Are Security Lending Fees Priced?

Theory and Evidence from the U.S. Treasury Market

Amrut Nashikkar*

November 26, 2007

Abstract

I study the extent to which security lending fees affect prices in the context of search frictions in the repo market, and make three main theoretical contributions to the literature on short-selling: 1) I incorporate heterogeneity in investors' access to the repo market, and show that securities become slowly "locked up" from the repo market as short interest builds up; 2) I provide testable predictions that distinguish whether short selling is due to hedging or arbitrage activity; 3) I show that when short-selling is driven by the hedging, the proportion of observed future lending fees that is priced is less than one and decreases as short interest builds up. I provide new stylized facts and test the model's implications using repo-rate data from regular US Treasury auctions. I find evidence that short-selling in on-the-run US treasuries is motivated by hedging rather than arbitrage activity. The hedging-based model matches a number of empirical patterns in prices and repo fees over regular auction cycles in US Treasury notes.

Keywords: Fixed Income, Repo Markets, Liquidity, Derivatives

*I am grateful for guidance provided by my dissertation committee - Marti Subrahmanyam (Chair), Lasse Pedersen, Raghu Sundaram, Stephen Figlewski, Robert Engle and Robert Whitelaw, and also from Jeff Wurgler, Yakov Amihud - that was instrumental in bringing this work to its current form. I am also grateful for initial discussions with Pierre-Olivier Weill on an earlier version of this paper. Thanks are also due to Rik Sen and Bryan Kelly for their help.

1 Introduction

This paper studies how securities lending fees affect securities prices. I contrast short-selling demand for a security that arises because it offers hedging benefits with short selling demand that arises because of arbitrage activity. I study the implications of both these channels in an equilibrium model where short sales in a security are restricted by the need to search for security lenders. Search frictions give rise to lending fees and price premia in securities. I show that when short selling is induced by hedging needs, the observed price premium is lower than the present value of all future lending fees through the investment horizon. In contrast, when it is driven by arbitrageurs trying to benefit from over-pricing in the security, the observed price premium is higher than the present value of lending fees. The model features heterogeneity in the participation of securities holders in repo markets, and explains why borrowers tend to perceive securities as being "difficult to locate" as short interest builds up. I show that the proportion of future lending fees incorporated into the price decreases as time passes and derive time-series patterns in prices, repo fees and short interest following the issuance of a security. The empirical tests of these predictions use "on-the-run" U.S. Treasury securities. The results imply that the short-selling of Treasury securities is primarily driven by hedging needs, rather than arbitrageurs benefiting from the on-the-run price premium.

Security borrowing and lending markets play a vital role in securities markets because agents desiring to sell a security short have to borrow it against collateral. A sale of the borrowed security to another investor creates the short position in that security. The need to sell securities short may arise because some agents desire to hedge against risk in other similar portfolios in which they have long positions, or because investors perceive that a security is being overvalued by investors holding a long position in the security. In either case, it can be costly to borrow securities when the demand to short-sell is high. When the demand to borrow securities for the purposes of short-selling is low, the collateral posted against the securities earns the risk-free interest rate, also called the general collateral rate. When the demand to sell a particular security short is high, the collateral earns an interest rate lower than the general collateral rate. This interest rate is called the special repo rate, and the security itself is said to be "on special". The difference between the general collateral rate and the special repo rate acts as an implicit security borrowing fee, and leads to costly short selling. In the US Treasury note market, the most recently issued and most liquid on-the-run securities tend to be on special in that they have high borrowing fees associated with them. A consequence of special repo rates is the presence of price differences between similar or even identical securities in asset markets, because holders of the security with heavy short selling demand have an extra stream of income in the form of the fees they earn by lending out their holdings of securities. Duffie (1996) argues that if this is the case, the price of the security should ex-ante get elevated by an amount that exactly offsets the fee that needs to be paid to borrow the security. Empirically, using data over a much smaller time period, Jordan and Jordan (1998) and Krishnamurthy (2002) show that this hypothesis cannot be rejected. In contrast, I show that with search frictions, the proportion of future lending fees incorporated into the price is less than one when short-selling is driven by hedging needs, and is greater than one when short-selling arises due to demand by buyers to hold securities at elevated prices. I also provide empirical support using a much longer time series of repo rates and U.S. Treasury prices.

In order to explain the intuition of the model, consider the following situation where one unit of the security is initially issued. The price of the security after two periods is known and agreed upon by all agents to be 100. Let us assume that the repo market clears in steps and the cash market is frictionless. There are two agents who wish to short the security (shorters 1 and 2), and three agents (buyer/lender A, B and C) willing to take long positions. The agents who short the security do so because otherwise they face a per-period hedging cost, which may be interpreted as arising out of their risk aversion and the need to hedge an existing portfolio. Let us also assume that the agents who take long positions are risk-neutral, but participation in the repo market entails some per period transaction costs, which differ between agents holding the security. Let us say these are 1, 5 and 10, for lenders A, B and C, respectively.

There are gains to both the borrowers and the lenders here. For simplicity, let us assume that in any period, borrowers compensate the lenders for at least their marginal costs of lending for that period - that is, security borrowers have all the pricing power. At any point of time, and for any schedule of (positive) future lending fees, agent A has the highest valuation of the security, because he faces the lowest transaction cost for lending. Hence at time 0, A buys the security. At the end of period one, buyer A lends the security to shorter 1 (who is randomly chosen from the two agents wishing to short), who in turn sells it to buyer B. At the end of period two, buyer B lends the security to shorter 2, who in turn sells the security to buyer C, who values it at 100. In such a setting, shorter 2 pays a lending fee of 5 to lender B. However, since lender B's cost of lending is 5, his ex-ante valuation of the security remains 100 (which is the price at which he buys the security from shorter A). That is to say, at the end of the first period, none of the future lending fees are incorporated into the price.

It should be noted, that for lender A to not recall his security and offer it to shorter 2, shorter 1 must pay lender A the same fee of 5 in the second period. Going back one step further, lender A is compensated in the first period by shorter 1 for his marginal cost of lending, which is one. The total lending fees that lender A expects to gain during this sequence of transactions is thus six (one in the first period and five in the second period), while he expects a total cost of two from his lending activity. Thus, ex-ante, lender A would be willing to buy the security at a price of 104. The lending fees that we would observe as this process unfolds would be 1 in the first period and 5 in the second period. The initial price would incorporate (4/6) or 66% of all the future lending fees observed starting in the first period, and the price at the end of the first period would incorporate only 0% of all the future lending fees observed starting in the second period. The initial price would decline, the per period fee would increase, and the proportion of future lending fees included in the price would decline. Neither lender B nor lender C can match the initial valuation of lender A, which is 104. This simple model does not explain how the measures of shorters and lenders evolve over time. However, it illustrates a situation where lenders have different marginal costs and the total short interest is more than the initial issue of the bond. Moreover, in this model, the hedgers get all the benefits of shorting the security, while we would expect these benefits to be shared between borrowers and lenders.

The model I present in the paper is more realistic because instead of having explicit transaction costs, it focuses on the difficulty of finding a security to borrow in the market for borrowing and lending securities, henceforth called the repo market, and its implications for lending fees and prices.[1] The costs are in the form of opportunity losses because a holder of the security cannot lend out her holdings instantaneously, but must wait to find a borrower of the security. There is no inherent risk to short selling.[2] The search-based framework that I use captures several stylized facts about the process of short selling. First, in order to obtain a short position in a security, an agent must first borrow it in the repo market, and sell it to another agent willing to buy the security at a price that is agreeable to the latter. Second, securities are traded in "market-lots," and in order to establish a short position of a certain size, the entire amount must be borrowed and delivered. Third, market participants often report that securities become scarcer to locate in the repo market when the existing short interest in the security is high.[3] Since every new short position creates a new long position, which is in itself can be potentially lent out, securities can become difficult to find only if the new buyers of the securities are more difficult to locate by future security borrowers. I capture this effect by modeling heterogeneity in terms of the participation in the repo-market by holders of securities.

In the setting of this paper, there are two conceptual channels that drive the demand to "short-sell." The first is the desire of some agents to hedge against unfavorable moves in other portfolios, or in their endowment streams. An example of such an agent could be a corporate bond desk that wishes to hedge interest rate risk but wants to make credit bets. This "hedging demand" to short-sell might cause the demand to borrow a security to be high relative to its available supply, and may lead to high specialness, and consequently price differences may arise even in situations where there are no optimistic beliefs amongst buyers about the intrinsic value of a security. In the second case, the short selling is motivated by elevated prices driven by buyers who have a hedging reason to have a long position in the on-the-run security. For instance, foreign central banks with need to place foreign reserves may push up prices of Treasuries, leading arbitrageurs to short sell. Henceforth, I refer to these two cases as the case of hedging demand to short sell and arbitrage

[1] Securities may be borrowed or lent in two kinds of transactions: a repo trade or a stock-loan. With minor technical differences, these two transactions are economically similar. The former is usually the mechanism by which Treasury securities and certain high quality bonds are lend, while the latter is typically used for risky corporate bonds and equities.

[2] Note that this is very different from other theories of convergence trading involving risky arbitrage, viz Shleifer and Vishny (1997),Xiong (2001) and Wurgler and Zhuravskaya (2002)

[3] See, i.a. Fisher (2003) and Molton (2004).

demand to short sell respectively.

In both cases, repo market frictions lead to a form of segmentation where the prices of the securities are above the intrinsic valuations of the buyers of the securities. However, when bargaining power is distributed across lenders and borrowers, I show that hedging demand to short-sell has different implications on the relationship between observed "over-pricing" and the observed lending fees, as compared to arbitrage demand to short sell. This leads to a simple test of determining which of the two cases is the main driver for fees in the repo market - in the former case, a hypothetical convergence trade makes *losses*, and the proportion of lending fees that are included in the price premium is less than one. The Duffie (1996) result that price premia are equal to observed future lending fees (or equivalently that convergence trades make zero profits) arises as two special cases. The first special case is when security borrowing is driven by hedging needs of short-sellers and the lenders of the security have no bargaining power. This is a degenerate case in that there are no price premia and no fees. The second special case is when security borrowing is driven by arbitrageur activity, that is there are agents who are willing to buy the security at a price that is greater than the "fundamental" price, and the borrowers of the security (who are arbitrageurs) have no bargaining power.

The basic approach of modeling search frictions in repo markets has been discussed in Duffie, Gârleanu, and Pedersen (2002) and Vayanos and Weill (2007).[4] Duffie, Gârleanu, and Pedersen (2002) model the dynamic equilibrium that arises when repo markets are constrained by search frictions, and there are differences in beliefs about the value of a security that will be revealed at a stochastic stopping time. Vayanos and Weill (2007) present a steady state model where both short-sellers and holders of securities derive hedging benefits. Other papers that discuss the theme of price premia and fees in repo markets, especially in the context of Treasury securities include Fleming and Garbade (2004) and Fleming and Garbade (2003). Of special relevance to my paper is the exposition of repo rate patterns over auction cycles given by Fisher (2002), who documents

[4]There is a long literature on the role of costly short selling and asset prices in the equity markets. See, i.a.,Lamont and Thaler (2003) for the role of stock-lending in carve-outs, Ofek and Richardson (2003) for the costly short selling and the dot com bubble, D'Avolio (2002) for the empirical determinants of special rebate rates in stocks and Evans, Geczy, Musto, and Reed (2005) for the roles of failures in the lending market for stocks. More recently, Cohen, Diether, and Malloy (2007) show that special rebate rates are not a sufficient statistics for over pricing in stocks.

a “rising and falling” pattern in lending fees implied by special repo rates in on-the-run Treasury notes as time passes after an auction, and a corresponding decline in the price premium. More recently, Graveline and McBrady (2005) argue that the repo rate specialness arises from hedging demand to short-sell rather than arbitrage demand to short sell by using empirical instruments that proxy for the two sources of demand, respectively.

This paper contributes to the literature by theoretically contrasting hedging demand to short sell with arbitrageur demand to short sell, incorporating heterogeneity in the extents to which holders of securities participate in repo markets, and explaining the joint evolution of prices, repo fees and short interest. I test the theoretical results using data on the price premia of on-the-run Treasury securities over off-the-run Treasury securities and their associated repo rates over several auction-cycles in the ten year Treasury note. The empirical tests I derive from the model support the conclusion that short positions in on-the-run Treasury securities are used by market participants to hedge other fixed income portfolios such as corporate bonds, agency securities, and interest rate derivatives, rather than to purely benefit from the on-the-run price premium. I show that not only does a pure convergence trade in on-the-run/off-the-run securities, on an average, make losses, but also the *dynamics* of the price premia and the observed specialness in the market can be adequately explained by the search-based model that I propose. In addition, consistent with the prediction of the model, as time passes since the auction, the price-premia of on-the-run issues incorporate a smaller and smaller proportion of the observed future specialness. The model is also realistic in the sense that it explains why market participants wishing to short-sell the on-the-run security find it increasingly difficult to locate the security as time passes since the issuance of the security.

