Are security lending fees priced? Theory and evidence from the U.S. treasury market

Amrut Nashikkar

October 24, 2007

Abstract

I present a dynamic model of short selling when the security borrowing for the short sale is restricted by search frictions. I analyze the case where short-selling demand is driven by hedging needs and contrast it with the case when short-selling demand arises from arbitrageur activity. Frictions lead to security borrowing fees and price premia arise from expected future fees in both cases. However, in the former, price premia are lower than the value of observed future borrowing fees. The Duffie(1996) result that prices exactly incorporate all future lending fees arises as a special case in the latter. The model features heterogeneity in the extents to which holders participate in repo markets, and explains why securities become “harder to locate” when the short interest is high. I apply the theoretical results to a time-series of repo-market fees and price premia in “on-the-run” U.S. treasury securities and show that repo fees in Treasury securities are primarily driven by hedging needs of short-sellers, rather than arbitrage activity. Additionally, I am able to match the dynamics of auction-cycle related patterns in prices and repo fees. I also show that the hedging benefits implied by specialness in U.S. treasuries are related to economy-wide risk factors, such as the return on the Fama-French book-to-market portfolio, the level of interest rates and the slope of the term-structure.

JEL Classification: G100 (General Financial Markets)
Keywords: Fixed Income, Repo Markets, Liquidity, Derivatives
1 Introduction

In order to create a short position in a security, it has to be borrowed and sold to another investor at a price that is agreeable to the latter. Thus, security borrowing and lending markets play a vital role in the financial markets because they allow agents to short sell securities for purposes such as hedging, or for speculating on anticipated price declines. Frictions in these markets can therefore have an impact on asset prices and can lead to short selling becoming costly when the demand to short-sell is high. This entails paying an implicit fee in the form of a special repo rate. The fee is can be thought of as “specialness”, which is the difference between the general collateral repo rate that the collateral would earn if the security was not in demand, and the lower special repo rate that the collateral earns because the security is in demand when the security is borrowed for the purpose of short selling it. 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 that has been sold short 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.

In this paper, I investigate situations where short sales in a security are restricted by the difficulty of finding a security to borrow in the market for borrowing and lending securities1, henceforth called the repo market. There is no inherent risk - the friction only arises from the difficulty of search 2. There are two channels that drive the demand to short-sell. The first, and an important driver in some markets, such as treasury securities, is the desire of some agents to hedge against unfavorable moves in other portfolios they hold. This “hedging demand” to short 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 even in situations where there are no optimistic

1Securities 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.

2A discussion of other papers that use the theme of search frictions in repo markets, viz Duffie, Garleanu and Pedersen (2002) and Vayanos and Weill (2007) is included in the literature review.
beliefs amongst holders about the intrinsic value of a security. The second case is the demand by arbitrageurs who seek to profit from an expected decline in the security’s prices. In the latter case, there is demand to *hold* the security at an inflated price, hence, the latter may also be thought of as short selling driven by the demand to hold the security. 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. This is because repo markets frictions allow holders of the security to extract rents from future borrowers, thus endogenously raising the price of the security above their own valuations.

In the setting of this paper, the Duffie (1996) result that price premia are equal to observed future lending fees (or 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, and the borrowers of the security have no bargaining power. When bargaining power is distributed across lenders and borrowers, I show that security-borrowing demand arising purely out of the hedging needs of short-sellers has different implications on the relationship between observed “over-pricing” and the observed lending fees, as compared to security-borrowing demand arising out of arbitrage. This leads to a simple test of determining whether fees in the repo market are driven by the hedging needs of short-sellers, rather than arbitrage activity - in the former case, a hypothetical convergence trade makes *losses*.

As a baseline case, I focus on demand to borrow securities being driven by the hedging needs of short-sellers. The case of short-selling driven purely by arbitrageur demand is a simple modification of this case. In the baseline 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 as on a certain termination date in the future. One of the securities is perfectly illiquid and hence cannot be sold short - the other is liquid and hence the demand to sell short is concentrated in the latter. The setting tries to capture several stylized facts about the process of short selling - 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. Secondly, 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. Thirdly, market participants often report that securities become scarcer
to locate in the repo market when the existing short interest in the security is high. Since every
new short position creates a new long position, which is in itself potentially lendable, securities can
become difficult to find only if the new buyers of the securities are harder to be located by borrowers
of the securities. I capture this effect by modeling heterogeneity in terms of the participation in
the repo-market by the “longs”.

I model 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 unfulfilled security borrowers, and the difficulty in finding a security lender. The price reflects the
valuation of the agent buying the security from the short seller, and not the valuation of existing
holders of the security. The fees are determined by the equilibrium of a bargaining game between
lenders and borrowers. Buyers 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 is high.
An obvious analogy is to a market with investors with a differing tax rates - low tax investors
in equilibrium have higher valuations for a highly taxable security (for example, one with more
income and lower capital gains), and when the supply of the security increases, progressively higher
tax rate investors become the marginal buyers. In the the setting I use, the supply of the security
increases over time because short interest builds up, and hence agents who participate less in the

\footnote{See, i.a. Fisher (2003) and Molton (2004).}
repo market progressively become marginal buyers. However, for borrowers in the repo market, this has an added implication - it 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 are those that stand to gain less from lending out their securities. This implies that the proportion of the observed future specialness that is incorporated into the price decreases.

In order to explain the intuition of the model, consider the following situation where there is one unit of security that is initially issued, whose price after two periods is known and agreed upon by all agents to be 100. We assume that the repo market clears in steps and the cash market is frictionless. There are two agents who wish to short the security (Shorter 1 and Shorter 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 assume that the agent who take long positions are risk-neutral, but participation in the repo market entails some per period transaction costs, which differ between the longs. Let us say these are 1, 5 and 10, for lender A, B and C, respectively. There are gains to shorting 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.

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 as 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 unfolded 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 the first period, and the price at the end of the first period incorporates only 0% of all the future lending fees observed starting the second period. The initial price declines, the per period fee increases, and the proportion of future lending fees priced in declines. It is easy to show that neither lender B, nor lender C can match the initial valuation of lender A, which is 104. This crude model does not explain how the measures of shorters and agents willing to lend the security evolve over time - it however, 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 a search-based model. There are no explicit transaction costs. The cost is 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. This allows us to jointly model both the dynamics of the allocations and the prices of the fees, which can be tested empirically.

