenders led to wild fluctuations in short interest. Thus, we also present results for the pre-crisis period to understand how these outlying observations may affect the results.

### *A. Determinants of Capital Flows*

We set the stage with some simple plots that show that there is a strong relationship between our capital measures and strategy returns and volatility. Figure 11 plots the 4-quarter moving average of our capital measures (the $\delta_t^k$ coefficients) versus annual strategy returns and realized volatilities over the same 4-quarter period. A few noticeable relationships stand out. First, there is a strong negative correlation between the level of *B/M* capital and past *HML* returns ($\rho$ = –0.27). Second, there is a strong positive correlation between the level of momentum capital and past *UMD* returns before 2001 ($\rho$ = 0.62). However, the correlation is more modest in the latter half of the sample ($\rho$ = 0.26). Third, there is also a strong negative relationship between the level of *B/M* capital and past *HML* return volatility ($\rho$ = –0.47) prior to 2008. The relationship between capital and realized volatility for momentum is less apparent over this period. However, from 2008-2010 there was a strong inverse relationship between momentum capital and realized *UMD* volatility.

### *A.1. Effects of Past Strategy Returns on Capital Flows*

A critical assumption of much of the literature on limits of arbitrage, beginning with Shleifer and Vishny (1997), is the existence of a performance-flow relationship.[18] If arbitrageurs suffer outflows after their trades move against them, then they may exacerbate the very mispricing they set out to arbitrage. In equilibrium, this may discourage arbitrageurs from attempting to arbitrage the mispricing ex ante.

Existing empirical work has focused on individual mutual funds and hedge funds and found such performance-flow relationships at the individual fund level.[19] A handful of papers have also found evidence of a positive performance flow relationship at the level of the aggregate mutual fund or hedge fund industry.[20] Why might we expect a performance-flow relationship to exist at the strategy level? First, fund managers may themselves chase performance across strategies. Second, a fund-level performance-flow may naturally lead to a strategy-level performance-flow relationship if end investors chase performance across funds that mix strategies in different proportions.

In Table 1, we regress capital flows in quarter $t$ on strategy returns in quarter $t$–1:

$$\Delta\delta_t^k = \alpha^k + \beta^k \cdot r_{t-1}^k + \varepsilon_t^k. \tag{3}$$

The $t$-statistics are computed using heteroskedasticty robust standard errors. There is little

---

[18] While Shleifer and Vishny (1997) take the performance-flow relationship as given, Barberis and Shleifer (2003) and Berk and Green (2004) micro-found it using as the result of performance chasing and rational updating about fund manager ability, respectively.

[19] Chevalier and Ellison (1997) and Sirri and Tufano (1998) find a convex performance-flow relationship for mutual funds. Ding, Liang, Gemansky, and Wermers (2009) find that the flow-performance relationship for hedge funds is also convex in the absence of share restrictions, but that the relation becomes concave in the presence of restrictions.

[20] Goetzmann and Massa (2003) find evidence of daily performance flow relationship using U.S. index funds. Specifically, outflows increase following down-market days. Wang and Zheng (2008) find a positive relation between quarterly aggregate hedge fund flows and past aggregate hedge fund returns using Lipper TASS data.

evidence of a quarterly performance-flow relationship for value strategies. In fact, the point estimate for $\beta^k$ is slightly negative. By contrast, we find a reliably positive performance-flow relationship for momentum at a quarterly frequency in the pre-2008 period. The magnitudes here seem reasonable. The estimates indicate that a 10% quarterly momentum return generates capital flows of 8.8 bps; the mean and standard deviation of quarterly momentum flows are 0.3 and 31 bps respectively.

*A.2. Effects of Strategy Volatility and Funding Constraints on Capital Flows*

Even in the absence of a performance-flow relationship, arbitrage may be limited if the leverage supplied to arbitrageurs is a function of past return volatility. For instance, in Brunnermeier and Pedersen (2009) a rise in volatility leads risk-averse lenders to raise margins on both long and short positions. As a result, leveraged arbitrageurs with limited capital are forced to scale back both long and short positions in order to meet margin requirements, potentially further raising volatility and margins in a "margin spiral."[21]

Arbitrageur positions are also decreasing in margins in Garleanu and Pedersen (2011). Furthermore, they argue that the difference between uncollateralized and collateralized short-term interest rates is a good proxy for the tightness of arbitrageurs' margins constraints.[22] Thus, we investigate whether short-side capital declines following a tightening of funding constraints, proxied using changes in the Treasury Eurodollar (TED) spread as in Frazzini and Pedersen (2010). Of course, without detailed micro-data on hedge fund leverage and margins, we cannot

[21] Since strategy-level volatility is persistent, standard mean-variance considerations would predict a similar relationship between volatility and capital. For instance, the quarterly auto-correlations of *HML* and *UMD* volatility realizations are 0.77 and 0.79, respectively, in our 1992-2010 sample. Thus, a rational arbitrageur would forecast high future volatility for strategy *k* if past volatility has been high and, if he has a short performance horizon, this would lead him to reduce his allocation to strategy *k*.

[22] This is because uncollateralized borrowing effectively relaxes an arbitrageur's margins constraint which collateralized borrowing does not. Thus, the difference between the uncollateralized rate and the collateralized rate equal the multiplier on the arbitrageur's margins constraint.

distinguish between these various mechanisms, but we can verify their common prediction.

Table 2 considers the effect of changes in strategy return volatility on strategy capital:

$$\Delta\delta_t^k = \alpha^k + \psi^k \cdot \Delta\sigma_{t-1}^k + \varepsilon_t^k. \tag{4}$$

There is evidence of the predicted negative relationship between capital and volatility for value. Regressing our capital flow measure on lagged changes in 1-quarter *HML* volatility yields a negative and significant coefficient. Again, the magnitudes seem reasonable. A 10% spike in annual *HML* volatility is associated with a 27 bps decline in our capital measure. For reference, the mean and standard deviation of the value flow measure are 1.8 bps and 24 bps respectively.

For momentum, there is no evidence of the hypothesized negative relationship. This may be due to the fact that realized 1-quarter volatility fluctuates more for momentum than for value. Changes in 1-quarter momentum volatility have a standard deviation that is 65% higher than that for changes in 1-quarter value volatility. However, there does appear to be a negative relationship between momentum capital and overall market volatility, which is driven by the financial crisis period.

When we examine the relationship between strategy capital and funding constraints, proxied by the Treasury Eurodollar (TED) spread, we find a strong negative relationship for momentum, but no relationship for value.

*A.3. Contagion and Spillovers Across Strategies*

We next investigate the effects of the returns on other strategies on the capital in a given strategy. This allows us to quantify the extent of wealth contagion or deleveraging spillovers across strategies. For instance, suppose there are two strategies, *A* and *B*, and that there is an initial adverse shock to the returns of strategy *A*. In the regression,

$$\Delta\delta_t^B = \alpha + \gamma \cdot r_{t-1}^A + \beta \cdot r_{t-1}^B + \varepsilon_t^B, \tag{5}$$

the coefficient $\gamma$ captures the effects of strategy-*A* returns on strategy-*B* capital flows. Limits-to-arbitrage or deleveraging stories would suggest that $\gamma$ will be large and positive when many arbitrageurs play both strategies *A* and *B*. By contrast, if strategies *A* and *B* are used by entirely distinct sets of arbitrageurs, $\gamma$ would be close to zero.

In Table 3 we regress $\Delta\delta_t^{MOM}$ on lagged of its own factor return (i.e., *UMD*), the market return, and HFR Indices tracking the performance of Event Driven and Equity Hedge (long/short) hedge funds. Hedge fund returns, rather than individual returns to other strategies, are likely to be the most powerful indicators of contagion because they "correctly" weight returns to other strategies. The table shows strong evidence that momentum capital flows, $\Delta\delta_t^{MOM}$, respond to several other factor returns in addition to $UMD_{t-1}$.[23] Column 3 shows that, holding $MKTRF_{t-1}$ and $UMD_{t-1}$ fixed, the effect of a one percentage point increase in the lagged return on HFR's Equity Hedge index ($EHEDGE_{t-1}$) is associated with a 4 bps increase in short interest for extreme losers. Thus, a one standard deviation increase in $EHEDGE_{t-1}$ is associated with a 19.5 bps increase in $\delta_t^{MOM}$ (the standard deviation of $\Delta\delta_t^{MOM}$ is 29 bps). In untabulated results, we find that the effect of $EHEDGE_{t-1}$ is nearly three times larger for negative returns than for positive returns. Column 4 finds a similar effect for the returns to HFR's Event Driven Index ($EVENT_{t-1}$). Both of these hedge fund returns are included in column 5, and the magnitudes of the effects fall but remain significant.

These results are suggestive of cross-strategy spillovers. When other equity strategies do

---

[23] We find little evidence that these other return factors help explain . One possibility is that there is a large group of arbitrageurs that only play value which may contrast with momentum. These value investors may have longer horizons and use relatively low leverage, so they are both willing and able to withstand wealth or contagion effects. To our knowledge, there are no pure momentum arbitrageurs. Momentum is a highly volatile strategy and so it is typically paired with other strategies to diversify away some of its idiosyncratic risk.

poorly, it seems that arbitrageurs liquidate momentum positions, presumably to meet margin requirements or capital redemptions. The sensitivity to event-driven returns is particularly suggestive since it is likely that only large multi-strategy hedge funds combine momentum and event-driven arbitrage.

It is interesting to note that when we control for hedge fund returns the coefficient on the market return becomes negative and significant. Holding fixed how hedge funds are doing, capital flows into momentum when the market does poorly. One possibility is that end investors find the purported low-$\beta$ of hedge funds more appealing when the overall market is doing poorly. This fits with anecdotal evidence that large quantities of capital flowed into hedge funds during the tech bust as institutions sought low-$\beta$ alternatives to equities. Alternatively, hedge funds may be trying to inflate their returns by attempting to time the market. When the market does poorly, they stop closet indexing and put their capital back into $\beta$-neutral long-short strategies.

*B. Arbitrageur Capital and Future Strategy Returns*

A third, less direct way to assess theories of limited arbitrage is to examine the relationship between arbitrage capital and future returns to anomaly strategies. If arbitrageurs are unconstrained, then they should increase their strategy capital allocations when they anticipate high returns going forward. If, on the other hand, they are constrained by binding capital or leverage constraints when expected returns are high, the relationship between capital allocations and future returns will be negative.

In Table 4 we forecast strategy returns over the following 4-quarters using capital flows:

$$r^k_{t\rightarrow t+4} = \mu^k + \varphi^k \cdot (\delta^k_t - \delta^k_{t-4}) + \varepsilon^k_{t\rightarrow t+4}, \tag{6}$$

Due to the overlapping returns, the *t*-statistics here are computed using Newey-West (1987) standard errors allowing for 6 lags. There is reliable evidence that recent capital flows negatively forecast future value returns. This result continues to hold even after controlling for the value spread, $VS_t$, and for past returns, $r_{t-4\to t}^{B/M}$, to capture the mean reversion in *HML* identified in Teo and Woo (2004). The effects in the table are economically large. The returns here are in percentage points and the coefficients are in basis points. Thus, the coefficient for $\delta_t^{B/M} - \delta_{t-4}^{B/M}$ implies that a 1 bp increase in our capital measure forecasts that future annual *HML* returns will decline by 0.13%.

This is unlikely to be purely a price pressure effect. An alternate interpretation is that returns to value are mean reverting and capital slowly chases value returns. This means that arbitrage capital wound up mistiming the value returns quite substantially. We examine this possibility in more detail. Regressing $\delta_{t+4}^{B/M} - \delta_t^{B/M}$ on $r_{t-4\to t}^{B/M}$ yields an estimated coefficient of 0.99 ($t$ = 4.67) which suggests a strong low-frequency performance flow relationship. This result is driven by events surrounding the tech bubble and bust. Brunnermeier and Nagel (2004) study this period, finding that most hedge funds rode the tech bubble, while short sellers timed it well.

Our measures reveal a more complicated story. Short sellers began shorting growth stocks during 1999, sustaining massive losses in the last two quarters of that year. While their short positions peaked at the height of the bubble in the first quarter of 2000, they closed out these positions too quickly during 2000. Perhaps capital withdrawals or tightening funding and margin constraints limited the ability of arbitrageurs to maintain these short positions. Whatever the reason, short sellers missed the rebound in value and collapse of growth stocks during late 2000 and 2001. They began aggressively shorting growth stocks again in late 2001 and 2002, after the rally in value and collapse of growth had already taken place.

Table 4 also examines the low-frequency relationship between returns and capital for momentum. As with value, regressing future *UMD* returns on past capital flows, $\delta_t^{MOM} - \delta_{t-4}^{MOM}$

reveals evidence that arbitrageurs have negative ability to time momentum at longer horizons in the pre-2008 period. However, this result disappears during the financial crisis period, as arbitrageurs appear to have successfully exited momentum before it incurred low returns.

## V. Conclusion

We propose a novel methodology for measuring strategy-level capital using time-series variation in the cross section of short interest. We find evidence suggesting that arbitrageurs have reacted to heightened competition by altering their strategies. In particular, they increasingly short moderate growth and moderate loser stocks, which we interpret as a shift towards using historically weaker signals of mispricing due to concerns that stronger signals have become over-crowded. Furthermore, quantitative investors have shifted away from shorting large stocks and into small stocks. We interpret this as a shift away from more informationally efficient, liquid stocks that is also driven by concerns about crowding.

Next we explore the determinants of capital flows into arbitrage strategies. We assess the evidence in favor of phenomena posited by the theoretical literature on the limits-to and destabilizing consequences of arbitrage, including responses to past returns (performance-flow), past volatility, and past returns in other strategies (cross-strategy spillovers). We find strong evidence of a performance-flow relationship for momentum, while capital flows for both value and momentum respond to volatility. For momentum, we also find evidence of spillovers using hedge fund return indices as a proxy for the performance of other arbitrage strategies. We also examine the forecasting power of capital flows for strategy returns and volatilities. For both value and momentum, lagged capital flows have strong negative forecasting power for returns, indicating that arbitrage capital has mistimed returns over our sample.

Our methodology for measuring strategy-level capital may be of independent interest to policymakers interested in detecting "crowded trades" because of the systemic risks they might pose. Existing approaches to detecting time-variation in crowding such as Adrian (2007), Pericoli and Sbracia (2010), and Pojarliev and Levich (2011) analyze changes in correlation structure of *ex post* returns. However, our approach may be better suited to detecting crowding *ex ante* because in relies on changing patterns in arbitrageurs positions.

## Appendix: Data Construction

Our timing conventions ensure all firm characteristics are publicly available as of the date on which short interest is measured. Below we provide detailed definitions for each of the anomaly sort variables used in the paper.

**Value (*B/M*)**: *B/M* deciles are refreshed quarterly, allowing for at least 3 months between the fiscal quarter-end when book equity is measured and the sort date. For instance, short interest observations for July, August, and September are associated with a *B/M* sorts performed at the end of June. These book-to-market ratios are based on market equity as of the end of the prior quarter (March) and on book-equity from fiscal quarters ending in the prior calendar quarter (January, February, or March). This is the quarterly analog of the familiar timing conventions established by Fama and French (1992). Book equity is defined as stockholder's equity, plus balance sheet deferred taxes and investment tax credits (when available), minus the book value of preferred stock.

We also sort firms on the basis of industry-adjusted *B/M* using the 48 Fama-French (1997) industries. Specifically, we subtract the 8-quarter moving average of aggregate industry *B/M* (industry book over industry market value) from each individual firm's book-to-market ratio.

**Price/return momentum**: 12-month return momentum deciles are based on cumulative returns from months *t*–12 to *t*–1. That is, we skip a month when computing past returns to avoid contaminating over measures with the short-term reversal phenomenon documented by Jegadeesh (1990). Momentum deciles are refreshed each month. For instance, short interest observations for July are associated with momentum sorts performed at the end of June. These sorts are based on the 11 month cumulative returns from July (of the previous year) through May.

**Earnings momentum**: For earnings momentum or post-earnings-announcement-drict (i.e. "PEADs") we follow Chan, Jegadeesh, and Titman (1996) and use a standardized unexpected earnings measure based on the seasonal random-walk earnings model. This "earnings surprise" is normalized by share price, $SUE_{it} = (EPS_{it} - EPS_{it-4}) / P_{it}$.

**Share issuance**: Following Fama and French (2008), we compute the year-over-year change in split-adjusted shares from quarterly Compustat data: $NS_{it} = \log[SHR_{it}^{Adj} / SHR_{it-4}^{Adj}]$

where $SHR_{it}^{Adj}$ is the product of common shares outstanding (*CSHOQ*) and Compustat's adjustment factor (*AJEXQ*).

**Accruals**: Balance sheet accruals in quarter *t* are defined as in Sloan (1996):

$$ACC_{it} = (\Delta CurrAssets_{it} - \Delta Cash_{it}) - (\Delta CurrLiab_{it} - \Delta STDebt_{it} - \Delta TaxesPayable_{it}) - Deprec_{it}.$$

Our measure of accruals is just then sum of balance sheet accruals over the past 4 quarters divided by average quarterly assets. The accruals measure and associated decile is refreshed each quarter following the timing conventions discussed above. Following Bergstresser and Philippon (2006), we also compute a cashflow-based measure of quarterly accruals as $EBXI_{it} - CFO_{it}$ where *EBXI* is reported earnings before extraordinary items and *CFO* is cash flows from continuing operations (operating cash-flows minus cash-flows from extraordinary items and discontinued operations).

**CAPM Residual Volatility**: $\sigma_{it}(\varepsilon)$ is the residual volatility from a trailing 24-month CAPM regression. In order to compute $\sigma_{it}(\varepsilon)$ we require that a firm has valid returns for at least 12 of the past 24 months.

**Distress**: We use the bankruptcy hazard rate estimated by Shumway (2001). The hazard model estimated by Shumway is $H_{it} = \exp[SHUM_{it}]/(1+\exp[SHUM_{it}])$ where

$$SHUM_{it} = -13.303 - 1.982 \cdot (NI/A)_{it} + 3.593 \cdot (L/A)_{it} - 0.467 \cdot RELSIZE_{it} - 1.809 \cdot (R_{it} - R_{Mt}) + 5.791 \cdot \sigma_{it}$$

*NI/A* is 4-quarter trailing net income over period-end total assets, *L/A* is total liabilities over total assets, *RELSIZE* is the log of a firm's market equity divided by the total capitalization of all NYSE and AMEX stocks, $R_{it} - R_{Mt}$ is firm's cumulative return over the prior 12-months minus the cumulative return on the value-weighted CRSP NYSE/AMEX index, and $\sigma_{it}$ is volatility of residuals from trailing 12-month market-model regression (treating the CRSP NYSE/AMEX index as the market return). This distress measured is refreshed each quarter.

**Asset Growth**: Following Fama and French (2008), we also compute measures of gross and net asset. Gross asset growth is simply the percentage change in assets over previous 4 quarters. Net asset growth is asset growth per split-adjusted share. Daniel and Titman (2006) and Fama and French (2008) argue that the forecasting ability of gross asset growth measures is driven largely by the net share issuance component as opposed to net growth component. These growth measures are refreshed each quarter.

## References


Adrian, Tobias (2007), "Measuring risk in the hedge fund sector," *Federal Reserve Bank of New York Current Issues in Economics and Finance*, Volume 13, Number 3, March/April 2007

Asness, Clifford and Ross Stevens, (1995), "Intra-Industry and inter-industry factors in the cross-section of expected stock returns, working paper Goldman Sachs Asset Management, Quantitative Research Group.

Aragon, George and Philip Strahan, (2010), "Hedge Funds as Liquidity Providers: Evidence from the Lehman Bankruptcy," working paper.

Barberis, Nicholas and Andrei Shleifer, (2003), "Style Investing," *Journal of Financial Economics* 68: 161-189.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, (2010). "Hedge Fund Stock Trading in the Financial Crisis of 2007-2008," working paper.

Bergstresser, Daniel and Thomas Philippon (2006), "CEO incentives and earnings management," *Journal of Financial Economics* 80: 511--529.

Berk, Jonathan B. and Richard C. Green (2004), "Mutual fund flows and performance in rational markets," *Journal of Political Economy* 112: 1269-1295.

Boehmer, Ekkehart, Charles M. Jones, and Xiaoyan Zhang, (2009), "Shackling short sellers: The 2008 shorting ban,"

Brunnermeier, Markus K., Nagel, Stefan (2004), "Hedge funds and the technology bubble," *Journal of Finance* 59: 2013--2040.

Brunnermeier, Markus K., and Lasse H. Pedersen (2009), "Market liquidity and funding liquidity," *Review of Financial Studies* 22(6): 2201-2238.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi (2008), "In search of distress risk," *Journal of Finance* 63: 2899-2938.

Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok (1996), "Momentum strategies," *Journal of Finance* 51: 1681-1713.

Chevalier, Judith, and Glenn Ellison, (1997), "Risk taking by mutual funds as a response to incentives." *Journal of Political Economy* 105: 1167–1200.

Cohen, Lauren, Karl Diether and Christopher Malloy (2007), "Supply and demand shifts in the shorting market," *Journal of Finance* 62: 2061-2096.

Cohen, Randolph and Christopher Polk (1996), "An investigation of the impact of industry factors in asset-pricing tests," working paper.

Cohen, Randolph, Christopher Polk, and Tuomo Vuolteenaho (2003), "The value spread," *Journal of Finance* 58: 609-641.

D'Avolio, Gene (2002), "The market for borrowing stock," *Journal of Financial Economics* 66: 271-306.

Daniel, Kent and Sheridan Titman (2006), "Market reaction to tangible and intangible information," *Journal of Finance* 61: 1605-1643.

Ding, Bill, Mila Getmansky, Bing Liang, and Russ Wermers, (2009), "Share restrictions and investor flows in the hedge fund industry," working paper.

Duffie, Darrell and Gustavo Manso (2007), "Information Percolation in Large Markets," *American Economic Review, Papers and Proceedings* 97: 203-209.

Duffie, Darrell, Semyon Malamud, and Gustavo Manso (2009), "Information Percolation with Equilibrium Search Dynamics," *Econometrica* 77: 1513-1574.

Fama, Eugene F., and Kenneth R. French (1992), "The cross-section of expected stock returns," *Journal of Finance* 47: 427--465.

Fama, Eugene F. and Kenneth R. French (1997), "Industry costs of equity capital," *Journal of Financial Economics* 43: 153-193.

Fama, Eugene F., and Kenneth R. French (2008), "Dissecting Anomalies," *Journal of Finance* 63: 1653-1678.

Frazzini, Andrea and Owen A. Lamont (2008), "Dumb money: mutual fund flows and the cross-section of stock returns," *Journal of Financial Economics* 88(2): 299-322.

Frazzini, Andrea and Lasse Heje Pedersen (2010), "Betting against beta," New York University working paper.

Garleanu, Nicolae and Lasse Heje Pedersen, (2011), "Margin-based asset pricing and deviations from the Law of One Price," forthcoming *Review of Financial Studies*.

Goetzmann, William N., and Massimo Massa, (2003), "Index funds and stock market growth," *Journal of Business*, 76(1), 1-28.

Graham, Benjamin, and David Dodd, (1934), *Security Analysis*, McGraw-Hill.

Gromb, Denis and Dimitri Vayanos, (2010), "Limits of arbitrage: The state of the theory," *Annual Review of Financial Economics* 2, 251-275

Hong, Harrison and Jeremy C. Stein (1999), "A unified theory of underreaction, momentum trading and overreaction in asset markets," *Journal of Finance* 54: 2143-2184.

Jegadeesh, Narasimhan (1990), "Evidence of predictable behavior of security returns," *Journal of Finance* 45: 881-898.

Khandani, Amir E., and Andrew W. Lo (2007), "What happened to the quants in August 2007?", MIT *Journal of Investment Management* 5(4): 29–78.

Khandani, Amir E., and Andrew W. Lo (2008), "What happened to the quants in August 2007? Evidence from factors and transactions data", MIT Sloan School working paper.

Lamont, Owen A. and Jeremy C. Stein (2004), "Aggregate short interest and market valuations," *American Economic Review Papers and Proceedings*, 94, 29-32.

Lewellen, Jonathan W. (2010), "Institutional investors and the limits of arbitrage," forthcoming *Journal of Financial Economics*.

Newey, Whitney K. and Kenneth D. West (1987), "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix," *Econometrica* 55: 703-708.

Pedersen , Lasse Heje (2009), "When everyone runs for the exit," *International Journal of Central Banking* 177-199.

Pericoli, Marcello and Massimo Sbracia (2010), "Crowded trades among hedge funds," Banca d'Italia working paper.

Pojarliev, Momtchil and Richard Levich (2011), "Detecting crowded trades in currency funds," *Financial Analysts Journal* 67:1.

Savor, Pavel, and Mario Gamboa-Cavazos (2011), "Holding on to your shorts: When do short sellers retreat?," University of Pennsylvania working paper.

Shleifer, Andrei and Robert W. Vishny (1997), "The limits of arbitrage," *Journal of Finance* 52: 35-55.

Shumway, Tyler (2001), “Forecasting bankruptcy more accurately: A simple hazard model,” *Journal of Business*, 101-124

Sirri, Erik and Peter Tufano (1998), “Costly search and mutual fund flows,” *Journal of Finance* 53: 1589-1622.

Sloan, Richard G. (1996), “Do stock prices fully reflect information in accruals and cash flows about future earnings?” *The Accounting Review*, 71(3), 289-315.

Stein, Jeremy (2008), “Conversations Among Competitors,” *American Economic Review* 98:5, 2150–2162.

Teo, Melvin and Sung-Jun Woo (2004), “Style effects in the cross-section of stock returns,” *Journal of Financial Economics*, 74(2), 367-398.

Thompson, Samuel B. (2011). “Simple formulas for standard errors that cluster by both firm and time,” *Journal of Financial Economics* 99:1-10.

Wang, Ashley and Lu Zheng, (2008), “Aggregate hedge fund flows and asset returns,” University of California, Irvine working paper.

**Figure 1: Average short interest ratios, 1992-2010.** This figure plots the monthly equal- and value- (i.e., market equity) weighted average short interest ratio for all stocks in our sample. The short interest ratio for stock *i* in month *t* is defined as $SR_{i,t} = SHORT_{i,t} / SHROUT_{i,t}$ where $SHORT_{i,t}$ is short interest as of the mid-month reporting date and $SHROUT_{i,t}$ is shares outstanding as of the reporting date.

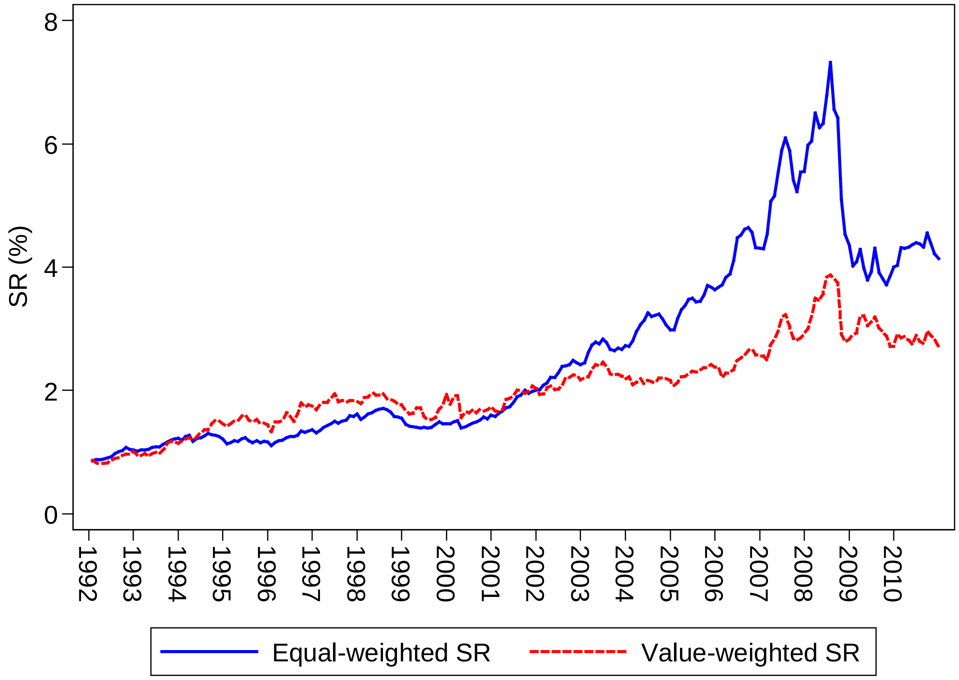



**Figure 2: Average short interest ratios by size decile.** This figure shows the average short interest ratio by NYSE size decile as of year-end 1995, 1999, 2003, 2005, 2007, and 2009.

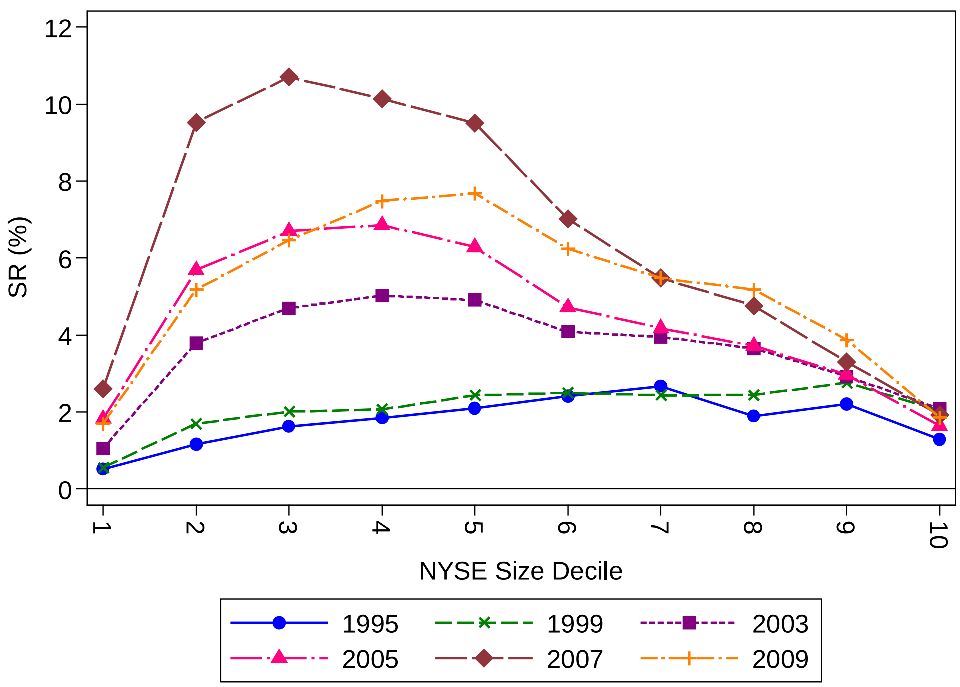
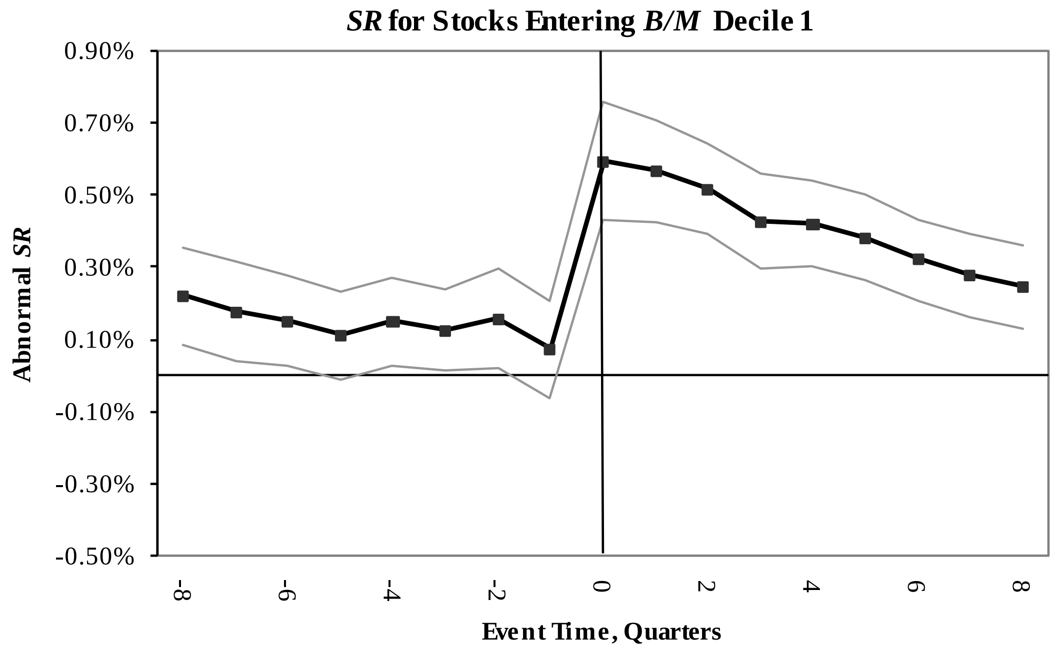


**Figure 3: Short interest for stocks entering the extreme *B/M* (growth) or momentum (loser) deciles.** The figure plots the "event time" coefficients which show the path of short interest for the typical stock entering the extreme growth or momentum deciles. Specifically, Panel A plots the $\delta^k$ for $k$ = -8, …,-1,0,+1,..,+8 obtained from estimating:

$$SR_{it} = \delta^{-8} 1_{it}^{-8}\{B/M\} + \cdots + \delta^{0} 1_{it}^{0}\{B/M\} + \cdots + \delta^{+8} 1_{it}^{+8}\{B/M\}$$
$$+\boldsymbol{\delta}^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\delta}^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\beta}'\mathbf{x}_{it} + Stock_i + Time_t + \varepsilon_{it}.$$

Panel B repeats this for the analogous specification for stocks entering the extreme momentum (i.e., past "loser") decile.

Panel A: *SR* for stocks entering *B/M* decile 1 (i.e., extreme growth stocks)




Panel B: *SR* for stocks entering momentum decile 1 (i.e., past return "losers")

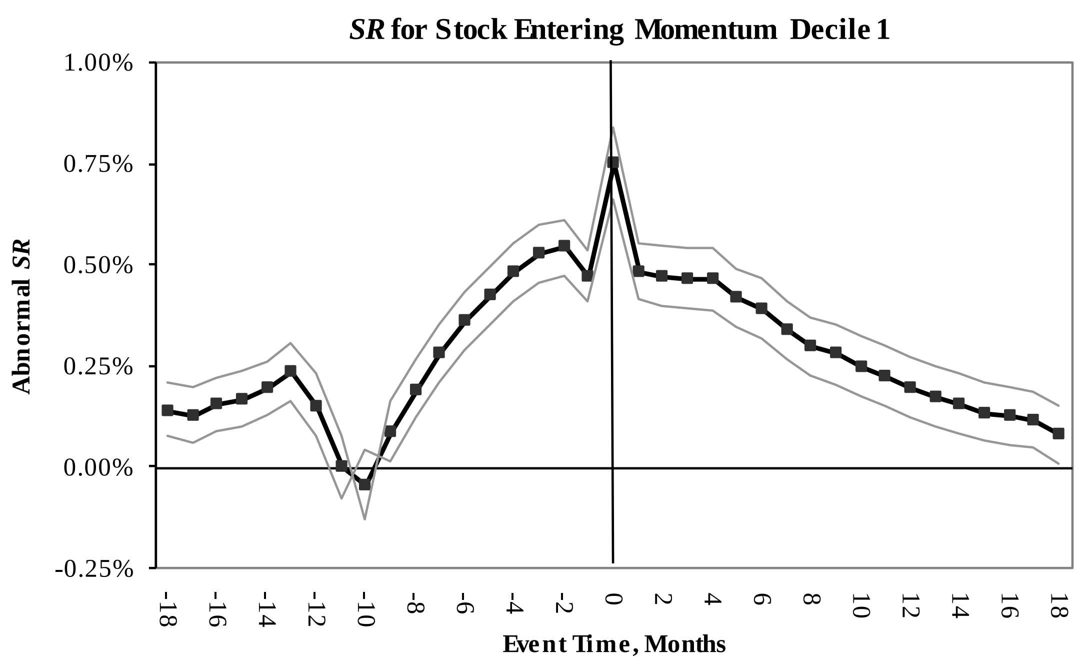


**Figure 4: Estimated capital intensities for value and momentum strategies.** The figure plots the time series of estimated coefficients on the extreme growth decile (        ) and extreme momentum loser decile (             from the following specification:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$

In Panel A, these regressions are estimated annually, pooling all observations in a given year. In Panel B, these regressions are estimating on a rolling quarterly basis, pooling all observations in a given 3 month period. Thus, both specifications also include a full set of month fixed effects. We compute confidence intervals for the estimated coefficients using standard errors that cluster by firm and, thus, are robust to serial correlation at the firm level.

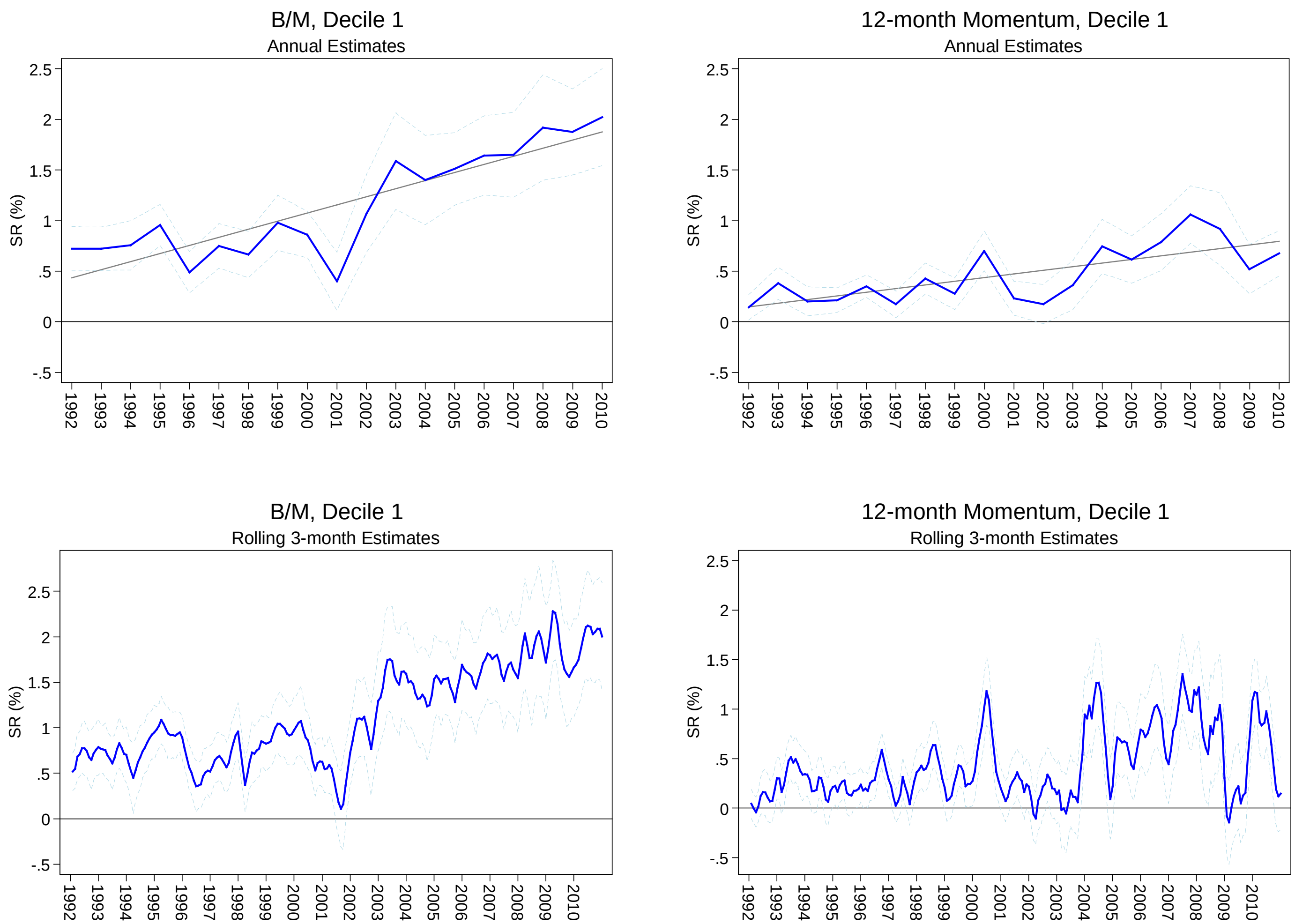


**Figure 5: Stylized depiction of the equity lending market.** The figure shows shorting demand and share lending supply in (*SR*,*Lending Fee*) space. Following D'Avolio (2002), we assume that shorting supply curves are kinked; being highly elastic for and inelastic beyond that kink. Here represents the fraction of institutional owners with active share lending programs. If so shorting supply is highly elastic, the stock is considered "general collateral" and the lending fee will typically be quite small. If and short sales constraints bind, the stock is said to be on "special" and short-sellers wishing to borrow shares will have to pay a larger fee. The figure shows the effect of an outward shift in both demand and supply for a stock that initially has a high level of institutional ownership. The figure suggests that short interest in stocks with high institutional ownership is unlikely to be affected by loosening supply constraints. For such stocks, it is likely that , so outward shifts in the kink or changes in the cost of shorting constrained stocks will not affect equilibrium short interest quantities.

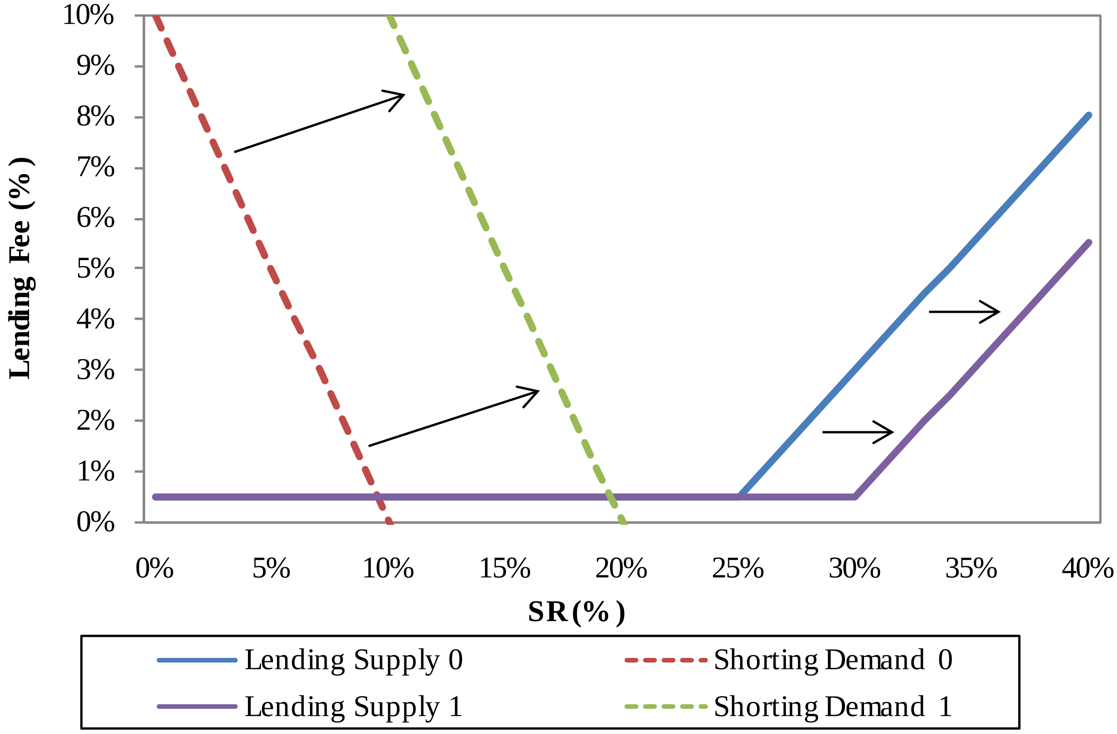
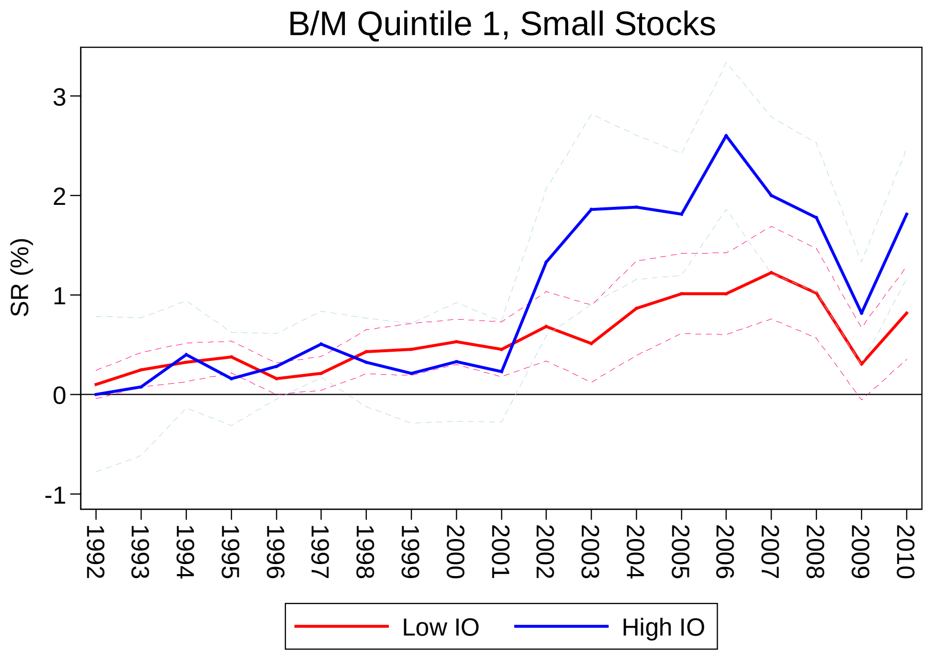


**Figure 6: Capital intensities for small stocks by institutional ownership.** The figure plots the time series of estimated coefficients on the extreme growth quintile and momentum quintile, allowing for separate effects by size and institutional ownership group:

$$SR_{it} = \sum_{io \in \{L,H\}} \sum_{me \in \{S,M,B\}} (\alpha_{tt(me,io)} + \boldsymbol{\delta}^{B/M}_{t(me,io)} \cdot \mathbf{1}^{B/M}_{it} + \boldsymbol{\delta}^{MOM}_{t(me,io)} \cdot \mathbf{1}^{B/M}_{it}) \times 1\{ME_{it} \in me\} \times 1\{IO_{it} \in io\} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$



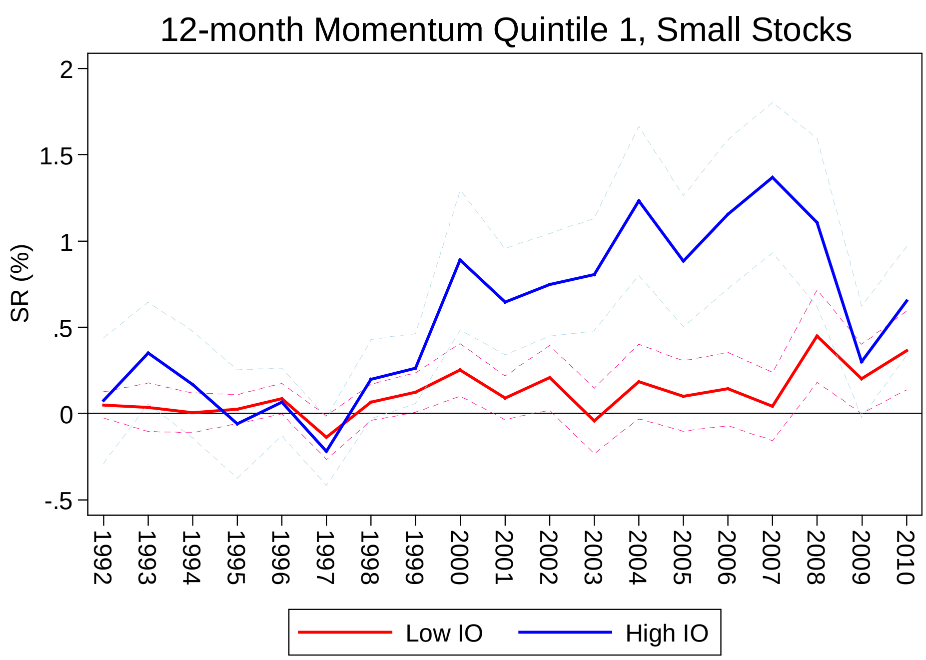
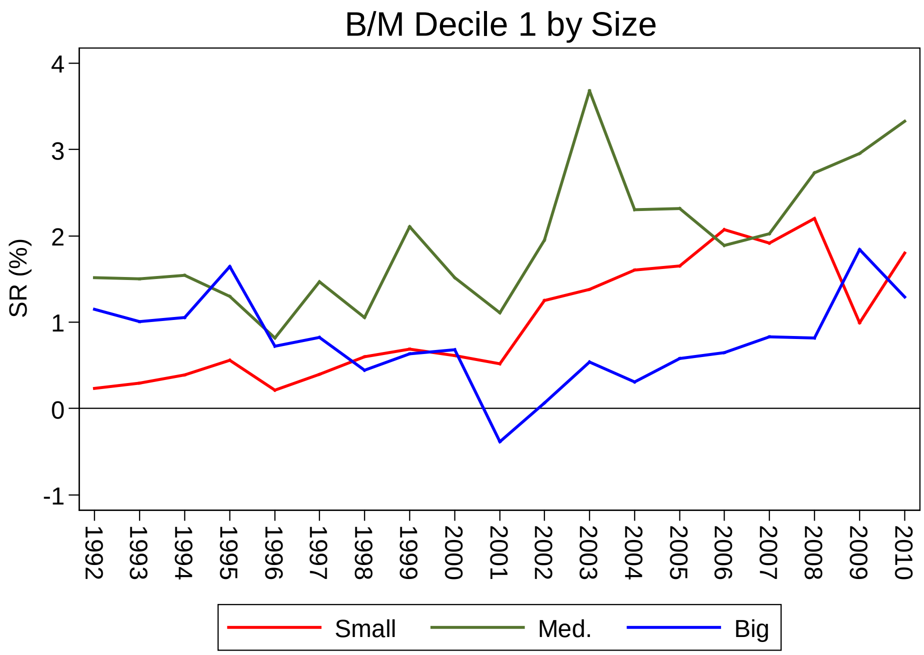



**Figure 7: Size interactions.** The figure plots the time series of estimated coefficients on the extreme growth decile and momentum decile, allowing for separate effects by size group:

$$\begin{aligned} SR_{it} = \alpha_t &+ \boldsymbol{\delta}_t^{B/M,SMALL} \cdot (\mathbf{1}_{it}^{B/M} \times 1_{it}^{SMALL}) + \boldsymbol{\delta}_t^{B/M,MED} \cdot (\mathbf{1}_{it}^{B/M} \times 1_{it}^{MED}) + \boldsymbol{\delta}_t^{B/M,BIG} \cdot (\mathbf{1}_{it}^{B/M} \times 1_{it}^{BIG}) \\ &+ \boldsymbol{\delta}_t^{MOM,SMALL} \cdot (\mathbf{1}_{it}^{MOM} \times 1_{it}^{SMALL}) + \boldsymbol{\delta}_t^{MOM,MED} \cdot (\mathbf{1}_{it}^{MOM} \times 1_{it}^{MED}) + \boldsymbol{\delta}_t^{MOM,BIG} \cdot (\mathbf{1}_{it}^{MOM} \times 1_{it}^{BIG}) \\ &+ \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}. \end{aligned}$$

These regressions are estimated annually.



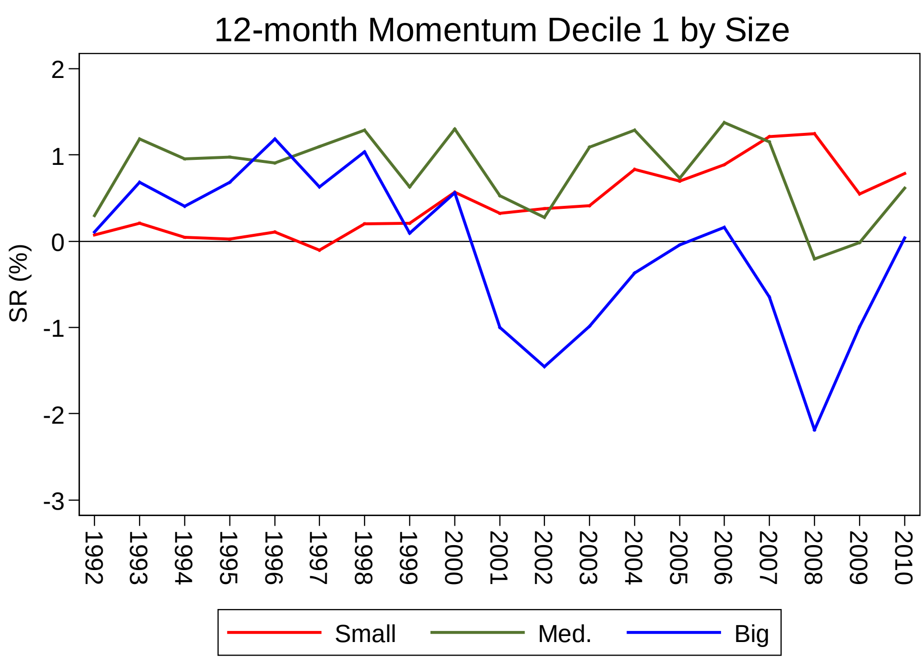
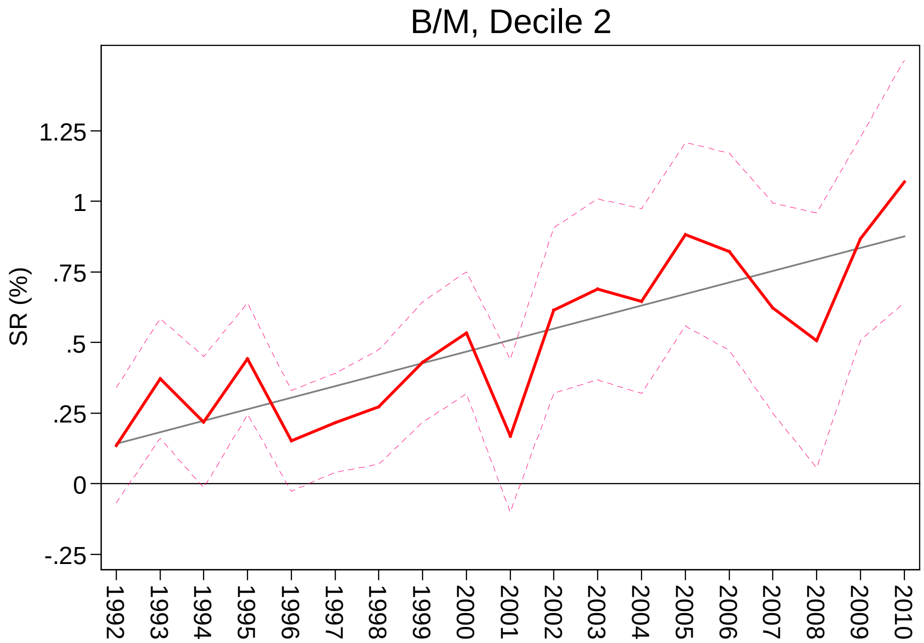



**Figure 8: Trading on lower quality signals.** The figure plots the time series of estimated coefficients on the growth decile 2 (        ) and momentum decile 2 (            from the following specification:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$

These regressions are estimated annually, pooling all observations in a given year, and include a full set of month fixed effects. We compute confidence intervals for the estimated coefficients using standard errors that cluster by firm and, thus, are robust to serial correlation at the firm level.



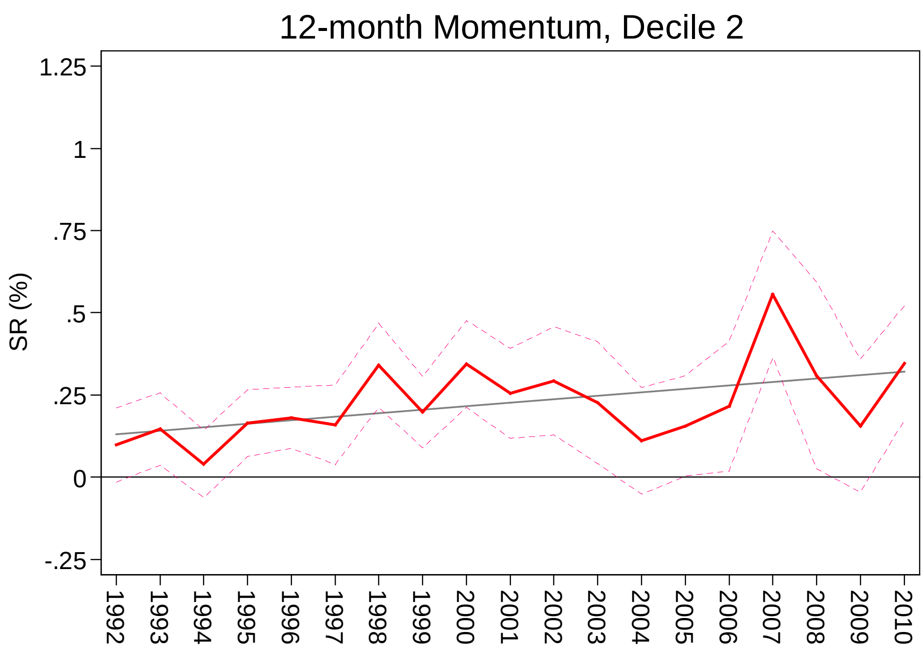



**Figure 9: Full cross-sectional relationships for momentum.** The figure plots the full set of estimated coefficients on the 10 momentum deciles dummies from the following specification:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$

These regressions are estimated annually, pooling all observations in a given year, and include a full set of month fixed effects. We compute confidence intervals for the estimated coefficients using standard errors that cluster by firm and, thus, are robust to serial correlation at the firm level.

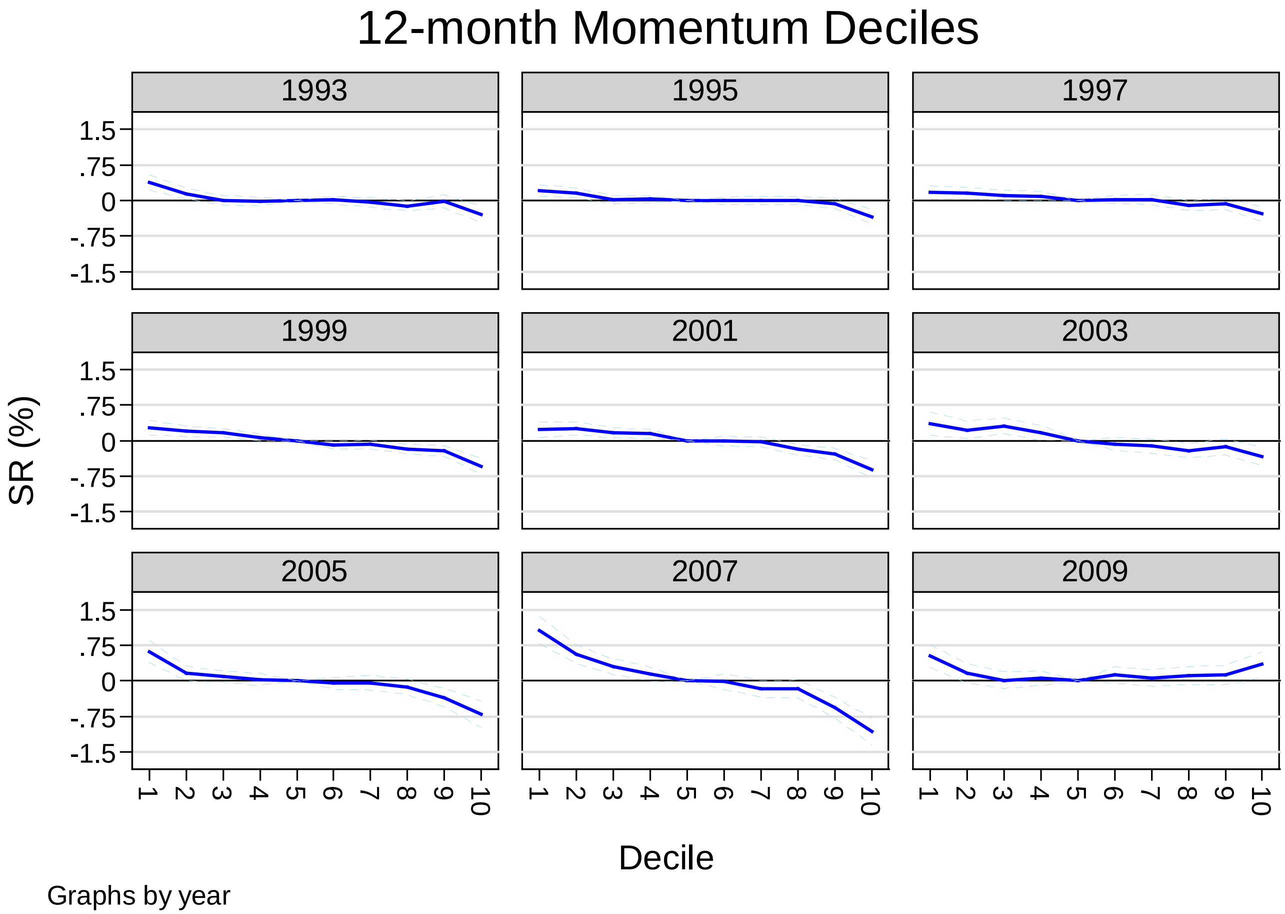


**Figure 10: Standard versus industry-adjusted value.** The figure plots the full set of estimated coefficients on the standard *B/M* decile dummies and industry-adjusted *B/M* decile dummies from the following specification:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{B/M(adj)} \cdot \mathbf{1}_{it}^{B/M(adj)} + \boldsymbol{\delta}_t^{MOM(12)} \cdot \mathbf{1}_{it}^{MOM(12)} + \boldsymbol{\delta}_t^{MOM(6)} \cdot \mathbf{1}_{it}^{MOM(6)} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$

These regressions are estimated annually, pooling all observations in a given year, and include a full set of month fixed effects. We compute confidence intervals for the estimated coefficients using standard errors that cluster by firm and, thus, are robust to serial correlation at the firm level.

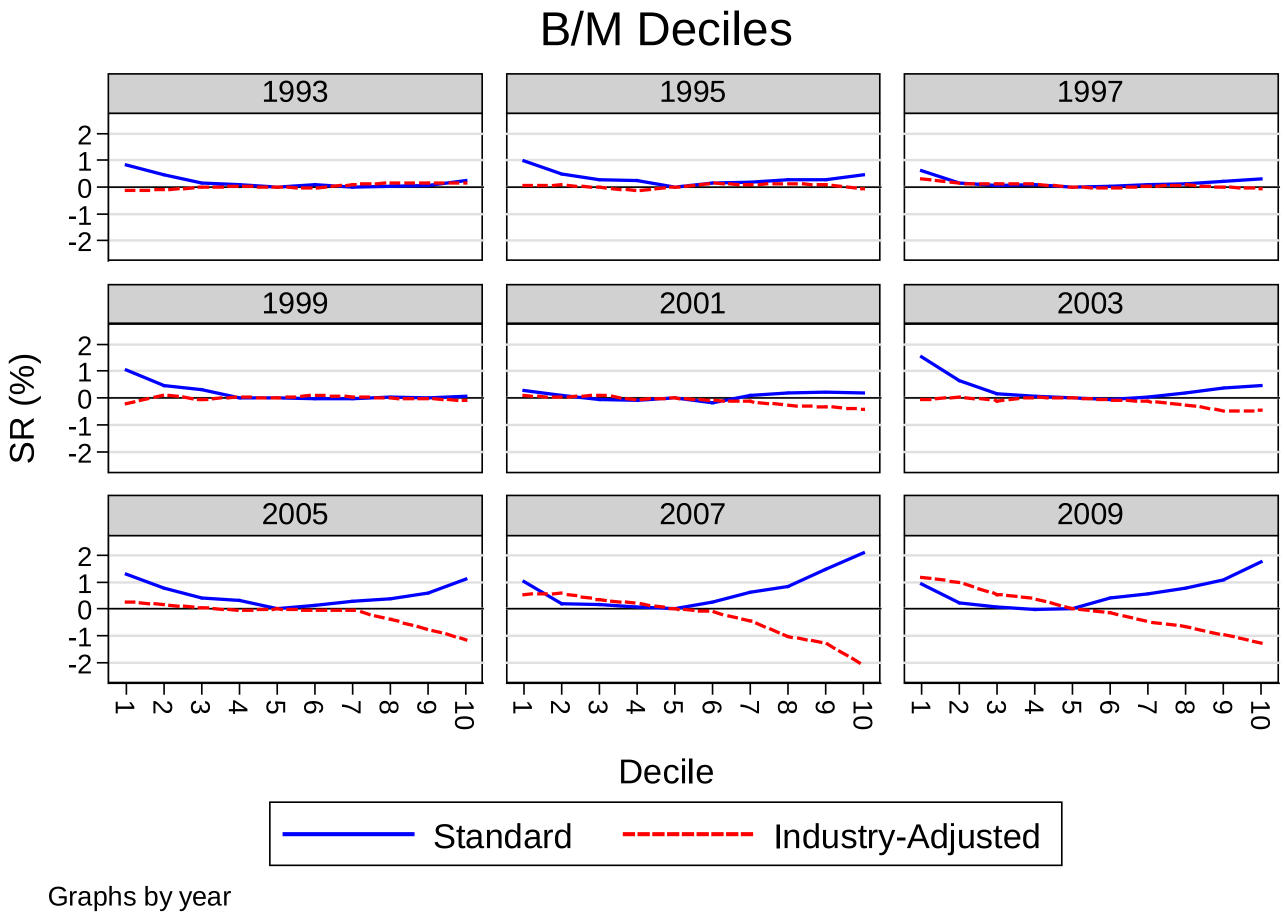


**Figure 11: Strategy capital, returns, and volatility.** The figure plots the 4-quarter moving average of the estimated quarterly coefficients on the extreme growth decile ( ) and extreme momentum loser decile ( versus annual strategy returns and realized volatilities over the same 4-quarter period.

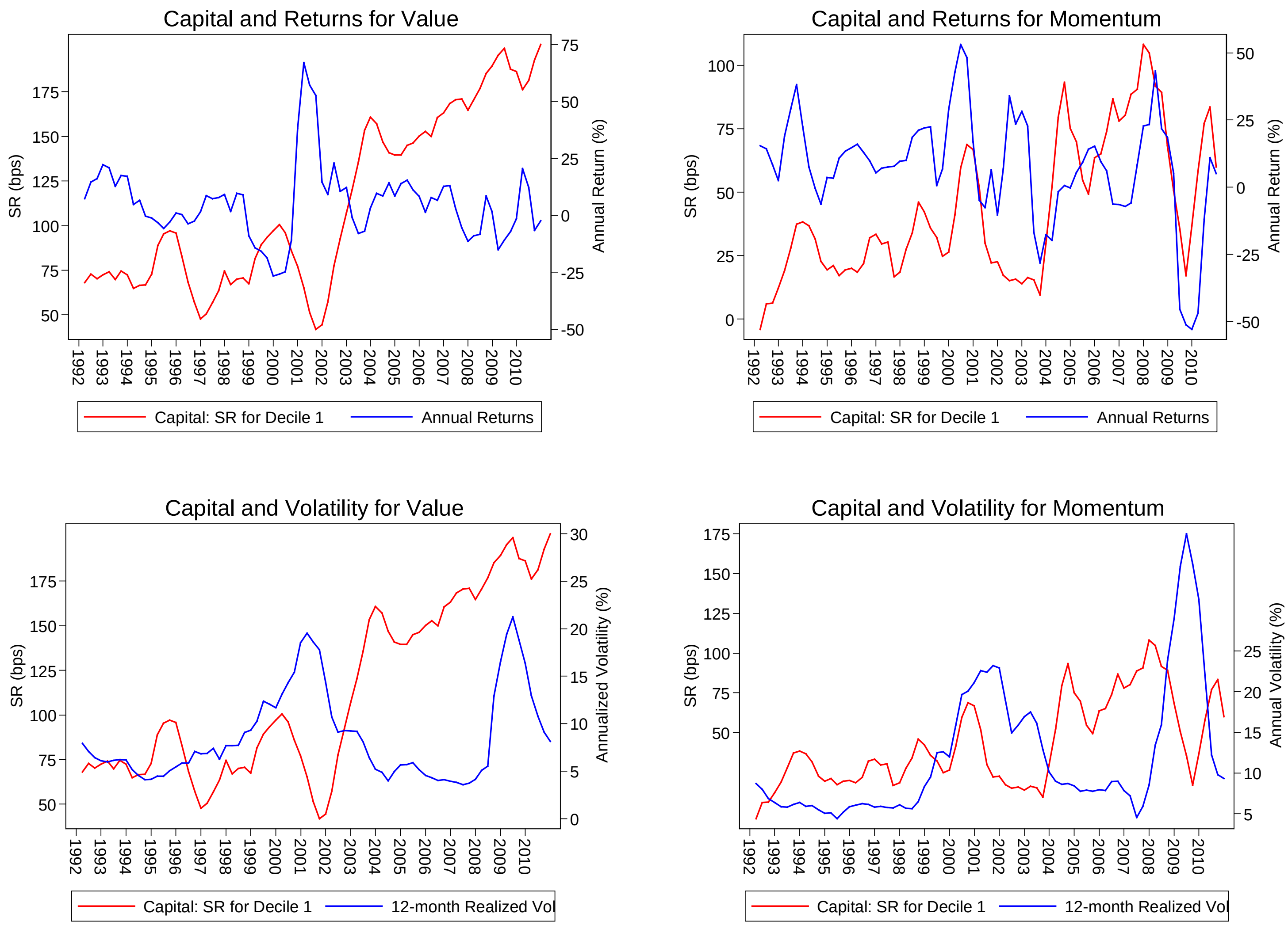


**Table 1: The effect of strategy returns on strategy-level capital flows.** This table shows regressions of the form:

$$\Delta\delta_t^k = \alpha^k + \beta^k \cdot r_{t-1}^k + \varepsilon_t^k.$$

for value and momentum. *t*-statistics are computed using heteroskedasticity robust standard errors.

| | **Value** | | | | **Momentum** | | | |
|---|---|---|---|---|---|---|---|---|
| | **Full Sample** | | **Pre-2008** | | **Full Sample** | | **Pre-2008** | |
| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** |
| | -0.460 | -0.567 | -0.377 | -0.623 | 0.239 | 0.563 | 0.881 | 1.093 |
| | [-1.17] | [-1.27] | [-1.07] | [-1.58] | [0.60] | [1.54] | [2.53] | [3.13] |
| | | -0.338 | | -0.406 | | 1.057 | | 0.777 |
| | | [-0.77] | | [-0.75] | | [2.55] | | [1.83] |
| Constant | 2.358 | 3.002 | 1.968 | 3.152 | -0.134 | -2.236 | -0.158 | -2.172 |
| | [0.84] | [0.93] | [0.72] | [0.91] | [-0.03] | [-0.59] | [-0.04] | [-0.53] |
| $T$ | 75 | 75 | 63 | 63 | 75 | 75 | 63 | 63 |
| $R^2$ | 0.018 | 0.032 | 0.013 | 0.027 | 0.005 | 0.082 | 0.057 | 0.097 |

**Table 2: The effect of strategy volatility on strategy-level capital flows.** This table shows regressions of the form:

$$\Delta\delta_t^k = \alpha^k + \psi^k \cdot \Delta\sigma_{t-1}^k + \varepsilon_t^k.$$

for value and momentum.  is the standard deviation of daily factor returns during quarter *t*. The quarterly and annual returns are in percentages and factor volatility measures are in annualized percentages. We measure the TED spread using the difference between the rate on 3-month Eurodollar deposits (i.e., 3-month LIBOR) and the yield on 3-month Treasury bill. Both rates are taken from the Federal Reserve H.15 release. *t*-statistics are computed using heteroskedasticity robust standard errors.

| | **Value** | | | | | | **Momentum** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Full Sample** | | | **Pre-2008** | | | **Full Sample** | | | **Pre-2008** | | |
| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** | **(9)** | **(10)** | **(11)** | **(12)** |
| | -2.655 | -3.742 | -3.325 | -3.209 | -3.484 | -3.478 | -0.811 | -0.413 | -0.940 | 0.651 | 0.891 | 1.140 |
| | [-4.10] | [-5.16] | [-3.98] | [-3.10] | [-2.29] | [-2.30] | [-0.83] | [-0.43] | [-1.08] | [0.81] | [1.06] | [1.50] |
| | | 1.044 | 0.417 | | 0.256 | 0.267 | | -0.793 | 0.548 | | -0.398 | -0.231 |
| | | [3.20] | [0.69] | | [0.27] | [0.28] | | [-2.84] | [0.87] | | [0.62] | [-0.35] |
| | | | 17.606 | | | -2.619 | | | -39.142 | | | -58.588 |
| | | | [1.81] | | | [-0.21] | | | [-2.88] | | | [-2.86] |
| Constant | 1.598 | 1.444 | 1.499 | 1.089 | 1.034 | 1.072 | -0.042 | 0.068 | -0.044 | 1.577 | 1.637 | 2.475 |
| | [0.61] | [0.58] | [0.61] | [0.41] | [0.40] | [0.40] | [-0.01] | [0.02] | [-0.01] | [0.42] | [0.44] | [0.71] |
| $T$ | 74 | 74 | 74 | 62 | 62 | 62 | 74 | 74 | 74 | 62 | 62 | 62 |
| $R^2$ | 0.150 | 0.229 | 0.264 | 0.157 | 0.159 | 0.159 | 0.023 | 0.053 | 0.155 | 0.010 | 0.013 | 0.122 |

**Table 3: Investigating contagion: The effect of other strategy returns on momentum capital flows.** This table shows regressions of the form:

$$\Delta\delta_t^{MOM} = \alpha + \gamma \cdot r_{t-1}^{A} + \beta \cdot UMD_{t-1} + \varepsilon_t^{MOM},$$

where is the lagged quarterly return on some other strategy *A*. *t*-statistics are computed using heteroskedasticity robust standard errors.

| | Full Sample | | | | | Pre-2008 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** | **(9)** | **(10)** |
| | 0.239 | 0.563 | 0.041 | 0.660 | 0.287 | 0.881 | 1.093 | 0.415 | 1.288 | 0.809 |
| | [0.60] | [1.54] | [0.11] | [1.94] | [0.84] | [2.53] | [3.13] | [1.04] | [4.11] | [1.90] |
| | | 1.057 | -1.324 | -0.673 | -1.520 | | 0.777 | -1.454 | -0.662 | -1.526 |
| | | [2.55] | [-1.83] | [-1.31] | [-2.14] | | [1.83] | [-1.90] | [-1.41] | [-2.07] |
| | | | 4.057 | | 2.600 | | | 3.937 | | 2.395 |
| | | | [4.07] | | [2.32] | | | [3.29] | | [1.75] |
| | | | | 4.348 | 2.640 | | | | 4.232 | 2.783 |
| | | | | [4.58] | [2.72] | | | | [4.04] | [2.54] |
| Constant | -0.134 | -2.236 | -10.611 | -13.028 | -14.155 | -0.158 | -2.172 | -11.085 | -14.718 | -15.848 |
| | [-0.03] | [-0.59] | [-2.59] | [-3.20] | [-3.53] | [0.04] | [0.53] | [-2.50] | [-3.32] | [-3.60] |
| *T* | 75 | 75 | 75 | 75 | 75 | 63 | 63 | 63 | 63 | 63 |
| $R^2$ | 0.005 | 0.082 | 0.209 | 0.205 | 0.238 | 0.057 | 0.097 | 0.215 | 0.222 | 0.251 |

**Table 4: The relationship between arbitrage capital and future returns.** This table shows regressions of the form:

$$r^k_{t\to t+4} = \mu^k + \varphi^k \cdot (\delta^k_t - \delta^k_{t-4}) + \varepsilon^k_{t\to t+4},$$

where is the 4-quarter annual return from quarter *t* to quarter *t*+4. is the value spread.

| | **Value** | | | | | | | | **Momentum** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Full Sample** | | | | **Pre-2008** | | | | **Full Sample** | | **Pre-2008** | |
| | **(1)** | **(2)** | **(3)** | **(4)** | **(5)** | **(6)** | **(7)** | **(8)** | **(9)** | **(10)** | **(11)** | **(12)** |
| | -0.125 | -0.123 | -0.147 | -0.144 | -0.135 | -0.132 | -0.161 | -0.170 | 0.032 | 0.050 | -0.109 | -0.085 |
| | [-1.99] | [-2.75] | [-1.61] | [-1.98] | [-2.14] | [-3.59] | [-1.68] | [-2.98] | [0.29] | [0.46] | [-1.97] | [-1.86] |
| | | 0.129 | | 0.129 | | 0.224 | | 0.253 | | | | |
| | | [1.93] | | [1.74] | | [2.71] | | [2.96] | | | | |
| | | | -0.173 | -0.171 | | | -0.202 | -0.297 | | -0.274 | | -0.300 |
| | | | [-0.86] | [-0.85] | | | [-0.92] | [-1.42] | | [-1.83] | | [-2.09] |
| Constant | 4.851 | -11.312 | 5.803 | -10.320 | 5.552 | -20.827 | 6.923 | -22.204 | 6.053 | 8.310 | 10.726 | 13.581 |
| | [1.42] | [-1.55] | [1.31] | [-1.36] | [1.50] | [2.22] | [1.37] | [-2.40] | [1.47] | [1.79] | [3.24] | [3.71] |
| *T* | 68 | 68 | 68 | 68 | 60 | 60 | 60 | 60 | 68 | 68 | 60 | 60 |
| $R^2$ | 0.063 | 0.172 | 0.091 | 0.194 | 0.072 | 0.255 | 0.111 | 0.215 | 0.004 | 0.074 | 0.073 | 0.159 |

## Internet Appendix: Additional Results

### A. Cross-Sectional Regression Summary Statistics

In this Appendix, we provide further details on our annual cross-sectional regressions which take the form given in equation (2). In Figure A1 we plot the cross-sectional $R^2$ and average number of stocks each month in our annual regressions. The $R^2$ hovers between 0.25 and 0.30 during the first 10 years of the sample before rising markedly in recent years. The increase in $R^2$ since 2000 is largely driven by the growing importance of size (*ME*) and institutional ownership (*IO*) in determining the cross-section short interest.

Figure A2 plots the coefficients for our six additional controls: institutional ownership, past turnover, trailing volatility, exchange dummies for NYSE and NASDAQ (AMEX is the omitted category), and a dummy indicating if the firm has convertible securities outstanding. First, we see that the impact of institutional ownership has increased over time and particularly since 2000. By 2007, a one percentage point increase in *IO* was associated with a 0.07% increase in *SR*. The growing impact of *IO* (i.e. the difference in *SR* between supply-constrained and unconstrained stocks) is consistent with a parallel shift in shorting demand for most stocks. When shorting demand is low, the supply constraint doesn't bind for either high or low *IO* stocks. However, when demand shifts out the constraint will bind for firms with low *IO*, but not for those with high *IO*. Thus, a broad increase in shorting demand (not captured by our other controls) would cause the difference in *SR* between high and low-*IO* firms to rise. Interestingly, the coefficient on *IO* fell significantly in 2009 and 2010, and the rolling monthly results show that the coefficient dropped sharply after September 2008. Anecdotally, this likely reflects the withdrawal of several large institutional investors from share lending programs in the late 2008 due to concerns about the re-investment portfolios of securities lenders.

Second, we see that *SR* is reliably increasing in past turnover. Throughout much of our sample a one percentage point increase in average turnover over the previous quarter is associated with a 0.20% to 0.25% increase in *SR*. The effect of past turnover on short interest declined somewhat in the late 1990s, recovered during the early 200s, and declined again following 2008. Third, we see that short-sellers were somewhat less willing to short highly volatile stocks during the 1990s. However, this effect has not been significant in recent years. Fourth, all else equal, we find that short interest is slightly lower for NYSE stocks when compared with AMEX stocks (the omitted group). Fifth, beginning in 2003, we find that *SR* has been noticeably larger for NASDAQ stocks than AMEX stocks. Finally, we find that the impact of having convertible securities outstanding grew steadily from 1991 to 2000. This effect declined following 2000 and from 2006-2008 there was little effect associated with having outstanding convertibles. However, the impact of convertibles on *SR* reemerged in 2009-2010.

### B. The Impact of Financial Stocks

In this section, we explore how the short interest patterns for the 2007-2009 financial crisis differ between financial and non-financial stocks. In Figure A3, we plot equal- and value-weighted short interest ratios separately for nonfinancial and financial stocks. On an equal-weighted basis, average short interest for *both* nonfinancial and financial stocks soared prior to the crisis, peaked in July 2008, and dropped sharply following the imposition of the short sales limitations for financial stocks in September 2008. While Figure A3 shows that short interest followed essentially parallel trends for nonfinancials and financials on an equal-weighted basis, the run-up in *SR* was far more pronounced for financial firms on a valued-weighted basis.

We next estimate our baseline cross-sectional specification in equation (2) separately for nonfinancial and financial stocks. In Figure A4, we plot the coefficients associated with the extreme growth and loser deciles. These figures replicate Figure 4 for all stocks separately for nonfinancial stocks in Panel A and financial stocks in Panel B. The figures for nonfinancials in Panel A are similar to those reported in the main text which makes sense given that the vast majority of all stocks are nonfinancial. The results for financials in Panel B show that there is little tendency to short financial stocks with low *B/M* (relative to other financial stocks). However, there was significant shorting of financial losers during the financial crisis and the monthly results suggest that the sharp contradiction in loser short interest in October 2008 may have been driven by the short interest ban for financials.

## C. Time-series Plots for Other Quantitative Strategies

### *C.1. Annual Short-Only Measures for Other Anomaly Strategies*

The plots below show the coefficients on our decile 1 dummies for other quantitative arbitrage strategies. First consider the *SUE* measure of earnings momentum. To obtain capital measures for earnings momentum, we add a full set of *SUE* decile dummies to our baseline model, estimating:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{SUE} \cdot \mathbf{1}_{it}^{SUE} + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}. \quad \text{(A1)}$$

Our estimates for other strategies are obtained by estimating analogous regressions. When computing deciles for these other strategies, we always code the deciles so that decile 1 is associated with abnormally low returns and decile 10 is associated with abnormally high returns.

In Figure A5, we plot our baseline short-side capital measure, $\delta_t^{k(1)}$, the boost to *SR* associated with being a decile 1 stock for quant signal *k* relative to the omitted decile 5. Given our coding conventions, we should expect to find $\delta_t^{k(1)} > 0$. Turning to the results, we see little effect from having poor earnings momentum (*SUE*). Large net stock issuance is associated with being highly shorted, particularly since 2000. This is consistent with the idea that arbitrageurs have added this signal to their models over this period. There is little effect from having high balance sheet accruals, though high cash-flow based accruals have been associated with high short interest in the years prior to the crisis. The Shumway (1997) distress measure has little effect on short interest, except for 2001 during the tech and telecom bust and 2008-2009 during the financial crisis. High CAPM residual volatility has been important in recent years, suggesting it too has gained popularity as a signal. Both gross and net asset growth have a consistent effect on short interest that has grown somewhat in recent years.

### *C.2. Annual Long-Short Capital Measures for Value and Momentum*

In Figures A6 below we plot the alternate measures of strategy capital described in the text. The lighter lines are the difference between our decile 1 dummies and our decile 10 dummies from regression (2) in the main text – i.e., $\delta_t^{k(1)} - \delta_t^{k(10)}$. As discussed above, the reluctance to short stocks that an arbitrage strategy recommends buying contains information about the amount capital playing that strategy. Given our coding of the deciles we expect to find $\delta_t^{k(1)} - \delta_t^{k(10)} > 0$. The darker lines are the results of running cross-sectional regressions of the form in equation (2) on raw characteristic deciles, rather than a full set of characteristic dummies. These regressions deliver a single coefficient for each characteristic summarizing its impact of *SR*. As seen below, while this "raw decile" approach imposes linearity on the mapping from characteristics to short interest, it appears to capture similar information from the long and short sides as that contained in the more flexible specification (2).

In Figure A6 we see that for *B/M* and momentum, the results using these short-long measures $\delta_t^{k(1)} - \delta_t^{k(10)}$ are quite similar to those presented in the main text, which only use decile

1 dummies $\delta_t^{k(1)}$. The one difference is that in recent years the level of momentum capital has generally been higher than the level of value capital. As discussed above, the reason is that in recent years investors have been extremely reluctant to short winners but not value stocks.

### *C.3. Annual Long-Short Capital Measures for Other Anomaly Strategies*

Turning to the other arbitrage strategies in Figure A7, we find an increase in the amount of capital playing *SUE* in recent years. This differs from the conclusion reached above using only the short side, which suggests that while arbitrageurs are reluctant to short stocks with positive earnings momentum, but they are not particularly eager to short stocks with negative earnings momentum. By contrast, the effects for net stock issuance (*NS*) are smaller than those seen above. This is because, somewhat surprisingly, we find that *SR* is also fairly high for large net repurchasers (i.e. $\delta_t^{NS(10)} > 0$). Turning to our accrual measures, we find that they have little effect, except for the period 2001-2004 which was marked by significant accounting scandals and 2006-2008. Next, we see that while the Shumway distress metric has a meaningful impact on *SR*, the effect is relatively constant over time with the exception of 2008. Residual CAPM volatility has little effect throughout most of the sample, but has gained popularity in 2009 and 2010. The effect of asset growth is similar to what we saw using only the short side.

**Figure A1: Regression summary statistics:** This figure plots the cross-sectional $R^2$ and average number of stocks each month in our annual regressions (see equation (2)).

**Figure A2: Coefficients for additional controls:** This figure plots the time series of coefficients for the additional controls used in our annual cross-sectional regressions (see equation (2)).