As a baseline, I focus on the case of hedging demand to short sell. The theoretical results for the case of arbitrage demand to short sell are a simple modification of this case. Some agents benefit from taking short positions in either one of two identical securities because this allows them to hedge the exposure of other portfolios they hold. There are no differences in beliefs about the values of two securities on a certain termination date in the future. One of the securities is perfectly illiquid in that cannot be sold short - the other security is perfectly liquid and hence the demand to sell short is concentrated in the latter. I theoretically derive the joint dynamics of prices, repo fees and short interest in the presence of hedging needs. These are determined by the valuations of marginal buyers at any point of time, the number of unfilled security borrowers, and the difficulty in finding a security lender. The price reflects the valuation of a marginal buyer who buys the security from the short sellers and *not* the valuation of existing holders of the security. The fees are determined as the equilibrium of a bilateral bargaining game between lenders and borrowers. Due to search frictions, buyers of a security cannot benefit instantaneously to the full extent of the borrowing fees that are observed in the market at that point - they need to wait before they can lend out their entire holdings. This implies that the proportion of the total lending fees that are incorporated as into the price is less than one.

When there are differences in terms of repo-market access within potential holders of the security, in equilibrium, it only makes sense for the high repo-access holders to buy the security initially when short interest is low, and for the low repo-access holders to buy the security after all the high-access "longs" have already become holders of the security and short interest has built up. In the setting that I use, the total number of long positions in the security increases over time because short interest builds up, and hence agents who participate less in the repo market progressively become marginal buyers. However, for borrowers in the repo market, this has an added implication - this makes it more and more difficult for the remaining borrowers of securities to access lenders after short interest builds up. In addition, the later buyers of the security stand to gain less from lending out their securities, because they have to wait for a longer time before they can start earning lending fees from their holdings. This implies that *the proportion of the* observed *future specialness that is incorporated into the price decreases.*

Finally, I link the implicit hedging benefits implied by on-the-run special repo rates with economy-wide risk factors. I find that several factors known to give rise to risk premia in the asset pricing literature seem to be directly related with special repo rates, after accounting for auction-specific variables. Prominent amongst these are the market excess return and the return on the Fama-French book-to-market portfolios. I find that when both the return on the market and the return on the book-to-market portfolio are high, the need to hedge interest rate risk as implied by the repo rate data is low.

This paper is divided into the following sections. Section 2 presents the model, and examines the properties of the dynamic equilibrium that arises between prices and lending fees in the context of search frictions. Section 3 describes the data, the empirical test design and the main empirical results. Section 4 concludes and outlines areas for further research.

2 Model

I first describe the set-up of the market for the base-line case. The base-line case mirrors the issuance of an on-the-run Treasury security where agents have hedging needs to short-sell. Some agents are willing to take up long positions in the security, while other agents have a need to acquire a short-position because this affords them certain hedging benefits for hedging interest rate risk in say, a corporate bond or another fixed income portfolio. A security is issued at time 0, and has a value of V(T) (known with certainty) at time T. The amount of the new security issued is S, and it is bought by a measure S of buyers at a price P(0). The buyers of this security then proceed to lend the security out to agents wishing to short it. The marginal buyers of the security are determined in equilibrium and the prices are characterized in section 2.3. There is another asset which has the same cash-flows as the recently issued asset, and a value V(T) at time T. I assume that this asset is perfectly illiquid and hence cannot be traded. In equilibrium, this also implies that the asset is held by agents who do not participate in the repo market at all and hold it to maturity, and its value is V(T) at all times.[5]

2.1 Borrowers

In the interval $t \in [0, T]$ there exists a set of agents desiring to acquire a short position in the security. The measure of these agents is given by D. Each agent is allowed to short only one unit of the security.[6] A hedger who puts a short position in the security earns a per period hedging benefit of x. [7]

[5]This may be thought of as a perfectly illiquid "off-the-run" security, which is not sold short.

[6]As in Vayanos and Weill (2007), this can be interpreted as a borrowing constraint or as risk aversion.

[7]Potentially, the per period benefit may itself be a function of time. However, this does not significantly add to the intuition of the model and makes it less tractable analytically.

In order to sell a security short, a borrower must first borrow it and sell it to a buyer. The cash market is frictionless, i.e. securities can be bought and sold instantly. However, borrowers must search for lenders in order to borrow the security. We define $\mu_{bo}(t)$ as the measure of borrowers who have not yet met a lender, and $\mu_s(t)$ as the measure of borrowers who have already shorted the security. For obvious reasons:

$$\mu_{bo}(t) + \mu_s(t) = D \tag{1}$$

2.2 Lenders

There exists a set of potential lenders with measure $\bar{\sigma}$. Lenders are differentiated by the rate at which they can be accessed by borrowers in the repo market, given by a decreasing function $\lambda : [0, \bar{\sigma}] \to [\lambda_0, \lambda_{\bar{\sigma}}]$ [8] In the numerical example which I use throughout this paper, the assumed distribution of λ is characterized by figure 1. With the hedging demand to borrow (D) set at 300, this corresponds to a probability of about 70% of encountering a borrower within the first day of acquiring a security for the highest access lender, and of about 30% for the lowest access lender.

However, potential lenders may not necessarily hold the security at any given point of time. Any potential lender will acquire the security only when her valuation of the security is at least as high as the price at which it can be bought. In subsection 2.3, I show that the agents who become marginal buyers of the security at any point of time are those who, out of the set of all possible holders who have not yet bought the security, find it easiest to lend out their holdings. As soon as an agent buys the security, she commences the search of a potential borrower.

Borrowers search for lenders in the repo market and are matched randomly in pairs with them. For a lender of type σ, at any time t, the arrival of a potential borrower is given by a Poisson process with an intensity of $\lambda_\sigma \mu^{bo}(t)$. Conditional on the fact that an agent of type σ has already bought a security, let the measure of agents who have not yet met a borrower in the interval defined by $\delta\sigma$ be given by $\mu^n_\sigma(t)\delta\sigma$. By the law of large numbers, the rate at which agents in the interval $\delta\sigma$ who have not yet lent out their securities meet potential borrowers is given by:

[8] The rate may be interpreted either as the ease with which borrowers can find the lender in the repo market, or the willingness of the lender to participate in the market.

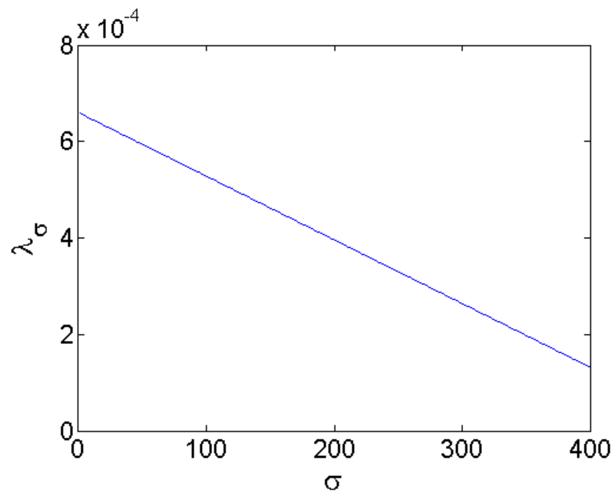

Figure 1: **Distribution of λ_σ against σ:** This figure shows the distribution of the lenders by the ease with which they can be searched in the repo market. λ denotes the per period intensity with which a particular agent holding the security can meet with a particular agent who desires to borrow the security. σ denotes the agents' "type" as characterized by the ease with which he can be accessed by a borrower. The agents' types are ordered in decreasing order of λ.

$$\dot{\mu^n_\sigma}(t)\delta\sigma = -\lambda_\sigma \delta\sigma \mu^n_\sigma(t)\mu^{bo}(t)$$

In the limit as $\delta\sigma \to 0$, this implies that:

$$\dot{\mu^n_\sigma}(t)d\sigma = -\lambda_\sigma d\sigma \mu^n_\sigma(t)\mu^{bo}(t) \tag{2}$$

where the $\dot{()}$ indicates a rate of change with respect to time. The implicit assumption is that all meetings between a potential lender and a potential borrower result in a transaction.

Because every time a security is lent out, it gets sold short and is bought by a new potential lender, we have the following condition:

$$\int_0^{\sigma(t)} \mu^n_\sigma(t)d\sigma = S \tag{3}$$

Let $\sigma(t)$ denote the type of agents who have become marginal buyers at time t. At time t, each borrower meets potential lenders lying in the interval $\delta\sigma$ with an intensity given by $\lambda_\sigma \mu^n_\sigma(t) d\sigma$. By the law of large numbers, this means that the total rate at which borrowers meet lenders of types $\in [0, \sigma(t)]$ is given by:

$$\dot{\mu^{bo}}(t) = -\mu^{bo}(t) \int_0^{\sigma(t)} \lambda_\sigma \mu^n_\sigma(t) d\sigma \tag{4}$$

The level of short interest at any time t is given by equation 1 as:

$$\mu_s(t) = D - \mu^{bo}(t)$$

Because the total sum of long positions ($\sigma(t)$) is equal to the sum of the initial float of the security and the total number of short positions, we have the following identity:

$$\begin{aligned} \sigma(t) &= S + \mu_s(t) \\ &= S + D - \mu^{bo}(t) \end{aligned} \tag{5}$$

The equations for the dynamics of the measures are independent of the equations for the valuations of the investors. This is possible if, once bought, a security is never sold, and once it has been lent out, it is not recalled till the termination date T. I justify why this happens in section 2.3. Equations 2 and 4, together with 5 and 3 define a system of equations that can be solved jointly for any value of S, D and $\lambda(\sigma)$. While an analytical solution for these equations is difficult for arbitrary $\lambda(\sigma)$, these are easy to solve numerically, since they just denote a system of integral equations. A numerical scheme for solving the system of equations is outlined in Appendix A. For the distribution of λ in figure 1 and with an initial value of S = 20 and D = 300, the evolution of available lenders of each type $\mu^n_\sigma(t)$ and the short interest $\mu^s(t)$ are plotted in figures 2 and 3 respectively.

Some explanation of the figures is in order. In figure 2, at time t = 0, a measure S of the highest rate lenders acquires the security. The function $\mu^n_\sigma(t)$ denotes the fraction of lenders in an interval $\partial\sigma$ around σ who have not yet lent out their securities, and at time 0, this fraction is 1 for

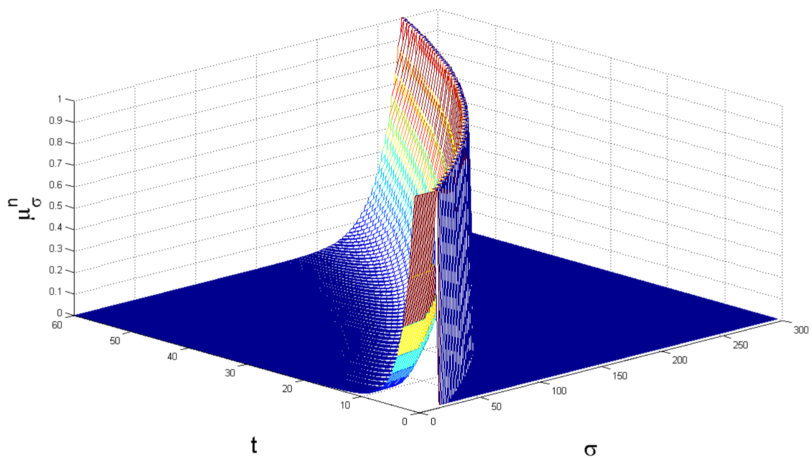

Figure 2: **Plot of μ^n_σ against time and σ** This figure shows a plot of the evolution of the measure of lenders who have not yet lent the security as a function of time and their type. The initial float, which is also equal to the sum of potential lenders at time zero, is 20. The total number of agents wishing to borrow the security is 300. The lenders are characterized by a distribution of their repo market availability that is given in Figure 1.

all values of $\sigma \leq S$. On the z-axis against time, we can see the type of potential lender $\sigma(t)$ who becomes the marginal buyer of the security at any time t. For the obvious reason that the marginal buyer has "just" bought the security, the fraction of marginal buyers who have not yet lent out their securities is always 1. Over time, each type of "long" encounters a borrower and lends out his security. That is the fraction $\mu^n_\sigma(t)$ reduces over time at a rate that depends both on λ_σ, and the number of borrowers available decreases, and by the identity in equation 5, the type of agent who has become the marginal buyer at any time t increases - that is, an agent with a lower and lower ease of participation in the repo market becomes the marginal buyer of the security.