In the empirical section of the paper, I apply the theoretical results to one of the most widely known convergence trades - that involving short selling an on-the-run treasury security to benefit from the expected decline in its price. Short positions in on-the-run treasury securities are commonly used by market participants to hedge other fixed income portfolios such as corporate bonds, agency securities, and interest rate derivatives. This hedging demand to short-sell is quite large - for instance, Vayanos and Weill (2007) estimate the short interest in a typical on-the-run five year note to be around 150%. Moreover, these securities are issued in more or less regular cycles, and thus their dynamics are observable and have clear auction-cyclical patterns. I show that not only
do these trades, on an average, lose money, but also that the dynamics of the price premia and the observed specialness in the market can be adequately explained by the search-based setting that I demonstrate in the theoretical section.

Finally, I link the implicit hedging benefits implied by on-the-run special repo rates with economy-wide factors related to risk. I find that several factors known to give rise to risk premia in the asset pricing literature seem to be directly related with these hedging benefits. Prominent amongst these are the market excess return, the return on the Fama-French book-to-market portfolios, and to a lesser degree, the momentum portfolio.

This paper is divided into the following sections. Section 2 includes a brief discussion of existing literature in this area, and where the theoretical and empirical results I present fit in. Section 3 presents the model intuition and the main model itself. The proofs of the propositions are in Appendix A and B. Section 4 describes the data, the test design I use to evaluate the predictions of the model, and discusses the implications. It also includes a discussion on economy-wide factors that seem to affect hedging needs, as measured by a stationary transformation of specialness. Section 5 concludes and outlines further research.

2 Overview

There is a long line of literature on the impact of short-sales constraints on the prices of securities. One of the earliest papers in this field is Miller (1977) who argues that in the presence of absolute short sales constraints, the price does not account for the valuations of the most pessimistic investors because they get left out of the market. Similar arguments are used by Jarrow (1981), and Figlewski (1981), who uses the level of short interest as a proxy for unobserved shorting demand, and argues that stocks that have high short interest tend to underperform. Lamont and Thaler (2003) use similar arguments for the Palm/3Com case. While these papers do not explicitly consider “costly” shorting, Ofek and Richardson (2003) show that during the dot-com boom, Internet stocks have lower rebate rates than non-internet stocks and they tend to underperform. The logic behind this is that “rational” agents seek to profit against irrational over-valuation of these stocks, by short selling them. This drives up the demand to borrow them, and the consequently borrowers of these
stocks end up having to pay higher fees. D’Avolio (2002) is one of the earliest papers to study the determinants of fees implied by rebate rates on stock borrowings. He finds that stocks with larger float and higher institutional earnings tend to be “easier” to borrow in the sense that they tend to have lower levels of specialness. The implicit argument is that the transaction cost of borrowing such securities is lower - which is in itself, fairly strong evidence of search frictions. Evans, Geczy, Musto, and Reed (2005) show the effect of rebate rate specialness in the equity borrowing and lending market on deviations in stock prices from synthetic stock prices implied by options trading on them. More recently Nashikkar and Pedersen (2007) show that corporate bonds that have shorting fees associated with them tend to be more expensive relative to their CDS contracts, than those that do not have shorting fees associated with them.

The earliest paper that explicitly models equilibrium in the repo market, and its effect on security prices is that of Duffie (1996) who presents a model of costly short sales. In this setting, lenders incur a transaction cost to lend a security. Investors pay an implicit fee to borrow the security to meet short selling obligations. There is heterogeneity in the costs that lenders incur in order to lend. In equilibrium the fee is determined by lending cost of the marginal lender of the security. The static equilibrium presented in Duffie (1996) is consistent with lenders starting off with endowments of securities rather than wealth. By itself, the model does not explain how these agents come to acquire the securities in the first place.5 However, the implication of this model is that prices fully incorporate all future lending fees, and is one of the null hypotheses against which I test my empirical results.

This paper has a setting that is similar to Duffie, Gârleanu, and Pedersen (2002), with the notable difference that in their setting, agents have different beliefs about the intrinsic value of a security that gets “revealed” at a termination date. The prices and repo fees are set by a bargaining process between optimistic and pessimistic investors who are constrained by the need to search for a counterpart. All agents are equally likely to borrow or lend in the repo market. In the setting of Duffie, Gârleanu, and Pedersen (2002), shorting is driven by differences in beliefs about the value of a security that gets revealed at a certain stochastic arrival time. Their model takes the

5Since different lenders have different marginal costs of lending, their ex-ante valuation inclusive of lending fees and net of lending costs would be different.
differences in valuations as given. Since different agents have different beliefs about the value of a
security, this setting is difficult to apply to an apparently risk-less convergence trade. In applying
this model to treasury markets, for instance, it is unclear why there would be significant differences
in beliefs between holders of any given treasury. Moreover, the convergence horizon in a typical
on the run/off the run trade is known to the market. Additionally, the dynamics of per period
lending fees in the model are inconsistent with observed dynamics in the treasury market, which is
an application that I am particularly interested in. This is because the fees in their model decline
as the short interest builds up.

The principal point of departure of the model presented in this paper from Duffie, Gărleanu,
and Pedersen (2002) is that it has no differences in the intrinsic valuations of the agents: All agents
have the same beliefs about the terminal value of the security. Rather, buyer heterogeneity is
characterized by differences in the ease with which borrowers of securities can approach them. This
is more in the spirit of Duffie (1996) where the heterogeneity of the “longs” is characterized by
differing marginal transaction costs of participation in the repo market, and better approximates
a market where some agents are constrained from lending their securities for either internal, or
regulatory reasons. Moreover, in the setting in this paper, the demand to short-sell is driven by
a per period hedging benefit that borrowers derive if they acquire a short position. This may be
thought of as a “reduced-form” approach of modeling the portfolio optimization problem of agents
whose endowment streams may be positively correlated with the security dividend process.