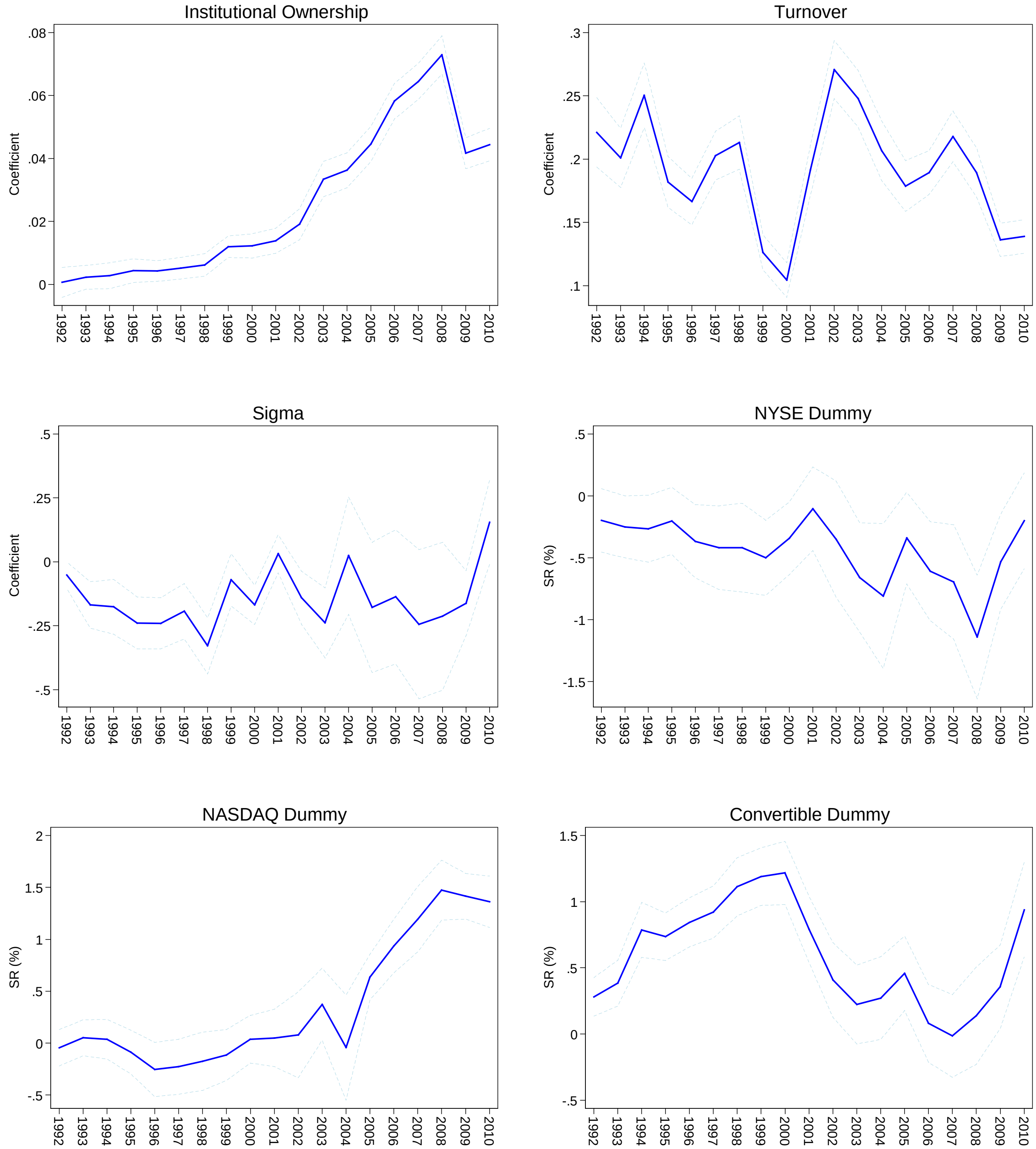


**Figure A3. Average short interest ratios for nonfinancial and financial stocks**. Panel A plots the monthly equal- and value- (i.e., market equity) weighted average short interest ratio for all nonfinancial stocks in our sample. Panel B plots short interest ratios for financial stocks. Financial stocks are stocks with Fama-French (1997) industry codes 44 (banking), 45 (insurance), 46 (real estate), or 47 (trading) based on their 48 industry classification scheme.

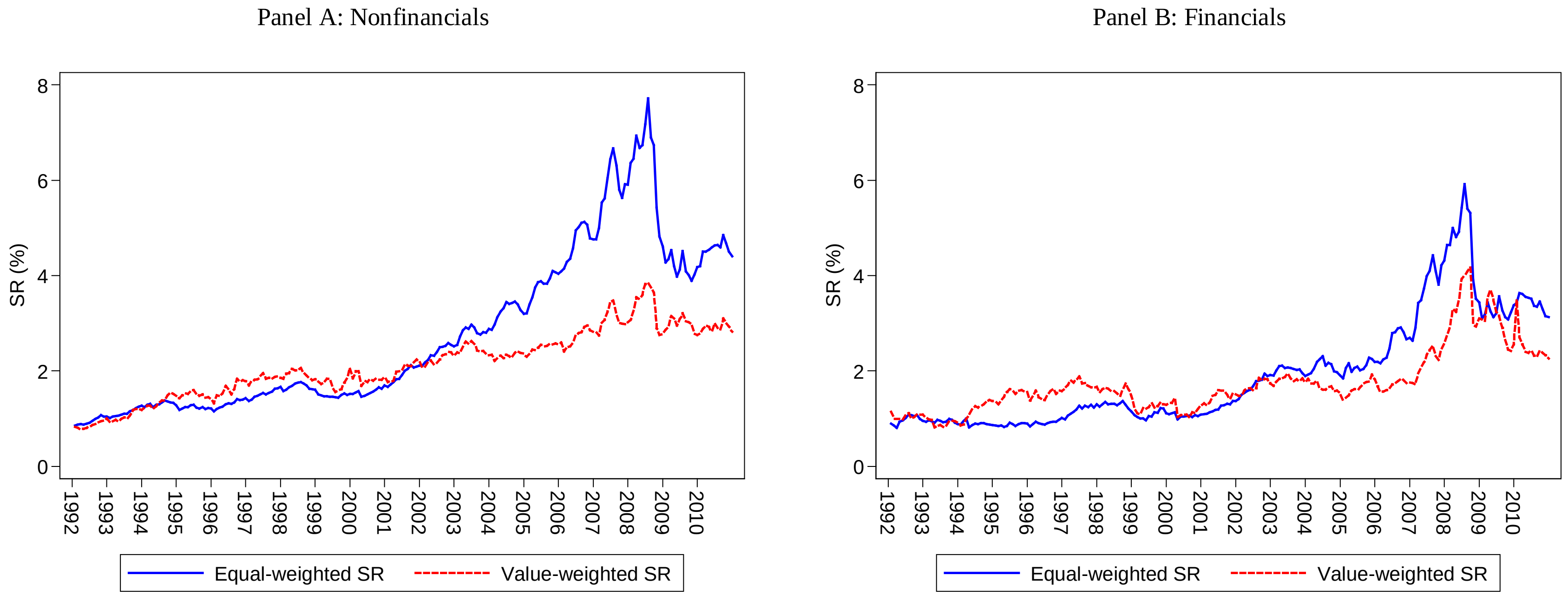


**Figure A4: Estimated capital intensities for value and momentum strategies for nonfinancial stocks and financial stocks.** The figure plots the time series of estimated coefficients on the extreme growth decile (        ) and extreme momentum loser decile (            from the following specification:

$$SR_{it} = \alpha_t + \boldsymbol{\delta}_t^{B/M} \cdot \mathbf{1}_{it}^{B/M} + \boldsymbol{\delta}_t^{MOM} \cdot \mathbf{1}_{it}^{MOM} + \boldsymbol{\delta}_t^{SIZE} \cdot \mathbf{1}_{it}^{SIZE} + \boldsymbol{\beta}_t' \mathbf{x}_{it} + \varepsilon_{it}.$$

These regressions are estimated annually, pooling all observations in a given year or on a rolling quarterly basis, pooling all observations in a given 3 month period. Both specifications also include a full set of month fixed effects. We compute confidence intervals using standard errors that cluster by firm and, thus, are robust to serial correlation at the firm level. In Panel A, we include only nonfinancial stocks. In Panel B, we include only financial stocks.

Panel A: Nonfinancial stocks only

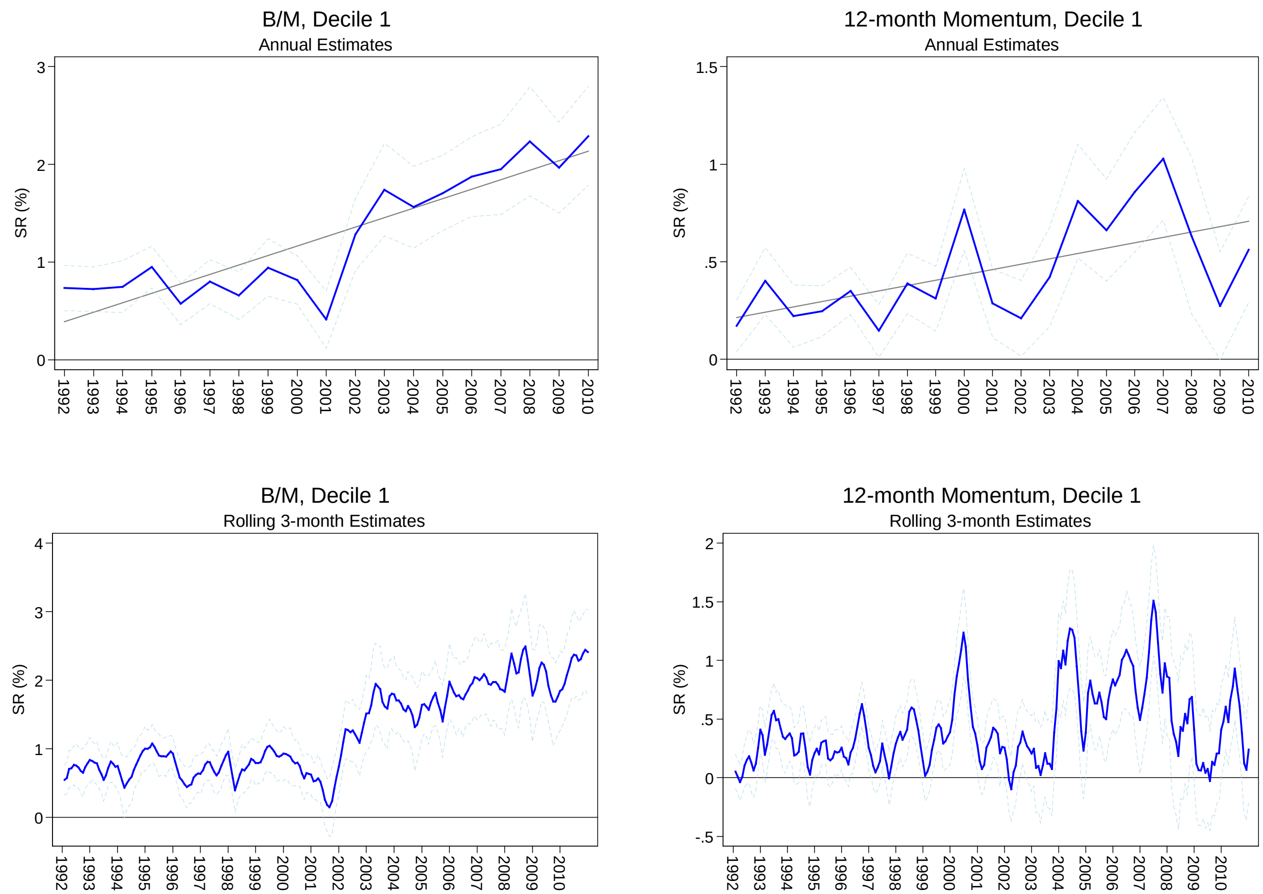


**Figure A4: Estimated capital intensities for value and momentum strategies for financial stocks and nonfinancial stocks (continued)**

Panel B: Financial stocks only

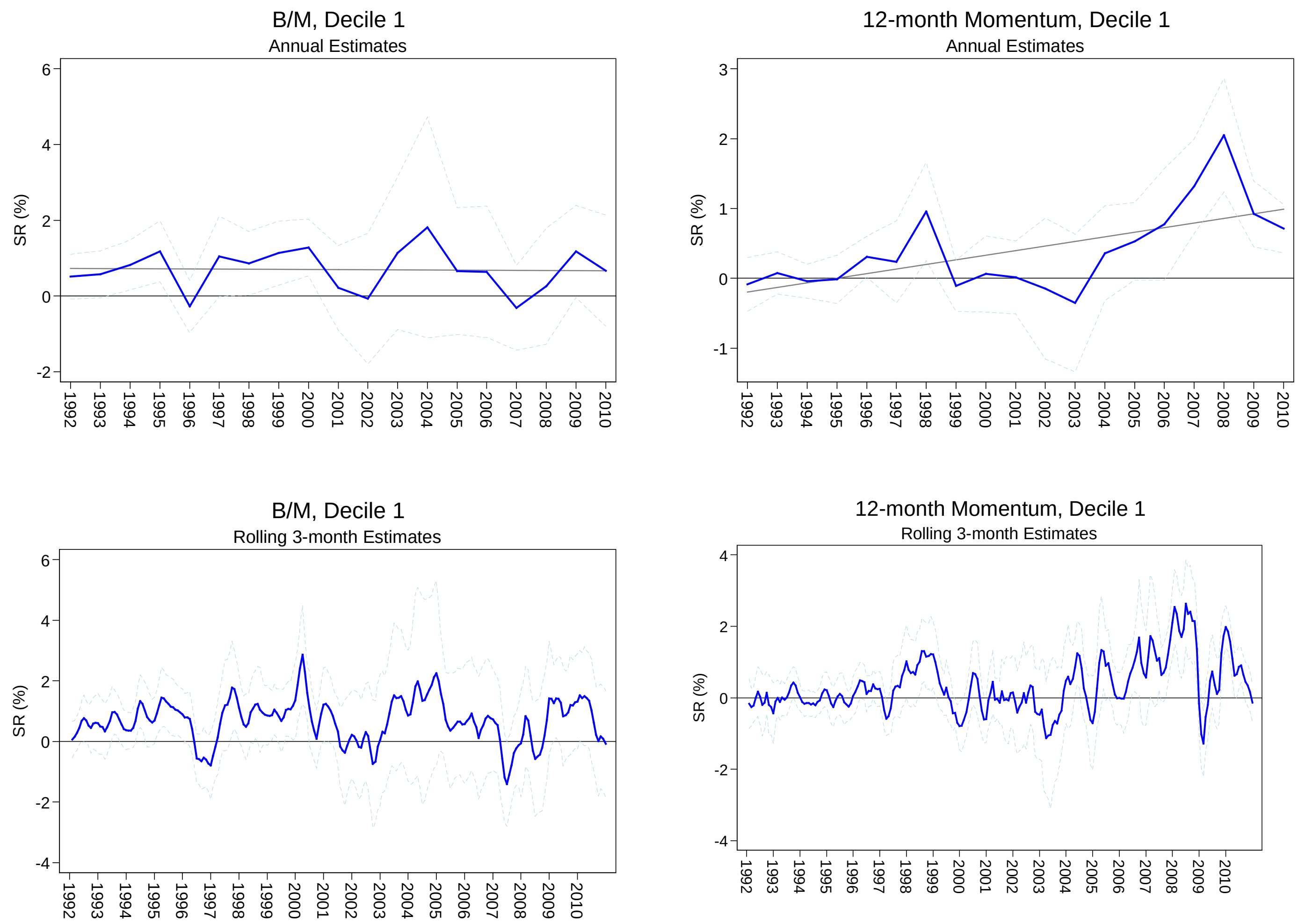


**Figure A5: Short-side capital measures for other anomaly strategies:** This figure plots the time series of for other anomaly strategies based on annual cross-sectional regressions (see equation (A1)).

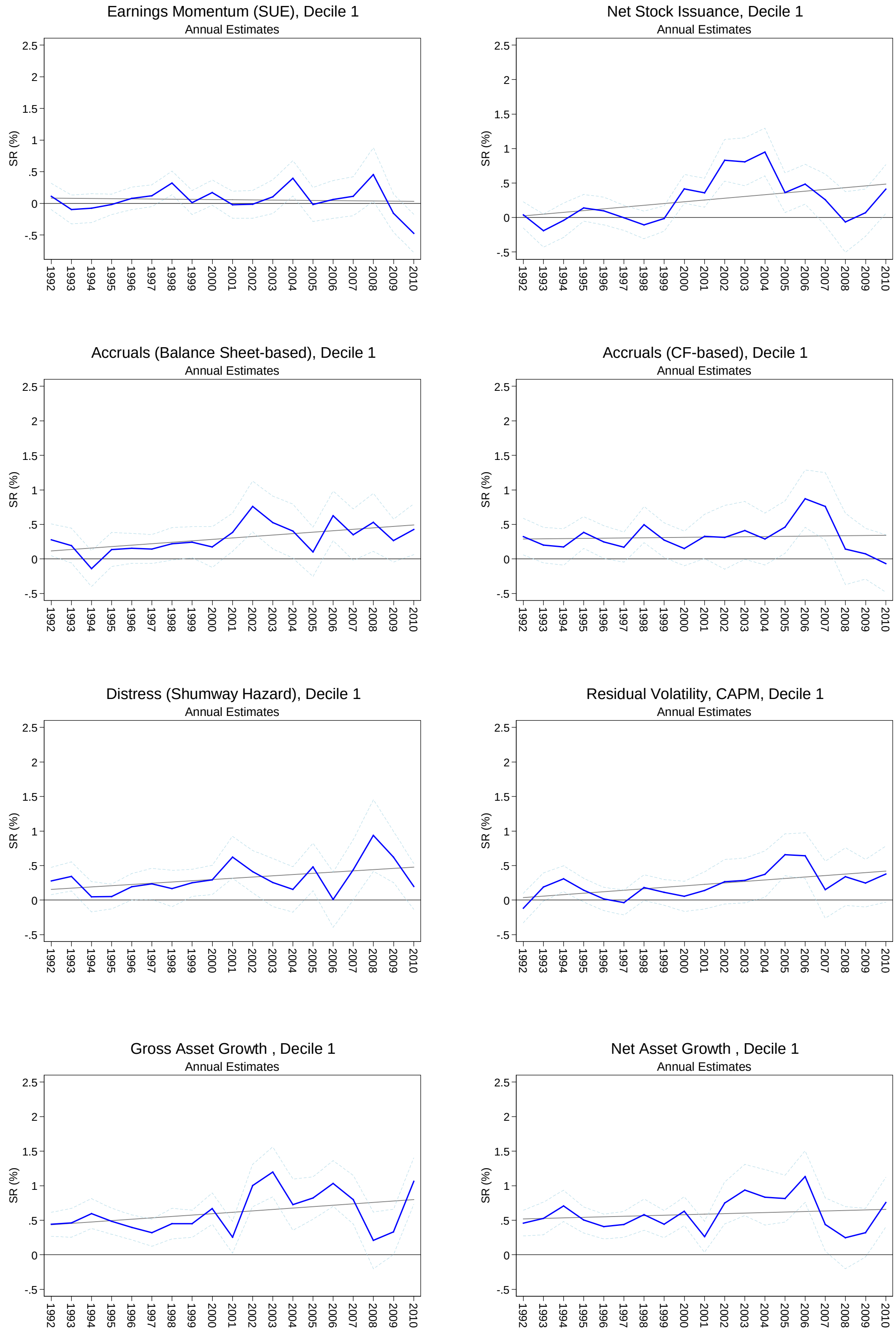


**Figure A6: Short-long capital measures for value and momentum:** This figure plots difference in coefficients between decile 1 and decile 10 from equation (2), i.e. , alongside the coefficient from a regression of *SR* on raw characteristic deciles.

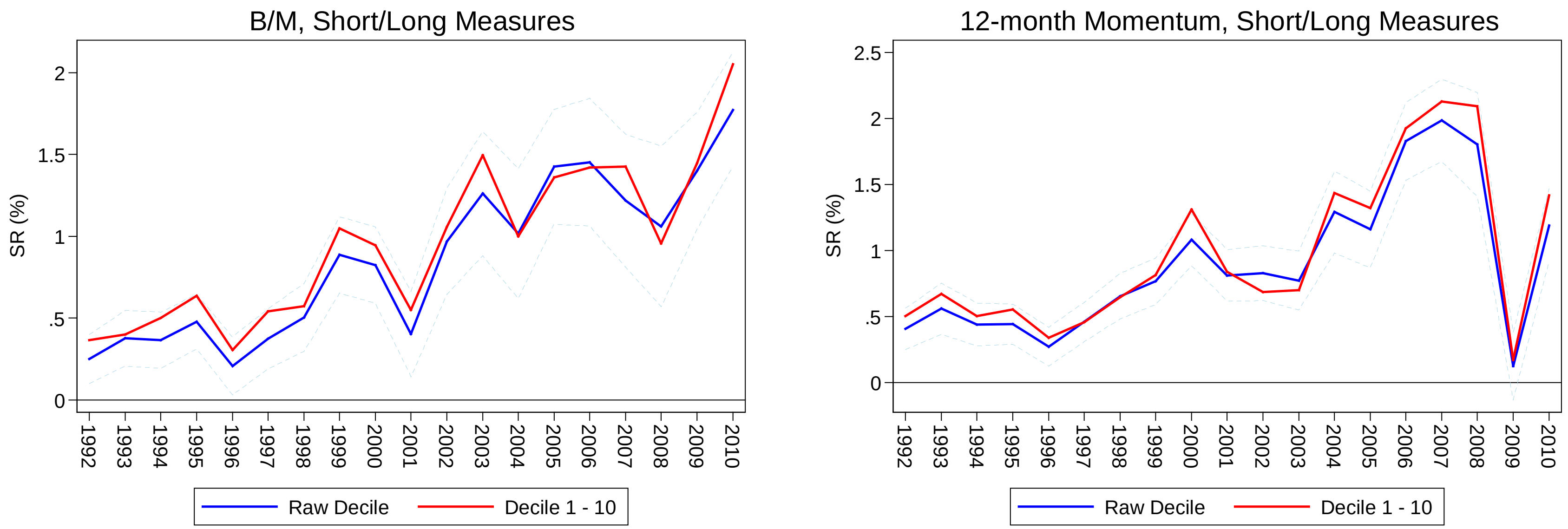
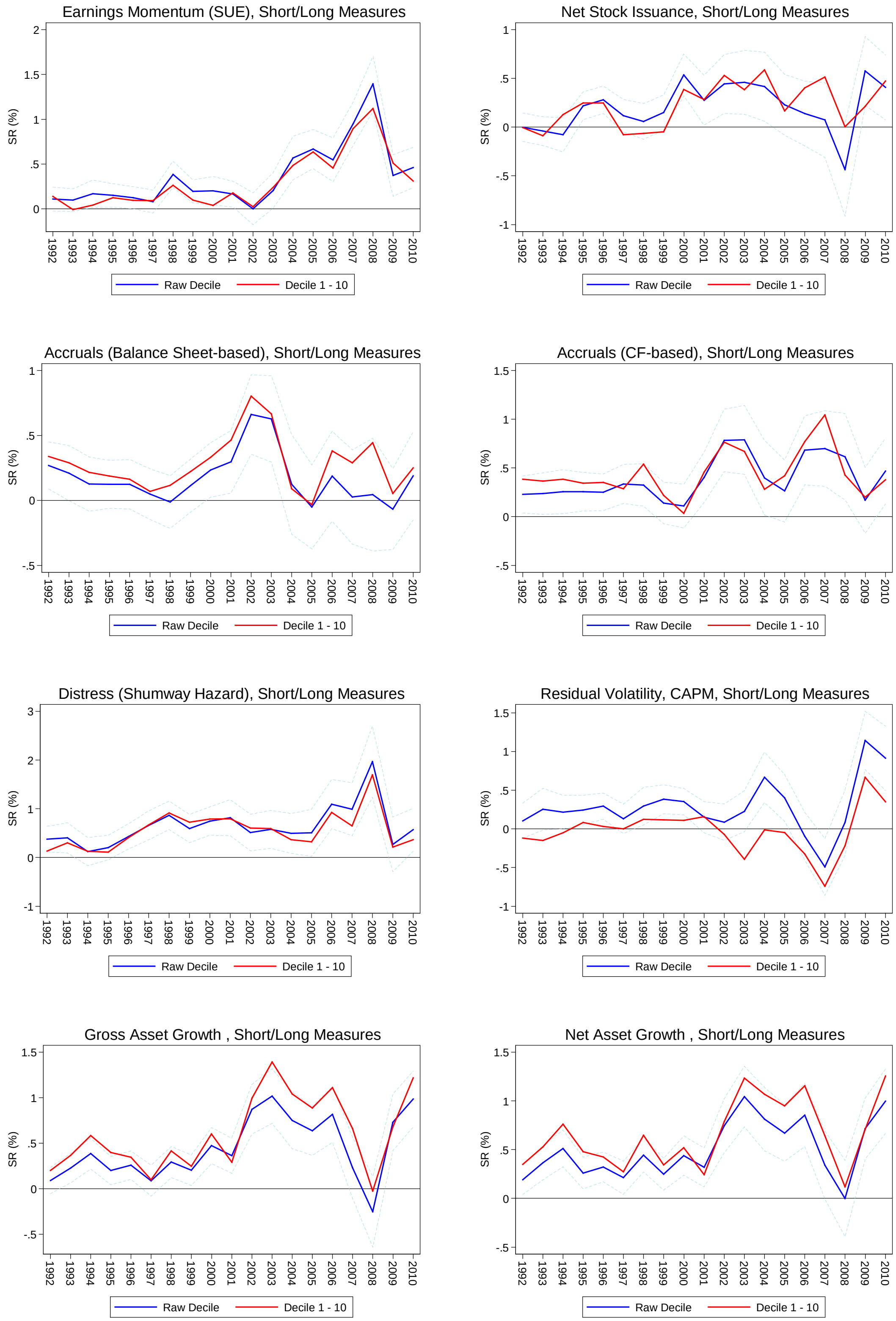


**Figure A7: Long-short capital measures for other anomaly strategies:** This figure plots difference in coefficients between deciles 1 and 10 from (A1) and the coefficient from a regression of *SR* on raw characteristic deciles.



# Illiquidity Premia in the Equity Options Market*

Peter Christoffersen — Rotman, CBS and CREATES
Ruslan Goyenko — McGill University
Kris Jacobs — University of Houston
Mehdi Karoui — McGill University

April 19, 2011

**Abstract**

Illiquidity is well-known to be a significant determinant of stock and bond returns. We report on illiquidity premia in equity option markets. An increase in option illiquidity decreases the current option price and predicts higher expected option returns. This effect is statistically and economically significant. It is robust across different empirical approaches and when including various control variables. The illiquidity of the underlying stock affects the option return negatively, consistent with a hedging argument: When stock market illiquidity increases, the cost of replicating the option goes up, which increases the option price and reduces its expected return.

JEL Classification: G12

Keywords: illiquidity; equity options; cross-section; option returns; option smile.

*We would like to thank IFM$^2$ and SSHRC for financial support. Christoffersen can be reached by phone at 416-946-5511 and via email to peter.christoffersen@rotman.utoronto.ca. Goyenko can be reached by phone on 514-398-5692 and via email to ruslan.goyenko@mcgill.ca. Jacobs can be reached by phone on 713-743-2826 and via email to kjacobs@bauer.uh.edu. Karoui can be reached by phone on 514-692-9155 and via email to mehdi.karoui@mail.mcgill.ca.

# 1 Introduction

The existing literature contains a wealth of evidence regarding illiquidity premia in stock and bond markets. It has been shown in both markets that illiquidity affects returns, with more illiquid assets having higher expected returns. In equity markets, Amihud and Mendelson (1986, 1989), Eleswarapu and Reinganum (1993), Brennan and Subrahmanyam (1996), Amihud (2002), Jones (2002), Pastor and Stambaugh (2003), and Acharya and Pedersen (2005) compare stock market illiquidity to ex-post returns on equities. In bond markets, Amihud and Mendelson (1991), Warga (1992), Boudoukh and Whitelaw (1993), Kamara (1994), Krishnamurthy (2002), Longstaff (2004), Goldreich, Hanke and Nath (2005), and Beber, Brandt and Kavajecz (2009) analyze the impact of bond illiquidity on expected bond returns.

There is also a growing body of evidence on the existence of illiquidity premia in other markets, see for instance Deuskar, Gupta, and Subrahmanyam (2011) for evidence on interest rate derivatives and Bongaerts, de Jong, and Driessen (2010) for evidence on the credit default swap market. Vijh (1990) measures liquidity premia and market depth in the equity options market, and George and Longstaff (1993) measure bid-ask spreads in index options and explain the nature of cross-sectional differences in these spreads. However, the literature has been mostly silent so far about the relationship between illiquidity and expected returns in equity option markets. This is surprising, because similar to stock and bond markets, market makers in option markets incur order processing and asymmetric information costs. George and Longstaff (1993) find that a substantial fraction of the bid-ask spread in option markets is attributed to premia compensating dealers for the risk of holding uncovered positions in illiquid options.

Our contribution is to study the effect of option and stock illiquidity on equity option returns. We document the statistical significance and economic magnitude of the impact of option illiquidity on option returns. We also estimate the effect of illiquidity in the underlying stocks on option returns. In a frictionless, complete-market model, the price of the option can be replicated by trading in the underlying asset and a risk free bond. If the underlying asset is illiquid, then the trading strategy replicating the price of the option is harder to implement and the illiquidity costs of this trade should affect the price and thus return of the option.

We establish our main results using cross-sectional Fama-MacBeth (1973) regressions for daily and weekly returns. We present univariate regressions but also multivariate regressions controlling for stock volatility, stock returns, lagged option returns, and other firm characteristics, as in Duan and Wei (2009). An increase in option illiquidity has a positive and significant impact on next period's option returns, across all moneyness and maturity categories. This evidence is consistent with the existence of an illiquidity premium in the options market, similar to the effect of stock illiquidity on stock returns reported by Amihud (2002). The effect is also economically significant: for example, a two standard deviation shock to out-of-the-money short-term call illiquidity results in a 2.37% change in the next day out-of-the-money short-term call returns. A two standard deviation shock to out-of-the-money short-term put illiquidity results in a 1.61% change in the next day out-of-the-money short-term put returns.

We find that the illiquidity of underlying stocks also has a significant impact on option prices. As expected, this effect is opposite to the effect of option illiquidity on option returns. A positive shock to stock illiquidity decreases next period's option returns. This finding is consistent with trading motivated by hedging considerations. Whenever stock market illiquidity increases, the higher stock transaction costs will increase the cost of replicating the option, which will increase the option price and reduce its expected return. This effect is also economically meaningful, although it is smaller compared to the impact of option illiquidity: for example, a two standard deviation shock to stock illiquidity would result in a 0.87% change in the next day short-term out of the money call returns and a 0.59% change in the next day short-term out of the money put returns. This is consistent with Cetin, Jarrow, Protter and Warachka (2006), who suggest that illiquidity of underlying stocks constitutes a significant part of option prices.

Analyzing the effects of illiquidity in the cross-section of option returns is empirically more challenging than analyzing the cross-section of stock returns, because of the strong dependence of option returns on the returns of the underlying. We therefore investigate the robustness of our results by analyzing the cross-section of implied volatilities in addition to the cross-section of returns. We find that both the illiquidity of the options and the underlying assets help explain the level of implied volatility, and that the sign of the effect is consistent with the evidence from the cross-section of returns. Moreover, option illiquidity significantly affects the slope of the implied volatility curve: the implied volatility curve is steeper for more illiquid option contracts.

Finally, we report time-series evidence for liquidity decile portfolios. We find that a contemporaneous increase in option illiquidity has a significantly negative effect on option prices, consistent with the cross-sectional evidence. This result is again similar to the effect of stock illiquidity on stock returns reported by Amihud (2002). A contemporaneous shock to option illiquidity decreases the current price and increases expected option returns to compensate traders for holding illiquid contracts.

To the best of our knowledge these results are new to the literature. The existing empirical evidence on equity option illiquidity is very limited. Using data from an interesting natural experiment, Brenner, Eldor and Hauser (2001) compare central bank issued and exchange traded options and report a 21% illiquidity discount for non-tradable central bank issued options. Cao and Wei (2010) document commonality in the illiquidity on equity option markets, but do not investigate the impact of illiquidity on option returns.[1]

The paper is organized as follows. Section 2 lays out our main hypotheses and discusses the theoretical literature on expected option returns. Section 3 describes the data and variables we use, in particular, the construction of option returns and illiquidity measures. Section 4 presents empirical results on the impact of illiquidity on the cross-section of option returns. Section 5 investigates the cross-section of implied volatilities and the slope of the implied volatility curve. Section 6 presents time-series evidence, and Section 7 concludes.

# 2 Illiquidity and Expected Option Returns

Motivated by the literature on liquidity risk in the bond and equity markets, we investigate the following hypotheses in our empirical work:

1. In the cross-section, illiquid options earn on average higher expected returns, supporting the existence of a positive illiquidity premium.

2. The illiquidity of the underlying stock negatively affects expected option returns, which is consistent with the following hedging argument: Higher stock transaction costs increase the cost of replicating the option, which increases the option price and reduces its expected return.

3. Option illiquidity and the illiquidity of the underlying stock are important determinants of the level and slope of the implied volatility curve.

4. In a time series analysis, lagged option illiquidity predicts future expected option returns and illiquidity shocks are negatively related to contemporaneous option returns, consistent with a positive illiquidity premium.

Before forging ahead with empirical tests of these hypotheses, we briefly review existing theoretical results on expected option returns. These results will be used to provide guidance in the design of the illiquidity tests.

[1] The equity option literature also contains related results on trading activity and demand pressures. Prominent papers include Garleanu, Pedersen, and Poteshman (2009), Easley, O'Hara, and Srinivas (1998), Lakonishok, Lee, and Poteshman (2007), Mayhew (2002), Pan and Poteshman (2006), and Roll, Schwartz, and Subrahmanyam (2010).

Mainstream option valuation theory assumes away illiquidity in option markets as well as in the market for the underlying and bond markets.[2] This is done in order to arrive at option valuation expressions that are deterministic functions of the underlying asset price and the interest rate as well as other variables, including volatility.

In the standard Black and Scholes (1973) model, the option price, $O$, for a non-dividend paying stock with price $S$ is a function of the strike price, $K$, the risk-free rate, $r$, maturity, $T$, and constant volatility, $\sigma$, which we can write

$$O = BS\left(S, K, r, T, \sigma\right) \tag{2.1}$$

Coval and Shumway (2001) show that in this basic model with constant risk-free rate and constant volatility, the expected instantaneous return on an option $E\left[R^{O}\right]$ is given by

$$E\left[R^{O}\right] = \left(r + (E\left[R^{S}\right] - r)\frac{S}{O}\frac{\partial O}{\partial S}\right)dt \tag{2.2}$$

where $E\left[R^{S}\right]$ is the expected return on the stock. The sensitivity of the option price to the underlying stock price (the option delta), denoted by $\frac{\partial O}{\partial S}$, will depend on the variables in (2.1). The delta is positive for call options and negative for puts. Thus the expected excess return on call options is positive and the expected excess return on put options is negative.

The presence of $E\left[R^{S}\right]$ and $\frac{\partial O}{\partial S}$ on the right-hand side of equation (2.2) shows that it is critical to properly control for the return on the underlying stock when regressing option returns on illiquidity measures.

In the Black-Scholes model, the risk-free rate is assumed to be constant across maturities. Bakshi, Cao and Chen (1997) show empirically that allowing for stochastic interest rates does not change the value of the option by much, compared to the simple use of maturity-specific risk-free rates in the Black-Scholes model. Thus we do not control for stochastic interest rates in our empirical analysis below.

The absence of stochastic volatility in the Black-Scholes model is much more critical. Hull and White (1987) and Scott (1987) develop option valuation models with stochastic volatility. Heston (1993) develops a stochastic volatility model that allows for correlation between the shock to returns and the shock to volatility, as well as for a volatility risk premium to compensate sellers of options for volatility risk. Broadie, Chernov, and Johannes (2009) and Duarte and Jones (2007) show that in a standard stochastic volatility model, the expected

[2] Black and Scholes (1973), Hull and White (1987), and Heston (1993) are classic examples of papers in this literature. See Jones (2006) for a detailed analysis of returns on S&P500 index options.

option return is given by

$$E\left[R^O\right] = \left(r + (E\left[R^S\right] - r)\frac{S}{O}\frac{\partial O}{\partial S} + \lambda\frac{\sigma}{O}\frac{\partial O}{\partial \sigma}\right)dt \tag{2.3}$$

where the sensitivity of the option price to volatility (the option vega), denoted by $\frac{\partial O}{\partial \sigma}$, is positive for all options, and where the price of volatility risk, $\lambda$, is negative because the added volatility risk increases the option value.[3] Equation (2.3) shows that it will be important to control for the dynamic volatility of the stock when regressing option returns on illiquidity measures.

The standard option valuation models discussed above do not allow for transactions costs or liquidity risk. A much smaller option valuation literature allows for illiquidity effects in the underlying asset. Prominent papers include Cetin, Jarrow and Protter (2004), Jarrow and Protter (2005), and Cetin, Jarrow, Protter and Warachka (2006). The latter paper shows that the Black-Scholes pricing model holds in the presence of liquidity costs associated with trading the underlying asset, but also that the optimal hedging strategy changes compared to Black-Scholes. Toft (1996) studies option valuation in the presence of trading costs. Constantinides and Perrakis (2002, 2007), Oancea and Perrakis (2007), and Constantinides, Jackwerth, and Perrakis (2009) rely on a stochastic dominance approach to characterize bounds on option prices. As this approach establishes option valuation bounds rather than option prices, expressions for the relationship between expected option returns and liquidity measures are not readily available.

In recent work, Bongaerts, de Jong and Driessen (2010) develop an equilibrium asset pricing model with liquidity risk where the underlying asset is in positive net supply and the derivative asset is in zero net supply. The model contains heterogeneous investors who differ with respect to their degree of risk-aversion, initial wealth and investment horizon. In a linear special case of their model, the expected option return can be derived as

$$E\left[R^O\right] = \delta_1 E\left[R^S\right] + \beta_2 E\left[IL^O\right] + \beta_3 E\left[IL^S\right] \tag{2.4}$$

where $IL^O$ is the illiquidity (in terms of transaction cost) of the option and $IL^S$ is the illiquidity of the underlying stock. Bongaerts, de Jong and Driessen (2010) show that when the less risk-averse investors have long positions in the option, the coefficient on $E\left[IL^O\right]$ is positive and the option buyers will earn a positive illiquidity premium. These investors are more sensitive to transaction costs and will therefore require compensation for illiquidity risk. The model is not conclusive with respect to the sign of the coefficient on $E\left[IL^S\right]$,

[3] The derivation of (2.3) assumes that the diffusion to volatility is linear in the volatility level.

which therefore remains an open question in the empirical analysis, to which we now turn.

# 3 Data and Variable Construction

## 3.1 Option Returns

We investigate the impact of option illiquidity as well as stock illiquidity on option returns. The construction of these two measures is complicated by the large number of option contracts and the need to construct stock illiquidity measures using high-frequency data. Moreover, data on option contracts for smaller firms is less readily available when researching longer time periods. We therefore limit ourselves to options data for S&P500 index constituents from OptionMetrics, which includes daily closing bid and ask quotes on American options, as well as their implied volatilities and deltas. By limiting ourselves to S&P500 firms, we bias our results towards not finding evidence of the importance of illiquidity. The sample period is January 1996 to December 2007. We limit the sample to firms that have options trading throughout the entire sample period. We implement this by verifying whether the firms have options trading on the first trading day of each calendar year in the sample, as well as the last day in our sample, December 31, 2007. This yields a sample of 341 firms.

We repeat our analysis for six different option samples. For each firm, we consider put and call options for two maturity categories: short-term, with time to maturity between 20 and 70 days, and long-term, with time to maturity between 71 and 180 days. Each maturity category is in turn divided according to moneyness into in-the-money (ITM), at-the-money (ATM), and out-of-the-money (OTM) options. We follow Driessen, Maenhout, and Vilkov (2009) and Bollen and Whaley (2004) and define moneyness according to the option delta from OptionMetrics,[4] which we denote by $\Delta$. OTM options are defined by $0.125 < \Delta \leq 0.375$ for calls and $-0.375 < \Delta \leq -0.125$ for puts. ATM options correspond to $0.375 < \Delta \leq 0.625$ for calls and $-0.625 < \Delta \leq -0.375$ for puts, and the ITM category is defined by $0.625 < \Delta \leq 0.875$ for calls and $-0.875 < \Delta \leq -0.625$ for puts.

Following Goyal and Saretto (2009) and Cao and Wei (2010), we apply filters to the option data and eliminate the following contracts: (i) prices that violate no-arbitrage conditions; (ii) observations with ask price lower than or equal to the bid price; (iii) options with open interest equal to zero; (iv) options with missing prices, implied volatilities or deltas; (v) options with prices lower than \$3 and bid-ask spread below \$0.05, or prices equal or higher than \$3 and bid-ask spread below \$0.10, on the grounds that the bid-ask spread is lower

[4] For American options, OptionMetrics relies on the Cox, Ross, and Rubinstein (1979) binomial tree model for computing implied volatilities and deltas.

than the minimum tick size which signals a data error. We have also re-run the empirical tests without imposing any filters, and the results are robust.

For all remaining options, our method for computing option returns follows Coval and Shumway (2001). We compute daily returns using quoted end-of-day bid-ask midpoints if quotes are available on the respective days. We compute equally-weighted average daily returns on a firm-by-firm basis for each moneyness and maturity category by averaging option returns for all available contracts. For each option category and for each firm, the return from $t$ to $t+1$ is defined by

$$R^O_{t+1} = \frac{1}{N}\sum_{n=1}^{N} \frac{O_{t+1}(K_n, T_n - 1) \times f_{t+1} - O_t(K_n, T_n) \times f_t}{O_t(K_n, T_n) \times f_t} \tag{3.1}$$

where $N$ is the number of available contracts in the particular category at time $t$ with legitimate quotes at time $t+1$. $O_t(K_n, T_n)$ is the mid-point quote, (ask+bid)/2, for an option with strike price $K_n$ and maturity $T_n$, and $f_t$ is the cumulative adjustment factor for splits or other distribution events, provided by OptionMetrics.

Weekly option returns are constructed similar to daily returns using Friday-to-Friday data wherever possible, and alternatively using a minimum of four daily returns.[5]

Figure 1 plots the daily option returns over time. Figure 1A contains the call option returns and Figure 1B has the put option returns. The short-term returns in the left panels are clearly more volatile than the long-term returns in the right panels. This is true for both calls and puts. All the option returns display volatility clustering and strong evidence of non-normality. As is typical of daily speculative returns, the mean is completely dominated by the dispersion.

Table 1 reports summary statistics. We first compute the respective statistics for each firm and report the average across firms. Table 1 shows that call returns on average are positive and put returns are negative, for daily data as well as weekly data in all categories. This is as expected from the option deltas as shown in (2.2). The option returns exhibit positive skewness and excess kurtosis in all categories as well, which is also as expected due to option gamma. Returns on OTM options are higher than returns on ITM options. They are also more variable and exhibit higher kurtosis. Returns on short-term options are higher and more variable than returns on long-term options, confirming the visual impression from Figure 1. The option returns display little evidence of serial dependence judging from the first-order autocorrelation, $\rho(1)$, but the absolute return autocorrelation $\rho^{abs}(1)$ is positive

[5] We try the following combinations: Friday-to-Friday, then Friday-to-Thursday, then Thursday-to-Friday, then Thursday-to-Thursday. If none of these are available then we discard the weekly observation for that option.

for all categories, confirming the volatility clustering apparent from Figure 1.

### 3.2 Illiquidity Measures for Stocks and Options

We investigate the impact on option returns of option illiquidity but also of illiquidity in the underlying stock market. There is an extensive literature on stock market illiquidity as we discussed in the introduction. We follow the convention in the literature and compute stock illiquidity as the effective spread obtained from high-frequency intraday TAQ (Trade and Quote) data. Specifically, for a given stock, the TAQ effective spread on the trade is defined as

$$IL_k^S = 2\left|\ln(P_k) - \ln(M_k)\right|, \tag{3.2}$$

where $P_k$ is the price of the $k^{th}$ trade and $M_k$ is the midpoint of the consolidated (from different exchanges) best bid and offer prevailing at the time of the $k^{th}$ trade. The daily stock's effective spread, $IL^S$, is the dollar-volume weighted average of all $IL_k^S$ computed over all trades during the day

$$IL^S = \frac{\sum_k DolVol_k IL_k^S}{\sum_k DolVol_k}$$

where the dollar-volume, $DolVol_k$, is the stock price multiplied by the trading volume.

The literature on equity option illiquidity is in its infancy, and therefore it is less clear how to define the option illiquidity measure. Furthermore, transaction prices to estimate effective spreads are not available for options. Similar to conventional illiquidity measures for stocks, we therefore measure illiquidity in the option market with relative quoted bid-ask spreads.[6] This is a transparent measure of illiquidity, and better alternatives are not readily available.[7] We compute relative quoted bid-ask spreads using end-of-day quoted bid and ask prices provided by Ivy DB OptionMetrics.[8] For each contract, we compute the daily relative quoted spread

$$IL_{t,n}^O = \frac{OA_t(K_n, T_n) - OB_t(K_n, T_n)}{O_t(K_n, T_n)} \tag{3.3}$$

where the prices $O_t(K_n, T_n)$, $OA_t(K_n, T_n)$, and $OB_t(K_n, T_n)$ are, respectively, the end of day closing mid-point, ask, and bid quotes reported in OptionMetrics, for an option with strike price $K_n$ and maturity $T_n$. Note $O_t(K_n, T_n) = (OA_t(K_n, T_n) + OB_t(K_n, T_n))/2$.

[6] For studies on stock market illiquidity that use relative bid-ask spreads, see for instance Hasbrouck and Seppi (2001), Huberman and Halka (2001), Chordia, Roll, and Subrahmanyam (2000, 2001), and Chordia, Sarkar, and Subrahmanyam (2005).

[7] Dollar quoted bid-ask spreads are not a good alternative as liquidity indicators, because they are mainly driven by the maturity and moneyness of the option contract. See Cao and Wei (2010) for a discussion.

[8] We use the following fields in OptionMetrics: "Best bid" defined as the best, or highest, closing bid price across all exchanges on which the option trades. Similarly for "Best offer".

The equally-weighted average spreads are then computed for each option category as

$$IL_t^O = \frac{1}{N} \sum\nolimits_{n=1}^{N} IL_{t,n}^O \tag{3.4}$$

where $N$ is the number of available contracts that are within the particular category at time $t$.

Given the data constraints, using quoted spreads as an alternative to effective spreads is reasonable. Battalio, Hatch and Jennings (2004), who use data for January 2000 through June 2002, which is part of our sample period, find that for large stocks the ratio of effective spread to quoted spread fluctuates between 0.8 and 1. Since our sample is limited to S&P500 firms, quoted spreads are a good substitute for effective spreads.

Panel A of Table 2 presents summary statistics for the relative bid-ask spread illiquidity measure using our cross-section of 341 firms. For the option illiquidity, $IL^O$, we find that short-term contracts are more illiquid than long-term contracts, regardless of whether the options are OTM, ATM, or ITM. For example, the average relative spread of short-term OTM call options is 34.02%, while for long-term OTM calls the average spread is 22.67%.[9]

Table 2 also shows that illiquidity is highest on average for OTM options, followed by ATM options, which in turn are more illiquid than ITM options. We therefore conclude there are strong moneyness and maturity effects in liquidity. In order to control for this, we will run our empirical tests separately on six different moneyness and maturity categories.

An alternative to the use of relative spreads as an illiquidity measure is Amihud's (2002) illiquidity measure, the price impact value, which is also considered by Bongaerts et al. (2010). We construct this measure for options as follows: For each day and for each option category we compute the average return and the average dollar volume across all available contracts. Dollar volume is computed as the bid-ask midpoint multiplied by trading volume. We then compute the ratio of the absolute return to the dollar volume for each day and average it for each week. This is similar Amihud's (2002) implementation, with the difference that we construct a weekly rather than a monthly measure.

Table 3 reports summary statistics for Amihud's illiquidity measure. Across option categories, OTM options have the highest price impact value. This is consistent with the evidence on relative spreads in Table 2, which shows that OTM options are most illiquid. Among other categories, similar to Table 2, ITM options are the least illiquid and ATM options are in between. The cross-sectional correlation between Amihud's price-impact measure and relative quoted bid-ask spreads ranges between 0.20 and 0.33 (not reported). This

[9] To put the magnitudes of these relative spreads in perspective, the average dollar spread in our sample is 22.9 cents for calls and 23.8 cents for puts. This is very similar to the 21.3 cent average dollar spread reported by Vijh (1990).

is comparable to the evidence presented in Goyenko et al. (2009) for stocks.

For our purpose, the disadvantage of Amihud's (2002) price impact measure is that our empirical investigation uses daily and weekly returns on options, because constructing monthly option returns is much less straightforward. Goyenko, Holden, and Trzcinka (2009) argue against aggregating Amihud's (2002) price impact measure at lower than monthly frequencies, on the grounds that it yields a noisy estimate of illiquidity. We therefore use the relative spreads for our main results. Nevertheless, we replicate our main results using Amihud's (2002) measure, and we obtain qualitatively similar results. These results are available from the authors on request.

Trading volume and open interest are sometimes used as illiquidity measures. Table 3 reports on option trading volume and open interest. Table 3 shows that, as is well-known in the empirical option valuation literature, open interest and volume are highest for ATM and OTM options and lower for ITM options. So, while ITM options are the cheapest to trade in a relative bid-ask sense, which is our measure of liquidity, the ATM and OTM options have the highest trading volume. This is consistent with existing literature, which suggests that volume is not informative about illiquidity in option markets. For instance, Mayhew (2002) argues that an option can be liquid even if it has low trading volume. This may occur if other options on the same stock are actively traded. In that case it is easy for a market maker to hedge the low-volume option with actively traded options at other strikes and maturities, as well as to hedge calls with puts and vice versa. Therefore, when thinking of illiquidity in terms of trading costs, one should not expect an obvious relationship between illiquidity and trading volume in option markets. The apparent incongruity between option volume and trading cost also has interesting parallels to the literature on stock market liquidity. Pastor and Stambaugh (2003) discuss the October 19, 1987 crash when the NYSE set a record in trading volume but where the stock market was highly illiquid from a trading cost perspective.

The right-most column in Table 2 shows that in our sample, stocks are on average substantially more liquid than options. The average relative bid-ask spread for stocks is 0.26% in our sample. This is lower than most estimates reported in the literature, which is due to the fact that our sample is limited to S&P500 firms, which are the most liquid. Panel A of Table 2 also indicates that option illiquidity is substantially more volatile than stock illiquidity.

Figure 2A depicts the evolution of our call illiquidity measure over time for all six option categories, and Figure 2B does the same for put illiquidity. For all six option categories, we report the average of the liquidity measure. Option illiquidity clearly declines over the sample period, but not in a monotonic fashion. As in Pastor and Stambaugh (2003), we see occasional large spikes in the illiquidity measures. The largest spike took place on September 17-18, 2001, which were the first days of trading after the September 11 attacks. Smaller spikes occur towards the end of the sample as the credit crisis gets underway.

The top panel of Figure 3 plots stock illiquidity over time. Stock illiquidity clearly decreases over time, which the literature attributes to decreases in tick size, the increase in electronic trading, and decimalization. There are illiquidity spikes associated with the 1997 Asian crisis, the 1998 LTCM collapse, 9/11, and the WorldCom bankruptcy in 2002.

In order to put our liquidity data into perspective, Figure 3 also plots the S&P500 index level (middle panel) and the VIX volatility index from the CBOE (bottom panel). The inverse relationship between market returns and volatility, the so-called "leverage effect" is evident when comparing the S&P500 level with VIX. Figure 3 also shows some evidence of dependence between spikes in stock illiquidity and spikes in the VIX.

Panels B and C of Table 2 show that option illiquidity has a sizeable positive correlation with stock market illiquidity for all option categories, with somewhat higher correlations for call options. This finding suggests co-movement between illiquidity in the two markets. Table 2 also indicates that the illiquidity of OTM call and put contracts is substantially more volatile than the other categories. Illiquidity of OTM short-term calls and puts is highly correlated, at 0.59 (not reported in Table 2). This supports the findings of George and Longstaff (1993), who suggest that traders regard call and put options as substitutes (via put-call parity) with trading activity in calls and puts being positively related to the bid-ask spreads in calls and puts.

### 3.3 Other Variables

We obtain daily stock returns, prices, and the number of outstanding shares from CRSP. Weekly stock returns are compounded from daily returns. Data on long-term debt and the par value of preferred stock, which are used to compute firm leverage, are from Compustat. The S&P 500 constituents are also from Compustat. The returns on the Fama-French and momentum factors are from Ken French's online data library.

## 4 Illiquidity and the Cross-Section of Option Returns

We investigate the cross-sectional relationship between option illiquidity and expected option returns. We proceed by running daily and weekly cross-sectional regressions, and subsequently testing the significance of the time-series means of the estimated coefficients, as in Fama and MacBeth (1973).

## 4.1 Computing Adjusted Option Returns

Variations in the price of the underlying security are by far the biggest determinant of returns, and it is important to account for this when analyzing determinants of option returns as we showed in equation (2.2). The common practice in the literature is to use delta-hedged option returns. While this transformation is appropriate for studying factors affecting option returns other than illiquidity, it creates a bias when testing the effect of illiquidity on option returns. In particular, Cetin, Jarrow, Protter and Warachka (2006) show that in a Black-Scholes economy with frictions, hedging does not eliminate the risk of the underlying stock. The hedging error due to the illiquidity of the underlying stock inflates option prices. Therefore, using delta-hedged returns biases our test results for stock illiquidity. We instead proceed as follows: we first run a cross-sectional regression of option returns on the returns of the underlying stock and their lagged values. We also include squared stock returns to control for the nonlinear dependence between the two variables

$$R_{i,t}^{O} = \delta_{0,t} + \delta_{1,t} R_{i,t}^{S} + \delta_{2,t} R_{i,t-1}^{S} + \delta_{3,t} \left( R_{i,t}^{S} \right)^2 + \varepsilon_{i,t}, \quad i = 1, 2, ...$$

and we refer to the residuals plus the intercept from these regressions as adjusted option returns, which we denote

$$\tilde{R}_{i,t}^{O} = \hat{\delta}_{0,t} + \hat{\varepsilon}_{i,t}$$

Below, we regress these adjusted option returns cross-sectionally on the illiquidity measures and a number of control variables.[10]

## 4.2 Capturing Liquidity Effects

Our treatment of illiquidity as an explanatory variable in the cross-section follows Amihud's (2002) investigation of expected stock returns, which is in turn inspired by the analysis of French, Schwert, and Stambaugh (1987). Table 2 reports average estimates of the first-order autocorrelation of individual illiquidity. The estimated values of $\rho(1)$ clearly indicate a rather persistent process, in line with the results for stock illiquidity reported by Amihud (2002).

We compute the lagged illiquidity measure, $IL_{i,t-1}^{O}$, as described in Section 3, for every firm in the sample and use it as a measure of expected liquidity.[11] Following Amihud (2002)

[10] For robustness we also run the regressions in one step, i.e. we control for stock return, lagged stock return and squared stock return on the right hand side together with the other control variables. The results are qualitatively similar, but obviously the regression R-square is much higher.

[11] The illiquidity measure, $IL_{i,t-1}^{O}$, described in Section 3, is based on the contracts available at time $t-1$. However, only contracts with returns available at time $t$ are considered in the computation of $IL_{i,t-1}^{O}$,

and French, Schwert, and Stambaugh (1987), we use ex-post realized returns as a measure of expected returns. We run cross-sectional regressions of returns between times $t-1$ and $t$ on the liquidity measure at time $t-1$; we also run similar regressions controlling for multiple other return determinants.

Since the illiquidity of the underlying asset can affect trading activity in the option markets via hedging pressures, we also include our measure of stock illiquidity $IL^S_{i,t-1}$.

## 4.3 Control Variables

We use a number of control variables in the liquidity regressions. To account for stale prices in the daily data, we include the lagged adjusted option return $R^O_{i,t-1}$ in the regression. Another important determinant of expected option returns is volatility, as we showed in equation (2.3). We estimate historical volatility from the daily stock return data using a simple $GARCH(1,1)$ model:

$$R^S_{i,t} = \mu_i + \sigma_{i,t-1} z_{i,t} \tag{4.1}$$

$$\sigma^2_{i,t} = \alpha_{0,i} + \alpha_{1,i}\sigma^2_{i,t-1} + \alpha_{2,i}\sigma^2_{i,t-1} z^2_{i,t-1} \tag{4.2}$$

where $R^S_{i,t}$ is the stock return, $\mu_i$ is the conditional mean, $\sigma^2_{i,t}$ is the conditional variance, and $z_{i,t}$ is a standard normal i.i.d. innovation.

Duan and Wei (2009) argue that the proportion of systematic risk affects the prices of individual options, and therefore option returns. We thus include $b_{t-1}$ in the regression, which is the square root of the R-square from the regression of stock returns on the Fama-French and momentum factors. Following Duan and Wei (2009), we obtain daily estimates of $b_{t-1}$ by using one-year rolling windows to run daily OLS regressions of the excess stock returns on the standard four equity factors (the market, size and book-to-market factors from Fama and French, 1993, and the momentum factor from Carhart, 1997). Furthermore, we control for firm-specific characteristics such as size and leverage which have been shown to affect the distribution of options prices, see for instance Dennis and Mayhew (2002) and Duan and Wei (2009). Following Duan and Wei (2009),we measure size using the natural logarithm of the firm's market capitalization. We define leverage as the sum of long-term debt and the par value of the preferred stock, divided by the sum of long-term debt, the par value of the preferred stock, and the market value of equity.

ensuring consistency between returns and illiquidity used in the regressions. As a robustness check, we repeat the tests using illiquidity based on all contracts, and the results are qualitatively very similar.

## 4.4 Firm-Level Results using Daily Returns

Our most general cross-sectional regression is motivated by the theoretical model in Bongaerts, de Jong and Driessen (2010), as discussed in equation (2.4) in Section 2. We run this regression with and without the control variables discussed above. The most general regression we consider is given by

$$\tilde{R}^O_{i,t} = \alpha_t + \beta_{1,t}\tilde{R}^O_{i,t-1} + \beta_{2,t}IL^O_{i,t-1} + \beta_{3,t}IL^S_{i,t-1} + \beta_{4,t}\sigma_{i,t-1} + \beta_{5,t}b_{i,t-1} + \beta_{6,t}\ln(size_{i,t-1}) + \beta_{7,t}lev_{i,t-1} + \varepsilon_{i,t} \tag{4.3}$$

We run this cross-sectional regression on every day $t$ using all firms available for a given moneyness/maturity category, and subsequently compute the time-series averages of the estimated coefficients.[12] These averages are reported in Table 4. To control for serial correlation, the Fama-MacBeth (1973) t-statistics are corrected according to the Newey and West (1987) procedure using twenty-two lags for daily data.

Panel A of Table 4 reports the results for daily call option returns for all moneyness/maturity categories, and Panel B reports on put options. For call contracts, option illiquidity $IL^O_{i,t-1}$ significantly predicts higher option returns the next day at the 1% significance level, across all maturity and moneyness categories. The coefficient on $IL^O_{i,t-1}$ is statistically significant when $IL^O_{i,t-1}$ is the only regressor, but also when including the control variables as in (4.3). Moreover, the $IL^O_{i,t-1}$ coefficient is not much affected when including the control variables. This suggests that option illiquidity is an independent determinant of option returns, and that its effect is not captured by other well-known determinants of option returns.

Ignoring option illiquidity is tantamount to overestimating option prices. The effect is also economically significant. For example, for OTM short-term options, the coefficient on $IL^O_{i,t-1}$ is 0.062. Table 2 indicates that the standard deviation for OTM short-term call option illiquidity is 0.191. Therefore, a two standard deviation positive shock to OTM short-term call option illiquidity would result in a 2.37% increase in the next-day return on the call option. This is a significant magnitude for daily changes in prices. The coefficient on $IL^O_{i,t-1}$ is higher for short-term contracts than for long-term contracts, implying that the illiquidity impact is especially pronounced for short-term options. Short-term OTM contracts have the highest illiquidity risk.

The positive predictive effect of option illiquidity on expected option returns is consistent with existing findings on the effect of stock illiquidity on stock returns (Amihud, 2002). The positive contemporaneous illiquidity shock decreases current prices and thus increases

[12] In all our tests, we require at least 30 firm-observations with all data available for each time t (day or week or month) to run a cross-sectional regression.

the expected return over the next period. Option markets are characterized by a positive illiquidity premium, because buyers of illiquid contracts seek higher expected returns.

These results have implications for the option valuation literature. Bakshi, Kapadia and Madan (2003) and others find that S&P500 index options are relatively more expensive than individual equity options, particularly in the case of short-term and OTM options: Index options display much larger risk-neutral kurtosis, (negative) skewness and volatility than equity options. This is regarded as somewhat of a puzzle because an index is a portfolio of equities and so one would expect index options to display less evidence of nonnormality than individual equity options. Our results suggest that this valuation difference could be driven by differences in liquidity. Index options are well-known to be much more liquid than individual equity options. Thus individual equity option prices are relatively more depressed by illiquidity than are index options. This is particularly true for short-term OTM options where the difference in pricing between index and equity options is the greatest.

Our results also have implications for option trading. In the well-known dispersion trade (see Driessen, Maenhout and Vilkov, 2009), investors sell index options, which are relatively expensive, and buy a portfolio of (cheaper) equity options as a hedge. This trade is commonly regarded as being driven mainly by correlation risk: When correlation increases, index options become relatively more expensive, which is bad for the dispersion seller. Our results suggest that this trade is also nontrivially exposed to liquidity risk because the equity options bought are much less liquid than the index options sold.

The illiquidity of the underlying stock $IL^S_{i,t-1}$ has a negative effect on expected call option returns, and this effect is statistically significant for all categories of short-term calls and for long-term ITM calls. Given the positive coefficient on $IL^O_{i,t-1}$, the negative coefficient on $IL^S_{i,t-1}$ is consistent with a hedging argument. When stocks become more illiquid, the higher stock transaction costs will increase the cost of replicating the option, which will increase the option price and reduce its expected return.

These results for short-term options are consistent with the evidence reported in Cetin, Jarrow, Protter and Warachka (2006), who suggest ITM options are the least exposed to the illiquidity of underlying stocks, since with ITM options most of the rebalancing of option payoff replicating portfolios occurs only as the stock price decreases. This argument suggests the largest effects for OTM options, where the replicating portfolio rebalancing occurs as the stock price changes in either direction, with ATM options somewhere in between. The coefficient on $IL^S_{t-1}$ is more negative for OTM options, -2.077, and the least negative for ITM options, -0.428, with the coefficient for ATM options being in between, -0.873. For long-term calls the effect of stock illiquidity is less pronounced. Our estimate of -2.077 implies that a two standard deviation shock to stock illiquidity results in a 87 basis point change next day for short-term out of the money call returns. Therefore, while the effect of stock illiquidity on call returns is small compared to that of option illiquidity, it is still economically meaningful.

Among other variables, the effect of the lagged option return $\tilde{R}^O_{i,t-1}$ is negative and significant, which indicates negative mean reversion in option returns, consistent with the evidence on stock returns at the daily frequency. The volatility of the underlying also has a negative and significant effect on expected option returns. This finding is consistent with an option pricing model allowing for stochastic volatility and negative volatility risk premium (e.g. Heston, 1993). As discussed in Section 2, in a stochastic volatility model, the expected option return, $E\left[R^O\right]$, is negatively related to volatility through the positive option vega, $\frac{\partial O}{\partial \sigma}$, and the negative price of volatility risk, $\lambda$, as in equation (2.3), which we repeat here for convenience

$$E\left[R^O\right] = \left(r + \left(E\left[R^S\right] - r\right)\frac{S}{O}\frac{\partial O}{\partial S} + \lambda\frac{\sigma}{O}\frac{\partial O}{\partial \sigma}\right)dt \tag{4.4}$$

Further, the proportion of systematic risk, $b_{i.t-1}$, is typically small and insignificant. Among firm-specific characteristics, size exhibits a strong influence while leverage is typically insignificant.

For the put options in Panel B, we obtain similar results for option illiquidity $IL^O_{t-1}$, which positively predicts next period put returns. This effect is significant across all moneyness and maturity categories in the univariate regression, but also when controlling for other variables. Also confirming the results for call options, the effect of put illiquidity on expected put returns is more pronounced for the short-term contracts compared to the long-term contracts. For example, the coefficient on $IL^O_{i,t-1}$ for OTM short-term contracts is 0.048, almost twice the coefficient for the long-term contracts, which is 0.023. In economic terms, the coefficient of 0.043 implies that a two standard deviation shock to OTM short-term put illiquidity results in a 1.61% change in the next day put return. This is also an economically meaningful number.

The results for stock illiquidity are also quite robust for the put options. The coefficient on $IL^S_{i,t-1}$ is negative in all six categories and significant in four of six categories.

The total risk $\sigma_{i,t-1}$ has a significant impact on put returns across all moneyness and maturity categories, whereas the effect from the share of systematic risk, $b_{t-1}$, is small and insignificant. Among other firm characteristics, size and leverage also seem to affect put returns, consistent with the evidence in Dennis and Mayhew (2002).

We verified the robustness of the results in Table 4 by using raw option returns and including current and lagged stock returns as regressors. This yields very similar results for the variables of interest. The resulting R-squares are of course much higher, as the stock return explains a significant part of the variation in option returns.

## 4.5 Firm-Level Results using Weekly Returns

Daily prices may be subject to problems such as stale quotes and microstructure noise. Table 5 therefore repeats the exercise from Table 4 using weekly data. To control for serial correlation, the Fama-MacBeth (1973) t-statistics are corrected according to the Newey and West (1987) procedure using eight lags. We classify the options as OTM, ATM, or ITM, as well as short-term and long-term according to their average delta and maturity over the week. The weekly results reported in Panel A of Table 5 for call options confirm the results from Panel A of Table 4. The coefficients on lagged option illiquidity, $IL^{O}_{i,t-1}$, are robustly positive, and the estimates are statistically significant. The coefficients on lagged stock illiquidity, $IL^{S}_{i,t-1}$, are negative and statistically significant. This is true for the univariate as well as the multivariate regressions.

The evidence on weekly put returns in Panel B of Table 5 also broadly confirms the results from daily returns in Panel B of Table 4. Option illiquidity is significantly positively related to option returns in four of the six categories in the multivariate regression. Stock illiquidity is strongly negatively related with option returns for all categories.

Overall, the evidence in Tables 4 and 5 documents a statistically and economically significant impact of option illiquidity on expected option returns. This effect is similar to the effect of stock illiquidity on expected stocks returns (Amihud, 2002) and suggests a positive illiquidity premium in equity option markets. The call option results are robust to controlling for lagged option returns, stock returns, and stock volatility, as well as stock illiquidity and firm-specific characteristics. Moreover, the illiquidity of the underlying stock has a significantly negative impact on expected call and put option returns.

## 4.6 Portfolio Results

In the regression approach used in Tables 4 and 5, noise in returns on individual option contracts may weaken inference. It is therefore of interest to confirm the results using different empirical techniques. A simple alternative approach is to sort firms in liquidity portfolio baskets, and investigate the resulting patterns in portfolio returns. This portfolio approach can reduce the noise in returns on individual contracts. Panel A of Table 6 presents portfolio results for daily call returns, and Panel B for daily put returns. Table 7 presents results for weekly data. At time $t-1$ (day or week) all options are sorted into liquidity deciles. Subsequently we compute the average option return, stock return, illiquidity and market capitalization for each decile portfolio at time $t$.

Consistent with the liquidity premium hypothesis, option returns are monotonically increasing from the most liquid decile portfolio to the most illiquid decile portfolio for both calls and puts. For the call options in Panel A, stock returns are increasing across decile portfolios, for daily as well as weekly data; for the put options in panel B, stock returns are decreasing across decile portfolios, as expected. Consistent with liquidity co-movement between stock and option markets, stock illiquidity monotonically increases with option illiquidity for both calls and puts, for weekly as well as daily data. It is also seen that the more illiquid firms are on average smaller.

We can use portfolios to investigate whether returns on different horizon investments outweigh the substantial transaction costs. Similar to Amihud and Mendelson (1986), we compute returns net of transaction costs, using the bid and ask quotes. The net return $R^O_{net}$-*Long* is computed as $(\text{bid}_t - \text{ask}_{t-1})/\text{ask}_{t-1}$ and $R^O_{net}$-*Short* is computed as $(-\text{ask}_t + \text{bid}_{t-1})/\text{bid}_{t-1}$. Not surprisingly, the net returns after-trading costs for both long and short option positions are negative, clearly indicating that at short horizons liquidity premia are absorbed by market frictions.

# 5 Illiquidity and Implied Volatility

In Section 4, we study the impact of option illiquidity on the cross-section of option returns. This is a natural starting point, because it is straightforward to build intuition for illiquidity's expected effects on returns. The existing literature on illiquidity in bond and stock markets also investigates the effects of illiquidity on returns, and provides a natural reference point. However, there are some obvious differences between the analysis of options markets and stock markets, and we have to keep these in mind when interpreting our results. Most importantly, even though an analysis of the effect of illiquidity on stock returns also needs to control for other return determinants, in the case of option returns an overriding concern is that the return on the underlying is the first-order determinant of option returns (see equation (2.2)). As explained above, we control for this in our empirical work in Section 4 by either using the residuals from a regression on stock returns in our analysis, or alternatively by including stock returns in the regression. But it is worthwhile to investigate if our results are robust to an alternative empirical setup.

For equity options, an alternative approach is provided by the analysis of implied volatilities. This is interesting from two perspectives. First, the analysis of implied volatilities is well-established in the option literature. In fact, the importance of some of the control variables used in (4.3) was previously demonstrated in the context of the study of the structure of implied volatilities, see for instance Bakshi, Kapadia, and Madan (2003) and Duan and Wei (2009). Deuskar, Gupta, and Subrahmanyam (2011), who study the effect of liquidity on bond options, exclusively use implied volatilities as left-side variables, presumably because of potential problems with the analysis of returns. Second, because the structure of implied volatilities can simply be thought of as a (nonlinear) transformation of the structure of option prices, its analysis can be easily linked to the illiquidity literature, which often presents its arguments in terms of prices rather than returns. For example, Amihud (2002) investigates the hypothesis that higher expected liquidity raises expected returns, which lowers prices, assuming that liquidity does not affect corporate cash flows.

We therefore investigate whether option illiquidity affects the structure of implied volatilities. Following Duan and Wei (2009), we investigate several aspects of the implied volatility curve by first estimating the following model for each firm $i$ for each moneyness and maturity category used in Section 4

$$iv_{i,t}(\chi_k, T_k) = \kappa_{i,t} + \theta_{i,t}(\chi_k - \bar{\chi}_k) + \eta_{i,t}(T_k - \bar{T}_k) + u^k_{i,t}, \qquad k = 1, 2, ..., K \tag{5.1}$$

where $iv_{i.t}(\chi_k, T_k)$ is the implied volatility for option $k$ with maturity $T_k$ and moneyness $\chi_k$ defined as the strike price over the stock price at time $t$. To ensure that sufficient contracts are available, we run the regression every month. Implied volatility and option characteristics are provided by Ivy DB OptionMetrics. We include only months with more than ten contracts available. $\bar{T}_k$ and $\bar{\chi}_k$ are the average time to maturity and moneyness, respectively, for each category. Using these regressions, we obtain for each firm $i$ a monthly time series $\kappa_{i,t}$ which corresponds to the estimated level of implied volatility, and a monthly time series $\theta_{i,t}$ which corresponds to the estimated moneyness slope of the implied volatility. We define $\widetilde{\kappa}_{i,t}$ as the residuals plus the intercept from the cross-sectional regression of $\kappa_{i,t}$ on the monthly volatility, estimated by the square root of the sum of squared daily returns for the month. This is needed in order to eliminate the first-order determinant of implied volatility, similar to the use of adjusted option returns in Section 4.

Using both call and put contracts, we first estimate the illiquidity impact on the level of implied volatility by running monthly cross-sectional Fama-MacBeth regressions for the following model

$$\widetilde{\kappa}_{i,t} = a_{0,t} + a_{1,t}IL^O_{i,t} + a_{2,t}IL^S_{i,t} + a_{3,t}R^S_{i,t} + a_{4,t}b_{i,t} + a_{5,t}\ln(size_{i,t}) + a_{6,t}lev_{i,t} + \varepsilon^k_{i,t} \tag{5.2}$$

where $R^S_{i,t}$ is the firm's stock return, $IL^O_{i,t}$ is the average for the month of daily option illiquidity, and $IL^S_{i.t}$ is the dollar-volume weighted average of daily stock illiquidity, respectively. The proportion of systematic risk averaged throughout the month is denoted by $b_{i,t}$ and defined as in Duan and Wei (2009). To capture size we use the last daily observation each month and to capture leverage we use the observation available in the previous quarter. As in Section 4, the regression is run using all firms available for a given moneyness/maturity category.

Table 8 presents the results of this approach for calls and puts respectively. Option illiquidity $IL_t^O$ negatively affects the level of implied volatility at the 1% significance level across all moneyness and maturity categories. These results are consistent with the positive predictive impact of option illiquidity $IL_{t-1}^O$ on option returns in Tables 4 and 5. An increase in illiquidity decreases current prices, and therefore also the level of implied volatility, and increases expected option returns. Moreover, stock illiquidity, $IL_t^S$, has a positive and significant impact on the level of implied volatility, which is also consistent with the results in Tables 4 and 5, and with a hedging argument. An increase in stock illiquidity facilitates trading in options to hedge long/short positions in more illiquid stocks. This causes an increase in contemporaneous option prices, i.e. the level of implied volatility. These findings suggest illiquidity spillovers between stock and option markets. Overall, we observe a strong and statistically significant effect of both option illiquidity and stock illiquidity on the level of implied volatility across all option categories. The robustness of this effect across all option categories suggests a systematic impact of illiquidity on option prices.

We next examine the effect of option illiquidity on the moneyness slope of the implied volatility curve. It is well known that the data exhibit a smile or smirk in the moneyness dimension, implying that the slope is sometimes negative and sometimes positive. We test the hypothesis that illiquidity increases the absolute value of the slope by estimating

$$|\theta_{i,t}| = c_{0,t} + c_{1,t}IL_{i,t}^O + c_{2,t}IL_{i,t}^S + c_{3,t}R_{i,t}^S + c_{4,t}b_{i,t} + c_{5,t}\ln(size_{i,t}) + c_{6,t}lev_{i,t} + \varepsilon_{i,t}^k \qquad (5.3)$$

where $\theta_{i,t}$ is obtained from equation (5.1).

Table 9 reports the regression results. Option illiquidity $IL_t^O$ significantly increases the implied volatility moneyness slope, while the effect of stock illiquidity $IL_t^S$ is less robust.

Overall, the results of Tables 8 and 9 suggest that option illiquidity is an important determinant of the structure of implied volatilities. Stock illiquidity is also an important determinant of the level and slope of implied volatility.

# 6 Option Illiquidity: Time Series Evidence

The cross-sectional results in Sections 4 and 5 provide substantial evidence of the importance of both option and stock illiquidity for option returns at the firm level. We now present time series evidence for portfolios. Portfolio-level time-series evidence can potentially yield additional insights as firm-specific risks are largely diversified away in this case. We use the time-series framework of French, Schwert, and Stambaugh (1987) and Amihud (2002).

Conducting portfolio-level time series analyses is more involved for options than for stocks. We proceed as follows. For each firm $i$ and for each period $t$, we compute the option return $R^O_{i,t}$ as the equally-weighted average return in (3.1) for all eligible contracts available at time $t$. For these contracts we also compute the average illiquidity at time $t$, denote it $IL_{i,t}$. Some contracts are not available at both times $t-1$ and time $t$, due to the data filters. Therefore, the illiquidity at time $t-1$ is only computed for the option contracts with quotes available to compute their returns at time $t$, which ensures that $R^O_{i,t}$, $IL_{i,t}$ and $IL_{i,t-1}$ are based on the same contracts. As in the cross-sectional regressions, we adjust option returns for variation in the price and volatility of the underlying stock. We do so by regressing the raw option return on the current and lagged stock return and squared stock return and using the residuals from this regression instead of the raw option returns.

Then, as in Section 4.6, we rank the returns into deciles based on $IL_{i,t-1}$, and for each decile, we compute the equally weighted average of $R^O_{i,t}$, $IL_{i,t}$ and $IL_{i,t-1}$. We also compute the equally weighted average of lagged stock illiquidity $IL^S_{i,t-1}$ for each decile portfolio.

We are interested in the time series dynamics of the effect of option illiquidity on option returns at the portfolio level. Following the methodology in Amihud (2002), we test the predictive power of option illiquidity on option returns as well as the effect of a contemporaneous and unexpected shock to option illiquidity on option returns. We estimate the illiquidity shock of each decile, $j$, in the following time series regression in logarithms, using weekly data:

$$\ln\left(IL_{j,t}\right) = \omega_{j,0} + \omega_{j,1}\ln\left(IL_{j,t-1}\right) + v^{IL}_{j,t} \tag{6.1}$$

We use the residuals from this regression as a proxy for unexpected shocks to option illiquidity, defined as $IL^u_{j,t} \equiv v^{IL}_{j,t}$. The effect of option illiquidity on option returns is subsequently estimated for each decile portfolio using the following regression

$$\widetilde{R}^O_{j,t} = \gamma_0 + \gamma_1\ln\left(IL'_{j,t-1}\right) + \gamma_2 IL^u_{j,t} + \gamma_3\ln\left(IL^S_{j,t-1}\right) + v_{j,t} \tag{6.2}$$

Based on our cross-sectional findings, we expect $\gamma_1$ to be positive and significant. Moreover, we expect $\gamma_1$ to monotonically increase from less illiquid to more illiquid portfolios since we expect the illiquidity impact to be higher for more illiquid assets, similar to Amihud's (2002) findings for stocks. Given that lagged illiquidity has a positive impact, the contemporaneous unexpected shock should have a negative effect on option returns, i.e. an unexpected positive illiquidity shock should decrease current option prices and thus increase expected option returns. Similar to the evidence on the impact of illiquidity on stock market returns (Amihud, 2002), we also expect the effect of $\gamma_2$ to be stronger, i.e. more negative, for more illiquid portfolios.

Finally, the expected sign for the effect of stock illiquidity on expected option returns can be motivated by the discussion in Cetin et al (2006). A positive illiquidity shock in the stock market increases the cost of the replicating portfolio and therefore increases the current option price. Since options become more expensive for more illiquid stocks, the expected return on these options decrease. Moreover, they decrease more for more illiquid options. We therefore expect $\gamma_3$ to be negative.

The illiquidity portfolio level results for all option categories are reported in Table 10. The results are more pronounced for the call options reported in Panel A. For both short-term and long-term OTM calls, $\gamma_1$ is positive and significant and increases with portfolio illiquidity. This coefficient is higher for short-term contracts, suggesting higher illiquidity premia for short-term calls. The unexpected illiquidity has a significantly negative effect on short-term OTM calls and the magnitude of this effect is monotonically increasing in portfolio illiquidity. This is similar to the effect of stock illiquidity on stock returns (Amihud, 2002). The unexpected illiquidity shock is only significant for high-illiquidity portfolios and for long-term OTM calls but its coefficient has the expected negative sign and monotonically increases in absolute value with portfolio illiquidity. The results on option illiquidity are qualitatively similar across ATM and ITM short-term and long-term contracts, but the effect is more pronounced for short-term contracts. This is consistent with Amihud and Mendelson's (1986) clientele effect, where the holders of longer term assets are able to amortize illiquidity costs due to longer holding periods and thus require lower compensation for bearing illiquidity costs. This assumes of course that long-term options are indeed held for longer periods on average.

We obtain similar results for put options in Panel B, but with the exception of OTM put options, the results for call options are stronger in terms of magnitude and significance than for put options. Even though the two are linked via put-call parity, for ATM and ITM contracts, call options appear to be more exposed to illiquidity in the option and stock markets.

Finally, we find that stock illiquidity has a negative and significant impact on expected option returns across both calls and puts and for different maturity and moneyness categories. The pattern of $\gamma_3$ across illiquidity portfolios is not monotone. It is higher in absolute value for medium-illiquidity portfolios and lower for extreme decile portfolios. This suggests that even though stock illiquidity does affect option returns, it represents a different type of risk than option illiquidity.

# 7 Conclusion

We present evidence on illiquidity premia in equity option markets. Using cross-sectional and time series evidence, we find an economically and statistically significant positive impact of option illiquidity on expected option returns. The cross-sectional results obtain in univariate regressions, as well as in multivariate regressions controlling for returns and volatility of the underlying equity, lagged option returns, and a variety of other variables. The results are robust across six different moneyness and maturity categories, and estimates obtained using the cross-section of implied volatilities confirm the positive impact of option illiquidity on option returns. Our results are similar to the findings of Amihud (2002), who reports a positive effect of stock illiquidity on stock returns. A shock to option illiquidity decreases the current price and increases expected option returns, thus compensating traders for holding illiquid contracts.

The illiquidity of the underlying stocks also has an economically significant impact on option returns. A positive shock to stock illiquidity increase current option prices and decreases expected option returns. This effect is consistent with an increase in hedging trades due to higher stock illiquidity: Whenever stock market illiquidity increases, the higher stock transaction costs increase the cost of replicating the option, which in turn increases the option price and reduces its expected return.

# References


[1] Acharya, V. and L. Pedersen, (2005), Asset Pricing with Liquidity Risk, Journal of Financial Economics, 77, 375-410.

[2] Amihud, Y. and H. Mendelson, (1986), Asset Pricing and the Bid-Ask Spread, Journal of Financial Economics, 17, 223-249.

[3] Amihud, Y. and H. Mendelson, (1989), The Effect of Beta, Bid-Ask Spread, Residual Risk, and Size on Stock Returns, Journal of Finance, 2, 479-486.

[4] Amihud, Y. and H. Mendelson, (1991), Liquidity, Maturity, and the Yields on U.S. Treasury Securities, Journal of Finance, 46, 1411-1425.

[5] Amihud, Y., (2002), Illiquidity and Stock Returns: Cross-Section and Time-Series Effects, Journal of Financial Markets, 5, 31-56.

[6] Bakshi, G., C. Cao and Z. Chen (1997), Empirical Performance of Alternative Option Pricing Models, Journal of Finance, 52, 2003-2049.

[7] Bakshi, G., N. Kapadia and D. Madan, (2003), Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options, Review of Financial Studies, 16, 101-143.

[8] Battalio, R., B. Hatch and R. Jennings, (2004), Toward a National Market System for U.S. Exchange-listed Equity Options, Journal of Finance, 59, 933-962.

[9] Beber, A., M. Brandt and K. Kavajecz, (2009), Flight-to-Quality or Flight-to-Liquidity? Evidence from the Euro-Area Bond Market, Review of Financial Studies, 22, 925-957.

[10] Black, F. and M. Scholes, (1973), The Pricing of Options and Corporate Liabilities, Journal of Political Economy, 81, 637-654.

[11] Bollen, N. and R. Whaley, (2004), Does Net Buying Pressure Affect the Shape of Implied Volatility Functions?, Journal of Finance, 59, 711–53.

[12] Bongaerts, D., F. De Jong and J. Driessen, (2010), Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market, Journal of Finance, forthcoming.

[13] Boudoukh, J. and R. Whitelaw, (1993), Liquidity as a Choice Variable: a Lesson from the Japanese Government Bond Market, Review of Financial Studies, 6, 265-292.

[14] Brennan, M. and A. Subrahmanyam, (1996), Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns, Journal of Financial Economics, 41, 441-464.

[15] Brenner, M., R. Eldor and S. Hauser, (2001), The Price of Options Illiquidity, Journal of Finance, 56, 789-805.

[16] Broadie, M., M. Chernov and M. Johannes, (2009), Understanding Index Option Returns, Review of Financial Studies, 22, 4493-4529.

[17] Cao, M. and J. Wei (2010), Commonality in Liquidity: Evidence from the Option Market, Journal of Financial Markets, 13, 20–48.

[18] Carhart, M., (1997), On Persistence of Mutual Fund Performance, Journal of Finance, 52, 57-82.

[19] Cetin, U., R. Jarrow, P. Protter and M. Warachka, (2006), Pricing Options in an Extended Black Scholes Economy with Illiquidity: Theory and Empirical Evidence, Review of Financial Studies, 19, 493–529.

[20] Cetin, U., R. Jarrow and P. Protter, (2004), Liquidity Risk and Arbitrage Pricing Theory, Finance and Stochastics, 8, 311–341.

[21] Cho, Y.-H., and R. Engle, (1999), Modeling the Impacts of Market Activity on Bid-Ask Spreads in the Option Market, NBER Working Paper 7331.

[22] Chordia, T., R. Roll and A. Subrahmanyam, (2000), Commonality in Liquidity, Journal of Financial Economics, 56, 3–28.

[23] Chordia, T., R. Roll and A. Subrahmanyam, (2001), Market Liquidity and Trading Activity, Journal of Finance, 56, 501–530.

[24] Chordia, T., A. Sarkar and A. Subrahmanyam, (2005), An Empirical Analysis of Stock and Bond Market Liquidity, Review of Financial Studies, 18, 85–129.

[25] Constantinides, G., J. Jackwerth and S. Perrakis (2009), Mispricing of S&P 500 Index Options, Review of Financial Studies, 22, 1247-1277.

[26] Constantinides, G. and S. Perrakis, (2002), Stochastic Dominance Bounds on Derivative Prices in a Multiperiod Economy with Proportional Transaction Costs, Journal of Economic Dynamics and Control, 26, 1323-1352.

[27] Constantinides, G. and S. Perrakis, (2007), Stochastic Dominance Bounds on American Option Prices in Markets with Frictions, Review of Finance, 11, 71-115.

[28] Coval, J. and T. Shumway, (2001), Expected Option Returns, Journal of Finance, 56, 983-1009.

[29] Cox, J., S. Ross and M. Rubinstein, (1979), Option Pricing: A Simplified Approach, Journal of Financial Economics, 7, 229-263.

[30] Dennis, P. and S. Mayhew, (2002), Risk-neutral Skewness: Evidence from Stock Options, Journal of Financial and Quantitative Analysis, 37, 471–93.

[31] Deuskar, P., A. Gupta and M. Subrahmanyam, (2011), Liquidity Effect in OTC Options Markets: Premium or Discount?, Journal of Financial Markets, 14, 127-160.