As Figure 3 shows, at the time of issuance the number of short-positions in the security are zero. The short interest increases rapidly at first as the borrowers of the security find lenders and put on short positions in the security. As time passes, the security is likely to be held more and more by agents who do not participate in the repo market. Moreover, the number of borrowers who have not yet undertaken short positions in the security also decreases. Because of this, the rate at which short interest builds up reduces with time, because meetings between borrowers and lenders become more and more infrequent.

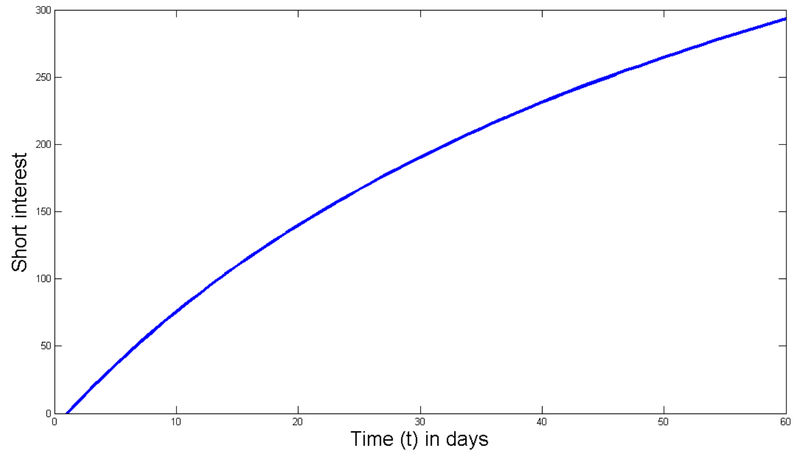

Figure 3: **Plot of Short interest (μ_t^s) against time** This figure shows a plot of the evolution of short interest as a function of time. The initial float, which is also equal to the sum of potential lenders at time zero, is 20. The total number of agents wishing to borrow the security is 300. The lenders are characterized by a distribution of their repo market availability that is given in Figure 1.

2.3 Valuation

When a potential lender and a borrower meet, they bargain over the lending fee to be paid. Since there are no recalls, the security remains lent until time T. Since a lender always has the option to lend a security to another agent at any point of time by recalling it, the per period fee w(t) at any time is the same for all borrower-lender matches. Hence the valuations of agents, once they become lenders, are the same for all types of holders of the security $\sigma \in [0, \sigma(t)]$. For this reason, we can suppress the type of an agent once she lends out her security. An agent who has already lent a security to a borrower and is earning a non-negative fee $w(t)$ on it has a valuation at any time t that is given by the following differential equation:

$$-dV^l(t) = w(t)dt \tag{6}$$

subject to the boundary condition:

$$V^l(T) = V(T)$$

which implies that:

$$V^l(t) = V(T) + \int_t^T w(s)ds \tag{7}$$

Now consider an agent of type σ who has bought a security and wishes to lend it. His valuation, $V_\sigma(t)$ is given by the following equation. This equation can be derived as a discrete time Bellman equation where the state of a holder of the security is indexed by whether or not she has lent it out and by taking the continuous time limit of the solution.

$$-dV_\sigma(t) = \lambda_\sigma \mu^{bo}(t)(V^l(t) - V_\sigma(t)) \tag{8}$$

subject to the boundary condition:

$$V_\sigma(T) = V(T)$$

The solution to this differential equation is given by:

$$V_\sigma(t) = \left(1 - \int_t^T \lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau} ds\right) V(T) + \int_t^T \lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau} V^l(s) ds \tag{9}$$

$$= V(T) + \int_t^T \lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau} \int_s^T w(\tau) d\tau ds \tag{10}$$

Equation 9 is intuitive. $\lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau}$ is the density function for the first arrival time for a Poisson Process whose intensity is time-varying according to the measure of borrowers available at any time t (the intensity is given by $\lambda_\sigma \mu^{bo}(t)$). The first term on the RHS is simply the probability of not meeting a borrower from time t to T. The second term gives the expected value of becoming a lender at any time between t and T. All that equation 10 implies is that the valuation of an agent of type σ is the sum of the value as on the termination date T, which is known with certainty, and the expected value of lending fees to be earned, once the agent meets a borrower. It is important to note that observations of $w(t)$ will occur because of the fees being paid by (other) borrowers who have already shorted the security. However, valuations will not incorporate all the observed fees, because every buyer has to wait before he finds a security borrower.

Note that subtracting 10 from 7 indicates that $V^l(t) - V_\sigma(t)$ is not dependent on V(T).

We now proceed to characterize the marginal buyers of the security at any time t. This is done in the following proposition and its corollaries.

Proposition 2.1. *The valuation of any agent who has not lent a security at time t is increasing in his repo market accessibility ($\lambda\sigma$)*

Intuitively, this is because agents with a higher rate of meeting borrowers can have a higher expected value of lending, because they are likely to meet borrowers before agents with a lower rate of meeting borrowers. The proof of the proposition is in appendix B. The proposition also implies the following:

Corollary 2.2. *The relationship between the price of the security and the expected future lending fees $L(t)$ at any time t is given by:*

$$P(t) = V(T) + \int_t^T \lambda_{\sigma(t)} \mu^{bo}(s) e^{-\int_t^s \lambda_{\sigma(t)} \mu^{bo}(\tau) d\tau} L(s) ds \tag{11}$$

where

$$L(s) = \int_s^T w(\tau) d\tau$$

For a buyer to buy the security, his valuation must at least be equal to the price of the security, because the cash market is perfectly competitive. A situation where agents who have a higher valuation than the price of the security cannot arise because the cash market is assumed to be frictionless - such agents would already have bought the security and started the lending search process. Agents belonging to type $\sigma > \sigma(t)$ at any time t will have a lower valuation than agents of type $\sigma(t)$ because of our assumption that λ is decreasing in σ.

Corollary 2.3. *A buyer of type $\sigma(t*)$ who buys a security at t^* never sells it at $t \in [t*, T]$.*

This follows from the fact that $V_{\sigma(t*)}(t) > V_{\sigma(t)}(t)$ owing to our assumption that λ is decreasing in σ and the fact that σ is increasing in t. Once bought a security is never sold to another (lower valuation) agent. The buyers simply enter the search process to find borrowers.

Corollary 2.4. *If an amount S of the security is auctioned at time t = 0, agents bidding in the auction are the set of highest lending rate agents such that* $\sigma(0) = S$*. The relation between the auction price and the measure of the marginal buyer in the auction is given by* $P(0) = V_{\sigma(0)}(0)$

Corollary 2.4 has an implication on the price of the security in the auction. Ceteris paribus, if S is smaller, $\sigma(0)$ is smaller and $\lambda(0)$ is higher implying a higher V(0). Needless to say, a lower value of S also has implications for L(t), the present value of future lending fees at any point of time. As we shall see, L(t) is also decreasing in S for all values of t, implying that auctions where lesser amounts of the security are issued will be at higher prices. This is also observed empirically by Nyborg and Sundaresan (1995) and Jordan and Jordan (1997).

Consider a borrower who has borrowed and shorted a security. His value $V^s(t)$ is given by the differential equation:

$$-\dot{V^s}(t) + \dot{P}_t = x(t) - w(t) \tag{12}$$

with the boundary condition:

$$V^s(T) = 0$$

This is because an agent who has sold the security short gains from any reduction in the price of the security, in addition to gaining the hedging benefit, while paying out the lending fee.

This implies that:

$$V^s(t) = x(T - t) - \int_t^T w(s)ds + P(t) - V(T) \tag{13}$$

Consider a borrower who has not yet borrowed a security. The value of the borrower at any time t is given by the differential equation:

$$-\dot{V^{bo}}(t) = \int_0^{\sigma(t)} \lambda_\sigma \mu_\sigma^n(t) d\sigma \left(V^s(t) - V^{bo}(t) \right) \tag{14}$$

with the boundary condition:

$$V^{bo}(T) = 0$$

which implies that:

$$\begin{aligned} V^{bo}(t) &= \int_t^T \int_0^{\sigma(s)} \lambda_\sigma \mu_\sigma^n(s) d\sigma e^{-\int_t^s \int_0^{\sigma(\tau)} \lambda_\sigma \mu_\sigma^n(\tau) d\sigma d\tau} \left[x(T-s) + P(s) - V(T) - \int_s^T w(\tau) d\tau \right] ds \\ &= \int_t^T \frac{-\dot{\mu^{bo}}(s)}{\mu^{bo}(s)} e^{-\int_t^s \frac{-\dot{\mu^{bo}}(\tau)}{\mu^{bo}(\tau)} d\tau} \left[x(T-s) + P(s) - V(T) - \int_s^T w(\tau) d\tau \right] ds \end{aligned} \tag{15}$$

Using the fact that $\frac{\dot{\mu^{bo}}(\tau)}{\mu^{bo}(\tau)}(d\tau) = d\left(ln(\mu^{bo}(\tau)) \right)$, this can be simplified to

$$V^{bo}(t) = \int_t^T \frac{-\dot{\mu^{bo}}(s)}{\mu^{bo}(t)} \left[x(T-s) + P(s) - V(T) - \int_s^T w(\tau) d\tau \right] ds \tag{16}$$

2.4 Bargaining

In this section, I model the bargaining that occurs when a lender meets a borrower at time t*. Borrowers and lenders meet and bargain over the lending fee schedule $w(\tau)$ where $t \le \tau \le T$. Let us assume that the lender is chosen with probability θ to make the first offer. She then offers a schedule that makes the borrower's benefit equal to his reservation value. With probability $(1-\theta)$, the borrower is chosen to make the first offer. In order to ensure that all types of agents get the same lending fee, we assume that there is information asymmetry. If all meetings are to result in a trade, the borrower has to offer a lending fee schedule that reflects the outside options of the highest valuation lender. Thus the expected present value of the lending fee is given by solving the following equation:

$$\theta\left(V^{s}(t)-V^{bo}(t)\right)=(1-\theta)\left(V^{l}(t)-V_{0}^{n}(t)\right) \tag{17}$$

Proposition 2.5. *If the borrowers have all the bargaining power, the per period fee is zero. The price of the security is constant and is equal to zero (the value at the termination date T).*

Proof in Appendix B. Intuitively, this is because when the borrowers have all the bargaining power, they set the gains from lending for the lenders equal to their reservation value. This is only possible for all values of t, if the gains from lending are zero.

Proposition 2.6. *If the lenders have all the bargaining power, the per period fee is given by x - dP(t), where dP(t) is the change in the price of the security.*

The proof is in Appendix B. For intuition, consider the fact that the total per period gain made by any agent short selling the security is x - dP(t). That is, the agent who sells one unit of the security short gains from both the hedging benefit and the reduction in the price of the security. When lenders have all the bargaining power, they are able to extract all benefits from the borrowers.

We are interested in the cases where some of the benefits due to shorting the security are shared between the lenders and the borrowers. For $0<\theta<1$, the total future lending fee is given by the following implicit equation.

$$\begin{aligned}L(t)=\theta&\left(P(t)-V(T)+x(T-t)-\int_{t}^{T}\frac{-\dot{\mu^{bo}}(s)}{\mu^{bo}(t)}\left[x(T-s)+P(s)-V(T)-L(s)\right]ds\right)\\&+(1-\theta)\int_{t}^{T}\lambda_{0}\mu^{bo}(s)e^{-\int_{t}^{s}\lambda_{0}\mu^{bo}(\tau)d\tau}L(s)ds\end{aligned} \tag{18}$$

where P(t) is given by equation 11.