In a recent theoretical paper, Vayanos and Weill (2007) present a steady state model in which
the borrowing process is driven by agents who get “hedging benefits” by shorting the security,
with a mechanism similar to the one I use. In their setting, holders of securities also derive some
per period benefits from holding a security, while the agents who short are those who derive some
(positive) per period benefit from taking a short position. They model two essentially identical
assets and show that liquidity differences can arise endogenously, and lead to all the short-selling
concentrating in the asset that has greater liquidity. Agents who wish to hedge by taking a short

---

6Note, however Pasquarello and Vega (2007) who present a model featuring differences in valuations arising out
of information asymmetry in the treasury market, immediately after auctions. More on this later.
7I defer an explicit derivation of this reduced form to prior literature in search and asset prices. See, i.a., Vayanos
position end up paying a fee to do so, which is determined by bargaining. They do not explicitly follow the dynamics of a security after it is issued. Their model also displays the feature of my model in that at any point of time, only a fraction of all the lending fees that are observed in the future are incorporated into the price. In addition, their calibrations indicate that a substantial part of the price premium observed is attributable to lending fees - the innate “liquidity” premium arising out of a demand to hold liquid assets is rather small. Since theirs is a steady state model, it cannot explain the dynamics of prices and repo specialness. In the dynamic setting that I use, the per period fee that I compute can be higher than the hedging benefit that short sellers derive, because in addition to paying fees to borrow the security, they benefit from its expected reduction in price. This feature is impossible to capture in a steady state model. The model I present also provides a natural explanation for why securities might get “locked away” and be perceived by the agents wishing to short-sell as being unavailable to borrow, when short interest is very high.

Very closely connected with the empirical studies I perform for the treasury repo market is the literature on asset pricing and liquidity, specifically as it relates to the on-the-run/off-the-run effect. In fact, this paper can be thought of as directly explaining the dynamics of on-the-run/off-the-run spreads in so far as they relate to a premium related to expected lending fees. I do not directly address the issue of liquidity differences between on-the-run and off-the-run - price premia do not arise out of liquidity differences in my model, but endogenously because of future lending fees. Krishnamurthy (2002) argues that the expected lending fees from specialness in on-the-run bonds are direct compensation for the fact that they have a liquidity premium, and hence lower expected returns. However, there has been recent literature\(^8\) that finds a poor relationship between liquidity differentials between on-the-run and off-the-run bonds, and the associated price premia. In my model, as long as there is a demand to short the liquid on the run security for purposes (such as hedging) over and above any motivation to exclusively profit from its anticipated price decline, price premia arise. An agent who does not have hedging needs never finds it profitable to short the security, in contrast with Krishnamurthy (2002) where arbitrageurs bring about an equilibrium between the repo market where the liquidity premium is exactly offset by future lending fees.

\(^8\)Refer Pasquareillo and Vega (2007)
The other main theme in the literature on the on-the-run/off-the-run effect in treasury securities has to do with liquidity premia related to transaction costs. The liquidity differences between treasuries are documented, amongst others, in Amihud and Mendelson (1991), who relate the liquidity premium between treasury bills and old notes maturing on the same day to their bid/ask spreads. Babbel, Merrill, Meyer, and de Villiers (2002) do a price impact study on large trades in off-the-run securities. They find that price impact in the on-run security is significantly lower than the price-impact in the off-the-run security. They argue that off-the-run treasury securities do not have close substitutes, or that even the liquid on-the-run securities cannot be used as substitutes for the off-the-run securities. The authors argue that this is evidence of a segmented market. They also find that price impact is greater for older securities. This implies that as a treasury ages, it becomes more and more “locked up” with investors that are unwilling to participate in the secondary market, and this drives up liquidity premia. Amihud and Mendelson (1986) argue that transactions costs create liquidity premia, and the premium reflects the differing expected returns for investors with different holding times who have to defray the transactions costs over differing lengths of time. There is an implicit clientele effect, in which securities that are more illiquid and are cheaper as a result are held in equilibrium by investors with longer holding periods. There have been several modifications to this approach. Amihud and Mendelson (1991) and Elton and Green (1998) study the differences between pairs of securities that have similar cash-flows and different prices. Elton and Green (1998) find that tax-effects can explain some of the differences between prices. Boudoukh and Whitelaw (1991) and Eom, Subrahmanyam, and Uno (1998) document the liquidity premium in Japanese Government Bonds. Longstaff (2004) compares Treasuries with Refcorp bonds, which have a similar default risk and finds evidence of a large liquidity premium in US treasuries. He argues that most of this is due to a “flight to liquidity” premium in Treasury securities, and that most of this premium cannot be explained by either transactions costs or by repo-rate differences between Refcorp bonds and treasury bonds.

It must be noted that the liquidity premium theory for the on-the-run/off-the-run effect has been contested in recent literature. In a recent paper, Pasquarello and Vega (2007) document that there is a poor relationship between liquidity of on-the-run and off-the-run bonds as measured
by their bid-ask spreads and the pricing premium observed. They present a model of information asymmetry in which sophisticated traders who have acquired the on-the-run bond skew the price of the bond toward their allocations. As evidence they find higher premia in on-the-run bonds, immediately after auction announcements, when dispersion of beliefs is high, and macroeconomic announcements are noisy. These results are not inconsistent with the model I present. Immediately after auction announcements, total future expected lending fees are high in my model. When macroeconomic announcements are noisy, the demand to short the bonds is also likely to be high because of hedging needs, leading to higher price premia. In this sense, one can view their theory as explaining the mechanism that generates the demand to short securities in the model I present.

One assumption that my model shares with Duffie, Gárleanu, and Pedersen (2002) is that the cash market for trading bonds is frictionless while the repo markets have frictions. This assumption has some strong justification in view of stylized facts and recent literature. The cash market for treasuries is one of the most liquid markets in the world, while the repo market is relatively more opaque. In a recent paper, Barclay, Hendershott, and Kotz (2006) find that while the market share of voice brokers in the cash market for trading treasury notes is around 19%, while it is around 61% for repo transactions. They argue that electronic brokers provide low-cost execution for routine trades, while voice brokers provide additional services that are most valuable for markets that are less active, and for more difficult trades. This is support for our assumption of liquid cash markets. Note that in absolute terms, both markets are highly liquid. In fact, Vayanos and Weill (2007) show that even search time of the order of an hour can generate significant specialness in the repo markets.