[32] Driessen, J., P. Maenhout and G. Vilkov, (2009), The Price of Correlation Risk: Evidence from Equity Options, Journal of Finance, 64, 1377–1406.

[33] Duan, J.-C and J. Wei, (2009), Systematic Risk and the Price Structure of Individual Equity Options, Review of Financial Studies, 22, 1981-2006.

[34] Duarte, J. and C. Jones, (2007), The Price of Market Volatility Risk, Working Paper, University of Southern California.

[35] Easley, D., M. O'Hara and P. Srinivas, (1998), Option Volume and Stock Prices: Evidence on Where Informed Traders Trade, Journal of Finance, 53, 432–465.

[36] Eleswerapu, V. and M. Reinganum, (1993), The Seasonal Behavior of the Liquidity Premium in Asset Pricing, Journal of Financial Economics, 34, 281-305.

[37] Fama, E. and K. French, (1993), Common Risk Factors in the Returns on Stocks and Bonds, Journal of Financial Economics, 33, 3-56.

[38] Fama, E. and J. MacBeth, (1973), Risk, Return, and Equilibrium: Empirical Tests, Journal of Political Economy, 71, 607-636.

[39] French, K., G. Schwert and R. Stambaugh, (1987), Expected Stock Returns and Volatility, Journal of Financial Economics, 19, 3-29.

[40] Garleanu, N., L. Pedersen and A. Poteshman, (2009), Demand-Based Option Pricing, Review of Financial Studies, 22, 4259-4299.

[41] George, T. and F. Longstaff, (1993), Bid–Ask Spreads and Trading Activity in the S&P 100 Index Options Market, Journal of Financial and Quantitative Analysis, 28, 381–397.

[42] Goldreich, D., B. Hanke and P. Nath (2005), The Price of Future Liquidity: Time-Varying Liquidity in the U.S. Treasury Market, Review of Finance, 9, 1-32.

[43] Goyal, A. and A. Saretto, (2009), Cross-Section of Option Returns and Volatility, Journal of Financial Economics, 94, 310-326.

[44] Goyenko, R., Holden, C. and C. Trzcinka, (2009), Do Liquidity Measures Measure Liquidity? Journal of Financial Economics, 92, 153-181

[45] Hasbrouck, J. and D. Seppi, (2001), Common Factors in Prices, Order Flows, and Liquidity, Journal of Financial Economics, 59, 383–411.

[46] Heston, S., (1993), A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, Review of Financial Studies, 6, 327-343.

[47] Huberman, G. and D. Halka, (2001), Systematic Liquidity, Journal of Financial Research, 2, 161–178.

[48] Hull, J. and A. White, (1987), The Pricing of Options on Assets with Stochastic Volatility, Journal of Finance, 42, 281-300.

[49] Jarrow, R. and P. Protter, (2005), Liquidity Risk and Risk Measure Computation, Working Paper, Cornell University.

[50] Jones, C., (2002), A Century of Stock Market Liquidity and Trading Costs, working paper, Columbia University.

[51] Jones, C., (2006), A Nonlinear Factor Analysis of S&P 500 Index Option Returns, Journal of Finance, 61, 2325–2363.

[52] Kamara, A., (1994), Liquidity, Taxes, and Short-Term Treasury Yields, Journal of Financial and Quantitative Analysis, 29, 403-417.

[53] Krishnamurthy, A., (2002), The Bond/Old-Bond Spread, Journal of Financial Economics, 66, 463-506.

[54] Lakonishok, J., I. Lee and A. Poteshman, (2007), Option Market Activity, Review of Financial Studies, 20, 813-857.

[55] Longstaff, F., (2004), The Flight-to-Liquidity Premium in U.S. Treasury Bond Prices, Journal of Business, 77, 511-526.

[56] Mayhew, S., (2002), Competition, Market Structure, and Bid-Ask Spreads in Stock Option Markets, Journal of Finance, 57, 931-958.

[57] Newey, W. and K. West, (1987), A Simple, Positive Semi-definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, 55, 703-708.

[58] Oancea, I. and S. Perrakis, (2007), Stochastic Dominance Option Pricing in Discrete and Continuous Time: an Alternative Paradigm, Working Paper, Concordia University.

[59] Pan, J., and A. Poteshman, (2006), The Information in Option Volume for Future Stock Prices, Review of Financial Studies, 19, 871-908.

[60] Pastor, L. and R. Stambaugh, (2003), Liquidity Risk and Expected Stock Returns, Journal of Political Economy, 113, 642-685.

[61] Roll, R., E. Schwartz, and A. Subrahmanyam, (2010), O/S: The Relative Trading Activity in Options and Stock, Journal of Financial Economics, 96, 1-17.

[62] Scott, L., (1987), Option Pricing when the Variance Changes Randomly: Theory, Estimators and Applications, Journal of Financial and Quantitative Analysis, 22, 419-438.

[63] Toft, K., (1996), On the Mean-Variance Tradeoff in Option Replication with Transactions Costs, Journal of Financial and Quantitative Analysis, 31, 233-263.

[64] Vijh, A., (1990), Liquidity of the CBOE Equity Options, Journal of Finance, 45, 1157-1179.

[65] Warga, A., (1992), Bond Returns, Liquidity, and Missing Data, Journal of Financial and Quantitative Analysis, 27, 605-617.

## Figure 1A
## Daily Call Option Returns

We plot daily returns on equally-weighted portfolios of call options. Option returns are computed from closing bid-ask price midpoints. For call options, OTM (out-of-the-money) corresponds to $0.125 < \Delta \leq 0.375$, ATM (at-the-money) corresponds to $0.375 < \Delta \leq 0.625$, and ITM (in-the-money) corresponds to $0.625 < \Delta \leq 0.875$. Short-term options have maturities between 20 and 70 days, whereas long-term options have maturities between 71 and 180 days. The option data are from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

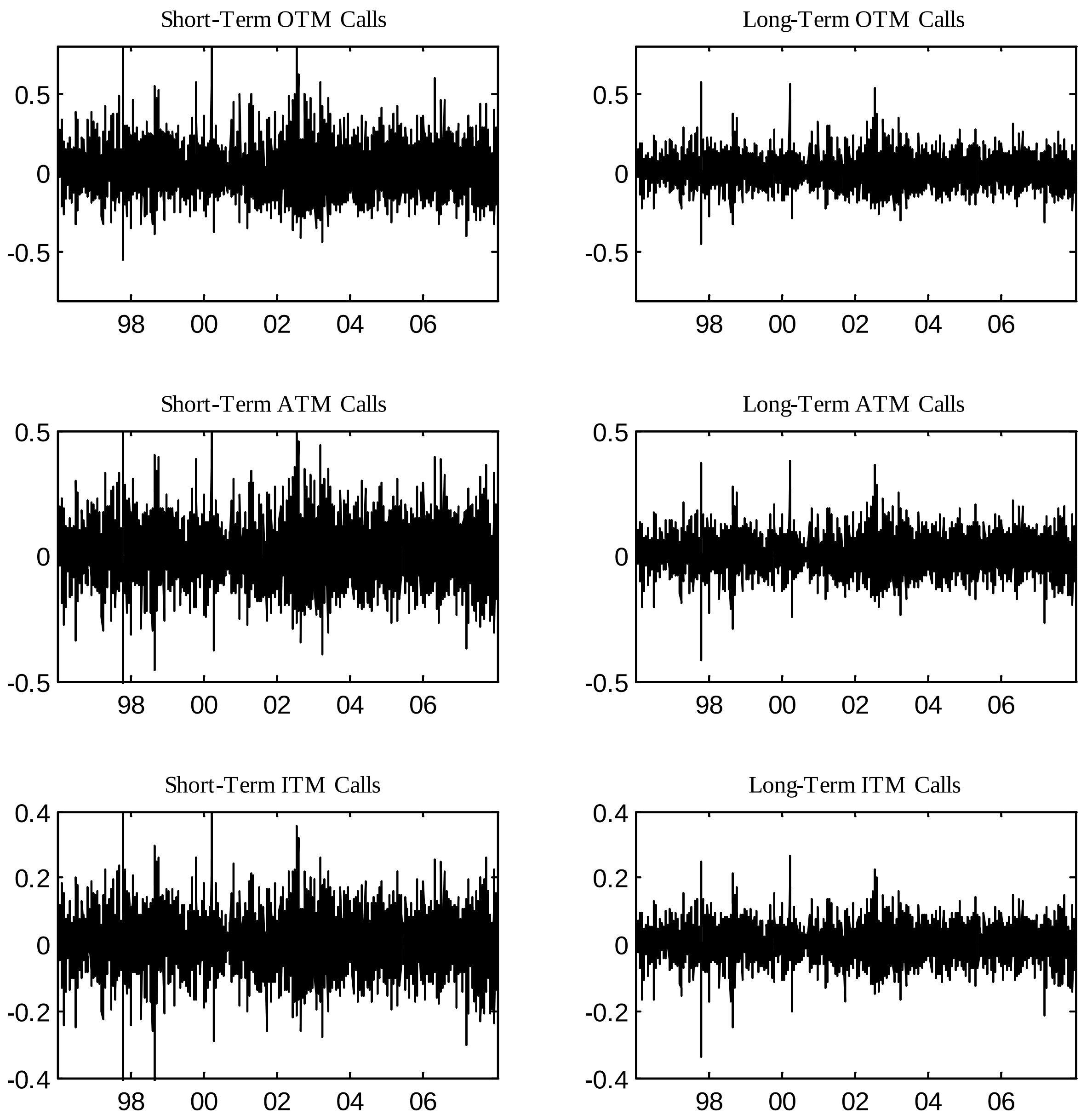
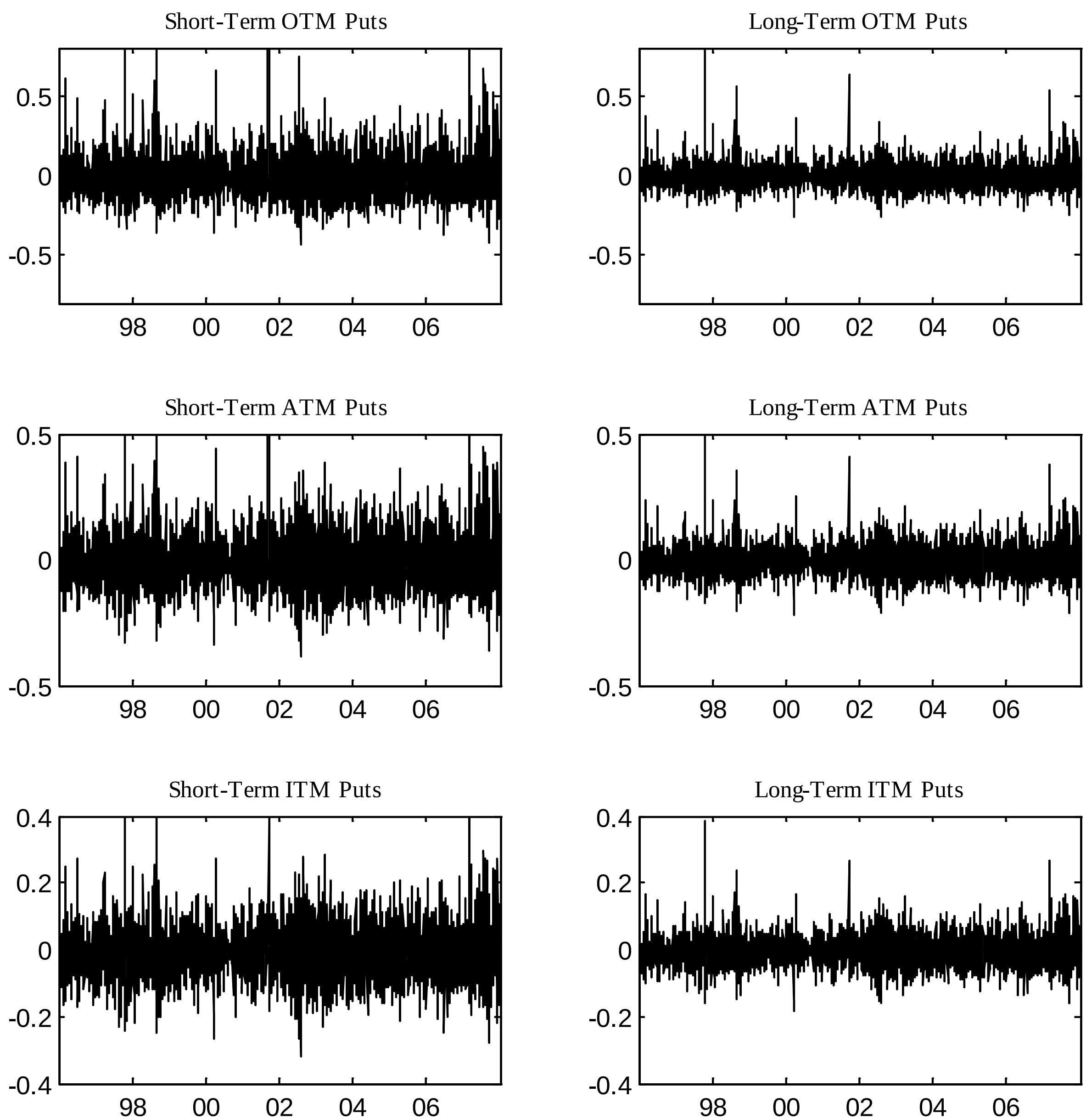


**Figure 1B**
**Daily Put Option Returns**

We plot daily returns on equally-weighted portfolios of put options. Option returns are computed from closing bid-ask price midpoints. For put options, OTM (out-of-the-money) corresponds to $-0.375 < \Delta \leq -0.125$, ATM (at-the-money) corresponds to $-0.625 < \Delta \leq -0.375$, and ITM (in-the-money) corresponds to $-0.875 < \Delta < -0.625$. Short-term options have maturities between 20 and 70 days, whereas long-term options have maturities between 71 and 180 days. The option data are from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.



## Figure 2A
## Aggregate Illiquidity for Call Options

We plot aggregate daily illiquidity measures for call options. The illiquidity measure is based on the average relative bid-ask spread, where ask and bid are end-of-day closing quoted ask and bid prices available from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

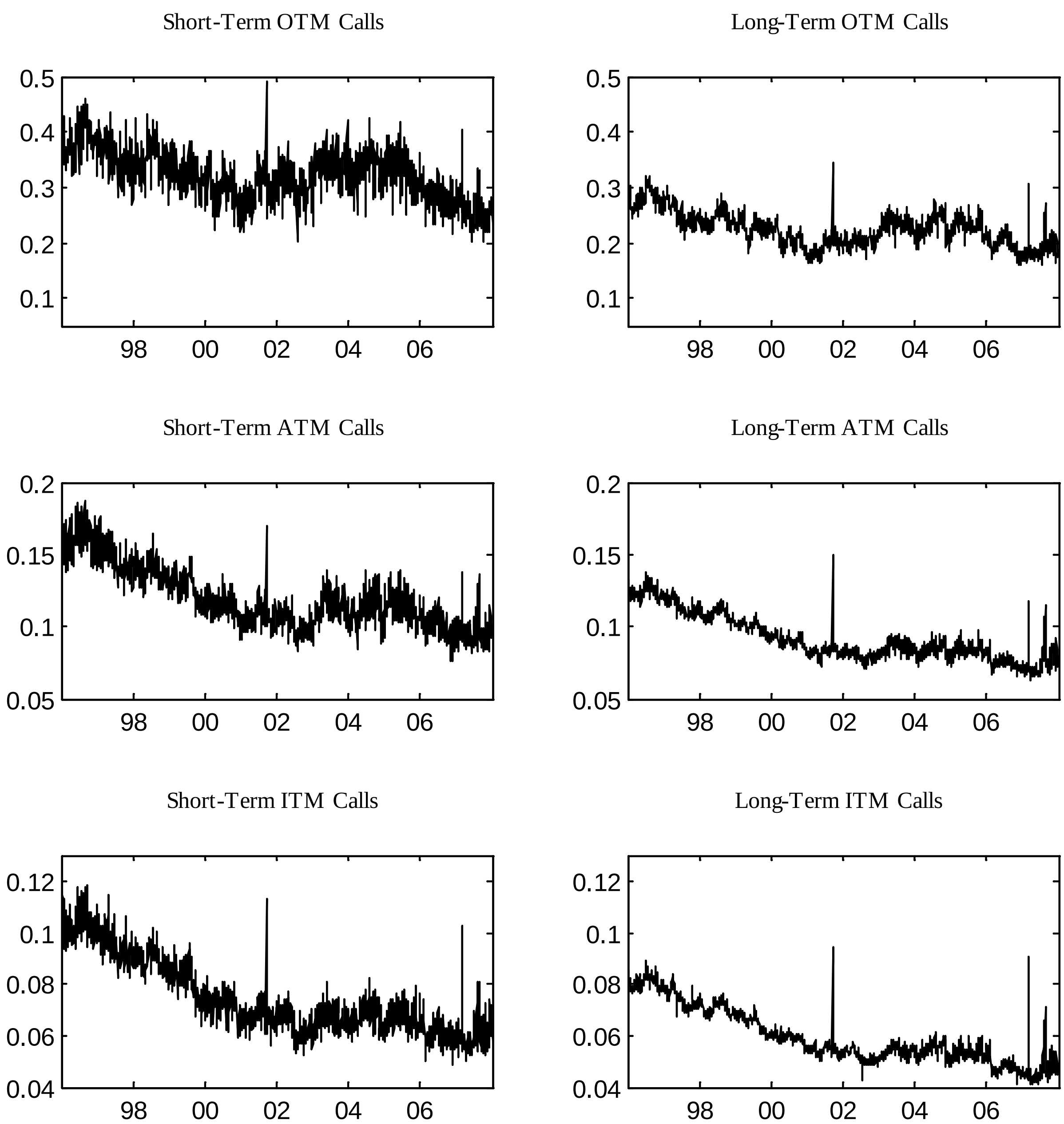


**Figure 2B**
**Aggregate Illiquidity for Put Options**

We plot aggregate daily illiquidity measures for put options. The illiquidity measure is based on the average relative bid-ask spread, where ask and bid are end of day closing quoted ask and bid prices available from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

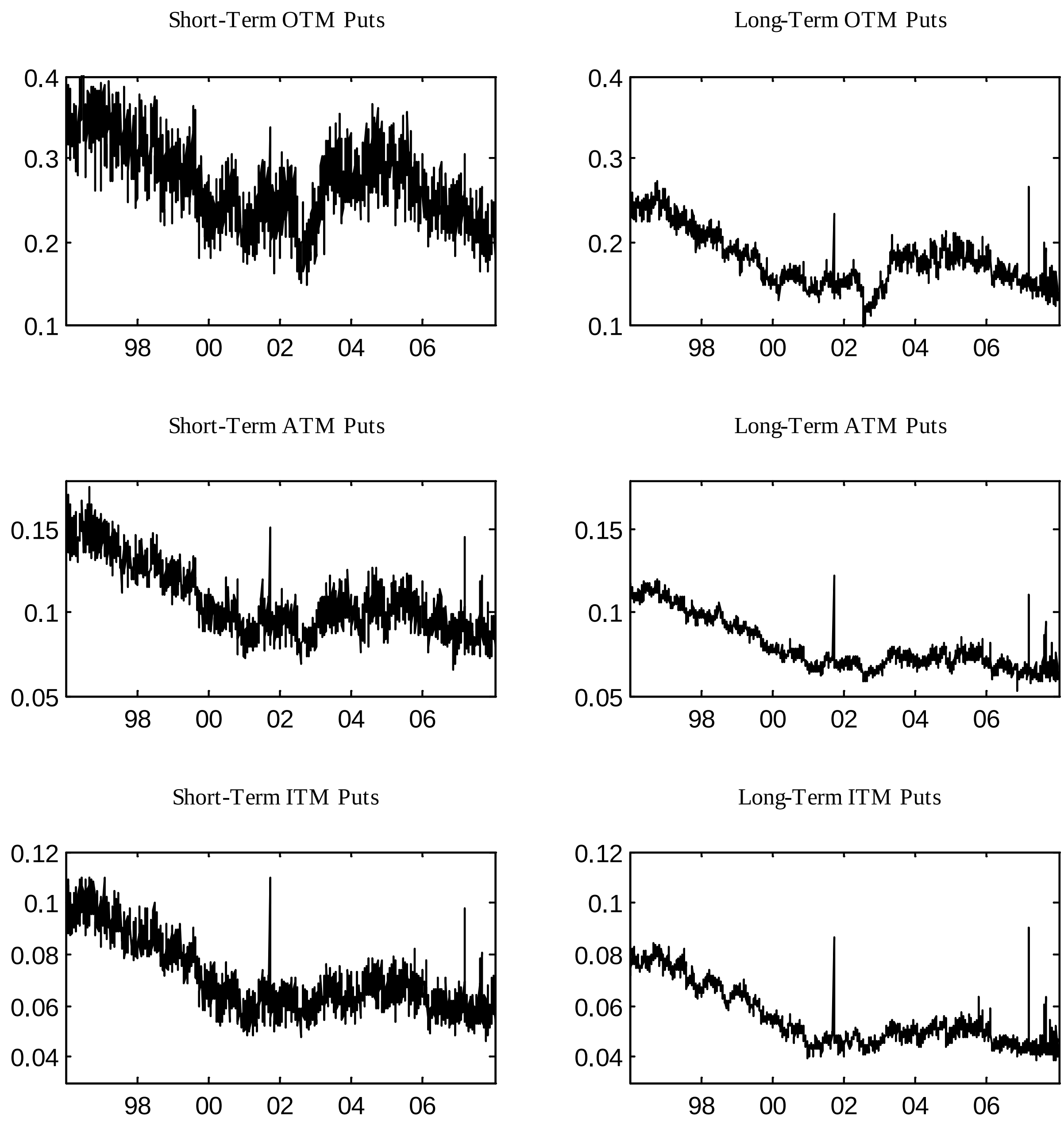
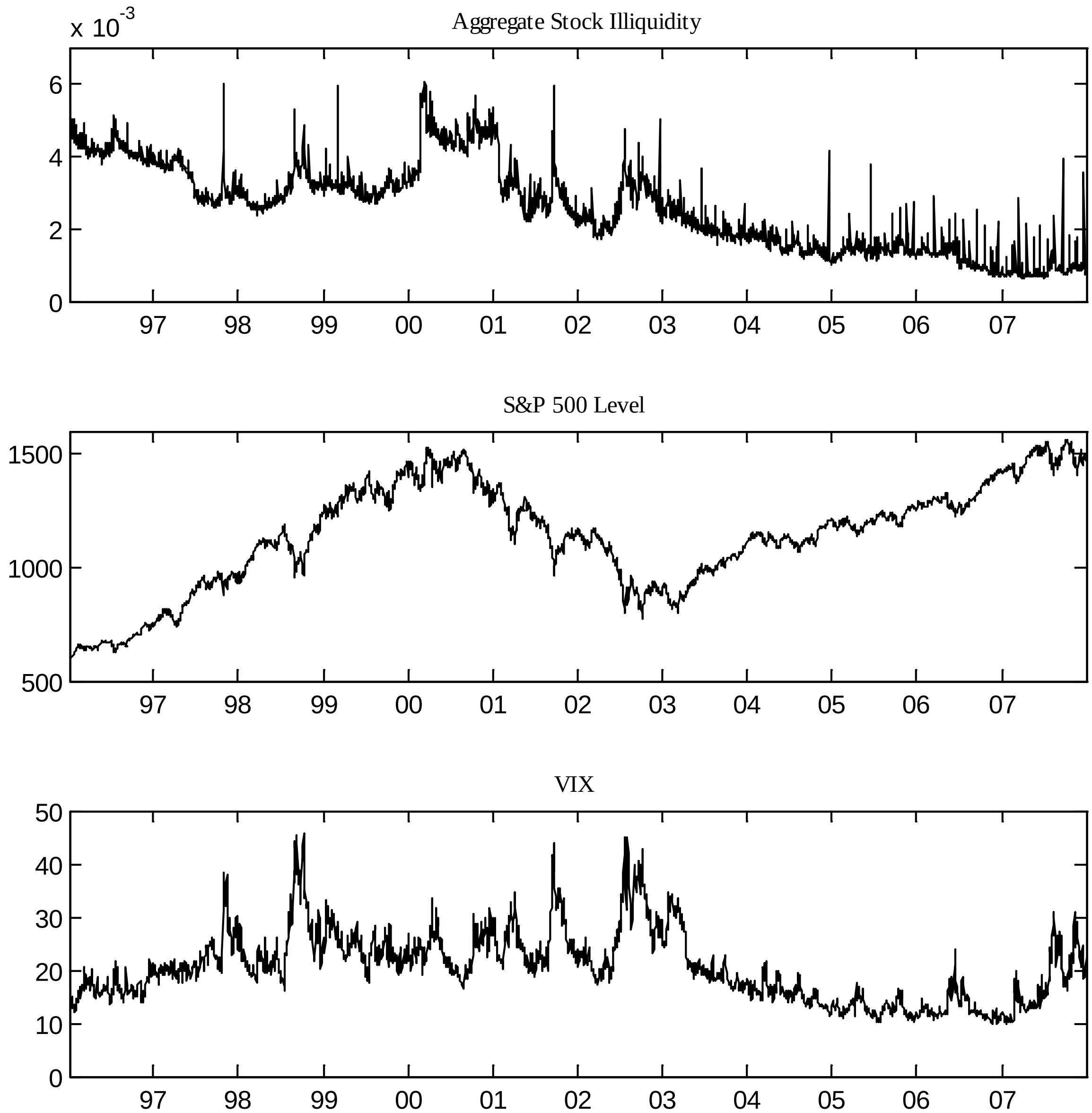


**Figure 3**
**Aggregate Stock Illiquidity, S&P 500 Index, and VIX**

We plot the aggregate stock illiquidity measure, the level of S&P 500 index, and the VIX. Stock illiquidity is estimated from TAQ (Trade and Quote) intra-day data as the dollar-volume-weighted average of effective relative spreads for each day. The sample period is from January 1996 through December 2007.



## Table 1
## Descriptive Statistics

We provide descriptive statistics for daily and weekly option returns. First we compute the descriptive statistics for each firm and then we take the cross-sectional averages of these statistics. We report the mean (in percentages), the standard deviation, the skewness, the kurtosis, the first-order autocorrelation of returns $\rho(1)$, and the first-order autocorrelation of absolute value of returns, $\rho^{abs}(1)$. The option returns are computed using closing bid-ask price midpoints. OTM (out-of-the-money) corresponds to $0.125 < \Delta \leq 0.375$ for calls and $-0.375 < \Delta \leq -0.125$ for puts. ATM (at-the-money) corresponds to $0.375 < \Delta \leq 0.625$ for calls and $-0.625 < \Delta \leq -0.375$ for puts. ITM (in-the-money) corresponds to $0.625 < \Delta \leq 0.875$ for calls and $-0.875 < \Delta \leq -0.625$ for puts. Short-term options have maturity between 20 and 70 days, whereas long-term options have maturity between 71 and 180 days. The option data are from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

| | Short-term | | | Long-term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| Daily Call Returns | | | | | | |
| Mean | 1.740 | 0.496 | 0.173 | 0.864 | 0.414 | 0.258 |
| Std | 0.361 | 0.256 | 0.177 | 0.217 | 0.157 | 0.110 |
| Skew | 2.941 | 1.347 | 0.742 | 1.940 | 1.107 | 0.667 |
| Kurt | 34.621 | 10.487 | 8.044 | 22.097 | 12.518 | 12.997 |
| $\rho(1)$ | -0.009 | -0.005 | -0.011 | -0.013 | -0.009 | -0.009 |
| $\rho^{abs}(1)$ | 0.061 | 0.065 | 0.082 | 0.087 | 0.093 | 0.105 |
| Avr. Nb. Firms | 221 | 213 | 236 | 262 | 291 | 300 |
| Daily Put Returns | | | | | | |
| Mean | -0.137 | -0.460 | -0.643 | -0.138 | -0.199 | -0.229 |
| Std | 0.313 | 0.227 | 0.166 | 0.175 | 0.129 | 0.098 |
| Skew | 4.039 | 1.466 | 0.645 | 3.555 | 1.411 | 0.748 |
| Kurt | 74.910 | 17.852 | 9.315 | 86.911 | 22.639 | 14.844 |
| $\rho(1)$ | -0.001 | -0.002 | -0.010 | 0.001 | 0.000 | -0.002 |
| $\rho^{abs}(1)$ | 0.065 | 0.065 | 0.073 | 0.086 | 0.087 | 0.087 |
| Avr. Nb. Firms | 232 | 199 | 194 | 301 | 274 | 227 |
| Weekly Call Returns | | | | | | |
| Mean | 10.020 | 2.572 | 0.172 | 4.287 | 1.975 | 1.148 |
| Std | 0.887 | 0.581 | 0.380 | 0.497 | 0.344 | 0.238 |
| Skew | 3.213 | 1.616 | 0.803 | 2.257 | 1.248 | 0.658 |
| Kurt | 22.379 | 8.766 | 5.273 | 16.497 | 8.249 | 5.537 |
| $\rho(1)$ | 0.003 | -0.008 | -0.022 | -0.011 | -0.025 | -0.035 |
| $\rho^{abs}(1)$ | 0.019 | 0.013 | 0.020 | 0.032 | 0.031 | 0.039 |
| Avr. Nb. Firms | 244 | 243 | 260 | 280 | 306 | 312 |
| Weekly Put Returns | | | | | | |
| Mean | -0.874 | -3.585 | -4.627 | -1.072 | -1.363 | -1.572 |
| Std | 0.739 | 0.508 | 0.357 | 0.398 | 0.284 | 0.212 |
| Skew | 3.668 | 1.713 | 0.796 | 2.980 | 1.354 | 0.722 |
| Kurt | 31.715 | 10.998 | 5.584 | 31.259 | 10.487 | 6.780 |
| $\rho(1)$ | 0.015 | 0.011 | 0.002 | 0.005 | -0.006 | 0.003 |
| $\rho^{abs}(1)$ | 0.034 | 0.023 | 0.014 | 0.052 | 0.039 | 0.041 |
| Avr. Nb. Firms | 252 | 229 | 220 | 311 | 290 | 245 |

## Table 2
## Illiquidity Measures

We present summary statistics for the illiquidity measures in percentages (in Panel A) and the correlations between the illiquidity measures for call and put options (in Panels B and C respectively). Stock illiquidity $IL^S$ is estimated from TAQ (Trade and Quote) intra-day data as the dollar-volume weighted average of the effective relative spread for each day. The option illiquidity measure $IL^O$ is based on the average relative bid-ask spread, where ask and bid are end-of-day closing quoted ask and bid prices available from Ivy DB OptionMetrics. For each firm and for each day, we compute the average of the relative bid-ask spreads of all the available options in a given category, and then we take the mean, the minimum, the maximum, the standard deviation and the first-order autocorrelation, $\rho(1)$, of these averages. We report the cross-sectional averages of these statistics in Panel A. We compute the cross-sectional correlations between the illiquidity measures on each day and report the time-series averages of these correlations in Panel B for call options and Panel C for put options. The option data are from Ivy DB OptionMetrics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

### Panel A: Descriptive Statistics

| | Options | | | | | | Stocks |
|---|---|---|---|---|---|---|---|
| | Short-term | | | Long-term | | | |
| | OTM | ATM | ITM | OTM | ATM | ITM | |
| $IL_t^O$ for call options | | | | | | | $IL_t^S$ |
| Mean | 34.02 | 12.97 | 7.86 | 22.67 | 9.48 | 6.05 | 0.26 |
| Min | 3.89 | 2.18 | 1.56 | 2.74 | 1.78 | 1.23 | 0.03 |
| Max | 123.40 | 48.33 | 30.85 | 91.45 | 35.35 | 22.54 | 3.85 |
| Std | 19.10 | 6.10 | 3.32 | 12.25 | 3.96 | 2.33 | 0.21 |
| $\rho(1)$ | 0.61 | 0.65 | 0.69 | 0.70 | 0.69 | 0.76 | 0.58 |
| $IL_t^O$ for put options | | | | | | | |
| Mean | 29.21 | 11.72 | 7.46 | 18.12 | 8.23 | 5.52 | |
| Min | 3.62 | 2.07 | 1.54 | 2.55 | 1.59 | 1.18 | |
| Max | 115.51 | 45.09 | 29.03 | 76.15 | 31.53 | 20.93 | |
| Std | 16.78 | 5.52 | 3.12 | 9.60 | 3.44 | 2.23 | |
| $\rho(1)$ | 0.67 | 0.69 | 0.68 | 0.77 | 0.74 | 0.75 | |

### Panel B: Correlation Matrix for Call Options

| | | $IL_t^O$ Short-term | | | $IL_t^O$ Long-term | | |
|---|---|---|---|---|---|---|---|
| | | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | ATM | 0.66 | | | | | |
| Short-term | ITM | 0.58 | 0.65 | | | | |
| $IL_t^O$ | OTM | 0.58 | 0.62 | 0.60 | | | |
| Long-term | ATM | 0.59 | 0.67 | 0.67 | 0.67 | | |
| | ITM | 0.64 | 0.70 | 0.74 | 0.67 | 0.74 | |
| $IL_t^S$ | | 0.24 | 0.25 | 0.26 | 0.30 | 0.26 | 0.29 |

### Panel C: Correlation Matrix for Put Options

| | | $IL_t^O$ Short-term | | | $IL_t^O$ Long-term | | |
|---|---|---|---|---|---|---|---|
| | | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | ATM | 0.69 | | | | | |
| Short-term | ITM | 0.60 | 0.64 | | | | |
| $IL_t^O$ | OTM | 0.69 | 0.69 | 0.64 | | | |
| Long-term | ATM | 0.65 | 0.69 | 0.67 | 0.72 | | |
| | ITM | 0.64 | 0.67 | 0.70 | 0.70 | 0.75 | |
| $IL_t^S$ | | 0.23 | 0.21 | 0.20 | 0.24 | 0.22 | 0.19 |

## Table 3
## Descriptive Statistics for Volume, Open Interest and Amihud's Illiquidity Measure

We present summary statistics for option volume, open interest and Amihud's Illiquidity measure for call options (Panel A) and put options (Panel B). For each firm and each day, we compute the average volume and open interest for all available options in a given category, and then we take the mean, the minimum, the maximum, the standard deviation and the first-order autocorrelation $\rho(1)$ of these averages. We compute Amihud's illiquidity measure on a weekly basis for each firm, and we take the mean, the minimum, the maximum, the standard deviation and the first-order autocorrelation $\rho(1)$ of these weekly measures. We report the cross-sectional averages of these statistics. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

**Panel A: Call Options**

| | Short-term | | | Long-term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| Amihud's Illiquidity Measure | | | | | | |
| Mean | 3.527 | 0.736 | 0.309 | 1.815 | 0.352 | 0.165 |
| Min | 0.004 | 0.001 | 0.000 | 0.002 | 0.001 | 0.000 |
| Max | 105.010 | 22.586 | 8.231 | 49.345 | 8.256 | 3.583 |
| Std | 8.729 | 1.846 | 0.727 | 4.072 | 0.691 | 0.329 |
| $\rho(1)$ | 0.077 | 0.126 | 0.107 | 0.117 | 0.156 | 0.118 |
| Volume | | | | | | |
| Mean | 198 | 271 | 108 | 88 | 114 | 34 |
| Min | 0 | 0 | 0 | 0 | 0 | 0 |
| Max | 10222 | 11246 | 8385 | 7004 | 8824 | 5261 |
| Std | 516 | 633 | 364 | 294 | 365 | 171 |
| $\rho(1)$ | 0.20 | 0.24 | 0.17 | 0.14 | 0.15 | 0.12 |
| Open Interest | | | | | | |
| Mean | 2295 | 2418 | 1743 | 2493 | 2587 | 1483 |
| Min | 2 | 3 | 1 | 2 | 12 | 3 |
| Max | 31024 | 30416 | 27355 | 27964 | 29252 | 22486 |
| Std | 3586 | 3731 | 3164 | 3443 | 3541 | 2497 |
| $\rho(1)$ | 0.84 | 0.85 | 0.86 | 0.89 | 0.91 | 0.89 |

**Panel B: Put Options**

| | Short-term | | | Long-term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| Amihud's Illiquidity Measure | | | | | | |
| Mean | 2.492 | 0.644 | 0.313 | 1.117 | 0.296 | 0.152 |
| Min | 0.003 | 0.001 | 0.001 | 0.002 | 0.000 | 0.000 |
| Max | 76.468 | 18.916 | 7.684 | 28.185 | 6.808 | 2.641 |
| Std | 6.421 | 1.666 | 0.761 | 2.448 | 0.623 | 0.304 |
| $\rho(1)$ | 0.072 | 0.100 | 0.084 | 0.100 | 0.112 | 0.087 |
| Volume | | | | | | |
| Mean | 141 | 151 | 59 | 56 | 54 | 16 |
| Min | 0 | 0 | 0 | 0 | 0 | 0 |
| Max | 9639 | 8598 | 5750 | 6309 | 6064 | 3045 |
| Std | 433 | 441 | 242 | 233 | 231 | 99 |
| $\rho(1)$ | 0.19 | 0.19 | 0.14 | 0.12 | 0.13 | 0.10 |
| Open Interest | | | | | | |
| Mean | 1689 | 1537 | 1075 | 1809 | 1580 | 940 |
| Min | 2 | 1 | 1 | 7 | 2 | 1 |
| Max | 27047 | 25244 | 20654 | 23298 | 22106 | 16473 |
| Std | 2835 | 2725 | 2179 | 2757 | 2567 | 1811 |
| $\rho(1)$ | 0.86 | 0.84 | 0.83 | 0.91 | 0.91 | 0.87 |

# Table 4
# Fama-MacBeth Regressions for Daily Adjusted Option Returns

We report the results of cross-sectional Fama-MacBeth regressions for daily adjusted call and put option returns ($\widetilde{R}^O$), i.e. the residuals plus the intercept from the regression of option returns on stock returns, lagged stock returns and squared stock returns. We include the lagged values of the following regressors: option illiquidity $IL^O$, the illiquidity of the underlying asset $IL^S$, the conditional volatility, which is estimated using a GARCH(1,1) model, the systematic risk proportion $b$, which corresponds to the square root of the $R^2$ from the regression of stock returns on Fama-French and momentum factors, and the logarithm of size and firm leverage. The option illiquidity measure $IL^O$ is based on the average relative bid-ask spread, where ask and bid are end-of-day closing quoted ask and bid prices available from Ivy DB OptionMetrics. Stock illiquidity is obtained as the dollar-volume average of the effective relative spreads from TAQ. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007. The symbols *, † and ‡ denote, respectively, significance at the 10%, 5% and 1% levels using Fama-MacBeth t-statistics with Newey-West correction for serial correlation, using 22 lags.

| | **Panel A: Call Options** | | | | | | **Panel B: Put Options** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Short-Term | | | Long-Term | | | Short-Term | | | Long-Term | | |
| | OTM | ATM | ITM | OTM | ATM | ITM | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL^O_{t-1}$ | 0.062‡ | 0.038‡ | 0.046‡ | 0.027‡ | 0.019‡ | 0.026‡ | 0.045‡ | 0.025‡ | 0.034‡ | 0.020‡ | 0.014‡ | 0.020‡ |
| **Adj $R^2$** | 0.012 | 0.013 | 0.018 | 0.010 | 0.010 | 0.013 | 0.011 | 0.013 | 0.015 | 0.011 | 0.012 | 0.014 |
| # Obs in CS (avr.) | 220 | 213 | 236 | 262 | 290 | 300 | 232 | 199 | 194 | 301 | 274 | 227 |
| # CS regressions | 2984 | 2984 | 2984 | 2982 | 2982 | 2982 | 2984 | 2983 | 2984 | 2982 | 2982 | 2982 |
| | | | | | | | | | | | | |
| $IL^S_{t-1}$ | -4.071‡ | -2.149‡ | -0.761‡ | -1.216‡ | -0.721‡ | -0.280‡ | -2.913‡ | -2.109‡ | -0.249 | -0.982‡ | -0.589‡ | -0.169† |
| **Adj $R^2$** | 0.009 | 0.010 | 0.009 | 0.008 | 0.010 | 0.010 | 0.010 | 0.013 | 0.014 | 0.009 | 0.012 | 0.014 |
| # Obs in CS (avr.) | 220 | 213 | 236 | 261 | 290 | 299 | 231 | 199 | 193 | 300 | 273 | 227 |
| # CS regressions | 2984 | 2984 | 2984 | 2982 | 2982 | 2982 | 2984 | 2983 | 2984 | 2982 | 2982 | 2982 |
| | | | | | | | | | | | | |
| $IL^O_{t-1}$ | 0.062‡ | 0.035‡ | 0.045‡ | 0.030‡ | 0.022‡ | 0.033‡ | 0.048‡ | 0.023‡ | 0.027‡ | 0.023‡ | 0.012‡ | 0.021‡ |
| $IL^S_{t-1}$ | -2.077‡ | -0.873‡ | -0.428‡ | -0.598‡ | -0.223‡ | -0.205‡ | -1.399‡ | -1.109‡ | -0.064 | -0.416‡ | -0.238‡ | -0.084 |
| $\tilde{R}^O_{t-1}$ | -0.032‡ | -0.015‡ | -0.010‡ | -0.031‡ | -0.014‡ | -0.011‡ | -0.030‡ | -0.013‡ | -0.011‡ | -0.027‡ | -0.014‡ | -0.012‡ |
| $\sigma_{t-1}$ | -0.112‡ | -0.049‡ | -0.014‡ | -0.041‡ | -0.018‡ | -0.005‡ | -0.064‡ | -0.031‡ | -0.007‡ | -0.019‡ | -0.011‡ | -0.003† |
| $b_{t-1}$ | -0.002 | 0.000 | 0.001 | 0.000 | 0.000 | 0.000 | -0.003* | -0.004‡ | -0.001 | -0.002 | -0.001 | 0.000 |
| ln $(size_{t-1})$ | 0.004‡ | 0.002‡ | 0.001‡ | 0.001‡ | 0.001‡ | 0.001‡ | 0.004‡ | 0.002‡ | 0.000‡ | 0.001‡ | 0.000‡ | 0.000‡ |
| $lev_{t-1}$ | -0.001 | 0.000 | -0.001† | -0.001 | 0.000 | 0.002 | -0.002† | -0.002† | 0.000 | -0.001 | 0.000 | 0.000 |
| **Adj $R^2$** | 0.060 | 0.075 | 0.080 | 0.057 | 0.071 | 0.080 | 0.065 | 0.087 | 0.094 | 0.067 | 0.088 | 0.098 |
| # Obs in CS (avr.) | 189 | 179 | 207 | 240 | 270 | 283 | 204 | 166 | 163 | 285 | 252 | 206 |
| # CS regressions | 2968 | 2968 | 2968 | 2965 | 2965 | 2965 | 2968 | 2965 | 2965 | 2965 | 2965 | 2965 |

## Table 5
## Fama-MacBeth Regressions for Weekly Adjusted Option Returns

We report the results of cross-sectional Fama-MacBeth regressions for weekly adjusted call and put option returns ($\widetilde{R}^{O}$), i.e. the residuals plus the intercept from the regression of option returns on stock returns, lagged stock returns and squared stock returns. We include the lagged values of the following regressors: option illiquidity $IL^{O}$, computed from relative daily quoted bid-ask spreads, $IL^{S}$ , the dollar-volume weighted average of daily stock illiquidity for the previous week, the conditional volatility of returns, computed as the square root of the sum of squared daily returns for the previous week, *b*, the average of daily systematic risk proportion for the previous week, the logarithm of the firm size and the firm leverage. We use the size observed on the last day of the previous week, and the leverage from the previous quarter. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007. The symbols *, † and ‡ denote, respectively, significance at the 10%, 5% and 1% levels using Fama-MacBeth t-statistics with Newey-West correction for serial correlation, using 8 lags.

| | **Panel A: Call Options** | | | | | | **Panel B: Put Options** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Short-Term | | | Long-Term | | | Short-Term | | | Long-Term | | |
| | OTM | ATM | ITM | OTM | ATM | ITM | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL^{O}_{t-1}$ | 0.211‡ | 0.101‡ | -0.004 | 0.076‡ | 0.059‡ | 0.048‡ | 0.159‡ | 0.012 | -0.081‡ | 0.059‡ | 0.000 | -0.023 |
| **Adj R$^2$** | 0.015 | 0.009 | 0.015 | 0.008 | 0.008 | 0.012 | 0.012 | 0.010 | 0.014 | 0.008 | 0.011 | 0.013 |
| # Obs in CS (avr.) | 244 | 243 | 260 | 279 | 305 | 311 | 251 | 229 | 220 | 311 | 290 | 244 |
| # CS regressions | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 |
| | | | | | | | | | | | | |
| $IL^{S}_{t-1}$ | -23.676‡ | -11.526‡ | -6.029‡ | -9.116‡ | -5.307‡ | -2.422‡ | -17.129‡ | -9.846‡ | -2.625‡ | -5.925‡ | -3.371‡ | -0.895* |
| **Adj R$^2$** | 0.010 | 0.011 | 0.011 | 0.012 | 0.012 | 0.013 | 0.011 | 0.012 | 0.013 | 0.012 | 0.014 | 0.018 |
| # Obs in CS (avr.) | 244 | 242 | 259 | 279 | 305 | 311 | 251 | 228 | 220 | 310 | 290 | 244 |
| # CS regressions | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 | 622 |
| | | | | | | | | | | | | |
| $IL^{O}_{t-1}$ | 0.262‡ | 0.193‡ | 0.073‡ | 0.122‡ | 0.132‡ | 0.105‡ | 0.192‡ | 0.055* | -0.100‡ | 0.091‡ | 0.041* | -0.032 |
| $IL^{S}_{t-1}$ | -20.355‡ | -7.431‡ | -3.664‡ | -9.239‡ | -3.888‡ | -1.872‡ | -14.206‡ | -5.626‡ | -0.151 | -4.969‡ | -2.169‡ | -0.173 |
| $\tilde{R}^{O}_{t-1}$ | 0.003 | 0.001 | -0.007‡ | -0.011‡ | -0.006‡ | -0.007‡ | -0.002 | -0.001 | -0.004† | -0.008† | -0.009‡ | -0.008‡ |
| $\sigma_{t-1}$ | -0.213‡ | -0.102‡ | -0.039‡ | -0.051‡ | -0.028‡ | -0.017‡ | -0.078‡ | -0.047‡ | -0.025‡ | -0.008 | -0.014‡ | -0.011‡ |
| $b_{t-1}$ | -0.015 | -0.004 | -0.004 | -0.012 | -0.005 | -0.001 | -0.042‡ | -0.034‡ | -0.010* | -0.015† | -0.008† | -0.001 |
| $\ln(size_{t-1})$ | 0.024‡ | 0.013‡ | 0.005‡ | 0.008‡ | 0.005‡ | 0.002‡ | 0.019‡ | 0.009‡ | 0.002‡ | 0.007‡ | 0.003‡ | 0.001* |
| $lev_{t-1}$ | 0.049‡ | 0.025‡ | 0.005† | 0.021‡ | 0.006‡ | 0.002 | 0.028‡ | 0.014‡ | 0.011‡ | 0.007† | 0.003* | 0.002 |
| **Adj R$^2$** | 0.058 | 0.050 | 0.054 | 0.049 | 0.054 | 0.060 | 0.055 | 0.057 | 0.060 | 0.052 | 0.066 | 0.071 |
| # Obs in CS (avr.) | 200 | 199 | 222 | 252 | 284 | 294 | 215 | 184 | 176 | 293 | 266 | 217 |
| # CS regressions | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |

## Table 6
## Daily Portfolio Strategies

We show portfolio sorting results for call options (Panel A) and put options (Panel B). Each day, we sort the firms into deciles based on their lagged option illiquidity $IL^O$. For each decile, we report (in percentages) the time-series average of raw option returns $R^O$, the net (after transaction costs) option returns, $R^O_{net}$-Long for the long position and $R^O_{net}$-Short for the short position, stock returns $R^S$, the option quoted relative bid-ask spread $IL^O$, the effective relative bid-ask spread $IL^S$ for the stock, and size in millions of dollars. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

### Panel A: Call Options

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | $2.Low$ | 4 | 6 | 8 | $10.High$ | $2.Low$ | 4 | 6 | 8 | $10.High$ |
| OTM | $R^O_t$ | 0.17 | 0.93 | 0.71 | 1.49 | 5.72 | 0.46 | 0.73 | 1.10 | 0.87 | 1.94 |
| | $R^O_{net}$-Long | -12.67 | -17.21 | -23.77 | -31.47 | -51.34 | -9.25 | -12.45 | -16.24 | -23.01 | -42.02 |
| | $R^O_{net}$-Short | -15.01 | -23.50 | -33.88 | -52.89 | -165.82 | -11.27 | -16.12 | -22.55 | -33.28 | -96.73 |
| | $R^S_t$ | 0.07 | 0.08 | 0.07 | 0.09 | 0.21 | 0.05 | 0.07 | 0.11 | 0.06 | 0.10 |
| | $IL^O_t$ | 15.87 | 21.92 | 30.92 | 42.55 | 77.80 | 11.01 | 14.62 | 19.48 | 28.00 | 57.50 |
| | $IL^S_t$ | 0.22 | 0.23 | 0.26 | 0.28 | 0.35 | 0.22 | 0.24 | 0.26 | 0.30 | 0.39 |
| | $size_t$ | 33189 | 23496 | 19543 | 13868 | 7880 | 38163 | 26299 | 18749 | 13073 | 7000 |
| ATM | $R^O_t$ | 0.06 | 0.26 | 0.43 | 0.54 | 1.43 | 0.34 | 0.41 | 0.45 | 0.46 | 0.84 |
| | $R^O_{net}$-Long | -6.32 | -8.02 | -9.92 | -12.76 | -22.42 | -5.01 | -6.34 | -7.85 | -10.02 | -16.34 |
| | $R^O_{net}$-Short | -6.88 | -9.33 | -12.08 | -16.03 | -34.53 | -6.00 | -7.68 | -9.57 | -12.24 | -22.29 |
| | $R^S_t$ | 0.07 | 0.05 | 0.07 | 0.07 | 0.09 | 0.05 | 0.07 | 0.07 | 0.06 | 0.08 |
| | $IL^O_t$ | 7.33 | 9.07 | 11.24 | 14.63 | 27.70 | 5.87 | 7.12 | 8.70 | 11.01 | 18.51 |
| | $IL^S_t$ | 0.22 | 0.23 | 0.24 | 0.27 | 0.37 | 0.22 | 0.24 | 0.25 | 0.29 | 0.38 |
| | $size_t$ | 38866 | 26397 | 19811 | 14445 | 7202 | 39591 | 26969 | 18087 | 12266 | 6447 |
| ITM | $R^O_t$ | 0.11 | 0.13 | -0.08 | 0.39 | 0.47 | 0.24 | 0.19 | 0.16 | 0.28 | 0.45 |
| | $R^O_{net}$-Long | -4.29 | -5.45 | -6.81 | -8.21 | -13.39 | -3.28 | -4.36 | -5.42 | -6.56 | -10.60 |
| | $R^O_{net}$-Short | -4.71 | -6.05 | -7.17 | -9.87 | -16.95 | -3.90 | -4.97 | -6.09 | -7.65 | -13.05 |
| | $R^S_t$ | 0.05 | 0.03 | 0.02 | 0.06 | 0.03 | 0.06 | 0.04 | 0.02 | 0.06 | 0.06 |
| | $IL^O_t$ | 4.82 | 5.88 | 7.03 | 9.00 | 14.79 | 3.75 | 4.74 | 5.77 | 7.01 | 11.46 |
| | $IL^S_t$ | 0.21 | 0.23 | 0.25 | 0.28 | 0.35 | 0.21 | 0.24 | 0.26 | 0.29 | 0.39 |
| | $size_t$ | 41649 | 29545 | 21687 | 14415 | 7302 | 40327 | 27161 | 19072 | 12798 | 5874 |

**Table 6 (continued)**
**Daily Portfolio Strategies**

**Panel B: Put Options**

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | $2. Low$ | 4 | 6 | 8 | $10. High$ | $2. Low$ | 4 | 6 | 8 | $10. High$ |
| OTM | $R_t^O$ | -0.51 | -0.30 | -0.15 | -0.13 | 3.02 | 0.15 | 0.05 | 0.06 | -0.01 | 0.47 |
| | $R_{net}^O$-Long | -11.70 | -15.77 | -20.68 | -28.39 | -47.97 | -8.14 | -10.90 | -14.06 | -18.89 | -35.23 |
| | $R_{net}^O$-Short | -12.24 | -18.38 | -26.44 | -40.75 | -127.51 | -9.27 | -12.55 | -16.83 | -23.94 | -63.84 |
| | $R_t^S$ | 0.06 | 0.05 | 0.03 | 0.05 | -0.06 | 0.03 | 0.04 | 0.05 | 0.05 | 0.05 |
| | $IL_t^O$ | 13.42 | 18.46 | 24.93 | 35.78 | 69.90 | 9.16 | 11.99 | 15.62 | 21.49 | 45.29 |
| | $IL_t^S$ | 0.22 | 0.23 | 0.25 | 0.28 | 0.33 | 0.22 | 0.24 | 0.25 | 0.28 | 0.36 |
| | $size_t$ | 36242 | 28023 | 21599 | 15759 | 8667 | 36849 | 27907 | 20584 | 13984 | 7312 |
| ATM | $R_t^O$ | -0.50 | -0.41 | -0.50 | 0.01 | -0.50 | 0.02 | -0.11 | -0.11 | -0.12 | -0.05 |
| | $R_{net}^O$-Long | -6.31 | -7.88 | -9.88 | -11.99 | -21.71 | -4.65 | -6.02 | -7.32 | -9.31 | -14.94 |
| | $R_{net}^O$-Short | -5.68 | -7.72 | -9.94 | -13.76 | -27.77 | -4.93 | -6.20 | -7.71 | -10.08 | -17.85 |
| | $R_t^S$ | 0.05 | 0.06 | 0.06 | 0.02 | 0.08 | 0.03 | 0.05 | 0.06 | 0.06 | 0.06 |
| | $IL_t^O$ | 6.57 | 8.14 | 10.24 | 13.04 | 25.08 | 5.06 | 6.24 | 7.55 | 9.67 | 15.99 |
| | $IL_t^S$ | 0.22 | 0.23 | 0.25 | 0.27 | 0.36 | 0.23 | 0.24 | 0.26 | 0.28 | 0.36 |
| | $size_t$ | 39024 | 30491 | 22602 | 16936 | 8830 | 38333 | 30040 | 21287 | 14210 | 7229 |
| ITM | $R_t^O$ | -0.30 | -0.51 | -0.27 | -0.43 | -0.60 | -0.06 | -0.14 | -0.17 | -0.19 | -0.23 |
| | $R_{net}^O$-Long | -4.49 | -5.83 | -6.69 | -8.46 | -13.68 | -3.24 | -4.33 | -5.35 | -6.55 | -10.46 |
| | $R_{net}^O$-Short | -4.08 | -5.14 | -6.62 | -8.37 | -14.77 | -3.23 | -4.26 | -5.31 | -6.62 | -11.32 |
| | $R_t^S$ | 0.06 | 0.11 | 0.07 | 0.10 | 0.13 | 0.05 | 0.06 | 0.08 | 0.08 | 0.11 |
| | $IL_t^O$ | 4.60 | 5.67 | 6.71 | 8.44 | 14.08 | 3.38 | 4.39 | 5.36 | 6.54 | 10.69 |
| | $IL_t^S$ | 0.23 | 0.24 | 0.26 | 0.28 | 0.35 | 0.23 | 0.25 | 0.27 | 0.30 | 0.38 |
| | $size_t$ | 40702 | 32555 | 25423 | 17446 | 8847 | 40411 | 32022 | 23561 | 16917 | 7070 |

## Table 7
## Weekly Portfolio Strategies

We show portfolio sorting results for calls (Panel A) and puts (Panel B). Each week, we sort the firms into deciles based on their lagged illiquidity $IL^O$. For each decile, we report (in percentages) the time-series average of raw option returns $R^O$, the net (after transaction costs) option returns ($R^O_{net}$-Long for the long position and $R^O_{net}$-Short for the short position), stock returns $R^S$, the option quoted relative bid-ask spread $IL^O$, the stock effective relative bid-ask spreads $IL^S$, and size, in millions of dollars. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007.

### Panel A: Call Options

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | $2.Low$ | 4 | 6 | 8 | $10.High$ | $2.Low$ | 4 | 6 | 8 | $10.High$ |
| OTM | $R^O_t$ | 3.02 | 4.85 | 6.80 | 10.62 | 25.97 | 3.09 | 3.12 | 3.76 | 4.92 | 7.82 |
| | $R^O_{net}$-Long | -10.20 | -13.52 | -18.16 | -23.25 | -35.59 | -6.82 | -10.37 | -13.88 | -19.56 | -37.04 |
| | $R^0_{net}$-Short | -18.24 | -27.61 | -40.37 | -62.64 | -195.71 | -14.12 | -18.85 | -25.54 | -38.03 | -104.13 |
| | $R^S_t$ | 0.25 | 0.41 | 0.56 | 0.69 | 1.28 | 0.30 | 0.32 | 0.37 | 0.47 | 0.61 |
| | $IL^0_t$ | 22.53 | 29.01 | 38.38 | 49.78 | 75.22 | 12.47 | 16.77 | 22.05 | 31.39 | 59.16 |
| | $IL^S_t$ | 0.21 | 0.23 | 0.25 | 0.28 | 0.34 | 0.21 | 0.23 | 0.26 | 0.30 | 0.39 |
| | $size_t$ | 35487 | 24695 | 20669 | 15097 | 8559 | 39044 | 26609 | 19270 | 13631 | 7386 |
| ATM | $R^O_t$ | 2.10 | 1.98 | 2.45 | 1.92 | 3.22 | 1.61 | 2.22 | 1.66 | 2.48 | 2.57 |
| | $R^O_{net}$-Long | -4.49 | -6.58 | -8.31 | -11.88 | -21.36 | -3.81 | -4.64 | -6.76 | -8.22 | -14.98 |
| | $R^0_{net}$-Short | -9.14 | -11.37 | -14.56 | -18.03 | -37.57 | -7.35 | -9.62 | -10.90 | -14.53 | -24.51 |
| | $R^S_t$ | 0.32 | 0.33 | 0.35 | 0.34 | 0.46 | 0.32 | 0.38 | 0.30 | 0.39 | 0.40 |
| | $IL^0_t$ | 9.28 | 11.47 | 14.44 | 18.99 | 35.18 | 6.23 | 7.51 | 9.24 | 11.65 | 20.10 |
| | $IL^S_t$ | 0.22 | 0.23 | 0.24 | 0.27 | 0.37 | 0.22 | 0.23 | 0.25 | 0.28 | 0.39 |
| | $size_t$ | 39486 | 27685 | 20045 | 14555 | 7511 | 40679 | 27471 | 18667 | 12519 | 6614 |
| ITM | $R^O_t$ | 0.63 | 0.79 | 0.20 | 0.18 | 0.08 | 0.95 | 1.31 | 1.17 | 1.43 | 1.69 |
| | $R^O_{net}$-Long | -3.82 | -4.85 | -6.60 | -8.48 | -13.96 | -2.62 | -3.30 | -4.47 | -5.49 | -9.45 |
| | $R^0_{net}$-Short | -5.29 | -6.79 | -7.53 | -9.73 | -16.80 | -4.65 | -6.16 | -7.17 | -8.89 | -14.41 |
| | $R^S_t$ | 0.25 | 0.28 | 0.21 | 0.25 | 0.29 | 0.29 | 0.35 | 0.33 | 0.37 | 0.42 |
| | $IL^0_t$ | 5.29 | 6.45 | 7.78 | 10.12 | 17.37 | 3.89 | 4.88 | 5.91 | 7.17 | 11.63 |
| | $IL^S_t$ | 0.21 | 0.23 | 0.25 | 0.28 | 0.36 | 0.21 | 0.23 | 0.25 | 0.29 | 0.39 |
| | $size_t$ | 41875 | 29476 | 21757 | 14436 | 7531 | 41879 | 27644 | 19737 | 12853 | 5928 |

**Table 7 (continued)**
**Weekly Portfolio Strategies**

## Panel B: Put Options

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | $2. Low$ | 4 | 6 | 8 | $10. High$ | $2. Low$ | 4 | 6 | 8 | $10. High$ |
| OTM | $R_t^O$ | -3.81 | -3.29 | -2.08 | -1.64 | 7.87 | -0.77 | -1.34 | -1.22 | -0.98 | -0.40 |
| | $R_{net}^O$-Long | -14.94 | -18.66 | -22.48 | -29.51 | -43.02 | -9.12 | -12.33 | -15.38 | -19.98 | -36.19 |
| | $R_{net}^0$-Short | -8.84 | -15.23 | -24.25 | -38.43 | -129.43 | -8.43 | -11.21 | -15.62 | -23.16 | -64.01 |
| | $R_t^S$ | 0.34 | 0.31 | 0.17 | 0.14 | -0.34 | 0.28 | 0.32 | 0.29 | 0.28 | 0.28 |
| | $IL_t^0$ | 19.03 | 25.58 | 33.23 | 45.79 | 74.32 | 10.25 | 13.47 | 17.58 | 24.21 | 50.16 |
| | $IL_t^S$ | 0.21 | 0.23 | 0.24 | 0.28 | 0.33 | 0.22 | 0.23 | 0.25 | 0.28 | 0.36 |
| | $size_t$ | 36413 | 28377 | 22343 | 16838 | 9556 | 37853 | 28216 | 20604 | 14338 | 7768 |
| ATM | $R_t^O$ | -2.93 | -3.66 | -3.20 | -3.59 | -5.11 | -0.92 | -1.38 | -1.40 | -1.25 | -1.99 |
| | $R_{net}^O$-Long | -8.77 | -11.10 | -12.58 | -15.63 | -26.20 | -5.61 | -7.29 | -8.61 | -10.44 | -16.83 |
| | $R_{net}^0$-Short | -3.28 | -4.44 | -7.26 | -10.26 | -23.17 | -4.00 | -4.92 | -6.41 | -8.95 | -15.95 |
| | $R_t^S$ | 0.32 | 0.38 | 0.31 | 0.29 | 0.37 | 0.26 | 0.35 | 0.35 | 0.32 | 0.37 |
| | $IL_t^0$ | 8.18 | 10.13 | 13.07 | 17.37 | 33.38 | 5.39 | 6.59 | 8.01 | 10.27 | 17.42 |
| | $IL_t^S$ | 0.22 | 0.23 | 0.24 | 0.27 | 0.35 | 0.22 | 0.23 | 0.25 | 0.28 | 0.36 |
| | $size_t$ | 39778 | 31261 | 22381 | 16994 | 8937 | 39042 | 29917 | 21032 | 14001 | 7340 |
| ITM | $R_t^O$ | -3.19 | -4.28 | -4.09 | -4.66 | -5.75 | -1.20 | -1.47 | -1.64 | -1.60 | -1.96 |
| | $R_{net}^O$-Long | -7.35 | -9.49 | -10.36 | -12.54 | -18.60 | -4.40 | -5.66 | -6.80 | -7.90 | -12.15 |
| | $R_{net}^0$-Short | -1.16 | -1.25 | -2.64 | -3.98 | -9.46 | -2.12 | -2.92 | -3.82 | -5.17 | -9.63 |
| | $R_t^S$ | 0.41 | 0.55 | 0.46 | 0.47 | 0.53 | 0.35 | 0.40 | 0.39 | 0.41 | 0.44 |
| | $IL_t^0$ | 5.19 | 6.35 | 7.59 | 9.80 | 17.50 | 3.58 | 4.58 | 5.57 | 6.74 | 11.11 |
| | $IL_t^S$ | 0.22 | 0.24 | 0.25 | 0.27 | 0.35 | 0.23 | 0.24 | 0.26 | 0.29 | 0.38 |
| | $size_t$ | 42684 | 33442 | 25285 | 17437 | 8929 | 41827 | 33106 | 23666 | 16708 | 7021 |

## Table 8
## Fama-MacBeth Regressions for the Level of Implied Volatility

For each month and for each category, we run the following regression using all observed options within the month. The regression is run separately for call and put options

$$iv_{i,t}(\chi_k, T_k) = \kappa_{i,t} + \Theta_{i,t}(\chi_k - \bar{\chi}_k) + \eta_{i,t}(T_k - \bar{T}_k) + \varepsilon_{i,t}^k, \quad k = 1,2, \ldots .K$$

where $iv_{i,t}(\chi_k, T_k)$ is the implied volatility for an option with moneyness $\chi_k$ and maturity $T_k$. The subscripts *t*, *i* and *k* correspond to month *t*, firm *i* and contract *k*, respectively. *K* is the number of contracts available for a given month and category. We consider only months for which *K* is larger than ten. For each firm *i*, we obtain a monthly time series for $\kappa_{i,t}$ which corresponds to the estimated level of implied volatility. Then, for each month *t*, we run the following regression

$$\tilde{\kappa}_{i,t} = a_{0,t} + a_{1,t}IL_{i,t}^O + a_{2,t}IL_{i,t}^S + a_{3,t}R_{i,t}^S + a_{4,t}b_{i,t} + a_{5,t}\ln(size_{i,t}) + a_{6,t}lev_{i,t} + \varepsilon_{i,t}, \quad i = 1,2, \ldots I$$

where $\tilde{\kappa}_{i,t}$ is the residual plus the intercept from the cross-sectional regression of $\kappa_{i,t}$ on the volatility $\sigma_t$. $IL_t^O$ is the monthly average of daily option illiquidity for the K contracts used to run the first regression, and $IL_t^S$ is the dollar-volume weighted average of daily stock illiquidity. $R_t^S$ is the monthly stock return. The option illiquidity is the relative bid-ask spread and the stock illiquidity is the effective bid-ask spread estimated from TAQ data. $b_t$ is the systematic risk proportion, which corresponds to the square root of the $R^2$ from the regression of stock returns on Fama-French and momentum factors. We use the monthly average of the daily systematic risk proportion. $\ln(size_t)$ and $lev_t$ are respectively the logarithm of firm size and the firm leverage. We use the firm size observed on the last day of the month and leverage from the last available quarter. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007. The symbols *, † and ‡ denote, respectively, significance at the 10%, 5% and 1% levels using Fama-MacBeth t-statistics with Newey-West correction for serial correlation, with 8 lags.

### Panel A: Call Options

| | Short-Term | | | Long-Term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | -0.102‡ | -0.381‡ | -0.803‡ | -0.121‡ | -0.547‡ | -1.006‡ |
| $IL_t^S$ | 13.418‡ | 11.651‡ | 13.956‡ | 11.880‡ | 11.490‡ | 13.174‡ |
| $R_t^S$ | -0.030‡ | -0.025† | -0.040‡ | -0.008 | -0.018 | -0.025* |
| $b_t$ | -0.002 | 0.012 | 0.018* | -0.001 | 0.009 | 0.013 |
| $\ln(size_t)$ | -0.018‡ | -0.019‡ | -0.019‡ | -0.015‡ | -0.018‡ | -0.017‡ |
| $lev_t$ | -0.012‡ | -0.014‡ | -0.007 | -0.017‡ | -0.013‡ | -0.005 |
| **Adj $R^2$** | 0.191 | 0.195 | 0.190 | 0.170 | 0.190 | 0.186 |
| # Obs in CS (avr.) | 272 | 257 | 276 | 289 | 310 | 313 |
| # CS regressions | 144 | 144 | 144 | 144 | 144 | 144 |

### Panel B: Put Options

| | Short-Term | | | Long-Term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | -0.129‡ | -0.417‡ | -0.733‡ | -0.204‡ | -0.673‡ | -0.958‡ |
| $IL_t^S$ | 12.665‡ | 11.507‡ | 14.199‡ | 11.755‡ | 10.689‡ | 11.849‡ |
| $R_t^S$ | -0.017 | -0.006 | -0.008 | -0.002 | 0.005 | 0.026† |
| $b_t$ | 0.020* | 0.016† | 0.012 | 0.009 | 0.012 | 0.007 |
| $\ln(size_t)$ | -0.018‡ | -0.020‡ | -0.021‡ | -0.015‡ | -0.019‡ | -0.019‡ |
| $lev_t$ | -0.011* | -0.015‡ | -0.009† | -0.007 | -0.013‡ | -0.016‡ |
| **Adj $R^2$** | 0.172 | 0.204 | 0.213 | 0.165 | 0.214 | 0.215 |
| # Obs in CS (avr.) | 274 | 241 | 234 | 314 | 296 | 248 |
| # CS regressions | 144 | 144 | 144 | 144 | 144 | 144 |

## Table 9
## Fama-MacBeth Regressions for the Moneyness-Slope of Implied Volatility

For each month and for each option category, we run the following regression using all observed options within the month. The regression is run separately for call and put options

$$iv_{i,t}(\chi_k, T_k) = \kappa_{i,t} + \theta_{i,t}(\chi_k - \bar{\chi}_k) + \eta_{i,t}(T_k - \bar{T}_k) + \varepsilon_{i,t}^k, \quad k = 1,2, \ldots . K$$

where $iv_{i,t}(\chi_k, T_k)$ is the implied volatility for an option with moneyness $\chi_k$ and maturity $T_k$. The subscripts *t*, *i* and *k* correspond to month *t*, firm *i* and contract *k*, respectively. *K* is the number of contracts available for the considered month and category. We consider only months for which *K* is larger than ten. For each firm *i*, we obtain a monthly time series for $\theta_{i,t}$ which corresponds to the estimated moneyness-slope of implied volatility. Then, for each month *t*, we run the following regression

$$|\theta_{i,t}| = c_{0,t} + c_{1,t}IL_{i,t}^O + c_{2,t}IL_{i,t}^S + c_{3,t}R_{i,t}^S + c_{4,t}b_{i,t} + c_{5,t}\ln(size_{i,t}) + c_{6,t}lev_{i,t} + \varepsilon_{i,t}, \quad i = 1,2, \ldots I$$

where $IL_t^O$ is the average across the month of daily option illiquidity of the K contracts used to run the first regression, and $IL_t^S$ is the dollar-volume weighted average of daily stock illiquidity. $R_t^S$ is the monthly stock return. The option illiquidity is the relative bid-ask spread, and the stock illiquidity is the effective bid-ask spread estimated from TAQ data. $b_t$ is the systematic risk proportion, which corresponds to the square root of the $R^2$ from the regression of stock returns on Fama-French and momentum factors. We take the monthly average of the daily systematic risk proportion. ln $(size_t)$ and $lev_t$ are respectively the logarithm of firm size and the firm leverage. We use the firm size observed on the last day of the month and the leverage from the last available quarter. The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007. The symbols *, † and ‡ denote, respectively, significance at the 10%, 5% and 1% levels using Fama-MacBeth t-statistics with Newey-West correction for serial correlation, using 8 lags.

### Panel A: Call Options

| | Short-Term | | | Long-Term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | 0.516‡ | 1.188‡ | 2.221‡ | 0.273‡ | 0.953‡ | 1.353‡ |
| $IL_t^S$ | 26.359‡ | 20.333‡ | 32.015‡ | -3.671† | -0.040 | 10.682‡ |
| $R_t^S$ | -0.033 | -0.045 | 0.093† | 0.050† | -0.040‡ | -0.044† |
| $b_t$ | -0.110‡ | -0.128‡ | -0.109‡ | -0.031† | -0.020* | 0.005 |
| ln $(size_t)$ | 0.022‡ | 0.024‡ | 0.056‡ | 0.001 | 0.013‡ | 0.022‡ |
| $lev_t$ | 0.125‡ | 0.094‡ | 0.137‡ | 0.057‡ | 0.063‡ | 0.108‡ |
| **Adj R²** | 0.125 | 0.074 | 0.078 | 0.101 | 0.080 | 0.090 |
| # Obs in CS (avr.) | 272 | 257 | 276 | 289 | 310 | 313 |
| # CS regressions | 144 | 144 | 144 | 144 | 144 | 144 |

### Panel B: Put Options

| | Short-Term | | | Long-Term | | |
|---|---|---|---|---|---|---|
| | OTM | ATM | ITM | OTM | ATM | ITM |
| $IL_t^O$ | 0.231‡ | 1.122‡ | 3.289‡ | 0.171‡ | 0.893‡ | 2.088‡ |
| $IL_t^S$ | 33.532‡ | 21.504‡ | 32.683‡ | 11.647‡ | 1.365 | 5.363† |
| $R_t^S$ | -0.042 | -0.083† | -0.199‡ | -0.056‡ | -0.032* | 0.008 |
| $b_t$ | -0.081* | -0.123‡ | -0.136‡ | 0.003 | -0.017 | -0.034‡ |
| ln $(size_t)$ | 0.039‡ | 0.017‡ | 0.035‡ | 0.020‡ | 0.009‡ | 0.002 |
| $lev_t$ | 0.158‡ | 0.097‡ | 0.116‡ | 0.099‡ | 0.056‡ | 0.048‡ |
| **Adj R²** | 0.071 | 0.060 | 0.108 | 0.085 | 0.072 | 0.124 |
| # Obs in CS (avr.) | 274 | 241 | 234 | 314 | 296 | 248 |
| # CS regressions | 144 | 144 | 144 | 144 | 144 | 144 |

## Table 10
## Time-Series Regressions for Weekly Option Returns

Each week, we sort the firms into deciles based on their lagged option illiquidity. The lagged illiquidity corresponds to the average of relative bid-ask spreads on the previous Friday of the contracts used to compute returns for the week. For each decile $j$, we take the average across firms of illiquidity, lagged illiquidity, and adjusted option returns, which are the residuals from time-series regression of option returns on current and lagged stock returns and squared stock returns. We thus obtain a weekly time-series for $IL_{j,t}$, and $\tilde{R}^O_{j,t}$ over the entire sample period. Then we run the following regression:

$$\ln(IL^O_{j,t}) = \omega_{j,0} + \omega_{j,1}\ln(IL^O_{j,t-1}) + v^{IL}_{j,t}.$$

Defining the unexpected illiquidity by $IL^u_{j,t} = v^{IL}_{j,t}$, we estimate the following time-series regression:

$$\tilde{R}^O_{j,t} = \gamma_0 + \gamma_1\ln(IL^O_{j,t-1}) + \gamma_2 IL^u_{j,t} + \gamma_3\ln(IL^S_{j,t-1}) + v_{j,t}$$

where $IL^S_{j,t-1}$ is lagged stock illiquidity.

The sample includes the S&P 500 constituents as of December 31, 2007 for which options trade throughout the entire sample period, which is from January 1996 through December 2007. The symbols *, † and ‡ denote, respectively, significance at the 10%, 5% and 1% levels with Newey-West correction for serial correlation, using 8 lags.

### Panel A: Call Options

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | 2. *Low* | 4 | 6 | 8 | 10. *High* | 2. *Low* | 4 | 6 | 8 | 10. *High* |
| OTM | $\ln(IL^O_{j,t-1})$ | 0.195‡ | 0.262‡ | 0.405‡ | 0.388‡ | 0.446‡ | 0.093‡ | 0.135‡ | 0.183‡ | 0.207‡ | 0.191‡ |
| | $IL^u_{j,t}$ | -0.053† | -0.061† | -0.089† | -0.149‡ | -0.342‡ | -0.012 | -0.034 | -0.027 | -0.078† | -0.139‡ |
| | $\ln(IL^S_{j,t-1})$ | -0.098‡ | -0.119‡ | -0.152‡ | -0.121‡ | -0.098‡ | -0.038‡ | -0.056‡ | -0.070‡ | -0.061‡ | -0.030‡ |
| | **Adj R$^2$** | 0.177 | 0.195 | 0.230 | 0.195 | 0.203 | 0.121 | 0.154 | 0.144 | 0.160 | 0.121 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |
| ATM | $\ln(IL^O_{j,t-1})$ | 0.062‡ | 0.111‡ | 0.136‡ | 0.196‡ | 0.103‡ | 0.034‡ | 0.054‡ | 0.072‡ | 0.126‡ | 0.073‡ |
| | $IL^u_{j,t}$ | -0.053‡ | -0.070‡ | -0.095‡ | -0.120‡ | -0.177‡ | -0.009 | -0.025 | -0.052† | -0.077‡ | -0.112‡ |
| | $\ln(IL^S_{j,t-1})$ | -0.053‡ | -0.070‡ | -0.086‡ | -0.087‡ | -0.039‡ | -0.022‡ | -0.034‡ | -0.043‡ | -0.048‡ | -0.012‡ |
| | **Adj R$^2$** | 0.124 | 0.173 | 0.204 | 0.210 | 0.227 | 0.070 | 0.117 | 0.136 | 0.160 | 0.150 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |
| ITM | $\ln(IL^O_{j,t-1})$ | 0.030‡ | 0.038‡ | 0.047‡ | 0.073‡ | 0.034 | 0.014‡ | 0.025‡ | 0.040‡ | 0.056‡ | 0.038* |
| | $IL^u_{j,t}$ | -0.029 | -0.058† | -0.077‡ | -0.114‡ | -0.170‡ | -0.017 | -0.043† | -0.047† | -0.075‡ | -0.104‡ |
| | $\ln(IL^S_{j,t-1})$ | -0.022‡ | -0.030‡ | -0.035‡ | -0.038‡ | -0.018‡ | -0.010‡ | -0.017‡ | -0.019‡ | -0.022‡ | -0.009† |
| | **Adj R$^2$** | 0.065 | 0.094 | 0.127 | 0.195 | 0.287 | 0.045 | 0.097 | 0.086 | 0.128 | 0.163 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |

## Table 10 (continued)
## Time-Series Regressions for Weekly Option Returns

### Panel B: Put Options

| | | Short-Term | | | | | Long-Term | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Decile | | $2. Low$ | 4 | 6 | 8 | $10. High$ | $2. Low$ | 4 | 6 | 8 | $10. High$ |
| OTM | $\ln(IL_{j,t-1}^{O})$ | 0.107‡ | 0.156‡ | 0.188‡ | 0.185‡ | 0.266‡ | 0.032‡ | 0.044‡ | 0.060‡ | 0.069‡ | -0.001 |
| | $IL_{j,t}^{u}$ | -0.101‡ | -0.130‡ | -0.183‡ | -0.211‡ | -0.388‡ | -0.112‡ | -0.115‡ | -0.149‡ | -0.186‡ | -0.274‡ |
| | $\ln(IL_{j,t-1}^{S})$ | -0.053‡ | -0.067‡ | -0.075‡ | -0.063‡ | -0.034† | -0.011* | -0.023‡ | -0.029‡ | -0.018‡ | 0.008 |
| | **Adj $R^2$** | 0.160 | 0.210 | 0.248 | 0.235 | 0.217 | 0.079 | 0.143 | 0.176 | 0.227 | 0.327 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |
| ATM | $\ln(IL_{j,t-1}^{O})$ | 0.041‡ | 0.031‡ | 0.055‡ | 0.013 | -0.007 | 0.018‡ | 0.020‡ | 0.020† | 0.011 | 0.002 |
| | $IL_{j,t}^{u}$ | -0.084‡ | -0.126‡ | -0.139‡ | -0.168‡ | -0.232‡ | -0.067‡ | -0.122‡ | -0.132‡ | -0.156‡ | -0.195‡ |
| | $\ln(IL_{j,t-1}^{S})$ | -0.033‡ | -0.040‡ | -0.047‡ | -0.034‡ | -0.011 | -0.014‡ | -0.013‡ | -0.017‡ | -0.013† | -0.002 |
| | **Adj $R^2$** | 0.119 | 0.182 | 0.177 | 0.272 | 0.338 | 0.067 | 0.103 | 0.129 | 0.141 | 0.269 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |
| ITM | $\ln(IL_{j,t-1}^{O})$ | 0.001 | -0.007 | -0.010 | -0.029† | -0.091‡ | 0.002 | -0.001 | -0.009 | -0.024‡ | -0.036‡ |
| | $IL_{j,t}^{u}$ | -0.076‡ | -0.088‡ | -0.128‡ | -0.154‡ | -0.206‡ | -0.037* | -0.062* | -0.108‡ | -0.132‡ | -0.198‡ |
| | $\ln(IL_{j,t-1}^{S})$ | -0.010† | -0.004 | -0.010 | -0.001 | 0.006 | -0.006† | -0.005 | -0.004 | 0.003 | 0.012‡ |
| | **Adj $R^2$** | 0.062 | 0.060 | 0.132 | 0.193 | 0.367 | 0.021 | 0.029 | 0.086 | 0.103 | 0.248 |
| | # Obs | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 | 621 |

# Learning from Prices, Liquidity Spillovers, and Market Segmentation *

Giovanni Cespa † and Thierry Foucault ‡

April 2011

**Abstract**

We describe a new mechanism that explains the transmission of liquidity shocks from one security to another ("liquidity spillovers"). Dealers use prices of other securities as a source of information. As prices of less liquid securities convey less precise information, a drop in liquidity for one security raises the uncertainty for dealers in other securities, thereby affecting their liquidity. The direction of liquidity spillovers is positive if the fraction of dealers with price information on other securities is high enough. Otherwise liquidity spillovers can be negative. For some parameters, the value of price information increases with the number of dealers obtaining this information. In this case, related securities can appear segmented, even if the cost of price information is small.

*Keywords:* Liquidity spillovers, Liquidity Risk, Contagion, Value of price information, Transparency, Colocation.

JEL Classification Numbers: G10, G12, G14

*A previous version of this paper was circulated under the title: "Dealer attention, liquidity spillovers, and endogenous market segmentation." We are grateful to Terry Hendershott (the AFA discussant) for his comments. We also thank Dimitri Vayanos, Dion Bongaerts, Mark van Achter and seminar participants at the 2011 AFA meeting, Erasmus University, the Copenhaguen Business School, theb ESADE-IESE workshop, the University of Naples, the Paris School of Economics, the School of Banking and Finance at UNSW, the 6th CSEF-IGIER Symposium on Economics and Institutions, and the workshop on "Acquisition, Use and Transmission of Private Information in Financial Markets" (European University Institute, June 2010) for helpful comments and suggestions. All errors are ours.

†Cass Business School, CSEF, and CEPR. E-mail: giovanni.cespa@gmail.com

‡HEC, School of Management, Paris, GREGHEC, and CEPR. Tel: (33) 1 39 67 95 69; E-mail: foucault@hec.fr

# 1 Introduction

The "flash crash" of May 6, 2010 provides a striking illustration of how a drop in the liquidity of one security can quickly propagate to other securities. As shown in the CFTC-SEC report on the flash crash, buy limit orders for the E-mini futures contract on the S&P 500 index vanished in a few minutes after 2:30 p.m. on May 6, 2010.[1] This evaporation of liquidity in the E-mini futures was soon followed by a similar phenomenon in the SPY Exchange Traded Fund (another derivative security on the S&P 500 index) and in the S&P 500 index component stocks (see Figure 1.12 in the joint CFTC-SEC report), resulting in a very high volatility in transaction prices (with some stocks trading as low as a penny or as high as $100, 000).

Why do such liquidity spillovers arise? Addressing this question is of broad interest. It can shed light on sudden and short systematic liquidity crises such as the flash crash. More generally, it can explain why liquidity co-varies across securities.[2] Co-movements in liquidity have important implications for asset pricing since they are a source of systematic risk (see for instance Acharya and Pedersen (2005), Korajczyk and Sadka (2008) and Amihud et al. (2005) for a survey). Yet, their cause(s) is not well understood. Co-variations in liquidity may be driven by systematic variations in the demand for liquidity (see Hendershott and Seasholes (2009) or Koch, Ruenzi and Starks (2010)) or systematic variations in the supply of liquidity. One possibility is that financing constraints constitute a systematic liquidity factor because they bind liquidity providers in different securities at the same time. This mechanism is formalized by Gromb and Vayanos (2002) and Brunnemeier and Pedersen (2007) and has received empirical support from analysis of NYSE stocks (see for instance, Coughenour and Saad (2004) or Comerton-Forde et al. (2010)). Another related explanation is that a drop in the capital available to financial intermediaries active in multiple securities can trigger an increase in risk aversion, impairing the supply of liquidity in these securities (as in Kyle and Xiong (2001)).

In this paper we analyze a new mechanism that generates co-movements in the supply of liquidity in different securities, even when dealers active in these securities are *distinct* and *not* simultaneously hit by a market wide shock. Dealers in a security often rely on the prices of other securities to set their quotes. For instance, dealers in a stock learn information from the prices of other stocks in their industry or stock index futures. We show that cross-security learning by dealers causes *liquidity spillovers* and thereby co-movements in liquidity.

To see this intuitively, consider a dealer in security $X$ who uses the price of security $Y$ as a source of information. Movements in the price of security $Y$ are informative because they reflect news about fundamentals known to dealers in security $Y$. However, this signal is noisy since price movements in security $Y$ also reflect transient price pressures due to uninformed trades. These transient price pressures account for a larger fraction of price volatility when the cost of

[1]See "Findings regarding the market events of May 6, 2010," CFTC-SEC joint report available at `http://www.sec.gov/news/studies/2010/marketevents-report.pdf`

[2]Evidence of co-variations in liquidity are provided in Chordia et al. (2000), Hasbrouck and Seppi (2001), Huberman and Halka (2001), Korajczyk and Sadka (2008), Corwin and Lipson (2011) for stocks and Chordia et al. (2005) for bonds and stocks.

liquidity provision for dealers in security $Y$ is higher.[3] For this reason, the informativeness of the price of security $Y$ for dealers in security $X$ is smaller when security $Y$ is less liquid.[4] Now suppose that a shock *specific* to security $Y$ decreases the cost of liquidity provision for dealers in this security (e.g., dealers in this security face less stringent limits on their positions). Thus, security $Y$ becomes more liquid and, for this reason, the price of security $Y$ becomes more informative for dealers in security $X$ (transient price pressures in security $Y$ contribute less to its volatility relative to news about fundamentals). As a result, inventory risk for dealers in security $X$ is lower and the cost of liquidity provision for these dealers declines as well. In this way, the improvement in liquidity for security $Y$ spreads to security $X$, as shown in Figure 1.

[**Insert Figure 1 about here**]

To formalize this intuition, we consider a model with distinct pools of risk averse dealers operating in two securities, $X$ and $Y$, with a two-factor structure. Dealers in a given market have identical information on one of the risk factors. However, dealers operating in different markets are informed on different risk factors. For this reason, dealers in one market can learn information about the risk factor on which they have no information by watching the price of the other security. We explore two cases: the case in which learning is *two-sided* (dealers in each security learn from each other's price) and the case in which learning is *one-sided* (the price of one security is informative for dealers in another security but not vice versa).[5] We refer to dealers who engage in cross-security price monitoring as being "pricewatchers." The fraction of pricewatchers associated with a security sets *the dealers' level of attention* to the other security.