Equation 18 together with equation 11 can be jointly solved to derive the prices and lending fees, given that the allocations are given by equations 1 and 3 respectively. Both equation 11 and equation 18 are recursive in the sense that at any given point of time t, we need to know only

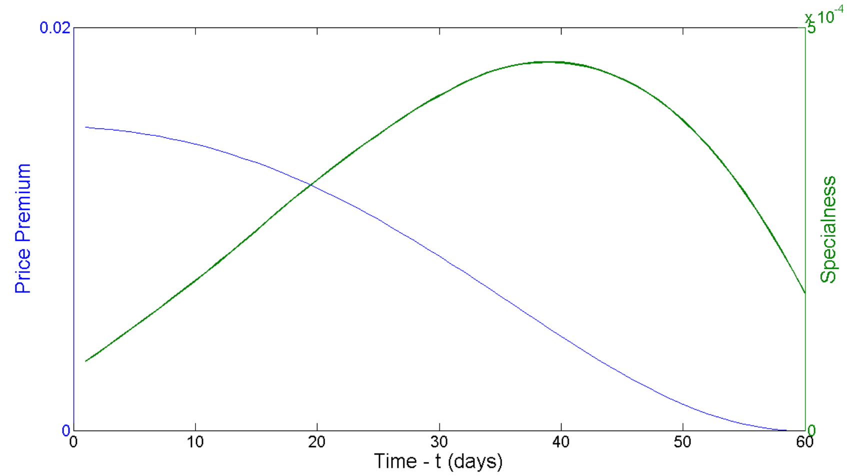

Figure 4: **Plot of the per-period lending fee and the price premium against time:** This figure shows a plot of the evolution of the per period lending fee $w(t)$ and the price premium $P(t) - V(T)$ as a function of time. The initial float, which is also equal to the sum of potential lenders at time zero, is 40. The total number of agents wishing to borrow the security is 300. The lenders are characterized by a distribution of the ease with which they can be accessed by borrowers that is given in Figure 1.

the future lending fee process $L(\tau)$ and the future price process $P(\tau)$, such that $t \leq \tau \leq T$. The per period lending fee $w(t)$ is simply the time derivative of the total future expected lending fee $L(t)$. The price premium $P(t) - V(T)$ and the per period fee $w(t)$ are plotted in Figure 4. The empirical counter part of this figure using the average price premium and the average specialness in the on-the-run ten year note is in Figure 7.

Proposition 2.7. *As time passes, the average time required by a borrower of security to locate a lender increases. That is, securities become "difficult" to locate as short interest builds up.*

The proof is outlined in Appendix B. If all the agents on the long side of the market are equally accessible to borrowers, then we would not expect the search time for any individual borrower to change. This is because every security that is sold short creates a corresponding new long position, which itself can be lent out, meaning that the total available supply of securities in the repo market is always equal to the initial issue size. If there are no differences in the search intensities with which borrowers can meet lenders, there will be no change in the average search time for an individual borrower. On the other hand, if buyers of the securities are differentiated by the ease with which borrowers may access them, the initial buyers of the security are all those agents who are easily

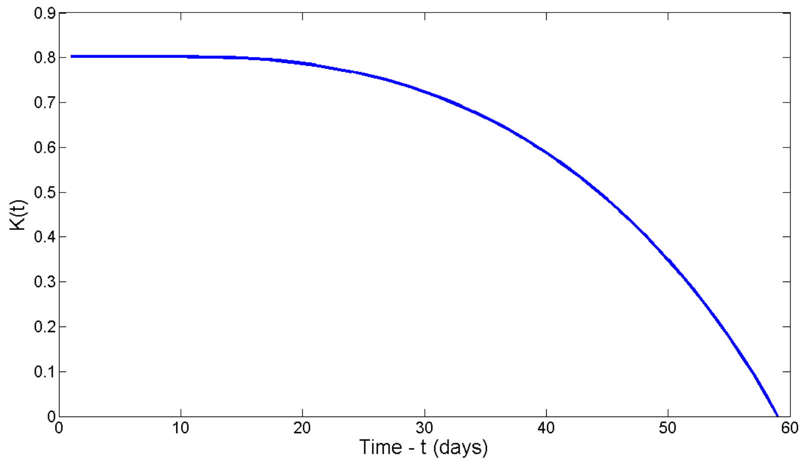

Figure 5: **Proportion of lending fees priced in:** This figure shows a plot of the proportion of all observed future lending fees $L(t)$ that are included as part of the price premium $P(t) - V(T)$. The initial float, which is also equal to the sum of potential lenders at time zero, is 40. The total number of agents wishing to borrow the security is 300. The lenders are characterized by a distribution of their repo market availability that is given in Figure 1.

accessed by borrowers, while later buyers of the security are those who are not easily accessed by the borrowers. Even though the total supply of the securities in the repo market remains the same, as time passes, some part of this supply ends up being held by agents who are "difficult to find" by borrowers of the security. This implies that the average search time increases as time passes since the issuance of a new Treasury security.

Proposition 2.8. *In the presence of search frictions and short selling driven by hedging needs, the proportion of the* observed *future lending fees that are incorporated into the price is less than one. Moreover, the proportion decreases with time. An arbitrageur who attempts to benefit from a short position in the security makes losses at all times.*

The proof of the proposition is outlined in Appendix B. Intuitively, this occurs for two reasons. The first is that the valuations of high repo-market access buyers are higher than the low repo-market access buyers. Hence, the former are the first to buy the security, and the latter buy the security security only after all the high repo-access buyers have acquired long positions in it. The second is that the number of borrowers available decreases as time passes. Because of this, an initial buyer of the security has to wait for a shorter time than an agent who buys the security later, and thus the opportunity loss from forgoing the observed lending fees in the market is lower for the former than for the latter. The fact that any buyer has to wait for some time before he lends out a security has two implications. First, not all of the observed lending fees are included into the price. Second, since the price at any given time is the valuation of the marginal buyer of the security, this implies that the proportion of future lending fees that is incorporated into the price of the security initially (when higher repo market access lenders are the marginal buyers) is higher than the proportion of future lending fees that is incorporated into the price later in time. Figure 5 shows the proportion of the future lending fees incorporated into the price $\frac{P(t)-V(T)}{L(t)}$ at any given time t, as an output of the model. This proposition implies that the proportion of observed future specialness that is incorporated in to the price of a security is less than one and decreases as time passes following issuance. This proposition forms the main basis of the empirical tests in this paper. Figure 8 shows an estimate of the proportion of future lending fees incorporated into the prices of ten year Treasury notes, as time passes after issuance. The method for computation of this coefficient is outlined in section 3.

Figure 6 shows the per period profit made by a hypothetical convergence trader who undertakes a short position in the security, and pays borrowing fees in order to maintain it. The profits are negative in all periods, indicating that a convergence trader makes losses. Figure 9 is the empirical counterpart of Figure 6, with the minor difference that it shows profits made over a weekly period. The detailed methodology for the computation of this figure is described in Section 3.

Proposition 2.9. ***Non-monotonicity of fees, prices and repo search frictions:***
If $\lambda_\sigma = 0$, $\forall\ 0 \leq \sigma \leq \bar{\sigma}$, *then* $P(t) = V(T)$ *and* $L(t) = 0$. *Similarly, as* $\lambda_\sigma \to \infty$, $\forall\ 0 \leq \sigma \leq \bar{\sigma}$, *then* $P(t) \to V(T)$ *and* $L(t) \to 0$.

For either the case of no frictions or absolute frictions, the price of the security at all times is equal to its expected value on termination date T. When there are no frictions, the lenders of the securities can lend instantaneously. Because the market has a whole has to have an overall long position in the security, there are always more lenders than borrowers. Hence the lenders compete the fee away to zero. If on the other hand, there are absolute frictions, then there are no fees to earn. Hence the price is the same as the valuation on the termination date. The only change between these two situations is in the welfare of the borrowers. In the former case, all their hedging

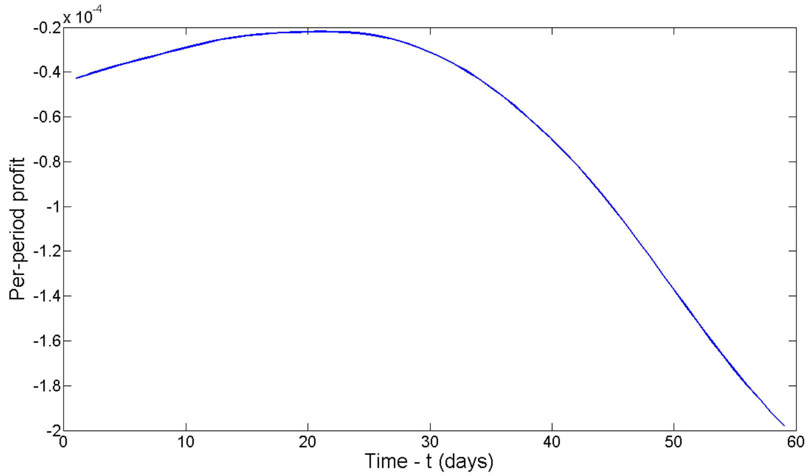

Figure 6: **Per period profits:** This figure shows a plot of the profit made every period by a hypothetical convergence trader who puts on a short position in the security. The initial float, which is also equal to the sum of potential lenders at time zero, is 40. The total number of agents wishing to borrow the security is 300. The lenders are characterized by a distribution of their repo market availability that is given in Figure 1.

needs are met, while in the latter case, none of their hedging needs are met.

Proposition 2.10. *The per period lending fee, the expected future lending fee, and the price premium $P(t) - V(T)$ at any point of time increase linearly in the hedging benefit of the short-sellers.*

Considering equation 18, and leaving the equations for the measures unchanged, we see that multiplying the hedging benefit by a constant $a > 1$ implies that the quantities L(t), and P(t) - V(T) also increase by the same amount. The proposition implies that both prices and the per period lending fees react to proportional changes in hedging benefits to the same degree. That is to say, if the hedging benefits that short sellers derive are high during a period, then price premia are high and at the same time specialness is also high. This allows us to perform the tests in section 3, where we compare price premia and specialness from different auction cycles in treasuries.

2.5 Discussion of the results

In this section, I make a few observations about the solution. Specifically, I focus on who benefits from the over pricing. Since the process of buying the security is competitive after the security has been issued, buyers on an average do not make any excess returns. This is because when an agent buys the security in the first place, she pays a price that is equal to all her expected earnings from the security, including any possibility of income from specialness. There are three entities that gain from the frictions:

1. Issuers of the security, who are able to sell their security at a price that is higher than its fundamental value. The price is elevated to the extent that the marginal initial buyer of the security expects to gain from lending fees. The issuer's gain per security issued is P(0) - V(T).

2. Initial buyers of the security who face lower search frictions in the market than the lowest valuation friction agent of type $\sigma(0) = S$. These agents, who lie in the interval $0 \leq \sigma \leq S$, have higher valuations because they can expect to lend more easily than the agent of type $\sigma = S$. Their gain per security is given by $V_\sigma(0) - P(0)$, where $0 \leq \sigma \leq S$.

3. Short sellers of the security, who derive some hedging benefits, as long as they have some bargaining power vis a vis the lenders.

Since buyers other than those who lie in $0 \leq \sigma \leq S$ do not stand to gain anything, even small unexpected costs related to shorting can disincentivise them.

The model also implies that the observed per period equilibrium special repo rate is not a sufficient statistic for the observed price premium. Only the expected *future* specialness is a sufficient statistic. In other words, the price premium reflects expected *future* specialness, and not the specialness that is observed at any given point of time. While prices and specialness are determined in equilibrium, the per period equilibrium specialness at any point of time may be associated with very different values of expected declines in future prices.

2.6 Demand to hold the security vs. Demand to hedge the security

In this section, I analyze the case where the *holders* of the security derive benefits from holding the security. In the case of Treasury securities, for example, this situation is analogous to a condition where agents who acquire long positions in a newly issued security are willing to hold the security at a higher price, because the new security affords them benefits over and above other Treasury securities, in terms of its higher liquidity. Consider an analog of the model described above. A security is issued at time 0, and its value at time T is expected to be V(T), which is agreed upon by all agents in the economy. However, in this case, it is the holders of the security who derive a per-period benefit of x from holding the recently issued and hence more liquid security. The price of the security is the highest initially because the hedging benefits are derived for the longest period, from time 0 to time T. I still assume that agents desiring a long position in the security are of different types, where the type of an agent is characterized by how easily the agent is accessible by potential borrowers in the repo market. Let us assume that there exists a number of arbitrageurs in measure D who wish to benefit from the expected decline in price by borrowing the security and selling it to other agents wishing to hold the security.