Special repo rates are commonly thought to have a natural lower bound of zero (or specialness has a natural upper bound in the form of the general collateral rate), because there are no explicit monetary costs to a failure to deliver. In my model there is no upper bound on the lending fee. In fact, Fleming and Garbade (2004) and Fleming and Garbade (2003) demonstrate instances when the special repo rate was actually negative\(^9\), and explore the implications of such situations. They argue that this is evidence of intangible costs arising from the failure of settlement in securities

\(^9\)This typically happens on “guaranteed delivery repo contracts”, which a lender enters into only when he can deliver the securities with certainty.
which have large short interest. Of special relevance to my paper is the exposition of repo rate patterns over auction cycles given by Fisher (2002), who demonstrates the rising specialness after as time passes after an auction, and the decline in the price premium.

More recently, Graveline and McBrady (2005) study the determinants of specialness in the repo market and argue that the repo rate specialness arises to a greater extent from demand to short-sell the more liquid security, than the demand to hold the security. They use several variables that act as proxies for shorting demand, such as implied volatility of interest rate caps, and large issuances of fixed rate corporate debt, and several proxies for liquidity based demand, such as the commercial paper spread and long term bond swap spreads, and show that that special repo rates are related to proxies for the demand to short securities, rather than the demand to hold liquid securities. The problem with this approach is that it is difficult to cleanly identify proxies for demand to hold securities from the demand to short-sell securities, since both are fundamentally related to the liquidity of the security. The results I present in this paper are perfectly consistent with their results.

3 Model

The basic idea behind the model is simple. It is that in a dynamic setting, at any given point of time, the price of a security is not the valuation (inclusive of lending fees) of an agent that has already become a lender, but the valuation of the agent who buys the security from the agent short-selling it. In a search-constrained repo market, this agent does not immediately realize future lending fees, but must wait before he can realize them. (The “transaction cost” interpretation of this is that a buyer of a security must incur some transaction costs before he becomes a lender.) Because of this, at any given point in time, we cannot expect the price to incorporate all of future lending fees. If buyers are characterized by heterogeneity in repo market participation, the valuation inclusive of future lending fees of an agent who participates in the repo market to a greater extent is higher. This means that when a security is first auctioned, it is likely to be bought principally
by agents who participate to a greater extent in the repo market. Over time, shorters acquire short positions by borrowing the security and selling it to agents who participate to a lower extent in the market (or have higher transaction costs). This has two implications. Firstly, the proportion of observed future lending fees that is incorporated into the price reduces (compare this with the Duffie (1996) setting where all future lending fees are incorporated into the price). Secondly, this reduces the ease with which future borrowers can borrow securities.

Typically, repo trades are settled on the same day with an exchange of cash and collateral happening simultaneously. On the other hand, most trades in the cash market are settled the next day. Because of this kind of staggered settlement, agents may sell a security even though they do not possess it. Agents will meet a commitment to meet a sale delivery obligation on the next day by borrowing securities through a reverse-repo for the special security. Sometimes, failures to settle may occur if the agent is unable to borrow the security. The model I describe has no delivery failures, because agents always borrow the security before they commit to selling the security. It may be thought of as describing “guaranteed delivery” repo contracts of the kind described by Fleming and Garbade (2003).

The model has a finite termination date when shorting activity in a particular security ceases. Hence, the total amount of future lending fees is always decreasing with time even though the per period fee may be increasing. In my model, the price of the security declines for two reasons. Firstly, the total future lending fee decreases. Secondly, the marginal buyers later in time end up having to wait a longer time before they lend their securities out (or face higher transaction costs in order to lend), both because these agents can be accessed with lesser ease by borrowers, and because then number of borrowers remaining is lower. I now describe the set-up of the model, and

\[10\] I abstract away from the issue of non-competitive bidding, although this is likely to have an impact in the sense that a larger amount of non-competitive bidding reduces the amount issued to competitive bidders.

\[11\] When the demand in the borrowing market is high, it may even be possible for such delivery failures to cascade. Sometimes these delivery failures form “daisy chains”, when the number of lenders in the market is limited, where in agent A sells a security to B for next day settlement, and hopes to borrow it from Agent C. Agent B sells it to Agent C for next day settlement, because he expects to receive the security on the next day. However, on the next day, agent A finds that agent C has no security to lend because C has not received the security from B, who, in turn, has not received it from A. Note that this is another kind of cost faced by agents who short sell securities.

\[13\] This can be thought of as the security going “off-the-run”, with the caveat that even securities that have recently gone off the run have significant repo-market specialness, and hence price premia. Hence the “termination date” in my model should be interpreted as being greater than the date when the security goes “off-the-run” in a literal sense.
its solution.

3.1 Set-up

I first describe the set-up of the market. A security is issued at time 0, and has a value (known with certainty) of $V(T)$ 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 3.4. There is a numeraire asset which has the same cash-flows as the recently issued asset. Thus it has a value $V(T)$ at time $T$. I assume that this asset is perfectly illiquid and is not short-sold. In equilibrium, this also implies that the asset is held by agents who do not participate in the repo market at all, and its price is $V(T)$ at all times\textsuperscript{14}.

3.2 Borrowers

In the interval $t \in [0, T]$ there exists a set of agents desiring to short the security. The measure of these agents is given by $D$. Each agent is allowed to short only one unit of the security\textsuperscript{15} An agent who shorts the security earns a per period hedging benefit of $x$.\textsuperscript{16} In order to short a security, 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:

$$\mu_{bo}(t) + \mu_{s}(t) = D \quad (1)$$

\textsuperscript{14}This may be thought of as a perfectly illiquid “off-the-run” security.\textsuperscript{15} As in Vayanos and Weill (2007), this can be interpreted as a borrowing constraint or as risk aversion.\textsuperscript{16} 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.
Figure 1: **Distribution of $\lambda$ against $\sigma$**: This figure shows the distribution of the lenders by the ease with which they can be searched in the repo market. $\lambda$ 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. $\sigma$ denotes the agents “type” as characterized by the ease with which he can be accessed by a borrower. The agents type are ordered in decreasing order of $\lambda$. The plot is of a distribution of $\lambda$ where some agents have relatively high repo market participation while the majority of them participate in the repo market with a relatively very low likelihood.