The model generates the spillover mechanism portrayed in Figure 1 and a rich set of implications. First, when learning is two-sided, an exogenous shock to the cost of liquidity provision in one security (say $Y$) is amplified by the propagation of this shock to the cost of liquidity provision in the other security (say $X$). Indeed, as learning is two-sided, the change in the liquidity of security $X$ feeds back on the liquidity of security $Y$, which sparks a chain reaction amplifying the initial shock. Hence, liquidity is fragile in our model: a small exogenous drop in the liquidity of one market can ultimately result in a disproportionately large drop in the liquidity of this market and other related markets.

[3] For stocks listed on the NYSE, Hendershott, Li, Menkveld and Seasholes (2010) show that 25% of the monthly return variance is due to transitory price changes. Interestingly, they also find that transient price pressures are stronger when market-makers' inventories are relatively large. This finding implies that price movements are less informative when dealers' cost of liquidity provision is higher, in line with our model.

[4] In this paper, we measure liquidity by the sensitivity of prices to market order imbalances, as in Kyle (1985). The market is more liquid when this sensitivity is low. Empirically, this sensitivity can be measured by regressing price changes on order imbalances (see for instance Glosten and Harris (1988) or Korajczyk and Sadka (2008)).

[5] For instance, consider dealers in a stock and dealers in stock index futures. The stock return is determined both by a systematic factor and an idiosyncratic factor whereas the stock index futures return is only driven by the systematic factor. Suppose that dealers in the stock index futures are well informed on the systematic factor. In this case, dealers in the stock can learn information about the systematic factor from the price of the stock index futures whereas dealers in the stock index futures have nothing to learn from the price of individual stocks. In this case learning is one sided.

Second, when learning is two-sided, the model can feature multiple equilibria with differing levels of liquidity. The reason is as follows. Suppose that dealers in security $X$ *expect* a drop in the liquidity of security $Y$. Then, dealers in security $X$ expect the price of security $Y$ to be noisier, which makes the market for security $X$ less liquid. But as a consequence, the price of security $X$ becomes less informative for dealers in security $Y$ and the liquidity of security $Y$ drops, which validates the expectation of dealers in security $X$. Hence, dealers' expectations about the liquidity of the other security can be *self-fulfilling*. For this reason, there exist cases in which, for the same parameter values, the liquidity of securities $X$ and $Y$ can be either relatively high or relatively low.[6] A sudden switch from a high to a low liquidity equilibrium is an extreme form of co-variation in liquidity and fragility since it corresponds to a situation in which the liquidity of several related securities dries up without an apparent reason.

Third, an increase in the fraction of pricewatchers in a security has an ambiguous impact on the liquidity of this security. On the one hand, this increase improves liquidity because pricewatchers require a smaller compensation for inventory risk (as they have more information). On the other hand, entry of new pricewatchers impairs liquidity because it exposes inattentive dealers (i.e., dealers without price information) to adverse selection. Indeed, pricewatchers bid relatively conservatively for the security when they receive bad signals and relatively aggressively when they receive good signals. As a result, inattentive dealers are more likely to end up with relatively large (small) holdings when the value of the security is low (large). In reaction to this winner's curse, inattentive dealers shade their bids, which reduces market liquidity. The net effect on liquidity is always positive when dealers' risk bearing capacity (i.e., dealers' risk tolerance divided by the variance of dealers' aggregate dollar inventory) is low enough. Otherwise, an increase in the fraction of pricewatchers can impair market liquidity when the fraction of pricewatchers is small.

Fourth, the exposure of inattentive dealers to adverse selection implies that liquidity spillovers can be *negative*. To see why, suppose that the liquidity of security $Y$ improves. This improvement implies that the price of security $Y$ conveys more precise information to pricewatchers in security $X$. Thus, the informational disadvantage of inattentive dealers increases and, as a result, the liquidity of security $X$ may drop. For this to happen, we show that the fraction of pricewatchers must be small enough and dealers' risk bearing capacity must be large.

In a last step, we endogenize the fraction of pricewatchers by introducing a cost of attention to prices. There are several possible interpretations for this cost. It may simply reflect the fact that monitoring the price of other securities requires attention (it is time consuming) and human dealers have limited attention.[7] More importantly maybe, real-time data on prices are costly to acquire. Data vendors (Reuters, Bloomberg, etc...) or trading platforms charge a fee for real time datafeed.[8] In particular, some market-makers can choose to pay a "co-location"

[6]There also exist cases in which the equilibrium is unique, even if learning is two-sided.

[7]Recent empirical papers (Corwin and Coughenour (2008), Boulatov et al. (2010) and Chakrabarty and Moulton (2009)) find that attention constraints for NYSE specialists have an effect on market liquidity. Thus, modelling dealer attention is important to understand liquidity.

[8]Market participants often complain about these data fees.For instance, the fee charged by Nasdaq for the dissemination of corporate bond prices has been very controversial. For accounts of these debates, see, for in-

fee to trading platforms in order to obtain the right to place their computers close to platforms' matching engines. In this way, they possess a split second advantage in accessing and reacting to changes in prices. Last, in the absence of real time price reporting (as for instance in some OTC markets), real time price information is available only to a few privileged dealers and very costly to collect for other participants.[9]

When learning is one-sided, the value of price information declines with the fraction of pricewatchers. Thus, the equilibrium fraction of pricewatchers is unique and inversely related to the cost of price information. When dealers' risk bearing capacity is low, a decrease in the cost of price information leads to an improvement in liquidity. Otherwise, liquidity is a U-shaped function of this cost. Indeed, for relatively high values of the cost of price information, a decrease in this cost triggers entry of a few pricewatchers, which is a source of adverse selection risk and impairs liquidity, as explained previously.

In contrast, when learning is two-sided, the value of monitoring the price of, say, security $X$ for dealers in security $Y$ can *increase* with the fraction of pricewatchers in either security (for some parameter values). The reason is as follows. As explained previously, if dealers' risk bearing capacity is low enough, an increase in the fraction of pricewatchers in security $Y$ makes this security more liquid. This improvement in liquidity spreads to security $X$, which makes the price of this security more informative. Thus, information on the price of security $X$ becomes more valuable for dealers in security $Y$. Furthermore, the value of information on the price of security $X$ for dealers in security $Y$ also increases in the fraction of pricewatchers in security $X$. Indeed, as the number of pricewatchers in security $X$ increases, the price of this security becomes more informative, which strengthens its informational value for dealers in security $Y$.

This finding is surprising since usually the value of financial information declines with the number of investors buying information (Grossman and Stiglitz (1980) or Admati and Pfleiderer (1986)). This principle does not necessarily apply to price information because the precision of price information increases in the number of dealers buying this information.

One consequence is that dealers' decisions to acquire price information on other securities are self-reinforcing both within and across markets. As a result, there can be multiple levels of attention in equilibrium for a fixed value of the cost of attention to prices. In particular, *for identical parameter values*, the markets for the two securities can appear well integrated (the fraction of pricewatchers is high) or segmented (the fraction of pricewatchers is low). As an illustration we construct an example in which, for a fixed correlation in the payoffs of both securities, the markets for securities $X$ and $Y$ are either fully integrated (all dealers are pricewatchers) or segmented (no dealer is a pricewatcher). For dealers in security $X$, monitoring the price of the other security does not have much value if there are no pricewatchers in security $Y$ and vice versa. Thus, the situation in which the two markets are segmented is self-sustaining

stance, "Latest Market Data Dispute Over NYSE's Plan to Charge for Depth-of-Book Data Pits NSX Against Other U.S. Exchanges," Wall Street Technology, May 21, 2007; the letter to the SEC of the Securities Industry and Financial Markets Association (SIFMA) available at `http://www.sifma.org/regulatory/comment_letters/41907041.pdf`, and "TRACE Market Data Fees go to SEC," Securities Industry News, 6/3/2002.

[9]For instance, a bond dealer may be an employee of a trading firm also active in credit default swaps (CDS). In this way, the dealer may be privy of information on trades in CDSs written on the bond.

and can persist even if the cost of attention declines.

The mechanism that leads to liquidity spillovers in our model generates predictions distinct from the mechanisms based on funding constraints or systematic shifts in risk aversion described in Brunnemeier and Pedersen (2008), Gromb and Vayanos (2002) or Kyle and Xiong (2001). In our model, funding restrictions or an increase in risk aversion for dealers in one asset class (e.g., stocks) can initially spark a drop in the liquidity of this class of assets. However, in contrast to other theories of co-variations in liquidity supply, our model predicts that this shock can spread to other asset classes (e.g., bonds) even if there is no tightening of funding constraints for dealers in other asset classes. The only requirement is that the prices of assets in the first class are used as a source of information to value assets in other classes. Furthermore, as explained previously, in our model liquidity spillovers can be negative while theories based on funding constraints imply positive liquidity spillovers.

Isolating the role of cross-asset learning in liquidity spillovers is challenging empirically because this mechanism can operate simultaneously with other sources of systematic variations in liquidity. One way to address this difficulty consists in studying the effects of changes in trading technologies that affect dealers' ability to learn from the prices of other assets. One strategy is to consider cases in which a security switches from an opaque trading system (e.g., an OTC market) to a more transparent trading system (a case in point is the implementation of post trade transparency in the U.S. bond market in 2002). In this case, dealers in related securities can more easily use the information conveyed by the price of the previously opaque security. This is similar to a decrease in the cost of price information in our model. Another approach is to study the effect of changes in co-location fees. Indeed, dealers who co-locate can be seen as pricewatchers in our model (they have very quick access to prices of other securities and can thereby make their strategies contingent on these prices). Hence, variations in co-location fees should also affect the fraction of pricewatchers. We develop predictions about the effects of such changes in trading technologies in the last part of the paper.

Our model is related to models of contagion (King and Wadhwani (1990), Kodres and Pritsker (2002), or Pasquariello (2007)) and cross-asset price pressures (Andrade, Chang and Seasholes (2008), Bernhardt and Taub (2008), Pasquariello and Vega (2009), Boulatov, Hendershott and Livdan (2010). These models describe various mechanisms through which a shock on investors' information or liquidity traders' demand in one security can affect the *prices* of other securities.[10] None of these models however studies the role of cross-asset learning in the transmission of a *liquidity shock* (i.e., a change in the sensitivity of price to order imbalances) in one security to other securities, as we do here. Our paper is also linked to the literature on the value of financial information (e.g., Grossman and Stiglitz (1980), Admati and Pfleiderer (1986)). We contribute to this literature by studying the value of securities price information. As explained previously, we show that price information is special in the sense that its value can increase with the number of investors buying this information, an effect which does not arise in standard models of information acquisition. In this respect, our paper adds to the few

[10] Most of these models build upon the multi-asset pricing models of Admati (1985) and Caballe and Krishnan (1994).

papers identifying conditions under which the value of financial information may increase with the number of informed investors (Barlevy and Veronesi (2000), Veldkamp (2006), Chamley (2007), and Ganguli and Yang (2009)).

The rest of the paper is organized as follows. Section 2 describes the model. In Section 3, we consider the case in which the fraction of pricewatchers is fixed and we show how liquidity spillovers and multiple equilibria arise in this set-up. In Section 4, we study how the value of price information depends on the fraction of pricewatchers and we endogenize this fraction. Section 5 discusses testable implications of the model and Section 6 concludes. Proofs are collected in the Appendix or the Internet Appendix.

# 2 The model

We consider two securities, denoted $D$ and $F$. These securities pay-off at date 2 and their payoffs, $v_D$ and $v_F$, are given by a factor model with two risk factors $\delta_D$ and $\delta_F$, i.e.,

$$v_D = \delta_D + d_D \times \delta_F + \eta, \quad (1)$$

$$v_F = d_F \times \delta_D + \delta_F + \nu. \quad (2)$$

The random variables $\delta_D$, $\delta_F$, $\eta$ and $\nu$ are independent and have a normal distribution, with mean zero. The variance of $\eta$ is denoted $\sigma_\eta^2$. We make additional parametric assumptions that simplify the exposition without affecting our conclusions. First, there is no idiosyncratic risk for security $F$ (i.e., $\nu = 0$). Second, the variance of the factors is normalized to one. Third, we assume that $d_F = 1$ and $d_D \in [0, 1]$, so that the payoffs of the two securities are positively correlated. To simplify notations, we therefore denote $d_D$ by $d$. When $d = 0$, the payoff of security $D$ does not depend on factor $\delta_F$. Thus, the price of security $F$ cannot convey new information to dealers in security $D$. In this case, we say that learning is one-sided.

Trades in securities $D$ and $F$ take place at date 1. In each market, there are two types of traders: *(i)* a continuum of risk-averse speculators and *(ii)* liquidity traders. The aggregate demand of liquidity traders in market $j$ is $u_j \sim N(0, \sigma_{u_j}^2)$. Liquidity traders' demands in both markets are independent and are absorbed by speculators. Hence, in the rest of the paper, we refer to speculators as *dealers* and to $u_j$ as the size of the *demand shock* in market $j$.

Dealers are specialized: they are active in only one security. In this way, we rule out co-movements in liquidity which arise simply because the same dealers are active in multiple securities.[11] Dealers specialized in security $j$ have perfect information on factor $\delta_j$ and no information on factor $\delta_{-j}$. However, they can follow the price of the other security to obtain information on this factor. We denote by $\mu_j$ the fraction of dealers specialized in security $j$ who monitor the price of security $-j$ and we refer to $\mu_j$, as the *level of attention* to security $-j$.

[11] In reality, dealer firms are active in multiple securities. However, these firms delegate trade-related decisions to individuals who operate on specialized trading desks. Naik and Yadav (2003) show empirically that the decision-making of these trading desks is largely decentralized (e.g., dealers' trading decisions within a firm are mainly driven by their own inventory exposure rather than the aggregate inventory exposure of the dealer firm to which they belong). Their results suggest that there is no direct centralized information sharing between dealers within these firms.

We refer to these dealers as being *pricewatchers*. Other dealers are called *inattentive dealers*. We use $W$ to index the decisions made by pricewatchers and $I$ to index the decisions made by inattentive dealers. The polar cases, in which there are either no pricewatchers in either market ($\mu_D = \mu_F = 0$) or all dealers are pricewatchers ($\mu_D = \mu_F = 1$) are called the "no attention case" and the "full attention case," respectively. Table 1 summarizes the various possible cases that will be considered in the paper.

| Attention/Learning | One-Sided: $d = 0$ | Two-Sided: $d > 0$ |
|---|---|---|
| No Attention | $\mu_D = \mu_F = 0$ | $\mu_D = \mu_F = 0$ |
| Limited Attention | $\mu_j > 0$ and $\mu_{-j} < 1$ | $\mu_j > 0$ and $\mu_{-j} < 1$ |
| Full Attention | $\mu_D = \mu_F = 1$ | $\mu_D = \mu_F = 1$ |

Table 1: Various Cases

Each dealer in market $j$ has a CARA utility function with risk tolerance $\gamma_j$. Thus, if dealer $i$ in market $j$ holds $x_{ij}$ shares of the risky security, her expected utility is

$$E\left[U\left(\pi_{ij}\right)|\delta_j, \mathcal{P}_j^k\right] = E\left[-\exp\left\{-\gamma_j^{-1}\pi_{ij}\right\}|\delta_j, \mathcal{P}_j^k\right], \tag{3}$$

where $\pi_{ij} = (v_j - p_j)x_{ij}$ and $\mathcal{P}_j^k$ is the *price information* available to a dealer with type $k \in \{W, I\}$ operating in security $j$.

As dealers submit price contingent demand functions, they all act as if they were observing the clearing price in their market. Thus, we have $\mathcal{P}_j^W = \{p_j, p_{-j}\}$ and $\mathcal{P}_j^I = \{p_j\}$. We denote the demand function of a pricewatcher by $x_j^W(\delta_j, p_j, p_{-j})$ and that of an inattentive dealer by $x_j^I(\delta_j, p_j)$.[12] In each period, the clearing price in security $j$, $p_j$, is such that the demand for this security is equal to its supply, i.e.,

$$\mu_j x_j^W(\delta_j, p_j, p_{-j})di + (1-\mu_j)x_j^I(\delta_j, p_j)di + u_j = 0, \quad \text{for } j \in \{D, F\}. \tag{4}$$

As in many other papers (e.g., Kyle (1985) or Vives (1995)), we will measure the level of illiquity in security $j$ by the sensitivity of the clearing price to the demand shock (i.e., $\partial p_j/\partial u_j$). In equilibrium, the aggregate inventory position of dealers in security $j$ after trading at date 1 is $-u_j$ and the total dollar value of this position at date 1 is $-u_j \times v_j$. The risk associated with this position for dealers in security $j$ can be measured by its variance conditional on information on risk factor $\delta_j$, i.e., $\sigma_{u_j}^2 \text{Var}[v_j|\delta_j]$. Thus, the ratio of dealers' risk tolerance to this variance (the total amount of risk taken by the dealers) is a measure of the risk bearing capacity of the

[12] As pricewatchers observe the price in security $-j$, they can make their trading strategy in security $j$ contingent on this price. Alternatively, one can assume that pricewatchers do not observe directly the price of security $-j$ but are allowed to place limit orders (a demand function) in security $j$ contingent on the price of other securities. Such indexed limit orders have been proposed by Black (1995) but are typically not offered by exchanges. See Cespa (2004) for an analysis of trading mechanisms that allow multi-price contingent orders.

market. We denote this ratio by $\mathcal{R}_j$:

$$\mathcal{R}_j = \frac{\gamma_j^2}{\sigma_{u_j}^2 \text{Var}[v_j|\delta_j]}. \tag{5}$$

The higher is $\mathcal{R}_j$, the higher is the risk bearing capacity of the dealers in security $j$. As we shall see this ratio plays an important role for some of our findings.

There are several ways to interpret the two securities in our model. For instance, as in King and Wadhwani (1990), securities $D$ and $F$ could be two stock market indexes for two different countries. Alternatively, they could represent a derivative and its underlying security. For instance, security $D$ could be a credit default swap (CDS) and security $F$ the stock of the firm on which the CDS is written. When $d = f = 1$ and $\sigma_\eta^2 = 0$, the payoff of the two securities is identical, as in Chowdry and Nanda (1991). In this case, the two securities can be viewed as the stock of a cross-listed firm and its American Depository Receipt (ADR) in the U.S. for instance. Factor $\delta_F$ can then be viewed as the component of the firm's cash-flows that comes from its sales in the U.S. In each of these cases, it is natural to assume that dealers have specialized information. For instance, dealers in country $j$ will be well informed on local fundamental news but not on foreign fundamental news as in King and Wadhwani (1990).[13]

# 3 Attention and liquidity spillovers

## 3.1 Benchmark: No attention

We first analyze the equilibrium in the no attention case ($\mu_D = \mu_F = 0$). For instance, the markets for securities $D$ and $F$ may be opaque so that dealers in each security can obtain information on the price of the other security only after some delay. Alternatively, the prices of each security are available in real time but accessing this information is so costly that no dealer chooses to be informed on the price of the other security (see Section 4).

**Lemma 1.** *(Benchmark) When $\mu_F = \mu_D = 0$, the equilibrium price in market $j$ is:*

$$p_j = \delta_j + B_{j0} u_j, \tag{6}$$

*with $B_{D0} = \gamma_D^{-1}(\sigma_\eta^2 + d^2)$ and $B_{F0} = \gamma_F^{-1}$.*

The sensitivity of the equilibrium price for security $j$ to the aggregate demand shock in this market, the illiquidity of security $j$, is given by $B_{j0}$ (we use index "0" to refer to the case in which $\mu_F = \mu_D = 0$). In the no attention case, the illiquidity of security $D$ is determined by parameters $\sigma_\eta^2$, $d$, and $\gamma_D$. We refer to these parameters as being the "liquidity fundamentals" of security $D$. Similarly, we refer to $\gamma_F$ as a liquidity fundamental of security $F$ since it only affects the illiquidity of security $F$. Illiquidity increases with dealers' risk aversion ($\gamma_j$ decreases) and uncertainty on the securities' payoffs ($\sigma_\eta^2$ increases).

[13] In the case of the CDS market, dealers in CDS are often affiliated with lenders and therefore better informed on the likelihood of defaults (and size of associated losses) than dealers in the stock market (see Acharya and Johnson (2007))

Importantly, in the benchmark case, there are no liquidity spillovers: a change in the illiquidity fundamental of one market does not affect the illiquidity of the other market. For instance, an increase in the risk tolerance of dealers in security $D$ makes this security more liquid but it has no effect on the illiquidity of the other security.[14] In contrast, with limited or full attention, a change in the illiquidity fundamental of one security will affect the illiquidity of the other security, as shown in the next sections.

## 3.2 Liquidity spillovers with full attention

In this section, we consider the case in which *all* dealers are pricewatchers, that is the full attention case ($\mu_D = \mu_F = 1$). The analysis is more complex than in the benchmark case as dealers in one security extract information about the factor that is unknown to them from the price of the other security. To solve this signal extraction problem, dealers must form beliefs on the relationship between clearing prices and risk factors. We will focus on equilibria in which these beliefs are correct, i.e., the rational expectations equilibria of the model. We first show that, in contrast to the benchmark case, the levels of illiquidity of both markets are interdependent and this interdependence leads to multiple equilibria (Section 3.2.1). We then provide an explanation for this finding and we show that that the interdependence in the illiquidity of securities $D$ and $F$ leads to liquidity spillovers: a shock to the illiquidity fundamental of one security propagates to the other security (Section 3.2.2). Finally, we show that when learning is two-sided, the total effect of a small shock on the illiquidity fundamental of one security can be much larger than the initial effect of such a shock (Section 3.2.3).

### 3.2.1 Equilibria with full attention

In our model, a linear rational expectations equilibrium is a set of prices $\{p^*_{j1}\}_{j\in\{D,F\}}$ such that

$$p^*_{j1} = R_{j1}\delta_j + B_{j1}u_j + A_{j1}\delta_{-j} + C_{j1}u_{-j}, \tag{7}$$

and $p^*_{j1}$ clears the market of asset $j$ for each realization of $\{u_j,\ \delta_j,\ u_{-j}, \delta_{-j}\}$ when dealers anticipate that clearing prices satisfy equation (7) and choose their trading strategy to maximize their expected utility (given in equation (3)). We say that the equilibrium is non-fully revealing if pricewatchers in security $j$ cannot infer perfectly the realization of risk factor $\delta_{-j}$ from observing the price of security $-j$. The sensitivity of the price in market $j$ to the demand shock in this market, i.e., the "illiquidity of market $j$," is measured by $B_{j1}$ in the full attention case. Index "1" is used to refer to the equilibrium when $\mu_D = \mu_F = 1$.

**Proposition 1.** *With full attention and $\sigma^2_\eta > 0$, there always exists a non-fully revealing linear rational expectations equilibrium. At any non-fully revealing equilibrium, $B_{j1} > 0$, $R_{j1} = 1$ and*

[14] In our model, a variation in risk tolerance of dealers in one security is just one way to vary the cost of liquidity provision for dealers in one asset class. In reality variations in this cost may be due to variations in risk tolerance, inventory limits or financing constraints for dealers in this asset class. The important point is that they do not directly affect dealers in other asset classes.

*the coefficients, $A_{j1}$ and $C_{j1}$ can be expressed as functions of $B_{j1}$ and $B_{-j1}$. Moreover*

$$B_{D1} = f_1(B_{F1}; \gamma_D, \sigma^2_\eta, d, \sigma^2_{u_F}) = \frac{\sigma^2_\eta}{\gamma_D} + \frac{d^2 B^2_{F1}\sigma^2_{u_F}}{\gamma_D(1 + B^2_{F1}\sigma^2_{u_F})}, \tag{8}$$

$$B_{F1} = g_1(B_{D1}; \gamma_F, \sigma^2_{u_D}) = \frac{B^2_{D1}\sigma^2_{u_D}}{\gamma_F(1 + B^2_{D1}\sigma^2_{u_D})}. \tag{9}$$

Proposition 1 shows that the illiquidities of securities $D$ and $F$ are interdependent since $B_{D1}$ is a function of $B_{F1}$ and vice versa. Moreover, all coefficients in the equilibrium price function can be expressed as functions of the illiquidity of securities $D$ and $F$. Thus, the number of non-fully revealing linear rational expectations equilibria is equal to the number of pairs $\{B^*_{D1}, B^*_{F1}\}$ solving the system of equations (8) and (9). In general, we cannot characterize these solutions analytically and therefore cannot solve for the equilibria in closed-form. However, we can find these solutions numerically. In Figure 2 we illustrate the determination of the equilibrium levels of illiquidity by plotting the functions $f_1(\cdot)$ and $g_1(\cdot)$ for specific values of the parameters.

[**Insert Figure 2 about here**]

The equilibria are the values of $B_{D1}$ and $B_{F1}$ at which the curves representing the functions $f_1(\cdot)$ and $g_1(\cdot)$ intersect. In panel (a) we set $\gamma_j = d = 1$, $\sigma_{u_j} = 2$, and $\sigma_\eta = 0.2$. In this case, we obtain three equilibria: one with a low level of illiquidity, one with a medium level of illiquidity and one with a relatively high level of illiquidity. In panels (b) and (c), we pick values of $\sigma_\eta$ or $d$ such that the correlation between the payoffs of securities $D$ and $F$ is smaller ($\sigma_\eta = 1$ in panel (b) while $d = 0.9$ in panel (c)). In this case, we obtain a unique equilibrium. More generally, when $d$ is low relative to $\sigma^2_\eta$, the model has a unique rational expectations equilibrium, as shown in the next corollary.

**Corollary 1.** *If $4d^2 < \sigma^2_\eta$ and $\mu_D = \mu_F = 1$ then there is a unique non-fully revealing rational expectations equilibrium.*

In particular, when learning is one sided ($d = 0$), there exists a unique non-fully revealing linear rational expectations equilibrium. Furthermore, in this case, we can characterize the equilibrium in closed-form (see Corollary 6 below).[15]

The case in which $\sigma^2_\eta = 0$ requires a separate analysis. In this case, it is still true that if there exists a non-fully revealing equilibrium then $B_{D1}$ and $B_{F1}$ solve the system of equations (8) and (9). However, in this case, the unique solution to this system of equations can be $B_{D1} = B_{F1} = 0$ so that a non-fully revealing equilibrium does not exist. As an example, consider the case in which the two securities are identical: $d = 1$, $\sigma^2_\eta = 0$, $\gamma_F = \gamma_D = \gamma$, $\sigma^2_{u_j} = \sigma^2_u$. We refer to this case as the *symmetric case.*

[15] The condition given in Corollary 1 is sufficient to guarantee the existence of a unique rational expectations equilibrium when all dealers are pricewatchers, but it is not necessary. Numerical simulations show that there exist multiple equilibria when $d$ is high relative to $\sigma^2_\eta$. Moreover it can be shown formally that the model has either one or three non-fully revealing rational expectations equilibria.

**Lemma 2.** *In the symmetric case with full attention, if $\sigma_u^2 > 4\gamma^2$, there are two non fully revealing linear rational expectations equilibria: a "High" illiquidity equilibrium and a "Low" illiquidity equilibrium. The levels of illiquidity in each of these equilibria are*

$$\text{High :} \qquad B^{H*} = \frac{\sigma_u + (\sigma_u^2 - 4\gamma^2)^{1/2}}{2\gamma\sigma_u}, \tag{10}$$

$$\text{Low:} \qquad B^{L*} = \frac{\sigma_u - (\sigma_u^2 - 4\gamma^2)^{1/2}}{2\gamma\sigma_u}, \tag{11}$$

*with $B^{H*} > B^{L*}$. If $\sigma_u^2 < 4\gamma^2$, a non-fully revealing equilibrium does not exist.*

#### 3.2.2 Cross-asset learning and liquidity spillovers

We now explain why cross-asset learning is naturally conducive to multiple equilibria and liquidity spillovers. To this end, it is useful to analyze in detail how dealers in one security extract information from the price of the other security. Our starting point is the following lemma.

**Lemma 3.** *With full attention, in any non-fully revealing linear rational expectations equilibrium,*

$$p_j^* = (1 - A_{j1}A_{-j1})\omega_j + A_{j1}p_{-j}^*, \text{ for } j \in \{F, D\}. \tag{12}$$

*where $\omega_j \equiv \delta_j + B_{j1}u_j$ for $j \in \{D, F\}$. Hence, $\omega_{-j}$ is a sufficient statistic for the price information, $\mathcal{P}_j^W = \{p_j^*, p_{-j}^*\}$, available to pricewatchers operating in security $j$.*

In other words, $\omega_{-j}$ is the signal about the risk factor $\delta_{-j}$ that pricewatchers operating in security $j$ extract from the price of security $-j$. In the absence of information on the price of security $-j$, the precision of the forecast formed by dealers in security $j$ about the payoff of security $j$ is $(\text{Var}[v_j|\delta_j])^{-1}$. In contrast, with access to price information, the precision of this forecast is[16]

$$\text{Var}[v_j|\delta_j, \omega_{-j}]^{-1} = (\text{Var}[v_j|\delta_j]\left(1 - \rho_{j1}^2\right))^{-1}, \tag{13}$$

where

$$\rho_{j1}^2 \overset{def}{=} \frac{E[v_j\omega_{-j}\,|\delta_j\,]^2}{\text{Var}[v_j\,|\delta_j\,]\text{Var}[\omega_{-j}]}. \tag{14}$$

Hence, the higher $\rho_{j1}^2$ is, the greater the informativeness of the signal conveyed by the price of security $-j$ to dealers in security $j$. For this reason, we refer to $\rho_{j1}^2$ as the informativeness of the price of security $-j$ about the payoff of security $j$ for dealers operating in security $j$. Using the definition of $\omega_j$, we obtain

$$\rho_{D1}^2 = \frac{d^2}{(\sigma_\eta^2 + d^2)(1 + B_{F1}^2\sigma_{u_F}^2)}, \tag{15}$$

$$\rho_{F1}^2 = \frac{1}{1 + B_{D1}^2\sigma_{u_D}^2}. \tag{16}$$

[16]This result follows from the fact that if $X$ and $Y$ are two random variables with normal distribution then $\text{Var}[X|Y] = \text{Var}[X] - \text{Cov}^2[X, Y]/\text{Var}[Y]$ and the fact that $E[\omega_{-j}\,|\delta_j\,] = 0$.

When $d = 0$, the price of security $F$ does not convey information to dealers in security $D$ ($\rho^2_{D1} = 0$) since the payoff of security $D$ does not depend on the risk factor known to dealers in security $F$. Using the expressions for $B_{j1}$ given in Proposition 1, we obtain that

$$B_{j1} = B_{j0}(1 - \rho^2_{j1}). \tag{17}$$

This observation yields the following result.

**Corollary 2.** *The markets for securities D and F are less illiquid with full attention than with no attention, i.e., $B_{j1} \leq B_{j0}$. Moreover, with full attention, an increase in the informativeness of the price of security $-j$ for dealers in security $j$ makes security $j$ more liquid, i.e.,*

$$\frac{\partial B_{j1}}{\partial \rho^2_{j1}} \leq 0. \tag{18}$$

The intuition for this result is straightforward. By watching the price of another security, dealers learn information. Hence, they face less uncertainty about the payoff of the security in which they are active. For this reason, with full attention, dealers require a smaller premium than with no attention to absorb a given demand shock (first part of the corollary) and this premium decreases with the informativeness of prices (last part of the corollary).

Price movements in security $j$ are driven both by news about factor $\delta_j$ and demand shocks specific to this security. The contribution of demand shocks to price variations becomes relatively higher when security $j$ becomes more illiquid. As a consequence the price of security $j$ becomes less informative for dealers in other markets when security $j$ becomes more illiquid. To see this, remember that the signal about factor $\delta_j$ conveyed by the price of security $j$ to dealers in security $-j$ is $\omega_j = \delta_j + B_{j1}u_j$. Clearly, this signal is noisier when $B_{j1}$ is higher, which yields the following result.

**Corollary 3.** *With full attention, an increase in the illiquidity of security $j$ makes its price less informative for dealers in security $-j$:*

$$\frac{\partial \rho^2_{-j1}}{\partial B_{j1}} \leq 0. \tag{19}$$

Corollaries 2 and 3 explain why the illiquidity of security $D$ and $F$ are interdependent when dealers in the two securities learn from each other's prices. Indeed, the illiquidity of security $-j$ determines the informativeness of the price of this security for dealers in security $j$ (Corollary 3) and as a result the illiquidity of security $j$ (Corollary 2).

This observation helps us to understand how multiple equilibria can arise when dealers learn from each other's prices. Consider dealers in security $F$. They do not directly observe the sensitivity of the price to demand shocks in security $D$, i.e., the illiquidity of security $D$. Hence, ultimately, the informativeness of the price of security $D$ for dealers in security $F$ depends on their belief regarding the illiquidity of security $D$. Similarly, the informativeness of the price of security $F$ for dealers in security $D$ depends on their belief regarding the illiquidity of security $F$. In sum, the illiquidity of security $j$ depends on the beliefs of the dealers active in this security about the illiquidity of security $-j$, which itself depends on the beliefs of its dealers about the illiquidity of security $j$. This loop leads to multiplicity as, for the same values of the exogenous parameters, various systems of beliefs can be self-sustaining.[17]

This circularity breaks down when dealers in security $D$ do not use the information contained in the price of security $F$ (either because $\mu_D = 0$ or because $d = 0$). In this case, the illiquidity of security $D$ is uniquely pinned down by its "fundamentals" ($\gamma_D$ and $\sigma_\eta^2$) and, as a result, the beliefs of dealers in security $F$ regarding the liquidity of security $D$ are uniquely defined as well (since dealers' expectations about the illiquidity of the other security must be correct in equilibrium). More generally, when $d$ is low relative to $\sigma_\eta^2$, security $D$ is not much exposed to factor $\delta_F$. Thus, the beliefs of dealers in security $D$ about the liquidity of security $F$ play a relatively minor role in the determination of the liquidity of security $D$ and, for this reason, the equilibrium is unique, as shown in Corollary 1.

The interdependence in the illiquidity of securities $D$ and $F$ has another implication. In contrast to the benchmark case, an exogenous change in the illiquidity of one market (due for instance to an increase in dealers' risk tolerance in this market) affects the illiquidity of the other market. We call this effect a *liquidity spillover*. To see this point, consider the effect of an increase in the risk tolerance of dealers in security $D$. The immediate effect of this increase is to make security $D$ more liquid as in the benchmark case. Hence, its price becomes more informative for dealers in security $F$ (Corollary 3), which then becomes more liquid (Corollary 2) because inventory risk for dealers in security $F$ is smaller when they are all better informed. Thus, the improvement in the liquidity of security $D$ spreads to liquidity $F$, although security $F$ experiences no change in its liquidity fundamentals.

More formally, consider the system of equations (8) and (9). Other things equal, an increase in the risk tolerance of dealers in security $D$ makes this security more liquid since $\partial f_1/\partial \gamma_D < 0$. In turn this improvement spreads to security $F$ because $\partial g_1/\partial B_{D1} \neq 0$. More generally, any exogenous change in the illiquidity of security $D$ will spill over to security $F$ because $\partial g_1/\partial B_{D1} \neq 0$. Similarly, an exogenous change in the illiquidity of security $F$ will spill over to security $D$ when $\partial f_1/\partial B_{F1} \neq 0$. The direction (positive/negative) of these liquidity spillovers is determined by the signs of $\partial g_1/\partial B_{D1}$ and $\partial f_1/\partial B_{F1}$.

**Corollary 4.** *With full attention, liquidity spillovers are always positive, i.e.,* $\partial f_1/\partial B_{F1} \geq 0$ *and* $\partial g_1/\partial B_{D1} > 0$. *Moreover when learning is one sided (*$d = 0$*), there is no spillover from security* $F$ *to security* $D$ *because the price of security* $F$ *conveys no information to dealers in security* $D$. *In contrast, when learning is two-sided (*$d > 0$*), liquidity spillovers operate in both directions.*

Intuitively, positive liquidity spillovers generate positive co-movements in illiquidity across-securities. In our model, illiquidity is not stochastic (it is a deterministic function of the parameters). However, we can create variations in illiquidity by picking randomly the exogenous

[17]Ganguli and Yang (2009) consider a single security model of price formation similar to Grossman and Stiglitz (1980). They show that multiple non-fully revealing linear rational expectations equilibria arise when investors have private information both on the asset payoff and the aggregate supply of the security. The source of multiplicity here is different since dealers have no supply information in our model.

parameters of the model (e.g., the risk tolerance of dealers in security $D$) and compute the resulting covariance for illiquidity of securities $F$ and $D$. Figure 5 in Section 3.3 provides an example that shows how positive liquidity spillovers result in positive covariation in liquidity.

### 3.2.3 Amplification: the illiquidity multiplier

With two-sided learning, liquidity spillovers operate in both directions. As a consequence, the total effect of a small change in the illiquidity fundamentals of one security is higher than the direct effect of these changes.

To see this consider the chain of effects that follows a small *reduction*, denoted by $\Delta\gamma_D < 0$, in the risk tolerance of dealers in security $D$. The *direct effect* of this reduction is to increase the illiquidity of security $D$ by $(\partial f_1/\partial\gamma_D)\Delta\gamma_D > 0$. As a consequence, the price of this security becomes less informative. Hence, dealers in security $F$ face more uncertainty and security $F$ becomes less liquid as well, although its liquidity fundamental ($\gamma_F$) is *unchanged*. The immediate increase in the illiquidity of security $F$ is equal to $(\partial g_1/\partial B_{D1})(\partial f_1/\partial\gamma_D)\Delta\gamma_D > 0$. When learning is two sided ($d > 0$), this increase in illiquidity for security $F$ leads to an even larger increase in the illiquidity of security $D$, starting a new vicious loop (as the increase in illiquidity for security $D$ leads to a further increase in illiquidity for security $F$ etc,...). As a result, the total effect of the initial decrease in the risk tolerance of dealers in security $D$ is an order of magnitude larger than its direct effect on the illiquidity of both securities. The next corollary formalizes this discussion.

**Corollary 5.** *Let*

$$\kappa \equiv \frac{1}{(1-(\partial g_1/\partial B_{D1})(\partial f_1/\partial B_{F1}))}, \tag{20}$$

*and assume that $d > 0$. With full attention, the total effects of a change in the risk tolerance of dealers in security $D$ is given by*

$$\underbrace{\frac{dB_{D1}}{d\gamma_D}}_{\textit{Total Effect}} = \kappa \underbrace{\frac{\partial f_1}{\partial\gamma_D}}_{\textit{Direct Effect}} < 0,$$

$$\underbrace{\frac{dB_{F1}}{d\gamma_D}}_{\textit{Total Effect}} = \kappa \underbrace{\frac{\partial g_1}{\partial B_{D1}}\frac{\partial f_1}{\partial\gamma_D}}_{\textit{Direct Effect}} < 0.$$

*and there always exists at least one equilibrium in which $\kappa > 1$.*

Thus, the initial effects of a small change in the risk tolerance of dealers in security $D$ are amplified by a factor $\kappa$. We call $\kappa$ the "illiquidity multiplier." This illiquidity multiplier can be relatively large when the illiquidity of each market is very sensitive to the illiquidity of the other market ($(\partial g_1/\partial B_{D1})(\partial f_1/\partial B_{F1})$ is high). In this sense, cross-asset learning is a source of fragility for financial markets.[18]

[18] Allen and Gale (2004) define a financial market as being fragile if "*small shocks have disproportionately large effects.*" (Allen and Gale (2004), page 1015).

Figure 3 illustrates this point for specific values of the parameters (in all our numerical examples we choose the parameter values such that there is a unique rational expectations equilibrium, except otherwise stated). It shows the value of $\kappa$ for various values of the idiosyncratic risk of security $D$ ($\sigma_\eta$) and the resulting values for the direct and total effects of a change in this risk tolerance on the illiquidity of securities $D$ and $F$, as a function of $\sigma_\eta$. In this example, the total drop in illiquidity of each security after a decrease in risk tolerance for dealers in security $D$ can be up to ten times bigger than the direct effect of this drop.

Table 2 provides another perspective on the illiquidity multiplier by showing the elasticity, denoted $\mathcal{E}_{B_{j1},\gamma_D}$, of illiquidity in each security to the risk tolerance of dealers in security $D$, i.e., the percentage change in illiquidity in each security for a one percent increase in the risk tolerance of dealers in security $D$. The table also shows the elasticity that would be obtained ($\hat{\mathcal{E}}_{B_{j1},\gamma_D}$) in the absence of the illiquidity multiplier (e.g., $\kappa = 1$ if $\mu_D = 0$). For instance, when $\gamma_D = 1.8$, a drop of 1% in the risk tolerance of dealers in security $D$ triggers an increase of 9% in the illiquidity of security $D$ and 14.9% in the illiquidity of security $F$. This is much larger than what would be obtained in the absence of bi-directional spillovers (e.g., if $\mu_D = 0$) since in this case the illiquidity of securities $D$ and $F$ would increase by only 1% and 1.5% respectively.

| $\gamma_D$ | $\kappa$ | $B_{D1}$ | $B_{F1}$ | Elasticities | | | |
|---|---|---|---|---|---|---|---|
| | | | | $\mathcal{E}_{B_{D1},\gamma_D}$ | $\hat{\mathcal{E}}_{B_{D1},\gamma_D}$ | $\mathcal{E}_{B_{F1},\gamma_D}$ | $\hat{\mathcal{E}}_{B_{F1},\gamma_D}$ |
| 2.2 | 1.54 | 0.19 | 2.11 | $-1.54$ | $-1.00$ | $-2.80$ | $-1.81$ |
| 2 | 2.16 | 0.23 | 2.87 | $-2.16$ | $-1.00$ | $-3.80$ | $-1.76$ |
| 1.8 | 9.94 | 0.36 | 5.94 | $-9.49$ | $-1.00$ | $-14.95$ | $-1.50$ |
| 1.62 | 2.35 | 0.57 | 11.01 | $-2.35$ | $-1.00$ | $-2.54$ | $-1.08$ |
| 1.46 | 1.65 | 0.70 | 13.41 | $-1.65$ | $-1.00$ | $-1.45$ | $-0.88$ |
| 1.31 | 1.39 | 0.82 | 15.29 | $-1.39$ | $-1.00$ | $-1.00$ | $-0.72$ |

Table 2: The table shows the impact of the illiquidity multiplier for different shocks to the risk aversion of dealers in market $D$. Other parameter values are $d = 1$, $\sigma_\eta = .62$, $\sigma_{u_F} = .1$, $\sigma_{u_D} = 1.6$, $\gamma_D = 1.8$, and $\gamma_F = 1/24$.

The corollary focuses on the effect of an increase in the risk tolerance of dealers in security $D$ but the effects of a change in the other exogenous parameters of the model ($\gamma_F$ and $\sigma^2_\eta$) are also magnified for the same reasons.

Last, we note that when the equilibrium is unique, it is necessarily such that $\kappa > 1$ (an implication of the last part of Corollary 5). When there are multiple equilibria, there is in general one equilibrium for which $\kappa < 0$. This equilibrium delivers "unintuitive"comparative statics.[19] For instance, in this equilibrium, a reduction in the risk tolerance of dealers in, say, security $D$ *increases* the liquidity of both securities. Such an equilibrium may exist because,

[19]It is possible to show that the model has three equilibria when it admits multiple equilibria. The equilibrium with $\kappa < 0$ is unstable whereas the two other equilibria (for which $\kappa > 1$) are stable.

as explained previously, the illiquidity of each security is in part determined by dealers' beliefs about the illiquidity of the other market. These beliefs may be disconnected from the illiquidity fundamentals of each security and yet be self-fulfilling.

## 3.3 Limited attention, adverse selection, and negative liquidity spillovers

We now turn our attention to the more general case in which $0 < \mu_D \leq 1$ and $0 < \mu_F \leq 1$. That is, we allow for limited attention by dealers in either security. In this case, the pricewatchers (dealers who monitor the price of the other security) have an informational advantage over inattentive dealers (dealers who do not monitor this price). This advantage is a source of adverse selection for inattentive dealers. This effect yields two new results: (a) liquidity spillovers can be negative and (b) an increase in the fraction of pricewatchers in one security can reduce the liquidity of this security when the fraction of pricewatchers is small. We now explain the intuition for these two results in more details. We proceed as follows. We first generalize Proposition 1 when attention is limited (Section 3.3.1). We then show that liquidity spillovers can be negative with limited attention and we provide a sufficient condition on the parameters for liquidity spillovers to be always positive (Section 3.3.2). Finally, we study the effect of a change in the fraction of pricewatchers in a security on the liquidity of this security (Section 3.3.3).

### 3.3.1 Equilibria with limited attention

As with full attention, a linear rational expectations equilibrium is a set of prices $\{p_j^*\}_{j\in\{D,F\}}$ such that

$$p_j^* = R_j\delta_j + B_j u_j + A_j\delta_{-j} + C_j u_{-j}, \tag{21}$$

and $p_j^*$ clears the market of asset $j$ for each realizations of $\{u_j, \delta_j, u_{-j}, \delta_{-j}\}$ when dealers anticipate that clearing prices satisfy equation (21) and choose their trading strategies to maximize their expected utility. The next proposition generalizes Proposition 1 when $0 < \mu_D \leq 1$ and $0 < \mu_F \leq 1$.

**Proposition 2.** *Suppose $\sigma_\eta^2 > 0$. With limited attention (i.e., $0 < \mu_D \leq 1$ and $0 < \mu_F \leq 1$), there always exists a non fully revealing linear rational expectations equilibrium. At any non-fully revealing equilibrium, $B_j > 0$, $R_j = 1$ and the coefficients $A_j$ and $C_j$ can be expressed as functions of $B_j$ and $B_{-j}$. Moreover*

$$B_j = B_{j0}(1-\rho_j^2) \times \frac{\gamma_j^2\mu_j\rho_j^2 + \sigma_{u_j}^2 \mathrm{Var}[v_j|\delta_j](1-\rho_j^2)}{\gamma_j^2\mu_j^2\rho_j^2 + \sigma_{u_j}^2 \mathrm{Var}[v_j|\delta_j](1-\rho_j^2)(1-\rho_j^2(1-\mu_j))}, \tag{22}$$

*where $\rho_D^2 \equiv d^2/((\sigma_\eta^2 + d^2)(1 + B_F^2\sigma_{u_F}^2))$ and $\rho_F^2 \equiv 1/(1 + B_D^2\sigma_{u_D}^2)$.*

Proposition 2 generalizes Proposition 1 when attention is limited. As in the full attention case, it can be shown that (i) pricewatchers in security $j$ extract a signal $\omega_{-j} = \delta_{-j} + B_{-j}u_{-j}$ from the price of security $-j$ and that (ii) variable $\rho_j^2$ is the informativeness of this signal. As the pricewatchers' trading strategy depends on the information they obtain from watching the price of security $-j$ (i.e., $\omega_{-j}$), the price of security $j$ partially reveals pricewatchers' private information.[20] Equation (21) implies that observing the price of security $j$ and risk factor $\delta_j$ is informationally equivalent to observing $\hat{\omega}_j \equiv A_j\delta_{-j} + B_j u_j + C_j u_{-j}$. Thus, in equilibrium, the information set of inattentive dealers, $\{\delta_j, p_j\}$, is informationally equivalent to $\{\delta_j, \hat{\omega}_j\}$. In what follows, we refer to $\hat{\omega}_j$ as inattentive dealers' price signal. Using the expressions for $A_j$ and $C_j$ (given in the proof of Proposition 2), we obtain that $\hat{\omega}_j = A_j\omega_{-j} + B_j u_j$. Hence, when $B_j > 0$, inattentive dealers' price signal is less precise than pricewatchers' price signal, which means that inattentive dealers in security $j$ are at an informational disadvantage compared to pricewatchers.

This disadvantage creates an adverse selection problem for the inattentive dealers. Indeed, relative to inattentive dealers, pricewatchers will bid aggressively when the price of security $-j$ indicates that the realization of the risk factor $\delta_{-j}$ is high and conservatively when the price of security $-j$ indicates that the realization of the risk factor $\delta_{-j}$ is low. As a consequence, inattentive dealers in one security will tend to have relatively large holdings of the security when its value is low and relatively small holdings of the security when its value is large. This bias in inattentive dealers' portfolio holdings is a source of adverse selection, which is absent when all dealers are pricewatchers. This new effect is key to understanding why liquidity spillovers may be negative in the limited attention case (see below).

Substituting $\rho_D^2$ and $\rho_F^2$ by their expressions in equation (22), we can express $B_j$ as a function of $B_{-j}$. Formally, we obtain:

$$B_D = f(B_F; \mu_D, \gamma_D, \sigma_\eta^2, d, \sigma_{u_F}^2), \tag{23}$$

$$B_F = g(B_D; \mu_F, \gamma_F, \sigma_{u_D}^2), \tag{24}$$

where the expressions for the functions $f(\cdot)$ and $g(\cdot)$ are given in the Appendix (see equations (A.26) and (A.28)). The linear rational expectations equilibria are completely characterized by the solution(s) of this system of equations. As in the full attention case and for the same reason, there might be multiple equilibria and we cannot in general provide an analytical characterization of these equilibria. Of course, when $\mu_D = \mu_F = 1$, the solutions to the previous system of equations are those obtained in the full attention case since this case is nested in the limited attention case.

### 3.3.2 When are liquidity spillovers positive?

As mentioned previously, liquidity spillovers from one security to the other can be negative when the fraction of pricewatchers in the latter security is relatively small. The intuition for negative spillovers is more easily seen when learning is one sided ($d = 0$) or when no dealers

[20]Pricewatchers' trading strategy (demand function) can be written as

$$x_j^W(p_j, \omega_{-j}) = a_j^W\,(E[v_j \mid \delta_j, p_{-j}] - p_j) = a_j^W(\delta_j - p_j) + b_j^W\omega_{-j},$$

where expressions for coefficients $a_j^W$ and $b_j^W$ are provided in the proof of Proposition 2.

in security $D$ are pricewatchers ($\mu_D = 0$). Indeed, in these cases, the price of security $F$ conveys no information to dealers in security $D$. Thus, the level of illiquidity in security $D$ is as in the benchmark case ($B_D = B_{D0}$) and the level of illiquidity in security $F$ is readily obtained by substituting this expression for $B_D$ in equation (22). Hence, there is a unique rational expectations equilibrium and we can characterize the equilibrium in closed form, which considerably simplifies the analysis. Remember that $\mathcal{R}_F$ is a measure of dealers' risk bearing capacity in security $F$ (see equation (5)). We obtain the following result.

**Corollary 6.** *With one-sided learning ($d = 0$) or no pricewacthers in security $D$ ($\mu_D = 0$), there is a unique linear rational expectations equilibrium where the levels of illiquidity of securities $D$ and $F$ are*

$$B_D = B_{D0}, \tag{25}$$

$$B_F = \frac{B_D^2 \sigma_{u_D}^2 (B_D^2 \sigma_{u_F}^2 \sigma_{u_D}^2 + \mu_F \gamma_F^2)}{\gamma_F(\mu_F^2 \gamma_F^2 (1 + B_D^2 \sigma_{u_D}^2) + B_D^2 \sigma_{u_D}^2 \sigma_{u_F}^2 (\mu_F + B_D^2 \sigma_{u_D}^2))}. \tag{26}$$

*In this equilibrium, liquidity spillovers from security $D$ to security $F$ are positive for all values of $\mu_F$ if $\mathcal{R}_F \leq 1$. In contrast, if $\mathcal{R}_F > 1$, liquidity spillovers from security $D$ to security $F$ are negative when $\mu_F < \hat{\mu}_F$ and positive when $\mu_F \geq \hat{\mu}_F$, where $\hat{\mu}_F$ is strictly smaller than one and defined in the proof of the corollary.*

When $\mu_D = \mu_F = 1$, the corollary describes the equilibrium obtained with full attention and one sided learning. In this case, as explained previously, liquidity spillovers from security $D$ to security $F$ are always positive. In contrast, when $\mu_F$ is small enough and $\mathcal{R}_F > 1$, liquidity spillovers from security $D$ to security $F$ can be negative.

To see why, consider a decrease in the risk tolerance of the dealers operating in security $D$ ($\gamma_D$ decreases). This decrease makes security $D$ less liquid and therefore less informative for pricewatchers in security $F$. Thus, uncertainty about the payoff of security $F$ increases. As with full attention, this "uncertainty effect" increases the illiquidity of security $F$. However, with limited attention, there is a countervailing effect that we call the "adverse selection effect." Indeed, as pricewatchers' private information is less precise, their informational advantage is smaller. As a consequence, inattentive dealers are less exposed to adverse selection. This effect reduces the illiquidity of security $F$. Intuitively the reduction in uncertainty has a small effect on illiquidity when (i) few dealers receive price information ($\mu_F < \hat{\mu}_F$) and (ii) when dealers' risk bearing capacity is high (i.e., $\mathcal{R}_F > 1$) since in this case uncertainty is not a big driver of illiquidity. When these conditions are met, the adverse selection effect prevails over the uncertainty effect. As a result the increase in the illiquidity of security $D$ *reduces* the illiquidity of security $F$. Otherwise, the uncertainty effect dominates and liquidity spillovers from security $D$ to $F$ are positive.

We now consider the more general case in which learning is two-sided ($d > 0$). The next corollary shows that liquidity spillovers in this case are positive if the fraction of pricewatchers in securities $D$ and $F$ is high enough.

**Corollary 7.** *Let*

$$\overline{\mu}_j = \max\left\{0, \frac{\mathcal{R}_j - 1}{\mathcal{R}_j}\right\}, \text{ for } j \in \{D, F\}. \quad (27)$$

*If $\mu_D \in [\overline{\mu}_D, 1]$ and $\mu_F \in [\overline{\mu}_F, 1]$ then liquidity spillovers from security $D$ to security $F$ and vice versa are positive for all values of $d$.*

Thus, the model will feature positive liquidity spillovers if the level of attention is higher than $\overline{\mu}_j$ for $j \in \{D, F\}$. This threshold is always less than one and can be as low as zero if dealers' risk bearing capacity is small enough in both markets, i.e., if $\mathcal{R}_j \leq 1$ for $j \in \{D, F\}$. In contrast, when the fraction of pricewatchers in security $j$ is less than $\overline{\mu}_j$, liquidity spillovers from security $-j$ to security $j$ can be negative for the reasons explained previously.

As an example, suppose that the parameter values are as follows: $\sigma_{u_F} = 0.1$, $\sigma_{u_D} = 1$, $\gamma_F = 1$, $d = 1$, $\mu_D = \mu_F = 0.1$, and $\sigma_\eta = 1$. In this case, $\overline{\mu}_D = 0$ while $\overline{\mu}_F = 0.9$. Thus, liquidity spillovers from security $F$ to security $D$ are positive while liquidity spillovers from security $D$ to security $F$ can be negative since $\mu_F < \overline{\mu}_F$ (Corollary 7). For instance Figure 4 considers the effect of an increase in the risk tolerance of dealers in security $D$. This increase reduces the illiquidity of security $D$ but it *increases* the illiquidity of security $F$ because liquidity spillovers from security $D$ to security $F$ are negative in this case.

**[Insert Figure 4 about here]**

Our model predicts the existence of positive or negative liquidity spillovers between securities. Empirically, these spillovers should translate into positive or negative co-movement in liquidity. We illustrate this point with the following experiment. For a given value of $\mu_F$, we compute the illiquidity of securities $F$ and $D$ assuming that $\gamma_D$ is uniformly distributed in $[0.5, 1]$ and setting $\sigma_{u_F} = \sigma_{u_D} = 1/2$, $\sigma_\eta = 2$, $\gamma_F = 1/2$. For these values of the parameters $\overline{\mu}_j = 0$ and liquidity spillovers are therefore positive. We then compute the covariance between the resulting equilibrium values for $B_D$ and $B_F$. Figure 5, Panel (a) and Panel (b) show this covariance as a function of $\mu_F$ when $d = 0$ and $d = 0.9$, respectively (for $\mu_D = 0.1$ and $\mu_D = 0.9$). In both cases, the covariance between the illiquidity of securities $D$ and $F$ is positive because illiquidity spillovers are positive. In panel (c) we set $\sigma_{u_F} = 0.1$, $d = 0.9$ and $\mu_D = 0.9$ while other parameters are unchanged. In this case liquidity spillovers from security $D$ to security $F$ can be negative when $\mu_F$ is smalle enough. As a result the covariance between the illiquidity of security $D$ and the illiquidity of security $F$ is negative for relative low values of $\mu_F$ and positive otherwise.

**[Insert Figure 5 about here]**

### 3.3.3 Is attention good for market liquidity?

We now study the relationship between the illiquidity of a security and the fraction of pricewatchers in this security. We already know that the illiquidity of security $j$ is always smaller with full attention than with no attention (see Corollary 2). However, as shown below, for small values of the fraction of pricewatchers, the illiquidity of a security may be strictly higher than with no attention. Hence, the relationship between illiquidity and attention is non monotonic. Again it is easier to establish this result when learning is one sided ($d = 0$) or when $\mu_D = 0$ since in these cases the equilibrium is unique and we can characterize it in closed-form. We obtain the following result.

**Corollary 8.** *Consider the cases in which learning is one sided ($d = 0$) or in which there are no pricewacthers in security $D$ ($\mu_D = 0$).*

1. *If $\mathcal{R}_F \leq 1$, an increase in attention by dealers in security $F$ reduces the illiquidity of this security.*
2. *If $\mathcal{R}_F > 1$, an increase in attention by dealers in security $F$ reduces the illiquidity of this security if $\mu_F \geq \mu_F^\bigstar$ and increases its illiquidity if $\mu_F < \mu_F^\bigstar$ where $0 < \mu_F^\bigstar < 1$ (see the appendix for the analytical expression of $\mu_F^\bigstar$).*

The impact of a change in the fraction of pricewatchers in security $F$ on the liquidity of this market is determined by both the adverse selection effect and the uncertainty effect, which play in opposite directions. On the one hand, an increase in the fraction of pricewatchers in security $F$ raises the exposure to adverse selection for inattentive dealers in security $F$. On the other hand, more dealers have relatively low inventory holding costs because more dealers are better informed about the payoff of security $j$. The first effect raises illiquidity while the second effect decreases illiquidity. As shown in Corollary 8, the second effect always prevails when the risk bearing capacity of dealers in security $F$ is less than one. When this condition is not satisfied, the adverse selection effect dominates when the fraction of pricewatchers is small ($\mu_F < \mu_F^\bigstar$). Hence, the relationship between the liquidity of security $F$ and the fraction of pricewatchers is non monotonic: it increases in the fraction of pricewatchers when this fraction is less than $\mu_F^\bigstar$, reaches a maximum when this fraction is equal to $\mu_F^\bigstar$ and then decreases.

When learning is two-sided, i.e., $d > 0$, the analysis of the impact of a change in attention in one market is more complex because liquidity spillovers operate in both directions. Hence, as explained in Section 3.2.3, the total impact of a change in the fraction of pricewatchers in one security on the illiquidity of this security is determined both by the *direct* impact of this change on illiquidity (measured by $(\partial f/\partial \mu_D)$ or $(\partial g/\partial \mu_F)$) and the *indirect* impact which accounts for the spillover effects described in the previous section. This indirect impact can be positive or negative depending on the direction of liquidity spillovers between the two markets. Signing the total impact of an increase in attention in one market on the level of illiquidity in both markets is therefore challenging. However, the next corollary shows that if $\mathcal{R}_j \leq 1$ then more attention leads to a more liquid market for both securities in at least one of the possible rational expectations equilibria of the model. When the equilibrium is unique, it must therefore have this property if $\mathcal{R}_j \leq 1$.

**Corollary 9.** *If $\mathcal{R}_j \leq 1$ for $j \in \{D, F\}$ then, other things equal, an increase in attention by dealers in security $j$ reduces the illiquidity of this security ($(\partial f/\partial \mu_D) < 0$ and $(\partial g/\partial \mu_F) < 0$).*

*Furthermore, there is always an equilibrium in which an increase in attention by dealers in security $j$ reduces the illiquidity of both securities in equilibrium.*

To save space, we provide the proof of this result in the Internet Appendix. We illustrate this corollary with a numerical example. We set $\sigma_\eta = 0.77$, $\sigma_{u_j} = 1$, $\gamma_j = 1$ and $d = 1$, so that learning is two-sided. In Figure 6, we plot the relationship between the illiquidity of security $D$ and the fraction of pricewatchers in this security for $\mu_D \in \{0.001, 0.002, \ldots, 1\}$ when $\mu_F = 0.5$ (panel (a)) and $\mu_F = 0.9$ (panel (b)) when $B_F$ is fixed at its equilibrium value for $\mu_D = 0.001$ (bold curve) and when $B_F$ adjusts to its equilibrium value for each value of $\mu_D$ (dotted curve). Thus, the bold curve represents the direct effect of a change in the fraction of pricewatchers in security $D$ (i.e., the effect holding constant the liquidity of security $F$) while the dotted curve represents the evolution of the equilibrium value of the illiquidity of security $D$, after accounting for spillover effects. The difference between the two curves shows the amount by which spillover effects magnify the direct effect of a change in attention on illiquidity.

**[Insert Figure 6 about here]**

Table 3 provides a summary of our main results when the level of attention in each market is exogenous.

| Panel A – One-sided learning $d = 0$ | | | | |
|---|---|---|---|---|
| Attention | Risk bearing capacity | Spillovers from from $D$ to $F$ | $\uparrow \mu_F$ on $B_D$ | $\uparrow \mu_F$ on $B_F$ |
| No | | No spillovers | No effect | No effect |
| Limited | $\mathcal{R}_F \leq 1$ | + | No Effect | − |
| | $\mathcal{R}_F > 1$ | + iff $\mu_F \geq \widehat{\mu}_F$ | No Effect | − iff $\mu_F \geq \mu_F^\star$ |
| Full | | + | N.A. | N.A. |
| **Panel B – Two-sided learning $d > 0$** | | | | |
| Attention | Risk bearing capacity | Spillovers from from $j$ to $-j$ | $\uparrow \mu_j$ on $B_j$ | $\uparrow \mu_j$ on $B_{-j}$ |
| No | | No spillovers | No effect | No effect |
| Limited | $\mathcal{R}_j \leq 1$ | + | − | − |
| | $\mathcal{R}_j > 1$ | Ambiguous/can be negative | Ambiguous | Ambiguous |
| Full | | + | N.A. | N.A. |

Table 3: Summary of the main findings with exogenous attention.

# 4 Endogenous attention

We now endogenize the level of dealers' attention to the prices of other securities, i.e., $\mu_j$. To this end we introduce a *cost of attention*, $C$ (see the introduction of the paper for interpretations of this cost).[21] We assume that dealers simultaneously make their decision to be a pricewatcher at date 0, before trades take place at date 1. Dealers who become pricewatchers pay the cost $C$. Other dealers do not pay this cost and cannot make their strategy contingent on the price in the other market. Once these decisions have been made, trades take place as described in the previous section.

Dealers' decisions to be a pricewatcher hinges on a comparison between the cost of attention and the value of attention, i.e., the informational value of the price of the other security. Let $\phi_j$ be the value of the information contained in the price of security $-j$ for dealers in security $j$ when a fraction $\mu_j$ of dealers in security $j$ are informed about the price of security $-j$. This value is the maximum fee that a dealer in security $j$ is willing to pay in order to observe the price of security $-j$, $p_{-j}$. This fee solves:

$$E\left[U\left((v_j - p_j)\, x_j^W - \phi_j\right)\right] = E\left[U\left((v_j - p_j)\, x_j^I\right)\right]. \tag{28}$$

In general, the solution to this equation depends on the level of illiquidity in security $-j$ since this level determines the informational content of the price of security $-j$. We stress this feature by explicitly writing $\phi_j$ as a function of the illiquidity of security $-j$: $\phi_j = \phi_j(\mu_j, B_{-j})$. In making their monitoring decisions, dealers take the fraction of pricewatchers as given. Hence, for a given fraction of pricewatchers in each market, a dealer in security $j$ chooses to monitor the price of security $-j$ if $\phi_j(\mu_j, B_{-j}) > C$ and abstains from monitoring this price if $\phi_j(\mu_j, B_{-j}) < C$. When $\phi_j(\mu_j, B_{-j}) = C$, the dealer is indifferent between monitoring the price of security $-j$ or not.

The fraction of pricewatchers in each market results from this cost-benefit analysis and is ultimately determined by the cost of attention. In the rest of this section, we study the effect of varying the cost of attention on the equilibrium fraction of pricewatchers and market illiquidity. This analysis yields two new insights. First, a decrease in the cost of attention can impair market liquidity. Second, when learning is two-sided, the value of attention for dealers in one security can increase both in the level of attention by dealers in the *same* security and dealers in the other security. As a consequence, dealers' attention decisions reinforce each other *and* multiple equilibria with differing levels of attention can arise for the same level of the cost of attention.

## 4.1 Attention decisions with one-sided learning

When $d = 0$, learning is one-sided: dealers in security $D$ learn no information from the price of security $F$. In this case, monitoring the price of security $F$ for dealers in security $D$ is worthless ($\phi_D(\mu_D, B_F) = 0$) and as a result all dealers in security $D$ optimally abstain from paying the cost of attention, i.e., $\mu_D = 0$. Thus, the level of illiquidity in security $D$ is as given in the

[21] In our analysis we take the cost of attention as being exogenous. In reality, part of this cost is determined by pricing decisions of data vendors (Bloomberg, Reuters, exchanges, etc...). An interesting extension of our paper would be to endogenize this cost by studying the optimal pricing policy of sellers of price information in our set-up. Cespa and Foucault (2009) study the optimal pricing policy for a monopolist seller of price information. But they restrict their attention to the case with a single security.

benchmark case, i.e., $B_D = \sigma_\eta^2/\gamma_D$ for all possible values of $\mu_F$. Hence, in this section we write $\phi_F(\mu_F, B_D)$ as $\phi_F(\mu_F)$ to simplify notation.

Using the specification of dealers' utility functions and the fact that all variables have a normal distribution, we obtain that[22]

$$\phi_F(\mu_F) = \frac{\gamma_F}{2} \ln\left(\frac{\text{Var}[v_F|\delta_F, \hat{\omega}_F]}{\text{Var}[v_F|\delta_F, \omega_D]}\right) > 0. \tag{29}$$

As explained in Section 3.3, pricewatchers in security $F$ obtain a signal $\omega_D$ about factor $\delta_D$ from monitoring the price of security $D$. The price information privately observed by pricewatchers leaks partially through the price of security $F$ as pricewatchers trade on this information, which conveys a signal $\hat{\omega}_F$ to inattentive dealers. However, this signal is less informative than the signal obtained by pricewatchers since price movements in security $F$ are also affected by the demand shock in this security. For this reason, pricewatchers can form a more precise forecast of the payoff of security $F$ than inattentive dealers, that is $\text{Var}[v_F|\delta_F, \hat{\omega}_F] > \text{Var}[v_F|\delta_F, \omega_D]$ and the value of being a pricewatcher is always strictly positive. Intuitively, the value of monitoring the price of security $D$ for dealers in security $F$ decreases in the fraction of pricewatchers in security $F$ because the leakage effect is stronger when the fraction of pricewatchers in security $F$ is higher. We establish this result in the next corollary.

**Proposition 3.** *If $d = 0$,*

$$\phi_F(\mu_F) = \frac{\gamma_F}{2} \ln\left(1 + \frac{\sigma_{u_F}^2 \sigma_{u_D}^2 B_D^2}{\gamma_F^2 \mu_F^2 (1 + B_D^2 \sigma_{u_D}^2) + \sigma_{u_F}^2 \sigma_{u_D}^4 B_D^4}\right). \tag{30}$$

*with $B_D = \sigma_\eta^2/\gamma_D$. Thus, the value of monitoring the price of security $D$ for dealers in security $F$ decreases in the fraction of pricewatchers in security $F$.*

Hence, with one sided learning, the value of acquiring price information declines with the fraction of dealers buying this information, as usual in models of information acquisition (e.g., Grossman and Stiglitz (1980) or Admati and Pfleiderer (1986)). Let $\mu_F^*(C)$ be the fraction of dealers in security $F$ who decide to pay the cost of attention. As $\phi_F(\mu_F)$ decreases in $\mu_F$, there are three possible cases:

1. If $\phi_F(1) > C$, then the value of monitoring the price of security $D$ for dealers in security $F$ exceeds the cost of monitoring even when all dealers pay the cost of monitoring. Thus, $\mu_F^*(C) = 1$.

2. If $\phi_F(0) < C$, then the value of monitoring the price of security $D$ for dealers in security $F$ is always lower than the cost of monitoring. Thus, $\mu_F^*(C) = 0$.

3. Otherwise, the equilibrium fraction of pricewatchers is such that dealers in security $F$ are just indifferent between monitoring the price of security $D$ or not. That is, $\mu_F^*(C)$ is the unique solution of $\phi_F(\mu_F) = C$.

[22]Our expression for the value of information is standard in models of information acquisition with normally distributed variables and CARA utility functions (see for instance Admati and Pfleiderer (1986)). Thus, for brevity we omit the derivation of this result, which can be obtained upon request.

We obtain the following result.

**Proposition 4.** *With one sided learning* ($d = 0$)*, the fraction* $\mu_F^*(C)$ *of dealers in security* $F$ *who monitor the price of security* $D$ *in equilibrium decreases in the cost of attention. This fraction is:*

1. $\mu_F^*(C) = 0$, *if* $C > \overline{C}$.

2. $\mu_F^*(C) = \sqrt{\frac{\sigma_{u_F}^2 \sigma_{u_D}^2 B_D^2 (1 - B_D^2 \sigma_{u_D}^2 (e^{2C/\gamma_F} - 1))}{\gamma_F^2 (1 + B_D^2 \sigma_{u_D}^2)(e^{2C/\gamma_F} - 1)}}$, *if* $\underline{C} \leq C \leq \overline{C}$.

3. $\mu_F^*(C) = 1$, *if* $C < \underline{C}$,

*where closed-form solutions for the thresholds* $\underline{C}$ *and* $\overline{C}$ *are given in the proof of the proposition and* $B_D = \sigma_\eta^2/\gamma_D$.

The illiquidity of security $F$ is in part determined by the fraction of pricewatchers in this security (see Section 3.3). As this fraction is itself determined by the cost of attention, the illiquidity of security $F$ is ultimately determined by the cost of attention. The next corollary describes the effect of a change in the cost of attention on the illiquidity of security $F$.

**Corollary 10.** *With one sided learning* ($d = 0$)*:*

1. *If* $\mathcal{R}_F \leq 1$ *then the illiquidity of security* $F$ *increases in the cost of attention for dealers active in this security.*

2. *If* $\mathcal{R}_F > 1$ *, there exists a value of* $C^* \in (\underline{C}, \overline{C})$ *such that the illiquidity of security* $F$ *increases in the cost of attention for dealers active in this security when* $C \leq C^*$ *and decreases in the cost of attention otherwise (the closed-form solution for* $C^*$ *is given in the proof of the corollary).*

A decrease in the cost of attention leads to an increase in the fraction of pricewatchers in security $F$ when learning is one-sided. As explained in Section 3.3, this evolution has an ambiguous effect on the illiquidity of security $F$. On the one hand, more attention reduces the uncertainty on the payoff of security $F$. On the other hand, inattentive dealers are more exposed to adverse selection if the attention of their competitors increases. As shown in Corollary 6, the uncertainty effect always dominates when $\mathcal{R}_F \leq 1$. Thus, in this case, a reduction in the cost of monitoring for dealers in security $F$ always improves the liquidity of this security. When $\mathcal{R}_F > 1$, the adverse selection effect dominates as long as the fraction of pricewatchers remains small, i.e., when $C$ is greater than $C^*$. Indeed, for this range of values for the cost of attention, only a few dealers choose to be pricewatchers. As a result, a small decline in the cost of attention reinforces the adverse selection risk for inattentive dealers and market liquidity deteriorates. Figure 7 illustrates the impact that a change in the cost of attention has on the fraction of pricewatchers, illiquidity, and the value of information with one-sided learning.

**[Insert Figure 7 about here]**

## 4.2 Attention decisions with two sided learning

We now consider the case in which $d > 0$, so that dealers in each security can learn information from the price of the other security. In this case, our main finding is that the value of price information for dealers in a given market can be increasing in the fraction of pricewatchers in both markets. This finding is counterintuitive since usually the value of financial information declines with the fraction of investors acquiring this information (see Grossman and Stiglitz (1980) or Admati and Pfleiderer (1986)). The value of price information has this property when learning is one-sided, as we have just shown in Proposition 4. In contrast, when learning is two-sided, price information is *special*: its value can increase in the number of investors who buy this information. As we shall see the main reason for this counter-intuitive result is that the value of price information tends to be higher for securities that are more liquid and securities tend to be more liquid when the fraction of pricewatchers is large enough.

Using again the dealers' utility functions specification and the fact that all variables are normally distributed, we obtain that the value of monitoring the price of security $-j$ for dealers in security $j$ is

$$\phi_j(\mu_j, B_{-j}(\mu_j, \mu_{-j})) = \frac{\gamma_j}{2} \ln\left(\frac{\text{Var}[v_j|\delta_j, \hat{\omega}_j]}{\text{Var}[v_j|\delta_j, \omega_{-j}]}\right), \tag{31}$$

where we stress the fact that the illiquidity of each market in equilibrium is a function of the fraction of pricewatchers in either market. To save space we provide the explicit expression for $\phi_j(\mu_j, B_{-j}(\mu_j, \mu_{-j}))$ in the Internet Appendix. For a fixed fraction of pricewatchers in market $-j$, we have

$$\frac{d\phi_j}{d\mu_j} = \underbrace{L_j}_{\text{Leakage effect}} + \underbrace{\Lambda_j}_{\text{Feedback effect}}. \tag{32}$$

with $L_j \equiv (\partial\phi_j/\partial\mu_j)$ and $\Lambda_j \equiv (\partial\phi_j/\partial B_{-j})(\partial B_{-j}/\partial\mu_j)$. Thus, the total effect of an increase in the fraction of pricewatchers in security $j$ on the value of being a pricewatcher is the sum of two effects: the *leakage effect* (that we described in the previous section) and the *feedback effect*, which is new as it arises only when learning is two-sided. To understand this feedback effect, consider an increase in the fraction of pricewatchers in security $D$ (the reasoning is symmetric for an increase in $\mu_F$). When $d > 0$, this increase affects the liquidity of security $D$ and thereby the liquidity of security $F$. In turn, the change in the liquidity of security $F$ feeds back on the value of monitoring this security since, as explained before, it affects the informativeness of the price of security $F$ for dealers in security $D$ if $d > 0$. The change in the value of information due to this feedback effect is measured by $\Lambda_D$. It is zero when learning is one-sided because in this case dealers in security $D$ learn no information from the price of security $F$ (hence $\partial\phi_D/\partial B_F = 0$).[23]

The total effect of an increase in the fraction of pricewatchers in security $j$ on the value of information in this market is positive if and only if the feedback effect outweighs the leakage effect

$$\Lambda_j > -L_j > 0. \tag{33}$$

[23]Moreover, $\partial B_D/\partial\mu_F = 0$ when $d = 0$ since the illiquidity of security $D$ is independent of $\mu_F$ in this case ($B_D = (\sigma_\eta^2/\gamma_D)$). Thus, $\Lambda_F = 0$ as well when learning is one-sided.

If the feedback effect dominates (i.e., condition (33) holds true), the value of being a pricewatcher in security $j$ increases in the fraction of pricewatchers in this security. Obviously, a necessary condition for this to happen is that the feedback effect is positive, which is a possibility when $\mathcal{R}_j \leq 1$. To see this, consider again the value of monitoring security $F$ for dealers in security $D$. When $\mathcal{R}_D \leq 1$, as shown in Corollary 9, an increase in the fraction of pricewatchers in security $D$ reduces the illiquidity of security $F$ ($\partial B_F/\partial \mu_D < 0$). As a consequence, the price of security $F$ becomes more informative for dealers in security $D$ and the value of monitoring this price is higher ($\partial \phi_D/\partial B_F < 0$), at least for some parameter values. In this case, the feedback effect for security $D$ is positive: $\Lambda_D > 0$.

We have not been able to delineate the set of parameters under which the feedback effect dominates the leakage effect. However, numerical simulations show that this set is not empty. To see this, consider Figure 8. Panel (a) on this figure plots the value of monitoring security $F$ for pricewatchers in security $D$ (i.e., $\phi_D(\mu_D, B_F)$) for two values of $\mu_F$. In both cases the value of observing the price of security $F$ increases with the fraction of pricewatchers in security $D$, which means that the feedback effect dominates the leakage effect.

**[Insert Figure 8 about here]**

Now consider the effect of a change in the fraction of pricewatchers located in market $-j$ on the value of monitoring this market for dealers in asset $j$. This cross-market monitoring effect is measured by

$$\frac{d\phi_j}{d\mu_{-j}} = \left( \frac{\partial \phi_j}{\partial B_{-j}} \frac{\partial B_{-j}}{\partial \mu_{-j}} \right). \tag{34}$$

As shown in Corollary 9, an increase in the fraction of pricewatchers in, say, security $D$ reduces the illiquidity of this security ($(\partial B_D/\partial \mu_D) < 0$) if $\mathcal{R}_D \leq 1$. In turn this effect makes the price of security $D$ more informative for dealers in security $F$ and increases the value of monitoring this price for dealers in security $F$ ($(\partial \phi_F/\partial B_D) < 0$). In this case, $(d\phi_F/d\mu_D) > 0$. That is, an increase in the fraction of pricewatchers in security $D$ makes the value of monitoring the price of security $D$ higher for dealers in security $F$.

Figure 8 illustrates the cross-market monitoring effect as well. First, consider panel (a) again. It shows that the value of monitoring the price of security $F$ for dealers in security $D$ is higher, all else being equal when $\mu_F = 0.9$ than when $\mu_F = 0.1$. Moreover, panel (b) shows that an increase in the fraction of pricewatchers in security $D$ makes the value of monitoring security $D$ higher for dealers in security $F$.

Thus, price information is special because the decision of each dealer to buy this information can reinforce each other both in the *same* market and across *different* markets.[24] The model shows that this happens in two distinct ways: (i) the value of being informed about the price of another security can increase in the fraction of dealers who follow this security ("within market

[24] The leakage effect implies that dealers' decisions to buy price information are "strategic substitutes": the acquisition of price information by one dealer reduces the value of being a pricewatcher for other dealers. In contrast, when positive, the feedback effect works to make dealers' decisions to buy price information "strategic complements": the acquisition of price information by one dealer strengthens the value of being a pricewatcher for other dealers operating in the same market.

complementarity") and (ii) the value of being informed about the price of another security can increase in the fraction of pricewatchers in this security ("cross market complementarity"). Both types of complementarity in dealers' monitoring decisions are absent when $d = 0$ and they do not necessarily both operate when $d > 0$ (in particular the leakage effect may prevail over the feedback effect even though the cross-market complementarity operates).

Now consider whether a dealer in market $j$ should become a pricewatcher. In making this decision, the dealer takes the fraction of pricewatchers in both markets as given. If $\phi_j(\mu_j, B_{-j}(\mu_D, \mu_F)) > C$, it is optimal for the dealer to be a pricewatcher since the value of monitoring the price in the other market is higher than the cost. If $\phi_j(\mu_j, B_{-j}(\mu_D, \mu_F)) < C$, it is optimal for the dealer not to monitor the price in the other market and finally, if $\phi_j(\mu_j, B_{-j}(\mu_D, \mu_F)) = C$, the dealer is just indifferent. Given these observations, the equilibrium fractions of pricewatchers in each market, $(\mu_D^*, \mu_F^*)$, are displayed in Table 4.

| $\mu_j^*, \mu_{-j}^*$ | When |
|---|---|
| $\mu_j^* = \mu_{-j}^* = 1$ | $\phi_j(1, B_{-j}(1,1)) > C$ for $j \in \{D, F\}$ |
| $\mu_j^* = 1$, $\mu_{-j}^* \in (0,1)$ | $\phi_j(1, B_{-j}(1, \mu_{-j}^*)) > C$ and $\phi_{-j}(\mu_{-j}^*, B_{-j}(1, \mu_{-j}^*)) = C$ |
| $\mu_j^*, \mu_{-j}^* \in (0,1)$ | $\phi_j(\mu_j^*, B_{-j}(\mu_j^*, \mu_{-j}^*)) = C$ for $j \in \{D, F\}$ |
| $\mu_j^* = 0, \mu_{-j}^* \in (0,1)$ | $\phi_j(0, B_{-j}(0, \mu_{-j}^*)) < C$ and $\phi_{-j}(\mu_{-j}^*, B_{-j}(1, \mu_{-j}^*)) = C$ |
| $\mu_j^*, \mu_{-j}^* = 0$ | $\phi_j(0, B_{-j}(0,0)) < C$ for $j \in \{D, F\}$. |

Table 4: The equilibrium fraction of pricewatchers in markets $j$ and $-j$.

Complementarities in attention decisions among dealers located in different markets lead to multiple equilibria for the levels of attention. Indeed, these complementarities imply that the value of cross-market monitoring will be relatively high when the fraction of pricewatchers in both markets is high and relatively low when the fraction of pricewatchers in both markets is low. Thus, for intermediate values of the cost of monitoring, there is room for multiple equilibria with various levels of market integration for the same values of the parameters (in particular the correlation of the payoffs of the two securities being fixed).

It is worth stressing that the multiplicity of possible attention levels in equilibrium is a phenomenon distinct from the multiplicity of rational expectations equilibria. Indeed, one may have a single linear rational expectations equilibrium for each possible level of attention in each security and yet multiple equilibrium levels of attention. As an example, consider the parameter values of Figure 8 again and suppose $C = 0.06$. For the parameter values in Figure 8, there is a unique non-fully rational expectations equilibrium for each value of $\mu_D$ and $\mu_F$. However, there are three possible pairs of equilibrium values for the levels of attention in each market: (i) $\mu_D^* = \mu_F^* = 1$, (ii) $\mu_D^* = 0$, $\mu_F^* = 1$ and (iii) $\mu_D^* \simeq 0.3$, $\mu_F^* = 1$. In all these equilibria, all dealers in security $F$ pay attention to the price of security $D$. In contrast, *for the same parameter values*, we can have an equilibrium in which dealers in security $D$ do not follow security $F$

($\mu_D^* = 0$), an equilibrium in which all dealers in security $D$ follow security $F$ ($\mu_D^* = 1$) or an equilibrium in which only a fraction of dealers in security $D$ buy price information on security $F$ ($\mu_D^* \simeq 0.3$). Thus, for the same fundamentals, dealers in security $D$ can appear to neglect the information contained in the price of security $F$ or to be very sensitive to this information.

We may also have situations in which, for the *same parameter values*, the markets for the two securities appear fully segmented because dealers in either market pay no attention to the other market ($\mu_D^* = \mu_F^* = 0$) or fully integrated because all dealers are pricewatchers ($\mu_D^* = \mu_F^* = 1$). To see this, consider the case in which the two markets are perfectly symmetric: $\gamma_F = \gamma_D = \gamma$, $d = 1$, $\sigma_\eta = 0$ and $\sigma_{u_F} = \sigma_{u_D} = \sigma_u$. In this case, we obtain (see the Internet Appendix for a derivation):

$$\phi_j(\mu_j, B_{-j}) = \frac{\gamma}{2} \ln\left(1 + \frac{B_{-j}^2 \sigma_u^4}{\gamma^2 \mu_j^2 (1 + B_{-j}^2 \sigma_u^2) + B_{-j}^4 \sigma_u^6}\right). \tag{35}$$

In this symmetric case, there are two non-fully revealing rational expectations equilibria if $\mu_D = \mu_F = 1$ (see Section 3.2). For the discussion, we focus on the high illiquidity equilibrium in which the level of illiquidity in markets $D$ and $F$ is $B^{H*}$ (given in equation (10)). In the symmetric case, parameters are identical in the two markets. Hence, by symmetry, we have $\phi_F(1, B^{H*}) = \phi_D(1, B^{H*})$ and $\phi_F(0, B_{F0}) = \phi_D(0, B_{D0})$. That is, the value of price information is identical in each market in the full attention case and in the no attention case, respectively. Let $\phi_0$ be the value of price information in the no attention case and let $\phi_1$ be the value of price information in the full attention case. Using equation (35), we obtain the following result.

**Proposition 5.** *In the symmetric case (i.e., $\gamma_F = \gamma_D = \gamma$, $d = 1$, $\sigma_\eta = 0$ and $\sigma_{u_F} = \sigma_{u_D} = \sigma_u$):*

1. *The value of monitoring prices in market $-j$ for dealers in market $j$ is strictly higher when $\mu_D = \mu_F = 1$ than when $\mu_D = \mu_F = 0$, that is, $\phi_1 > \phi_0$ for $j \in \{H, L\}$.*

2. *Moreover if the cost of attention is such that $\phi_0 < C < \phi_1$, then the cases in which all dealers are pricewatchers ($\mu_D^* = \mu_F^* = 1$) and no dealers are pricewatchers ($\mu_D^* = \mu_F^* = 0$) are possible equilibria.*

The first part of the proposition shows that in the symmetric case the value of monitoring is always higher when all dealers are pricewatchers than when no dealers are pricewatchers. For this reason, for the *same* parameters value, the markets for securities $F$ and $D$ can be either fully integrated (all dealers in each market account for the price information available in the other market) or fully segmented, as claimed in the second part of the proposition. As an illustration, suppose that $\sigma_\delta = \sigma_u = 1$, $\gamma = 1/2$. In this case, we have

$$\phi_0 = \frac{\gamma}{2} \ln\left(1 + \frac{\gamma^2}{\sigma_u^2}\right) \approx 0.055, \;\; \phi_1 = \frac{\gamma}{2} \ln\left(1 + \frac{(B^{H*})^2 \sigma_u^4}{\gamma^2 (1 + (B^{H*})^2 \sigma_u^2) + (B^{H*})^4 \sigma_u^6}\right) \approx 0.127.$$

Thus, for any value of $C \in [0.055, 0.127]$, the markets for securities $F$ and $D$ can be either fully segmented or fully integrated, depending on whether dealers in both markets coordinate on the high or the low attention equilibrium. The liquidity of both markets and the informativeness of prices are higher if dealers coordinate on the high attention equilibrium. Interestingly, in this case, the markets can remain segmented even if the cost of attention decreases, unless it falls below $C = 0.055$.