In such a situation, the evolution of measures of lenders and borrowers will be identical to that described in sections 2.1 and 2.2. If the security is issued in an amount S, the initial set of buyers of the security in the auction will be S agents who are most easily accessible by borrowers in the repo market, out of a set of all possible agents who could potentially desire to hold a long position in the security. Moreover, solving the Bellman equations for valuations of the different types of agents gives the following equations for a lender of the security, and the non-lender of the security of type σ:

$$V^l(t) = V(T) + x(T-t) + L(t) \tag{19}$$

$$V_\sigma(t) = V(T) + x(T-t) + \int_t^T \lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau} L(s)ds \tag{20}$$

Equation 20 also implies that the price of the security is given by valuation of the marginal buyer $\sigma(t)$ at time t.

$$P(t) = V(T) + x(T-t) + \int_t^T \lambda_{\sigma(t)} \mu^{bo}(s) e^{-\int_t^s \lambda_{\sigma(t)} \mu^{bo}(\tau)d\tau} L(s)ds \tag{21}$$

The valuations of short-sellers and potential borrowers, respectively are given by the following equations

$$V^s(t) = P(t) - V(T) - L(t) \tag{22}$$

$$V^{bo}(t) = \int_t^T \frac{-\dot{\mu^{bo}}(s)}{\mu^{bo}(t)} \left[P(s) - V(T) - L(s)\right] ds \tag{23}$$

Proposition 2.11. *When short selling is driven by the arbitrage activity, and there is demand to hold the security because holders of the security derive benefits, the price premium is greater than the value of future lending fees.*

To prove this, let us denote by $P^*(t)$ the quantity $P(t) - x(T-t)$. The lending fees in this case are the same as those in a situation where the short-sellers derive the hedging benefits and the price of the security is $P^*(t)$. The only difference between these two cases is that the price in the latter case is higher by the quantity $x(T-t)$. The borrowers in this situation are all arbitrageurs trying to benefit from the reduction in the price. In this case for $L(t) > 0$, it has to be the case that $V^s(t) > 0$, otherwise the borrowers lose on their short positions and there is no short-selling. However, this is possible only if $P(t) - V(T) > L(t)$, that is the price premium is greater than the observed future lending fee. A special case of the latter is the situation where arbitrageurs have no bargaining power - that is $\theta = 0$, in which case, using an argument similar to proposition 2.5, we can show that $w(t) = -dP(t)$. This is the Duffie(1996) result.

Proposition 2.11 taken together with proposition 2.8 has significant empirical implications. Graveline and McBrady (2005) use variables that are related to the demand to hedge Treasury securities, such as implied volatilities of interest rate options, and the issuance of fixed rate corporate debt and show that they explain some of the variation in repo-specialness in on-the-run Treasury securities. However, it is rather difficult to argue that these variables are completely unrelated to the demand to hold the on-the-run Treasury security. The implications of the model I present here are much easier to test. One simply needs to answer the question of whether a trade that tries to benefit from the price premium in the on-the-runs security by shorting it, and paying the implicit repo-fee, is profitable on average or not. If it is not, then it is most likely that the shorting fees are higher than the price premium in the security, and that most of the shorting demand is driven by agents who wish to short for hedging purposes. In the same vein, results from the literature on short-selling equities, e.g. Evans, Geczy, Musto, and Reed (2005) show that mispricing opportunities between options and the underlying equity can be profitably traded on, net of borrowing fees, by short selling the underlying security and creating a synthetic long position in the option. It implies that specialness in these kinds of stocks is driven primarily by arbitrageurs taking short positions to benefit from the expected decline in price.

3 Empirical results

In this section, I describe an application of the model to a typical on-the-run/off-the-run convergence trade in Treasury securities. The principal reason for using Treasury securities is that they are issued in auction cycles that are fairly regular and known to market participants. Thus their dynamics can be easily observed in the data.

In a typical Treasury auction, dealers bid for the security in uniform price auctions and initially acquire long positions. Other agents, including other dealers themselves, have a need to hedge their exposures to other fixed income instruments that are affected by interest rate risk, such as fixed rate corporate bonds, mortgages, and interest rate swaps. These agents borrow Treasury securities through reverse-repo, and short-sell them to other agents, to acquire short positions. The overwhelming majority of reverse-repos are overnight in nature, because of the risk that the agent who acquires the short position might face a squeeze. Overnight repo positions are typically rolled over by borrowing securities from the lender at the special repo rate prevailing in the market.

The first implication of the model is that short-interest in a security increases as time passes since issuance. Primary dealers in treasury securities typically maintain short positions in treasury securities in order to hedge their exposures to other fixed income instruments. While data on primary dealer positions in individual securities is not available, their maturity-wise positions are published at weekly intervals by the Federal Reserve Bank of New York since January, 2001. My conversations with market participants reveal that most short positions by primary dealers are in on-the-run treasury securities. For example, primary dealer short positions in the seven to ten year maturity bucket published by the Federal Reserve Bank of New York are to a large extent

Primary Dealer positions - 10 year bucket	
Days since issue	-100.69
	(-2.44)**
C	-42914.98
	(-5.26)**
AR(1)	0.60
	(1.69)
AR(2)	0.28
	(0.81)

Table 1: **Time series regression of weekly primary dealer positions in the ten year maturity bucket:** This table shows the results of an ARIMA estimation with two AR terms on primary dealer positions (in millions of dollars) in the ten year maturity bucket for US Treasury securities as reported by the Federal Reserve Bank of New York at weekly intervals for the period from July 2001 to June 2005. Figures in parentheses denote robust z-statistics. The dependent variables are the number of calendar days passed since the issuance of the most recent ten year Treasury note and a constant term.

driven by short positions in the on-the-run ten year note. As a test of whether short interest in the on-the-run security increases as time passes since issuance, I run a simple ARIMA regression of the Fed reported primary dealer positions in the seven to ten year maturity bucket, with time since issuance of the last on-the-run security as a dependent variable. I introduce two AR terms to control for auto-correation.

The results are presented in Table 1. First, it is interesting to note that the average position of Primary dealers in the ten year note is a negative, indicating that primary dealers usually maintain short positions. Second, the coefficient on the time since issuance term is negative. Since a negative sign represents a short position, this is evidence that the short interest in the on-the-run ten year note increases as time passes since its auction. The results of Table 1 may be compared with the output of the model that is presented in Figure 3.

Overnight special repo rates for on-the-run Treasury securities are published by GovPx and are available via the Bloomberg data service. I use daily data on special repo rates for 10 year on-the-run securities from 11/6/1995 to 6/1/2007. The reason for choosing ten year notes is that these securities have been issued over the sample period following regular cycles. These data represent the volume weighted average rate observed by the inter-dealer brokers that contribute to the GovPx system for all overnight reverse repo trades observed in the on-the-run bonds of a given maturity on any day. There are some periods for which data on special repo rates are missing. Most notably, this occurs after periods of market crises such as the week following September 11,

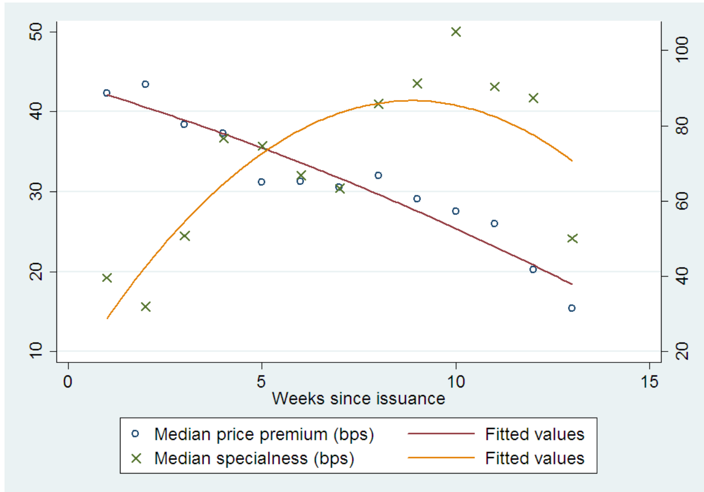

Figure 7: **Specialness and Price Premium in 10 year Treasury note:** This figure shows the price premium and the specialness , by days passed since issue, for an average quarterly issuance cycle for the ten year Treasury note. The price premia and the specialness are computed as averages across 38 quarterly ten year note auction cycles in the period from November 1995 to June 2007. The price premium for each note is computed as the reduction in price it will have to undergo in order to have the same spread compared to a USD-LIBOR swap of equivalent maturity, as off-the-run note issued two auctions prior. The specialness is computed as the difference between the special repo rate as published by GovPx , and the general Treasury collateral rate, also published by GovPx, for the same day.

2001. In these periods, I use indicative bid-side rates that are also collected by GovPx on the same day. Occasionally, the special repo rate for the day happens to be more than the general collateral rate. This could be because of data problems, and the fact that the general collateral rate itself might change during the course of a day. Such occurrences are rare. When this happens, I set the special repo rate to the general collateral rate, because special repo rates can never be more than the general collateral rates.[9]

I also use end of the day Treasury Prices from GovPx. Specifically, I use prices of on-the-run, first off-the-run, second off-the-run and third-off-the run securities. I use the Eurodollar swap curve as a benchmark against which I evaluate the prices of each of these securities. Since the Treasury yield curve itself is contaminated by the presence of expected future repo earnings, I evaluate each note against a hypothetical swap of the same maturity and the same characteristics as the bond, and then compare their spreads. Part of the difference between swap rates and Treasury yields arises because of the credit risk implicit in the floating rate LIBOR benchmark against which swaps

[9]This constitutes an arbitrage opportunity where an agent lends out the special security in a general collateral repo on which he pays the general collateral repo rate, and simultaneously reverses in the same security on the collateral for which he gets paid the special repo rate, thus making a riskless profit.

are valued. However, there is no reason why the term structure of credit risk underlying the swap benchmark is at all materially different for the ten year point and the nine year and nine months point, as would be the case for a typical auction cycle for the ten year note. Hence I believe that my constructed price premium measure is free from any biases that may be caused by the slope of the yield curve.

Figure 7 shows the specialness and price premium of the ten year on-the-run note against the 3rd off-the-run note evaluated against the swap rate over an average quarterly auction cycle in my sample. The price premium on the on-the-run bond is the reduction in price it will have to go through in order to have the same spread when evaluated against the swap curve as the third off-the-run bond. It is clear that an on-the-run effect persists into the second cycle. Since we want to compare the valuation of the on-the-run note with another note that has no specialness, we choose the premium of a note over the third off-the-run note in our study.

3.1 Test Design

Proposition 2.8 implies that

$$P(t) = E_t(P(T)) + K_t E_t(L_t) \tag{24}$$

or

$$E_t(P(t) - P(T)) = K_t E_t(L_t) \tag{25}$$

where P(t) is the price premium in an on-the-run Treasury security at any given day t after issuance of the bond, P(T) is the price premium at the end of the auction cycle, and $E_t(P(T))$ is its expectation at time t, and L_t is the cumulative value of observed specialness from time t to T, where the specialness for each day is discounted at the short term risk free rate (which I assume to be the same as the general collateral rate for the day) and $E_t(L_t)$ is its expectation at time t. Our hypothesis is that K_t is less than one and decreases as time passes. We intend to test against the null of $K_t = 1$.

Some previous studies in this literature, most notably Jordan and Jordan (1997) run pooled regressions of the following form:

$$s(t) = KL_t + \sum_{i=1}^{I} K_i v_i(t)$$

where $v_i(t)|_{i=1}^{I}$ is a set of variables that may be expected to affect the spreads, such as the demand to hold liquid assets. Note that the *realized* value of L_t is on the right hand side of this regression. They find that the coefficient K is not significantly different than one. However there are three problems with regressions of this form.

1. **Non-stationarity:** By definition, both the variables on the left hand side and the right hand side are non-stationary with respect to t. This leads to spurious inference in a regression.

2. **Information set:** The reason L_t should not be used on the right hand side of this regression is because the realized value of specialness is not in the information set of the investor at time t. Even assuming that the realized value of total specialness is the sum of its expected value at time t and some randomly distributed error term, we have an errors in variables problem on the right hand side variable. Besides, this answers the wrong question. We are interested in testing a hypothesis on the *expected* lending fees that is included in the price of the bond, not the actual realized fees, which could depend on the left hand side variable.

3. **Convergence:** As is clear from Figure 7, the on-the-run/off-the-run spread does not go to zero at the end of the auction cycle, but persists into the subsequent auction cycle. A trader who short sells the security from the beginning of the auction cycle to the issuance of the next note does not benefit to the full extent of the observed price premium - rather only to the extent that the price premium converges toward zero over the length of the auction cycle.

It is difficult to correct for these issues when the data-set has a limited number of auction cycles. However, with a large number of auction cycles, it is possible to design an econometric procedure that is free from biases caused due to non-stationarity. Denote by i the auction cycle, and by t the number of days passed since the security was issued. Then, under the null, we can perform the following estimation.

$$K_t \equiv \frac{\sum_{i=1}^{I}(P^i(t) - P^i(T))}{\sum_{i=1}^{I}(L(t))} \tag{26}$$

In appendix C, I show why this is a consistent estimator for K_t and outline a method to derive the standard error of this estimate.