### 3.3 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 $\lambda : [0, \bar{\sigma}] \rightarrow [\lambda_0, \lambda_\bar{\sigma}]$. In the numerical example which I use throughout this paper, the assumed distribution of $\lambda$ is characterized by figure 1. With the hedging demand to borrow ($D$) set a 300, this corresponds to about 70% probability of encountering a borrower within the first day of acquiring a security for the highest access lender, and 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 is trading. In subsection 3.4 I show that the agents who will become marginal buyers of the security at any point of time are those who find it easiest to lend out their holdings out of the set of all possible holders who have not yet bought the security. 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.

---

17The 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.
For a lender of type $\sigma$, at any time $t$, the arrival process 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 $\sigma$ 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:

$$\dot{\mu}^n_{\sigma}(t) \delta \sigma = -\lambda_{\sigma} \delta \sigma \mu^n_{\sigma}(t) \mu^{bo}(t)$$  \hspace{1cm} (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. Also, because every time a security is lent out, it gets shorted and bought by a new potential lender, we have the following condition:

$$\int_{0}^{\sigma(t)} \mu^n_{\sigma}(t) \delta \sigma = S$$  \hspace{1cm} (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 $\lambda_{\sigma} \mu^n_{\sigma}(t) d\sigma$. This means that the total rate at which borrowers meet lenders of types $\in [0, \sigma(t)]$ is given by:

$$\mu^{bo}(t) = -\mu^{bo}(t) \int_{0}^{\sigma(t)} \lambda_{\sigma} \mu^n_{\sigma}(t) \delta \sigma$$  \hspace{1cm} (4)

The level of short interest at any time $t$ is given by 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:

\[
\sigma(t) = S + \mu_s(t) = S + D - \mu^{lo}(t)
\]  

(5)

Notice that 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, an assumption that I justify in section 3.4. 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 \(\lambda\) 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. Examine 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 \(\sigma\) who have not yet lent out their securities, and at time 0, this fraction is 1 for 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 \(\lambda_\sigma\), 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.

### 3.4 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 till time \(T\). Since a lender always has the option to
Figure 2: **Plot of $\mu_n^s$ against time and $\sigma$** 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.

Figure 3: **Plot of Short interest ($\mu^s_t$) 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.
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, is 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 \( t \) is given by the following differential equation:

\[
-dV^l(t) = w(t)dt
\]  

subject to the boundary condition:

\[
V^l(T) = V(T)
\]

which implies that:

\[
V^l(t) = V(T) + \int_t^T w(s)ds
\]  

Now consider an agent of type \( \sigma \) 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 taking the continuous time limit of the solution.

\[
-dV_\sigma(t) = \lambda_\sigma \mu^{bo}(t)(V^l(t) - V_\sigma(t))
\]

subject to the boundary condition:

\[
V_\sigma(T) = V(T)
\]
The solution to this differential equation is given by:

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

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

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 \(\sigma\) 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 there will be observations of \(w(t)\) occurring because of the fees being paid by (other) borrowers who have already shorted the security. However, not all the observed fees will be incorporated into the valuation, because frictions cause the agent 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 3.1.** The valuation of any agent who has not lent a security at time \(t\) is increasing in his access \(\lambda \sigma\)

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 theorem is in appendix B. The theorem also implies the following:
Corollary 3.2. The price of the security at any time $t$ is $V_{\sigma(t)}(t)$. That is

$$P(t) = V(T) + \int_t^T \lambda_{\sigma(t)} \mu^{b_0}(s) e^{-\int_t^s \lambda_{\sigma(t)} \mu^{b_0}(\tau) d\tau} L(s) ds$$

(11)

where

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

For a buyer to buy the security, his valuation must be not less than 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 $\lambda$ is decreasing in $\sigma$.

Corollary 3.3. 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 $\lambda$ is decreasing in $\sigma$ and the fact that $\sigma$ 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 3.4. 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 $\sigma(0) = S$, and the price is given by $P(0) = V_{\sigma(0)}(0)$

Corollary 3.4 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 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:

\[
-V^s(t) + \dot{P}_t = x(t) - w(t)
\]  

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:

\[ V^s(t) = x(T-t) - \int_t^T w(s) ds + P(t) - V(T) \]  

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

\[
-V^{bo}(t) = \int_0^{\sigma(t)} \lambda_{\sigma} \mu_{\sigma}^{\sigma}(t) d\sigma \left( V^s(t) - V^{bo}(t) \right)
\]  

with the boundary condition:

\[ V^{bo}(T) = 0 \]

which implies that:
\[ V^{bo}(t) = \int_t^T \int_0^{\sigma(s)} \lambda_s \mu^{\alpha}(s) d\sigma e^{-\int_t^s \int_0^{\sigma(\tau)} \lambda_s \mu^{\alpha}(\tau) d\tau} \left[ x(T - s) + P(s) - V(T) - \int_s^T w(\tau) d\tau \right] ds \]

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

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

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

\[ (16) \]

3.5 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 \( t \leq \tau \leq T \). Let us assume that the lender is chosen with probability \( \theta \) to make the first offer. He then offers a schedule that makes the borrowers benefit from lending 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 the highest valuation lender. Thus the expected present value of the lending fee is given by solving the following equation:

\[ \theta \left( V^b(t) - V^{bo}(t) \right) = (1 - \theta) \left( V^l(t) - V^n_0(t) \right) \]  

\[ (17) \]

**Proposition 3.5.** 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 to their reservation value. This is only possible for all values of \( t \), if the gains from lending are zero.
Figure 4: Plot of \( w(t) \) against time: This figure shows a plot of the evolution of the per period lending fee \( w(t) \) 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 per period hedging benefit that agents who short derive is 0.02.

**Proposition 3.6.** 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 shorter gains from 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.

\[
L(t) = \theta \left( P(t) - V(T) + x(T - t) - \int_t^T \frac{\mu^b(t)}{\mu^b(s)} \left[ x(T - s) + P(s) - V(T) - L(s) \right] ds \right) + (1 - \theta) \int_t^T \lambda_0 \mu^b(s)e^{-\int_s^t \lambda_0 \mu^b(\tau)d\tau} L(s)ds
\]

where \( P(t) \) is given by equation 11. Proof in Appendix B.