In summary, when learning is two sided, the value of price information can increase in the fraction of pricewatchers. This property means that dealers' decisions to monitor the price of another security are complements both within and across markets. That is, they reinforce each other. As a consequence, multiple equilibria with differing levels of attention are sustainable and two securities may appear segmented even though the correlation of their payoffs is high and the cost of monitoring is relatively low.

This result has interesting implications. First, it implies that fads, traditions, or other coordination devices may determine the degree of integration between two securities, independently of the correlation in the payoffs of these securities. Second, a decrease in the cost of attention (due for instance to better information linkages between markets) does not necessarily entail greater market integration, unless the cost is very low. Third, dealers operating in related but opaque segments may undervalue the benefit of greater market integration. Indeed, in the low attention equilibrium, the value of getting price information is low. Thus, data vendors will perceive a weak demand and will therefore lack incentives to collect and disseminate price information. In this case, regulatory intervention is needed. A case in point is the U.S. corporate bond market where real time dissemination of bond prices took off only under regulatory pressure (see Bessembinder et al. (2006)).

# 5 Testable implications

One way to test whether cross-security learning is a source of liquidity spillovers is to consider changes in trading technologies that affect dealers' ability to learn from the prices of other securities. According to our model, these changes should affect the extent of liquidity spillovers across securities and the levels of liquidity for these securities. In contrast, theories of liquidity co-movements based on market wide changes in dealers' risk bearing capacity (e.g., Brunnermeier and Pedersen (2009)) make no predictions on such changes in trading technologies. In the rest of the paper, we illustrate this approach with two thought experiments.

## 5.1 From opaque to transparent markets

Suppose that the markets for securities $D$ and $F$ are opaque so that the cost of obtaining information on the prices of securities $D$ and $F$ is high. In this case, the fraction of pricewatchers in both securities is low. Let us denote the fraction of pricewatchers in this environment by $\mu_j^b$ for $j \in \{D, F\}$. Now suppose that the market for security $D$ becomes transparent while the market for security $F$ remains opaque. After this switch, the fraction of pricewatchers in security $D$ remains unchanged whereas the fraction of pricewatchers in security $F$ is higher since transparency reduces the cost of acquiring information on the price of security $D$. That is $\mu_D^b = \mu_D^a$ but $\mu_F^a > \mu_F^b$ where $\mu_j^a$ is the fraction of pricewatchers in security $j$ *after* the switch to a new trading system for security $F$. To simplify the discussion, let us assume that

$\mu_D^b = \mu_D^a = 0$. In this case the model has a unique rational expectations equilibrium for all values of $\mu_F$ and we can use Corollaries 6 and 8 to develop predictions about the effects of this change in market design.

In this case, if dealers' risk bearing capacity in security $F$ is relatively low ($\mathcal{R}_F \leq 1$), the liquidity of security $F$ should increase after the market for security $D$ becomes transparent (see Corollary 8), even though the market structure for security $F$ is identical before and after the change affecting the other security. Moreover, co-variation in liquidity between securities $D$ and $F$ should be positive and greater than before the change in market design as explained in Section 3.3 (Corollary 6).

If instead, dealers' risk bearing capacity in security $F$ is relatively large ($\mathcal{R}_F > 1$) and the fraction of pricewatchers in security $F$ remains small ($\mu_F^a < \mu_F^\bigstar$) then the liquidity of security $F$ should decrease after the market for security $D$ becomes transparent (Corollary 8). The reason is that the transparency of security $D$ exposes inattentive dealers active in security $F$ to adverse selection by giving an informational advantage to pricewatchers (see Section 3.3.3). Moreover, in this case, liquidity spillovers from security $D$ to security $F$ will be negative (Corollary 6).

The implementation of the TRACE system in the U.S. corporate bond market is a field experiment close to the thought experiment we just described. Until 2002, the U.S. corporate bond market was very opaque: the price of each transaction was known only to the parties involved in the transaction. This situation changed when the SEC required dissemination of transaction prices for a subset of bonds through a reporting system called TRACE. This requirement initially applied to 498 bonds and was implemented in July 2002. Bessembinder et al. (2006) study the effects of this reform of the bond market on the liquidity of TRACE eligible bonds (security $D$ in our thought experiment) and non-TRACE-eligible bonds (security $F$).[25] Interestingly, Bessembinder et al. (2006) find a significant increase in liquidity for non-TRACE eligible bonds, as predicted by our model (see Table 3, page 272 in Bessembinder et al.(2006)).

The model makes the additional prediction, which to our knowledge has yet to be tested, that the liquidity of non-TRACE bonds should become more sensitive to changes in the liquidity of TRACE bonds after the implementation of TRACE. This prediction can be tested by analyzing the lead-lag relationships between measures of liquidity for TRACE-eligible bonds and non-TRACE bonds. The model implies that a shock to the liquidity of TRACE bonds should have a greater effect in absolute value on the liquidity of non-TRACE bonds after the implementation of TRACE and that the direction of this effect might be negative if few dealers in non-TRACE bonds watch the prices of TRACE bonds.

Bessembinder et al.(2010) also finds that the liquidity of the TRACE eligible bonds increases. This finding is consistent with the model as well. To see this suppose now that both the markets for securities $D$ and $F$ become transparent. If $\mathcal{R}_j \leq 1$ for both securities or $\mu_j$ is high enough then the liquidity of both securities is higher in the transparent system, for all values of the fraction of pricewatchers (see Corollary 9).

[25] Edward et al. (2005) and Goldstein et al. (2007) also consider the effects of greater transparency in the U.S. bond markets. However, they do not analyze the effects of greater transparency on non-eligible bonds.

## 5.2 Co-location fees

The recent years have witnessed a growth of so called "high frequency market-makers" (e.g., GETCO, Optiver, etc...), who use highly automated strategies. These market-makers often use price information available about one security to take positions in other securities. For instance, they monitor quote updates in stock index futures and use this information to set their quotes in other securities.

The case in which $d = 0$ can be used to analyze this type of trading strategy. Indeed, in this case we can interpret security $D$ as providing information on a market wide risk factor ($\delta_D$) and security $F$ as a security that loads on this factor and another factor ($\delta_F$). We interpret pricewatchers in security $F$ as high frequency market-makers: they watch in real-time the price of security $D$ and use this information to determine their position in security $F$.

As explained in the introduction, high frequency market-makers obtain price information faster than other market participants by co-locating their computers close to trading platforms' matching engines, at a cost equal to the co-location fee charged by the platform.[26] Thus, the co-location fee is one component of the cost of price information.

Now suppose that the co-location fee declines. In this case, Proposition 4 implies that the number of high frequency market-makers should increase since the cost of price information declines. If the risk bearing capacity of high-frequency market-makers is low ($\mathcal{R}_F \leq 1$), entry of new pricewatchers should improve the liquidity of security $F$. Moreover, liquidity spillovers from security $D$ to security $F$ should be positive and stronger after the reduction in the co-location fee (see Corollary 10 and Figure 5).

If instead the risk bearing capacity of high-frequency market-makers is high ($\mathcal{R}_F > 1$), the scenario is more complex. If $C > C^*$, entry of new high frequency market-makers increases the exposure to adverse selection for other dealers in security $F$. Thus, the liquidity of security $F$ should drop after the reduction in the co-location fee (see Corollary 10). Moreover, liquidity spillovers from security $D$ to security $F$ can be negative in this case. Indeed, an improvement in liquidity for security $D$ allows pricewatchers in security $F$ to obtain more precise information. Thus, if the fraction of pricewatchers remains small, the risk of adverse selection for inattentive dealers increases and the liquidity of security $F$ drops following an increase in liquidity for security $D$.

Jovanovic and Menkveld (2010) study entry of a high frequency market-maker in Dutch stocks traded on Chi-X (a European trading platform). They show empirically that following this entry, quotes in Chi-X become relatively more informative on price innovations in the Dutch index futures.[27] Moreover, the liquidity of the stocks in which the high frequency market-maker is active improves. This is consistent with the model when $\mathcal{R}_F \leq 1$. In this case the model makes the additional prediction that the liquidity of Dutch shares should become more sensitive

[26] This fee can be significant. For instance, the monthly fee for this service for stocks listed on NYSE Amex is as high as \$61,000 per month in 2011. See NYSE Amex equities price list 2011 at `http://www.nyse.com/pdfs/amex-equities-prices.pdf`.

[27] Hendershott and Riordan (2010) also find empirically that high frequency traders make the market more informationally efficient.

to changes in the liquidity of the Dutch index futures after entry of the high-frequency market-maker.

# 6 Conclusions

In this paper we analyze a new mechanism that explains the transmission of liquidity shocks in one market to another market ("liquidity spillovers"). Central to this mechanism is the fact that dealers in one security often use the price of other securities as a source of information to set their quotes. The price of a security conveys a noisier signal about fundamentals when the market for this security is less liquid. As a result, a drop in the liquidity of one security propagates to other securities because it increases the level of uncertainty for dealers in all other securities. This propagation of the initial liquidity shock makes all prices less informative, which amplifies the initial drop in liquidity. For this reason, even small initial shocks on market liquidity in one asset class can ultimately result in large market wide changes in liquidity.[28]

The model provides several additional insights:

1. Liquidity spillovers are not necessarily positive. The direction of these spillovers depends on the fraction of dealers with price information on other securities. When this fraction is large, liquidity spillovers are positive. In contrast, liquidity spillovers can be negative when price information is only available to a relatively small number of dealers and dealers' risk bearing capacity is large.

2. A decrease in the cost of price information can increase market illiquidity if it triggers too small an increase in the fraction of dealers who acquire information on the price of other securities.

3. The value of price information can increase, for some parameter values, with the fraction of dealers buying this information. For this reason, for the same parameter values, multiple levels of segmentation (high, medium or low) between securities can be sustained in equilibrium.

Future work should study the implications of our model for asset pricing. The model implies that the extent of liquidity co-movements between assets is in part determined by the cost of acquiring price information. Hence, liquidity risk and therefore risk premia should be sensitive to changes in trading technologies that affect this cost, as explained in the last part of our paper. Moreover the model implies that the liquidity of some securities could covary negatively

[28]In line with this transmission mechanism, the CFTC-SEC report on the Flash crash emphasizes the role that uncertainty on the cause (transient price pressures or changes in fundamentals) of the large price movements in the E-mini futures on the S&P500 played in the evaporation of liquidity during the Flash crash. The authors of this report write (on page 39): "*market makers that track the prices of securities that are underlying components of an ETF are more likely to pause their trading if there are price-driven or data-driven integrity questions about those prices. Moreover extreme volatility in component stocks makes it very difficult to accurately value an ETF in real-time. When this happens, market participants who would otherwise provide liquidity for such ETFs may widen their quotes or stop providing liquidity [...].*" This is consistent with our model in which the liquidity of a security drops when prices of other securities become less reliable as a source of information.

with the liquidity of other securities. These securities should therefore provide a good hedge against market wide variations in liquidity and offer negative risk premia for this risk. Do such securities exist in reality? Do they have the characteristics that our model predicts (relatively few well informed dealers with high risk bearing capacity)? We leave these questions for future research.

# A Appendix

**Proof of Lemma 1**

If $\mu_D = 0$ then all dealers in security $D$ only observe factor $\delta_D$ when they choose their demand function. As dealers have a CARA utility function, it is immediate that their demand function in this case is

$$x_D^I(\delta_D) = \gamma_D \frac{E[v_D|\delta_D] - p_D}{\mathrm{Var}[v_D|\delta_D]} = \gamma_D \frac{\delta_D - p_D}{\sigma_\eta^2 + d^2}. \tag{A.1}$$

Using the clearing condition, we deduce that the clearing price is such that:

$$p_D = \delta_D + \left(\frac{\sigma_\eta^2 + d^2}{\gamma_D}\right) u_D = \delta_D + B_{D0} u_D.$$

A similar reasoning yields the expression of the clearing price for security $F$. □

**Proof of Proposition 1**

This proposition is a special case of Proposition 2, which considers the more general case in which $\mu_j$ can take any value. This proposition is proved below. □

**Proof of Lemma 2**

In the symmetric case, we can proceed as in the proof of Proposition 2 to show that a non-fully revealing linear rational expectations equilibrium exists if and only if the system of equations (8) and (9) has at least one strictly positive solution. Solving this sytem shows that this is the case if and only if $\sigma_u^2 \geq 4\gamma^2$ and that in this case the system of equations (8) and (9) has two solutions: $B_D^* = B_F^* = B^{H*}$ and $B_D^* = B_F^* = B^{L*}$. Otherwise, the unique solution of this system is $B_{D1}^* = 0$ and $B_{F1} = 0$. Hence, there is no non-fully revealing linear rational expectations equilibria when $\sigma_u^2 < 4\gamma^2$. □

**Proof of Lemma 3**

See Step 1 in the proof of Proposition 2. □

**Proof of Corollary 1**

From Step 3 in the proof of Proposition 2, we deduce that when $\mu_D = \mu_F = 1$, there is a unique non-fully revealing equilibrium if and only if $\Psi_1'(B_{D1}) < 0$, $\forall B_{D1}$. Using the expression for $\Psi_1(\cdot)$ (equation (A.32)), we obtain

$$\Psi_1'(B_{D1}) = -\gamma_D \gamma_F^2 (1 + B_{D1}^2 \sigma_{u_D}^2)^2 + \\ 4B_{D1}\sigma_{u_D}^2(\sigma_\eta^2 - \gamma_D B_{D1}) \left(\gamma_F^2(1 + B_{D1}^2\sigma_{u_D}^2) + B_{D1}^2 \sigma_{u_D}^2 \sigma_{u_F}^2\right) + B_{D1}^3 \sigma_{u_D}^4 \sigma_{u_F}^2 \gamma_D (4\gamma_D^{-1} d^2 - B_{D1}).$$

Remember that when $\mu_D = \mu_F = 1$, $B_{D1} > \sigma_\eta^2/\gamma_D$ (see Step 3 in the proof of Proposition 2). Hence, if $4d^2/\gamma_D \leq \sigma_\eta^2/\gamma_D$ then $\Psi_1'(B_{D1}) < 0$.

□

**Proof of Corollary 2**

The result follows immediately from equation (17) □

**Proof of Corollary 3**

The result follows immediately from equations (15) and (16). □

**Proof of Corollary 4**

The result follows immediately from the definition of functions $f_1(\cdot)$ and $g_1(\cdot)$ in Proposition 1. □

**Proof of Proposition 2**

**Step 1.** We show below (Step 2) that if $p_j^* = R_j\delta_j + B_ju_j + A_j\delta_{-j} + C_ju_{-j}$ is a rational expectations equilibrium then $R_j = 1$ and $C_j = A_jB_{-j}$. Hence, in a rational expectations equilibrium, the price in market $j$ can be written $p_j^* = \omega_j + A_j\omega_{-j}$, where $\omega_j = \delta_j + B_ju_j$. Thus, $\{\delta_j,\, \omega_{-j}\}$ is a sufficient statistic for $\{\delta_j, p_{-j}, p_j\}$. Clearly, the equilibrium is non-fully revealing if and only if $B_j > 0$. Moreover, $\{\delta_j, \hat{\omega}_j\}$ is a sufficient statistic for $\{\delta_j, p_j\}$, where $\hat{\omega}_j = B_ju_j + A_j\omega_{-j}$ and since $\omega_{-j} = p^*_{-j} - A_{-j}\omega_j$, we can also write the equilibrium price in market $j$ as

$$p_j^* = \omega_j + A_j(p^*_{-j} - A_{-j}\omega_j) = (1 - A_jA_{-j})\omega_j + A_jp^*_{-j}.$$

These observations prove Lemma 3.

**Step 2. Equilibrium in market $j$.**
**Pricewatchers' demand function.** A pricewatcher's demand function in market $j$, $x_j^W(\delta_j, p_j, p_{-j})$, maximizes

$$E\left[-\exp\left\{-\left((v_j - p_j)x_j^W\right)/\gamma_j\right\}|\delta_j, p_j, p_{-j}\right].$$

We deduce that

$$\begin{aligned} x_j^W(\delta_j, p_j, p_{-j}) &= \gamma_j\left(\frac{E[v_j|\delta_j, p_{-j}, p_j] - p_j}{\text{Var}[v_j|\delta_j, p_{-j}]}\right) \\ &= a_j^W(E[v_j|\delta_j, p_{-j}, p_j] - p_j), \end{aligned} \tag{A.2}$$

with $a_j^W = \gamma_j\text{Var}[v_j|\delta_j, p_{-j}]^{-1}$.

As $\{\delta_D,\, \omega_F\}$ is a sufficient statistic for $\{\delta_D, p_F, p_D\}$, we deduce (using well-known properties of normal random variables) that

$$\begin{aligned} E[v_D|\delta_D, p_F, p_D] &= E[v_D|\delta_D, \omega_F] \\ &= \delta_D + \frac{d}{(1 + B_F^2\sigma_{u_F}^2)}\omega_F, \end{aligned} \tag{A.3}$$

and

$$a_D^W = \frac{\gamma_D}{\text{Var}[v_D \mid \delta_D, \omega_F]} \tag{A.4}$$

$$= \gamma_D\left(\frac{1 + B_F^2\sigma_{u_F}^2}{d^2B_F^2\sigma_{u_F}^2 + \sigma_\eta^2(1 + B_F^2\sigma_{u_F}^2)}\right) \tag{A.5}$$

$$= \frac{\gamma_D}{\text{Var}[v_D|\delta_D](1 - \rho_D^2)}, \tag{A.6}$$

where $\rho_D^2 \equiv d^2/((\sigma_\eta^2 + d^2)(1 + B_F^2\sigma_{u_F}^2))$. Thus,

$$x_D^W(\delta_D, \omega_F) = a_D^W(\delta_D - p_D) + b_D^W \omega_F,$$

where

$$\begin{aligned} b_D^W &= \frac{\gamma_D}{\mathrm{Var}[v_D|\delta_D, \omega_F]} \frac{\mathrm{Cov}[v_D, \omega_F]}{\mathrm{Var}[\omega_F]} \\ &= d a_D^W \left( \frac{1}{1 + B_F^2 \sigma_{u_F}^2} \right). \end{aligned} \tag{A.7}$$

Similarly, for pricewatchers in security $F$ we obtain

$$x_F^W(\delta_F, \omega_D) = a_F^W(\delta_F - p_F) + b_F^W \omega_D, \tag{A.8}$$

where $\omega_D = \delta_D + B_D u_D$, and

$$a_F^W = \gamma_F \left( \frac{1 + B_D^2 \sigma_{u_D}^2}{B_D^2 \sigma_{u_D}^2} \right) = \frac{\gamma_F}{\mathrm{Var}[v_F|\delta_F](1 - \rho_F^2)}, \quad b_F^W = a_F^W \frac{1}{1 + B_D^2 \sigma_{u_D}^2}, \tag{A.9}$$

where $\rho_F^2 \equiv (1 + B_D^2 \sigma_{u_D}^2)^{-1}$.

**Inattentive Dealers.** An inattentive dealers' demand function in market $j$, $x_j^I(\delta_j, p_j)$, maximizes:

$$E\left[ -\exp\left\{ -\left( (v_j - p_j)\, x_j^I \right)/\gamma_j \right\} |\delta_j, p_j \right].$$

We deduce that

$$\begin{aligned} x_j^I(\delta_j, p_j) &= \gamma_j \left( \frac{E[v_j|\delta_j, p_j] - p_j}{\mathrm{Var}[v_j|\delta_j, p_{-j}]} \right) \\ &= a_j^I(E[v_j|\delta_j, p_j] - p_j), \end{aligned} \tag{A.10}$$

with $a_j^I = \gamma_j \mathrm{Var}[v_j|\delta_j, p_{-j}]^{-1}$.

As $\{\delta_D, \hat{\omega}_D\}$ is a sufficient statistic for $\{\delta_D, p_D\}$, we deduce (using well-known properties of normal random variables) that

$$\begin{aligned} E[v_D|\delta_D, p_D] &= E[v_D|\delta_D, \hat{\omega}_D] \\ &= \delta_D + \frac{d A_D}{A_D^2(1 + B_F^2 \sigma_{u_F}^2) + B_D^2 \sigma_{u_D}^2} \hat{\omega}_D, \end{aligned} \tag{A.11}$$

and

$$a_D^I = \frac{\gamma_D}{\mathrm{Var}[v_D \mid \delta_D, \hat{\omega}_D]} \tag{A.12}$$

$$= \gamma_D \frac{A_D^2(1 + B_F^2 \sigma_{u_F}^2) + B_D^2 \sigma_{u_D}^2}{d^2(A_D^2 B_F^2 \sigma_{u_F}^2 + B_D^2 \sigma_{u_D}^2) + \sigma_\eta^2(A_D^2(1 + B_F^2 \sigma_{u_F}^2) + B_D^2 \sigma_{u_D}^2)}. \tag{A.13}$$

Thus,

$$x_D^I(\delta_D, \hat{\omega}_D) = a_D^I(\delta_D - p_D) + b_D^I \hat{\omega}_D,$$

where

$$b_D^I = \frac{\gamma_D}{\text{Var}[v_D|\delta_D, \hat{\omega}_D]} \frac{\text{Cov}[v_D, \hat{\omega}_D]}{\text{Var}[\hat{\omega}_D]}$$
$$= a_D^I \frac{dA_D}{A_D^2(1 + B_F^2\sigma_{u_F}^2) + B_D^2\sigma_{u_D}^2}. \tag{A.14}$$

Similarly, for market $F$ we obtain:

$$x_F^I(\delta_F, \hat{\omega}_F) = a_F^I(\delta_F - p_F) + b_F^I\hat{\omega}_F, \tag{A.15}$$

where

$$a_F^I = \gamma_F \frac{A_F^2(1 + B_D^2\sigma_{u_D}^2) + B_F^2\sigma_{u_F}^2}{A_F^2 B_D^2\sigma_{u_D}^2 + B_F^2\sigma_{u_F}^2}, \quad b_F^I = a_F^I \frac{A_F}{A_F^2(1 + B_D^2\sigma_{u_D}^2) + B_F^2\sigma_{u_F}^2}. \tag{A.16}$$

**Clearing price in market $j$.** The clearing condition in market $j \in \{D, F\}$ imposes

$$\mu_j x_j^W(\delta_j, p_j, p_{-j}) + (1 - \mu_j)x_j^I(\delta_j, p_j) + u_j = 0.$$

Let $a_j = \mu_j a_j^W + (1 - \mu_j)a_j^I$. Using equations (A.2) and (A.10), we solve for the clearing price and we obtain

$$p_j^* = \delta_j + \left(\frac{\mu b_j^W + (1 - \mu_j)b_j^I A_j}{a_j}\right)\omega_{-j} + \left(\frac{(1 - \mu_j)b_j^I B_j + 1}{a_j}\right)u_j, \tag{A.17}$$

Remember that we are searching for an equilibrium such that $p_j^* = R_j\delta_j + B_j u_j + A_j\delta_{-j} + C_j u_{-j}$. We deduce from equation (A.17) that in equilibrium, we must have $R_j = 1$,

$$B_j = \left(\frac{(1 - \mu_j)b_j^I B_j + 1}{a_j}\right), \; A_j = \left(\frac{\mu b_j^W + (1 - \mu_j)b_j^I A_j}{a_j}\right), \text{ and } C_j = A_j B_{-j}.$$

Thus

$$B_j = \frac{1}{a_j - (1 - \mu_j)b_j^I}, \quad \text{for } \; j \in \{D, F\}, \tag{A.18}$$
$$A_j = \mu_j B_j b_j^W, \quad \text{for } \; j \in \{D, F\}. \tag{A.19}$$

Coefficients $A_j$ and $C_j$ ultimately depend on the coefficients $\{B_j, B_{-j}\}$. Hence, the equilibrium is fully characterized once coefficients $B_j$ and $B_{-j}$ are known as claimed in the proposition. Substituting (A.6) in (A.7) and rearranging we obtain

$$b_D^W = d\gamma_D \frac{1}{d^2 B_F^2\sigma_{u_F}^2 + \sigma_\eta^2(1 + B_F^2\sigma_{u_F}^2)}. \tag{A.20}$$

Using (A.19) (for $j = D$) and (A.20), we can rewrite (A.14) as

$$b_D^I = a_D^I \frac{d^2\mu_D\gamma_D(d^2 B_F^2\sigma_{u_F}^2 + \sigma_\eta^2(1 + B_F^2\sigma_{u_F}^2))}{B_D(\mu_D^2 d^2\gamma_D^2(1 + B_F^2\sigma_{u_F}^2) + \sigma_{u_D}^2(d^2 B_F^2\sigma_{u_F}^2 + \sigma_\eta^2(1 + B_F^2\sigma_{u_F}^2))^2)}. \tag{A.21}$$

Similarly, using (A.19) (for $j = D$) and (A.20), we can rewrite (A.13) as

$$a_D^I = \frac{\gamma_D\left(\mu_D^2 d^2\gamma_D^2(1+B_F^2\sigma_{u_F}^2)+\sigma_{u_D}^2(d^2B_F^2\sigma_{u_F}^2+\sigma_\eta^2(1+B_F^2\sigma_{u_F}^2))^2\right)}{(d^2B_F^2\sigma_{u_F}^2+\sigma_\eta^2(1+B_F^2\sigma_{u_F}^2))(\mu_D^2d^2\gamma_D^2+\sigma_{u_D}^2(\sigma_\eta^2+d^2)(\sigma_\eta^2(1+B_F^2\sigma_{u_F}^2)+d^2B_F^2\sigma_{u_F}^2))} \tag{A.22}$$

$$= \frac{\gamma_D\left(\mu_D^2\gamma_D^2\rho_D^2+\sigma_{u_D}^2(d^2+\sigma_\eta^2)(1-\rho_D^2)^2\right)}{(d^2+\sigma_\eta^2)(1-\rho_D^2)(\mu_D^2\gamma_D^2\rho_D^2+\sigma_{u_D}^2(d^2+\sigma_\eta^2)(1-\rho_D^2))}. \tag{A.23}$$

Inserting (A.23) in (A.21) yields after some algebra

$$b_D^I = \gamma_D^2\frac{d^2\mu_D}{B_D(\mu_D^2d^2\gamma_D^2+\sigma_{u_D}^2(\sigma_\eta^2+d^2)(\sigma_\eta^2(1+B_F^2\sigma_{u_F}^2)+d^2B_F^2\sigma_{u_F}^2))}. \tag{A.24}$$

We can now replace (A.6), (A.23) and (A.24) in (A.18) and, after some tedious algebra, we obtain

$$B_D = f(B_F;\mu_D,\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2), \tag{A.25}$$

where

$$f(B_F;\mu_D,\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2) = \frac{B_{D0}(1-\rho_D^2)\left(\mu_D\gamma_D^2\rho_D^2+(\sigma_\eta^2+d^2)\sigma_{u_D}^2(1-\rho_D^2)\right)}{\rho_D^2\mu_D^2\gamma_D^2+\sigma_{u_D}^2(\sigma_\eta^2+d^2)(1-\rho_D^2)(1-\rho_D^2(1-\mu_D))}, \tag{A.26}$$

with $\rho_D^2 = d^2/((\sigma_\eta^2+d^2)(1+B_F^2\sigma_{u_F}^2))$ and $B_{D0} = (\sigma_\eta^2+d^2)/\gamma_D$. In a similar way we obtain

$$B_F = g(B_D;\mu_F,\gamma_F,\sigma_{u_D}^2), \tag{A.27}$$

where

$$g(B_{D1};\mu_F,\gamma_F,\sigma_{u_D}^2) = \frac{B_{F0}(1-\rho_F^2)\left(\mu_F\gamma_F^2\rho_F^2+\sigma_{u_F}^2(1-\rho_F^2)\right)}{\rho_F^2\mu_F^2\gamma_F^2+\sigma_{u_F}^2(1-\rho_F^2)(1-\rho_F^2(1-\mu_F))}, \tag{A.28}$$

with $\rho_F^2 = (1+B_D^2\sigma_{u_D}^2)^{-1}$ and $B_{F0} = \gamma_F^{-1}$. Last, as $\mathrm{Var}[v_D|\delta_D] = \sigma_\eta^2 + d^2$ and $\mathrm{Var}[v_F|\delta_F] = 1$, we obtain that

$$B_j = B_{j0}(1-\rho_j^2)\times\frac{\gamma_j^2\mu_j\rho_j^2+\sigma_{u_j}^2\mathrm{Var}[v_j|\delta_j](1-\rho_j^2)}{\gamma_j^2\mu_j^2\rho_j^2+\sigma_{u_j}^2\mathrm{Var}[v_j|\delta_j](1-\rho_j^2)(1-\rho_j^2(1-\mu_j))}, \tag{A.29}$$

as claimed in the proposition.

**Step 3. Existence of a non-fully revealing equilibrium with full attention ($\mu_j = 1$).** Let $f_1(B_{F1};\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2) \equiv f(B_F;1,\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2)$ and $g_1(B_{D1};\gamma_F,\sigma_{u_D}^2) \equiv g(B_{D1};1,\gamma_F,\sigma_{u_D}^2)$. When $\mu_D = \mu_F = 1$, we deduce from equations (A.25) and (A.27) that a non-fully rational expectations equilibrium exists if and only if the following system of equations has a strictly positive solution

$$B_{D1} = f_1(B_{F1};\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2) = \frac{\sigma_\eta^2}{\gamma_D}+\frac{d^2B_{F1}^2\sigma_{u_F}^2}{\gamma_D(1+B_{F1}^2\sigma_{u_F}^2)}, \tag{A.30}$$

$$B_{F1} = g_1(B_{D1};\gamma_F,\sigma_{u_D}^2) = \frac{B_{D1}^2\sigma_{u_D}^2}{\gamma_F(1+B_{D1}^2\sigma_{u_D}^2)}. \tag{A.31}$$

Note that $B_{F1} > 0$ if and only if $B_{D1} > 0$. Let

$$\Psi_1(B_{D1}) \equiv f_1(g_1(B_{D1};\gamma_F,\sigma_{u_D}^2);\gamma_D,\sigma_\eta^2,d,\sigma_{u_F}^2) - B_{D1}.$$

Substituting the expression for $B_{F1}$ in equation (A.30), we deduce that the equilibrium levels for the illiquidity of security $B_{D1}$ solve $\Psi_1(B_{D1}) = 0$. Thus, a non-fully revealing equilibrium exists if and only if $\Psi_1(B_{D1}) = 0$ has at least one strictly positive root. Using the expression for $g_1(B_{D1}; \gamma_F, \sigma^2_{u_D})$, we obtain

$$\Psi_1(B_{D1}) = \left(\sigma^2_\eta - \gamma_D B_{D1}\right)\left(\gamma^2_F(1 + B^2_{D1}\sigma^2_{u_D})^2 + B^4_{D1}\sigma^4_{u_D}\sigma^2_{u_F}\right) + d^2 B^4_{D1}\sigma^4_{u_D}\sigma^2_{u_F}, \tag{A.32}$$

which is a polynomial of degree 5 in $B_{D1}$. Observe that $\Psi_1(\cdot)$ is continuous and

$$\Psi_1\left(\frac{\sigma^2_\eta}{\gamma_D}\right) \geq 0, \quad \Psi_1\left(\frac{\sigma^2_\eta + d^2}{\gamma_D}\right) < 0.$$

Thus, (A.32) has at least one solution $B^*_{D1}$ in the interval $[\sigma^2_\eta/\gamma_D, (\sigma^2_\eta + d^2)/\gamma_D]$. As $\sigma^2_\eta > 0$, this proves existence of a non-fully revealing equilibrium when $\mu_D = \mu_F = 1$.

**Step 4. Existence of a non fully revealing equilibrium with limited attention ($\mu_j <$ 1).**

With limited attention, we deduce from equations (A.25) and (A.27) that a non-fully revealing equilibrium exists if and only if the following equation has one strictly positive solution

$$\Psi(B_D) \equiv f(g(B_D; \mu_F, \gamma_F, \sigma^2_{u_D}); \mu_D, \gamma_D, \sigma^2_\eta, d, \sigma^2_{u_F}) - B_D = 0.$$

Calculations show that $\Psi(\cdot)$ is an odd-degree polynomial in $B_D$ with negative leading coefficient. Hence,

$$\lim_{B_D \to \infty} \Psi(B_D) = -\infty,$$

while, for $\sigma^2_\eta > 0$,

$$\Psi(0) = \gamma^{12}_F \mu^8_F \sigma^2_\eta (d^2 \gamma^2_D \mu_D + \sigma^2_\eta \sigma^2_{u_D}(d^2 + \sigma^2_\eta)) > 0.$$

Thus, there always exists a strictly positive value $B^*_D$, such that $\Psi(B^*_D) = 0$ when $\sigma^2_\eta > 0$. □

**Proof of Corollary 5**

**Step 1:** The total effect of a change in $\gamma_D$ on the illiquidity of security $D$ is given by

$$\frac{dB_{D1}}{d\gamma_D} = \frac{\partial f_1}{\partial \gamma_D} + \frac{\partial f_1}{\partial B_F}\frac{dB_F}{d\gamma_D}.$$

As

$$\frac{dB_{F1}}{d\gamma_D} = \frac{\partial g_1}{\partial B_D}\frac{dB_{D1}}{d\gamma_D},$$

and $(\partial g_1/\partial B_{D1})(\partial f_1/\partial B_{F1}) > 0$ (since $d > 0$), we deduce that:

$$\frac{dB_{D1}}{d\gamma_D} = \kappa \frac{\partial f_1}{\partial \gamma_D},$$
$$\frac{dB_{F1}}{d\gamma_D} = \kappa \left(\frac{\partial g_1}{\partial B_{D1}}\frac{\partial f_1}{\partial \gamma_D}\right),$$

with $\kappa = 1 - ((\partial g_1/\partial B_{D1})(\partial f_1/\partial B_{F1}))$.

**Step 2:** Now we prove that there alwauys exists at least one non fully revealing rational expectations equilibrium for which $\kappa > 1$. Let

$$h_1(B_{D1}) \equiv f_1(g_1(B_{D1}; \gamma_F, \sigma^2_{u_D}); \gamma_D, \sigma^2_\eta, d, \sigma^2_{u_F}).$$

Note that

$$\frac{\partial h_1}{\partial B_{D1}} = \frac{\partial f_1}{\partial B_{F1}} \frac{\partial g_1}{\partial B_{D1}}.$$

Hence, if $h_1'(B_{D1}) < 1$ at an equilibrium value for $B_{D1}$ then there exists at least one equilibrium in which $\kappa > 1$. Remember that the equilibrium values for $B_{D1}$ solve (see Step 3 in the proof of Proposition 2)

$$\Psi_1(B_{D1}) \equiv h_1(B_{D1}) - B_{D1} = 0.$$

Hence, the roots of the polynomial $\Psi_1(B_{D1})$ are the possible equilibrium values for the illiquidity of security $D$. Using equation (A.32), we obtain

$$\begin{aligned}\Psi_1(B_{D1}) = & -B^5_{D1}\gamma_D\sigma^4_{u_D}(\gamma^2_F + \sigma^2_{u_F}) + B^4_{D1}\sigma^4_{u_D}(\gamma^2_F\sigma^2_\eta + (d^2 + \sigma^2_\eta)\sigma^2_{u_F}) \\ & - 2B^3_{D1}\gamma_D\gamma^2_F\sigma^2_{u_D} + 2B_{D1}\gamma^2_F\sigma^2_\eta\sigma^2_{u_D} - B_{D1}\gamma_D\gamma^2_F + \gamma^2_F\sigma^2_\eta.\end{aligned}$$

Using Descartes' rule of signs, we obtain that $\Psi_1(\cdot)$ has five, three or one positive root. These roots correspond to the intersections of the function $h_1(B_{D1})$ with the 45-degree line. As $h_1(0) = \sigma^2_\eta/\gamma_D > 0$ and,

$$h_1'(B_{D1}) = \frac{4B^3_D d^2\gamma^2_F\sigma^4_{u_D}\sigma^2_{u_F}(1 + B^2_D\sigma^2_{u_D})}{\gamma_D(\gamma^2_F(1 + B^2_D\sigma^2_{u_D})^2 + B^4_D\sigma^4_{u_D}\sigma^2_{u_F})^2} > 0,$$

the function $h_1(B_{D1})$ cuts for the first time the 45-degree line from above. Hence, at this intersection point, we must have $h_1'(B_{D1}) < 1$. Let $B^{L*}_{D1}$ be this intersection point. When the equilibrium is unique, the equilibrium level of illiquidity must be $B^{L*}_{D1}$ as otherwise $h_1(\cdot)$ would never cut the 45-degree line and therefore an equilibrium would not exist. When there are multiple equilibria, $B^{L*}_{D1}$ is the lowest level of illiquidity for security $D$ among all non-fully revealing equilibria since this is the lowest positive root of $\Psi_1(B_{D1})$. Thus, there always exists an equilibrium in which $h_1'(B_{D1}) < 1$ at the equilibrium value for $B_{D1}$.

□

**Proof of Corollary 6**

**Step 1:** For the expressions for the illiquidity levels in securities $D$ and $F$, see the paragraph that precedes the corollary.

**Step 2:** For the second part, we differentiate $B_F$ with respect to $B_D$ and we obtain that

$$\frac{\partial B_F}{\partial B_D} = \frac{2B_{D\mu_F\sigma^2_{u_D}}(\gamma^4_F\mu^2_F + B^2_D\gamma^2_F\sigma^2_{u_D}\sigma^2_{u_F}(2\mu_F - B^2_D(1-\mu_F)\sigma^2_{u_D}) + B^4_D\sigma^4_{u_D}\sigma^4_{u_F})}{\gamma_F((\gamma_F\mu_F)^2(1 + B^2_D\sigma^2_{u_D}) + B^2_D\sigma^2_{u_D}\sigma^2_{u_F}(\mu_F + B^2_D\sigma^2_{u_D}))^2}. \quad \text{(A.33)}$$

The numerator of this expression contains a quadratic polynomial in $\mu_F$ with two real roots. Let $\mathcal{P}(\mu_F)$ be this polynomial. One root of $\mathcal{P}(\mu_F)$ is always negative. The other root is

$$\hat{\mu}_F = \frac{B_D^2\sigma_{u_D}^2\sigma_{u_F}\left(-(2+B_D^2\sigma_{u_D}^2)\sigma_{u_F} + \sqrt{4\gamma_F^2 + B_D^2\sigma_{u_D}^2\sigma_{u_F}^2(4+B_D^2\sigma_{u_D}^2)}\right)}{2\gamma_F^2}.$$

As the leading coefficient on $\mathcal{P}(\mu_F)$ (i.e., the coefficient on $\mu_F^4$) is positive, we deduce that $(\partial B_F/\partial B_D)$ is positive if and only if $\mu_F > \hat{\mu}_F$. Direct calculations show that $\hat{\mu}_F \leq 0$, if $\mathcal{R}_F \leq 1$. Thus, in this case, $(\partial B_F/\partial B_D)$ is positive for all values of $\mu_F$. Otherwise $\hat{\mu}_F > 0$ and $(\partial B_F/\partial B_D) < 0$ if and only if $\mu_F < \hat{\mu}_F$. This implies that $\hat{\mu}_F < 1$, as otherwise liquidity spillovers would be negative even when $\mu_F = 1$ (which we know is impossible from Corollary 4). □

**Proof of Corollary 7**

First observe that a change in $B_{-j}$ only affects the illiquidity of security $j$ through its effect on $\rho_j^2$. As $\rho_j^2$ decline in $B_{-j}$, we deduce that liquidity spillovers from security $j$ to security $-j$ are positive if and only if $(\partial B_j/\partial\rho_j^2) < 0$. Now we show that $\mu_j \geq \overline{\mu}_j$ is a sufficient condition for this to be the case. Observe that $B_j = B_{j0}(1-\rho_j^2)G(\mu_j,\rho_j^2)$ with

$$G(\mu_j,\rho_j^2) \equiv \frac{\gamma_j^2\mu_j\rho_j^2 + \sigma_{u_j}^2\text{Var}[v_j|\delta_j](1-\rho_j^2)}{\gamma_j^2\mu_j^2\rho_j^2 + \sigma_{u_j}^2\text{Var}[v_j|\delta_j](1-\rho_j^2)(1-\rho_j^2(1-\mu_j))}. \tag{A.34}$$

Therefore, we have:

$$\frac{\partial B_j}{\partial\rho_j^2} = -B_{0j}G(\mu_j,\rho_j^2) + B_{0j}(1-\rho_j^2)\frac{\partial G}{\partial\rho_j^2}, \tag{A.35}$$

Now observe that:

$$\frac{\partial G(\mu_D,\rho_D^2)}{\partial\rho_D^2} = \frac{(\sigma_\eta^2+d^2)(1-\mu_D)(1-\rho_D^2)\sigma_{u_D}^2(\gamma_D^2\mu_D(1+\rho_D^2) + (\sigma_\eta^2+d^2)(1-\rho_D^2)\sigma_{u_D}^2)}{(\gamma_D^2\mu_D^2\rho_D^2 + \sigma_{u_D}^2\text{Var}[v_D|\delta_D](1-\rho_D^2)(1-\rho_D^2(1-\mu_D)))^2} > 0.$$

Inserting this expression and the expression for $G(\mu_D,\rho_D^2)$ in equation (A.35), we obtain after some algebra

$$\frac{\partial B_D}{\partial\rho_D^2} = -\frac{\text{Var}[v_D|\delta_D]\mu_D}{\gamma_D(\gamma_D^2\mu_D^2\rho_D^2 + \sigma_{u_D}^2\text{Var}[v_D|\delta_D](1-\rho_D^2)(1-\rho_D^2(1-\mu_D)))^2}\times$$
$$(\gamma^4\mu_D^2\rho_D^4 + \sigma_{u_D}^2\text{Var}[v_D|\delta_D](1-\rho_D^2)(\text{Var}[v_D|\delta_D](1-\rho_D^2)\sigma_{u_D}^2 - \gamma_D^2(1-\mu_D-\rho_D^2(1+\mu_D)))).$$

As $\rho_D^2 < 1$, we deduce that the sign of $(\partial B_D/\partial\rho_D^2)$ is the opposite of the sign of

$$\mu_D - \left(\frac{\mathcal{R}_D - 1}{\mathcal{R}_D}\right)\left(\frac{1-\rho_D^2}{1+\rho_D^2}\right),$$

which is positive if $\mu_D \geq \overline{\mu}_D$. We deduce that $(\partial f/\partial B_F) > 0$ if $\mu_D > \overline{\mu}_D$. A similar reasoning shows that $(\partial g/\partial B_D) > 0$ if $\mu_F > \overline{\mu}_F$. □

**Proof of Corollary 8**

Using the expression for $B_F$ in the one sided case (see equation (25)), we obtain

$$\frac{\partial B_F}{\partial \mu_F} = -\frac{B_D^2\sigma_{u_D}^2(\gamma_F^4\mu_F^2 - B_D^4\sigma_{u_D}^4\sigma_{u_F}^2(\gamma_F^2(1-2\mu_F) - \sigma_{u_F}^2) + B_D^2\gamma_F^2\mu_F\sigma_{u_D}^2(\gamma_F^2\mu_F + 2\sigma_{u_F}^2))}{\gamma_F((\gamma_F\mu_F)^2(1+B_D^2\sigma_{u_D}^2) + B_D^2\sigma_{u_D}^2\sigma_{u_F}^2(\mu_F + B_D^2\sigma_{u_D}^2))^2}. \tag{A.36}$$

The sign of this derivative is the same as the sign of its numerator, which is a quadratic polynomial in $\mu_F$ with a positive leading coefficient. Hence, its sign is positive for all values of $\mu_F$ that are larger, in absolute value, than the two real roots of this polynomial. Upon inspection, the first of these roots is always negative, whereas the other root is

$$\mu_F^\bigstar = \frac{-B_D^2\sigma_{u_D}^2\sigma_{u_F}\left(\sigma_{u_F}(1+B_D^2\sigma_{u_D}^2) - ((1+B_D^2\sigma_{u_D}^2)(\gamma_F^2 + B_D^2\sigma_{u_D}^2\sigma_{u_F}^2))^{1/2}\right)}{\gamma_F^2(1+B_D^2\sigma_{u_D}^2)}.$$

We observe that $\mu_F^\bigstar \leq 0$ if and only if $\mathcal{R}_F \leq 1$. Thus, in this case, $(\partial B_F/\partial \mu_F) < 0$, as claimed in Part 1 of the corollary. When $\mathcal{R}_F > 1$, we have $\mu_F^\bigstar > 0$ and $(\partial B_F/\partial \mu_F) > 0$ if and only if $\mu_F < \mu_F^\bigstar$, as claimed in the second part of the corollary. Last we observe that $\mu_F^\bigstar < 1$ as otherwise the illiquidity of security $F$ would be smaller with full attention than with no attention, which is never true (see Corollary 4). □

**Proof of Proposition 3**

Using the notations introduced in the proof of Proposition 2, we have

$$\mathrm{Var}[v_F|\delta_F, \hat{\omega}_F] = \gamma_F(a_F^I)^{-1},$$
$$\mathrm{Var}[v_F|\delta_F, \omega_F] = \gamma_F(a_F^W)^{-1},$$

where

$$a_F^W = \gamma_F\left(\frac{1+B_D^2\sigma_{u_D}^2}{B_D^2\sigma_{u_D}^2}\right), \quad a_F^I = \gamma_F\left(\frac{\mu_F^2\gamma_F^2(1+B_D^2\sigma_{u_D}^2) + B_D^4\sigma_{u_D}^4\sigma_{u_F}^2}{B_D^2\sigma_{u_D}^2(\mu_F^2\gamma_F^2 + B_D^2\sigma_{u_D}^2\sigma_{u_F}^2)}\right).$$

We deduce that

$$\phi_F(\mu_F, B_D) = \frac{\gamma_F}{2}\ln\left(\frac{a_F^W}{a_F^I}\right),$$

and the expression for $\phi_F(\mu_F, B_D)$ given in the corollary follows. It is then immediate that $\partial\phi_F(\mu_F)/\partial\mu_F < 0$. □

**Proof of Proposition 4**

As explained in the text, the fraction of pricewatchers in equilibrium is zero iff $\phi_F(0) < C$. Using equation (30), we deduce that this condition is satisfied iff $C > \overline{C}$ where

$$\overline{C} = \frac{\gamma_F}{2}\ln\left(1 + \frac{1}{\sigma_{u_D}^2 B_D^2}\right).$$

Similarly, the fraction of pricewatchers in equilibrium is one iff $\phi_F(1) > C$. Using equation (30), we deduce that this condition is satisfied iff $C < \underline{C}$ where:

$$\underline{C} = \frac{\gamma_F}{2}\ln\left(1 + \frac{\sigma_{u_F}^2\sigma_{u_D}^2B_D^2}{\gamma_F^2(1+B_D^2\sigma_{u_D}^2) + \sigma_{u_F}^2\sigma_{u_D}^4B_D^4}\right).$$

Otherwise the fraction of pricewatchers in equilibrium solves $\phi_F(\mu_F, B_D) = C$ and we obtain the expression for $\mu_F^*(C)$ by inverting $\phi_F(\mu_F)$ given in equation (30). □

**Proof of Corollary 10**

For a given value of $C$, the level of illiquidity of security $F$ is given by $B_F(\mu_F^*(C))$ where $B_F(\cdot)$ is given in equation (26) when $d = 0$. Thus:

$$\frac{\partial B_F}{\partial C} = \left.\frac{\partial B_F}{\partial \mu_F}\right|_{\mu_F=\mu_F^*(C)} \left(\frac{\partial \mu_F^*(C)}{\partial C}\right).$$

We know that $(\partial\mu_F^*(C)/\partial C) \leq 0$ (Proposition 4). Moreover, using equation (26), we deduce that when $d = 0$, $(\partial B_F/\partial \mu_F) < 0$ if and only if $\mu_F > \widehat{\mu}_F$ where

$$\widehat{\mu}_F = \left(\frac{\sigma_\eta^4 \sigma_{u_D}^2 \sigma_{u_F}}{\gamma_F}\right)\sqrt{\frac{\max\{\gamma_F^2 - \sigma_{u_F}^2 \mathrm{Var}[v_F|\delta_F], 0\}}{\gamma_D^2 + \sigma_\eta^4\sigma_{u_D}^2}}.$$

Thus, when $\gamma_F^2 \leq \sigma_{u_F}^2 \mathrm{Var}[v_F|\delta_F]$, $\widehat{\mu}_F = 0$ and $(\partial B_F/\partial\mu_F)|_{\mu_F=\mu_F^*(C)} < 0$. It follows that $(\partial B_F/\partial C) > 0$. When $\gamma_F^2 > \sigma_{u_F}^2 \mathrm{Var}[v_F|\delta_F]$ then $\widehat{\mu}_F > 0$. As $\mu_F^*(C)$ decreases with $C$ from one to zero over $[\underline{C}, \overline{C}]$, there exists a value $C^* \in (\underline{C}, \overline{C})$ such $\mu_F^*(C) = \widehat{\mu}_F$ and $\mu_F^*(C) < \widehat{\mu}_F$ iff $C > C^*$. Thus, in this case, $(\partial B_F/\partial\mu_F) < 0$ iff $C < C^*$. The second part of the corollary follows. □

**Proof of Proposition 5**

We have

$$\phi_j(1, B^{H*}) = \frac{\gamma}{2}\ln\left(1 + \frac{(B^{H*})^2\sigma_u^4}{\gamma^2(1 + (B^{H*})^2\sigma_u^2) + (B^{H*})^4\sigma_u^6}\right), \tag{A.37}$$

and

$$\phi_j(0, B_{j0}) = \frac{\gamma}{2}\ln\left(1 + \frac{\sigma_\delta^2}{B_{j0}^2\sigma_u^2}\right) = \frac{\gamma}{2}\ln\left(1 + \frac{\gamma^2}{\sigma_u^2}\right)$$

Thus,

$$\phi_j(1, B^{H*}) > \phi_j(0, B_{j0}) \Leftrightarrow \frac{(B^{H*})^2\sigma_u^4}{\gamma^2(1 + (B^{H*})^2\sigma_u^2) + (B^{H*})^4\sigma_u^6} > \frac{\gamma^2}{\sigma_u^2}. \tag{A.38}$$

We deduce that $\phi_j(1, B^{H*}) > \phi_j(0, B^*(0))$ if and only if

$$-\gamma^2\sigma_u^6(B^{H*})^4 + (\sigma_u^4 - \gamma^4)\sigma_u^2(B^{H*})^2 - \gamma^4 > 0. \tag{A.39}$$

Using the expression for $B^{H*}$ given in equation (10), we obtain that

$$(B^{H*})^2 = \frac{(B^{H*}\sigma_u^2 - \gamma)}{\gamma\sigma_u^2}. \tag{A.40}$$

Thus, we can rewrite condition (A.39) as

$$-\gamma\sigma_u^2(B^{H*}\sigma_u^2 - \gamma)^2 + (\sigma_u^4 - \gamma^4)(B^{H*}\sigma_u^2 - \gamma) - \gamma^5 > 0.$$

It can be checked that this inequality holds true if $B^{H*}$ belongs to

$$\left(\frac{\gamma}{\sigma_u^2} + \frac{\sigma_u^4 - \gamma^4 - ((\sigma_u^4 - \gamma^4)^2 - 4\gamma^6\sigma_u^2)^{1/2}}{2\gamma\sigma_u^4}, \frac{\gamma}{\sigma_u^2} + \frac{\sigma_u^4 - \gamma^4 + ((\sigma_u^4 - \gamma^4)^2 - 4\gamma^6\sigma_u^2)^{1/2}}{2\gamma\sigma_u^4}\right).$$

Straighforward calculations show that this is the case when $\sigma_u^2 > 4\gamma^2$, which is required for the existence of a symmetric equilibrium.

**Part 2:** Suppose that $\mu_D^* = \mu_F^* = 1$. Then in this case, the value of monitoring market $j$ for a dealer in security $-j$, given the actions of other dealers, is $\phi_1$. As this value is higher than $C$, monitoring is optimal. Hence $\mu_D^* = \mu_F^* = 1$ is an equilibrium. Now suppose that $\mu_D^* = \mu_F^* = 0$. Then in this case, the value of monitoring market $j$ for a market-maker in market $-j$, given the actions of other dealers, is $\phi_0$. As this value is lower than $C$, not monitoring is optimal. Hence $\mu_D^* = \mu_F^* = 0$ is an equilibrium. □

# References


[1] Acharya V.V. and T. Johnson (2007). "Insider trading in credit derivatives," *Journal of Financial Economics*, 84, 110–141.

[2] Acharya V.V. and Pedersen L.H. (2005). "Asset pricing with liquidity risk," *Journal of Financial Economics*, 77, 375-410.

[3] Admati, A. R. (1985). "A noisy rational expectations equilibrium for multiple asset securities markets," *Econometrica*, 53, 629–657.

[4] Admati, A. R. and P. Pfleiderer (1986). "A monopolistic market for information," *Journal of Economic Theory*, 39, 400–438.

[5] Amihud, Y., Mendelson, H. and L. H. Pedersen (2005). "Liquidity and asset prices," *Foundations and Trends in Finance*, 1, 269–364.

[6] Allen, F. and Gale, D. (2004). "Financial fragility, liquidity, and asset prices," *Journal of the European Economic Association*, 6, 1015–1048.

[7] Andrade, S., Chang, C., and M. Seasholes (2008). "Trading imbalances, predictable reversals, and cross-stock price pressure," *Journal of Financial Economics*, 88, 406–423.

[8] Barlevy, G., and P. Veronesi, (2000). "Information acquisition in financial markets," *Review of Economic Studies*, 67, 79–90.

[9] Bernhardt, D. and D. Taub (2008). "Cross-asset speculation in stock markets," *Journal of Finance*, 63, 2385–2427.

[10] Bessembinder, H., Maxwell, W., and K. Venkataraman (2006). "Market transparency, liquidity externalities, and institutional trading costs in corporate bonds," *Journal of Financial Economics*, 82, 251–288.

[11] Black, F. (1995)."Equilibrium Exchanges," *Financial Analysts Journal*, 51, 23–29.

[12] Boulatov, A., Hatch, B., S. Johnson, and A. Lei (2009). "Dealer attention, the speed of quote adjustments to information, and net dealer revenue," *Journal of Banking and Finance*, 33, 1531–1542.

[13] Boulatov, A., Hendershott, T., and Livdan, D. (2010). "Informed trading and portfolio returns," Working paper, Haas School of Business, University of California, Berkeley.

[14] Brunnermeier, M. and L. H. Pedersen (2009). "Market liquidity and funding liquidity," *Review of Financial Studies*, 22, 2202–2236.

[15] Caballé, J., and M. Krishnan, 1994, "Imperfect Competition in a Multi-Security Market with Risk Neutrality," *Econometrica*, 62, 695–704.

[16] Cespa, G. (2004). "A comparison of stock market mechanisms," *Rand Journal of Economics*, 35, 803–824.

[17] Cespa, G. and T. Foucault (2009). "Insider-Outsider, transparency and the value of the ticker," *CEPR Discussion Paper #6794*.

[18] Chakrabarty, B. and P. Moulton (2009). "Earning more attention: the impact of market design on attention constraints," mimeo, Fordham Graduate School of Business.

[19] Chamley, C. (2007). "Complementarities in information acquisition with short-term trades," *Theoretical Economics*, 441-467.

[20] Comerton-Forde, C., Jones, C., Hendershott, T., Seasholes, M., and P. Moulton (2010). "Time variation in liquidity: The role of market maker inventories and revenues," *Journal of Finance*, 65, 295–331.

[21] Corwin, S. and J. Coughenour (2008). "Limited attention and the allocation of effort in securities trading," *Journal of Finance*, 63, 3031–3067.

[22] Corwin, S. and M. Lipson (2011). "Order Characteristics and the Sources of Commonality in Prices and Liquidity," *Journal of Financial Markets*, 14, 47-81.

[23] Coughenour, J. and M. Saad (2004). "Common market-makers and commonalities in liquidity," *Journal of Financial Economics*, 73, 37–69.

[24] Chordia, T., Roll R. and A. Subrahmanyam (2000). "Commonality in liquidity," *Journal of Financial Economics*, 56, 3–28.

[25] Chordia, T., Sarkar A., and A. Subrahmanyam (2005). "An empirical analysis of stock and bond market liquidity," *Review of Financial Studies*, 18, 85–129.

[26] Chowdry, B. and V. Nanda (1991). "Multimarket trading and market liquidity," *Review of Financial Studies*, 4, 483–511.

[27] Edwards, A., Harris, L., and M. Piwowar (2005). "Corporate bond markets transparency and transaction costs," *Journal of Finance*, 62, 1421–1451.

[28] Ganguli, J.V. and Yang, L. (2009), "Complementarities, multiplicity, and supply information," *Journal of the European Economic Association*, 7, 90–115.

[29] Glosten, L., and L. Harris, (1988). "Estimating the Components of the Bid-Ask Spread," *Journal of Financial Economics*, 21, 123–142.

[30] Goldstein, M., Hotchkiss, E. and E. Sirri (2007) "Transparency and liquidity: A controlled experiment on corporate bonds," *Review of Financial Studies*, 20, 235–273.

[31] Gromb, D. and D. Vayanos (2002) "Equilibrium and welfare in markets with financially constrained arbitrageurs," *Journal of Financial Economics*, 66, 361–407.

[32] Grossman, S. and J. Stiglitz (1980). "On the impossibility of informationally efficient markets," *American Economic Review*, 70, 393-408.

[33] Hasbrouck, J. and D. J. Seppi (2001). "Common factors in prices, order flows, and liquidity," *Journal of Financial Economics* 59, 383–411.

[34] Hendershott, T. and Riordan, R. (2010). "Algorithmic trading and information," Working paper, University of California at Berkeley.

[35] Hendershott, T., Li, S., Menkveld, A., and Seasholes, M. (2010). "Risk sharing, costly participation, and monthly returns," Working paper, University of California at Berkeley.

[36] Hendershott, T. and Seasholes, M. (2009). "Market predictability and non informational trading," Working paper, University of California at Berkeley.

[37] Huberman, G. and D. Halka (2001). "Systematic Liquidity," *Journal of Financial Research*, 24, 161–178.

[38] Jovanovic, B. and A. Menkveld (2010). "Middlemen in limit order markets," available at `http://papers.ssrn.com/sol3/`.

[39] King, M. and S. Wadhwani (1990). "Transmission of volatility between stock markets," *Review of Financial Studies*, 3, 5–33.

[40] Kyle, A. and W. Xiong (2001). "Contagion as a wealth effect," *Journal of Finance*, 56, 1401–1440.

[41] Koch, A., S. Ruenzi, and L. Starks (2010). "Commonality in liquidity: a demand-side explanation," Working paper, University of Texas.

[42] Kodres, L., and M. Pritsker (2002). "A rational expectations model of financial contagion," *Journal of Finance*, 57, 769–799.

[43] Korajczyk, R. and R. Sadka (2008). "Pricing the commonality across alternative measures of liquidity," *Journal of Financial Economics*, 87, 45–72.

[44] Kyle A. (1985), "Continuous auctions and insider trading," *Econometrica*, 53, 1315–1335.

[45] Naik, N. Y. and P. K. Yadav (2003), "Do dealers manage inventory on a stock-by-stock or a portfolio basis?" *Journal of Financial Economics*, 69, 325–353.

[46] Pasquariello, P. (2007). "Imperfect competition, information heterogeneity, and financial contagion," *Review of Financial Studies*, 20, 391–426.

[47] Pasquariello, P. and C. Vega (2009). "Strategic cross-trading in the U.S. stock market," Working paper, University of Michigan.

[48] Vives, X. (1995). "Short-term investment and the informational efficiency of the market," *Review of Financial Studies*, 8, 125–160.

[49] Veldkamp, L. (2006). "Media frenzies in markets for financial information," *American Economic Review*, 96, 577–601.

# Figures

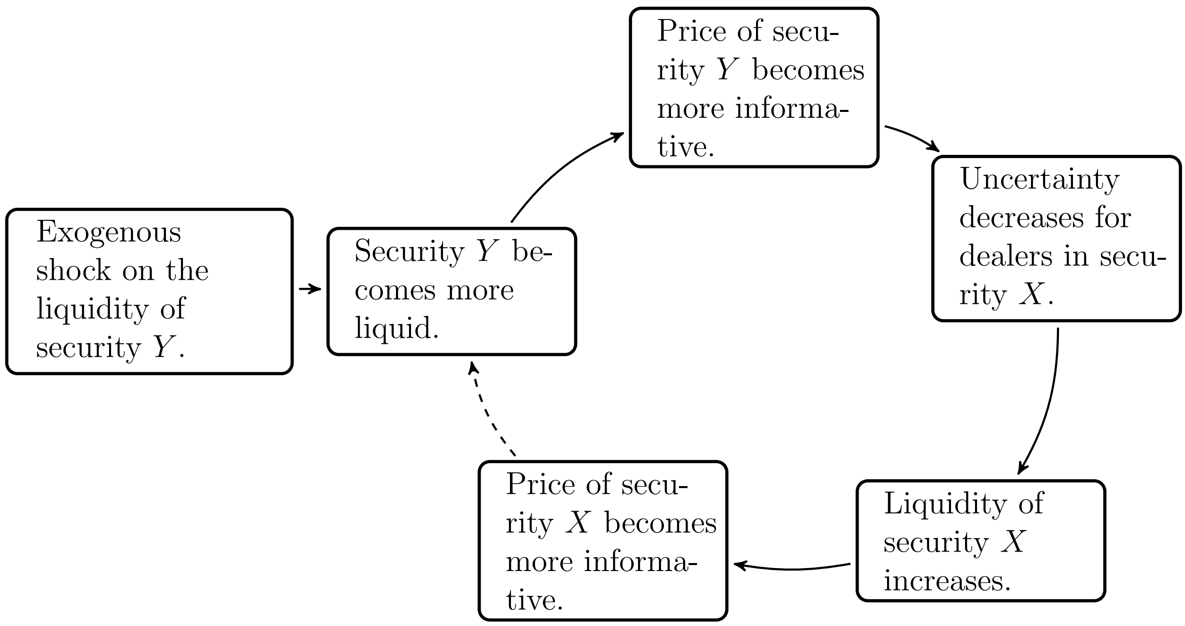
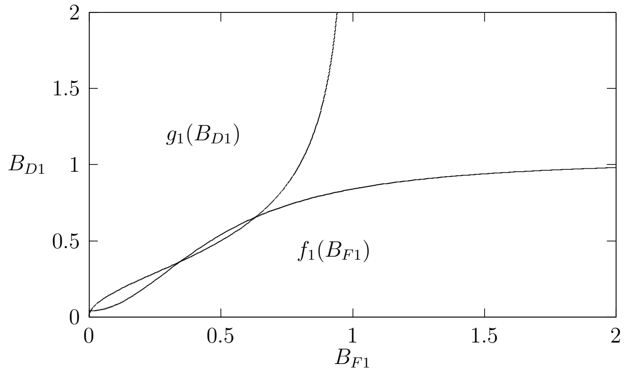
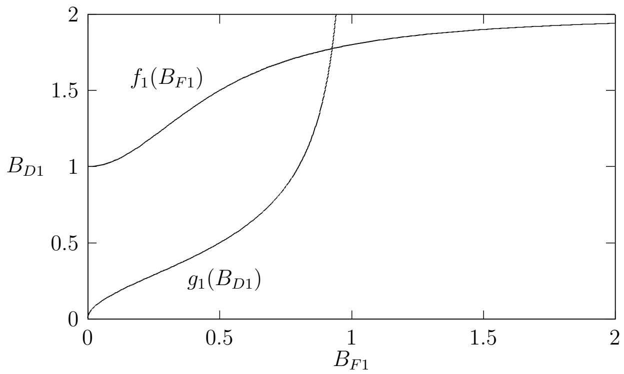


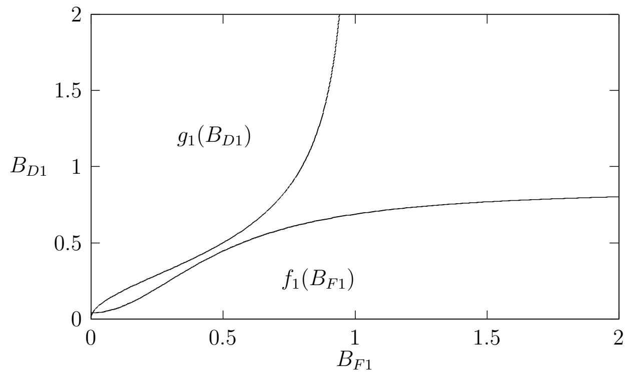

Figure 1: Cross-asset learning and liquidity spillovers.




(a)




(b)




(c)

Figure 2: Equilibrium determination with full attention: multiplicity (panel (a)) and uniqueness (panel (b) and (c)). Parameters' values are as follows: $\gamma_j = d = 1$, $\sigma_{u_j} = 2$, and $\sigma_\eta = .2$ (panel (a)), while in panel (b) we set $\sigma_\eta = 1$ and in panel (c) we set $d = 0.9$.

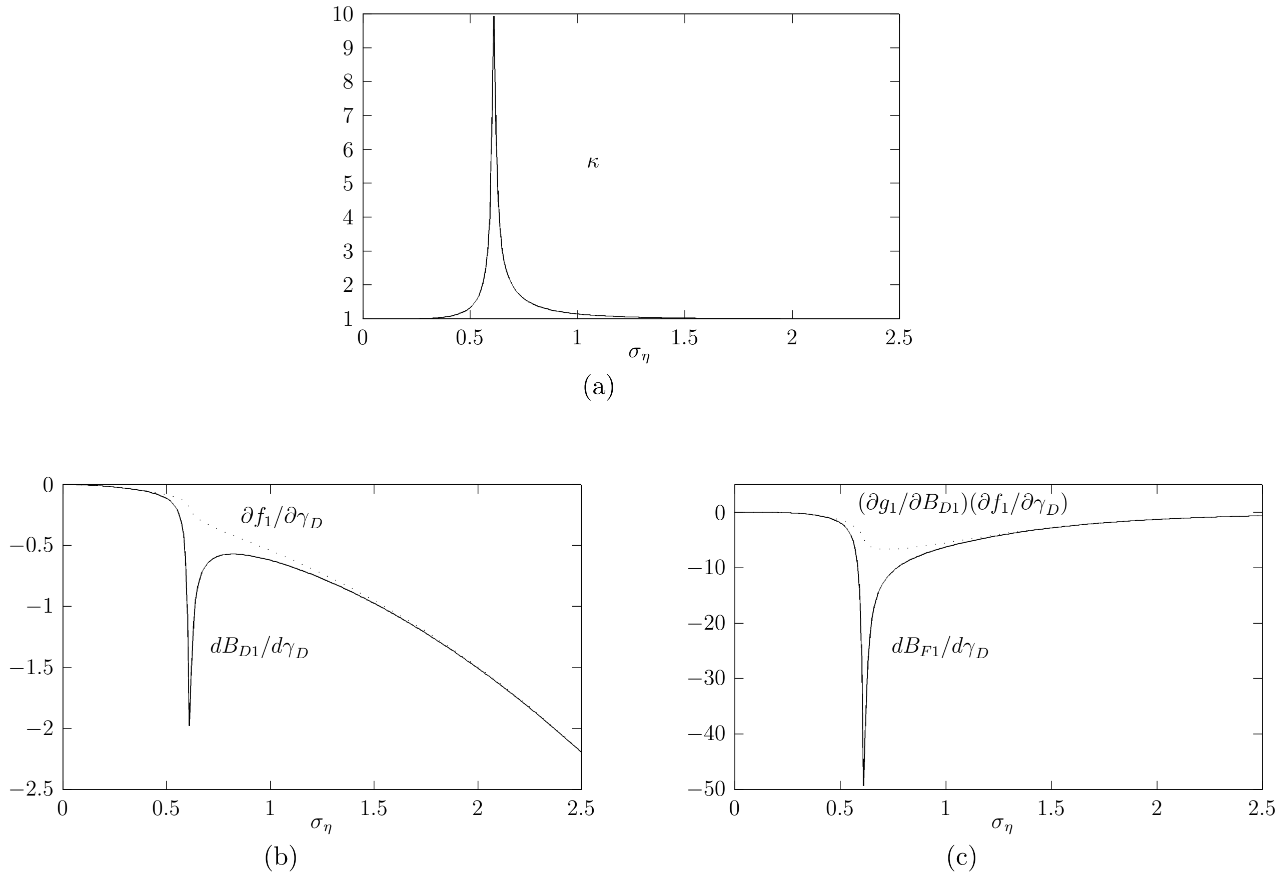



Figure 3: Illiquidity multiplier. In panel $(a)$ we plot $\kappa$ as a function of $\sigma_\eta$. Panels $(b)$ and $(c)$ show the direct effect (dotted line) and total effect (plain line) of a change in the risk tolerance of the dealers in security $D$ on the illiquidity of securities $D$ and $F$, respectively as a function of $\sigma_\eta$. Other parameter values are $\sigma_{u_F} = .1$, $\sigma_{u_D} = 1.6$, $\gamma_D = 1.8$, and $\gamma_F = 1/24$.

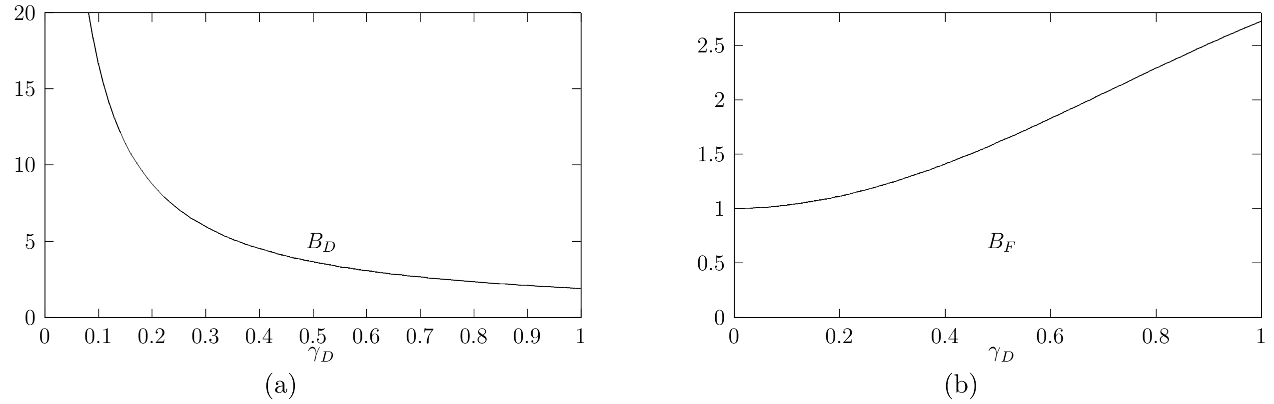



Figure 4: Negative liquidity spillovers. Parameters' values are as follows: $\sigma_{u_F} = .1$, $\sigma_{u_D} = 1$, $\gamma_F = 1$, $d = 1$, $\mu_F = \mu_D = .1$, $\sigma_\eta = 1$, and $\gamma_D \in \{.01, .02, \ldots, 1\}$.

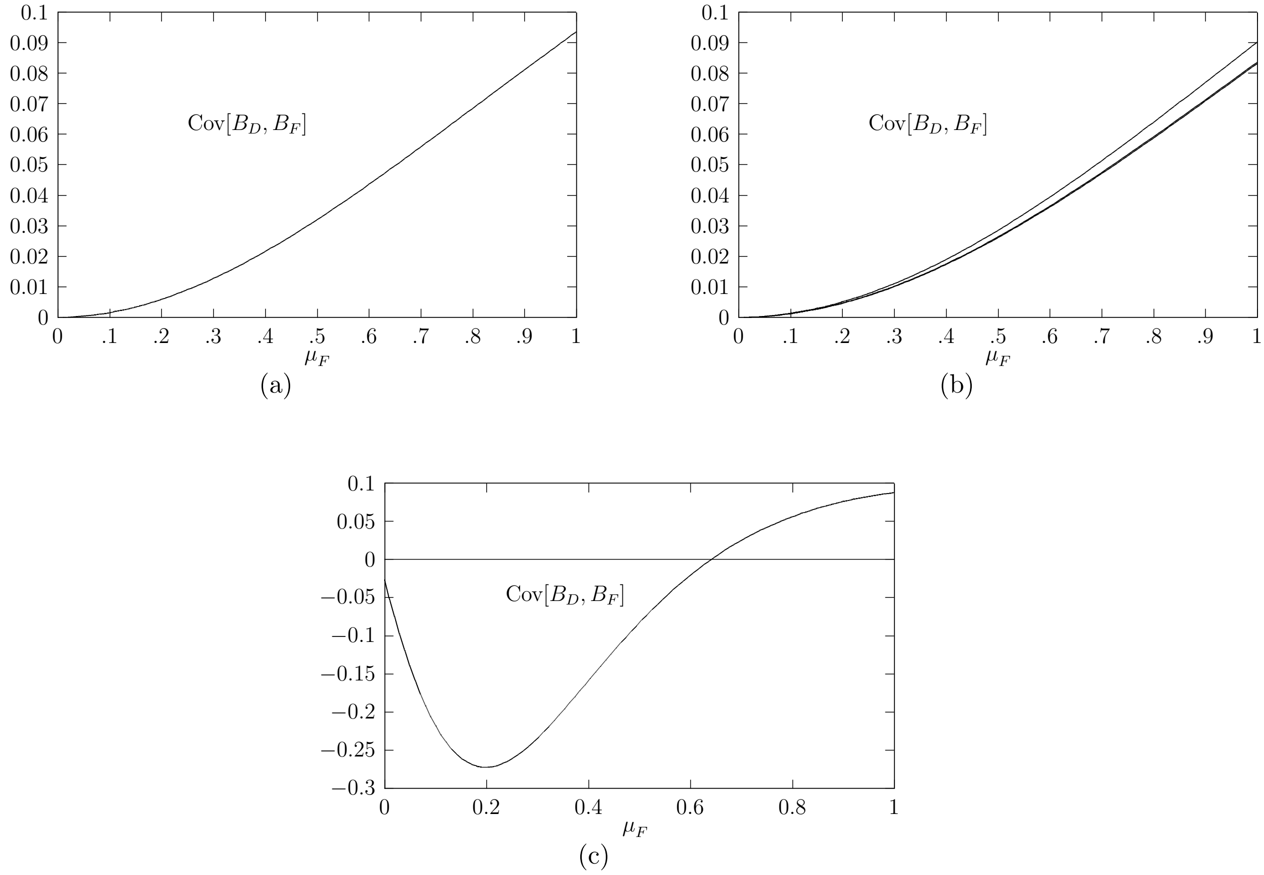
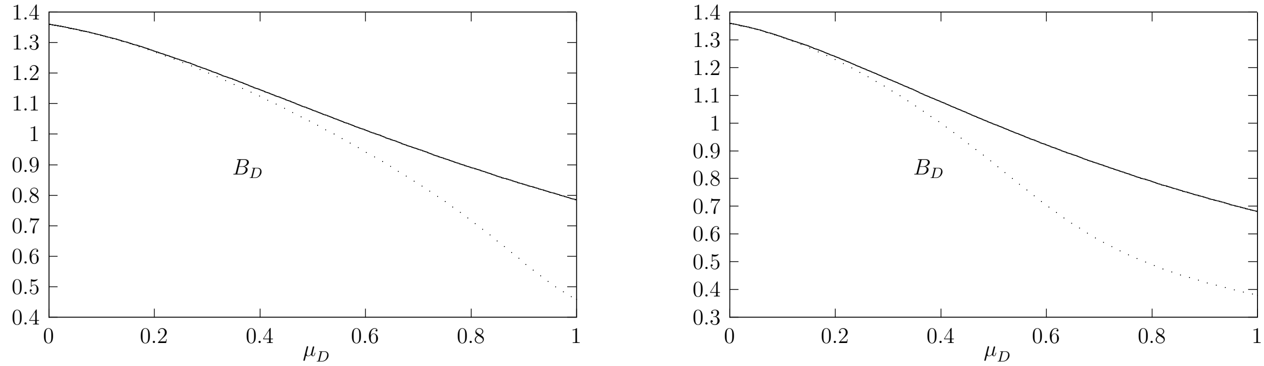



Figure 5: Comovement in illiquidity. The figure displays the covariance between the illiquidity of security $F$ and the illiquidity of security $D$ as a function of $\mu_F$ when $d = 0$ (panel (a)) and $d = 0.9$ (panels (b) and (c)). In panel (b) the covariance between the illiquidity of the two securities is higher when $\mu_D = 0.9$ (light curve) than when $\mu_D = 0.1$ (bold curve), for all values of $\mu_F > 0$. Other parameter values are $\sigma_{u_F} = \sigma_{u_D} = 1/2$, $\sigma_\eta = 2$, $\gamma_F = 1/2$, and $\mu_D \in \{0.1, 0.9\}$ for panels (a) and (b), while in panel (c) we set $\sigma_{u_F} = 0.1$, $d = \mu_D = 0.9$ and keep the other parameters' values fixed.




(a) (b)

Figure 6: The figure displays the illiquidity of security $D$ as a function of $\mu_D$ when $\mu_F = 0.5$ (in panel (a)) and when $\mu_F = 0.9$ (panel (b)) when $B_F$ is fixed at its equilibrium value for $\mu_D = 0.001$ (bold curve) and when instead it adjusts to its equilibrium value for each value of $\mu_D$ (dotted curve). The difference between the two curves shows the amount by which spillover effects magnify the direct effect of a change in attention on illiquidity. Parameters' values are as follows: $\sigma_{u_D} = \sigma_{u_F} = 1$, $\sigma_\eta = 0.77$ and $d = \gamma_D = \gamma_F = 1$.

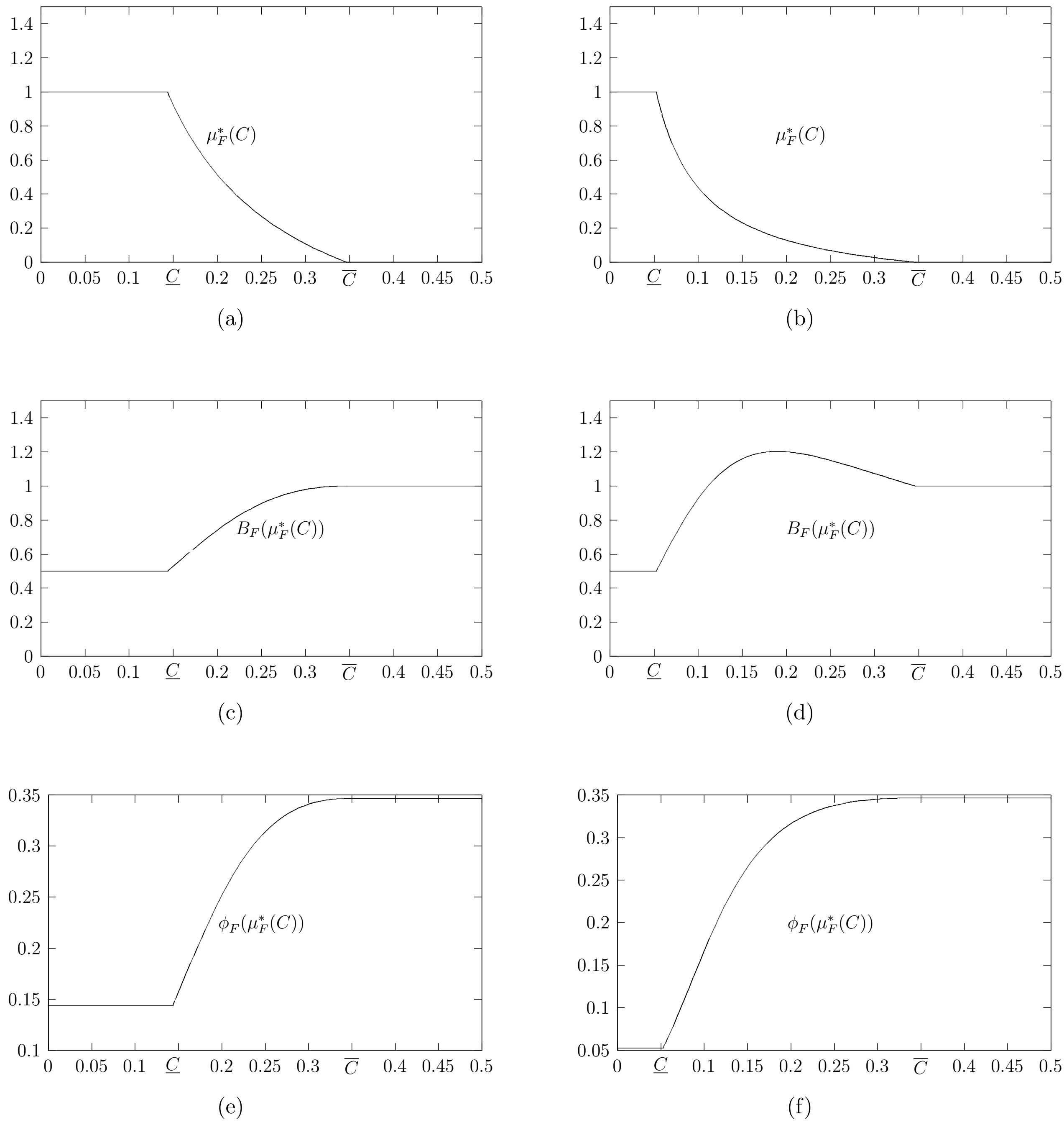



Figure 7: Impact of a change in the cost of attention on the fraction of pricewatchers, illiquidity, and the value of information with one-sided learning. Case with $\mathcal{R}_F \leq 1$ (panels (a), (c), and (e)), and case with $\mathcal{R}_F > 1$ (panels (b), (d), and (f)). Parameters' values are as follows: $\sigma_{u_D} = 1$, $\gamma_F = \gamma_D = 1$, $d = 0$, and $\sigma_\eta = 1$, with $\sigma_{u_F} = 1$ in panels (a), (c), and (e) whereas $\sigma_{u_F} = 0.5$ in panels (b), (d), and (f).

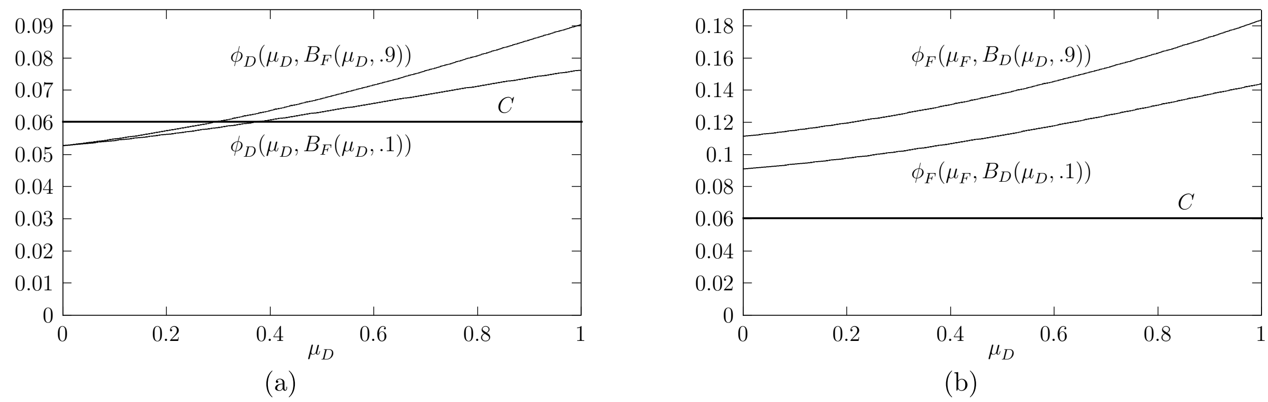



Figure 8: Positive feedback effect and cross-market monitoring effect. In panel (a) we plot $\phi_D$ as a function of $\mu_D$, for $\mu_F \in \{0.1, 0.9\}$. In panel (b) we plot $\phi_F$ as a function of $\mu_D$, for $\mu_F \in \{0.1, 0.9\}$. Other parameter values are as follows: $\sigma_\eta = 1$, $\sigma_{u_F} = \sigma_{u_D} = 1$, $\gamma_F = \gamma_D = 1$, and $d = 1$.

# *Subsidizing Liquidity: The Impact of Make/Take Fees on Market Quality*[*]

Katya Malinova and Andreas Park[†]
University of Toronto

April 26, 2011

**Abstract**

In recent years most equity trading platforms moved to subsidize the provision of liquidity. Under such a make/take fee structure, submitters of limit orders typically receive a rebate upon execution of their orders, and the exchange covers its costs by charging a higher fee for market orders. Trading rebates have, arguably, been a major facilitator for the emergence of algorithmic trading. We study the impact of this, now prevalent, fee structure on market quality, market efficiency, and trading activity by analyzing the introduction of liquidity rebates on the Toronto Stock Exchange. Using a proprietary dataset, we find that the liquidity rebate structure leads to decreased spreads, increased depth, increased volume, and intensified competition in liquidity provision. Explicitly accounting for exchange fees and rebates, we further find that trading costs for market orders did not decrease and that per share revenues for liquidity providers increase, despite the reduced bid ask spreads and increased competition. Finally, we find no evidence for changes in intermediation or market efficiency.

*JEL Classification: G12, G14.*

*Keywords:* Liquidity credits, market quality, trading.

[*]Financial support from the SSHRC (grant number 410101750) is gratefully acknowledged. We thank Gustavo Bobonis and Martin Wagener for insightful discussions and Zuhaib Chungtai, Andrew Bolyaschevets, James Cheung, Steve El-Hage, and Michael Brolley for valuable research assistance. We also thank the Toronto Stock Exchange (TSX) for providing us with a database and Alex Taylor for providing insights into the data. The views expressed here are those of the authors and do not necessarily represent the views of the TMX Group.

[†]E-mail: `katya.malinova@utoronto.ca` (corresponding author) and `andreas.park@utoronto.ca`.

The equity trading landscape has changed dramatically over the last decade. Worldwide, most public markets moved away from human interactions and are now organized as electronic limit order books, where traders either post passive limit orders that offer to trade a specific quantity at a specific price or submit active market(able) orders that "hit" posted limit orders. Posters of passive limit orders provide, or "make", liquidity, submitters of active market orders "take" liquidity. In contrast to traditional intermediated markets, limit order books rely on the voluntary provision of liquidity and must offer enough of it to attract trading. As a result, it is now the industry standard to subsidize passive trading volume.