There is another representation to equation 25, which is in the form of the profits over a short horizon Δ, made by a hypothetical convergence trader following a zero wealth strategy with a short position in the on-the-run security and a long position in an off-the-run security which has no specialness. The quantities of the short and the long position are pinned down by the need to make the portfolio duration neutral. This is an alternative approach, and is followed by Krishnamurthy (2002). Using data on twelve auction cycles using the thirty year bond, they show that such a strategy makes zero returns on an average. I replicate this approach and validate my results. The null hypothesis to be tested with this approach can be established from the implication of proposition 2.8 that a hypothetical convergence trade always makes losses when short selling is driven by hedging needs:

$$E_t\left(P(t) - P(t+\Delta) - (L_t - L_{t+\Delta})\right) < 0 \tag{27}$$

Under the null hypothesis implied by Duffie (1996) and Krishnamurthy (2002), the left hand side of equation 27 is equal to zero at all times. Under the alternative hypothesis of the model, it is less than zero for all values of t.

There are 42 regularly spaced auction cycles for the ten year note in the data. I ignore the reopening of the note that occurs after a month. Such reopenings have the potential to add some distortion to the time series of specialness, because the effective supply of the security increases. However, the amount issued in the reopenings is typically much smaller than the amount initially issued. I also ignore other non-periodic reopenings that occur from time to time. For the entire sample period, I do not have data on open market operations conducted by the Treasury from time to time, which might change the available float of a particular security. However the effect of these is not likely to be material[10].

3.2 Empirical Tests

The data series of the price premium in the ten year note and its specialness between November 1995 and June 2007 has two auction cycle frequencies giving us 42 auction cycles in all - 38 cycles with a quarterly frequency, and four with a semi-annual frequency. We first split the data up according to the length of the auction cycles. To ensure that we are dealing with roughly comparable auction cycles, we only use data from the quarterly auction cycles. The time series of price premium and specialness in the quarterly auction cycles is thus a non-stationary series with a recurring frequency of anywhere between 57 and 68 working days. One way of addressing non-stationarity is to fit an explicit functional form that depends on the days past the auction cycle, and on other variables that we may expect to affect specialness and spreads and to simultaneously estimate the model for the entire data-set. It is difficult for us to guess the functional form to use. Also, in order to test equation 24, we need to form expectations of both the reduction in the price premium, and the cumulative lending fees from any day in the auction cycle to the end of the auction cycle. We follow a three-step procedure:

1. We group daily data by weeks since issuance. That is, we compute the average price premium over the off the run Treasury during a week, and the total cumulative fees that are paid from the middle of the week to the date the next security is issued. This is done so as to reduce any problems related to bid-ask bounce and asynchronous price observations. This gives us a data set which has, for every auction cycle and for weeks since issuance, a measure of the price premium for the week and a measure of the cumulative repo specialness from the middle of that week till the end of the auction cycle. Between auction cycles, there is still a variation in the number of weeks that ranges from 12 to 14.

2. For every auction cycle and week, we compute the change in the price premium that took

[10] The data on open market operations is available from July 2005 onwards from the Federal Reserve Bank of New York. The frequency of permanent open market operations is low, and they typically involve notes or bonds other than the on-the-run securities. For example, for the ten year note, the total number of permanent open market operations that the Fed undertook was 58. Of these, only three featured the purchase of the on-the-run Treasury security, and the size of the purchase was, on an average, about USD 105 million.

place from a particular week to the last week in the auction cycle. Since on-the-run/off-the-run spreads decrease in most auction cycles, this change is positive, and represents the profits made by a convergence trader who did not face any costs of establishing a short position.

3. We form sub-samples of the two variables above by weeks since the last auction. That is, we form subgroups of the series where each group contains only the data-points applicable to a certain number of weeks past a particular auction date. By construction, the sub-samples are stationary. For each value of weeks since last auction, we compute the cumulative average cost of shorting owing to specialness and the cumulative average earnings from the reduction in the price premium. We then compute the ratio of the two, which is the sample estimate of K_t in equation 26. We also compute the standard error of this estimate, using the method outlined in Appendix C. This enables us to answer the question: "Looking at the same day in different auction cycles, what proportion of the expected lending fees did the pricing premium reflect on that day?"

Table 2 shows the results of this estimation, along with t-stats for two null hypotheses - the first is that the ratio K_t is zero. The second null hypothesis is that the coefficient K_t is equal to 1. The coefficient on the expected value of future lending fees is plotted in Figure 8. For most weeks, I can clearly reject the null that specialness is *not* priced in. It is also apparent that the ratio of the convergence profit to the total amount of fees paid in the form of specialness is less than one for all weeks. For two weeks, it is statistically different from 1. Moreover, the ratio is decreases with time - starting at a value of around 0.8 in the immediate weeks after issuance, and declining to 0.3 as the issuance of the next security draws closer. Note that the standard errors are particularly wide for the later weeks, because the ratio is computed using two very small quantities, both of which are measured with a considerable degree of error. The correspondence between 8 and 5 is clear. Furthermore, Table 3 shows a regression of the estimated proportion on the weeks past the auction of the security for which the ratio is computed. Even with the small number of data points available, the relationship between this ratio and the time since issuance of the security is negative and statistically significant. The decline in this ratio supports our theory that the agents who buy the bond later into the auction cycle do not avail of the entire amount of the (typically higher)

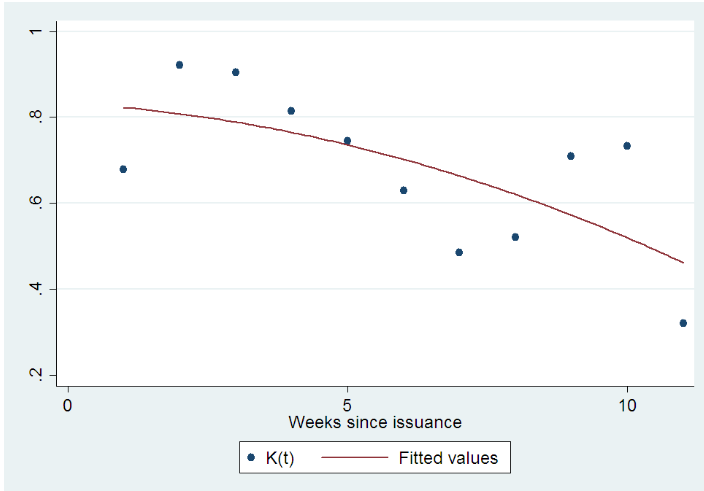

Figure 8: **Coefficient on expected total future lending fees priced in:** This figure shows the results for the computation of the average convergence in the price premium, against the average future lending fees, as on any week after issuance. The ratio denotes the sample estimate of the quantity K_t where $K_t = \frac{E_t(P(t)-P(T))}{E_t(L(t))}$. We separately compute K_t for each value of t. The methodology for computation of this estimate is as explained in section 3. The price premium is computed as the price reduction a note will have to go through to have the same spread over the swap rate as an off the run note issued two auctions prior. Specialness is computed as the difference between the special repo rate as published by GovPx and the Treasury general collateral rate for the same day. The data consists of 38 auction cycles for the ten year note, with a quarterly frequency, over a period from November 1995 to June 2007.

lending fees that are observed, and they price the bonds accordingly.

The second test involves the weekly profits on the convergence trade, and is presented in table 4. In almost all weeks, the mean profit on a convergence trade involving short selling the on-the-run security makes losses. However, because of the noisiness of the data series, we are able to reject the null hypothesis that the profit on the convergence trade is zero only for two out of the eleven weeks over which we compute the profits. Figure 9 graphically depicts the values and the confidence intervals of the weekly profits. As can be seen, the profits are more negative in the latter half of the auction cycle when repo-specialness is high, as compared to the beginning of the auction cycle. This lends further support to the results presented above.

3.3 Economic drivers of hedging benefits

In this section, I describe some of the economic drivers of hedging benefits that give rise to specialness in the repo rates in the Treasury market. Recall that our model implies that specialness and price premia are both linear in hedging benefits derived by agents in the economy when they hedge

Week	Price convergence (%) (1)	Dollar value of fees paid (%) (2)	Ratio (3) = (1)/(2)
1	0.135	0.199	0.679
			(3.48)**
			(1.65)*
2	0.177	0.192	0.922
			(4.80)**
			(0.41)
3	0.162	0.179	0.905
			(4.47)**
			(0.47)
4	0.136	0.167	0.814
			(3.73)**
			(0.85)
5	0.110	0.148	0.745
			(3.18)**
			(1.09)
6	0.081	0.129	0.629
			(2.64)**
			(1.55)
7	0.057	0.118	0.485
			(2.03)**
			(2.15)**
8	0.052	0.099	0.521
			(1.78)*
			(1.64)
9	0.057	0.080	0.710
			(2.16)**
			(0.88)
10	0.046	0.063	0.733
			(2.13)**
			(0.78)
11	0.013	0.042	0.322
			(0.60)
			(1.27)
12	0.004	0.024	0.181
			(0.21)
			(0.95)

Table 2: **Proportion of future lending fees priced in:** This table shows the results for the computation of the average convergence in the price premium, against the average future lending fees, as on any week after issuance. The ratio denotes the sample estimate of the quantity K_t where $K_t = \frac{E_t(P(t)-P(T))}{E_t(L(t))}$. We separately compute K_t for each value of t. Two t-stats are shown. The first t-stat is for the null that $K_t = 0$. The second t-stat is for the null that $K_t = 1$. Standard errors are computed using the Delta method, using the sample variance of $P^i(t) - P^i(T))$, $L^i(t)$, and the covariance between $P^i(t) - P^i(T)$ and $L^i(t)$, where i is an index that denotes the auction cycle. The methodology is as explained in section 3. The price premium P(t) is computed as the price reduction a note will have to go through to have the same spread over the swap rate as an off the run note issued two auctions prior. Specialness is computed as the difference between the special repo rate as published by GovPx and the Treasury general collateral rate for the same day. The data consists of 38 auction cycles for the ten year note, with a quarterly frequency, over a period from November 1995 to June 2007.

Proportion of future lending fees priced in	
Weeks since issuance (t)	-0.047
	(3.61)**
Constant	0.942
	(9.85)**
Observations	12
R-squared	0.57

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

Table 3: **Proportion of future lending fees priced in by weeks past issuance:** This table shows the results of a regression of the ratio of price premium to the dollar value of lending fees paid, with the number of weeks after the issuance of the security that the ratio is computed for. The ratio denotes the sample estimate of the quantity K_t where $K_t = \frac{E_t(P(t)-P(T))}{E_t(L(t))}$. We separately compute K_t for each value of t. The figures in parentheses are t-stats against the null hypothesis that the coefficient is 0. The price premium P(t) is computed as the price reduction a note will have to go through to have the same spread over the swap rate as an off the run note issued two auctions prior. Specialness is computed as the difference between the special repo rate as published by GovPx and the Treasury general collateral rate for the same day. The data consists of 38 auction cycles for the ten year note, with a quarterly frequency, over a period from November 1995 to June 2007.

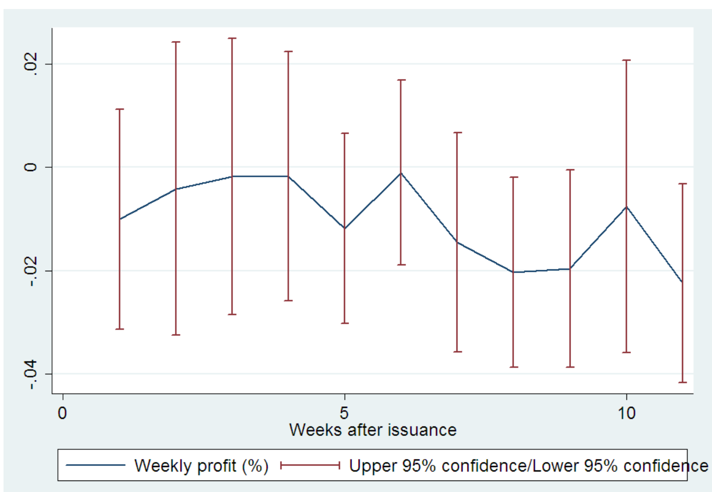

Figure 9: **Weekly profits by weeks since issuance:** This figure shows the means and confidence intervals for the profit made by a duration neutral strategy involving a notional initial amount of 100 dollars, with a short position in the on-the-run security and a long position in an equivalent amount of the third off-the-run security, by week after issuance of the on-the-run security. The portfolio consists of a short position of 1 unit in the on-the-run security, and a long position in an equivalent amount of the third-off-the run security, and overnight general collateral, so that the overall duration of the portfolio is zero. The interest rate received on the collateral for the security borrowing to maintain the short position is the special repo rate for 10 year on-the-run securities as reported by GovPx. It is assumed that the long position in the off-the-run security is funded by lending them out at the general Treasury collateral repo rate. The data consists of 38 auction cycles for the ten year note, with a quarterly frequency, over a period from November 1995 to June 2007.