**Proposition 3.7.** 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
Figure 5: **Plot of $L(t)$ against time**: This figure shows a plot of the evolution of the total future lending fee $L(t)$ 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 their repo market availability that is given in figure 1. The per period hedging benefit that agents who short derive is 0.02.

Figure 6: **Plot of price premium $P(t)$ against time** This figure shows a plot of the price premium (the difference between current price and the value $V(T)$ on the date of termination). 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. The per period hedging benefit that agents who short derive is 0.02.
$P(t) \rightarrow V(T)$ and $L(t) \rightarrow 0$.

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 there are borrowers. Hence the lenders compete the fee away to zero. If, on the other hand, there are absolute frictions, there are no fees to earn, and hence the price is the same as the valuation on the termination date. The only thing that changes between these two situation is the welfare of the borrowers. In the former case, they get to meet all their hedging needs, and end up deriving all the benefits from hedging, while in the latter case, they do not meet any of their hedging needs.

**Proposition 3.8.** 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.

Proof in Appendix B. Intuitively, considering equation 18, and leaving the equations for the measures unchanged, we see that the 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, that if the hedging benefits that short sellers derive are high during a period, price premia are high and at the same time specialness is also high. This allows us to perform the tests in section 4, where in we compare price premia and specialness from different auction cycles in treasuries.

### 3.6 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 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 dis-incentivize 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 future price declines.

### 3.7 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, and show that the dynamics are similar to short-selling being driven by hedging demand, in terms of the measures, with an important implication for the price premium observed. 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. Holders of the security derive a per-period benefit of $x$. The price of the security initially is highest because the hedging benefits are derived for the longest period. There exists a number of arbitrageurs in measure D who
arbitrage against the expected decline in price, by borrowing the security and selling it to other agents wishing to hold the security.

It is easy to see that in such a situation, the evolution of measures of lenders and borrowers will be identical to that described in sections 3.2 and 3.3. Appendix C shows that the valuations of a lender of the security, and the non-lender of the security of type $\sigma$ are given by:

\begin{equation}
V_l(t) = V(T) + x(T-t) + L(t)
\end{equation}

\begin{equation}
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
\end{equation}

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

\begin{equation}
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
\end{equation}

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

\begin{equation}
V^s(t) = P(t) - V(T) - L(t)
\end{equation}

\begin{equation}
V^{bo}(t) = \int_t^T \frac{-\mu^{bo}(s)}{\mu^{bo}(t)} [P(s) - V(T) - L(s)] ds
\end{equation}

Denote by $P^*(t)$ the quantity $P(t) - x(T-t)$, and it is easy to see that the lending fees in this case are the same as a situation where it is the short-sellers who 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 $x(T-t)$. 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 3.5, we can show that $w(t) = -dP(t)$, which is the Duffie(1996) result. Also, $V^s(t) > 0$ implies that $P(t) - V(T) > L(t)$, that is the price premium is greater than the observed future lending fee. The borrowers in this situation are all convergence traders, who make profits.
This has significant empirical implications. Graveline and McBrady (2005) use variables that 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 an 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 mis-pricing 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.

4 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 wherein he finds it impossible to buy or borrow the security and deliver it back to the agent who has lent him the security on the delivery date. However, it is possible to roll-over a repo position by borrowing securities from the lender, at the special repo rate prevailing in the market.

We proceed to make the following argument: A security can get locked away and become unavailable to borrowers, only in a situation where the buyers of the security do not care about fees from lending. Since rationally, we would not expect them to price any lending gains, the price at which they buy will not account for lending fees. Thus securities getting “locked away”, and hence rising specialness will only be consistent with a decreasing fraction of observed future lending fees being priced in.

Data on special repo rates for on-the-run treasury securities is published by GovPx and is available via the Bloomberg data service. I use daily data on special repo rates for 5 year and 10 year on-the-run securities from 11/6/1995 to 6/1/2007. The reason for choosing ten year and five year bonds is that these securities have been issued over the sample period following regular cycles. These special repo rate 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, 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 they happen, I set the special repo rate to the general collateral rate, because the special repo rates can never be more than the general collateral rates.

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 the third-off-the run securities. I use the swap curve against

---

18 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 risk-less profit.
3 month LIBOR 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 are are valued. However, there is no reason to suppose that the term structure of credit risk
underlying the swap benchmark is any 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.

Figure 7 shows the specialness and price premium of the ten year on the run against the 3rd off
the run note evaluated against the swap rate over an average three month auction cycle for the ten
year security in my sample. The price premium on the on-the-run bond as 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 a bond with another bonds that has no specialness, we
choose the premium of a note over the third off-the-run note in our study.

4.1 Test Design

Recall equation 11 which implies that

\[ P(t) = V(T) + \int_t^T \lambda_\sigma(t)\mu^{b_0}(s)e^{-\int_t^s \lambda_\sigma(t)\mu^{b_0}(\tau)d\tau} L(s)ds \]

where

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

Our intention is to test this against the hypothesis that

\[ P(t) = V(T) + L(t) \]
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 auction cycles ago. 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.

Since \( L_t \) is always decreasing in \( t \), at any given point of time, the second term on the right hand side can be expressed as a fraction of \( L_t \). Furthermore, since both \( \mu^{bo} \) and \( \lambda_{\sigma(t)} \) are decreasing with time, we would expect this fraction to decrease. Thus, we can write equation 24, interpreting our measure of the price premium as a “relative” price measure as:

\[
P(t) = E_t(P(T)) + K_t E_t(L_t)
\]

or

\[
E_t(P(t) - P(T)) = K_t E_t(L_t)
\]  

(25)

where \( P(t) \) is the price premium 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 \( E_t(L_t) \). Our hypothesis is that \( K_t \) is less than one and decreasing 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:

\[
s(t) = KL_t + \sum_{i=1}^{f} K_i v_i(t)
\]

where \( v_i(t) |_{i=1}^{f} \) is a set of variables that may be expected to affect the spreads, such as the demand to hold liquid assets. Note that the \textit{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 the left hand side and the right hand side are non-stationary with respect to \( t \). This will lead to spurious inference.

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 \textit{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 to from the beginning the end of the auction cycle 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.