This practice, known as make/take fees, is controversial. It has been argued that the subsidies caused excessive intermediation by attracting algorithmic traders that solely focus on capturing fee rebates.[1] Moreover, while some market-making firms are in favour of liquidity subsidies, other market participants have voiced concerns that make/take fees could result in excessive costs for liquidity takers.[2] To the best of our knowledge, there is no empirical study that conclusively addresses advantages and disadvantages of make/take fees. The present study aims to fill this gap.

Our analysis is based on trading fee changes on the Toronto Stock Exchange (TSX) and uses a proprietary database.[3] The TSX phased in the liquidity fee rebates on two distinct dates, introducing them on October 01, 2005 for all securities that were interlisted with NASDAQ or AMEX and on July 01, 2006 for the remainder of the securities. We study the 2005 change,[4] after which an active marketable order incurred a per share fee of \$.004 and a passive limit order that is "hit" received a per share fee rebate of \$.00275.

[1]See "Rise of the machines: Algorithmic trading causes concern among investors and regulators", The Economist July 30th 2009.

[2]See, for instance, the comments for the make/take fee structure in the options markets sent to the SEC by GETCO at http: //www.getcollc.com/images/uploads/getco_comment_090208.pdf, or the petition by Citadel in favor of a fee cap at http: //www.sec.gov/rules/petitions/2008/petn4-562.pdf. Responding to these concerns, the SEC even imposed a 30-cent ceiling on stock exchanges for 100-share equity trades.

[3]TSX Inc. holds copyright to the data, all rights reserved. It is not to be reproduced or redistributed. TSX Inc. disclaims all representations and warranties with respect to this information, and shall not be liable to any person for any use of this information.

[4]The 2006 event also involved a change in the fees for the NASDAQ and AMEX interlisted securities, making it difficult to isolate the effect of liquidity rebates.

Active orders for stocks that did not move to this rebate structure incurred a cost of 1/55 of 1% (1.8 basis points) of the dollar value of the transaction and passive orders were free. To put the make/take fees into perspective, the median end July 2005 closing price in our sample of 73 companies that were interlisted with NASDAQ and AMEX is $6.08. The per share taker fee of $0.004 translates into a fee of 6.58 basis points at the median, the passive side's per share rebate of $.00275 translates into 4.52 basis points at the median.

Our empirical strategy is an event study on the introduction of the fee rebates. Since the change affected the incentives for liquidity provision for only a subset of companies, we are able to control for market wide conditions by matching securities that were affected with securities that were not. We then perform tests using a difference-in-differences approach to capture the marginal impact of the fee structure change on market quality, trader welfare, volume, and competition for liquidity provision.

We assess market quality by standard bid-ask spread, depth and market efficiency measures. We find that, compared to the control group, securities that were interlisted on NASDAQ or AMEX experienced a decrease in their time weighted quoted spreads of 12.1 basis points and an increase in their quoted depth.[5] Studying autocorrelations of midquote returns, and the 5/30 minute and 15/30 minute variance ratios to detect changes in market efficiency, we find no effect. We thus conclude that the fee rebates improve liquidity offered throughout the day and that there is no evidence that they affect market efficiency.

A liquidity taker's welfare is commonly measured by the transaction costs, which are proxied by the effective spread. For a buyer initiated transaction, the effective spread is twice the difference between the average per share price and the prevailing midpoint of the quoted bid and offer prices. We observe a marked decline in effective spreads, which indicates that liquidity makers passed on some of their fee rebate to takers. After adjusting the effective spread to account for the exchange fees, however, we find no evidence that

[5] Bid-ask spreads on the TSX are, on average, larger than those on U.S. exchanges, even though the TSX is one of the world's largest exchanges by market capitalization and trading volume. Since 2005, however, spreads have fallen substantially.

transaction costs have declined — instead we identify a (statistically weak) increase.

A liquidity maker's per share revenue is commonly proxied by the magnitude of the price reversal after a transaction, and it is measured by the realized spread. For a buyer initiated transaction, the realized spread is twice the difference between the average per share price and the midpoint of the quoted bid and offer prices several minutes after the transaction. Here, too, we observe a decline in the spread.

The decrease in the spreads suggests that liquidity providers pass on some of their rebate to liquidity takers. One question is whether competition is so fierce that the entire rebate gets "competed away". To fully capture the revenue benefit to liquidity providers, we adjust the realized spread to include the fee rebate. We find that the total revenues to liquidity makers actually increased and that this effect is particularly pronounced for stocks with low competition for liquidity provision.

A key objective of subsidizing liquidity provision is for the exchange to attract more volume. We indeed find an increase in volume, which is somewhat surprising considering that transaction costs actually went up. A potential criticism of fee rebates is that an increase in volume may be caused merely by increased intermediation. The argument is that to capture liquidity rebates, an intermediary such as an algorithm "injects" itself between two (cost insensitive) traders who would have otherwise transacted on their own. As our data allows us to identify orders that originate from clients, we can study intermediation by analyzing the fraction of client to non-client trades. If there are relatively more client to non-client trades, then the higher volume is at least partly due to an increase in intermediation. Yet we do not find any change in the fraction of client to non-client trades and are left with the puzzle that both volume and transaction costs have increased.

Finally, with the introduction of fee rebates, ceteris paribus, it becomes cheaper to post limit orders. It is then imaginable that institutions see the introduction of rebates as an opportunity to enter the market for liquidity provision. To asses the extent of competition, we count the number of improvements in the best bid and offer prices and depth, the number of liquidity providing market participants that are involved in transactions, and we compute the Herdindahl Index of market concentration. The latter, also known as the Herfindahl-Hirschman Index,[6] is widely used as a proxy for the competitiveness of a given industry — for instance, the U.S. Department of Justice and the Federal Trade Commission use it to assess the effects of a merger on competition — and it is computed as the sum of the squared market shares. The higher the index, the lower the level of competition. When it comes to trading, the good provided is liquidity. In traditional dealer markets, market share in liquidity is synonymous with market share in volume and the Herfindahl index for the concentration of market making is computed based on dealers' shares of volume (see Ellis, Michaely, and O'Hara (2002) and Schultz (2003)). In an electronic limit order book such as the TSX, liquidity is supplied by passive orders. We thus measure a trader's market share as the fraction of limit order volume that this trader provides.

We find a significant increase in the number of improvements in the bid ask spread and depth, which we show to be driven by improvements in depth. The number of spread improvements, on the other hand, declines. Since the average depth also increases, we conclude that after the fee change, traders compete more aggressively on depth. We further show that the increase in the number of quote improvements is driven by two factors. First, traders compete more aggressively for liquidity provision, as is implied by a decrease in the Herfindahl Index. Second, we find (weak) evidence that the fee rebates attract new entry in the market for liquidity provision.

To summarize our results, we find that competition, particularly on depth, intensifies. Although liquidity providers lower spreads in response to the fee change, their per share revenues increase, taking rebates into account. This hints at the possibility that competition in prices is less relevant than competition for market share in liquidity provision.

Colliard and Foucault (2011) provide theoretical guidance for the effects of a fee change. They show that trader welfare is affected only by the total fee, i.e. the sum of maker and taker fees, and that the make/take fee composition has no impact, provided

[6]See, e.g. Tirole (1988); see also Hirschman (1964) for a discussion of the origin of the index.

the tick size is zero, because quotes adjust to neutralize any fee redistribution. In our study, the total fee increases for stocks with low prices and declines for stocks with high prices. Since the fees change for all stocks, we cannot address changes in the composition. However, we do find support for Colliard and Foucault's theoretical prediction that an increase in the total fee decreases taker welfare. Furthermore, our findings support their prediction that the bid-ask spread decreases in the take fee and increases in the make fee.

Foucault, Kadan, and Kandel (2009) find theoretically that the optimal make/take fee composition depends on the relative levels of competition among the liquidity providers and liquidity demanders, and on the relative monitoring costs for these two groups. They argue that the lower fee (or a rebate) on the liquidity makers will increase the trading rate and aggregate welfare only under some conditions (for instance, when liquidity providers have higher monitoring costs than liquidity demanders, or when the level of competition among liquidity providers is low compared to that among liquidity demanders). When these conditions are not satisfied, the optimal make/take fee structure would impose higher fees on makers rather than on takers. Finally, our work also relates to Degryse, Van Achter, and Wuyts (2011) who theoretically study the impact of clearing and settlement fees on liquidity and welfare.

The next section reviews trading on the TSX and the details of the fee changes. Section 2 describes the data, the sample selection, and the regression methodology. Section 3 discusses results on market quality and efficiency. Section 4 describes trader welfare, Section 5 presents results on volume and intermediation, Section 6 discusses competition. Section 7 concludes. Tables and figures are appended.

# 1 The Toronto Stock Exchange and its Trading Fees

## 1.1 Trading on the TSX

The Toronto Stock Exchange (TSX) has been an electronic-only trading venue since it closed its physical floor in 1997. In 2005, the TSX was the seventh largest exchange world-wide in terms of market capitalization of traded securities and twelfth largest in dollar trading volume.[7]

Trading on the TSX is organized in an upstairs-downstairs structure. Orders can be filled by upstairs brokers (usually these are very large orders), who have price improvement obligations, or they can be cleared via the consolidated (electronic) limit order book. The TSX limit order book generally follows the so-called price-time priority.[8] It is constructed by sorting incoming limit orders lexicographically, first by their price ("price priority") and then, in case of equality, by the time of the order arrival (with the earlier orders enjoying the "time priority"). Transactions in the limit order book occur when active orders — market orders (orders to buy or sell at the best available price) or marketable limit orders (e.g. a buy limit order with a price higher than the current best ask) — are entered into the system. Unpriced market orders occur very infrequently on the TSX, and in what follows we will use the term "active order" to for the marketable portion of an order, and we use "passive order" for a standing limit order that is hit by an active order. Active orders "walk the book", i.e., if the order size exceeds the number of shares available at the best bid or offer price, then the order continues to clear at the next best price.

All orders must be sent to the TSX by registered brokers (the Participating Organizations (P.O.)). Trading is organized by a trading software (the trading engine), and our data is the audit trail of the processing of the trading engine. We describe the data in more detail in Section 2. Orders of sizes below round lot size (for the companies in our sample this size is 100 shares) are cleared by the equity specialist, referred to as the Registered Trader (RT). Similarly, portions of orders that are not multiples of the round lot size (e.g. 99 shares of a 699 share order) will be cleared by the RT, after the round lot portion of the order has cleared (e.g. the 99 shares of a 699 share order will clear after, and only if, the 600 shares have cleared). Furthermore, the RT has the obligation to provide minimum fills when there are no standing limit orders, but the RT's powers

[7]Source: *World Federation of Security Exchanges*.

[8]One exception to this rule is a so-called unintentional cross, where time priority is overruled if active and passive orders are submitted by the same broker.

are small compared to those of the NYSE designated market maker (formerly referred to as the specialist),[9] and the RT is involved in only about 1.3-1.4% of the dollar volume in our sample (see Table 3).

The TSX with its public, electronic limit order book thus largely relies on its users to voluntarily supply liquidity by posting limit orders. This system contrasts traditional systems where dealers are institutionally obliged to make a market.

## 1.2 Details of the Change in Trading Fees

The TSX was a monopolist for equity trading in Canada during our sample period, and the lack of market fragmentation allows us to isolate the impact of liquidity rebates. When fee rebates were introduced in Europe or the U.S., on the other hand, these markets were already beginning to fragment.

The TSX phased in the liquidity rebates on two discrete dates, introducing them on October 01, 2005 for the TSX companies that were interlisted on NASDAQ or AMEX; on July 01, 2006 all remaining companies switched; we focus on the 2005 change of fees.[10]

Prior to October 01, 2005, all TSX securities were subject to the so-called value-based trading fee system, under which the active side of each transaction incurred a fee based on the dollar amount of the transaction (1/50 of 1% a the dollar-amount in the months immediately preceding October 01) and the passive side incurred no fee or rebate. On October 01, TSX-listed securities that were also interlisted with NASDAQ and AMEX switched to a volume-based trading regime, under which for each traded share the active side had to pay a fee of $.004 and the passive side obtained a rebate on its exchange fees of $.00275. All other securities remained at the prevailing value-based regime, although, the fees were slightly reduced — after October 01, 2005, active orders incurred a fee of

[9] Subject to tight rules, the RT has the right to participate in orders to unload a pre-existing inventory position that she or he built up in the process of providing liquidity to markets. The RT has no informational advantage over other traders.

[10] We restrict attention to the 2005 change for two reasons: first, in 2006 there was a change in the level of fees simultaneously with the switch to a make/take fee structure. Second, a difference-in-differences analysis in 2006 has less statistical power because the treatment group, non-interlisted securities, is much larger than the control group, interlisted stocks.

1/55 of 1% of the dollar-amount of the transaction and passive orders remained free. The value based taker fee per trade is capped at \$50, the volume based taker fee and maker rebate are capped at \$100 and \$50, respectively.

Compared to the old value based fee structure, the new volume based billing yields the TSX higher per share fee revenue for securities that trade below \$6.875. Liquidity takers pay less for securities that trade above \$22.[11] To put these fees into perspective, the median closing price at the end of July 2005 in our sample of the companies that were interlisted with NASDAQ and AMEX is \$6.08. Under the "old" value-based system, the per share taker fee is 1.8 basis points (which is \$0.00111 at the median), there was no maker fee or rebate, and thus the TSX's per share revenue is 1.8 basis points. Under the "new" volume based billing, the taker fee is \$0.004 (or 6.58 basis points at the median), the passive side's rebate is \$.00275 per share (or approximately 4.52 basis points at the median), and thus the TSX's revenue at the median price is about 2 basis points.

# 2 Data, Sample Selection, and Methodology

## 2.1 Data Sources

Our analysis is based on a proprietary dataset, provided to us by the Toronto Stock Exchange (TSX). Data on market capitalization, monthly volume, splits, and (inter-) listing status is obtained from the monthly TSX e-Reviews publications. Data on the CBOE's volatility index VIX is from Bloomberg. We analyze the effect of the fee structure change by looking at a 4 month window (2 months before and 2 months after the introduction of the liquidity rebates), from August 01, 2005 to November 30, 2005. The TSX participating organizations are billed at the end of each month, and the event window was chosen to include the month immediately following the change as well as one month after the first bill that was based on the new fee structure. We exclude trading days that have no

[11] Total fees coincide for the price $p$ that solves $p \times 1/55 \times 1\% = (\$.004 - \$.00275)$, active fees coincide for the price $p$ that solves $p \times 1/55 \times 1\% = \$.004$.

or limited U.S. trading (an example is the U.S. Thanksgiving and the Friday following it); information on scheduled U.S. market closures is obtained from the NYSE Calendar. We further exclude October 11, 2005 and November 21, 2005 as the TSX data included several recording errors for these days.

The TSX data that is provided to us is the input-output of the central trading engine, and it includes all messages that are sent to and from the brokers. The data contains public and private information for all orders, cancellations and modifications sent to the limit order book, public and private information on all trade reports, and details on dealer (upstairs) crosses. Further, the data contains all the system messages and user notifications, for instance, announcements about changes in the stock status, such as trading halts and freezes, announcements about estimated opening prices, indications that there is too little liquidity in the book (the spread is too wide), and so on.

Each message consists of up to 500 subentries, such as the date, ticker symbol, time stamp, price, volume, and further information that depends on the nature of the message. For instance, order submission, notification and cancellation messages contain information about the order's price, total and displayed volume, the orders's time priority, broker ID, trader ID, order number (new and old for modifications), information about the nature of the account (e.g. client, inventory or equity specialist), information about whether an order is submitted anonymously or whether the broker number is to be displayed in the TSX pay-for data feed,[12] information about whether an order is a short sale, and some further details that we do not exploit in this project.

For each order that is part of the trade, the data additionally contains the volume of the transaction as well as the public (as sent to the data feeds) and private (the actual) remaining volumes, information on whether an order was filled by a registered trader and where it was executed (e.g. in the public limit order book, with a specialist outside the limit order book (for oddlots), in the market for special terms orders, or crossed by a

[12]In accordance with Canadian regulations, the choice of whether to attribute the order to a particular dealer remains with the dealer. Submitting a non-anonymous order may be advantageous for time priority reasons. Traders can also specify that they do not want to clear against an anonymous order.

broker). The liquidity supplier rebates only affect trades that clear via the limit order book. Consequently, we exclude opening trades, oddlot trades, dealer crosses, trades in the special terms market, and trades that occur outside normal trading hours.

Importantly for the construction of the liquidity and competition measures, the transaction data specifies the active (liquidity demanding) and passive (liquidity supplying) party, thus identifying each trade as buyer-or seller-initiated. Finally, one useful system message is the "prevailing quote". It identifies the best bid and ask quotes as well as the depth at the best quotes, and it is sent each time there is a change in the best quotes or the depth at these quotes. This message allows us to precisely identify the prevailing quote at each point in time.

## 2.2 Sample Selection

We construct our sample as follows. Out of 3,000+ symbols that trade on the TSX, we include only common stock and exclude debentures, preferred shares, notes, rights, warrants, capital pool companies, stocks that trade in US funds, companies that are traded on the TSX Venture and on the NEX market, exchange traded funds, and trust units. We require that the companies had positive volume in July 2005, according to the TSX e-Review, and were continuously listed between July 2005 and November 2005. We further exclude securities that had stock splits, that were under review for suspension, that had substitutional listings, and that had an average daily midquote below $1.

Differently to commonly applied filters, we retain companies with dual class shares. This is due to a peculiarity of the Canadian market, where, as of August 2005, an estimated 20-25% of companies listed on the TSX made use of some form of dual class structure or special voting rights, whereas in the United States, only about 2% of companies issue restricted voting shares (see Gry (2005)). We exclude Nortel (symbol: NT) because it was involved in a high profile accounting scandal at the time of our sample period (along with Worldcom and Enron). Finally, we omit companies that have insufficient trading for the computation of major liquidity measures; specifically, we require that there is enough data to compute the realized spread for 95% of the 80 trading days that comprise our sample.

We determine a company's interlisted status from the TSX e-Reviews. We then classify companies as "interlisted with NASDAQ or AMEX" in our 2005 sample if they were interlisted with NASDAQ or AMEX from August to November 2005 and non-interlisted with NASDAQ and AMEX if they were not interlisted from August to November. Companies that changed their (inter-)listing status during the sample period or for which the status was unclear were omitted from the sample.

We are then left with 73 NASDAQ and AMEX interlisted companies and 374 TSX only and NYSE interlisted companies. In what follows, we will refer to companies that are interlisted with NASDAQ and AMEX as "interlisted", and we will refer to companies that are listed only on the TSX or that are interlisted with NYSE as "non-interlisted".

## 2.3 Matched Sample

We construct the matched sample as follows. Using one-to-one matching without replacement, we determine a unique non-interlisted match for each of the interlisted securities based on closing price, market capitalization, and a level of competition for liquidity provision, as measured by the Herfindahl Index (formally defined in the next subsection).

One-to-one matching without replacement based on closing price and market capitalization has been shown to be the most appropriate method to test for difference in trade execution costs; see Davies and Kim (2009). We additionally include a measure of competition as a matching criterium, for three reasons. First, our treatment group, the interlisted securities, is not a random sample, and liquidity provision in the average interlisted stock is systematically more competitive than in the average TSX only stock, even controlling for market capitalization. Second, the focus of this study is not only trade execution costs but also other variables that are affected by competition, such as traders' behavior, welfare and the levels of intermediation.[13] Finally, we aim to identify the impact of the introduction of the liquidity rebates, and according to Foucault, Kadan, and Kandel (2009), who study the make/take fees theoretically, this impact depends on the level of competition among traders.

We randomize the order of matching by sorting the stocks in the treatment group (i.e. the interlisted securities) alphabetically by symbol. The match for each treatment group security $i$ is then defined to be a control group security $j$ that minimizes the following matching error:

$$matcherror_{ij} := \left|\frac{p_i - p_j}{p_i + p_j}\right| + \left|\frac{MC_i - MC_j}{MC_i + MC_j}\right| + \left|\frac{HHI_i - HHI_j}{HHI_i + HHI_j}\right|, \tag{1}$$

where $p_i$, $MC_i$, and $HHI_i$ denote security $i$'s July 2005 closing price, market capitalization as of the end of July 2005, and the average July 2005 value of the Herfindahl Index at the broker level, respectively. Tables 14 and 15 contain the list of interlisted companies and their matches.

## 2.4 Measuring Competition: The Herfindahl Index

We quantify competition among traders by the Herfindahl Index. The index is widely used to assess market concentration and it computed as the sum of the squared market shares. We study the market for liquidity provision. In an electronic limit order book, liquidity is provided by passive orders and a trader's market share is the fraction of passive limit order volume that this trader provides.[14] The Herfindahl Index for different levels of liquidity providing entities (e.g., broker, trader) per day $t$ per security $i$ is

$$HHI_{it} = \sum_{k=1}^{n_t} \left( \frac{passive\ volume_{it}^k}{\sum_{k=1}^{n_t} passive\ volume_{it}^k} \right)^2, \tag{2}$$

[13] When matching only on price and market capitalization, the results for most liquidity measures, including spreads (the variable of interest in Davies and Kim (2009)), are similar.

[14] Weston (2000), Ellis, Michaely, and O'Hara (2002) and Schultz (2003) use the Herfindahl Index of market concentration to assess competition for market making in dealer markets; their indices are based on NASDAQ dealers' shares of volume.

where $n_t$ is the number liquidity providing entities on day $t$ in security $i$ and $\textit{passive volume}_{it}^k$ is the $k-$th entity's total passive volume for that day and security. Higher values of the index correspond to higher levels of market concentration and thus to lower levels of competition (value 1 corresponds to monopolistic liquidity provision).

We consider two levels of liquidity providing entities, namely, the broker and the trader level. At the broker level, the passive volume per security per day is the total intraday passive volume of that broker, excluding dealer crosses. The "broker level HHI" does not differentiate between trades that brokers post by client request and that they post on their own accounts to make a market. To better understand the behavior of institutions that provide liquidity on an ongoing basis, we compute the index for traders that trade in and out of their inventories; in our data such trades stem from either an inventory or a equity specialist account. We refer to the latter index as the "trader level HHI."

We also compute the number of liquidity providing brokers and liquidity providing inventory traders to shed some light on possible changes in competition indices.

## 2.5 Panel Regression Methodology

For each security in our sample and for each of their matches, we compute a number of liquidity and market activity measures for the 4 month window around the event date (2 months before and after October 01, 2005). Our panel regression analysis employs a difference in differences approach and thus controls for market-wide fluctuations. To additionally control for U.S. events that may affect interlisted securities differentially, we include the CBOE volatility index VIX in our regressions. For each measure, we run the following regression[15]

$$\textit{dependent variable}_{it} = \beta_0 + \beta_1 \textit{fee change}_t + \beta_2 \mathsf{VIX}_t + \sum_{j=1}^{8} \beta_{2+j} \textit{control variable}_{ij} + \epsilon_{it},$$

[15]This regression methodology is similar to that in Hendershott and Moulton (2011). We discuss an alternative methodology in the appendix.

where *dependent variable*$_{it}$ is the time $t$ realization of the measure for treatment group security $i$ less the realization of the measure for the $i$th control group match; *fee change*$_t$ is an indicator variable that is 1 after the event date and 0 before; $\mathsf{VIX}_t$ is the closing value of CBOE's volatility index for day $t$, and *control variable*$_{ij}$ are security level control variables for the company and its match: the log of the market capitalization, the log of the closing price, and the share turnover and the daily midquote return volatility in the month before the event window, July 2005.[16] Summary statistics for our treatment and control group are in Table 2.

We conduct inference in all regressions in this paper using double-clustered Cameron, Gelbach, and Miller (2011) standard errors, which are robust to both cross-sectional correlation and idiosyncratic time-series persistence.[17] For brevity we display only the estimates for the coefficient $\beta_1$ on the fee change dummy, and we omit the estimates for the constant as well as estimates for the coefficients on VIX and on the controls. The number of observations roughly equals the number of companies in the treatment group multiplied with the number of trading days in our sample periods (correcting for a small number of missing observations when a company or its match did not trade for a day), at most 5,840 observations.

**Regressions for Subsamples.** In addition to analyzing the impact of the fee structure change on the entire sample, we estimate the effects separately for the groups of treatment companies above and below the median with respect to pre-sample (July 31, 2005) market capitalization, total July 2005 trading volume (in shares), and the average July 2005 Herfindahl index of market concentration at the broker level. Medians of market capitalization, volume, and the Herfindahl Index are, respectively, \$475 million, 1.795

[16]In unreported regressions we further controlled for company fixed effects. We also used dynamic instead of the July 2005 static controls for prices. In both cases, the results are similar.

[17]Cameron, Gelbach, and Miller (2011) and Thompson (2010) developed the double-clustering approach simultaneously. We follow the former and employ their programming technique. See also Petersen (2009) for a detailed discussion of (double-) clustering techniques.

million shares, and 0.2296 (Table 2). We estimated the following equations

$$\begin{aligned} dependent\ variable_{it} &= \beta_0 + \beta_1 fee\ change_t \times above\ median_i \\ &+\beta_2 fee\ change_t \times below\ median_i + \beta_3 above\ median_i \\ &+\beta_4 \mathsf{VIX}_t + \sum_{j=1}^{8} \beta_{4+j} control\ variable_{ij} + \epsilon_{it}, \end{aligned} \quad (3)$$

where $above\ median_i$ is an indicator variable that equals 1 if security $i$ has market capitalization (or trading volume, TSX share of volume, Herfindahl index) above the median; similarly for the variable $below\ median_i$.

Furthermore, as we explain in Section 1.2, under the new volume-based make/take fee structure liquidity takers pay lower fees for stocks that trade at high prices (above \$22). We thus estimated the effects separately for stocks with July 31 closing prices above and below \$22, where the regression equation is the same as (3), except $above\ median_i$ equals 1 if security $i$'s July 31 closing price is above \$22; likewise for $below\ median_i$. We will henceforth refer to a closing price of \$22 as the "break-even price." Similarly, in Section 1.2 we also explain that the total fees, i.e. taker fee minus maker rebate, increase for securities that trade at prices below \$6.875 and otherwise decrease. We thus study subsamples of securities with July 31 closing prices above and below \$6.875.

We report only the estimates of interest, i.e. the estimated coefficients on the interaction terms $fee\ change_t \times above\ dummy_i$ and $fee\ change_t \times below\ dummy_i$. Results from tests for differences in the coefficients are indicated in the respective tables.

# 3 Market Quality

## 3.1 Quoted Liquidity

We measure quoted liquidity using time and trade weighted quoted spreads and depth. The *quoted spread* is the difference between the best price at which someone is willing to buy, or the offer price, and the best price at which someone is willing to sell, or the bid price. We express the spread measures in basis points as a proportion of a prevailing quote midpoint. *Share depth* is defined as average of the number of shares that can be traded on the bid and offer side; the *dollar depth* is the dollar amount that can be traded at the bid and the offer. We use logarithms of the depth measures to ensure a more symmetric distribution since several Canadian companies, particularly, non-interlisted ones, historically have very large depth. High liquidity refers to large depth and small spreads.

The trade weighted spread and depth are the prevailing spread and depth averaged over transactions, and they capture the impact of the fee change on executions. The time weighted measures additionally reflect the availability of liquidity throughout the day.

**Results.** Figure 1 shows a marked decline in the quoted spread after the event date and an increase in the dollar depth. The summary statistics in Table 3 paint a similar picture, and our panel regressions further confirm these observations. The panel regression results for the change in the quoted spread are in the first two columns of Table 4. The first column depicts the time weighted quoted spreads, the second column displays the trade weighted quoted spreads.

The average price for interlisted companies on September 30, 2005, was \$12.07, the median price was \$5.66. The size of the rebate in 2005 was ¢.275 per share, which translates into 4.56 and 9.72 basis points at the average and median prices, respectively, for a round-trip transaction (i.e., a simultaneous passive buy and sell). We observe that the estimate on the time weighted quoted spread declines by 12.09 basis points, the trade weighted quoted spread declines by 9.34 basis points. The latter is roughly the amount of the rebate at the median price and around double the rebate at the mean price. These results are significant at the 1% level.

When considering subsamples, we find that significant effects arise for stocks that trade below the break-even price for market orders, \$22, for all levels of competition, market capitalization, and total fees, and for stocks that have high volume. Further, the coefficient estimates differ significantly for subsamples with respect to the break-even price.

Table 5 displays the results of our panel regressions on depth. We find that time and trade weighted share and dollar depth all increase significantly. Further, these increases are significant in the subsamples of securities with prices below the break-even price for market orders, with prices above the break-even price for total fees, with high competition, with high market capitalization, and with low trading volume.

In summary, quoted liquidity improves in that spreads become tighter and more shares/dollar volume can be traded at the best bid and offer prices.

## 3.2 Effective Liquidity

Quoted liquidity only measures posted conditions, whereas effective liquidity captures the conditions that traders decided to act upon. The costs of a transaction to the liquidity demander are measured by the *effective spread*, which is is the difference between the transaction price and the midpoint of the bid and ask quotes at the time of the transaction. For the $t$-th trade in stock $i$, the proportional effective spread is defined as

$$espread_{ti} = 2q_{ti}(p_{ti} - m_{ti})/m_{ti}, \tag{4}$$

where $p_{ti}$ is the transaction price, $m_{ti}$ is the midpoint of the quote prevailing at the time of the trade, and $q_{ti}$ is an indicator variable, which equals 1 if the trade is buyer-initiated and $-1$ if the trade is seller-initiated. Our data includes identifiers for the active and passive side for each transaction, thus precisely signing the trades. Further, our data is message by message, as processed by the trading engine, and it includes quote changes. The prevailing quote is thus precisely identified as the last quote before the transaction.

The change in liquidity provider profits is measured by decomposing the effective spread into its permanent and transitory components, namely the *price impact* and the *realized spread*,

$$espread_{ti} = priceimpact_{ti} + rspread_{ti}. \tag{5}$$

The price impact reflects the portion of the transaction costs that is due to the presence of informed liquidity demanders, and a decline in the price impact would indicate a decline in adverse selection. The realized spread reflects the portion of the transaction costs that is attributed to liquidity provider revenues. In our analysis we use the five-minute realized spread, which assumes that liquidity providers are able to close their positions at the quote midpoint five minutes after the trade. The proportional five-minute realized spread is defined as

$$rspread_{ti} = 2q_{ti}(p_{ti} - m_{t+5\text{ min},i})/m_{ti}, \tag{6}$$

where $p_{ti}$ is the transaction price, $m_{ti}$ is the midpoint of the quote prevailing at the time of the $t$-th trade, $m_{t+5\text{ min},i}$ is the midpoint of the quote 5 minutes after the $t$-th trade, and $q_{ti}$ is an indicator variable, which equals 1 if the trade is buyer-initiated and $-1$ if the trade is seller-initiated.

**Results.** Figure 2 plots the 5-day moving averages of the effective spread and the price impact for each of our the treatment group of interlisted and their control group matches. The figure suggests that the change in the fee structure led to a decrease in the effective spread, and it also indicates a decline in the price impact. The summary statistics in Table 3 point to significant improvement of liquidity, and the panel regressions confirm this observation.

The third column of Table 4 shows that after the fee change effective spreads fell significantly, by about 10 basis points. We further find significant effects in subsamples with prices below the break-even price of $22, for low market capitalization, high trading volume, and all levels of competition. Coefficients for the subsample estimates differ significantly for below vs. above the break-even price.

The fourth column of Table 4 displays our regression results for realized spreads. We find that 5-minute realized spreads decline by 5.23 basis points. In subsamples we find significant effects for prices blow the break-even price, high competition, and high volume. The price impact, listed in the fifth column of Table 4 declines by 5 basis points. In subsamples we find significant effects for prices blow the break-even price, low competition, low market capitalization, and high volume.

The decline in transaction costs, as measured by the effective spread, can be due to liquidity makers foregoing some of their revenue, or it can be attributed to a change in trade informativeness. We conclude that the liquidity providers share some portion of the rebate by lowering their revenue and also that adverse selection declines. The decline in adverse selection is consistent with the idea that narrower spreads attract new, price-sensitive uninformed traders and informed traders with weaker information. Our findings on an increase in volume that we discuss in Section 5 further support this idea.

With perfect competition for liquidity provision, liquidity makers would pass on their credits to liquidity takers across the board. We find, however, that the effective spread declines *only* for the subsample of securities that have higher per share fees for liquidity takers under the new volume based make/take fee system compared to the old value-based billing. Since the realized spread also declines significantly for this subsample, we conclude that liquidity providers only pass on their rebates for the subset of securities that experienced an increase in liquidity takers fees.

Colliard and Foucault (2011) provide some theoretical guidance for the effects of a fee change. Their model predicts that the bid-ask spread decreases in the take fee and increases in the make fee. In our study, the make fee declines (from 0 to $-\$.00275$ per share), and we find that spreads decline, as predicted (see Table 4). The take fee, on the other hand, increases for stocks with low prices and declines for stocks with high prices. Consistent with the theoretical predictions, we find that spreads decline for low price stocks, and that the coefficient for high price stocks is insignificantly different from 0.

## 3.3 Market Efficiency

We measure market efficiency with two standard proxies, the return autocorrelation and the variance ratio. Specifically, we analyzed the impact of the liquidity rebate structure on the first order autocorrelations of 5-, 15-, and 30-minute midquote returns, and the 5/30 minute and 15/30 minute variance ratios, as described in Campbell, Lo, and MacKinley (1997), calculated for each security each day. Prices that follow a random walk, should have a return autocorrelation of zero. Autocorrelations are negative on average, thus an increase in autocorrelation or a decrease in its absolute value would signify improved market efficiency. The 5-minute/30-minute variance ratio is six times the 5-minute variance of midquote returns divided by the 30-minute variance of midquote returns; similarly for the 15-/30 minute variance ratios. The variance ratio evaluates whether short-term price changes are reversed on average. Such reversals, if they exist, would indicate that over short horizons, trades cause prices to deviate from the (efficient) equilibrium price. As there is usually some excess volatility, the variance ratio is commonly greater than one, and thus a decline in the variance ratio would indicate improved market efficiency.

Table 6 displays the results of our panel regressions the impact of the fee change on autocorrelations and variance ratios.[18] We do not find significant effects for any of the measures.

# 4 Trader Welfare

The effective spread is often considered to be the best measure for transaction costs. The spread does not, however, include exchange fees. To determine a liquidity demander's welfare, it is important to explicitly account for these fees. We thus compute

$$\textit{fee adjusted espread}_{ti} = (2q_{ti}(p_{ti} - m_{ti}) + 2 \times \textit{exchange fee}_{ti})/m_{ti}, \tag{7}$$

where $\textit{exchange fee}_{ti}$ is the per share fee to remove liquidity. Before the change of fees it is $1/50 \times 1\% \times p_{ti}$ for all securities, and after the change it is $1/55 \times 1\% \times p_{ti}$ for non-interlisted stocks and \$0.004 for interlisted stocks.

Similarly, the realized spread is considered to measure the benefit to the liquidity

[18] The table displays the results using signed autocorrelations; results for absolute values are similar.

provider. To explicitly account for liquidity rebates, we compute

$$\textit{rebate adjusted rspread}_{ti} = (2q_{ti}(p_{ti} - m_{t+5 \text{ min},i}) + 2 \times \textit{fee rebate}_{ti})/m_{ti}, \tag{8}$$

where $\textit{fee rebate}_{ti}$ is the per share maker fee rebate. It is 0 for all securities before the fee change. After the change it is 0 for non-interlisted stocks and \$.00275 for interlisted stocks.

**Results.** Focussing only on effective and realized spreads and omitting exchange fees may give the misleading impression that liquidity demanders unambiguously benefit while liquidity takers obtain reduced revenue. Figure 3 shows instead that after the fee change, the passive side benefited, and it indicates that the costs for the active side did not decrease.

Table 7 shows the regression results for fee and rebate adjusted spreads. We find that the fee adjusted effective spreads increase, although the significance is only at the 10% level. The table also shows that total liquidity provider revenues increase, and thus the liquidity rebates more than compensate the liquidity providers for the revenue that is passed on to liquidity demanders. Furthermore, there are stark differences in revenues between low and high competition and low and high price stocks.[19]

Colliard and Foucault (2011) predict that the fee adjusted effective spread (the "cum fee" spread in their paper) increases in the total fee. In our case, total fees decline for stocks priced below \$6.875 (see Section 1.2). Consistent with the theoretical predictions, we find that for the subsample with prices below \$6.875, exchange fee adjusted effective spreads increase. For prices above \$6.875, the coefficient is negative, but statistically insignificant. Further, the difference in the subsample coefficients is statistically significant.

[19]The increase for low price stocks is probably in part caused by the fact that the fixed amount rebate has a stronger relative impact when the price is low.

# 5 Volume

One key question is whether changes in fees have any effect on trading behavior. If traders engage in the same transactions irrespective of the exchange fees, then the change in fees is merely redistributive and has no impact on aggregate welfare.

To detect changes in behavior, we study the impact of the fee change on the number of shares traded, the dollar amount of all trades, and the number of transactions. We further decompose these numbers into volume that stems from clients and non-clients to understand if there are changes in intermediation.

**Aggregate Volume.** Table 8 displays our results on volume and the number of transactions, measured in logarithms. Our results suggest that the fee change increases volume, dollar volume, and the numbers of transactions.

**Intermediated Volume.** One possible explanation for the increase in volume is an increase in intermediation. When traders are not overly sensitive to transaction costs, an intermediary, such as an algorithm programmed to take advantage of fee rebates, may be able to inject itself between two traders who would have otherwise transacted on their own. We proxy for the extent of intermediation by the fraction of volume that occurs between a client and an intermediary.[20] Table 10 shows our findings on intermediated trades and indicates no change in the extent of intermediation.

**Market Participation.** The increase in volume could also stem from the entry of new traders. We study changes in market participation by analyzing client volume. Table 9 displays our findings and shows that client volume increases significantly. This finding is consistent with the result on the decreased price impact if one believes that the reduced spreads attract price sensitive or less well informed traders. New entry is, however, somewhat surprising because transaction costs did not decline (Section 4).

[20] Our data identifies client trades as well as equity specialist, broker inventory, and option market maker trades. We classify all parties other than clients as intermediaries.

# 6 Competition in Liquidity Provision

With the introduction of fee rebates, ceteris paribus, it becomes cheaper to post limit orders. It is then imaginable that institutions see the introduction of rebates as an opportunity to enter the market for liquidity provision. To asses the extent of competition, we count the number improvements of the best bid and offer prices and depth, the number of liquidity providing market participants that are involved in transactions, and we compute the Herdindahl Index of market concentration (introduced in Section 2.4).

## 6.1 Improvements in the Quoted Bid-Ask Spread and Depth.

The first column in Table 12 summarizes our findings on the total number of spread and depth improvements. We find a significant increase in the number of improvements, which indicates increased competition. The second and third columns show that this increase is driven by improvements in depth, while the number of spread improvements declines. Since the average depth also increased, we conclude that after the fee change, traders compete more aggressively on depth.

The decline in the number of spread improvements is consistent with our finding that average depth increases. As depth increases, fewer trades walk the book and there may be fewer opportunities to improve the spread after the book was depleted. Furthermore, since quoted spreads decline, there is less room for improving the spread.

Our findings on the increase in the number of quote improvements are consistent with Foucault, Kadan, and Kandel (2009) who predict, in particular, that the liquidity providers' monitoring activity increases as their fee decreases.

## 6.2 Market Participation and Concentration.

The increase in the number of quote improvements could be driven by two factors: first, existing traders may compete more aggressively, and second, the liquidity rebates may have attracted new traders. The Herfindahl Index at the trader level, which we focus on here, is based the shares of passive volume that traders provide from their inventory, and it captures the first factor.

The first column of Table 11 displays our results on the trader level HHI. The decline in the index signifies reduced market concentration and increased competition. Looking at the subsample of stocks that trade below $22, we find that competition increases significantly. This finding is consistent with the significant increase in depth that we observe there.

To assess market participation, we count the number of liquidity providing brokerages and the number of liquidity providing inventory traders. The number of brokers per security per day is the number of unique broker IDs that were on the passive side of transactions. The number of inventory traders is the number of unique trader IDs that traded on an inventory or equity specialist accounts and that were on the passive side of transactions. Table 2 shows for interlisted stocks that the median numbers of brokers and inventory traders were 12 and 4, respectively. Columns two and three in Table 11 reveal that the number of brokers and traders both increased after the change, although the coefficient on the number of traders is significant only at the 10% level.

We thus conclude that competition in the market for liquidity provision increased and that this increase is at least in part driven by market entry.

# 7 Conclusion

The introduction of fee rebates for passive volume on the Toronto Stock Exchange led to a substantial decline in bid-ask spreads, an increase in depth and an increase in volume. The changes in spreads are consistent with theoretical predictions, but the increase in volume is puzzling since transaction costs, accounting for both the spread and the exchange fees, did not go down. That being said, the increase in volume is consistent with a theoretical prediction of Colliard and Foucault (2011) who find a positive relation of trading fees and volume for some parameter values.

We also find that after the introduction of the fee rebates, liquidity providers compete more aggressively for market share in the "make" market. Furthermore, even though liquidity providers lower their spreads in response to the fee change, when taking rebates into account, liquidity providers' per share revenues increase. These two findings together suggest that competition in depth is at least as important as competition in spreads.

# Appendix: Alternative Methodology

**Alternative Specification.** Our main regression equation uses as dependent variables the time realization of various measures for treatment group security less the realization of the measure for the control group match. An alternative differences in differences approach is to regress the levels directly on the event and the interlisting status as the main effects and on the interaction of these two. The coefficient on the latter is then the variable of interest. Specifically, the alternative regression equation is

$$\begin{aligned} \textit{dependent variable}_{it} &= \beta_0 + \beta_1 \textit{fee change}_t + \beta_2 \textit{interlisted}_i + \beta_3 \textit{fee change}_t \times \textit{interlisted}_i \\ &\quad + \beta_4 \mathsf{Volatility}_t + \sum_{j=1}^{4} \beta_{3+j} \textit{control variable}_{ij} + \epsilon_{it}, \qquad (9) \end{aligned}$$

where *dependent variable*$_{it}$ is the time $t$ realization of the measure security $i$; *fee change*$_t$ is an indicator variable that is 1 after the event date and 0 before; *interlisted*$_i$ is an indicator variable that is 1 if the security is interlisted and 0 otherwise; $\mathsf{Volatility}_t$ is the closing value of a volatility index for day $t$, and *control variable*$_{ij}$ are security level control variables for the company: a variable that relates to the price of security $i$, the log of the market capitalization on July 31, 2005, and the share turnover and the daily midquote return volatility in the month before the event window, July 2005. The variable of interest for our study is $\beta_3$.

We ran regression (9) for several variations of the security price based control: the log of the closing price on July 31, 2005 (as in the main text), the midquote for stock $i$ on day $t$ as well as its logarithm, the return of stock $i$ from day $t-1$ to $t$, the return for stock $i$ from day $t-2$ to $t-1$. We further used two volatility indices: the CBOE's VIX and the TMX's MVX. The MVX is based on the implied volatility of index options on the TSX60 stock index and it is highly correlated ($> 70\%$) with the VIX.

The regression results using this alternative specification are similar.

# References


Bessembinder, H., and K. Venkataraman (2004): "Does an electronic stock exchange need an upstairs market?," *Journal of Financial Economics*, 73(1), 3–36.

Bloomfield, R., M. O'Hara, and G. Saar (2005): "The make or take decision in an electronic market: Evidence on the evolution of liquidity," *Journal of Financial Economics*, 75, 165–199.

Cameron, A. C., J. B. Gelbach, and D. L. Miller (2011): "Robust Inference with Multi-Way Clustering," *Journal of Business Economics and Statistics*, forthcoming.

Campbell, J. Y., A. W. Lo, and A. C. MacKinley (1997): *The Econometrics of Financial Markets*. Princeton University Press.

Colliard, J.-E., and T. Foucault (2011): "Securities market structure, trading fees and investors' welfare," working paper, HEC Paris.

Davies, R. J., and S. S. Kim (2009): "Using matched samples to test for differences in trade execution costs," *Journal of Financial Markets*, 12(2), 173 – 202.

Degryse, H., M. Van Achter, and G. Wuyts (2011): "Internalization, Clearing and Internalization, Clearing and Settlement, and Stock Market Liquidity," Discussion paper, Erasmus University Rotterdam.

Ellis, K., R. Michaely, and M. O'Hara (2002): "The Making of a Dealer Market: From Entry to Equilibrium in the Trading of Nasdaq Stocks," *The Journal of Finance*, 57(5), pp. 2289–2316.

Foucault, T., O. Kadan, and E. Kandel (2009): "Liquidity Cycles and Make/Take Fees in Electronic Markets," Discussion paper, EFA 2009 Bergen Meetings Paper.

Gry, T. (2005): "Dual-Class Share Structures and Best Practices in Corporate Governance," Staff Report PRB 05-25E, Staff of the Parliamentary Information and Research Service (PIRS).

Hasbrouck, J. (2007): *Empirical Market Microstructure*. Oxford University Press.

Hendershott, T., C. Jones, and A. Menkveld (2010): "Does Algorithmic Trading Improve Liquidity?," *Journal of Finance*, forthcoming.

Hendershott, T., and P. Moulton (2011): "Automation, Speed, and Stock Market Quality: The NYSE's Hybrid," *Journal of Financial Markets*, forthcoming.

Hirschman, A. O. (1964): "The Paternity of an Index," *The American Economic Review*, 54(5), p. 761.

Hollifield, B., R. A. Miller, P. Sandas, and J. Slive (2006): "Estimating the Gains from Trade in Limit-Order Markets," *The Journal of Finance*, 61(6), pp. 2753–2804.

O'Hara, M., and M. Ye (2010): "Is Market Fragmentation Harming Market Quality?," working paper, Cornell University.

Petersen, M. A. (2009): "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches," *Review of Financial Studies*, 22(1), 435–480.

Schultz, P. (2003): "Who makes markets," *Journal of Financial Markets*, 6(1), 49–72.

Smith, B. F., D. A. S. Turnbull, and R. W. White (2001): "Upstairs Market for Principal and Agency Trades: Analysis of Adverse Information and Price Effects," *The Journal of Finance*, 56(5), 1723–1746.

Thompson, S. B. (2010): "Simple formulas for standard errors that cluster by both firm and time," *Journal of Financial Economics*, In Press, Corrected Proof, –.

Tirole, J. (1988): *The Theory of Industrial Organization*. The MIT Press.

Weston, J. P. (2000): "Competition on the Nasdaq and the Impact of Recent Market Reforms," *Journal of Finance*, 55(6), 2565–2598.

**Table 1**
**Summary Statistics on Trading Activity for Interlisted Companies and their Non-Interlisted Matches**

The table lists aggregate trading volume numbers for the August-November 2005 sample period for NASDAQ/AMEX-interlisted companies and their respective matches. Percentage numbers are for the share that the respective number has of total volume.

| | | NASDAQ/AMEX interlisted | | Non-interlisted | |
|---|---|---|---|---|---|
| Total volume (excluding special terms market) | Share volume | 1,847,794,191 | | 2,140,879,197 | |
| | Dollar volume | $ 20,517,866,297 | | $ 26,768,731,058 | |
| | Transactions | 1,966,642 | | 1,451,526 | |
| Intraday | | 1,313,804,000 | 71.1% | 1,349,823,200 | 63.0% |
| | | $ 14,726,937,292 | 71.8% | $ 15,962,222,831 | 59.6% |
| | | 1,808,270 | 91.9% | 1,247,051 | 85.9% |
| Open | | 28,873,204 | 1.6% | 46,924,654 | 2.2% |
| | | $ 356,600,562 | 1.7% | $ 584,311,868 | 2.2% |
| | | 32,269 | 1.6% | 48,900 | 3.4% |
| Afterhours | | 87,457,828 | 4.7% | 107,148,290 | 5.0% |
| | | $ 2,180,634,369 | 10.6% | $ 2,215,202,425 | 8.3% |
| | | 21,516 | 1.1% | 17,704 | 1.2% |
| Dealer crosses | | 413,080,078 | 22.4% | 631,005,919 | 29.5% |
| | | $ 3,056,619,162 | 14.9% | $ 7,753,556,056 | 29.0% |
| | | 5,248 | 0.3% | 7,595 | 0.5% |
| Oddlots | | 4,579,081 | 0.2% | 5,977,134 | 0.3% |
| | | $ 197,074,912 | 1.0% | $ 253,437,878 | 0.9% |
| | | 99,339 | 5.1% | 130,276 | 9.0% |
| Equity specialist | | 66,763,881 | 3.6% | 92,300,034 | 4.3% |
| (all trades, including oddlots) | | $ 276,512,711 | 1.3% | $ 362,617,083 | 1.4% |
| | | 269,071 | 13.7% | 325,678 | 22.4% |
| Number of market orders | | 1,240,327 | | 779,492 | |
| Non-client market order volume | | 493,981,000 | 27% | 393,193,700 | 18% |
| Non-client market order transactions | | 585,996 | 30% | 293,166 | 20% |
| Client market order volume | | 819,823,000 | 73% | 956,629,500 | 82% |
| Client market order transactions | | 1,222,274 | 70% | 953,885 | 80% |

## Table 2
## Pre-sample Summary Statistics of Interlisted Companies and their Matches

The table lists selected summary statistics for the NASDAQ/AMEX-interlisted companies and their matches for the pre-sample month of July. Unless otherwise specified, the numbers are average per day per company. The letter M signifies millions. intraday volume refers to transactions that occur in the open market during regular trading hours (9:30-16:00), excluding oddlot trades, special terms orders and dealer crosses.

| | | NASDAQ/AMEX interlisted | Non-interlisted |
|---|---|---|---|
| Total July intraday volume in shares | Mean | 2,837,000 | 3,784,000 |
| | StE | (4,426,000) | (9,333,000) |
| | Median | 1,308,000 | 1,857,000 |
| Total July intraday dollar volume | | $37.1M | $39.7M |
| | | ($95M) | ($125M) |
| | | $8.617M | $12.4M |
| Total July transactions | | 4,407 | 3,320 |
| | | (6413) | (5209) |
| | | 2,354 | 1,870 |
| Closing price end July 2005 | | $ 11.95 | $ 12.13 |
| | | (17.30) | (17.09) |
| | | $ 6.08 | $ 6.12 |
| Market capitalization end July 2005 | | $1,330M | $1,500M |
| | | ($4,540M) | ($6,020M) |
| | | $475M | $392M |
| Time weighted quoted spread (in bps) | | 73.76 | 93.83 |
| | | (52.87) | (60.03) |
| | | 60.77 | 90.18 |
| Time weighted quoted spread (in cents) | | ¢4.781 | ¢6.271 |
| | | (¢4.644) | (¢5.210) |
| | | ¢3.525 | ¢4.578 |
| Time weighted dollar depth | | $15,196 | $20,759 |
| | | (13,173) | (16,632) |
| | | $11,786 | $16,825 |
| Herfindahl Index, broker level | | 0.235 | 0.249 |
| | | (0.075) | (0.081) |
| | | 0.23 | 0.247 |
| Herfindahl Index, trader level | | 0.476 | 0.592 |
| | | (0.171) | (0.209) |
| | | 0.471 | 0.607 |
| Number of brokers | | 12.73 | 12.16 |
| | | (5.384) | (5.504) |
| | | 11.9 | 11.45 |
| Number of market making traders | | 5.88 | 4.576 |
| | | (5.176) | (5.536) |
| | | 4.15 | 3.1 |

## Table 3
## Summary Statistics of Interlisted Companies and their Matches: Before and After the Change of Fees

The table lists selected summary statistics for the NASDAQ/AMEX-interlisted companies and their matches for the sample period August-November 2005, per day per company. All measures for spreads and transaction costs are in basis points of the prevailing midquote. The standard errors presented for the difference-in-differences are adjusted by factor $\sqrt{73}$; * indicates significance at the 10% level, ** at the 5% level, and *** at the 1% level.

| | | Treatment group of NASDAQ/AMEX interlisted stocks | | Control group of non-interlisted stocks | | |
|---|---|---|---|---|---|---|
| | | Before | After | Before | After | Diff-in-Diff |
| Intraday dollar volume (in logs) | Mean | 13.08 | 13.1 | 13.36 | 13.22 | 0.169** |
| | StE | (1.594) | (1.643) | (1.412) | (1.503) | (0.081) |
| | Median | 12.97 | 13.04 | 13.25 | 13.2 | |
| Time weighted quoted spread | | 70.59 | 71.63 | 88.96 | 103.2 | -13.25*** |
| | | (50.510) | (52.070) | (55.090) | (65.890) | (3.430) |
| | | 53.86 | 66.71 | 84.47 | 87.17 | |
| Effective spread | | 60.58 | 62.18 | 79.28 | 91.93 | -11.05*** |
| | | (42.310) | (43.570) | (49.520) | (60.660) | (3.112) |
| | | 45.18 | 60.35 | 77.23 | 81.42 | |
| Time weighted dollar depth (in logs) | | 9.364 | 9.38 | 9.722 | 9.637 | 0.101*** |
| | | (0.667) | (0.714) | (0.576) | (0.618) | (0.038) |
| | | 9.336 | 9.279 | 9.725 | 9.591 | |
| Exchange fee adjusted effective spread | | 64.58 | 81.72 | 83.28 | 95.56 | 4.862 |
| | | (42.300) | (56.170) | (49.520) | (60.660) | (3.445) |
| | | 49.18 | 72.6 | 81.22 | 85.05 | |
| Rebate adjusted realized spread | | 18.19 | 30.8 | 39.16 | 43.05 | 8.726*** |
| | | (20.940) | (32.510) | (31.300) | (38.090) | (2.717) |
| | | 12.75 | 21.11 | 34.87 | 35.4 | |
| Client to non-client trades as a fraction of total volume | | 45.7% | 46.0% | 41.3% | 40.7% | 0.78% |
| | | (0.081) | (0.092) | (0.092) | (0.103) | (0.009) |
| | | 46.3% | 46.3% | 40.8% | 40.8% | |
| Herfindahl Index, trader level | | 0.449 | 0.428 | 0.596 | 0.607 | -0.0317** |
| | | (0.174) | (0.170) | (0.213) | (0.214) | (0.014) |
| | | 0.464 | 0.424 | 0.608 | 0.606 | |

## Table 4
## Panel Regressions Results for Marginal Changes in Bid-Ask Spreads

Dependent variables are treatment group value minus control group value for time weighted and trade weighted quoted spread, effective spread, and 5-minute realized spread and price impact. All spreads and the price impact are measured in basis points of the prevailing midquote.

**Specifications that apply to this and all subsequent tables.** The treatment group in 2005 are the NASDAQ and AMEX interlisted securities. Each dependent variable is regressed on a dummy variable set equal to one for dates after October 01, 2005 and zero before, daily market volatility as measured by the CBOE VIX index, and the following control variables for the security and its match: log(market capitalization) and log(price) at July 31, 2005, and dollar turnover and return volatility in July 2005. Coefficients for volatility, control variables, and the constant are not reported for brevity. The full sample for 2005 is 73 securities. Standard errors are in parentheses; * indicates significance at the 10% level, ** at the 5% level, **+ at the 2%, and *** at the 1% level. Standard errors are robust to time series and cross-sectional correlation. Results other than the full sample are split by the median for the control group for July 2005 market capitalization ($475M), total volume (1.795M shares), and the Herfindahl Index (.2296). The break-even price for market orders is $22, for higher prices, market orders are cheaper under the new regime. The break-even price for maker plus taker fees is $6.875; for higher prices, the total fee is lower under the new regime. We report only the coefficient estimates for the interaction terms; see Section 2.5 for the full specification for the estimated equation. We test for equality of coefficients, where "Yes" indicates that we reject the hypothesis.

| | time weighted quoted spread | trade weighted quoted spread | effective spread | 5 min realized spread | 5 min price impact |
|---|---|---|---|---|---|
| full sample | -12.0928*** | -9.3401*** | -10.0538*** | -5.2311** | -5.0015**+ |
| | (3.4968) | (2.8437) | (3.0374) | (2.3381) | (2.0678) |
| — Break-even Price for Market Orders — | | | | | |
| above $22 | -1.6723 | -1.2625 | -1.0578 | -1.7872 | 0.7251 |
| | (5.1254) | (3.7694) | (4.0810) | (2.5619) | (3.2556) |
| below $22 | -13.7469*** | -10.6247*** | -11.4844*** | -5.7810** | -5.9144**+ |
| | (3.8598) | (3.1530) | (3.3793) | (2.6466) | (2.3920) |
| Different Coefficient? | Yes** | Yes** | Yes** | – | – |
| — Herfindahl Index — | | | | | |
| low competition | -13.1704** | -10.4836** | -11.4119** | -3.9198 | -7.8241** |
| | (5.9463) | (4.7917) | (4.9770) | (3.9657) | (3.7789) |
| high competition | -11.0444*** | -8.2329*** | -8.7383**+ | -6.4991**+ | -2.2761 |
| | (3.5352) | (3.0805) | (3.5879) | (2.7042) | (2.6600) |
| Different Coefficient? | – | – | – | – | – |
| — Market Capitalization — | | | | | |
| above median | -7.3514*** | -4.5349** | -4.8206* | -2.5039* | -2.3538 |
| | (2.7585) | (2.2464) | (2.7051) | (1.3444) | (2.3949) |
| below median | -16.7061*** | -14.0315*** | -15.1628*** | -7.9036* | -7.5986** |
| | (6.0966) | (4.9732) | (5.2298) | (4.3566) | (3.7889) |
| Different Coefficient? | – | Yes* | Yes* | – | – |
| —Break-even Price for Total Fees — | | | | | |
| above $6.875 | -10.2158** | -6.7037** | -6.7714** | -5.4727**+ | -1.3784 |
| | (4.5753) | (3.0908) | (3.3447) | (2.1683) | (1.9157) |
| below $6.875 | -13.6414*** | -11.5233*** | -12.7719*** | -5.0470 | -7.9979** |
| | (4.9228) | (4.3463) | (4.6598) | (3.9422) | (3.7080) |
| Different Coefficient? | – | – | – | – | – |
| — Share Trading Volume — | | | | | |
| above median | -15.5000*** | -12.5071*** | -14.3684*** | -6.8560**+ | -7.5999*** |
| | (4.8567) | (3.7086) | (4.0492) | (2.9331) | (2.6160) |
| below median | -8.7777* | -6.2767 | -5.8707 | -3.6713 | -2.4701 |
| | (4.7358) | (4.1754) | (4.4619) | (3.6929) | (3.4007) |
| Different Coefficient? | – | – | – | – | – |

**Table 5**
**Panel Regressions for Depth at the Best Bid and Offer Prices**

Dependent variables are treatment group value minus control group value for the trade weighted and time weighted depth. Depth is measured in the log of the number of shares and the log of the dollar amount. Specifications for the panel regression and significance levels are as in Table 4.

| | share depth throughout the day | share depth at transaction | $ depth throughout the day | $ depth at transaction |
|---|---|---|---|---|
| full sample | 0.0898**+ | 0.0837**+ | 0.1133*** | 0.1070*** |
| | (0.0369) | (0.0360) | (0.0394) | (0.0384) |
| | — Break-even Price for Market Orders — | | | |
| above $22 | 0.0992 | 0.0688 | 0.0414 | 0.0111 |
| | (0.0777) | (0.0633) | (0.0985) | (0.0838) |
| below $22 | 0.0884** | 0.0861** | 0.1247*** | 0.1223*** |
| | (0.0402) | (0.0398) | (0.0419) | (0.0414) |
| Different Coefficient? | – | – | – | – |
| | — Herfindahl Index — | | | |
| low competition | 0.0562 | 0.0397 | 0.0865 | 0.0695 |
| | (0.0498) | (0.0501) | (0.0533) | (0.0526) |
| high competition | 0.1226**+ | 0.1263*** | 0.1393*** | 0.1434*** |
| | (0.0489) | (0.0458) | (0.0522) | (0.0496) |
| Different Coefficient? | – | – | – | – |
| | — Market Capitalization — | | | |
| above median | 0.1294*** | 0.1229*** | 0.1304**+ | 0.1241*** |
| | (0.0482) | (0.0450) | (0.0511) | (0.0476) |
| below median | 0.0513 | 0.0454 | 0.0967* | 0.0903 |
| | (0.0512) | (0.0514) | (0.0551) | (0.0549) |
| Different Coefficient? | – | – | – | – |
| | —Break-even Price for Total Fees — | | | |
| above $6.875 | 0.1445*** | 0.1339*** | 0.1587*** | 0.1483*** |
| | (0.0481) | (0.0439) | (0.0505) | (0.0464) |
| below $6.875 | 0.0448 | 0.0422 | 0.0759 | 0.073 |
| | (0.0495) | (0.0505) | (0.0536) | (0.0539) |
| Different Coefficient? | – | – | – | – |
| | — Share Trading Volume — | | | |
| above median | 0.0559 | 0.0563 | 0.0982* | 0.0985* |
| | (0.0507) | (0.0493) | (0.0523) | (0.0511) |
| below median | 0.1228**+ | 0.1103** | 0.1280**+ | 0.1153** |
| | (0.0484) | (0.0479) | (0.0534) | (0.0520) |
| Different Coefficient? | – | – | – | – |

## Table 6
## Panel Regressions for Market Efficiency Measures

Dependent variables are treatment group value minus control group value for the $x$-minute autocorrelation and $x$-/$y$-minute variance ratios. Details on these measures are in Section 3. Specifications for the panel regression and significance levels are as in Table 4.

| | 5-minute autocorrelation | 15-minute autocorrelation | 30-minute autocorrelation | 5/30-minute variance ratio | 15/30-minute variance ratio |
|---|---|---|---|---|---|
| full sample | 0.0026 | 0.0061 | 0.0018 | -0.0062 | 0.0083 |
| | (0.0068) | (0.0082) | (0.0098) | (0.0081) | (0.0082) |
| | | — Break-even Price for Market Orders — | | | |
| above $22 | -0.0214 | -0.0145 | -0.0014 | -0.0118 | -0.0157 |
| | (0.0155) | (0.0187) | (0.0200) | (0.0166) | (0.0243) |
| below $22 | 0.0064 | 0.0095 | 0.0023 | -0.0053 | 0.0122 |
| | (0.0072) | (0.0085) | (0.0109) | (0.0093) | (0.0089) |
| Different Coefficient? | – | – | – | – | – |
| | | — Herfindahl Index — | | | |
| low competition | 0.0017 | 0.0162 | 0.0038 | -0.0082 | -0.0032 |
| | (0.0101) | (0.0102) | (0.0145) | (0.0146) | (0.0117) |
| high competition | 0.0035 | -0.0032 | -0.0000 | -0.0042 | 0.0190* |
| | (0.0088) | (0.0123) | (0.0132) | (0.0104) | (0.0098) |
| Different Coefficient? | – | – | – | – | – |
| | | — Market Capitalization — | | | |
| above median | 0.0048 | -0.0136 | 0.0006 | 0.0007 | 0.0104 |
| | (0.0090) | (0.0115) | (0.0132) | (0.0137) | (0.0108) |
| below median | 0.0004 | 0.0262*** | 0.0031 | -0.0132 | 0.0061 |
| | (0.0106) | (0.0093) | (0.0149) | (0.0112) | (0.0122) |
| Different Coefficient? | – | – | Yes*** | – | – |
| | | —Break-even Price for Total Fees — | | | |
| above $6.875 | 0.0091 | -0.0101 | 0.0035 | 0.0088 | 0.0159 |
| | (0.0094) | (0.0115) | (0.0150) | (0.0130) | (0.0113) |
| below $6.875 | -0.0029 | 0.0201** | 0.0004 | -0.0190* | 0.0018 |
| | (0.0089) | (0.0103) | (0.0124) | (0.0107) | (0.0116) |
| Different Coefficient? | – | – | – | – | – |
| | | — Share Trading Volume — | | | |
| above median | 0.0026 | 0.0016 | 0.0032 | -0.0105 | 0.0070 |
| | (0.0077) | (0.0123) | (0.0162) | (0.0108) | (0.0092) |
| below median | 0.0027 | 0.0107 | 0.0005 | -0.0019 | 0.0096 |
| | (0.0106) | (0.0107) | (0.0113) | (0.0135) | (0.0126) |
| Different Coefficient? | – | – | – | – | – |

## Table 7
## Panel Regressions for Transaction Costs and Rebate Benefits

Dependent variables are treatment group value minus control group value for proportional effective spreads, adjusted for active order exchange fees, and realized 5 minute spreads, adjusted for exchange fee rebates as described in (7) and (8). Costs and benefits are measured in basis points of the prevailing midquote. Specifications for the panel regression and significance levels are as in Table 4.

| | exchange fee adjusted effective spreads | rebate adjusted realized 5 minute spreads |
|---|---|---|
| full sample | 5.6538* | 8.0544*** |
| | (3.3209) | (2.5238) |
| | — Break-even Price for Market Orders — | |
| above $22 | -2.4563 | -0.2473 |
| | (4.1147) | (2.6138) |
| below $22 | 6.9440* | 9.3770*** |
| | (3.7366) | (2.8674) |
| Different Coefficient? | Yes* | Yes**+ |
| | — Herfindahl Index — | |
| low competition | 9.8563* | 13.1860*** |
| | (5.4822) | (4.3179) |
| high competition | 1.5773 | 3.0968 |
| | (3.8412) | (2.7434) |
| Different Coefficient? | – | Yes** |
| | — Market Capitalization — | |
| above median | 2.2950 | 4.8913*** |
| | (3.1257) | (1.5057) |
| below median | 8.9320 | 11.1541**+ |
| | (5.7850) | (4.7828) |
| Different Coefficient? | – | – |
| | —Break-even Price for Total Fees — | |
| above $6.875 | -4.411 | -1.345 |
| | (3.2416) | (1.9756) |
| below $6.875 | 13.9760*** | 15.8351*** |
| | (5.0483) | (4.0637) |
| Different Coefficient? | Yes*** | Yes*** |
| | — Share Trading Volume — | |
| above median | 1.7195 | 6.6988** |
| | (4.1914) | (2.9735) |
| below median | 9.4669* | 9.3517** |
| | (5.1549) | (4.1513) |
| Different Coefficient? | – | – |

## Table 8
## Panel Regressions for Volume and Transactions

Dependent variables are treatment group value minus control group value for the logarithms of share volume, dollar volume and the number of transactions. Note that an incoming active order can trigger multiple transactions. Specifications for the panel regression and significance levels are as in Table 4.

| | volume in shares | dollar volume | trans-actions |
|---|---|---|---|
| full sample | 0.1709** | 0.1945**+ | 0.20*** |
| | (0.0752) | (0.0821) | (0.06) |
| | — Break-even Price for Market Orders — | | |
| above $22 | -0.0136 | -0.0719 | 0.06 |
| | (0.1854) | (0.1925) | (0.12) |
| below $22 | 0.2003**+ | 0.2369*** | 0.22*** |
| | (0.0806) | (0.0879) | (0.07) |
| Different Coefficient? | – | – | – |
| | — Herfindahl Index — | | |
| low competition | 0.0929 | 0.1228 | 0.12 |
| | (0.1052) | (0.1135) | (0.08) |
| high competition | 0.2466**+ | 0.2640**+ | 0.27*** |
| | (0.0988) | (0.1088) | (0.09) |
| Different Coefficient? | – | – | – |
| | — Market Capitalization — | | |
| above median | 0.1285** | 0.1297* | 0.17*** |
| | (0.0629) | (0.0685) | (0.05) |
| below median | 0.2124 | 0.2577* | 0.22** |
| | (0.1296) | (0.1402) | (0.11) |
| Different Coefficient? | – | – | – |
| | —Break-even Price for Total Fees — | | |
| above $6.875 | 0.1871* | 0.2013* | 0.20** |
| | (0.1134) | (0.1168) | (0.1000) |
| below $6.875 | 0.1576 | 0.1890* | 0.20**+ |
| | (0.0979) | (0.1106) | (0.0800) |
| Different Coefficient? | – | – | – |
| | — Share Trading Volume — | | |
| above median | 0.0830 | 0.1255 | 0.16** |
| | (0.0825) | (0.0930) | (0.08) |
| below median | 0.2574** | 0.2626** | 0.23**+ |
| | (0.1184) | (0.1274) | (0.09) |
| Different Coefficient? | – | – | – |

## Table 9
## Panel Regressions for Total Volume by Trader Type

Dependent variables are treatment group value minus control group value for the logarithms of share volume, dollar volume and transactions, split by client and non-client orders. Both the active and passive sides of a trade are counted, and thus the sum of client and non-client volume is twice the daily volume. Specifications for the panel regression and significance levels are as in Table 4.

| | share volume | | dollar volume | | transactions | |
|---|---|---|---|---|---|---|
| | non-client | client | non-client | client | non-client | client |
| full sample | 0.2007*** | 0.1511* | 0.2245*** | 0.1745** | 0.2160*** | 0.1716**+ |
| | (0.0769) | (0.0781) | (0.0827) | (0.0850) | (0.0624) | (0.0693) |
| | — Break-even Price for Market Orders — | | | | | |
| above $22 | 0.0210 | -0.0358 | -0.0341 | -0.0939 | 0.1164 | 0.0089 |
| | (0.1434) | (0.2084) | (0.1436) | (0.2151) | (0.1176) | (0.1473) |
| below $22 | 0.2286*** | 0.1808** | 0.2648*** | 0.2172**+ | 0.2317*** | 0.1974*** |
| | (0.0855) | (0.0825) | (0.0919) | (0.0899) | (0.0693) | (0.0745) |
| Different Coefficient? | Yes*** | – | Yes* | – | Yes** | – |
| | — Herfindahl Index — | | | | | |
| low competition | 0.1050 | 0.0809 | 0.1369 | 0.1106 | 0.1738** | 0.0710 |
| | (0.1027) | (0.1109) | (0.1107) | (0.1191) | (0.0823) | (0.0910) |
| high competition | 0.2929*** | 0.2191** | 0.3088*** | 0.2365** | 0.2566*** | 0.2690*** |
| | (0.1064) | (0.1003) | (0.1141) | (0.1105) | (0.0876) | (0.0918) |
| Different Coefficient? | – | – | – | – | – | – |
| | — Market Capitalization — | | | | | |
| above median | 0.1327** | 0.1137 | 0.1336** | 0.1150 | 0.1505*** | 0.1623*** |
| | (0.0608) | (0.0716) | (0.0638) | (0.0776) | (0.0518) | (0.0629) |
| below median | 0.2678** | 0.1876 | 0.3141** | 0.2326 | 0.2805*** | 0.1806 |
| | (0.1341) | (0.1316) | (0.1437) | (0.1422) | (0.1061) | (0.1143) |
| Different Coefficient? | – | – | – | – | – | – |
| | —Break-even Price for Total Fees — | | | | | |
| above $6.875 | 0.2100* | 0.1644 | 0.2252** | 0.1783 | 0.2135** | 0.1682 |
| | (0.1135) | (0.1170) | (0.1122) | (0.1217) | (0.0942) | (0.1034) |
| below $6.875 | 0.1930* | 0.1401 | 0.2239** | 0.1715 | 0.2181*** | 0.1742** |
| | (0.1004) | (0.1019) | (0.1142) | (0.1137) | (0.0791) | (0.0876) |
| Different Coefficient? | – | – | – | – | – | – |
| | — Share Trading Volume — | | | | | |
| above median | 0.0632 | 0.0911 | 0.1049 | 0.1336 | 0.1215* | 0.1736* |
| | (0.0805) | (0.0901) | (0.0894) | (0.1005) | (0.0704) | (0.0886) |
| below median | 0.3364*** | 0.2105* | 0.3429*** | 0.2156* | 0.3093*** | 0.1704* |
| | (0.1222) | (0.1203) | (0.1306) | (0.1291) | (0.0966) | (0.0978) |
| Different Coefficient? | Yes* | – | – | – | – | – |

## Table 10
## Panel Regressions on the Fraction of Intermediated Trades

Dependent variables are treatment group value minus control group value for the client to non-client fraction of total volume. Specifications for the panel regression and significance levels are as in Table 4.

| | share volume | dollar volume | transactions |
|---|---|---|---|
| full sample | 1.070 | 1.070 | 0.690 |
| | (0.980) | (0.980) | (0.850) |
| | — Break-even Price for Market Orders — | | |
| above $22 | 0.910 | 0.910 | 0.930 |
| | (1.930) | (1.920) | (2.150) |
| below $22 | 1.100 | 1.100 | 0.650 |
| | (1.070) | (1.070) | (0.880) |
| Different Coefficient? | – | – | – |
| | — Herfindahl Index — | | |
| low competition | 0.970 | 0.970 | 1.780 |
| | (1.440) | (1.440) | (1.200) |
| high competition | 1.170 | 1.180 | -0.370 |
| | (1.190) | (1.190) | (1.020) |
| Different Coefficient? | – | – | – |
| | — Market Capitalization — | | |
| | 0.000 | 0.000 | -0.730 |
| above median | (1.050) | (1.050) | (1.060) |
| | 2.120 | 2.120 | 2.080* |
| below median | (1.580) | (1.590) | (1.180) |
| Different Coefficient? | – | – | Yes* |
| | —Break-even Price for Total Fees — | | |
| above $6.875 | 0.0126 | 0.0127 | 0.0042 |
| | (0.0112) | (0.0112) | (0.0119) |
| below $6.875 | 0.0091 | 0.0091 | 0.0091 |
| | (0.0148) | (0.0148) | (0.0109) |
| Different Coefficient? | – | – | – |
| | — Share Trading Volume — | | |
| above median | -0.230 | -0.230 | -0.550 |
| | (1.290) | (1.290) | (0.950) |
| below median | 2.340* | 2.330* | 1.890 |
| | (1.300) | (1.300) | (1.230) |
| Different Coefficient? | – | – | Yes* |

## Table 11
## Panel Regressions on Competition Indicators

Dependent variables are treatment group value minus control group value for the trader level Herfindahl Index, the number of liquidity providing brokers and the number of liquidity providing traders that trade on inventory accounts. The Herfindahl Index is defined in (2), the number of brokers is the number of broker IDs that are on the passive side of trades, the number of inventory traders is the number of trader IDs that are on the passive side of trades while using their inventory account. All measures are per stock per day. A decrease in the Herfindahl Index indicates a decrease in market concentration and thus an increase in competition for liquidity provision. Specifications for the panel regression and significance levels are as in Table 4.

| | trader level Herfindahl Index | number of brokers | number of inventory traders |
|---|---|---|---|
| full sample | -0.0350**+ | 0.6903** | 0.5007* |
| | (0.0144) | (0.3410) | (0.2617) |
| | — Break-even Price for Market Orders — | | |
| above $22 | 0.0325 | -0.6283 | 0.1746 |
| | (0.0362) | (0.5968) | (0.6439) |
| below $22 | -0.0459*** | 0.8996**+ | 0.5525* |
| | (0.0144) | (0.3717) | (0.2886) |
| Different Coefficient? | Yes** | Yes** | – |
| | — Herfindahl Index — | | |
| low competition | -0.0305* | 0.6522 | 0.2476 |
| | (0.0180) | (0.4214) | (0.2154) |
| high competition | -0.0392* | 0.7274 | 0.7470* |
| | (0.0209) | (0.5009) | (0.4404) |
| Different Coefficient? | – | – | – |
| | — Market Capitalization — | | |
| above median | -0.0114 | 0.0573 | 0.4874 |
| | (0.0163) | (0.3202) | (0.4215) |
| below median | -0.0588*** | 1.3062**+ | 0.5137* |
| | (0.0212) | (0.5500) | (0.2657) |
| Different Coefficient? | Yes** | Yes** | – |
| | —Break-even Price for Total Fees — | | |
| above $6.875 | -0.0146 | 0.2103 | 0.1426 |
| | (0.0217) | (0.5037) | (0.3695) |
| below $6.875 | -0.0521*** | 1.0863**+ | 0.7962** |
| | (0.0168) | (0.4351) | (0.3523) |
| Different Coefficient? | – | – | – |
| | — Share Trading Volume — | | |
| above median | -0.0334* | 0.3055 | 0.4937 |
| | (0.0187) | (0.4320) | (0.4318) |
| below median | -0.0367* | 1.0647** | 0.5075* |
| | (0.0204) | (0.4944) | (0.2698) |
| Different Coefficient? | – | – | – |

## Table 12
## Panel Regressions for Improvements in the Best Bid and Offer

Dependent variables are treatment group value minus control group value for the total number of improvements at the best bid and offer (BBO) as well as its decomposition into the number of improvements with regards to prices and depth. Specifically, the number of improvements in the BBO is computed, for each stock and day, by counting the number of times that there is an increase in the number of shares available at the bid or offer for a fixed or an improved prices and the number of times that the bid is increased or the offer decreased. Specifications for the panel regression and significance levels are as in Table 4.

| | Number of BBO improvements | spread improvements | depth improvements | Number of BBO changes |
|---|---|---|---|---|
| full sample | 102.2**+ | -54.3*** | 156.5*** | 236.3*** |
| | (41.2) | (9.8) | (47.3) | (58.0) |
| | | — Break-even Price for Market Orders — | | |
| above $22 | 76.5 | -179.9** | 256.4 | 127.0 |
| | (230.1) | (78.6) | (220.0) | (285.7) |
| below $22 | 106.3*** | -34.4*** | 140.7*** | 253.6*** |
| | (41.2) | (9.1) | (39.5) | (82.8) |
| Different Coefficient? | – | Yes* | – | – |
| | | — Herfindahl Index — | | |
| low competition | -4.2 | -48.9*** | 44.7**+ | 31.1 |
| | (28.4) | (17.1) | (18.9) | (50.7) |
| high competition | 205.8*** | -59.5*** | 265.3*** | 435.9*** |
| | (72.0) | (15.2) | (82.9) | (99.4) |
| Different Coefficient? | Yes*** | – | Yes*** | Yes*** |
| | | — Market Capitalization — | | |
| above median | 189.2*** | -71.6*** | 260.8*** | 406.2*** |
| | (73.1) | (16.3) | (83.8) | (97.5) |
| below median | 17.6 | -37.5** | 55.1** | 71.0 |
| | (37.3) | (17.0) | (27.5) | (72.4) |
| Different Coefficient? | Yes** | – | Yes**+ | Yes*** |
| | | —Break-even Price for Total Fees — | | |
| above $6.875 | 170.1* | -90.8*** | 260.9*** | 393.1*** |
| | (89.5) | (24.3) | (89.2) | (133.1) |
| below $6.875 | 46.2** | -24.2*** | 70.4*** | 106.9** |
| | (21.6) | (4.0) | (26.2) | (52.7) |
| Different Coefficient? | – | Yes**+ | Yes** | – |
| | | — Share Trading Volume — | | |
| above median | 208.3*** | -38.2** | 246.5*** | 383.2*** |
| | (73.2) | (16.6) | (78.8) | (100.4) |
| below median | -1 | -70.0*** | 69.0* | 93.4 |
| | (53.4) | (23.8) | (37.5) | (94.7) |
| Different Coefficient? | Yes**+ | – | Yes** | Yes** |

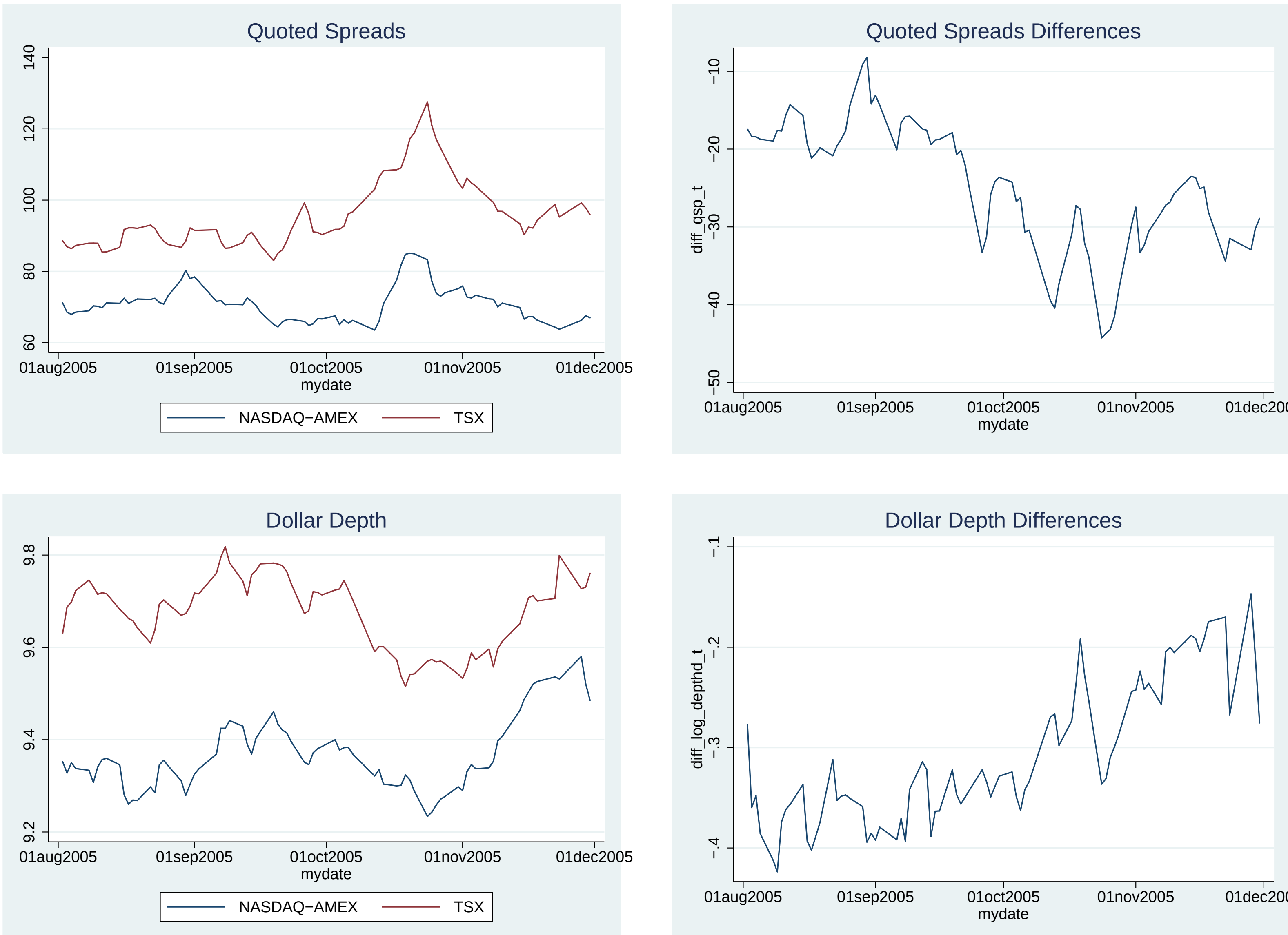
## Table 13
## Panel Regressions on the Equity Specialist's Trading Activity

Dependent variables are treatment group value minus control group value for measures of trading activity of the equity specialist (registered trader): the total active and passive share volume, dollar volume, and the number of transactions. Volume is in logarithms. Specifications for the panel regression and significance levels are as in Table 4.

| | share volume | | dollar volume | | transactions | |
|---|---|---|---|---|---|---|
| | passive | active | passive | active | passive | active |
| full sample | 0.1997**+ | 0.0791 | 0.1884*** | 0.0450 | 5.68**+ | 2.51** |
| | (0.0835) | (0.0938) | (0.0606) | (0.0751) | (2.42) | (1.27) |
| | | | — Break-even Price for Market Orders — | | | |
| above $22 | 0.0294 | 0.0610 | 0.0967 | -0.0388 | 7.08 | 5.74 |
| | (0.2002) | (0.1481) | (0.0858) | (0.0904) | (10.46) | (5.71) |
| below $22 | 0.2255*** | 0.0976 | 0.2020*** | 0.0726 | 5.45** | 2.00 |
| | (0.0869) | (0.1108) | (0.0674) | (0.0902) | (2.36) | (1.30) |
| Different Coefficient? | – | – | – | – | – | – |
| | | | — Herfindahl Index — | | | |
| low competition | 0.0641 | -0.0436 | 0.1044 | -0.0815 | 0.36 | -0.02 |
| | (0.0852) | (0.1201) | (0.0726) | (0.1133) | (2.71) | (0.94) |
| high competition | 0.3121**+ | 0.1675 | 0.2588*** | 0.1315 | 10.83*** | 4.98** |
| | (0.1230) | (0.1213) | (0.0847) | (0.0895) | (3.73) | (2.37) |
| Different Coefficient? | Yes* | – | – | – | Yes** | Yes* |
| | | | — Market Capitalization — | | | |
| above median | 0.1253 | -0.0028 | 0.1469**+ | -0.0317 | 6.96** | 3.44 |
| | (0.0917) | (0.0754) | (0.0601) | (0.0552) | (3.50) | (2.25) |
| below median | 0.2763** | 0.1907 | 0.2296** | 0.1533 | 4.42 | 1.62 |
| | (0.1297) | (0.1863) | (0.0999) | (0.1539) | (3.24) | (1.38) |
| Different Coefficient? | – | – | – | – | – | – |
| | | | — Share Trading Volume — | | | |
| above median | 0.1036 | 0.0229 | 0.1042 | -0.0180 | 7.74* | 4.91** |
| | (0.1028) | (0.1134) | (0.0792) | (0.0838) | (4.13) | (2.26) |
| below median | 0.3019**+ | 0.1739 | 0.2766*** | 0.1407 | 3.69* | 0.17 |
| | (0.1179) | (0.1408) | (0.0819) | (0.1217) | (2.19) | (1.20) |
| Different Coefficient? | – | – | – | – | – | Yes* |

**Figure 1**
**Quoted Liquidity: Spreads and Depth**

The top left panel plots the time-weighted quoted spreads for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The bottom left panel plots depth at the best bid and offer prices. The top and bottom right panels plot the differences of, respectively, quoted spreads and depth for interlisted securities vs. their non-interlisted matches. All plots are 5-day moving averages. Spreads are measured in basis points of the midpoint, depth is measured in the logarithm of the average dollar amount available for trading at the best bid and offer prices.