Week since issuance	Average profit on convergence trade (bps)	t-stat
1	-0.998	(0.92)
2	-0.411	(0.29)
3	-0.174	(0.13)
4	-0.171	(0.14)
5	-1.181	(1.27)
6	-0.09	(0.11)
7	-1.445	(1.34)
8	-2.025	(2.17)**
9	-1.963	(2.03)**
10	-0.756	(0.53)
11	-2.24	(2.29)**

Table 4: **Average Weekly Profits:** This table shows, by week since issuance of the security, the average weekly profit made on a convergence trade consisting of shorting the ten year on-the-run note and going long on the ten year note issued two auctions prior. The data consists of weekly profit computed on the convergence trade for 38 quarterly auction cycles of the ten year note issued from November 1995 to June 2007. The profit is computed as the realized profit on the strategy, net of fees due to specialness of the on-the-run security paid on the security. I assume that the amount of collateral in order to borrow the on-the-run security is equal to its price. The dollar amount of daily fees are computed using an actual/360 convention, as the difference between the general collateral repo rate and the special repo rate for the on-the-run security, for every day as reported by GovPx. I compute the weekly overall profits, and then group the observations by weeks passed since issuance of the security. For each group, I compute the mean and the standard error of the profits.

against unfavorable moves in interest rates. From this point of view, it is natural to conjecture that repo rate specialness could be related to factors that reflect risk aversion in general in the economy.

I investigate relationships between the need to hedge and economy wide factors that are related to risk premia. The problem with analyzing the raw data series of repo rates is that they are dependent on auction cycle variables such as the length of the auction cycle, and the time since the last on-the-run security was issued. In order to compute a measure that is comparable across time, we need to control for the dependence of special repo rates on variables that are directly related to auction cycle dynamics. I do this by performing a regression of the monthly average repo specialness on the length of the auction cycle, the time since last issuance, and the float of the security issued. This may be thought of as controlling for variables related to the availability of the security in the lending market. These residuals may be autocorrelated, but are consistent. In the second stage, I use the residuals from the first stage regression. I average them at the monthly level and regress the monthly average of these residuals on the return on the market, the return on the Fama-French size and book to market portfolios and on the momentum portfolio for the month. The results are presented in Table 5.

The results of the regression are revealing. There is an expected high correlation between specialness and the risk free rate. This is because the risk free rate acts as a natural upper bound on the level of the special repo rates, since special repo rates cannot be higher than the per period risk-free rate. There seems to be a negative correlation between the excess return on the market and the specialness in the on-the-run security. Seen in the context of the model, this seems to imply that specialness seems to be lower during times when the market excess return is high and vice versa. However, the relationship is not statistically significant. This result parallels the result by Krishnamurthy (2002) who documents that the profit on the on-the-run/ off-the-run convergence trade seems to have a payoff similar to a put option on the market. The results on the Fama-French size portfolio are uncorrelated with specialness. However, the returns on the HML portfolio appear to have a strong relationship with specialness. Specialness is high when low book to market ratio firms have higher returns than high book to market ratio firms. This indicates that special repo rates price in an economy-wide risk premium that is related to the premium on value vs. growth firms. While this seems to be a risk-premium effect, I am unable to make clear statements about why this effect arises in the first place. It is possible that times when the growth firms outperform value firms are also times when there is a high amount of issuance of fixed rate corporate debt. There could also be an economy wide risk premium - the same effects that drive returns on low book to market ratio firms cause agents to derive higher benefits to hedging against adverse moves in interest rates. I leave this question as a matter of future research.

Hedging benefits are an important driver of specialness, and hence price premia. Specialness is an additional source of income for agents who hold on-the-run securities. The negative relationship between specialness and economy-wide sources of risk could have an important bearing on these price premia. Our regressions show that specialness is high during times of market distress. This means that part of the so called "flight to liquidity" effect could arise from not just the desire of agents to hold liquid securities, but because of an increased demand to hedge against adverse moves - when this demand is high, such agents are willing to pay higher fees in order to short the securities, and hence the price of these securities get elevated.

	Specialness (bps)
Issue date	96.723
	(5.27)**
Days since issuance	2.35
	(11.19)**
Auction cycle length	1.003
	(12.78)**
Square of Days from auction	-0.017
	(9.64)**
Log (float)	-124.301
	(12.09)**
Constant	2,076.42
	(12.11)**
Observations	3062
R-squared	0.12

	Monthly residual specialness (bps)
mktrf	-141.1
	(0.95)
smb	-215.7
	(1.36)
hml	-645.4
	(2.88)**
rf	7935.6
	(2.40)*
umd	-111
	(1.14)
Constant	-20
	(2.10)*
Observations	134
R-squared	0.10

Robust t statistics in parentheses
* significant at 5%; ** significant at 1%

Table 5: **Relationship between hedging benefits and Fama French risk factors** This table shows the results of a two step procedure. In the first step, I normalize daily specialness in the on-the-run Treasury security by regressing the daily specialness on auction-cycle dependent variables, such as the days passed since the security was issued, the square of days passed (to account for non-linearity), the log of the float of the security issued, the total length of the auction cycle till the next security is going to be issued in days, and a dummy variable that accounts for any jumps in on-the-run specialness on the day a new on-the-run security is issued. I do this for issuances of the ten year note from the period November 1995 to July 2007. There are 42 auction cycles in all, in this period. Next, I compute the residuals from these regressions. By construction, these residuals do not have any time-series patterns within them. I take the average unexplained specialness for a month, and regress it on the excess return on the market, the risk-free rate, the return on the size portfolio, the return on the book to market portfolio and the return on the momentum portfolio.

4 Conclusion

I present a model of short selling in the presence of hedging needs where security borrowing is restricted by search frictions. Search frictions allow lenders of security to extract rents from borrowers, and this gives rise to price differences between similar securities. I present two cases. In the first case, the demand to sell a security short is driven by the need by market participants to hedge other similar portfolios. In the second case, the demand arises because some agents have a need to hold the security at an elevated price, and arbitrageurs try to benefit from the "overpricing" by selling the security short. In both cases, prices of securities are above the "intrinsic" valuations of the buyers of those securities, because buyers can expect to earn lending fees on their positions in the future. However, I show that in the former case, the fees than an agent has to pay in order to maintain the short position are greater than the expected reduction in price over the convergence horizon. In such a situation, arbitrageurs stay out of the market. The only agents who short-sell securities are those who derive some kind of implicit hedging benefit because they short the security. Furthermore, if there is heterogeneity in terms of how likely holders of the security are to participate in the repo market, this results in a situation where agents who are more likely to participate in the repo market become buyers of the security earlier, when short interest in the security is low. As time passes, the marginal buyers of the security are those who can expect a lower proportion of fees by lending out their securities. Hence, the price progressively incorporates a lower and lower proportion of future shorting fees. The model gives simple testable predictions on whether short selling in a security is driven by hedging demand, or by arbitrage activity.

I apply my model to convergence trades between on-the-run and off-the-run securities in the Treasury market. Previous literature has argued that on average, such convergence trades make zero profits after accounting for shorting fees in the form of repo-specialness. I show that such convergence trades make losses on average. Furthermore, I am able to match observed patterns in special repo rates over regular treasury auction cycles. Observed specialness increases and then falls over the auction cycle of an on-the-run security, while the observed price premium between the on-the-run and off-the-run securities typically declines. I show that this phenomenon is consistent with repo market heterogeneity. On-the-run securities become more special as time passes because when short interest is increases, the buyers of securities tend to those agents who participate to a lesser degree in the repo market, and hence cannot be easily accessed by future borrowers. I also show that the proportion of observed lending fees that are priced in to the on-the-run security is lower after a greater amount of time has elapsed after issuance, which is consistent with phenomenon that the marginal buyers of these securities do not avail of the benefits of lending them out and earning fees on them.

Next, I investigate the economic drivers of these hedging benefits. I find that the hedging needs implied by special repo rates are negatively related to economy-wide risk-factors, such as excess market return, the return on momentum portfolios, and most importantly on the Fama-French book to market portfolio.

Appendix A

Solving for the measures

I denote by the function $F(\sigma, t)$ the measure of agents of type σ who have not yet lent out their securities. Denote by $Q(\sigma, t)$ as an indicator function that takes the value 1 if an agent of type σ has become the marginal buyer at time t, and zero otherwise. Then we have the following identities:

$$\int_0^{\bar{\sigma}} Q(\sigma, t) d\sigma = \sigma(t) \tag{28}$$

The evolutions of the measures are given by:

$$\begin{aligned} \dot{\sigma(t)} &= (S + D - \sigma(t)) \int_0^{\bar{\sigma}} \lambda_\sigma F(\sigma, t) Q(\sigma, t) d\sigma \\ F(\dot{\sigma}, t) &= -\lambda_\sigma F(\sigma, t) Q(\sigma, t)(S + D - \sigma(t)) \end{aligned} \tag{29}$$

Integrating the first equation over a small time interval from t to s gives us:

$$ln\left(\frac{S + D - \sigma(s)}{S + D - \sigma(t)}\right) = \int_t^s \int_0^{\bar{\sigma}} \lambda_\sigma F(\sigma, \tau) Q(\sigma, \tau) d\sigma d\tau \tag{30}$$

where $F(\sigma, s)$ can be expressed in terms of $F(\sigma, t)$ as

$$ln\left(\frac{F(\sigma, s)}{F(\sigma, t)}\right) = -\int_t^s \lambda_\sigma Q(\sigma, \tau)(S + D - \sigma(\tau)) d\tau \tag{31}$$

The solution scheme for this set of equations aims to find a functional fixed point for $\sigma(\tau)$. The nature of the problem lends itself to a contraction mapping procedure. Given $\sigma(t)$, assume that a starting solution is available for the functional form for $\sigma(\tau)$ over the interval given by $t \leq \tau \leq s$. Denote this solution as $\sigma^i(\tau)$. The aim is to get an update of this solution so that the distance metric $sup_t^s|\sigma^i(\tau) - \sigma^{i+1}(\tau)| \to 0$

Given $\sigma^0(\tau)$, the indicator $Q(\sigma, \tau)$ is well defined by equation 28. Given $Q(\sigma, \tau)$, $F(\sigma, t)$ and $\sigma^0(\tau)$, $F(\sigma, \tau)$ can be computed using equation 31. Given $F(\sigma, \tau)$ and $Q(\sigma, \tau)$, compute $\sigma^{i+1}(\tau)$ using equation 30. Compute $sup_t^s|\sigma^i(\tau) - \sigma^{i+1}(\tau)|$. Using $\sigma^{i+1}(\tau)$ as the next starting step, iterate till the distance metric converges to a value below tolerance.

Appendix B

Proof of proposition 2.1

Suppose $\lambda_{\sigma(1)} < \lambda_{\sigma(2)} \Rightarrow V_{\sigma(1)}(t) \geq V_{\sigma(1)}(t)$ over some region $[\underline{T} \leq t \leq \overline{T}\}$, with the equality holding at $\overline{T}$ Then, we have

$$\lambda_{\sigma(1)}\mu^{bo}(t)(L(t) - V_{\sigma(1)}(t)) < \lambda_{\sigma(2)}\mu^{bo}(t)(L(t) - V_{\sigma(2)}(t))$$
$$\Rightarrow -\dot{V_{\sigma(1)}}(t) < -\dot{V_{\sigma(2)}}(t)$$
$$\Rightarrow \dot{V_{\sigma(1)}}(t) > \dot{V_{\sigma(2)}}(t)$$

Integrating from t to $\overline{T}$, and using the fact that $V_{\sigma(1)}(\overline{T}) = V_{\sigma(2)}(\overline{T})$,

$$V(\overline{T}) - V_{\sigma(1)}(t) > V(\overline{T}) - V_{\sigma(2)}(t)$$
$$\Rightarrow V_{\sigma(1)}(t) < V_{\sigma(2)}(t)$$

A contradiction.

Proof of proposition 2.5

If borrowers have all the bargaining power, then borrowers set the gain from lending for any lender they encounter equal to their reservation value. Consider a lender of type σ who encounters a borrower. The reservation value of this lender is $V_\sigma(t)$, while his value after he becomes a lender is given by $V(T) + \int_t^T w(s)ds$. Thus if borrowers have all the bargaining power, we have the identity:

$$V_\sigma(t) = V(T) + \int_t^T w(s)ds$$

or

$$\int_t^T \lambda_\sigma \mu^{bo}(s) e^{-\int_t^s \lambda_\sigma \mu^{bo}(\tau)d\tau} \int_s^T w(\tau)d\tau ds = \int_t^T w(s)ds \tag{32}$$

For $\lambda_\sigma < \infty$, the left hand side of 32 is greater than or equal to the right hand side, with equality at $w(t) = 0 \forall\ 0 \le t \le T$. Hence the lending fee is zero if borrowers have all the bargaining power.