While it is difficult to correct for both these issues when the data-set is limited and has a limited number of auction cycles, with a large number of auction cycles, it is possible to correct
for these issues. I do it in the following way. 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.

\[ K_t \equiv \frac{\sum_{i=1}^{t}(P^i(t) - P^i(T))}{\sum_{i=1}^{t}(L(t))} \]  

(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 \( \Delta \), made by a hypothetical convergence trader following a zero wealth strategy with
a short position in the on-the-run security and a short 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). He shows, using data on twelve auction cycles using the thirty year bond,
that such a strategy makes zero returns on an average. I replicate this approach and validate
my results, as a robustness check. The null hypothesis to be tested with this approach can be
established by writing equation 25 as:

\[ E_t(P(t) - P(t + \Delta)) = K_tE_t(L_t) - K_{t+\Delta}E_t(L_{t+\Delta}) \]

where \( K_{t+\Delta} < K_t < 1 \).

Let \( K_{t+\Delta} = K_t - M_t \), then

\[ E_t(P(t) - P(t + \Delta) - (L_t - L_{t+\Delta})) = (K_t - 1)E_t(L_t - L_{t+\Delta}) + M_tE_tL_{t+\Delta} \]

(27)

Under the null hypothesis of \( K_t = K_{t+\Delta} = 1 \), which is implied by Duffie (1996) and Krishnamurthy
(2002), the left hand side of equation 27 is equal at all times to zero. Under the alternative
hypothesis of my model, it is less than zero, and decreases as t increases.

There are 42 regularly spaced auction cycles for the ten year note in the data that I use. 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\textsuperscript{19}.

4.2 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
are with a quarterly frequency, and four are with a semi-annual frequency. I first split the data
up according to the length of the auction cycles. To ensure that we are dealing with roughly
comparable auction cycles, I 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 other variables that we may expect to affect specialness and spreads, estimate the model for
the entire data-set simultaneously. 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 thus follow a three-step procedure:

1. We first group daily data by weeks since issuance. That is compute the average price premium
   over the off the run treasury during a week, and the total cumulative fees that are paid from
   them middle of the week to the date the next security is issued. This is done so as to reduce

\textsuperscript{19}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.
any problems related to bid-ask bounds and asynchronicity between the price observations, which are as of the end of any day (with settlement on the next day), with the repo rates, which are weighted averages of transactions that take place over the course of a day. 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. Next, we compute, for every auction cycle and week, the change in the price premium that took 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 cycle, this change is positive, and represents the profits made by a convergence trader who did not face any repo specialness.

3. Form sub-samples of the two variables above by days since the last auction. That is, I form subgroups of the series where each group contains only the data-points applicable to a certain number of days past a particular auction date. That is, I compare the the first day’s forecast from the first auction cycle, with the first days forecast from the second auction cycle, and so on. By construction, the sub-samples are stationary. For each value of days since last auction, I compute the average reduction in the price premium between that day and the day the next security was issued, and the average cost of shorting owing to specialness, and compute the ratio of the two. The standard errors for the ratio are computed using the delta method, using the variance covariance matrix of the realizations within each of the sub-samples. That is , I compute the sample estimate of \( K_t \), appealing to equation 26. I 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 1 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
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 $K_t = \frac{E_t\left(P(t) - P(T)\right)}{E_t(L(t))}$. We separately compute $K_t$ for each value of $t$. The confidence band indicates estimated values of $K(t)$ at 95% confidence intervals. The standard errors for these 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 4. 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.

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 decreasing 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 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) 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 2. 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, I am able to reject the null hypothesis that the profit on the convergence trade is zero only for two out of the eleven weeks
<table>
<thead>
<tr>
<th>Week</th>
<th>Convergence profit (%)</th>
<th>Dollar value of fees paid</th>
<th>Ratio (3) = (1)/(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.135</td>
<td>0.199</td>
<td>0.679 (3.48)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.65)*</td>
</tr>
<tr>
<td>2</td>
<td>0.177</td>
<td>0.192</td>
<td>0.922 (4.80)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.41)</td>
</tr>
<tr>
<td>3</td>
<td>0.162</td>
<td>0.179</td>
<td>0.905 (4.47)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.47)</td>
</tr>
<tr>
<td>4</td>
<td>0.136</td>
<td>0.167</td>
<td>0.814 (3.73)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.85)</td>
</tr>
<tr>
<td>5</td>
<td>0.110</td>
<td>0.148</td>
<td>0.745 (3.18)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.09)</td>
</tr>
<tr>
<td>6</td>
<td>0.081</td>
<td>0.129</td>
<td>0.629 (2.64)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.55)</td>
</tr>
<tr>
<td>7</td>
<td>0.057</td>
<td>0.118</td>
<td>0.485 (2.03)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(2.15)**</td>
</tr>
<tr>
<td>8</td>
<td>0.052</td>
<td>0.099</td>
<td>0.521 (1.78)*</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.64)</td>
</tr>
<tr>
<td>9</td>
<td>0.057</td>
<td>0.080</td>
<td>0.710 (2.16)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.88)</td>
</tr>
<tr>
<td>10</td>
<td>0.046</td>
<td>0.063</td>
<td>0.733 (2.13)**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.78)</td>
</tr>
<tr>
<td>11</td>
<td>0.013</td>
<td>0.042</td>
<td>0.322 (0.60)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.27)</td>
</tr>
<tr>
<td>12</td>
<td>0.004</td>
<td>0.024</td>
<td>0.181 (0.21)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.95)</td>
</tr>
</tbody>
</table>

Table 1: **Ratio of the expected convergence and expected future lending fees, by weeks after issuance:**

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 4. 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 auction 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.

over which I compute the profits. Figure 9 graphically depicts the coefficients 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.

### 4.3 Economic drivers of hedging costs

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 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
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 2: **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 auction cycles previously. 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.

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.

In the second stage, I use the residuals from the first stage regression at the monthly level and regress 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 3.