## Figure 2
## Effective Liquidity: Price Impacts and Effective Spreads

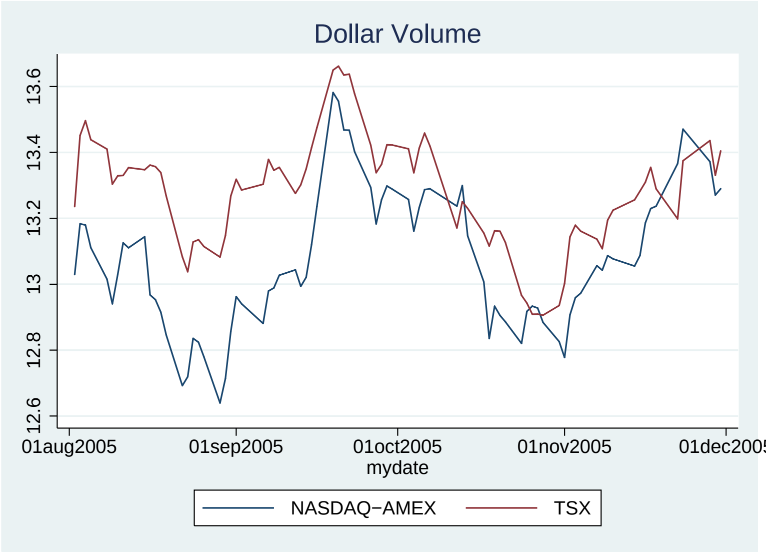
The left panel plots the trade-weighted effective spread for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The bottom left panel plots the trade-weighted 5-minute price impact. The top and bottom right panels plot the differences of, respectively, effective spreads and price impact for interlisted securities vs. their non-interlisted matches. All plots are 5-day moving averages. Spreads and price impact are measured in basis points of the midpoint.

## Figure 3
## Plots of Trade Execution Costs for Active Orders and Benefits for Passive Orders

The left panel plots the trade-weighted exchange fee adjusted effective spread for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The bottom left panel plots the trade-weighted 5-minute rebate adjusted realized spread. The top and bottom right panels plot the differences of, respectively, adjusted effective and realized spreads for interlisted securities vs. their non-interlisted matches. All plots are 5-day moving averages. Spreads are measured in basis points of the midpoint.

**Figure 4**
**Plots of Dollar Volume**

The left panel plots the average daily intra-day dollar volume (all trades against standing orders in the limit order book) for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The right panel plots the differences of the average dollar volume for interlisted securities vs. their non-interlisted matches. All plots are 5-day moving averages. Dollar volume is in logarithm.



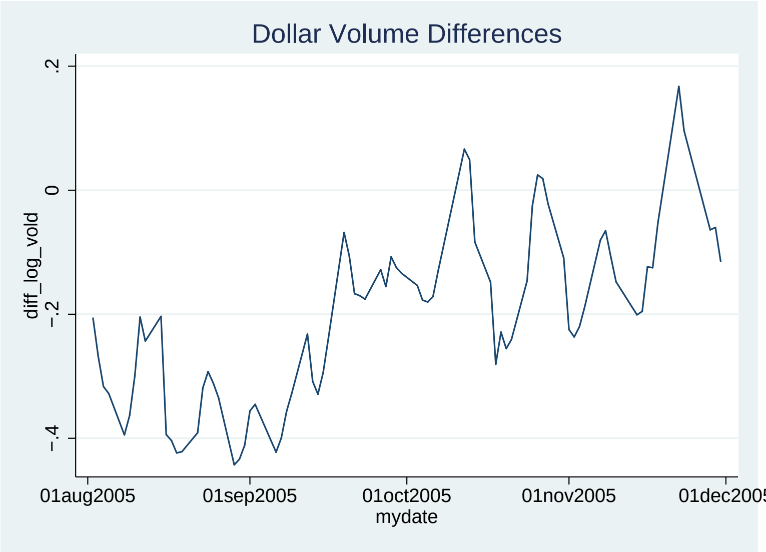



**Figure 5**
**Plots of the Herfindahl Index**

The left panel plots the average of the per day per stock trader level Herfindahl Index (see Section 2.4) for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The right panel plots the differences of the trader level HHIs for interlisted securities vs. their non-interlisted matches. All plots are 5-day moving averages.

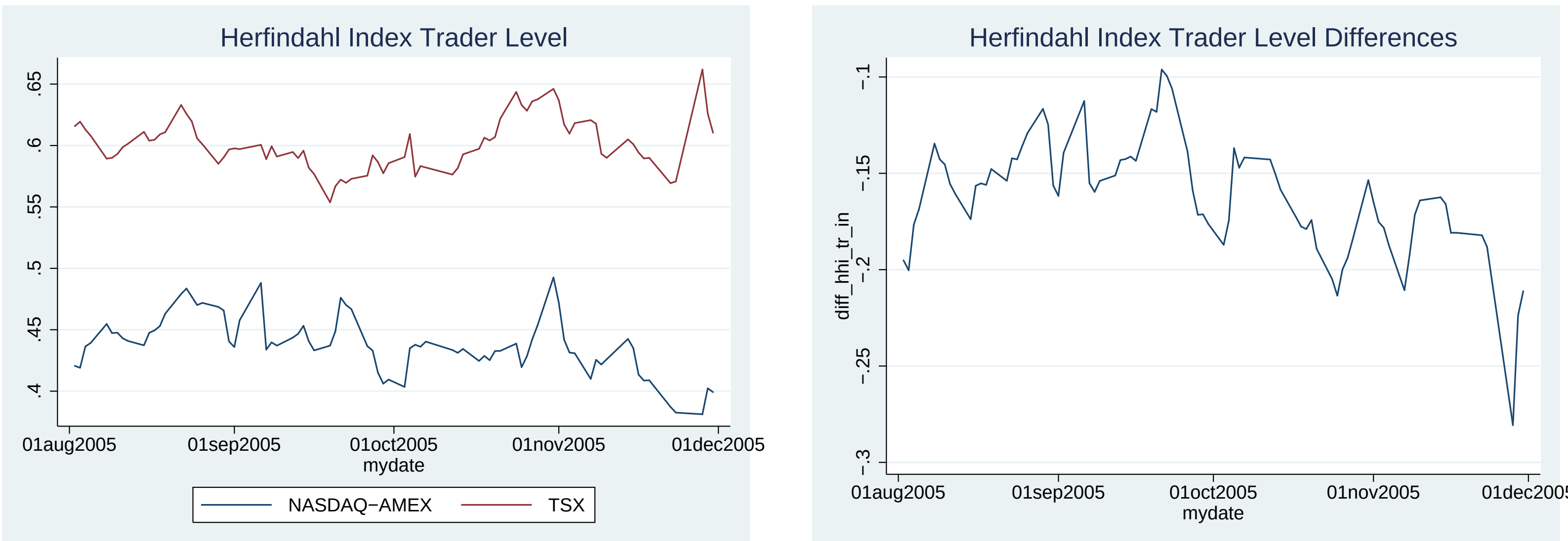
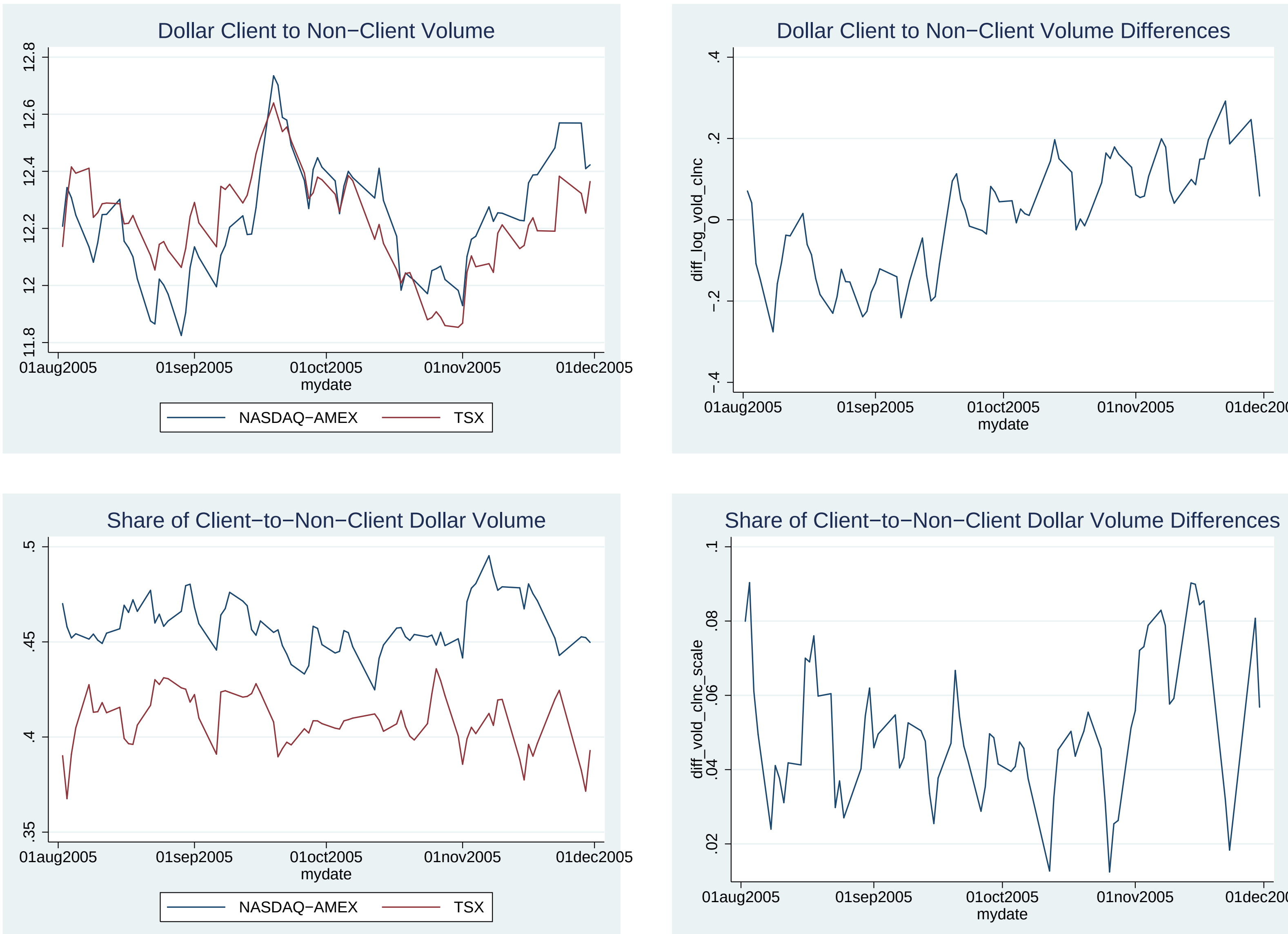


**Figure 6**
**Plots of Intermediated Dollar Volume**

The top left panel plots the daily dollar volume of trades between clients and non-clients for the group of NASDAQ/AMEX interlisted securities and their matches (labelled as "TSX"). The bottom left panel plots the fraction of such intermediated trades of the total dollar volume. The top and bottom right panels plot the differences of, respectively, levels and fractions of intermediated dollar volume for interlisted securities vs. their non-interlisted matches.



## Table 14
## List of all interlisted companies and their non-interlisted matches, Part I

| Treatment Group: Interlisted with AMEX or NASDAQ | | Control group match: non-interlisted | |
|---|---|---|---|
| ABZ | ABER DIAMOND CORPORATION | SBY | SOBEYS INC. |
| AEZ | AETERNA ZENTARIS INC. | ITX | ITERATION ENERGY LTD. J |
| ANP | ANGIOTECH PHARMACEUTICALS INC. | AGF.NV | AGF MANAGEMENT LTD. CL 'B' NV |
| ARZ | AURIZON MINES LTD. J | ENE | ENDEV ENERGY INC. |
| ATY | ATI TECHNOLOGIES INCORPORATED | TA | TRANSALTA CORPORATION |
| AXP | AXCAN PHARMA INC. | IMN | INMET MINING CORPORATION |
| BEV | BENNETT ENVIRONMENTAL INC. | STY | STYLUS ENERGY INC. |
| BGO | BEMA GOLD CORPORATION J | UTS | UTS ENERGY CORPORATION |
| BLD | BALLARD POWER SYSTEMS INC. | IUC | INTERNATIONAL URANIUM CORPORATION J |
| BRA | BIOMIRA INC. | CEK | CASPIAN ENERGY INC. J |
| CBJ | CAMBIOR INC. | NS | NORSKE SKOG CANADA LIMITED |
| CEF.NV.A | CENTRAL FUND OF CANADA LTD. CL 'A' NV | SWP | SASKATCHEWAN WHEAT POOL INC. |
| CLG | CUMBERLAND RESOURCES LTD. J | ANO | ANATOLIA MINERALS DEVELOPMENT LIMITED J |
| COM | CARDIOME PHARMA CORP. | KEC | KICK ENERGY CORPORATION J |
| CRY | CRYPTOLOGIC INC. | AAH | AASTRA TECHNOLOGIES LIMITED |
| CSN | COGNOS INC. | CTR.NV | CANADIAN TIRE CORP. LTD. CL 'A' NV |
| DAX | DRAXIS HEALTH INC. | IXL | INNOVA EXPLORATION LTD. J |
| DII.SV | DOREL INDUSTRIES INC. CL 'B' SV | AGA | ALGOMA STEEL INC. |
| DSG | DESCARTES SYSTEMS GROUP INC. (THE) | GWE | GREY WOLF EXPLORATION INC. |
| DSM | DESERT SUN MINING CORP. J | ARG | AMERIGO RESOURCES LTD. J |
| ECG | ENVOY COMMUNICATIONS GROUP INC. | EDV | ENDEAVOUR MINING CAPITAL CORP. ORDINARY J |
| ELD | ELDORADO GOLD CORPORATION | BBD.MV.A | BOMBARDIER INC. CL 'A' MV |
| EXF.SV | EXFO ELECTRO-OPTICAL ENGINEERING INC. SV | QUA | QUADRA MINING LTD. |
| FMI | FORBES MEDI-TECH INC. | WF | WHITE FIRE ENERGY LTD. |
| FNX | FNX MINING COMPANY INC. | ATA | ATS AUTOMATION TOOLING SYSTEMS INC. |
| FRG | FRONTEER DEVELOPMENT GROUP INC. J | CSY | CSI WIRELESS INC. |
| FSV.SV | FIRSTSERVICE CORPORATION SV | CCL.NV.B | CCL INDUSTRIES INC. CL 'B' NV |
| GAC | GEAC COMPUTER CORPORATION LTD. | HBC | HUDSON'S BAY COMPANY |
| GAM | GAMMON LAKE RESOURCES INC. J | FAP | ABERDEEN ASIA-PACIFIC INCM INVESTMENT CO LTD. |
| GSC | GOLDEN STAR RESOURCES LTD. | OIL | OILEXCO INCORPORATED J |
| HUM | HUMMINGBIRD LTD. | MRG | MERGE CEDARA EXCHANGECO LIMITED EXCHANGEABLE |
| HYG | HYDROGENICS CORPORATION | SGF | SHORE GOLD INC. J |
| IDB | ID BIOMEDICAL CORPORATION | KFS | KINGSWAY FINANCIAL SERVICES INC. |
| IE | IVANHOE ENERGY INC. | UEX | UEX CORPORATION J |
| IMG | IAMGOLD CORPORATION | LIM | LIONORE MINING INTERNATIONAL LTD. |
| IMO | IMPERIAL OIL LTD. | RY | ROYAL BANK OF CANADA |

## Table 15
## List of all interlisted companies and their non-interlisted matches, Part II

| Treatment Group: Interlisted with AMEX or NASDAQ | | Control group match: non-interlisted | |
|---|---|---|---|
| IMX | IMAX CORPORATION | GND | GENNUM CORPORATION |
| IOL | INTEROIL CORPORATION J | CCA.SV | COGECO CABLE INC. SV |
| KRY | CRYSTALLEX INTERNATIONAL CORPORATION J | TBC | TEMBEC INC. |
| MAE | MIRAMAR MINING CORPORATION | IVW | IVERNIA INC. J |
| MEC.SV.A | MAGNA ENTERTAINMENT CORP. CL 'A' SV | ITP | INTERTAPE POLYMER GROUP INC. |
| MFL | MINEFINDERS CORPORATION LTD. J | CYT | CRYOCATH TECHNOLOGIES INC. |
| MPV | MOUNTAIN PROVINCE DIAMONDS INC. J | COB.SV.A | COOLBRANDS INTERNATIONAL INC. CL 'A' SV |
| MR | METALLICA RESOURCES INC. J | ACA | ASHTON MINING OF CANADA INC. |
| MX | METHANEX CORPORATION | MNG | MERIDIAN GOLD INC. |
| NG | NOVAGOLD RESOURCES INC. J | PTI | PATHEON INC. |
| NGX | NORTHGATE MINERALS CORPORATION | DY | DYNATEC CORPORATION |
| NNO | NORTHERN ORION RESOURCES INC. J | TRE | SINO-FOREST CORPORATION |
| NRM | NEUROCHEM INC. | SWG | SOUTHWESTERN RESOURCES CORP. J |
| NSU | NEVSUN RESOURCES LTD. J | CDV | COM DEV INTERNATIONAL LTD. |
| ONC | ONCOLYTICS BIOTECH INC. | CNH | CINCH ENERGY CORP. J |
| OTC | OPEN TEXT CORPORATION | RUS | RUSSEL METALS INC. |
| OZN | OREZONE RESOURCES INC. J | ZL | ZARLINK SEMICONDUCTOR INC. |
| PAA | PAN AMERICAN SILVER CORP. | CRW | CINRAM INTERNATIONAL INC. |
| PCR | PERU COPPER INC. J | SMF | SEMAFO INC. J |
| PDL | NORTH AMERICAN PALLADIUM LTD. | IFP.SV.A | INTERNATIONAL FOREST PRODUCTS LTD. CL 'A' SV |
| QLT | QLT INC. | BVI | BLACKROCK VENTURES INC. |
| RIM | RESEARCH IN MOTION LIMITED | WN | WESTON LTD. GEORGE |
| RNG | RIO NARCEA GOLD MINES LTD. | MAL | MAGELLAN AEROSPACE CORPORATION |
| SNG | CANADIAN SUPERIOR ENERGY INC. J | BGC | BOLIVAR GOLD CORP. J |
| SOY | SUNOPTA, INC. | SGB | STRATOS GLOBAL CORPORATION |
| SSO | SILVER STANDARD RESOURCES INC. | RRZ | RIDER RESOURCES LTD. |
| SVN | 724 SOLUTIONS INC. | RVE | ROCKYVIEW ENERGY INC. |
| SW | SIERRA WIRELESS, INC. | FE | FIND ENERGY LTD. |
| TEO | TESCO CORPORATION | KCO | KERECO ENERGY LTD. |
| TGL | TRANSGLOBE ENERGY CORPORATION J | WLE | WESTERN LAKOTA ENERGY SERVICES INC. |
| TLC | TLC VISION CORPORATION | CGS.SV | CANWEST GLOBAL COMMUNICATIONS CORP. SV |
| TNX | TAN RANGE EXPLORATION CORPORATION J | WPT | WESTPORT INNOVATIONS INC. |
| VAS | VASOGEN INC. | VIA | VIRGINIA GOLD MINES INC. J |
| WED | WESTAIM CORPORATION (THE) | WTN | WESTERN CANADIAN COAL CORP. J |
| YM | YM BIOSCIENCES INC. J | DDS | LABOPHARM INC. |
| YRI | YAMANA GOLD INC. J | AGI | ALAMOS GOLD INC. J |
| ZIC | ZI CORPORATION | TOS | TSO3 INC. J |

# Notes on Bonds: Liquidity at all Costs in the Great Recession

David Musto
Greg Nini
Krista Schwarz *

April 27, 2011

VERY PRELIMINARY AND INCOMPLETE

**Abstract:** We address the connection between market stress and asset pricing by analyzing a large and systematic discrepancy arising among off-the-run Treasuries. We first show that bonds traded for much less than notes with matching maturity and coupon, over five percent less in December 2008. We then ask how the small differences between these securities, in particular their liquidity, could project to so large a gap. We gauge the potential for long/short arbitrage with repo and fails data indicating the frictions arbitrageurs encountered, and then with daily transactions data we relate the demand for the expensive but liquid note to the cross section of insurers' liquidity needs.

*The Wharton School, University of Pennsylvania, 3620 Locust Walk, Philadelphia PA 19104. Department of Finance: David Musto, (215) 898-2323, musto@wharton.upenn.edu. Department of Risk & Insurance: Greg Nini, greg30@wharton.upenn.edu, (215) 898-0310. Department of Finance: Krista Schwarz, (215) 898-6087, kschwarz@wharton.upenn.edu. We are grateful to the Dean's Research Fund of the Wharton School data funding. We are particularly grateful to Michael Bulboff, who provided outstanding research assistance. All remaining errors are our own.

# Notes on Bonds: Liquidity at all Costs in the Great Recession

**Abstract:** We address the connection between market stress and asset pricing by analyzing a large and systematic discrepancy arising among off-the-run Treasuries. We first show that bonds traded for much less than notes with matching maturity and coupon, over five percent less in December 2008. We then ask how the small differences between these securities, in particular their liquidity, could project to so large a gap. We gauge the potential for long/short arbitrage with repo and fails data indicating the frictions arbitrageurs encountered, and then with daily transactions data we relate the demand for the expensive but liquid note to the cross section of insurers' liquidity needs.

The market for U.S. Treasury debt is the largest, most liquid, and safest securities market in the world. The total amount of publicly-held, Treasury-issued debt currently stands at $9 trillion, and daily trading volume typically exceeds $500 million.[1] Not surprisingly, pricing anomalies in the Treasury market are infrequent, short-lived, and well-studied when they do occur. For instance, typically there is a pricing difference between off-the-run and on-the-run securities; the most recently issued coupon security of a particular maturity tends to be slightly more expensive than previously-issued securities of the same original maturity. However, Krishnamurthy (2002) shows that the trading profits from entering into a convergence trade that is short the on-the-run security and long the off-the-run security are largely offset by the cost of borrowing the on-the-run. He concludes that there are no consistent arbitrage profits to be made from these pricing differences.

In this paper, we document the occurrence of a large pricing anomaly in the Treasury market that created arbitrage profits that would not have been offset by borrowing costs. Specifically, we show that a large yield spread developed between securities originally issued as thirty-year bonds and securities originally issued as ten-year notes, even though the securities share the same maturity date. For a several month period toward the end of 2008, original issue bonds became substantially cheaper than original issue notes, even after adjusting for differences in coupons. We show that the pricing anomaly was large, reaching a level as high as 5 percent, and it remains even after correcting for the difference in funding costs as measured by repurchase agreement (repo) rates.

[1] The source for Treasury securities outstanding and trading volume is the Securities Industry and Financial Markets association, available at www.sifma.org.

The pricing anomaly happens during a period of significant market turmoil, when liquidity was in particularly high demand. The spread that we document is highly correlated with other measures of market liquidity, including spreads between on-the-run and off-the-run Treasury securities and average bid-ask spreads on all Treasury securities. We conjecture that the pricing anomaly we document is related to small differences in liquidity that became magnified during the financial crisis. We show that bonds generally have lower trading volume and wider bid-ask spreads, which widened further during the crisis, suggesting that old thirty-year bonds are less liquid than ten-year notes. There are three reasons for this difference in liquidity. First, smaller amounts of these bonds were originally issued. Second, the majority of bonds typically are stripped and held in stripped form for the remainder of their lives, whereas stripping of notes is much more limited (Jordan, Jorgensen and Kuipers (2000)). This further reduces the amount of the bond that is immediately available to trade. Third, we conjecture that the bond may be disproportionately held by longer-term investors.

We interpret the apparent mispricing of Treasury securities from the perspective of the "limits to arbitrage" literature. Although repo rates and bid-ask spreads did reduce the profits available from trading against the mispricing, we show that the magnitude of the discrepancy provided a clear arbitrage opportunity from the perspective of a hold-to-maturity investor. The "limits to arbitrage" literature presents an explanation for why an arbitrageur may still shy away from such a trade. The essence of the story - described in detail in papers such as Shleifer and Vishny (1998), Gromb and Vayanos (2002) and Vayanos and Vila (2009) - is that arbitrageurs are risk-averse, have a short-horizon, or are capital constrained. All of these frictions can prevent an arbitrageur from taking the perspective of a hold-to-maturity investor, making the mispricing less attractive. In particular, as noise traders move prices for reasons unrelated to an assets fundamentals, arbitrageurs are tempted to trade the security to take advantage of the mispricing. However, arbitrageurs have to assume the risk that the noise traders could make the mispricing worse in the short-run might, which subjects the arbitrageur to interim losses. Because the arbitrageur is risk-averse in the short-run or has limited capital, this risk can deter the arbitrageur from acting. This is true even if the mispricing would make the arbitrageur a sure profit, if he/she could hold out indefinitely.

In addition to the on-the-run/off-the-run spread that is studied extensively in Krishnamurthy (2002), other anomalies in the Treasury market have also been documented. Amihud and Mendelson (1991) and Kamara (1994) compare bills and notes with less than six months remaining to an identical maturity date, so that both are effectively zero-coupon securities. During their sample, the notes were consistently cheaper (traded at higher yields) than the bills. Amihud and Mendelson argue that the price differential represents a premium for the greater liquidity of bills, but Kamara (1994) suggests that the difference owes in part to the differential tax treatment that existed at the time. In the paper most closely related to ours, Strebulaev (2003) compares the yields of coupon securities (Treasury notes and bills) with different original-issue tenors but with identical maturity dates. Although Strebulaev confirms that bills tend to be more expensive than similar notes, he finds that standard liquidity proxies are not correlated with bill-note pricing differences. Moreover, he finds no evidence of systematic pricing differences within Treasury notes, leading him to conclude that liquidity differences are not the source of the note-bill anomaly.

Many pricing anomalies have been interpreted in the context of this type of argument, and it has potential to apply to the Treasury market as well. In fact, the Treasury market is an ideal setting to cleanly establish the existence of an arbitrage and test some empirical implications. Treasury securities all share identical credit risk and do not differ in priority, but some securities are more easily traded than others. Bonds are cheaper than notes, perhaps because some investors have a preference for the greater liquidity of notes, but arbitrage capital is normally available to keep the prices of notes and bonds aligned. The crisis was a time in which arbitrage capital was withdrawn, and this led the spread between note and bond yields to skyrocket. Treasury market anomalies (not confined to the bond-note spread) during the crisis are discussed in these terms by Hu, Pan and Wang (2010). By identifying specifically which Treasury securities are relatively cheap and which are relatively expensive, we can then investigate the types of participants that are either exploiting the arbitrage or acting as noise traders based on their trading activity in these particular Treasury securities at the time that the pricing divergence widened.

The plan for the remainder of this paper is as follows. In section 2, we describe the note-bond pricing anomaly that is the focus of this paper. Section 3 relates pricing anomalies in the crisis to characteristics of individual coupon securities, arguing that the cross-section of yields can be accounted for in terms of liquidity differences. Section 4 evaluates the characteristics of investors that were exploiting the arbitrage opportunity, and those that were making it worse. Section 5 concludes.

## I. The Arbitrage

We begin by illustrating an example of the pricing anomaly that explore, shown in Figure 1. The figure shows the spread between the yields to maturity on two Treasury securities, both maturing on February 15, 2015. We take the difference between an original-issue 30 year bond and an original issue 10 year note. The bond was originally issued in 1985 with a coupon of 11.25 percent, and the note was originally issued in 2005 with a coupon of 4 percent. Throughout 2005, 2006, and much of 2007, the bond has a slightly higher yield-to-maturity than the note, by about few basis points, on average. In late 2007 and early 2008, however, the bond yield climbed substantially relative to the note yield. The difference spiked in the fall of 2008, when the spread of the bond yield over note yield climbed to 80 basis points, representing a price difference of about five dollars per 100 dollars face value. It is worth noting that both securities were well off-the-run during the time when the yield spread widened most notably.

This yield spread documented in Figure 1 does not necessarily represent an arbitrage opportunity, since the note has a lower coupon than the bond and thus a longer duration. But with an upward sloping yield curve, as was the case during 2008, the difference in coupons should result in a relatively *higher* yield to maturity for the note relative to the bond; Figure 1 shows that the note has a lower yield to maturity.

To conduct a more precise comparison between the pricing of the note and the bond, we create a synthetic portfolio of the bond and a Treasury STRIP to exactly match the cash flows of the note. Specifically, for a note with coupon rate $C_n$ and a bond with coupon rate $C_b$ (both maturing on the same day), we form a portfolio that puts weight $C_n / C_b$ on the bond and weight $1 - C_n / C_b$ in a STRIP maturing on the maturity date of the bond. This portfolio will have identical cash flows to the note, which lets us compare the prices of two assets that generate identical cash flows. In our empirical analysis below, we show that the price of the note compared with the price of the bond and STRIP portfolio is usually close to zero but grew significantly during the period of the financial crisis before returning back close to zero in 2009.

A trading strategy that bet on convergence of the prices of two portfolios would have made positive profits. Moreover, these profits would be riskless, as long as the positions could be held until maturity, and as long as funding costs or other frictions in implementing the trade would not exceed the difference in returns on the portfolio. In particular, the cost of shorting the more expensive note could wipe out any profits created by convergence in the prices, although we show below that this is not the case. In the absence of such significant frictions, the strategy of betting on convergence could be scaled to produce enormous profits.

In this paper we consider nine bond-note pairs similar to the example displayed in Figure 1. In all cases, the original-issue note becomes expensive relative to the portfolio comprised of the original-issue bond and STRIP with identical maturity date, although the size of the pricing gap varies across the pairs. We explicitly incorporate the cost of forming the short position in the note using repo rates that account for any specialness in shorting a particular security. We interpret any remaining difference as potential arbitrage profits that would be available to a hold-to-maturity investment position.

## II. Apparent Pricing Anomalies

Our analysis begins by recognizing that the period between August 2007 and May 2009 represented a period of significant market turmoil. For example, Hu, Pan, and Wang (2010) document that deviations in Treasury yields from a smooth yield curve hit a record high in the weeks following the Chapter 11 filing of Lehman Brothers in September 2008. Hu, Pan, and Wang construct a measure of illiquidity based on the average deviation of Treasury prices from those based on a smooth yield curve and show that this measure provides a useful proxy for illiquidity and is a priced risk factor. We adopt their measure of illiquidity and show that deviations from the smooth curve were systematic: original-issue 30-year bonds became cheap relative to original-issue 10 year notes. The systematic nature of the pricing deviations leads us to create the bond-note pairs that we explore further in the next section.

*II.1 Cheap and Expensive Securities*

This subsection addresses the question of which securities became relatively cheap during the crisis and which were relatively expensive. To address this question, we compare actual Treasury prices with those implied from a parametric zero-coupon yield curve fitted to the set of all coupon securities. We use the parameter estimates provided by the Federal Reserve Board, who every day fit the six-parameter model of instantaneous forward rates of Svensson (1994) to observed prices on coupon treasury securities.[2] With the parameter estimates, we can compute the fitted price of each security on every calendar day and compute the difference between observed prices and the fitted price. We denote the difference as the pricing error, which by construction has mean close to zero across all securities.[3] We use the CRSP daily Treasury database for our treasury security prices.

Figure 2 shows the average pricing error for all securities which were originally issued as thirty-year bonds, ten-year notes and five-year notes. Prior to the summer of 2007, average pricing errors were close to zero, and there is very little difference between thirty-year bonds and ten-year notes. Beginning in the fall of 2007 and extending through early summer of 2009, a notable pattern emerges. The thirty-year bonds became cheap relative to the smooth curve, and the ten-year notes became expensive. Notably, the pricing errors on the five-year notes do not

[2] See Gurkaynak, Sack, and Wright (2006) for a discussion of the methodology. See the following website for the data: http://www.federalreserve.gov/pubs/feds/2006/200628/200628abs.html

[3] The mean is not exactly zero because prices are a non-linear function of forward rates.

show systematic time series deviations. As we will show next, part of the pattern can be explained by the amount of the bonds outstanding, but there will still remain a significant pattern that thirty-year bonds became cheap relative to ten-year notes.

To further explore the determinants of the pricing error, we estimate a panel regression of the pricing errors on individual securities onto a variety of security characteristics. The regression is of the form

$$e_{it} = \alpha + \beta' X_i + \varepsilon_{it},$$

where $e_{it}$ denotes the pricing error for the $i$th security on day $t$, and $X_i$ is a vector of bond-specific characteristics. We use two sets of independent variables. First, as two proxies for liquidity, we include the size of the issue and the quoted bid-ask spread. We measure size as the log of the original amount of the bond issued, and use the log of the dollar value of the bid-ask spread.[4] Second, we include dummy variables indicating whether the security was originally issued as a thirty-year bond, a ten-year note, or a five-year note, with the excluded category including seven-year, three-year, and two-year notes. The regression includes observations during 2005 through 2010 and includes all coupon securities with remaining time to maturity of at least one year and no more than ten years; the smooth yield curve fits best within this range. We also run the regression on a sub-sample of observations during the crisis period, which we define as lasting from the fourth quarter of 2007 through the second quarter of 2009. Since the same security appears many times in the sample, standard errors account for potential serial correlation in residuals.

[4] We will eventually account for buybacks and re-openings by allowing the amount outstanding to vary by day.

The results are shown in Table 1. The coefficient on the amount issued is positive, suggesting that larger issues tend to be relatively expensive. Similarly, securities with larger bid-ask spreads tend to trade at higher prices, although the effect is fairly small and only significantly different from zero during the crisis period. We view these results as suggesting that differences in liquidity, as proxied by issue size and bid-ask spreads, lead to systematic differences pricing, with more liquid issues trading cheaper than less liquid issues. Interestingly, the effect of issue size and bid-ask spreads is much stronger during the crisis, suggesting that liquidity differences were exacerbated during the crisis. Even after controlling for these liquidity proxies, the dummy variables for the original-issue term of the security confirm that ten-year notes became expensive relative to thirty-year bonds. The difference in estimated coefficients on the thirty-year bonds and the ten-year notes is large and statistically significantly different from zero in both samples. During the crisis, the difference in coefficients exceeds 1, meaning that, on average over the seven quarters, bonds were more than 1 percent cheaper than notes.

We do not have a compelling reason why the notes became rich relative to the bonds, but we conjecture that unobserved differences in liquidity are the underlying source, which was exacerbated during the crisis. Although an interesting area for future research, the underlying reason is unimportant for our subsequent analysis. At maturity of the proposed trading strategy, both securities are equally liquid, so from the perspective of a hold-to-maturity investor, any liquidity differences do not matter. What matters for us is that bonds systematically cheapened relative to notes, which creates the potential arbitrage that we explore.

## III. The Arbitrage Strategy

In this section, we describe how we construct two portfolios with identical cash flows and show that, in normal times, the two portfolios have very similar prices. We then document the pricing anomaly that emerges and show that funding costs did not reach levels that would overwhelm the arbitrage profits.

### *III.1 The Bond-Note Pairs*

We construct nine pairs of securities with the same maturity dates that were originally issued as ten- and thirty-year Treasuries. We consider only nominal, non-callable Treasury securities, of which the February 2015 securities (described above) are an example. Prior to 1985, the U.S. Treasury Department exclusively issued callable thirty-year securities, so we use only bonds issued after 1984. We also restrict our sample to notes that were issued prior to the summer of 2008, so that all of the bond/note pairs exist during the peak of the financial crisis. With these restrictions, we are left with nine bond/note Treasury pairs with identical maturity dates ranging from February 2015 to May 2018.

For each pair, we construct a portfolio that is short the note, long a fraction of the bond to match the coupons of the note, and long a Treasury STRIP to match the principal payment at maturity (as discussed above). This portfolio is constructed to have zero cash flows after origination, so it should not have any cost or benefit at origination. We view any money received at origination as an arbitrage opportunity.

*III.2 Accounting for the Cost of Funding*

In the classic "convergence trade" that we describe above, an arbitrageur would take a long position in the cheap security (the bond) and a short position in the expensive one (the note). In reality, it is often expensive to short some securities, a friction discussed by Duffie (1996) and Krishnamurthy (2002), among others. The only way to take a short position in a Treasury issue is to enter into a repo contract where one lends out cash and takes the security as collateral. The lender can then sell the collateral immediately, betting that the price will fall, intending to buy it back at the close of the repo contract, hopefully at a cheaper price. An investor wishing to bet on an anomaly in the Treasury market must short the expensive security in this way. At the same time, the investor can buy the cheap security, typically using the security as collateral to borrow money to finance the purchase. In most repo transactions, any Treasury security is considered to be acceptable collateral, and the corresponding interest rate on the loan is known as the general collateral (GC) interest rate.

In some cases, repo cash borrowers may deliver any Treasury security as collateral, leading particularly expensive issues to not be delivered in GC agreements. However, some repo agreements specify the precise issue that must be used as collateral and must be returned at the end of the repo contract. When one security is unusually expensive, demand from investors wishing to short it can drive down the repo rate on that security to a level below the GC rate, and the security is referred to as "special." Securities that are expensive in the cash market are typically "special" in the repo market, meaning that the cost of shorting them is particularly high. When Treasury securities become special, the repo rate on the particular security is known as the security's special repo rate, which will be lower than the GC repo rate. When this happens, an investor betting on an anomaly in the Treasury market will receive a lower interest rate on his/her loan of cash (collateralized by the expensive security) than he/she must pay to borrow to buy the cheap security, which will be at the GC repo rate. The spread between GC repo and specials rate could in principle wipe out the profitability of the convergence trade.

Indeed, Krishnamurthy (2002) shows that the profits on the convergence trade between on-the-run and off-the-run bonds are roughly wiped out by the gap between the corresponding repo rates. Although the spread between these two bonds systematically converges over time, the average profits of this trade are close to zero due to the cost of shorting the newly issued bond. Krishnamurthy argues therefore that there is no genuine arbitrage opportunity.

It is particularly important for us to account for funding costs, since anecdotal evidence suggests that funding became quite difficult during the crisis. Strains in the repo market likely made it hard to short comparatively expensive Treasury securities. Additionally, an institutional feature of the Treasury repo market made investors relatively reluctant to lend out their securities when GC rates because very low, as they did following Lehman's bankruptcy. Specifically, the lack of a penalty for failing to deliver on a repo transaction created a bound of zero on the specials rate, which could have prevented the market from clearing without excess demand or supply. [5]

[5] A market participant will lend funds against a security that is priced "special" only to meet an obligation to deliver that security. Until May 2009, the penalty for a failure to deliver a security into a transaction was that the security was to be delivered the next day at the same price. This is equivalent to giving the buyer of the security an interest free loan. This would be preferable to borrowing at a negative specials rate. So specials rates cannot normally go below zero. Due to massive fails in the repo market, and the resulting drop in securities lend via repos, the Treasury Market Practices Group (TMPG), a self-governing industry group, proposed a penalty fails rate, which was backed by the Federal Reserve. The explicit penalty in failing to deliver a security was introduced in May 2009 as Max(3-FFT,0), where FFT is the base of the Federal Reserve's target rate. In a zero policy rate environment, this rule levies a 3 percent penalty rate on a fail.

Using data on repo transactions from a large interdealer broker, we show that the profits available from the bond-note convergence trade we propose would have been much larger than the costs of funding the trade. Figure 3 plots the monthly return on the convergence trade (ignoring funding costs) along with the level of the special-GC spread (the funding cost) for the bond-note pair maturing on February 15, 2015. Although funding costs do rise during the crisis, the picture shows that the pricing differences were substantially larger than the funding costs. Even at the peak divergence in prices of the underlying securities, the repo funding costs remain below 15 basis points per month. Monthly returns, however, are much larger, in some cases exceeding 2 percent at the peak. Per $1000 principal, funding costs reach a maximum of $1 per month. Raw returns peak at $14.1 per $1000 principal in December 2008, following Lehman's bankruptcy filing in September 2008.

*III.3 Time Series Pattern of Arbitrage*

We next explore the time periods when the arbitrage grew to its widest levels, focusing on aggregate liquidity and limits to arbitrage. We conjecture that the risk aversion of potential arbitrageurs increased and arbitrage capital was withdrawn from the market. If so, the pricing error should be correlated with other systemic liquidity indicators.

To investigate this further, we run a daily time-series regression of the average pricing error across our nine bond-note pairs on several measures of aggregate liquidity. We use the LIBOR-OIS spread, the repo bid-ask spread, and the GC repo rate. The results are shown in Table 2. To account for the significant serial correlation in the pricing errors, we use Newey-West standard errors with a lag-length of 30.

The results suggest that the pricing error is significantly correlated with the measures of aggregate liquidity. The coefficient on the LIBOR-OIS spread is large and positive, indicating that broader funding strains were correlated with the anomaly in the Treasury market. The coefficient on the repo bid-ask spread is also positive, suggesting that the strains in the repo market also happened coincidentally with the pricing anomalies. Finally, the GC rate is significantly negative, which corroborates the notion that a lower GC spread makes lending expensive securities in the repo market less attractive, which in turn prevents arbitrageurs from bringing prices back into line.

## IV. Investor Response?

The pricing of Treasury securities in the crisis represented an arbitrage opportunity. Based on the "limits to arbitrage" paradigm, we suggest that this reflects a lack of arbitrage capital willing to take short-run risk to wait for the long-run gain. In this section, we explore the trading behavior of insurance companies, who are potential long-term investors that could profit from the arbitrage.

### *IV.1 Trading and Holdings Data*

We have a dataset consisting of transactions-level data showing all buys and sells of Treasury securities for all U.S. registered insurance companies, who report such transactions in Schedule D within their statutory regulatory filings. For each trade in the dataset, we know the insurance company conducting the trade, along with the date, size, and direction of the transaction. We also have several characteristics of the insurers, including several measures of their capital, including their financial strength ratings and their size. We also construct three additional variables based on the trading history of each insurer. In sum, we consider six cross-sectional characteristics of each insurer:

(i) *Buy-and-hold indicator*. A dummy that is one if that insurer is a "buy and hold" insurer, i.e. never conducts a sell transaction of Treasury securities.
(ii) *Horizon*. This is the average number of days that an insurer holds a given Treasury security.
(iii) *Churn*. This is the ratio of transactions volume relative to holdings over all Treasury securities for each investor. A lower value corresponds to a less active trader.
(iv) *Size*. This is the amount of assets held by the insurer.
(v) *Investment Grade*. A dummy that is one if the insurer's best rating is classified as investment grade.
(vi) *Premium-to-Asset* ratio. This is a leverage measure for each insurer.

For each insurer $i$ in month $t$, we construct the net purchases of notes less the net purchases of bonds, which we denote as $NP_{i,t}$. This is a measure of the propensity to engage in the arbitrage trade, and we relate the measure to the size of the pricing error and the cross-sectional characteristics of each insurer. In particular, we conduct a regression of the form:

$$NP_{i,t} = \alpha_i + \beta_i PE_t + \varepsilon_{i,t} \quad (1)$$

We further assume that the intercept and slope coefficients in this regression are linear functions of the characteristics of the insurer, collected in a vector $X_i$:

$$\alpha_i = a + b' X_i, \quad \beta_i = c + d' X_i \quad (2)$$

Substituting (2) into (1) gives:

$$NP_{i,t} = a + b' X_i + cPE_t + d' X_i PE_t + \varepsilon_{i,t} \quad (3)$$

which we then estimate as a pooled regression. The main object of interest is the interaction coefficient, $d$. This tells us whether a particular characteristic of an insurer makes the insurer more or less likely to buy the note when it becomes particularly expensive.

Table 3 reports the results from estimating equation (3) with each of the six different insurer characteristics separately. As can be seen, longer-horizon investors are on net sellers of the expensive note, which suggests that they are behaving as arbitrageurs. This is consistent with the finding of Coval and Stafford (2007) in equity markets. Non-investment grade and more leveraged insurers are also sellers of the expensive note, which means that these are the insurers who were in effect exploiting the arbitrage opportunity. Perhaps they had to sell Treasuries quickly in order to raise cash, and chose to do so by selling the relatively expensive and liquid notes.

## V. Conclusions

In normal times, the pricing of different Treasury securities is internally consistent. Two different Treasury coupon securities with different coupon rates but the same maturity date will have almost identical yields. Indeed, one can form a portfolio combining either one of these coupon securities with a set of STRIPS such that the portfolio has exactly the same payoffs as the other security. The portfolio and the security should—and normally do—have almost exactly the same price; otherwise one could create riskless profits that should not exist in a well-functioning market. However, starting with the onset of the financial crisis in August 2007, and then accelerating after the collapse of Lehman in the fall of 2008, these arbitrage relationships broke down dramatically. Bonds that were originally issued as thirty-year bonds that had 6-9 years to maturity became much cheaper than bonds originally issued with a ten-year maturity, even though both had nearly the same maturity date.

In the canonical theoretical model of persistent arbitrage opportunities, Shleifer and Vishny (1997) show that risk-aversion and bounded capital can explain why arbitrageurs are limited in their ability to prevent the emergence of pricing anomalies. In their model, "noise traders" have a liquidity-based motivation for trading that may cause prices to deviate from their fundamental value. Arbitrageurs trade against the noise traders to offset the deviations, but risk-aversion and limited capital can prevent the arbitrageurs from completely offsetting the divergence. The model explains why pricing discrepancies, and apparent arbitrage opportunities, can persist for some time. This paper aims to give some empirical content to the Shleifer and Vishny (1997) model by characterizing the nature of the noise traders and arbitrageurs and offering clues as to their motivation.

Studying the unusual pricing of Treasury securities at times of market stress gives us useful insights into the behavior of fixed income markets at times when there are distressed asset sellers. The Treasury market environment allows for particularly clean analytical results and interpretation of these issues, but the lessons learned should have applicability to other fixed income securities and perhaps even to different asset classes.

**References**


Amihud, Yakov and Haim Mendelson (1986). Asset Pricing and the Bid-Ask Spread. *Journal of Financial Economics 17*, 223-249.

Coval, Joshua and Erik Stafford (2007). Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics 86*, 479-512.

Duffie, Darrell (1996). Special Repo Rates. *Journal of Finance 51*, 493-526.

Gurkaynak, Refet, Brian Sack, and Jonathan Wright (2006). The U.S. treasury yield Curve: 1961 to the Present. Finance and Economics Discussion Series Paper 2006-28.

Gromb, Denis and Dimitri Vayanos (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics* 66, 361-407.

Hu, Xing, Jun Pan and Jiang Wang (2010). Noise as Information for Illiquidity. NBER Working Paper, 16468.

Jordan, Bradford D., Randy D. Jorgensen and David R. Kuipers (2000). The Relative Pricing of U.S. Treasury STRIPS: empirical evidence. *Journal of Financial Economics 56*, 89-123.

Kamara, Avraham (1994). Liquidity, Taxes, and Short-Term Yields. *Journal of Financial and Quantitative Analysis 29*, 403-417.

Krishnamurthy, Arvind (2002). The bond/old-bond spread. *Journal of Financial Economics 66*, 463-506.

Shleifer, Andrei, and Robert W. Vishny (1997). The Limits of Arbitrage, *Journal of Finance 52*, 35-55.

Strebulaev, Ilya (2002). Liquidity and Asset Pricing: Evidence from the U.S. Treasury Securities Market, Working Paper, Stanford University.

Svensson, Lars (1994), "Estimating and Interpreting Forward Rates: Sweden 1992-4," National Bureau of Economic Research Working Paper #4871.

Vayanos, Dimitri and Jean-Luc Vila (2009). A preferred-habitat theory of the term structure of interest rates, Working Paper, London School of Economics.

**Figure 1 – The Arbitrage**

This figure presents the time series of the difference between the yields to maturity on two Treasury securities: an original-issue 30 year bond and an original-issue 10 year note. Both securities mature on February 15, 2015. The bond was, originally issued in 1985 with a coupon of 11.25 percent; the note was originally issued in 2005 with a coupon of 4 percent.

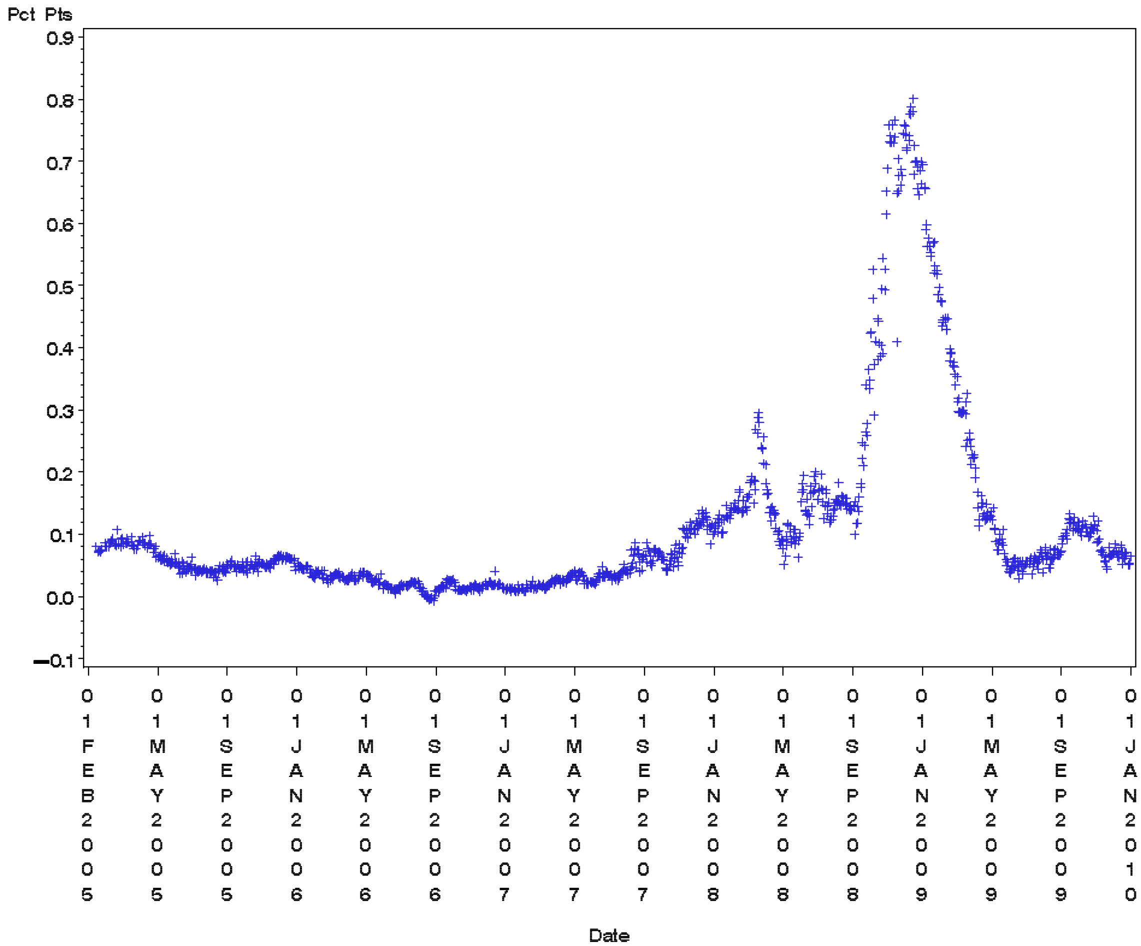


**Figure 2 – Pricing Errors by Original-Issue Maturity**

This figure presents the one-month rolling averages of the pricing errors across three original-issue maturity buckets: thirty-year bonds, ten- year notes, and five-year notes. The pricing error is defined as the difference between the actual price of the security and the fitted price based on a smooth forward rate yield curve. The vertical axis is measured in percentage points.

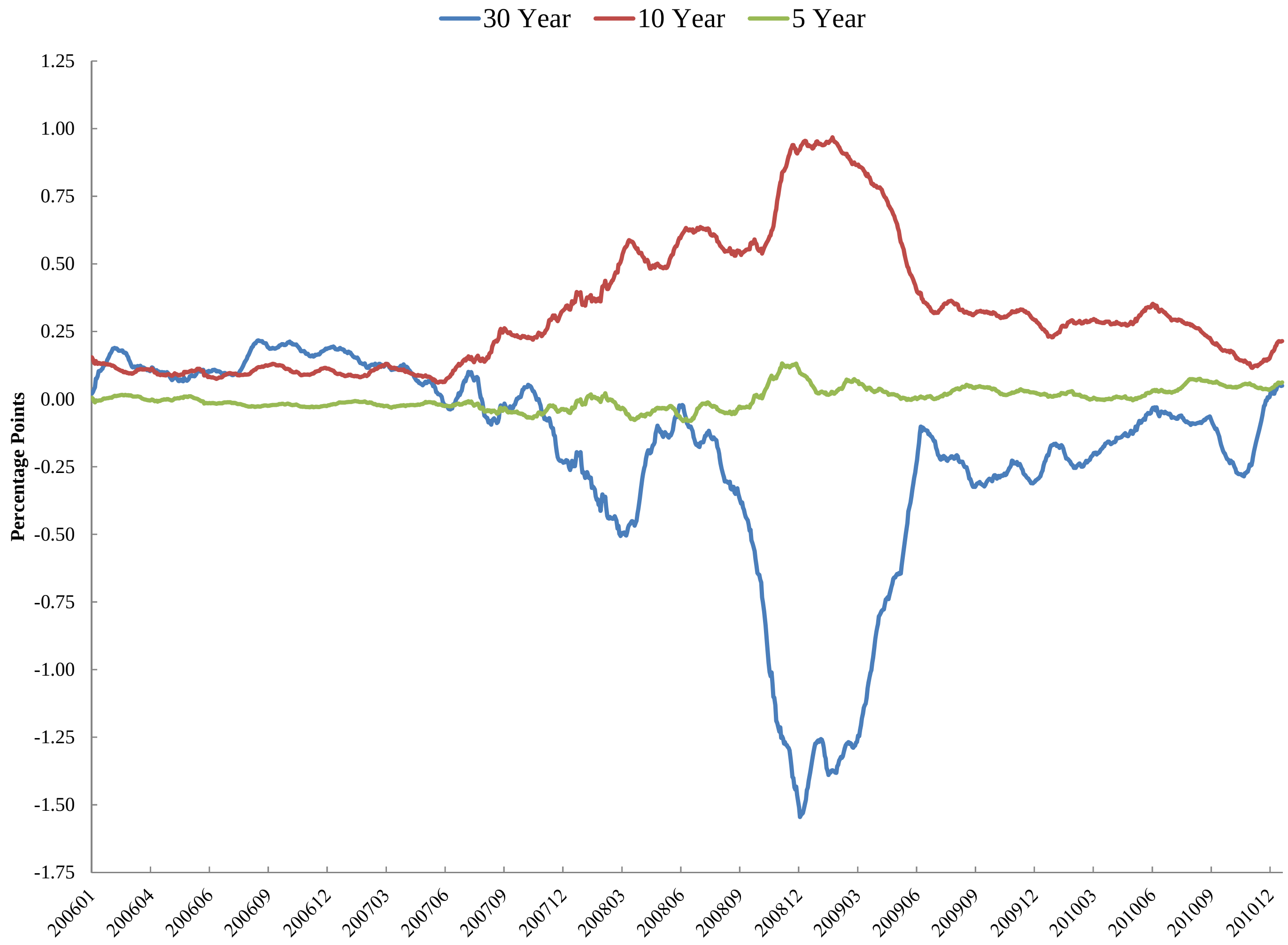


**Figure 3 – Arbitrage Profits vs Funding Costs**

This figure presents the monthly return on the convergence trade (ignoring the special-GC spread) and the level of the special-GC spread for the bond-note pair maturing on February 15, 2015.

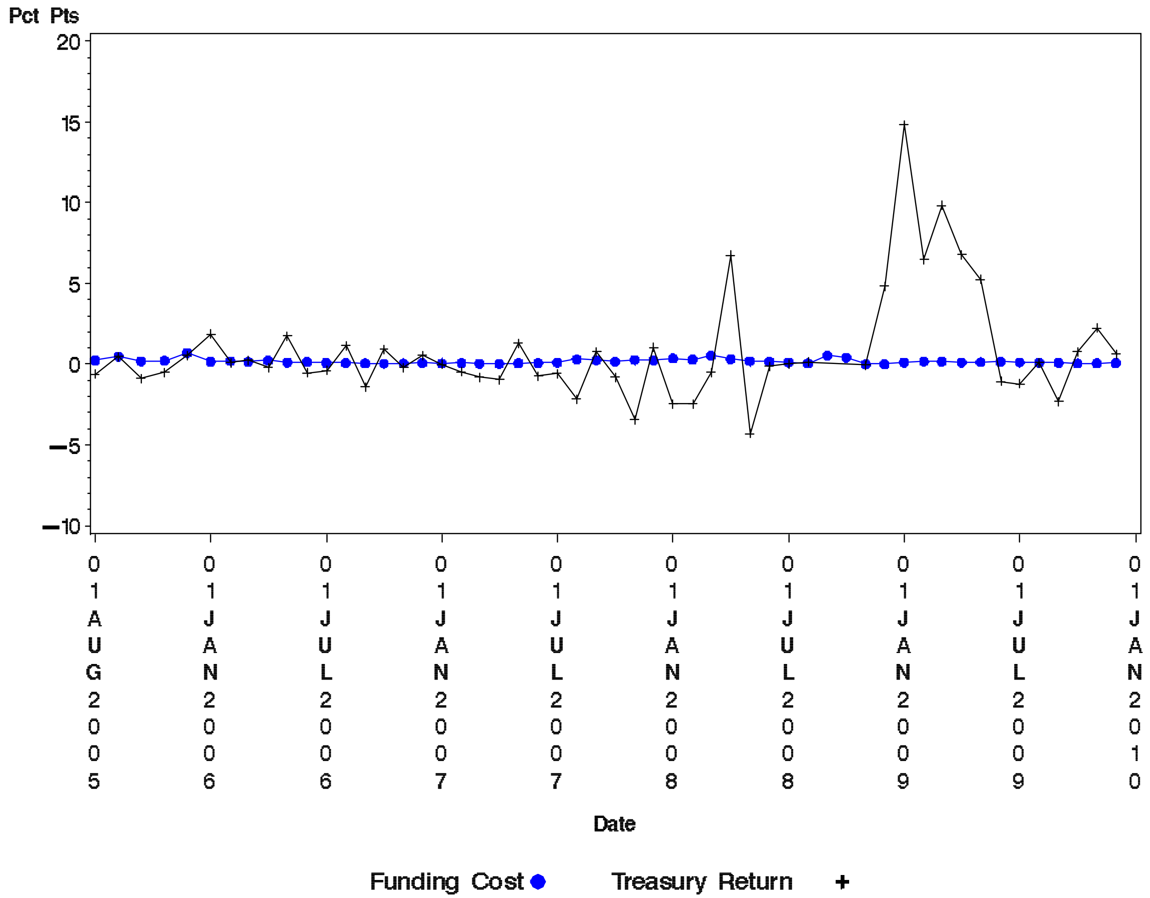


**Table 1 – Cross-Sectional Characteristics of Pricing Errors**

This table presents a regression of pricing errors on several bond characteristics: ln(outstanding) is the log of the dollar amount of the bond outstanding, ln(bid-ask) is the log of the dollar difference in quoted bid and ask prices, and the other three variables are dummy variables indicating the original issue maturity of the bond. The pricing error is defined as the difference between the actual price of the security and the fitted price based on a smooth forward rate yield curve. The sample period is January 1, 2005 through December 31, 2010. The crisis period is from September 1, 2007 to June 30, 2009. Standard errors (in parentheses) account for clustering within bond cusip and arbitrary heteroskedasticity; ** (*) denotes estimates that are statistically significantly different from zero at the 1(5)-percent level.

| | Dependent Variable: Pricing Error | |
|---|---|---|
| | Full Sample | Crisis Period |
| Intercept | -2.232** | -4.505** |
| | (0.747) | (1.792) |
| ln(outstanding) | 0.136** | 0.281** |
| | (0.044) | (0.107) |
| ln(bid-ask) | 0.034 | 0.092** |
| | (0.022) | (0.036) |
| Original issue 30-year | -0.144 | -0.378** |
| | (0.075) | (0.121) |
| Original issue 10-year | 0.342** | 0.671** |
| | (0.041) | (0.079) |
| Original issue 5-year | 0.070** | 0.147** |
| | (0.023) | (0.054) |
| R-Square | .231 | .407 |
| Observations | 149,228 | 46,192 |

**Table 2 – Time Series Characteristics of Pricing Errors**

This table presents a daily time regression of average portfolio pricing errors for nine bond-note pairs on several macro measures of liquidity: the average bid-ask spread on repo transactions (Repo B/A Spread), the repo rate for general collateral Treasuries (GC Repo Rate), and the spread between Libor and the overnight indexed swap rate (Libor-OIS Spread). For each bond-note pair, the pricing error is the price difference between the note and a bond plus a STRIP that gives the identical cash flows to the note. The sample period is January 1, 2005 through December 31, 2010. The crisis period is from September 1, 2007 to June 30, 2009. Newey-West Standard errors (with 30 lags) are in parentheses; ** (*) denotes estimates that are statistically significantly different from zero at the 1(5)-percent level.

| | Dependent Variable: Portfolio Pricing Error | | | |
|---|---|---|---|---|
| Repo B/A Spread | 15.80** | | | -2.58 |
| | (7.67) | | | (2.61) |
| GC Repo Rate | | -4.45** | | -3.89** |
| | | (1.06) | | (1.01) |
| Libor-OIS Spread | | | 14.03** | 9.19** |
| | | | (2.90) | (2.11) |
| R-Square | .231 | .406 | | |
| Observations | 149,231 | 46,194 | | |

**Table 3 – Who Engages in the Arbitrage**

This table presents a pooled regression of net purchases of notes less bonds on the pricing error interacted with various characteristics of the insurance companies. Standard errors (in parentheses) account for clustering within insurer and arbitrary heteroskedasticity; ** (*) denotes estimates that are statistically significantly different from zero at the 1(5)-percent level.

| | Dependent Variable: Net Purchases |
|---|---|
| Buy and hold | -12.8** |
| | 3.51 |
| Horizon | 7.2** |
| | 1.80 |
| Churn | -3.9** |
| | 1.07 |
| Assets | -3.5** |
| | 1.22 |
| I-grade | -19.2** |
| | 5.00 |
| Premium/Assets | 9.2** |
| | 2.12 |

# Trading Frenzies and Their Impact on Real Investment

Itay Goldstein, Emre Ozdenoren, and Kathy Yuan*

January 22, 2011

**Abstract**

We study a model where a capital provider learns from the price of a firm's security in deciding how much capital to provide for new investment. This feedback effect from the financial market to the investment decision gives rise to trading frenzies, where speculators all wish to trade like others, generating large pressure on prices. Coordination among speculators is sometimes desirable for price informativeness and investment efficiency, but speculators' incentives push in the opposite direction, so that they coordinate exactly when it is undesirable. We analyze the effect of various market parameters on the likelihood of trading frenzies to arise.

*We thank Alessandro Lizzeri (the editor), three anonymous referees, Roland Benabou, Josh Coval, Alex Edmans, Joao Gomes, Johan Hombert, Peter Kondor, Miklos Koren, Alessandro Pavan, Haresh Sapra, Lars Stole, Anjan Thakor, Dimitri Vayanos, Liyan Yang, and seminar and conference participants at AFA Meeting, CEU, LSE, MBS, NBER Corporate Finance Meeting, NYU, Paris School of Economics, Tel Aviv University, UCL, University of Chicago, University of Toronto, Utah Winter Finance Conference, Wharton, and Yale for their comments. We thank Leonid Spesivtsev for excellent research assistance. Goldstein is from the University of Pennsylvania; e-mail: itayg@wharton.upenn.edu. Ozdenoren is from the London Business School and CEPR; e-mail: eozdenoren@london.edu and emreo@umich.edu. Yuan is from the London School of Economics and CEPR; e-mail: K.Yuan@lse.ac.uk.

Trading frenzies in financial markets occur when many speculators rush to trade in the same direction leading to large pressure on prices. Financial economists have long been searching for the sources of trading frenzies, asking what causes strategic complementarities in speculators' behavior. This phenomenon is particularly puzzling given that the price mechanism in financial markets naturally leads to strategic substitutes, whereby the expected change in price caused by speculators' trades makes others want to trade in the opposite direction.

We argue in this paper that the potential effect that financial-market trading has on the real economy, i.e., on firms' cash flows, may provide the mechanism for trading frenzies to arise. Intuitively, suppose that speculators in the financial market short sell a stock, leading to a decrease in its price. Since the stock price provides information about the firm's profitability, it affects decisions by various agents, such as capital providers. Seeing the decrease in price, capital providers update downwards their expectation of the firm's profitability. This weakens the firm's access to capital and thus hurts its performance.[1] As a result, the firm's value decreases, and short sellers are able to make a profit. This creates a source for complementarities, whereby the expected change in value caused by speculators' trades makes others want to trade in the same direction, and generates a trading frenzy.

We develop a model to study and analyze this phenomenon. In particular, we study an environment where a capital provider decides how much capital to provide to a firm for the purpose of making new real investment. The decision of the capital provider depends on his assessment of the productivity of the proposed investment. In his decision, the capital provider uses two sources of information: his private information and the information aggregated by the price of the firm's security which is traded in the financial market. The reliance of capital provision on financial-market prices establishes the effect that the financial market has on the real economy. We refer to this effect as the 'feedback effect'.[2]

The financial market in our model contains many small speculators trading a security, whose payoff is correlated with the cash flow obtained from the firm's investment. Speculators trade on the basis of information they have about the productivity of the investment. They have access to two signals: the first signal is independent across speculators (conditional on the realization of the productivity), while the second one is correlated among them.[3] The

[1]Other agents that may be affected by the information in the price are managers, employees, customers, etc.

[2]In our model the financial market is a secondary market, and hence the only feedback from it to the firm's cash flow is informational; there is no transfer of cash from the market to the firm.

[3]In our model, the correlation is perfect, but this is not essential.

correlated signals introduce common noise in information into the model, which can be due to a rumor, for example. A trading frenzy occurs when speculators put large weight on the correlated signal relative to the idiosyncratic signal, and so they tend to trade similarly to each other.

To close the model, we introduce noisy price-elastic supply in the financial market. The market is cleared at a price for which the demand from speculators equals the exogenous supply. The endogenous price, in turn, reflects information about the productivity of the investment, as aggregated from speculators' trades. But, given the structure of information and trading, the information in the price contains noise from two sources – the noisy supply and the common noise in speculators' information. The information in the price is then used by the capital provider, together with his private information, when making the decision about capital provision and investment.

Analyzing the weight speculators put on the correlated signal relative to the idiosyncratic signal, we shed light on the determinants of trading frenzies. In a world with no strategic effects, this weight is naturally given by the ratio of precisions between the correlated and the idiosyncratic signals. But, in the equilibrium of our model, there are two strategic effects that shift the weight away from this ratio of precisions. The first effect is the usual outcome of a price mechanism. When speculators put weight on the correlated information, this information gets more strongly reflected in the price, and then the incentive of each individual speculator to put weight on the correlated information decreases. This generates strategic substitutes and pushes the weight that speculators put on the correlated information below the ratio of precisions.[4] The second effect arises due to the feedback effect from the price to the capital provision decision. When speculators put weight on the correlated information, this information gets to have a stronger effect on the capital provision to the firm and hence on the real value of its traded security. Then, the incentive of each speculator to put weight on this information increases. This leads to strategic complementarities that make speculators put a larger weight on the correlated signal.

This second effect is what causes a trading frenzy, leading speculators to put large weight on their correlated information, and to trade in a coordinated fashion. When this effect dominates, our model generates a pattern that looks like a 'run' on a stock by many speculators, who are driven by common noise in their correlated signals (e.g. rumor), leading to a price decline, lack of provision of new capital, and collapse of real value. This echoes some highly publicized events such as the bear raid on Overstock.com in 2005 or the bear raids

[4]Strategic substitutes due to the price mechanism appear in various forms in the literature on financial markets. See, for example, Grossman and Stiglitz (1980).

on Bear Stearns and Lehman Brothers in 2008.

We investigate what circumstances increase the tendency for a trading frenzy in our model. First, we show that speculators are more likely to trade in a coordinated fashion when the supply in the financial market is more elastic with respect to the price. This can be interpreted as a more liquid market. In such a market, the strategic substitutes due to the price mechanism are weak, as informed demand is easily absorbed by the elastic supply without having much of a price impact. Hence, speculators tend to put more weight on correlated information and trade more similarly to each other. Second, we find that when there is small variance in the supply function, i.e., when there is small variance in noise/liquidity trading in the financial market, speculators tend to put large weights on their correlated signals and thus to act in a coordinated fashion. This is because in these situations, the capital provider relies more on the information in the price since the price is less noisy, and so the feedback effect from the market to the firm's cash flows strengthens, increasing the scope of strategic complementarities. Third, the precision of various sources of information also plays an important role in shaping the incentive to rely on correlated vs. uncorrelated information. Intuitively, there will be more coordination when speculators' correlated signals are sharper and when their uncorrelated signals are noisier. Interestingly, there will be more coordination when the capital provider has less precise information of his own, as then the feedback from the market to his decision is stronger.

Another question we ask is whether trading frenzies are good or bad for the efficiency of the capital provision decision. We find that they are sometimes good and sometimes bad, and that there is a conflict between the level of coordination in equilibrium and the one that maximizes the efficiency of the capital provision decision. The efficiency of the capital provision decision is maximized when the informativeness of the price is highest. It turns out that when there is high variance of noise/liquidity trading in the market, higher degree of coordination among speculators increases price informativeness. This is because, in noisy markets, coordination among speculators is beneficial in suppressing the noise in liquidity trading that reduces the informativeness of the price. In such markets, trading frenzies among speculators are actually desirable because they enable decision makers to detect some trace of informed trading in a market subject to large volume of liquidity trading and noise. On the other hand, when the market is less noisy, the importance of coordination among speculators declines, and the additional noise that coordination adds via the excess weight that speculators put on their correlated information (which translates into weight on common noise) makes coordination undesirable. Hence, the conflict arises because high levels of coordination are desirable in noisy markets, but in equilibrium, speculators coordinate more in less noisy markets.

Finally, our model assumes that speculators submit market orders, as in Kyle (1985), i.e., they do not condition on the price when they trade. This is consistent with many real-world situations where traders are prepared to take on price risk to achieve greater immediacy in trading. In the penultimate section of the paper, we extend the model to allow speculators to condition on the price. Strategic interactions disappear if speculators observe exactly the same message from the market that is observed by the capital provider. But, this assumption is not very realistic, as capital providers are exposed to various sources of market information –e.g., future prices (if the capital provider acts with a lag), prices from other markets, or even rumours – that are not all perfectly observable to speculators when they trade. Hence, we develop a version of the model where the speculators condition their trades on the price, but the capital provider is exposed to another piece of information correlated with the price. We show how strategic interactions reemerge and discuss when they lead to frenzies like in our main model.

Our paper builds on a small, but growing, branch of models in financial economics that consider the feedback effect from trading in financial markets to corporate decisions. The basic motivation for this literature goes back to Hayek (1945), who posited that market prices provide an important source of information for various decision makers. Empirical evidence for this link is provided by Luo (2005) and Chen, Goldstein, and Jiang (2007). On the theoretical side, earlier contributions to this literature include Fishman and Hagerty (1992), Leland (1992), Khanna, Slezak, and Bradley (1994), Boot and Thakor (1997), Dow and Gorton (1997), Subrahmanyam and Titman (1999), and Fulghieri and Lukin (2001).

Several recent papers in this literature are more closely related to the mechanism in our paper. Ozdenoren and Yuan (2008) show that the feedback effect from asset prices to the real value of a firm generates strategic complementarities. In their paper, however, the feedback effect is modeled exogenously and is not based on learning. As a result, their paper does not deliver the implications that our paper delivers on the effect of liquidity and various information variables on coordination and efficiency. Khanna and Sonti (2004) also model feedback exogenously and show how a single trader can increase the value of his existing inventory in the stock by trading to affect the value of the firm. Goldstein and Guembel (2008) do analyze learning by a decision maker, and show that this might lead to manipulation of the price by a single potentially informed trader. Hence, the manipulation equilibrium in their paper is not a result of strategic complementarities among heterogeneously informed traders. Dow, Goldstein, and Guembel (2007) show that the feedback effect generates complementarities in the decision to produce information, but not in the trading decision.

More generally, our paper is related to the literature on informational externalities in financial markets. In particular, Vives (1993), Amador and Weil (2011) and others show how the reliance of agents on public information imposes negative externalities on others, as it reduces the efficiency of learning. Our paper shows how the weight on common information increases due to strategic complementarities that emerge as a result of the informational feedback from the market to real investment. Other papers explore other sources of complementarities in financial markets. For example see, Froot, Scharfstein, and Stein (1992), Hirshleifer, Subrahmanyam, and Titman (1994), Bru and Vives (2002), Veldkamp (2006a and 2006b), Ganguli and Yang (2009), Amador and Weil (2011), and Garcia and Strobl (2011).

Our paper is most closely related to Goldstein, Ozdenoren, and Yuan (2011) and Angeletos, Lorenzoni, and Pavan (2010). Both of these papers derive endogenous complementarities as a result of learning from the aggregate action of agents. To analyze trading frenzies and their impact on real investments, we embed this mechanism in a model of financial markets where a capital provider learns from the price to make an investment decision. Modeling the financial market explicitly enriches the problem in various ways. For example, having a price mechanism introduces strategic substitutes that coexist with the strategic complementarities in the model. Also, the ability of speculators to learn from the aggregate action in the rational-expectations-equilibrium extension gives rise to other effects mentioned above.[5] Hence, our model is substantially different from the above mentioned models. In terms of results, our model generates new insights in the context of our study, such as the effect of supply elasticity and noise trading on coordination in financial markets. We also derive new results on the difference between the equilibrium level of coordination and the efficient level of coordination.

The remainder of this paper is organized as follows. In Section 1, we present the model setup and characterize the equilibrium of the model. In Section 2, we solve the model. Section 3 analyzes the determinants of coordination among speculators in our model. In Section 4, we discuss the implications for the efficiency of investments and the volatility of prices and investments. In Section 5, we extend the model to allow speculators to condition their trades on the price. Section 6 concludes. All proofs are provided in the appendix.

[5] Another technical detail that we highlight in the description of the model is the use of log-normal distributions, which is necessary in a setting of feedback from financial-market prices to investment decisions.

# 1 Model

The model has one firm and a traded asset. There is a capital provider who has to decide how much capital to provide to the firm for the purpose of making an investment. There are three dates, $t = 0, 1, 2$. At date 0, speculators trade in the asset market based on their information about the fundamentals of the firm. At date 1, after observing the asset price and receiving private information, the capital provider of the firm decides how much capital the firm can have and the firm undertakes investment accordingly. Finally, at date 2, the cash flow is realized and agents get paid.

## 1.1 Investment

The firm in this economy has access to a production technology, which at time $t = 2$ generates cash flow $\tilde{F}I$. Here, $I$ is the amount of investment financed by the capital provider, and $\tilde{F} \geq 0$ is the level of productivity. Let $\tilde{f}$ denote the natural log of productivity, $\tilde{f} = \ln \tilde{F}$. We assume that $\tilde{f}$ is unobservable and drawn from a normal distribution with mean $\bar{f}$ and variance $\sigma_f^2$. We use $\tau_f$ to denote $1/\sigma_f^2$. As will become clear later, assuming a log-normal distribution for the productivity shock $\tilde{F}$ enables us to get a tractable closed-form solution.

At time $t = 1$ the capital provider chooses the level of capital $I$. Providing capital is costly and the capital provider must incur a private cost of: $C(I) = \frac{1}{2}cI^2$, where $c > 0$. This cost can be thought of as the cost of raising the capital, which is increasing in the amount of capital provided, or as effort incurred in monitoring the investment (which is also increasing in the size of the investment). The capital provider's benefit increases in the cash flow generated by the investment. To ease the exposition, we say that he captures the full amount $\tilde{F}I$.[6] The capital provider chooses $I$ to maximize the value of the cash flow from investing in the firm's production technology minus his cost of raising capital $C(I)$, conditional on his information set, $\mathcal{F}_l$, at $t = 1$:

$$I = \arg\max_I E[\tilde{F}I - C(I)|\mathcal{F}_l]. \tag{1}$$

The solution to this maximization problem is:

$$I = \frac{E[\tilde{F}|\mathcal{F}_l]}{c}. \tag{2}$$

The capital provider's information set, denoted by $\mathcal{F}_l$, consists of a private signal $\tilde{s}_l$ and the asset price $P$ observed at date 0 (we will elaborate on the formation of $P$ next).

[6]As we discuss below, the model will generate similar results if we assume that the capital provider gets a portion of the cash flow: $\beta\tilde{F}I$. But, then we would need to carry another parameter, $\beta$.

That is, $\mathcal{F}_l = \{\tilde{s}_l, P\}$. The private signal $\tilde{s}_l$ is a noisy signal about $\tilde{f}$ with precision $\tau_l$: $\tilde{s}_l = \tilde{f} + \sigma_l \tilde{\epsilon}_l$, where $\tilde{\epsilon}_l$ is distributed normally with mean zero and standard deviation one and $\tau_l = 1/\sigma_l^2$. Later, we will conduct comparative statics with respect to the precision of the capital provider's private signal. It is important to emphasize that even though our capital provider learns from the information in the price, he still may have good sources of private information. In fact, his signal can be more precise than other signals in the economy. Despite this, he still attempts to learn from the market, as agents in the market have other signals that are aggregated by the price.

## 1.2 Speculative Trading

The traded asset is a claim on the payoff from the firm's investment $\tilde{F}I$, which is realized at the final date $t = 2$. The price of this risky asset at $t = 0$ is denoted by $P$. One way to think about the traded asset is as a derivative, whose payoff is tied to the return from the investment. It can also be viewed as equity, to the extent that the value of the firm is $\tilde{F}I$ (and so the cost of the investment $C(I)$ is privately incurred by the capital provider).[7] It should be noted that, no matter what the nature of the asset is, our market is a secondary market with no cash transfers to the firm. The only effect of the market on the firm will be via the information revealed in the trading process.

In the market, there is a measure-one continuum of heterogeneously informed risk-neutral speculators indexed by $i \in [0, 1]$. Each speculator is endowed with two signals about $\tilde{f}$ at time 0. The first signal, $\tilde{s}_i = \tilde{f} + \sigma_s \tilde{\epsilon}_i$, is privately observed where $\tilde{\epsilon}_i$ is independently normally distributed across speculators with mean zero and unit variance. The precision of this signal is denoted as $\tau_s = 1/\sigma_s^2$. The second signal is $\tilde{s}_c = \tilde{f} + \sigma_c \tilde{\epsilon}_c$. This signal is observed by all speculators and $\tilde{\epsilon}_c$ is independently and normally distributed with mean zero and unit variance and $\tau_c = 1/\sigma_c^2$.[8]

[7] In this case, we could analyze our model assuming that this value is shared between the capital provider and shareholders, such that the former receives $\beta \tilde{F}I$ and the latter receive $(1 - \beta)\tilde{F}I$. This would not change our results, but will add complexity due to the additional parameter $\beta$. Hence, we omit $\beta$ in the paper.

[8] The assumption that the second signal is a common signal greatly simplifies the analysis. However, it is not necessary. The necessary element is that the noise in the information observed by speculators has a common component that cannot be fully teased out by the capital provider. In Goldstein, Ozdenoren, and Yuan (2010), we analyzed an alternative setup, where the second signal is specified as a heterogenous private signal with a common noise component $\tilde{\epsilon}_c$ and an agent-specific noise component $\tilde{\epsilon}_{2i}$. That is, $\tilde{s}_{ci} = \tilde{f} + \sigma_c \tilde{\epsilon}_c + \sigma_{\epsilon 2} \tilde{\epsilon}_{2i}$, where $\tilde{\epsilon}_c$ and $\tilde{\epsilon}_{2i}$ are independently normally distributed variables with mean zero and variance one. That paper, however, was simpler on other dimensions, as there was no price formation for the traded asset.

Each speculator can buy or sell up to a unit of the risky asset. The size of speculator $i$'s position is denoted by $x(i) \in [-1, 1]$. This position limit can be justified by limited capital and/or borrowing constraints faced by speculators.[9] Due to risk neutrality, speculators choose their positions to maximize expected profits. A speculator's profit from shorting one unit of the asset is given by $P - \tilde{F}I$, where $\tilde{F}I$ is the asset payoff and $P$ is the price of the asset. Similarly, a speculator's profit from buying one unit of the asset is given by $\tilde{F}I - P$.

Formally, speculator $i$ chooses $x(i)$ to solve:

$$\max_{x(i)\in[-1,1]} x(i) E\left[\tilde{F}I - P|\mathcal{F}_i\right], \tag{3}$$

where $\mathcal{F}_i$ denotes the information set of speculator $i$ and consists of $\tilde{s}_i$ and $\tilde{s}_c$. Since each speculator has measure zero and is risk neutral, an informed speculator optimally chooses to either short up to the position limit, or buy up to the position limit. We denote the aggregate demand by speculators as $X = \int_0^1 x(i)\, di$, which is given by the fraction of speculators who buy the asset minus the fraction of those who short the asset. For now, we assume that speculators do not observe the price when they trade, and hence they submit market orders, as in Kyle (1985). We discuss the role of this assumption in the extension in Section 5.

## 1.3 Market Clearing

At date 0, conditional on his information, each speculator submits a market order to buy or sell a unit of the asset to a Walrasian auctioneer. The Walrasian auctioneer then obtains the aggregate demand by speculators $X$ and also a noisy supply curve from uninformed traders, and sets a price to clear the market. The noisy supply of the risky asset is exogenously given by $Q(\tilde{\xi}, P)$, a continuous function of an exogenous demand shock $\tilde{\xi}$ and the price $P$. The supply curve $Q(\tilde{\xi}, P)$ is strictly decreasing in $\tilde{\xi}$, and increasing in $P$, that is, it is upward sloping in price. The demand shock $\tilde{\xi} \in \mathbb{R}$ is independent of other shocks in the economy, and $\tilde{\xi} \sim N(0, \sigma_\xi^2)$. As always, we denote $\tau_\xi = 1/\sigma_\xi^2$. The usual interpretation of noisy supply/demand is that there are agents who trade for exogenous reasons, such as liquidity or hedging needs. They are usually referred to as "noise traders". Several papers in the finance literature have explicitly endogenized the actions of these traders in simpler settings, but doing so here will significantly complicate the model. Our paper derives interesting comparative statics with respect to the amount of noise trading in the market (captured by $\sigma_\xi^2$).

[9] The specific size of this position limit on asset holdings is not crucial for our results. What is crucial is that informed speculators cannot take unlimited positions; if they do, strategic interaction among informed speculators will become immaterial.

To solve the model in closed form, we assume that $Q(\tilde{\xi}, P)$ takes the following functional form:

$$Q(\xi, P) = 1 - 2\Phi\left(\tilde{\xi} - \alpha \ln P\right), \tag{4}$$

where $\Phi(\cdot)$ denotes the cumulative standard normal distribution function. The parameter $\alpha$ captures the elasticity of the supply curve with respect to the price. It can be interpreted as the liquidity of the market: when $\alpha$ is high, the supply is very elastic with respect to the price, and so large informed demand is easily absorbed in the price without having much of a price impact. This notion of liquidity is similar to that in Kyle (1985), where liquidity is considered high when the informed trader has a low price impact. The basic features assumed in (4), i.e., that the supply is increasing in price and also has a noisy component, are standard in the literature. It is also common in the literature to assume particular functional forms to obtain tractability. The specific functional form assumed here is close to that in Dasgupta (2007) and Hellwig, Mukherji, and Tsyvinski (2006) .

## 1.4 Equilibrium

We now turn to the definition of equilibrium.

Definition 1: [Equilibrium with Market Orders ] An equilibrium consists of a price function, $P(\tilde{f}, \tilde{\epsilon}_c, \tilde{\xi}) : \mathbb{R}^3 \to \mathbb{R}$, an investment policy for the capital provider $I(\tilde{s}_l, P) : \mathbb{R}^2 \to \mathbb{R}$, strategies for speculators, $x(\tilde{s}_i, \tilde{s}_c) : \mathbb{R}^2 \to [-1, 1]$, and the corresponding aggregate demand $X(\tilde{f}, \tilde{\epsilon}_c)$, such that:

- For speculator $i$, $x(\tilde{s}_i, \tilde{s}_c) \in \arg\max_{x(i)\in[-1,1]} x(i)E\left[\tilde{F}I - P|\tilde{s}_i, \tilde{s}_c\right]$;
- The capital provider's investment is $I(\tilde{s}_l, P) = E\left[\tilde{F}|\tilde{s}_l, P\right]/c$.
- The market clearing condition for the risky asset is satisfied:

$$Q(\tilde{\xi}, P) = X(\tilde{f}, \tilde{\epsilon}_c) \equiv \int x(\tilde{f} + \sigma_s \tilde{\epsilon}_i, \tilde{f} + \sigma_c \tilde{\epsilon}_c) d\Phi(\tilde{\epsilon}_i). \tag{5}$$

Definition 2: A *linear monotone equilibrium* is an equilibrium where $x(\tilde{s}_i, \tilde{s}_c) = 1$ if $\tilde{s}_i + k\tilde{s}_c \geq g$ for constants $k$ and $g$, and $x(\tilde{s}_i, \tilde{s}_c) = -1$ otherwise.

In words: in a monotone linear equilibrium, a speculator buys the asset if and only if a linear combination of his signals is above a cutoff $g$, and sells it otherwise. In the rest of the paper we focus on linear monotone equilibria.

## 2 Solving the Model

In this section, we explain the main steps that are required to solve our model. Restricting attention to a linear monotone equilibrium, we first use the market clearing condition to determine the asset price. We then characterize the information content of the asset price to derive the capital provider's belief on $\tilde{f}$ based on $\{P, \tilde{s}_l\}$ and solve for the optimal investment problem. Finally, given the capital provider's investment rule and the asset pricing rule, we solve for individual speculators' optimal trading decision.