Proof of proposition 2.6

Setting $\theta = 1$ in equation 18 gives us

$$L(t) = \left(P(t) - V(T) + x(T-t) - \int_t^T \frac{-\dot{\mu^{bo}}(s)}{\mu^{bo}(t)} \left[x(T-s) + P(s) - V(T) - L(s) \right] ds \right)$$

$$\Rightarrow$$

$$P(t) - V(T) + x(T-t) - L(t) = 0$$

$$\forall\ 0 \le t \qquad \le T$$

The result follows from differentiating with respect to t.

Proof of proposition 2.7

It suffices to show that the rate at which a borrower may encounter a lender decreases as time passes. The rate B(t) at which a borrower meets a lender at any time t is given by

$$B(t) = \int_0^{\sigma(t)} \lambda_\sigma \mu_\sigma^n(t) d\sigma \tag{33}$$

Consider some $t_2 > t_1$. Then,

$$B(t_2) = \int_0^{\sigma(t_2)} \lambda_\sigma \mu_\sigma^n(t_2) d\sigma$$
$$B(t_1) = \int_0^{\sigma(t_1)} \lambda_\sigma \mu_\sigma^n(t_1) d\sigma \tag{34}$$

We have

$$\sigma(t_2) > \sigma(t_1) \tag{35}$$

From the fact that λ_σ is decreasing in σ, this implies that

$$\begin{aligned} &\lambda\sigma > \lambda\sigma(t_1) \\ &\quad \forall\ \sigma < \sigma(t_1) \\ &\quad > \lambda\sigma \\ &\quad \forall\ \sigma(t_1) < \sigma < \sigma(t_2) \end{aligned} \tag{36}$$

Equation 3 implies that

$$\begin{aligned} &\mu_\sigma^n(t_1) - \mu_\sigma^n(t_2) > 0 \\ &\quad \forall\ 0 \le \sigma \le \sigma(t_1) \end{aligned} \tag{37}$$

From equation 5,

$$\int_0^{\sigma(t_2)} \mu_\sigma^n(t_2) d\sigma = \int_0^{\sigma(t_1)} \mu_\sigma^n(t_1) d\sigma = S \tag{38}$$

Rewriting equation 38, we get

$$\int_0^{\sigma(t_1)} (\mu_\sigma^n(t_1) - \mu_\sigma^n(t_2)) d\sigma = \int_{\sigma(t_1)}^{\sigma(t_2)} \mu_\sigma^n(t_2) d\sigma \tag{39}$$

Multiplying the LHS and RHS of 39 by $\lambda_{\sigma(t_1)}$, and by conditions 37 and 36, we get

$$\begin{aligned}\int_0^{\sigma(t_1)} \lambda_\sigma(\mu_\sigma^n(t_1) - \mu_\sigma^n(t_2))d\sigma &> \int_0^{\sigma(t_1)} \lambda_{\sigma(t_1)}(\mu_\sigma^n(t_1) - \mu_\sigma^n(t_2))d\sigma \\ &= \int_{\sigma(t_1)}^{\sigma(t_2)} \lambda_{\sigma(t_1)}\mu_\sigma^n(t_2)d\sigma \\ &> \int_{\sigma(t_1)}^{\sigma(t_2)} \lambda_\sigma\mu_\sigma^n(t_2)d\sigma \end{aligned} \tag{40}$$

This implies that $B(t_2) < B(t_1)$, that is the rate at which a borrower meets a lender decreases.

Proof of proposition 2.8

Consider a hypothetical buyer of the security who does not face search frictions when searching for borrowers. That is, consider a buyer of type σ_∞ such that $\lambda_\sigma^\infty \to \infty$. Since such a buyer can instantly lend out his security, his valuation at time t is given by:

$$V^\infty(t) = V(T) + L(t) \tag{41}$$

If $\sigma(t)$ denotes the marginal buyer at any time t, then by proposition 2.1,

$$\begin{aligned}V_{\sigma(t)}(t) &< V^\infty(t) \\ &< V(T) + L(t)\end{aligned}$$

Using Equation 11 this implies that:

$$P(t) - V(T) < L(t)$$

or

$$\frac{P(t) - V(T)}{L(t)} < 1 \tag{42}$$

The second part of the proposition follows from noting that, for $t_2 > t_1$, $\sigma(t_2) > \sigma(t_1)$, which implies that $\lambda_{\sigma(t_2)} < \lambda_{\sigma(t_1)}$. By proposition 2.1, this implies that $V_{\sigma(t_1)(t_2)} > V_{\sigma(t_1)(t_1)}$. For obvious reasons:

$$\frac{V_{\sigma(t_1)(t_2)} - V(T)}{L(t_2)} > \frac{V_{\sigma(t_1)(t_1)} - V(T)}{L(t_2)} \tag{43}$$

Define $K_{\sigma(t_1)}(t) = \frac{V_{\sigma(t_1)}(t) - V(T)}{L(t)}$. Since $V_{\sigma(t_1)(t)} \leq V^l(t)$ for all values of t and t_1, with the equality holding *only* at $t_1 = t = T$, we have $-d(V_{\sigma(t_1)(t)}) > w(t)dt$. Multiplying the left hand side by $L(t)$ and the right hand side by $V_{\sigma(t_1)}(t)$ preserves the inequality, and thus we get

$$-d(V_{\sigma(t_1)(t)})L(t) > w(t)V_{\sigma(t_1)}(t)dt \tag{44}$$

Finally, by taking the derivative of $K_{\sigma(t_1)}(t)$, equation 44 implies that

$$\frac{d(K_{\sigma(t_1)}(t))}{dt} < 0 \tag{45}$$

This implies that

$$\frac{V_{\sigma(t_1)(t_1)} - V(T)}{L(t_1)} > \frac{V_{\sigma(t_1)(t_2)} - V(T)}{L(t_2)} \tag{46}$$

Equations 43 and 46 taken together imply the second result.

Also, taking the derivative of equation 42, we get

$$-d(P(t)) > w(t)dt \tag{47}$$

This means that a hypothetical convergence trade makes losses.

Proof of proposition 2.9

Setting $\lambda_\sigma = 0$ in equation 9 implies that

$$P(t) = V(T)$$

Equation 18 can hold only at $w(t) = 0$, similar to the logic used in the proof of proposition 2.6. Similarly, suppressing the index for the type of the lender, $\lambda \to \infty$ implies that

$$V^l(t) - V(t) \to 0$$

which because of equation 17 implies that

$$V^{bo}(t) - V^s(t) \to 0$$

which, again is possible for all t, only if $w(t) \to 0$.

Appendix C

Recall equation 25 which implies that:

$$E_t(P(t) - P(T)) = K_t E_t(L_t)$$

For each auction cycle i, it implies that:

$$E_t^i(P^i(t) - P^i(T)) = K_t E_t^i(L^i(t))$$

Taking the unconditional expectations of both sides,

$$E(E_t^i(P^i(t) - P^i(T))) = K_t E(E_t^i(L^i(t)))$$

Now, appealing to the law of iterated expectations,

$$\begin{aligned} E(E_t^i(P^i(t) - P^i(T))) &\equiv \frac{1}{I}\sum_{i=1}^{I}(P^i(t) - P^i(T)) \\ E(E_t^i(L^i(t))) &\equiv \frac{1}{I}\sum_{i=1}^{I}(L(t)) \\ K_t &\equiv \frac{\sum_{i=1}^{I}(P^i(t) - P^i(T))}{\sum_{i=1}^{I}(L(t))} \end{aligned}$$

The standard errors around K_t can be formed using the Delta method, as follows. Denote by $\epsilon_P^i(t)$ as the difference between the quantity $E_t^i(P^i(t) - P^i(T))$, and its realization $P^i(t) - P^i(T)$, and $\epsilon_L^i(t)$ as the difference between $E_t^i(L^i(t))$ and its realization $L^i(t)$. Let $E_t^i(P^i(t) - P^i(T))$ be distributed with mean $E(E_t^i(P^i(t) - P^i(T)))$, and variance $\sigma_p^2 = E\left(E_t^i(P^i(t) - P^i(T)) - E(E_t^i(P^i(t) - P^i(T)))\right)^2$.

Then,

$$
\begin{aligned}
\sigma_p^2 &= E\left(E_t^i(P^i(t) - P^i(T)) - E(E_t^i(P^i(t) - P^i(T)))\right)^2 \\
&= E\left(E_t^i(P^i(t) - P^i(T))^2\right) - 2E\left(E_t^i(P^i(t) - P^i(T))E\left(E_t^i(P^i(t) - P^i(T))\right)\right) \\
&+ E\left(E_t^i(P^i(t) - P^i(T))\right)^2 \\
&= E(\epsilon_P^i(t)^2) \qquad (48)
\end{aligned}
$$

Similarly let $E_t^i(L^i(t))$ be distributed with mean $E(E_t^i(L^i(t)))$ and variance σ_l^2. A similar argument shows that

$$\sigma_l^2 = E(\epsilon_L^i(t)^2) \qquad (49)$$

and the covariance term is

$$\sigma_l p = E(\epsilon_L^i(t)\epsilon_P^i(t)) \qquad (50)$$

The quantities on the right hand side of 48, 49 and 50 can be estimated from their sample moments. Appealing to the Delta method, the asymptotic variance of K_t is given by:

$$
\begin{aligned}
Asy.Var(K_t) \equiv & \frac{1}{\left(\frac{1}{I}\sum_{i=1}^{I}(L(t))\right)^2}\sigma_l^2 \\
& + \frac{\left(\frac{1}{I}\sum_{i=1}^{I}(P^i(t) - P^i(T))\right)^2}{\left(\frac{1}{I}\sum_{i=1}^{I}(L(t))\right)^4}\sigma_p^2 \\
& + 2\frac{\left(\frac{1}{I}\sum_{i=1}^{I}(P^i(t) - P^i(T))\right)}{\left(\frac{1}{I}\sum_{i=1}^{I}(L(t))\right)^3}\sigma_l p
\end{aligned}
$$

References

Cohen, L., K.B. Diether, and C.J. Malloy, 2007, Supply and demand shifts in the shorting market, The Journal of Finance (forthcoming).

D'Avolio, Gene, 2002, The market for borrowing stock, *Journal of Financial Economics* 66, 271–306.

Duffie, Darrell, 1996, Special repo rates, *Journal of Finance* 51, 493–526.

———, Nicolae Gârleanu, and Lasse Heje Pedersen, 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics* 66, 307–339.

Evans, Richard B., Christopher C. Geczy, David K. Musto, and Adam V. Reed, 2005, Failure is an Option: Impediments to Short Selling and Options Prices, *SSRN eLibrary.*

Fisher, M, 2002, Special repo rates-an introduction, *Federal Reserve Bank of Atlanta Economic Review.*

Fleming, MJ, and K Garbade, 2003, The specials market for us treasury securities and the federal reserves securities lending, Federal Reserve Bank of New York.

———, 2004, Repurchase agreements with negative interest rates, *Federal Reserve Bank of New York Current Issues in Economics and Finance.*

Graveline, JJ, and MR McBrady, 2005, Who makes on the run treasuries special?, Working Paper (Stanford Graduate School of Business).

Jordan, BD, and SD Jordan, 1997, Special repo rates: An empirical analysis, *The Journal of Finance.*

Krishnamurthy, A, 2002, The bond/old-bond spread, *Journal Of Financial Economics* 66, 567–574.

Lamont, O.A., and R.H. Thaler, 2003, Can the market add and subtract? mispricing in tech stock carve-outs, *Journal of Political Economy* 111, 227–268.

Nyborg, K.G., and S. Sundaresan, 1995, Discriminatory versus uniform treasury auctions - evidence from when-issued transactions, *The Journal of Finance* 50(3), 999–1000.

Ofek, Eli, and Matthew Richardson, 2003, Dotcom mania: The rise and fall of internet stock prices, *The Journal of Finance* 58, 1113–1138.

Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *The Journal of Finance* 52, 35–55.

Vayanos, D., and P.O. Weill, 2007, A search-based theory of the on-the-run phenomenon, Journal of Finance (Forthcoming).

Wurgler, J., and E. Zhuravskaya, 2002, Does arbitrage flatten demand curves for stocks?, *Journal of Business* 75, 583–608.

Xiong, W., 2001, Convergence trading with wealth effects: an amplification mechanism in financial markets, *Journal of Financial Economics* 62, 247–292(46).