In a linear monotone equilibrium, speculators short the asset whenever $\tilde{s}_i + k\tilde{s}_c \leq g$ or, equivalently, $\sigma_s \tilde{\epsilon}_i \leq g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c$. Hence, their aggregate selling can be characterized by: $\Phi\left(\left(g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c\right)/\sigma_s\right)$. Conversely, they purchase the asset whenever $\tilde{s}_i + k\tilde{s}_c \geq g$ or, equivalently, $\sigma_s\tilde{\epsilon}_i \geq g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c$. Hence, their aggregate purchase can be characterized by $1 - \Phi\left(\left(g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c\right)/\sigma_s\right)$. The net holding from speculators is then:

$$X\left(\tilde{f}, \tilde{\epsilon}_c\right) = 1 - 2\Phi\left(\frac{g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c}{\sigma_s}\right). \tag{6}$$

The market clearing condition together with equation (4) indicate that

$$1 - 2\Phi\left(\frac{g - (1+k)\tilde{f} - k\sigma_c\tilde{\epsilon}_c}{\sigma_s}\right) = 1 - 2\Phi\left(\tilde{\xi} - \alpha\ln P\right). \tag{7}$$

Therefore the equilibrium price is given by

$$P = \exp\left(\frac{(1+k)\tilde{f} + k\sigma_c\tilde{\epsilon}_c - g + \sigma_s\tilde{\xi}}{\alpha\sigma_s}\right) = \exp\left(\frac{\tilde{f} + k\tilde{s}_c - g + \sigma_s\tilde{\xi}}{\alpha\sigma_s}\right), \tag{8}$$

which is informationally equivalent to

$$z(P) \equiv \frac{g + \alpha\sigma_s\ln P}{1+k} = \tilde{f} + \frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{1}{1+k}\sigma_s\tilde{\xi} = \left(\frac{1}{1+k}\right)\tilde{f} + \frac{k}{1+k}\tilde{s}_c + \frac{1}{1+k}\sigma_s\tilde{\xi}. \tag{9}$$

From the above equation, we can see that $z(P)$, which is a sufficient statistic for the information in $P$, provides some information about the realization of the productivity shock $\tilde{f}$. Yet, the signal $z(P)$ is not fully revealing of $\tilde{f}$, as it is also affected by the noise in the common signal $\tilde{\epsilon}_c$ and by the noisy demand $\tilde{\xi}$. Since the capital provider observes $z(P)$, he will use it to update his belief about the productivity. Note that $z(P)$ is distributed normally with a mean of $\bar{f}$. The variance of $z(P)$ given $\tilde{f}$ is $\sigma_p^2 = (k/(1+k))^2\sigma_c^2 + (1/(1+k))^2\sigma_s^2\sigma_\xi^2$. Hence, we denote the precision of $z(P)$ as a signal for $\tilde{f}$ as:

$$\tau_p = 1/\sigma_p^2 = \frac{(1+k)^2\tau_c\tau_\xi\tau_s}{k^2\tau_\xi\tau_s + \tau_c}. \tag{10}$$

After characterizing the information content of the price, we can derive the capital provider's belief on $\tilde{f}$. That is, conditional on observing $\tilde{s}_l$ and $z(P)$, the capital provider believes that $\tilde{f}$ is distributed normally with mean $\left(\tau_f \bar{f} + \tau_l \tilde{s}_l + \tau_p z(P)\right) / (\tau_f + \tau_l + \tau_p)$ and variance $1/(\tau_f + \tau_l + \tau_p)$. Then, using the capital provider's investment rule in equation (1) and taking expectations, we can express the level of investment as:

$$\begin{aligned} I &= \frac{1}{c} E[\tilde{F} | \tilde{s}_l = s_l, P] = \frac{1}{c} E[\exp\left(\tilde{f}\right) | \tilde{s}_l = s_l, P] \\ &= \frac{1}{c} \exp\left(\frac{\tau_f \bar{f} + \tau_l s_l + \tau_p z(P)}{\tau_f + \tau_l + \tau_p} + \frac{1}{2(\tau_f + \tau_l + \tau_p)}\right). \end{aligned} \tag{11}$$

Given the capital provider's investment policy in (11) and the price in (8), we can now write speculator $i$'s expected profit from buying the asset given the information that is available to him (shorting the asset would give the negative of this):

$$\begin{aligned} E\left[\tilde{F} I - P | \tilde{s}_i, \tilde{s}_c\right] &= \frac{1}{c} E\left[\exp\left(\frac{\tau_f \bar{f} + \tau_l s_l + \tau_p z(P)}{\tau_f + \tau_l + \tau_p} + \frac{1}{2(\tau_f + \tau_l + \tau_p)} + \tilde{f}\right) | \tilde{s}_i, \tilde{s}_c\right] \\ &\quad - E\left[\exp\left(\frac{\tilde{f} + k\tilde{s}_c - g + \sigma_s \tilde{\xi}}{\alpha \sigma_s}\right) | \tilde{s}_i, \tilde{s}_c\right]. \end{aligned} \tag{12}$$

Note that we made use here of the fact that $\tilde{F} = \exp\left(\tilde{f}\right)$. This is where using the natural log of the productivity parameter plays a key role. Using the properties of the exponential function, we can express the value of the firm $\tilde{F} I$ as $\frac{1}{c} \exp\left(\left(\tau_f \bar{f} + \tau_l s_l + \tau_p z(P) + 1/2\right) / (\tau_f + \tau_l + \tau_p) + \tilde{f}\right)$, where the expression in parentheses is linear in $\tilde{f}$. This enables us to get a linear closed-form solution, which would otherwise be impossible in a model of feedback.

Conditional on observing $\tilde{s}_i$ and $\tilde{s}_c$, speculator $i$ believes that $\tilde{f}$ is distributed normally with mean $\left(\tau_f \bar{f} + \tau_s \tilde{s}_i + \tau_c \tilde{s}_c\right) / (\tau_f + \tau_s + \tau_c)$ and variance $1/(\tau_f + \tau_s + \tau_c)$. Hence, substituting for $z(P)$ (from (9)) and taking expectations, equation (12) can be rewritten as:

$$E\left[\tilde{F} I - P | \tilde{s}_i, \tilde{s}_c\right] = \frac{1}{c} \exp(a_0 + a_1 \tilde{s}_i + a_2 \tilde{s}_c) - \exp(b_0 + b_1 \tilde{s}_i + b_2 \tilde{s}_c), \tag{13}$$

where the coefficients $a_0$, $a_1$, $a_2$, $b_0$, $b_1$, and $b_2$ are functions of $k$ and of the model's parameters. Explicit expressions for these coefficients are provided in the proof of Proposition 1 in the appendix.

A speculator will choose to buy the asset if and only if (13) is positive. Rearranging and taking logs leads to the following condition:

$$\tilde{s}_i + B(k)\, \tilde{s}_c \geq C(k) \tag{14}$$

where $B(k) = (a_2 - b_2) / (a_1 - b_1)$ and $C(k) = (b_0 - a_0 + \ln c) / (a_1 - b_1)$.[10] Function $B(k)$ can be thought of as the best response of a speculator to other speculators' weight on the

[10] Here, we assume that $a_1 - b_1 > 0$. This is verified later in the proof of Proposition 1.

correlated signal. That is, if all speculators in the economy put a relative weight $k$ on the correlated signal when deciding whether to attack or not, the best response for a speculator is to put the weight $B(k)$ on his correlated signal. The symmetric equilibrium is solved when $B(k) = k$. Recall that $a_1$, $a_2$, $b_1$, and $b_2$ are also functions of $k$, and hence the equilibrium condition $B(k) = k$ leads to a third-order polynomial. Analyzing this polynomial, we obtain the result in the following proposition. All proofs are in the Appendix.

PROPOSITION 1: For a high enough level of supply elasticity $\alpha$, there exists a monotone linear equilibrium characterized by weight $k^* > 0$ that speculators put on the common signal. This equilibrium is unique when the precision of the prior $\tau_f$ is sufficiently small.

The weight $k^*$ that speculators put on the common signal in equilibrium captures the degree of coordination in their trading decisions. When $k^*$ is high, speculators put a large weight on the common information when deciding whether to sell or buy the asset. This leads to large coordination among them and gives rise to a trading frenzy. In the upcoming sections, we develop a series of results on the determinants of coordination and its implications for the efficiency of the investment decision and for the volatility of prices. We focus on the case of large supply elasticity (large $\alpha$) and imprecise prior (small $\tau_f$), for which we know that there exists a unique equilibrium.

# 3 The Determinants of Speculators' Coordination

The weight that speculators put on the common signal in this model is affected by the degree to which there are strategic complementarities or strategic substitutes among them. To see the sources of the two strategic effects, recall from (3), that a speculator's expected profit is $x(i) E\left[\tilde{F}I - P|\mathcal{F}_i\right]$. When other speculators put more weight on the common signal, this signal gets to have a stronger effect on the price $P$, as well as on the real value of the security $\tilde{F}I$ (since the capital provider's investment decision is affected by the price). The first effect pushes the speculator to put a lower weight on the common signal, since relying on the common signal more heavily implies paying a high price when buying and a low price when selling. On the other hand, the second effect pushes the speculator to put a higher weight on the common signal, since relying on the common signal more heavily implies buying a security with high value and selling one with low value. Hence, the source of strategic substitutes in our model is the price mechanism, which is usual in models of financial markets, while the source of strategic complementarities is the feedback effect to the real value of the security.

In a world without these strategic effects, the weight that speculators put on the common signal relative to the private signal would be equal to the ratio of precisions between the signals: $\tau_c/\tau_s$. But, with strategic effects, the equilibrium weight on the common signal $k^*$ reflects the sum of the strategic effects on top of the precisions ratio; where the strategic substitutes due to the price mechanism push $k$ down and the strategic complementarities due to the feedback effect push it up. In the rest of this section, we formally isolate the various determinants of coordination to understand the impact of each factor on the equilibrium level of coordination.

## 3.1 Impact of Learning by the Capital Provider

Suppose that there is no feedback effect from prices to real values, because the capital provider does not learn from the price. In this case, the capital provider's decision on how much capital to provide becomes (this equation is analogous to equation (11) in the main model):

$$I = \frac{1}{c}E[\tilde{F}|\tilde{s}_l = s_l] = \frac{1}{c}\exp\left(\frac{\tau_f \bar{f} + \tau_l s_l}{\tau_f + \tau_l} + \frac{1}{2(\tau_f + \tau_l)}\right). \tag{15}$$

We again solve for the linear monotone equilibrium where speculators buy the asset if and only if $\tilde{s}_i + k_{BM}\tilde{s}_c \geq g_{BM}$ (the subscript $BM$ stands for 'benchmark'), and purchase the asset otherwise. Given the investment rule in (15), the expected profit for speculator $i$ from buying the asset, given the information available to him, becomes (this equation is analogous to equation (12) in the main model):

$$\begin{aligned} E[\tilde{F}I - P|\tilde{s}_i, \tilde{s}_c] &= E\left[\frac{1}{c}\exp\left(\frac{\tau_f \bar{f} + \tau_l s_l}{\tau_f + \tau_l} + \frac{1}{2(\tau_f + \tau_l)}\right)\tilde{F}|\tilde{s}_i, \tilde{s}_c\right] \\ &\quad -E\left[\exp\left(\frac{1}{\alpha\sigma_s}\left(\tilde{f} + k_{BM}\tilde{s}_c - g_{BM} + \sigma_s\tilde{\xi}\right)\right)|\tilde{s}_i, \tilde{s}_c\right]. \end{aligned} \tag{16}$$

For a speculator who buys the asset, (16) must be positive. Taking expectation and rearranging, we can see that a speculator buys the asset if and only if $\tilde{s}_i + B_{BM}(k)\,\tilde{s}_c \geq C_{BM}$ where[11]

$$B_{BM}(k) = \frac{\tau_c}{\tau_s} - \frac{\frac{\sqrt{\tau_s}}{\alpha}k}{\frac{\tau_s}{\tau_f+\tau_s+\tau_c}\left(\frac{\tau_f+2\tau_l}{\tau_f+\tau_l} - \frac{\sqrt{\tau_s}}{\alpha}\right)}. \tag{17}$$

[11]The expression for $C_{BM}$ and other details are in the proof of Proposition 2.

Solving $B_{BM}(k) = k$, as in the main model, we obtain the equilibrium weight that speculators put on the common signal in the case of no feedback effect from price to real investment:

$$k_{BM} = \frac{\left(\left(1 - \frac{\sqrt{\tau_s}}{\alpha}\right)\tau_f + \left(2 - \frac{\sqrt{\tau_s}}{\alpha}\right)\tau_l\right)\tau_c}{\left(\frac{\sqrt{\tau_s}}{\alpha}\right)(\tau_f + \tau_l)(\tau_c + \tau_f + \tau_s) + \left(\left(1 - \frac{\sqrt{\tau_s}}{\alpha}\right)\tau_f + \left(2 - \frac{\sqrt{\tau_s}}{\alpha}\right)\tau_l\right)\tau_s}. \tag{18}$$

Inspecting (18), we can see that $k_{BM}$ is lower than $\tau_c/\tau_s$, and that it approaches $\tau_c/\tau_s$ as $\alpha$ gets very large. The intuition is as follows: $\tau_c/\tau_s$ represents the ratio of precisions between the common signal and the idiosyncratic signal. This is the relative weight that speculators would put on the common signal if there were no strategic interactions. In a world without a feedback effect, the only strategic interaction between the speculators comes from the price mechanism, which generates strategic substitutes that reduce $k_{BM}$ below $\tau_c/\tau_s$. As $\alpha$ gets very large, this effect weakens, since the supply is highly elastic in the price, and so the price is not strongly affected by speculators' trades. Hence, speculators converge to the weight of $\tau_c/\tau_s$.

The following proposition summarizes the properties of $k_{BM}$ and its relation to the equilibrium weight $k^*$ in the main model.

Proposition 2: If the capital provider does not learn from the price when making lending decisions, the weight speculators put on the common signal $k_{BM}$ is given by (18). For a high enough level of supply elasticity $\alpha$, $k_{BM}$ is strictly below the equilibrium weight $k^*$ that speculators put on the common signal in the main model (with a feedback effect).

We can see that when we shut down the feedback effect from the price to real investment, the weight that speculators put on the common signal decreases. This is in line with our discussion above, according to which the feedback effect from prices to real investment is the source of complementarity in speculators' strategies, making them want to put more weight on the common signal. Hence, the feedback effect is the cause of trading frenzies in our model.

For illustration, we plot the best response function for our main model (as in equation (14)) and for the benchmark case (as in equation (17)) in Figure (1). In the figure, the intersections of $B(k)$ and $B_{BM}(k)$ with the 45-degree line establish the equilibrium weights $k^*$ and $k_{BM}$, respectively. As we see in the figure, $B(0) = B_{BM}(0) = \tau_c/\tau_s$. That is, in both cases, if other speculators put no weight on the common signal, a speculator finds it optimal to use the ratio of precisions between the common signal and the idiosyncratic signal as the weight for the common signal. This is because when other speculators do not put weight on

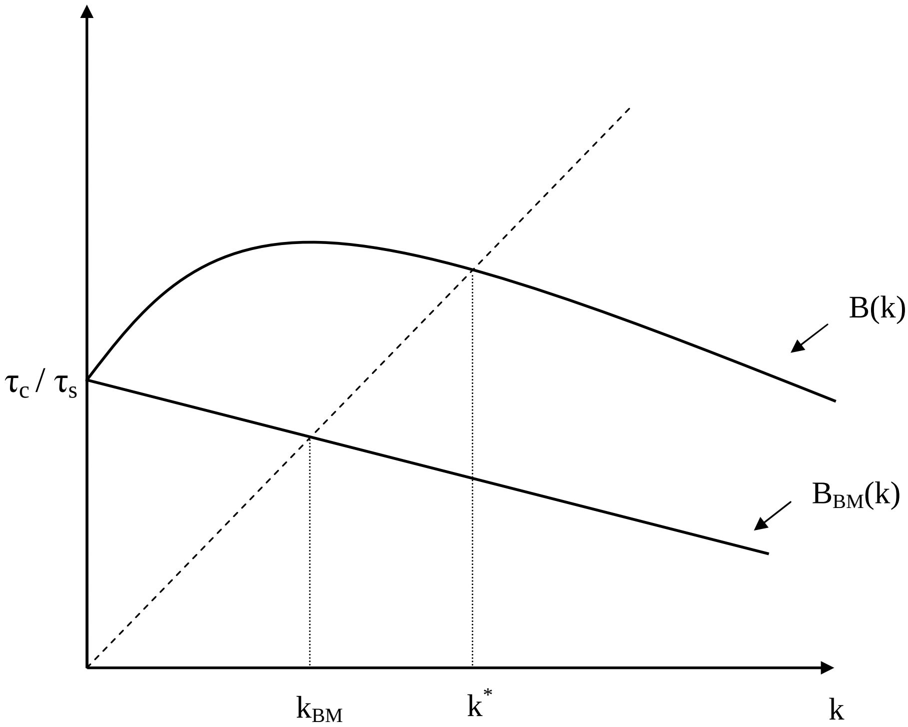



Figure 1: Best Response: $B(k)$ and $B_{BM}(k)$

the common signal, this signal is essentially like a private signal and hence it gets weighted solely based on its precision.

Once $k$ increases above 0, strategic substitutability from the price mechanism emerges in the benchmark model. Indeed, the best response $B_{BM}(k)$ is a decreasing function of $k$: when others put more weight on the common signal, this signal gets more strongly reflected in the price, making an individual speculator reduce the weight he puts on the common signal. By contrast, in our main model, in addition to strategic substitutability from the price mechanism, strategic complementarity also emerges due to the feedback effect. For $\alpha$ large enough, the effect from strategic complementarity dominates that from strategic substitutability, resulting in $B(k)$ increasing above $\tau_c/\tau_s$. As the figure shows, this results in a higher equilibrium weight on the common signal in the main model than in the benchmark model, which is proved formally in the proof of Proposition 2.

## 3.2 Impact of Supply Elasticity

The parameter $\alpha$ captures the elasticity of supply with respect to price in our model. When $\alpha$ is high, the supply of shares is very sensitive to the price, meaning that an increase in demand by informed traders is quickly absorbed in the market, so that informed trading does not have a large price impact. As mentioned above, $\alpha$ can then be interpreted as a measure of liquidity, and our model can be used to tell what is the effect of liquidity on trading frenzies. The following proposition tells us that the extent to which speculators coordinate on the common signal increases in the level of liquidity $\alpha$.

PROPOSITION 3: The equilibrium level of coordination $k^*$ is increasing in the supply elasticity $\alpha$, and for $\alpha$ large enough $k^*$ is greater than the precisions ratio $\tau_c/\tau_s$.

In illiquid markets, order flows have a large effect on the price. Then, when speculators put more weight on the common signal, this signal has a substantial effect on the price, and so other speculators want to put less weight on the common signal. This effect decreases as $\alpha$ goes up and liquidity improves. Hence, in liquid markets there is a greater tendency for coordination and trading frenzies. As the proposition shows, when $\alpha$ is large enough, the weight on the common signal increases beyond the ratio of precisions $\tau_c/\tau_s$.

## 3.3 Impact of Noise Trading

Noise trading is captured in our model by the variable $\tilde{\xi} \sim N(0, \sigma_\xi^2)$. A high level of $\sigma_\xi^2$ implies that the market is exposed to large levels of noise trading. In the literature on financial markets, this introduces noise to the price, and in the presence of a feedback effect, it makes it harder to base investment decisions on the price. In our model, we examine the effect of noise trading on speculators' coordination. As we will see later, this will have further implications for the informativeness of the price.

PROPOSITION 4: For a high enough level of supply elasticity $\alpha$, the equilibrium weight $k^*$ that speculators put on the common signal is decreasing in the variance of noise trading $\sigma_\xi^2$.

The intuition here goes as follows: With high variance in the noise demand, there is high variance in the market price for reasons that are not related to speculators' trades. As a result, the reliance of the capital provider on the information in the price decreases. This weakens the feedback effect and hence the strategic complementarities among speculators, leading to a lower level of $k^*$.

It is worth noting that changes in the position limits of speculators will have similar effects to changes in the variance of noise trading. For example, if speculators could choose positions in the range $[-2, 2]$ (instead of $[-1, 1]$, assumed in the paper), they would have more impact on the capital provider's decision for a given level of $\sigma_\xi^2$ and thus would put a larger weight on the common signal in equilibrium. Hence, the effect of loosening speculators' trading constraints is similar to that of reducing the variance of noise trading.

### 3.4 Impact of the Information Structure

We now establish comparative statics results on the effect of the informativeness of various signals on the equilibrium level of coordination. The results are summarized in the next proposition.

PROPOSITION 5: For a high enough level of supply elasticity $\alpha$, the equilibrium level of coordination $k^*$ decreases in the precision of speculators' private signals $\tau_s$, increases in the precision of their common signal $\tau_c$, and decreases in the precision of the capital provider's signal $\tau_l$.

These results are intuitive. Speculators put more weight on the common signal relative to the private signal when the common signal is more precise ($\tau_c$ is higher) and the private signal is less precise ($\tau_s$ is lower). Hence, trading frenzies are more likely when the common information becomes more precise relative to speculators' idiosyncratic sources of information. Less obvious is the result that the tendency for coordination among speculators decreases when the capital provider has more precise information ($\tau_l$ is higher). The reason is that when the capital provider has more precise information, he relies less on the price, and so the feedback effect from markets to real decisions weakens, and there is less scope for strategic complementarities.

### 3.5 A Note on the Nature of the Traded Security

Before moving to the next section, we would like to discuss the nature of the traded security. Our model assumes that the traded security is a claim on the cash flow from the investment $\tilde{F}I$. As we note in Section 1, this can be interpreted as a derivative, or, under some conditions, as equity of the traded firm (in which case it is simple to change the model so that the traded security is a claim on some portion $\beta$ of the cash flow $\tilde{F}I$).

The key feature of the traded security is that its cash flow depends on the investment decision $I$. This introduces a feedback loop between the financial market and the real economy, whereby the price affects the investment decision, and the investment decision is reflected in the price. This feedback loop is the crucial element for our result on strategic complementarities and trading frenzies. To illustrate this, note that if the traded security was a claim on the fundamental $\tilde{F}$, there would be no feedback loop and no frenzies. When speculators trade on $\tilde{F}$, the value of the security is exogenous and hence does not depend on speculators' behavior; this eliminates the strategic interaction that is central to our paper. It is worth noting that a security on $\tilde{F}$ might also not be easy to implement, since $\tilde{F}$ is not an easily verifiable cash flow (unlike $\tilde{F}I$, which is the cash flow from the investment). Indeed, most real-world financial securities are similar to the security we describe here in that they provide a claim on a cash flow that depends on fundamental and firm action.

Another possible security that features a feedback loop is one that provides a claim on the net return from the investment $\tilde{F}I - C(I)$. Unfortunately, we are unable to solve a model with this traded security. To see this, go back to (12). The expected value of the security for a speculator $E\left[\tilde{F}I|\tilde{s}_i, \tilde{s}_c\right]$ is expressed there as one exponential term (given our log-normal distributions), which is crucial for our ability to find a linear solution. If the traded security was $\tilde{F}I - C(I)$, we would have two exponential terms, which would render the steps for finding a linear solution impossible. We think that the basic message of our model – the emergence of strategic interactions among speculators due to the informational feedback from the security to the investment decision – will not change under this alternative security, albeit some of the details may change. As mentioned, with the existing techniques, such a model is unsolvable.

# 4 Coordination, Investment Efficiency, and Non-Fundamental Volatility

In this section, we explore the effect that coordination has on the efficiency of investment decisions and on market volatility. To analyze investment efficiency, we look at the ex ante expected net benefit of investment (i.e. expected net benefit before any of the signals are realized given the prior belief that $\tilde{f}$ is normally distributed with mean $\bar{f}$ and precision $\tau_f$) from the perspective of the capital provider. We keep the information structure the same as before, and in particular, in the interim stage we allow the capital provider to obtain information only from his private signal and the price. So our efficiency criterion is given by:

$$E_0\left[\max_I E\left[\tilde{F}I - \frac{1}{2}cI^2\,|\tilde{s}_l = s_l, P\right]\right], \tag{19}$$

where a speculator purchases the asset if $\tilde{s}_i + k\tilde{s}_c \geq g$ and shorts it otherwise (for constant $k$ and $g$) and $P$ is the market clearing price. We denote the optimal level of coordination $k_{OP}$ to be the one that maximizes investment efficiency as in (19).

The following proposition characterizes $k_{OP}$, and how it is linked to the accuracy of the information inferred from the market price, $\tau_p$:

Proposition 6: The level of coordination that maximizes investment efficiency is $k_{OP} = \tau_c / (\tau_s \tau_\xi)$, which also maximizes the precision of the price $\tau_p$.

The capital provider cares about the events in the security market only to the extent that they affect the quality of the information he has when making the investment decision. Hence, the level of coordination that maximizes investment efficiency is the one that maximizes the accuracy of the information in the market price. Examining the expression for the price signal in (9), we can see that there is a tradeoff in setting the level of coordination. The tradeoff arises because there are two sources of noise in the price, one coming from the noise trading $\widetilde{\xi}$ and the other one from the noise in the common signal $\tilde{\epsilon}_c$. (The first source of noise becomes more prominent when speculators' private information is noisy – $\tau_s$ is low – because then noise trading becomes relatively more important.) A high level of coordination reduces the effect of the first source of noise – as coordinated speculative trading helps overcoming the large volume of noise trading – and increases the effect of the second source of noise – as coordinated speculative trading increases the weight on the common signal. Therefore, the optimal level of coordination will be high when the potential damage from noise trading is high ($\tau_\xi$ and $\tau_s$ are low) or when the potential damage from noise in the common signal is low ($\tau_c$ is high). Then, $k_{OP} = \tau_c / (\tau_s \tau_\xi)$.

It is interesting to compare the optimal level of coordination characterized here with the level of coordination that is obtained in equilibrium. From Proposition 4 we know that in equilibrium speculators coordinate more when the variance in the noise trading is low ($\tau_\xi$ is high). A high $\tau_\xi$ implies that speculators' trades have more effect on the capital provider's decision, increasing the scope of strategic complementarities. Yet, this is exactly when coordination is not desirable for the efficiency of the investment. Hence, there is a sharp contrast between the profit incentives of speculators and the efficiency of the investment. Speculators coordinate more exactly when it is inefficient to do so. The following proposition summarizes the comparison between the optimal level of coordination and the equilibrium level of coordination.

Proposition 7: For a high enough level of supply elasticity $\alpha$, there exists $\bar{\tau}_\xi$ such that the level of coordination that maximizes investment efficiency is greater than the equilibrium level of coordination ($k_{OP} > k^*$) when the precision of the noise trading distribution $\tau_\xi$ is below $\bar{\tau}_\xi$. Similarly, $k_{OP} < k^*$ for $\tau_\xi > \bar{\tau}_\xi$.

The proposition says that speculators coordinate too much in markets with less noise trading and coordinate too little in markets with more noise trading. Interestingly, this implies that trading frenzies are only sometimes undesirable. When there is high variation in noise trading, price informativeness would improve if speculators coordinated their trades more to provide a signal that overcomes the effect of noise trading. Yet, it is exactly in this case that they find coordination less profitable in equilibrium.

We close this section by noting some of the implications of inefficient coordination levels. Deviations from the optimal level of coordination $k_{OP}$ are manifested in our model by higher levels of non-fundamental volatility. We define this as volatility that does not come from the variability in fundamental. The following proposition establishes the link between the level of coordination and non-fundamental volatility of price and investment.

PROPOSITION 8: (a) Non-fundamental volatility of asset price is minimized at $k = k_{OP}$ (where its value is $1/(\tau_c + \tau_s \tau_\xi)$).

(b) Similarly, non-fundamental volatility of investment is minimized at $k = k_{OP}$ (where its value is $1/(\tau_l + \tau_c + \tau_s \tau_\xi)$).

This proposition indicates that the strategic interactions among speculators in the financial markets often lead to non-fundamental volatility in prices as well as real activities. The source of this non-fundamental volatility could come from either too low coordination (that is, when the market is characterized by a high amount of noise trading) or too high coordination (that is, when the market has low noise trading and the noise in the correlated signals among speculators is high). Note that non-fundamental volatility is difficult to measure since it is defined as the volatility that does not come from fundamentals, while the volatility of fundamentals is unobservable (the volatility of cash flow is observable, but includes volatility due to noise). Hence, this notion is interesting mostly for theoretical reasons.

# 5 A Model where Speculators Learn from the Price

So far in the paper, we assumed that speculators in the financial market submit market orders that are not conditioned on the price. This assumption is common in the literature on financial markets, going back to Kyle (1985). It is also consistent with many situations in the real world, as speculators often face price risk when they trade without knowing the exact price (speculators may prefer this strategy over fully conditioning on the price, as this provides them greater immediacy in trading). In this section, we explore the importance of this assumption for our main result, which is the emergence of strategic complementarities among speculators due to the informational feedback from the price of the security to the investment decision.

As we explained before, the mechanism behind the strategic complementarities in our paper goes as follows: when speculators put more weight on the common signal, this signal gets to have a greater effect on the value of the security via the information conveyed by the price to the capital provider. Then, given this behavior of other speculators, each individual speculator finds it optimal to put more weight on the common signal himself. When speculators observe the price or fully condition on the price in this framework (as in Grossman and Stiglitz (1980)), strategic complementarities disappear. Once they observe the price, speculators know the message transmitted from the market to the capital provider, and so conditional on the price, they do not care what other speculators are doing. Indeed, one can show that if we just add the assumption that speculators fully condition on the price to our model, the weight that speculators put on the common signal will always be fixed at the precisions ratio: $\tau_c/\tau_s$ (as it is in the case of no strategic effects).

But, the above rational-expectations framework makes a very strong assumption: that the exact aggregate message from the market that is observed by the capital provider is also observed by each and every speculator. This eliminates the higher-order beliefs that are important for our complementarities to arise. This assumption is also not very realistic. In the real world, capital providers (or other decision makers) may receive a market-related signal that is not perfectly observable to traders. For example, when the timing of traders' trading decision and the capital provider's investment decision do not coincide, their information sets may not be perfectly correlated. This happens when traders condition their trade on the current price while the capital provider who acts with a lag has access to information from future prices or from other correlated markets. Alternatively, the capital provider and traders may share the same aggregate information source but their degrees of exposure to the source are different. Introducing such elements would imply that the speculators do not observe perfectly the aggregate message(s) used by the capital provider in making his investment decisions. This introduces back the higher-order beliefs that are crucial for our strategic complementarities to arise.

To analyze this formally in a tractable way, we now introduce two additions to the model. First, speculators observe the price and learn from it when they trade, just like in the traditional rational-expectations-equilibria (REE) literature (e.g., Grossman and Stiglitz (1980)). Second, conditional on fundamental, the capital provider's private source of information is correlated with the noise in the speculators' common signal. Specifically, his private source of information is now $\tilde{s}_l = \tilde{f} + \sigma_{cp}\tilde{\epsilon}_c$. Therefore, he is exposed to the noise in the common signal observed by speculators, $\tilde{\epsilon}_c$, but this is multiplied by a different coefficient $\sigma_{cp}$ (as opposed to $\sigma_c$ in the speculators' common signal), so he does not observe exactly the same common signal received by speculators $\tilde{s}_c$. Note that we could add another source of idiosyncratic noise to the capital provider's signal, and this would not have a qualitative effect on the results. As always, we denote the precision of the capital provider's signal as $\tau_{cp} = 1/\sigma_{cp}^2$. Finally, to maintain tractability, we now assume that the fundamental $\tilde{f}$ is distributed uniformly over the real line (i.e., $\tau_f$ approaches 0). Other than these changes, the model remains the same as in the previous sections.

The assumption that the capital provider observes a signal that is correlated with the speculators' common signal is consistent with the motivation described above. That is, the idea is that the capital provider observes an aggregate message from the market (the combination of the price $P$ and $\tilde{s}_l$) which is different from what is observed by any speculator when trading. The signal $\tilde{s}_l = \tilde{f} + \sigma_{cp}\tilde{\epsilon}_c$ can thus be thought of as a reduced form of observing the price from another market or from later rounds of trade, as these will also be affected by the common noise that exists in the market (recall that we could add an additional source of noise to $\tilde{s}_l$). Alternatively, the assumption that both $\tilde{s}_l$ and $\tilde{s}_c$ are affected by the common noise $\tilde{\epsilon}_c$ may capture other realistic scenarios. If the common signal in the market $\tilde{s}_c$ represents rumors, then the capital provider may be observing some version of these rumors. Or, the capital provider may report a noisy version of his signal $\tilde{s}_l$, which is observed among traders in the form of $\tilde{s}_c$.

We now turn to solve and analyze the extended model. As before, we consider monotone linear strategies where the speculators put weight on $\tilde{s}_i$, $\tilde{s}_c$, and now also on the price $P$. That is, speculators short the asset whenever $\tilde{s}_i + k\tilde{s}_c + m\ln P \leq g$ and buy it otherwise. The parameters $k$, $m$, and $g$ are determined endogenously. Following the steps in the main model, the net holding from speculators is then:

$$X\left(\tilde{f}, \tilde{\epsilon}_c\right) = 1 - 2\Phi\left(\frac{g - (1+k)\,\tilde{f} - k\sigma_c\tilde{\epsilon}_c - m\ln P}{\sigma_s}\right). \tag{20}$$

Since the market supply is $1 - 2\Phi\left(\tilde{\xi} - \alpha\ln(P)\right)$, we use the market clearing condition to express the price:

$$P = \exp\left(\frac{(1+k)\,\tilde{f} + k\sigma_c\tilde{\epsilon}_c + \sigma_s\tilde{\xi} - g}{\sigma_s\alpha - m}\right) = \exp\left(\frac{\tilde{f} + k\tilde{s}_c + \sigma_s\tilde{\xi} - g}{\sigma_s\alpha - m}\right).$$

The sufficient statistic for the information in $P$, $z\,(P)$, is now expressed as:

$$z(P) \equiv \frac{g + (\sigma_s\alpha - m)\ln(P)}{1+k} = \tilde{f} + \frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{\sigma_s}{1+k}\tilde{\xi} = \left(\frac{1}{1+k}\right)\tilde{f} + \frac{k}{1+k}\tilde{s}_c + \frac{\sigma_s}{1+k}\tilde{\xi}.$$

The capital provider makes his decision based on $z(P)$ and $\tilde{s}_l$, while speculators make their decisions based on $z(P)$, $\tilde{s}_i$, and $\tilde{s}_c$. Denoting the equilibrium weight that speculators put on the common signal $k^{**}$, in the proposed linear equilibrium, the equilibrium weight $k^{**}$ solves $B_P\left(k^{**}\right) = k^{**}$, where $B_P\left(\cdot\right)$ is a best-response function defined similarly to that in the main model. The following proposition derives conditions under which a unique equilibrium exists.

Proposition 9: There is a unique equilibrium if $\tau_{cp} > \tau_c$ (the capital provider's signal is more precise than the speculators' common signal) and $\tau_s$ (the precision of speculators' private signals) or $\tau_\xi$ (the precision of the noise trading distribution) are small enough or if $\tau_{cp}$ (the precision of capital provider's signal) is large enough.

In the rest of this section we will assume that the equilibrium is unique. As in the main model, we now compare the coordination level $k^{**}$ with the one that would be obtained in a benchmark where the capital provider does not learn from the price (but speculators do). In the benchmark, we say that speculators short the asset whenever $\tilde{s}_i + k_N \tilde{s}_c + m_N \ln P \leq g_N$ and buy otherwise. The following proposition compares the weight on the common signal $k_N$ in the benchmark model and the equilibrium weight $k^{**}$.

Proposition 10: When the capital provider does not learn from the price when making the lending decision, the weight that speculators put on the common signal $k_N$ is strictly below the equilibrium weight $k^{**}$ they put when the capital provider learns from the price, if and only if the following condition is true:

$$-2\tau_c \left( \frac{\sqrt{\tau_{cp}}}{\sqrt{\tau_c}} - 1 \right) + \tau_s \left(1 + \tau_\xi\right) < 0, \tag{21}$$

i.e., if $\tau_{cp} > \tau_c$ (the capital provider's signal is more precise than the speculators' common signal) and $\tau_s$ (the precision of speculators' private signals) or $\tau_\xi$ (the precision of the noise trading distribution) are small enough.

To understand the role of strategic complementarities, we plot the best response function in the REE model $B_P(\cdot)$ (as in equation (29) in the appendix) and in the benchmark REE model $B_N(\cdot)$ where the capital provider does not learn from the price (as in equation (32) in the appendix) in the following figure and compare the slopes of these two functions.

In the benchmark case, there is no strategic complementarity. In fact, the slope of $B_N(k)$, is $\partial B_N(k)/\partial k = -\tau_\xi$. That is, when others put a larger weight on the common signal, a speculator's best response is to reduce his weight on the common signal. This is intuitive since when others put more weight on the common signal, the price as a signal for the fundamental becomes more correlated with the common signal. This causes a speculator to reduce the weight he puts on the common signal, since some of this information is already embedded in the price.

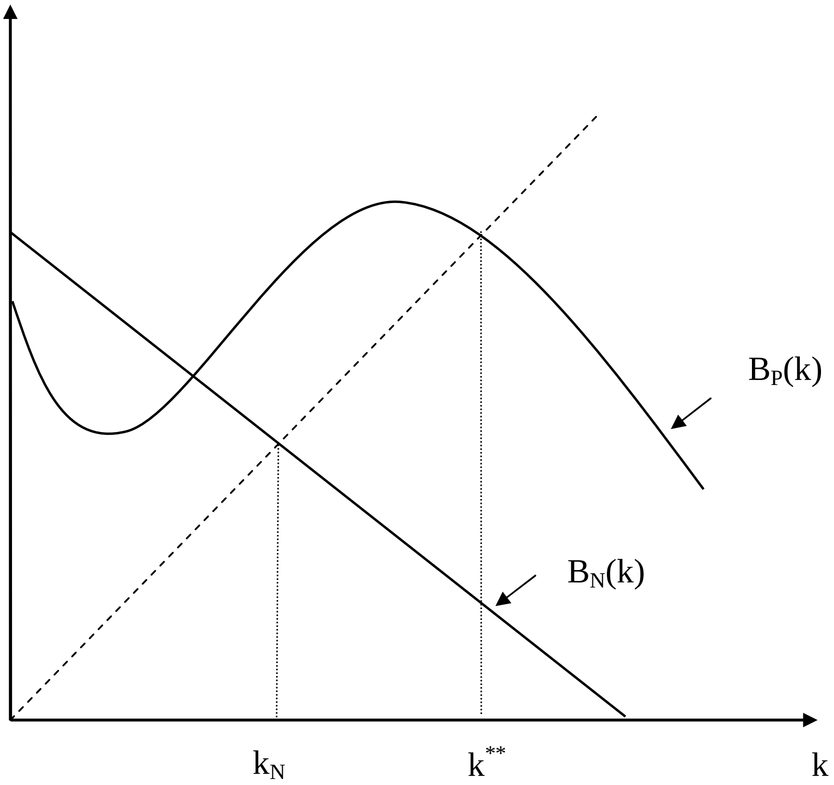



Figure 2: Best Response: $B_P(k)$ and $B_N(k)$

However, when the capital provider learns from the price, a speculator's weight on the common signal is sometimes increasing in others' weight on the common signal. Indeed the slope of $B_P(\cdot)$ in the figure is sometimes positive. The reason is that the weight that others put on the common signal affects the reliance of the capital provider on his own signal $\tilde{s}_l$ and hence affects the real value of the security. In particular, when $k$ goes up, the price becomes less precise to the capital provider and he shifts weights from the price to $\tilde{s}_l$ in his investment decision. Then, a speculator knows that the common signal $\tilde{s}_c$, which is correlated with the capital provider's signal $\tilde{s}_l$, will be more strongly correlated with the investment decision of the capital provider, and this increases the incentive of the speculator to rely on $\tilde{s}_c$. This creates the strategic complementarities when the capital provider learns from the price. Putting this together with the opposite effect in the benchmark (when the capital provider does not learn from the price), the above proposition develops the condition in equation (21) under which $k$ is higher when the capital provider learns from the price. Note that this condition is satisfied when the precision of the capital provider's signal ($\tau_{cp}$) is large relative to the precision of the common signal ($\tau_c$), the precision of the noise trading ($\tau_\xi$), and the precision of the speculator's private signal ($\tau_s$). The latter three precisions are related to the precision of the price since price aggregates speculators' private and common signals as well as the noise trading. Therefore, this condition is related to how heavily the capital provider relies on $\tilde{s}_l$ versus the price in making investment decisions. When the capital provider is prone to rely on $\tilde{s}_l$ more heavily, speculators coordinate more in equilibrium benefiting from his reliance on $\tilde{s}_l$.

In summary, for strategic complementarities to arise in our model due to the informational feedback from prices to capital provision, it is important that speculators do not observe the exact message that the capital provider receives from the market. In our main model, developed in previous sections, this was obtained because the speculators did not observe the price that the capital provider learns from. In the model developed in this section, we let the speculators observe the price, but assume that the capital provider, in addition to observing the price, observes something else which is correlated with the price and is not observed by speculators. This restores the high-order beliefs of the basic model and allows for strategic complementarities to arise.

# 6 Conclusion

We study strategic interactions among speculators in financial markets and their real effects. Two opposite strategic effects exist. On the one hand, speculators wish to act differently from each other as a certain action by other speculators changes the price in a way that reduces the profit for other speculators from this action. On the other hand, due to the feedback effect from the price to the real investment, a certain action by speculators changes the real value of the firm in a way that increases the incentive of other speculators to take this action. This creates a basis for trading frenzies, where speculators rush to trade in the same direction, putting pressure on the price and on the firm's value. We characterize which effect dominates when and analyze the resulting level of coordination in speculators' actions.

The interaction among speculators affects the informational content of the price. Since prices affect real investment in our model, we can ask what level of coordination is most efficient for real investment. In general, speculators' incentives to coordinate go in opposite direction to the optimal level of coordination. Speculators want to coordinate more when there is a low amount of noise trading, but this is when coordination is less desirable from an efficiency point of view. Hence, our model shows that there is always either too much or too little coordination, and this reduces the efficiency of investment and creates excess volatility in the price.

Interestingly, our paper is also related to an old debate on whether speculators stabilize prices. The traditional view is that by buying low and selling dear, rational speculators stabilize prices. Hart and Kreps (1986) argue that when speculators can hold inventories and there is uncertainty about preferences, speculative activity may cause excess price movement. Our paper contributes to this literature by pointing out that when speculative activity has an effect on real investments, speculators might coordinate on correlated sources of information, and create excess volatility in prices. In our model, this reduces the efficiency of real investments.

# References


**Amador, Manuel and Pierre-Olivier Weill**, "Learning from Prices: Public Communication and Welfare," *Journal of Political Economy*, 2011, *forthcoming.*

**Angeletos, George-Marios, Guido Lorenzoni, and Alessandro Pavan**, "Beauty Contests and Irrational Exuberance: A Neoclassical Approach," 2010. MIT working paper.

**Boot, Arnoud and Anjan Thakor**, "Financial system architecture," *Review of Financial Studies*, 1997, *10*, 693–733.

**Bru, Lluis and Xavier Vives**, "Informational externalities, herding and incentives," *Journal of Institutional and Theoretical Economics*, 2002, *158*, 91–105.

**Chen, Qi, Itay Goldstein, and Wei Jiang**, "Price informativeness and investment sensitivity to stock price," *Review of Finacial Studies*, 2007, *20* (3), 619–650.

**Dasgupta, Amil**, "Coordination and delay in global games," *Journal of Economic Theory*, 2007, *134*, 195–225.

**Dow, James and Gary Gorton**, "Stock market efficiency and economic efficiency: Is there a connection?," *Journal of Finance*, 1997, *52*, 1087–1129.

**_ , Itay Goldstein, and Alexander Guembel**, "Incentives for information production in markets where prices affect real investment," 2007. Working Paper.

**Fishman, Mike J. and Kathleen M. Hagerty**, "Insider trading and the efficiency of stock prices," *RAND Journal of Economics*, 1992, *23*, 106–122.

**Froot, Kenneth, David Scharfstein, and Jeremy Stein**, "Herd on the street: informational inefficiencies in a market with short-term speculation," *Journal of Finance*, 1992, *47*, 1461–1484.

**Fulghieri, Paolo and Dmitry Lukin**, "Information production, dilution costs, and optimal security design," *Journal of Financial Economics*, 2001, *61*, 3–42.

**Ganguli, Jayant Vivek and Liyan Yang**, "Complementarities, multiplicity, and supply information," *Journal of the European Economic Association*, 2009, *7*, 90–115.

**Garcia, Diego and Gunter Strobl**, "Relative Wealth Concerns and Complementarities in Information Acquisition," *Review of Financial Studies*, 2011, *24*, 169–207.

**Goldstein, Itay and Alexander Guembel**, "Manipulation and the allocational role of prices," *Review of Economic Studies*, 2008, *75*, 133–164.

_ **, Emre Ozdenoren, and Kathy Yuan**, "Learning and Complementarities in Speculative Attacks," *Review of Economic Studies*, 2011, *forthcoming.*

**Grossman, Sanford and Joseph Stiglitz**, "On the impossibility of informationally efficient markets," *American Economic Review*, 1980, *70*, 393–408.

**Hayek, Friedrich**, "The use of knowledge in society," *American Economic Review*, 1945, *35*, 519–530.

**Hellwig, Christian, Arijit Mukherji, and Aleh Tsyvinski**, "Self-fulfilling currency crises: The role of interest rates," *American Economic Review*, 2006, *96* (5), 1769–87.

**Hirshleifer, David, Avanidhar Subrahmanyam, and Sheridan Titman**, "Security analysis and trading patterns when some investors receive information before others," *Journal of Finance*, 1994, *49* (5), 1665–1698.

**Khanna, Naveen and Ramana Sonti**, "Value creating stock manipulation: Feedback effect of stock prices on firm value," *Journal of Financial Markets*, 2004, *7* (3), 237–270.

_ **, Steve L. Slezak, and Michael H. Bradley**, "Insider trading, outside search and resource allocation: Why firms and society may disagree on insider trading restrictions," *Review of Financial Studies*, 1994, *7* (3), 575–608.

**Kyle, Albert S.**, "Continuous Auctions and Insider Trading," *Econometrica*, 1985, *53*, 1315–1335.

**Leland, Hayne**, "Insider trading: Should it be prohibited?," *Journal of Political Economy*, 1992, *100*, 859–887.

**Luo, Yuanzhi**, "Do insiders learn from outsiders? Evidence from mergers and acquisitions," *Journal of Finance*, 2005, *60*, 1951–1972.

**Ozdenoren, Emre and Kathy Yuan**, "Feedback effects and asset prices," *Journal of Finance*, 2008, *63*, 1939–1975.

**Subrahmanyam, Avanidhar and Sheridan Titman**, "The going-public decision and the development of financial markets," *Journal of Finance*, 1999, *54*, 1045–1082.

**Veldkamp, Laura**, "Information markets and the comovement of asset prices," *Review of Economic Studies*, 2006, *73*, 823–845.

_ , "Media frenzies in markets for financial information," *American Economic Review*, 2006, *96*, 577–601.

**Vives, Xavier**, "How Fast Do Rational Agents Learn?," *Review of Financial Studies*, 1993, *60*, 329–347.

# Appendix

**Proof of Proposition 1:** Based on (12) and the updating done by the speculator based on his information, the coefficients in (13) are given as follows:

$$\begin{aligned}
a_0 &= \frac{\tau_f \bar{f} + \frac{1}{2}}{\tau_f + \tau_l + \tau_p} + \frac{\left(\tau_f + 2\tau_l + \tau_p\left(1 + \frac{1}{1+k}\right)\right)}{\tau_f + \tau_l + \tau_p}\left(\frac{\tau_f \bar{f}}{\tau_f + \tau_s + \tau_c}\right) \\
&+ \left(\frac{\tau_f + 2\tau_l + \left(1 + \frac{1}{1+k}\right)\tau_p}{\tau_f + \tau_l + \tau_p}\right)^2 \frac{1}{2(\tau_f + \tau_s + \tau_c)} + \frac{1}{2}\left(\frac{\tau_l}{\tau_f + \tau_l + \tau_p}\right)^2 \sigma_l^2 \\
&+ \frac{1}{2}\left(\frac{\tau_p}{\tau_f + \tau_l + \tau_p}\right)^2\left(\frac{1}{1+k}\right)^2 \sigma_s^2 \sigma_\xi^2,
\end{aligned}$$

$$a_1 = \frac{\left(\tau_f + 2\tau_l + \tau_p\left(1 + \frac{1}{1+k}\right)\right)}{\tau_f + \tau_l + \tau_p}\frac{\tau_s}{\tau_f + \tau_s + \tau_c},$$

$$a_2 = \frac{\tau_p \frac{k}{1+k}}{\tau_f + \tau_l + \tau_p} + \frac{\left(\tau_f + 2\tau_l + \tau_p\left(1 + \frac{1}{1+k}\right)\right)}{\tau_f + \tau_l + \tau_p}\frac{\tau_c}{\tau_f + \tau_s + \tau_c},$$

$$b_0 = \frac{1}{\alpha\sigma_s}\left(\frac{\tau_f \bar{f} + \frac{1}{2}\frac{1}{\alpha\sigma_s}}{\tau_f + \tau_s + \tau_c} - g\right) + \frac{1}{2\alpha^2}\sigma_\xi^2,$$

$$b_1 = \frac{1}{\alpha\sigma_s}\frac{\tau_s}{\tau_f + \tau_s + \tau_c},$$

$$b_2 = \frac{1}{\alpha\sigma_s}\left(\frac{\tau_c}{\tau_f + \tau_s + \tau_c} + k\right).$$

Note that

$$a_1 - b_1 = \frac{\tau_s}{\tau_f + \tau_s + \tau_c}\left(\frac{\left(\tau_f + 2\tau_l + \tau_p\left(1 + \frac{1}{1+k}\right)\right)}{\tau_f + \tau_l + \tau_p} - \frac{1}{\alpha\sigma_s}\right).$$

A sufficient condition for $a_1 - b_1 > 0$ is that $\alpha > \sqrt{\tau_s}$. Recall that $B(k) = (a_2 - b_2)/(a_1 - b_1)$. Substituting, we obtain:

$$B(k) = \frac{\frac{\tau_p \frac{k}{1+k}}{\tau_f+\tau_l+\tau_p} + \frac{\left(\tau_f+2\tau_l+\tau_p\left(1+\frac{1}{1+k}\right)\right)}{\tau_f+\tau_l+\tau_p}\frac{\tau_c}{\tau_f+\tau_s+\tau_c} - \frac{1}{\alpha\sigma_s}\left(\frac{\tau_c}{\tau_f+\tau_s+\tau_c} + k\right)}{\frac{\tau_s}{\tau_f+\tau_s+\tau_c}\left(\frac{\left(\tau_f+2\tau_l+\tau_p\left(1+\frac{1}{1+k}\right)\right)}{\tau_f+\tau_l+\tau_p} - \frac{1}{\alpha\sigma_s}\right)}.$$

Simplifying $B(k) - k = 0$ we get:

$$\begin{aligned}
0 = &\left[\frac{1}{\tau_s}\frac{(k+1)(\tau_c + \tau_f + \tau_s)(\tau_f + \tau_l + \tau_p)}{\tau_p + (k+1)\tau_l + \left(1 - \frac{\sqrt{\tau_s}}{\alpha}\right)(1+k)(\tau_f + \tau_l + \tau_p)}\right] \\
&\left(\frac{\tau_p}{\tau_f + \tau_l + \tau_p}\frac{k}{1+k} + \left(\frac{\tau_f + 2\tau_l + \left(1 + \frac{1}{1+k}\right)\tau_p}{\tau_f + \tau_l + \tau_p}\right)\left(\frac{-\tau_s k + \tau_c}{\tau_f + \tau_s + \tau_c}\right) - \frac{\sqrt{\tau_s}}{\alpha}\left(\left(\frac{-\tau_s k + \tau_c}{\tau_f + \tau_s + \tau_c}\right) + k\right)\right).
\end{aligned}$$

The term in square brackets is strictly positive for $\alpha > \sqrt{\tau_s}$. So the equilibrium condition can be simplified to:

$$\begin{aligned} 0 &= \frac{\tau_p}{\tau_f+\tau_l+\tau_p}\frac{k}{1+k} + \left(\frac{\tau_f+2\tau_l+\left(1+\frac{1}{1+k}\right)\tau_p}{\tau_f+\tau_l+\tau_p}\right)\left(\frac{-\tau_s k+\tau_c}{\tau_f+\tau_s+\tau_c}\right) - \left(\frac{-\tau_s k+\tau_c}{\tau_f+\tau_s+\tau_c}\right) - k \\ &\quad + \left(1-\frac{1}{\alpha\sigma_s}\right)\left(\frac{\tau_f+\tau_c}{\tau_f+\tau_s+\tau_c}k + \frac{\tau_c}{\tau_f+\tau_s+\tau_c}\right). \end{aligned}$$

Recall that $\tau_p = \left((1+k)^2\tau_c\tau_\xi\tau_s\right)/\left(k^2\tau_\xi\tau_s+\tau_c\right)$. We denote $r \equiv \tau_\xi\tau_s$. Substituting for $\tau_p$ and simplifying, the right-hand side becomes:[12]

$$\begin{aligned} H(k) &= -k^3\left((\tau_c+\tau_f+\tau_s)(\tau_c+\tau_f+\tau_l)+\tau_l\tau_s\right) - \tau_c k^2\left(\tau_c+\tau_f-\tau_l+2\tau_s\right) \\ &\quad - \tau_c k\left(\tau_s-\tau_c\right) + \tau_c^2 - \frac{1}{r}\left(\tau_c k\left(\tau_c\tau_f+\tau_c\tau_l+\tau_f\tau_l+\tau_f\tau_s+2\tau_l\tau_s+\tau_f^2\right)-\tau_c^2\tau_l\right) \\ &\quad + \left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)\left((\tau_c+\tau_f)(\tau_c+\tau_f+\tau_l)k^3+\tau_c(3\tau_c+3\tau_f+\tau_l)k^2\right. \\ &\quad \left.+\tau_c(\tau_f+3\tau_c)k+\tau_c^2\right) + \left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)\frac{\tau_c}{r}\left((\tau_f+\tau_l)(\tau_c+\tau_f)k+\tau_c(\tau_f+\tau_l)\right). \end{aligned} \tag{22}$$

For an equilibrium, we need $H(k) = 0$.

First, we focus on existence of an equilibrium with $k > 0$. $H(k)$ has a positive root if and only if

$$\alpha > \sqrt{\tau_s}\frac{\tau_f+\tau_l+r}{\tau_f+2\tau_l+2r}.$$

To see this, note that the coefficient for $k^3$ is always negative, implying that the value of $H(k)$ becomes negative as $k$ becomes large. So, there exists a strictly positive root for the polynomial if its value at $k=0$ is strictly positive. This condition is given by the above inequality. If the inequality is violated, the value of the polynomial is negative at $k=0$. Its derivative at $k=0$ is given by

$$\begin{aligned} &-\tau_c(\tau_s-\tau_c) - \frac{1}{r}\left(\tau_c\left(\tau_c\tau_f+\tau_c\tau_l+\tau_f\tau_l+\tau_f\tau_s+2\tau_l\tau_s+\tau_f^2\right)\right) \\ &+\left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)\tau_c(\tau_f+3\tau_c) + \left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)\frac{\tau_c}{r}\left((\tau_f+\tau_l)(\tau_c+\tau_f)\right). \end{aligned}$$

At $\frac{\sqrt{\tau_s}}{\alpha} \geq \frac{\tau_f+2\tau_l+2r}{\tau_f+\tau_l+r}$ the derivative is negative. This means that $H(k)$ is decreasing at $k=0$ for $\frac{\sqrt{\tau_s}}{\alpha} \geq \frac{\tau_f+2\tau_l+2r}{\tau_f+\tau_l+r}$. Moreover, the second derivative is negative when $\frac{\sqrt{\tau_s}}{\alpha} \geq \frac{\tau_f+2\tau_l+2r}{\tau_f+\tau_l+r}$, and thus the expression will keep decreasing. Therefore the polynomial cannot have a positive root.

[12] The simplification is achieved by dividing by $r$ and multiplying through with $(\tau_c+\tau_f+\tau_s)\left(\tau_c\tau_f+\tau_c\tau_l+\tau_c r+\tau_c k^2 r+\tau_f k^2 r+k^2\tau_l r+2\tau_c k r\right)$.

For uniqueness, we need to check the sign of the discriminant of $H(k)$ for $\alpha$ large and $\tau_f$ small. Letting $\alpha$ go to infinity and $\tau_f$ go to zero we obtain the discriminant as $-\frac{\tau_c^3}{r^3}$ times

$$\begin{aligned}
&\left(\begin{array}{c} 64\tau_c^4\tau_l + 128\tau_c^3\tau_l^2 + 160\tau_c^3\tau_l\tau_s + 64\tau_c^2\tau_l^3 + 416\tau_c^2\tau_l^2\tau_s + 144\tau_c^2\tau_l\tau_s^2 \\ +284\tau_c\tau_l^2\tau_s^2 + 56\tau_c\tau_l\tau_s^3 + 8\tau_l\tau_s^4 \end{array}\right) r^3 \qquad (23)\\
&+\left(\begin{array}{c} 64\tau_c^5\tau_l + 192\tau_c^4\tau_l^2 + 160\tau_c^4\tau_l\tau_s + 192\tau_c^3\tau_l^3 + 608\tau_c^3\tau_l^2\tau_s + 144\tau_c^3\tau_l\tau_s^2 + 64\tau_c^2\tau_l^4 \\ +448\tau_c^2\tau_l^3\tau_s + 440\tau_c^2\tau_l^2\tau_s^2 + 56\tau_c^2\tau_l\tau_s^3 + 416\tau_c\tau_l^3\tau_s^2 + 80\tau_c\tau_l^2\tau_s^3 + 8\tau_c\tau_l\tau_s^4 + 48\tau_l^2\tau_s^4 \end{array}\right) r^2 \\
&+\left(128\tau_c\tau_l^4\tau_s^2 - 32\tau_c^2\tau_l^3\tau_s^2 - 16\tau_c^2\tau_l^2\tau_s^3 - 52\tau_c^3\tau_l^2\tau_s^2 - 64\tau_c\tau_l^3\tau_s^3 + 32\tau_c\tau_l^2\tau_s^4 + 96\tau_l^3\tau_s^4\right) r \\
&+\left(64\tau_l^4\tau_s^4 + 32\tau_c\tau_l^3\tau_s^4\right).
\end{aligned}$$

The coefficient of $r^2$ in (23) is strictly positive so the quadratic part of (23) is minimized at:

$$r = -\frac{128\tau_c\tau_l^4\tau_s^2 - 32\tau_c^2\tau_l^3\tau_s^2 - 16\tau_c^2\tau_l^2\tau_s^3 - 52\tau_c^3\tau_l^2\tau_s^2 - 64\tau_c\tau_l^3\tau_s^3 + 32\tau_c\tau_l^2\tau_s^4 + 96\tau_l^3\tau_s^4}{2\left(\begin{array}{c} 64\tau_c^5\tau_l + 192\tau_c^4\tau_l^2 + 160\tau_c^4\tau_l\tau_s + 192\tau_c^3\tau_l^3 + 608\tau_c^3\tau_l^2\tau_s + 144\tau_c^3\tau_l\tau_s^2 + 64\tau_c^2\tau_l^4 \\ +448\tau_c^2\tau_l^3\tau_s + 440\tau_c^2\tau_l^2\tau_s^2 + 56\tau_c^2\tau_l\tau_s^3 + 416\tau_c\tau_l^3\tau_s^2 + 80\tau_c\tau_l^2\tau_s^3 + 8\tau_c\tau_l\tau_s^4 + 48\tau_l^2\tau_s^4 \end{array}\right)}.$$

Substituting this back to the quadratic above we find that the minimized value is:

$$\begin{aligned}
&\frac{1}{2}\frac{\tau_l^3\tau_s^4}{\left(\begin{array}{c} 8\tau_c^5 + 24\tau_c^4\tau_l + 20\tau_c^4\tau_s + 24\tau_c^3\tau_l^2 + 76\tau_c^3\tau_l\tau_s + 18\tau_c^3\tau_s^2 + 8\tau_c^2\tau_l^3 + 56\tau_c^2\tau_l^2\tau_s \\ +55\tau_c^2\tau_l\tau_s^2 + 7\tau_c^2\tau_s^3 + 52\tau_c\tau_l^2\tau_s^2 + 10\tau_c\tau_l\tau_s^3 + \tau_c\tau_s^4 + 6\tau_l\tau_s^4 \end{array}\right)} \\
&\times\left(\begin{array}{c} 343\tau_c^6 + 2352\tau_c^5\tau_l + 1176\tau_c^5\tau_s + 5376\tau_c^4\tau_l^2 + 6944\tau_c^4\tau_l\tau_s + 1344\tau_c^4\tau_s^2 + 4096\tau_c^3\tau_l^3 \\ +13312\tau_c^3\tau_l^2\tau_s + 6448\tau_c^3\tau_l\tau_s^2 + 512\tau_c^3\tau_s^3 + 8192\tau_c^2\tau_l^3\tau_s + 9984\tau_c^2\tau_l^2\tau_s^2 + 1984\tau_c^2\tau_l\tau_s^3 \\ +5120\tau_c\tau_l^3\tau_s^2 + 2048\tau_c\tau_l^2\tau_s^3 + 128\tau_c\tau_l\tau_s^4 + 192\tau_l^2\tau_s^4 \end{array}\right)
\end{aligned}$$

which is strictly positive. Since the quadratic term is strictly positive at its minimum, it is positive for all $r$. Since $r^3$ term is positive for $r > 0$ as well, (23) is strictly positive for all $r > 0$. That is, the discriminant is strictly negative for large enough $\alpha$ and small enough $\tau_f$, and hence $H(k) = 0$ has a unique root. QED.

**Proof of Propositions 2:** First, we derive $k_{BM}$. Based on (16) and taking expectations, we see that a speculator buys the asset when:

$$\begin{aligned}
&\ln\left(\frac{1}{c}\right) + \frac{\tau_f\bar{f} + \frac{1}{2}}{\tau_f + \tau_l} + \left(\frac{\tau_f + 2\tau_l}{\tau_f + \tau_l}\right)\left(\frac{\tau_f\bar{f} + \tau_s\tilde{s}_i + \tau_c\tilde{s}_c}{\tau_f + \tau_s + \tau_c}\right) \qquad (24)\\
&+\left(\frac{\tau_f + 2\tau_l}{\tau_f + \tau_l}\right)^2\frac{1}{2(\tau_f + \tau_s + \tau_c)} + \frac{1}{2}\left(\frac{\tau_l}{\tau_f + \tau_l}\right)^2\sigma_l^2 \\
\geq\ &\frac{1}{\alpha\sigma_s}\left(\frac{\tau_f\bar{f} + \tau_s\tilde{s}_i + \tau_c\tilde{s}_c + \frac{1}{2}\frac{1}{\alpha\sigma_s}}{\tau_f + \tau_s + \tau_c} + k_{BM}\tilde{s}_c - g_{BM}\right) + \frac{1}{2\alpha^2}\sigma_\xi^2.
\end{aligned}$$

Rearranging (24), a speculator buys the asset when $\tilde{s}_i + B_{BM}(k_{BM})\,\tilde{s}_c \geq C_{BM}$ where

$$B_{BM}(k_{BM}) = \frac{\tau_c}{\tau_s} - \frac{\frac{\sqrt{\tau_s}}{\alpha}k_{BM}}{\frac{\tau_s}{\tau_f+\tau_s+\tau_c}\left(\frac{\tau_f+2\tau_l}{\tau_f+\tau_l} - \frac{\sqrt{\tau_s}}{\alpha}\right)}$$

and

$$\begin{aligned} C_{BM} \;=\; & \frac{1}{\left(\frac{\tau_f+2\tau_l}{\tau_f+\tau_l} - \frac{1}{\alpha\sigma_s}\right)\left(\frac{\tau_s}{\tau_f+\tau_s+\tau_c}\right)} \\ & \left(\ln c + \frac{1}{\alpha\sigma_s}\left(\frac{\tau_f\bar{f} + \frac{1}{2}\frac{1}{\alpha\sigma_s}}{\tau_f+\tau_s+\tau_c} - g_{BM}\right) - \frac{\tau_f\bar{f}+\frac{1}{2}}{\tau_f+\tau_l} - \left(\frac{\tau_f+2\tau_l}{\tau_f+\tau_l}\right)\left(\frac{\tau_f\bar{f}}{\tau_f+\tau_s+\tau_c}\right)\right. \\ & \left. - \left(\frac{\tau_f+2\tau_l}{\tau_f+\tau_l}\right)^2 \frac{1}{2(\tau_f+\tau_s+\tau_c)} + \frac{1}{2\alpha^2}\sigma_\xi^2 - \frac{1}{2}\left(\frac{\tau_l}{\tau_f+\tau_l}\right)^2\sigma_l^2\right). \end{aligned}$$

Setting $B_{BM}(k_{BM}) = k_{BM}$ leads to the expression for $k_{BM}$ in (18).

Now, we show that in the main model (with feedback effect) $k^* > \tau_c/\tau_s$ for $\alpha$ large enough. To see this note that $H(\tau_c/\tau_s) > 0$ for $\alpha$ large enough. Since $H(k)$ has a unique root and crosses the axis from above, the conclusion follows. Next, note that $k_{BM} < \tau_c/\tau_s$ and thus $k_{BM} < k^*$ for $\alpha$ large enough. QED.

**Proof of Proposition 3:** We showed in the proof of Proposition 2 that $k^* > \tau_c/\tau_s$ for $\alpha$ large enough. By inspecting (22), we can see that $H(k)$ shifts up as $\alpha$ increases, so its unique root $k^*$ increases in $\alpha$. QED.

**Proof of Proposition 4:**

Consider the following terms involving $1/r$ in $H(k)$ in (22):

$$\begin{aligned} & -\frac{1}{r}\left(\tau_c k(\tau_c\tau_f + \tau_c\tau_l + \tau_f\tau_l + \tau_f\tau_s + 2\tau_l\tau_s + \tau_f^2) - \tau_c^2\tau_l\right) \\ & + \left(1 - \frac{\sqrt{\tau_s}}{\alpha}\right)\frac{\tau_c}{r}\left((\tau_f+\tau_l)(\tau_c+\tau_f)k + \tau_c(\tau_f+\tau_l)\right). \end{aligned}$$

For $\alpha$ large enough, these terms are negative iff $k$ exceeds $\tau_c/\tau_s$. So for $k > \tau_c/\tau_s$, $H(k)$ shifts up as $r$ goes up. By Proposition 3, for $\alpha$ large enough, $k^*$ which implicitly depends on $r$ exceeds $\tau_c/\tau_s$ for all $r$. Since $H(k)$ crosses the axis once from above at $k^*$, we see that $k^*$ must be increasing in $r$. Since increasing $\sigma_\xi$ and $r$ are inversely related, an increase in $\sigma_\xi$ leads to a decrease in $k^*$. QED.

**Proof of Proposition 5**

Let

$$D(k) = -3\left((\tau_f+\tau_c+\tau_l)(\tau_f+\tau_c+\tau_s)+\tau_l\tau_s\right)k^2 - 2\tau_c(\tau_f+\tau_c-\tau_l+2\tau_s)k$$
$$+\tau_c(\tau_c-\tau_s) - \frac{1}{r}\left(\tau_c(\tau_f\tau_c+\tau_f\tau_l+\tau_f\tau_s+\tau_c\tau_l+2\tau_l\tau_s+\tau_f^2)\right)$$
$$+\left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)\left(3(\tau_f+\tau_c)(\tau_f+\tau_c+\tau_l)k^2+2\tau_c(3\tau_f+3\tau_c+\tau_l)k+\tau_c(\tau_f+3\tau_c)\right)$$
$$+\frac{\tau_c}{r}\left(1-\frac{\sqrt{\tau_s}}{\alpha}\right)(\tau_f+\tau_c)(\tau_f+\tau_l)$$

Note that $\partial H/\partial k = D(k)$. When the equilibrium is unique $D(k^*) < 0$ since $H(k)$ crosses zero from above.

To see that for $\alpha$ large $\frac{\partial k^*}{\partial \tau_s} < 0$,note for $\alpha$ large, $\frac{\partial k^*}{\partial \tau_s}$ is arbitrarily close to

$$\frac{\frac{1}{\tau_s^2\tau_\xi}\left(\tau_c^2\tau_f+2\tau_c^2\tau_l+\tau_c\tau_s^2\tau_\xi k^*+2\tau_c\tau_s^2\tau_\xi\left(k^*\right)^2+\left(\tau_c\tau_s^2\tau_\xi+\tau_f\tau_s^2\tau_\xi+2\tau_l\tau_s^2\tau_\xi\right)\left(k^*\right)^3\right)}{D(k^*)} < 0.$$

To see that for $\alpha$ large $\frac{\partial k^*}{\partial \tau_c} > 0$,note for $\alpha$ large, $\frac{\partial k^*}{\partial \tau_c}$ is arbitrarily close to

$$\frac{\left(\begin{array}{c}\tau_s\left(k^*\right)^3-2(2\tau_c+\tau_f+\tau_l-\tau_s)\left(k^*\right)^2-(8\tau_c+\tau_f-\tau_s)k^*-4\tau_c\\ -\frac{1}{r}\left(-\tau_s(\tau_f+2\tau_l)k^*+2\tau_c\tau_f+4\tau_c\tau_l\right)\end{array}\right)}{D\left(k^*\right)}.$$

Also,

$$\tau_c\left(-\tau_s\left(k^*\right)^3+2\left(2\tau_c+\tau_f+\tau_l-\tau_s\right)\left(k^*\right)^2+\left(8\tau_c+\tau_f-\tau_s\right)k^*+4\tau_c\right.$$
$$\left.+\frac{1}{r}\left(-\tau_s\left(\tau_f+2\tau_l\right)k^*+2\tau_c\tau_f+4\tau_c\tau_l\right)\right) > -\tau_s\left(\tau_c+\tau_f+2\tau_l\right)\left(k^*\right)^3$$
$$+2\tau_c\left(\tau_f+\tau_l-\tau_s+\tau_c\right)\left(k^*\right)^2+\tau_c\left(\tau_f-\tau_s+4\tau_c\right)k^*+2\tau_c^2$$
$$+\frac{\tau_c}{r}\left(\tau_c\tau_f+2\tau_c\tau_l-\tau_s\left(\tau_f+2\tau_l\right)k^*\right)$$

and the right hand side of the above inequality is arbitrarily close to $H(k^*)$ which is equal to zero.

Finally, to see that for $\alpha$ large $\frac{\partial k^*}{\partial \tau_l} < 0$,not for $\alpha$ large, $\frac{\partial k^*}{\partial \tau_l}$ is arbitrarily close to

$$\frac{\frac{2}{r}(k^*\tau_s-\tau_c)(r\left(k^*\right)^2+\tau_c)}{D\left(k^*\right)} < 0$$

since $k^* > \tau_c/\tau_s$. QED

**Proof of Proposition 6:**

We substitute $I$ equation (2) into equation (19) and compute the expectations:

$$
\begin{aligned}
& \frac{1}{c}E\left[\exp\left(\tilde{f}\right)\exp\left(\frac{\tau_f\bar{f}+\tau_l s_l+\tau_p z\left(P\right)}{\tau_f+\tau_l+\tau_p}+\frac{1}{2(\tau_f+\tau_l+\tau_p)}\right)\right] \\
& -\frac{1}{2}\frac{1}{c}E\left[\exp\left(2\left(\frac{\tau_f\bar{f}+\tau_l s_l+\tau_p z\left(P\right)}{\tau_f+\tau_l+\tau_p}+\frac{1}{2(\tau_f+\tau_l+\tau_p)}\right)\right)\right] \\
= & \frac{1}{c}E\left[\exp\left(2\tilde{f}+\frac{\tau_f\left(\bar{f}-\tilde{f}\right)+\tau_l\sigma_l\tilde{\epsilon}_l+\tau_p\left(z\left(P\right)-\tilde{f}\right)}{\tau_f+\tau_l+\tau_p}+\frac{1}{2(\tau_f+\tau_l+\tau_p)}\right)\right] \\
& -\frac{1}{2}\frac{1}{c}E\left[\exp\left(2\left(\tilde{f}+\frac{\tau_f\left(\bar{f}-\tilde{f}\right)+\tau_l\sigma_l\tilde{\epsilon}_l+\tau_p\left(z\left(P\right)-\tilde{f}\right)}{\tau_f+\tau_l+\tau_p}+\frac{1}{2(\tau_f+\tau_l+\tau_p)}\right)\right)\right] \\
= & \frac{1}{c}\frac{1}{2}\exp\left(2\bar{f}+\frac{1}{\tau_f}\frac{\tau_f+2\tau_l+2\tau_p}{\tau_f+\tau_l+\tau_p}\right).
\end{aligned}
$$

Therefore the maximization problem can be viewed as maximizing the following expression in $k$ :

$$\exp\left(\frac{\tau_f+2\tau_l+2\tau_p}{\tau_f+\tau_l+\tau_p}\right),$$

and this is equivalent to maximizing $\tau_p$ which is maximized at $\tau_c/\left(\tau_s\tau_\xi\right)$. QED.

**Proof of Proposition 7:**

For $\alpha$ large enough $H(k)$ evaluated at $k_{OP}=\tau_c/r$ is approximately:

$$\frac{\tau_c^2}{r^3}\left(\tau_c+r\right)\left(2\tau_c r-\tau_c\tau_s+2\tau_f r-\tau_f\tau_s+2\tau_l r-2\tau_l\tau_s-r\tau_s+2r^2\right)$$

which is negative for $r$ small. Moreover it may be decreasing in $r$ for $r$ small but eventually increases and becomes positive. This means that there is a cutoff $\bar{r}$ for $r$ such that for $r<\bar{r}$ we have $k^*<k_{OP}$ and for $r>\bar{r}$ we have $k^*>k_{OP}$. QED.

**Proof of Proposition 8:** (a) The market clearing price is

$$P=\exp\left(\frac{\left(1+k\right)\tilde{f}+k\sigma_c\tilde{\epsilon}_c-g+\sigma_s\tilde{\xi}}{\alpha\sigma_s}\right),$$

and its non-fundamental volatility can be written as the volatility of the following:

$$z(P)-\tilde{f}=\frac{g+\alpha\sigma_s\ln(P)}{1+k}-\tilde{f}=\frac{k}{1+k}\sigma_c\tilde{\epsilon}_c+\frac{\sigma_s}{1+k}\tilde{\xi}.$$

It is straightforward to show that when $k=k_{OP}=\tau_c/\left(\tau_s\tau_\xi\right)$, its non-fundamental volatility is the lowest and is

$$\text{Non-Fundamental Volatility (Asset Price)}=\frac{1}{\tau_c+\tau_s\tau_\xi}.$$

(b) We know that:

$$I = \frac{1}{c}\exp\left(\frac{\tau_f \bar{f} + \tau_l s_l + \tau_p\left(\tilde{f} + \frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{\sigma_s}{1+k}\tilde{\xi}\right)}{\tau_f + \tau_l + \tau_p} + \frac{1}{2(\tau_f + \tau_l + \tau_p)}\right).$$

Taking logs on both sides, we obtain:

$$\ln I = \ln\left(\frac{1}{c}\right) + \left(\frac{\tau_f \bar{f} + \tau_l s_l + \tau_p\left(\tilde{f} + \frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{\sigma_s}{1+k}\tilde{\xi}\right)}{\tau_f + \tau_l + \tau_p} + \frac{1}{2(\tau_f + \tau_l + \tau_p)}\right).$$

We can define the non-fundamental volatility of the real investment as the volatility of the following:

$$\frac{(\tau_f + \tau_l + \tau_p)\left(\ln I - \ln\left(\frac{1}{c}\right)\right) - \frac{1}{2} - \tau_f \bar{f}}{\tau_l + \tau_p} - \tilde{f} = \frac{\tau_l \sigma_l \epsilon_l + \tau_p\left(\frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{\sigma_s}{1+k}\tilde{\xi}\right)}{\tau_l + \tau_p}$$

It is straightforward to show that when $k = k_{OP} = \tau_c / (\tau_s \tau_\xi)$, $\tau_p = \tau_c + \tau_s \tau_\xi$,and the non-fundamental volatility of the real investment is the lowest which is

$$\text{Non-Fundamental Volatility (Real Investment)} = \frac{1}{\tau_l + \tau_c + \tau_s \tau_\xi}.$$

QED.

**Proof of Proposition 9:** In this proof we use the notation $\rho = \sqrt{\tau_{cp}/\tau_c}$. Since $\tau_{cp} > \tau_c$, we have $\rho > 1$. We start with the capital provider's decision. The capital provider updates his belief based on observing:

$$z(P) = \tilde{f} + \frac{k}{1+k}\sigma_c\tilde{\epsilon}_c + \frac{\sigma_s}{1+k}\tilde{\xi} \text{ and } \tilde{s}_l = \tilde{f} + \sigma_{cp}\tilde{\epsilon}_c.$$

This is equivalent to observing $\tilde{s}_l$ and:

$$z(P) - \frac{k}{1+k}\frac{\sigma_c}{\sigma_{cp}}\tilde{s}_l = \left(1 - \frac{k}{1+k}\frac{\sigma_c}{\sigma_{cp}}\right)\tilde{f} + \frac{\sigma_s}{1+k}\tilde{\xi}.$$

Capital provider's conditional belief on $\tilde{f}$ is distributed normally with mean:

$$\left(\frac{\left(1 - \frac{k}{1+k}\rho\right)}{\left(\frac{k}{1+k}\rho - 1\right)^2 + \frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right) z(P) + \left(1 + \frac{\left(\frac{k}{1+k}\rho - 1\right)}{\left(\frac{k}{1+k}\rho - 1\right)^2 + \frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\tilde{s}_l$$

and variance:

$$\Omega = \frac{1}{\tau_s\tau_\xi\left(k(\rho - 1) - 1\right)^2 + \tau_{cp}}.$$

Using the capital provider's investment rule and taking expectations, we can express the level of investment as:

$$\begin{aligned} I &= \frac{1}{c}E[\tilde{F}|s_l,P] = \frac{1}{c}E[\exp\left(\tilde{f}\right)|s_l,P] \qquad (25) \\ &= \frac{1}{c}\exp\left(\left(\frac{\left(1-\frac{k}{1+k}\rho\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)z(P)+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)s_l+\frac{1}{2\Omega}\right) \end{aligned}$$

Given the capital provider's investment policy in (25) and the price, we can now write speculator $i$'s expected profit from buying the asset given the information that is available to him (shorting the asset would give the negative of this):

$$E\left[\tilde{F}I-P|s_i,s_c,P\right] = \qquad (26)$$

$$\frac{1}{c}E\left[\exp\left(\begin{array}{c}\tilde{f}+\frac{1}{2\Omega}+ \\ \left(\frac{\left(1-\frac{k}{1+k}\rho\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)z(P)+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)s_l\end{array}\right)\middle|\begin{array}{c}s_i,\\ s_c,\\ P\end{array}\right]-P$$

To solve for the speculators' conditional expectation, note that since speculators know both $P$ and $s_c$, that means that they observe:

$$g+(\sigma_s\alpha-m)\ln(P)-ks_c=\tilde{f}+\sigma_s\tilde{\xi}. \qquad (27)$$

Therefore, conditional on observing $s_i$, $s_c$ and $P$ speculator $i$ believes that $\tilde{f}$ is distributed normally with mean

$$\frac{\tau_s}{\tau_s+\tau_c+\tau_s\tau_\xi}s_i+\frac{\tau_c}{\tau_s+\tau_c+\tau_s\tau_\xi}s_c+\frac{\tau_s\tau_\xi}{\tau_s+\tau_c+\tau_s\tau_\xi}\left(g+(\sigma_s\alpha-m)\ln(P)-ks_c\right)$$

and precision $\tau_s+\tau_c+\tau_s\tau_\xi$. Moreover,

$$\sigma_{cp}\tilde{\epsilon}_c=\frac{\sigma_{cp}}{\sigma_c}\left(s_c-\tilde{f}\right).$$

Now we take expectation in (26) and note that a speculator would purchase the asset if and only if his expected profit is no less than zero:

$$\frac{1}{c}\exp\left(\begin{array}{c}\left(\frac{\left(1-\frac{k}{1+k}\rho\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)z(P)+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\frac{1}{\rho}s_c \\ +\left(\frac{\tau_s}{\tau_s+\tau_c+\tau_s\tau_\xi}s_i+\frac{\tau_c}{\tau_s+\tau_c+\tau_s\tau_\xi}s_c+\frac{\tau_s\tau_\xi}{\tau_s+\tau_c+\tau_s\tau_\xi}\left(g+(\sigma_s\alpha-m)\ln(P)-ks_c\right)\right) \\ \left(1+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\left(1-\frac{1}{\rho}\right)\right) \\ +\text{variance terms}\end{array}\right)$$

$$-P\geq 0 \qquad (28)$$

Condition (28) can be rewritten as

$$s_i + B_P(k)\, s_c \geq A \ln P + B,$$

where $B_P(k)$ is the best response function given by

$$B_P(k) = \tag{29}$$
$$\frac{\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\frac{1}{\rho}+\left(\frac{\frac{\tau_{cp}}{\rho^2}-k\tau_s\tau_\xi}{\tau_s+\frac{\tau_{cp}}{\rho^2}+\tau_s\tau_\xi}\right)\left(1+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\left(1-\frac{1}{\rho}\right)\right)}{\left(\frac{\tau_s}{\tau_s+\frac{\tau_{cp}}{\rho^2}+\tau_s\tau_\xi}\left(1+\left(1+\frac{\left(\frac{k}{1+k}\rho-1\right)}{\left(\frac{k}{1+k}\rho-1\right)^2+\frac{1}{(1+k)^2}\frac{\tau_{cp}}{\tau_s\tau_\xi}}\right)\left(1-\frac{1}{\rho}\right)\right)\right)}.$$

Setting $B_P(k) - k = 0$ gives us the following equilibrium condition:

$$\frac{C(k)}{\tau_s\tau_\xi\rho\,(k(\rho-1)-1)\,(2k(\rho-1)-1)+\tau_{cp}(2\rho-1)} = 0 \tag{30}$$

where

$$\begin{aligned} C(k) &= \left(-2\tau_s^2\tau_\xi\rho^2(\rho-1)^2(\tau_\xi+1)\right)k^3+\left(\tau_s\tau_\xi(\rho-1)\left(-\tau_{cp}+2\tau_{cp}\rho+4\tau_s\rho^2+4\tau_s\tau_\xi\rho^2\right)\right)k^2 \\ &\quad+\left(2\tau_{cp}\tau_s\tau_\xi-2\tau_s^2\tau_\xi\rho^2-2\tau_{cp}\tau_s\tau_\xi\rho^2-2\tau_{cp}\tau_s\tau_\xi\rho-2\tau_s^2\tau_\xi^2\rho^2-2\tau_{cp}\tau_s\rho^2+\tau_{cp}\tau_s\rho\right)k \\ &\quad+\left(2\tau_{cp}^2+\tau_{cp}\tau_s\rho+\tau_{cp}\tau_s\tau_\xi+\tau_{cp}\tau_s\tau_\xi\rho\right). \end{aligned}$$

First we show that the denominator of (30) is strictly positive for $\tau_s$ or $\tau_\xi$ small or $\tau_{cp}$ large enough. This is because the denominator is minimized at $k = 3/(4(\rho-1))$ and the value of the denominator at that point is $(2\rho-1)\tau_{cp} - (1/8)\tau_s\tau_\xi\rho$ which is strictly positive if $\tau_s$ or $\tau_\xi$ small or $\tau_{cp}$ large enough. So the equilibrium condition becomes $C(k) = 0$. Finally, we show that $C(k) = 0$ has a unique strictly positive root by verifying that the discriminant of $C(k)$ is negative if $\rho > 1$ and $\tau_s$ or $\tau_\xi$ are small or $\tau_{cp}$ is large. QED.

**Proof of Proposition 10:** In this proof we use the notation $\rho = \sqrt{\tau_{cp}/\tau_c}$ that was introduced in the previous proof. Following steps similar to the above we see that a speculator buys the asset if and only if

$$\frac{1}{c}\exp\left(\frac{\sqrt{\tau_c}}{\sqrt{\tau_{cp}}}s_c+\left(2-\frac{\sqrt{\tau_c}}{\sqrt{\tau_{cp}}}\right)\left(\begin{array}{c}\frac{\tau_s}{\tau_s+\tau_c+\tau_s\tau_\xi}s_i+\frac{\tau_c}{\tau_s+\tau_c+\tau_s\tau_\xi}s_c \\ +\frac{\tau_s\tau_\xi}{\tau_s+\tau_c+\tau_s\tau_\xi}\left(g_N+(\sigma_s\alpha-m_N)\ln(P)-k_N s_c\right)\end{array}\right) \atop +\text{variance terms}\right)-P\geq 0. \tag{31}$$

Condition (31) can be rewritten as

$$s_i + B_N(k)\, s_c \geq A_N \ln P + B_N,$$

where $B_N(k)$ is the best response function given by:

$$B_N(k) = \frac{\left(2 - \frac{1}{\rho}\right)\left(\frac{\frac{\tau_{cp}}{\rho^2} - k\tau_s\tau_\xi}{\tau_s + \frac{\tau_{cp}}{\rho^2} + \tau_s\tau_\xi}\right) + \frac{1}{\rho}}{\left(2 - \frac{1}{\rho}\right)\left(\frac{\tau_s}{\tau_s + \frac{\tau_{cp}}{\rho^2} + \tau_s\tau_\xi}\right)}. \tag{32}$$

In the benchmark equilibrium, $k_N$ solves $B_N(k_N) = k_N$. Thus,

$$k_N = \left(\frac{1}{2\rho - 1}\right)\left(\frac{\tau_{cp}}{\tau_s(1 + \tau_\xi)}\frac{2}{\rho} + 1\right).$$

Recall that $k^{**}$ satisfies $C(k^{**}) = 0$ where $C(\cdot)$ is defined in the proof of Proposition 9. Note that

$$C(k_N) = -\frac{2}{\tau_s}\frac{\tau_\xi\left(\tau_{cp} + \tau_s\rho^2 + \tau_s\tau_\xi\rho^2\right)^2\left(2\tau_{cp} - 2\tau_{cp}\rho + \tau_s\rho^2 + \tau_s\tau_\xi\rho^2\right)}{\rho^2(2\rho - 1)^3(\tau_\xi + 1)^2}.$$

Since $C(0) > 0$ and $C(\cdot)$ crosses zero from above this implies that $k^* > k_N$ if and only if $C(k_N) > 0$. That is if and only if

$$2\tau_{cp} - 2\tau_{cp}\rho + \tau_s\rho^2 + \tau_s\tau_\xi\rho^2 < 0.$$

Substituting $\sqrt{\tau_{cp}/\tau_c}$ for $\rho$, we obtain the condition in the statement of the proposition.

QED.