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

<table>
<thead>
<tr>
<th>Monthly residual specialness (bps)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mktrf</td>
<td>-141.1</td>
</tr>
<tr>
<td></td>
<td>(0.95)</td>
</tr>
<tr>
<td>smb</td>
<td>-215.7</td>
</tr>
<tr>
<td></td>
<td>(1.36)</td>
</tr>
<tr>
<td>hml</td>
<td>-645.4</td>
</tr>
<tr>
<td></td>
<td>(2.88)**</td>
</tr>
<tr>
<td>rf</td>
<td>7935.6</td>
</tr>
<tr>
<td></td>
<td>(2.40)*</td>
</tr>
<tr>
<td>umd</td>
<td>-111</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
</tr>
<tr>
<td>Constant</td>
<td>-20</td>
</tr>
<tr>
<td></td>
<td>(2.10)*</td>
</tr>
<tr>
<td>Observations</td>
<td>134</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.10</td>
</tr>
</tbody>
</table>

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

Table 3: **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.
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. The relationship is not statistically significant, though. This results 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 seem to be 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.

If hedging benefits are an important driver of specialness, and hence price premia, and 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 during bad times. 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.

5 Conclusion

I present a model where price differences between similar securities arise endogenously because of short-sales frictions. The model features frictionless cash markets, and repo markets that are
constrained by the need to search for counter-parties. Such price differences are apparently riskless, but cannot be arbitraged away. This is because the fees that an agent has to pay in order to maintain the short position are more than the expected reduction in price over the time convergence between the prices of those securities is expected to take place. 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 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. 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.

I apply the model to convergence trades between on-the-run and off-the-run securities in the treasury market. Previous literature has argued that on an average, such convergence trades make zero profits once shorting fees in the form of repo-specialness is accounted for. I show that such convergence trades make negative profits, on an average.

Furthermore, I am able to match observed patterns in the dynamics of repo specialness within an auction cycle. Observed specialness tends to be higher as more time passes since the security is issued, 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. The reason on-the-run securities become more special as time passes since the security is auctioned because the holders of securities tend to be agents who participate to a lesser degree in the repo market. Because of this agents wishing to sell those securities short in order to gain hedging benefits, find it difficult to do so, and are willing to pay a higher per-period fee. I also show that the proportion of observed lending fees that are priced in to the on-the-run security is lower, which is consistent with the marginal buyers of those 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 repo-specialness, once auction-cycle related dynamics have been addressed, is 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. Since specialness is a form on income for a holder of the on-the-run securities, and is partly priced in, the negative correlation might explain why on-the-run securities have higher prices during times of market distress - rather than arising out of the demand by agents to hold liquid securities, part of it could be driven by the demand by risk averse agents wishing to hedge their portfolios against interest rate risk, and their willingness to pay high shorting fees in order to do so.
Appendix A

Solving for the measures

I denote by the function $F(\sigma, t)$ the measure of agents of type $\sigma$ 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 $\sigma$ has become the marginal buyer at time $t$, and zero otherwise. Then we have the following identities:

$$\int_0^{\bar{\sigma}} Q(\sigma, t) d\sigma = \sigma(t)$$  \hspace{1cm} (28)

$$\int_0^{\bar{\sigma}} \lambda_\sigma F(\sigma, t) Q(\sigma, t) d\sigma = 1$$ \hspace{1cm} (29)

The evolutions of the measures are given by:

$$\dot{\sigma}(t) = (S + D - \sigma(t)) \int_0^{\sigma} \lambda_\sigma F(\sigma, t) Q(\sigma, t) d\sigma$$

$$\dot{F}(\sigma, t) = -\lambda_\sigma F(\sigma, t) Q(\sigma, t)(S + D - \sigma(t))$$ \hspace{1cm} (30)

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

$$\ln \left( \frac{S + D - \sigma(s)}{S + D - \sigma(t)} \right) = \int_t^s \int_0^{\sigma} \lambda_\sigma F(\sigma, \tau) Q(\sigma, \tau) d\sigma d\tau$$ \hspace{1cm} (31)

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

$$\ln \left( \frac{F(\sigma, s)}{F(\sigma, t)} \right) = -\int_t^s \lambda_\sigma Q(\sigma, \tau)(S + D - \sigma(\tau)) d\tau$$ \hspace{1cm} (32)

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 $\sup_{i} \left| \sigma^{i}(\tau) - \sigma^{i+1}(\tau) \right| \to 0$

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 32. Given $F(\sigma, \tau)$ and $Q(\sigma, \tau)$, compute $\sigma^{i+1}(\tau)$ using equation 31. Compute $\sup_{i} \left| \sigma^{i}(\tau) - \sigma^{i+1}(\tau) \right|$. 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 3.1

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

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

$$\Rightarrow -V_{\sigma(1)}(t) < -V_{\sigma(2)}(t)$$

$$\Rightarrow V_{\sigma(1)}(t) > V_{\sigma(2)}(t)$$

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

$$V(\bar{T}) - V_{\sigma(1)}(t) > V(\bar{T}) - V_{\sigma(2)}(t)$$

$$\Rightarrow V_{\sigma(1)}(t) < V_{\sigma(2)}(t)$$

A contradiction.

Proof of proposition 3.5

If borrowers have all the bargaining power, then borrowers set the gain from lending for any lender they encounter to their reservation value. Consider a lender of type $\sigma$ 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

\[
\int_t^T \lambda_\sigma \mu^{\pi_0}(s)e^{-\int_s^t \lambda_\sigma \mu^{\pi_0}(\tau)d\tau} \int_s^T w(\tau)d\tau ds = \int_t^T w(s)ds \tag{33}
\]

For \( \lambda_\sigma < \infty \), the left-hand side of 33 is greater than or equal to the right-hand side, with equality at \( w(t) = 0 \forall 0 \leq t \leq T \). Hence the lending fee is zero if borrowers have all the bargaining power.

**Proof of proposition 3.6**

Setting \( \theta = 1 \) in equation 18 gives us

\[
L(t) = \left( P(t) - V(T) + x(T - t) - \int_t^T \frac{-\mu^{\pi_0}(s)}{\mu^{\pi_0}(t)} [x(T - s) + P(s) - V(T) - L(s)] ds \right)
\]

\[
\Rightarrow
P(t) - V(T) + x(T - t) - L(t) = 0 \\
\forall 0 \leq t \leq T
\]

The result follows from differentiating with respect to \( t \).

**Proof of proposition 3.7**

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 3.6. 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:

\[ E_i^i(P^i(t) - P^i(T)) = K_t E_i^i(L^i(t)) \]

Taking the unconditional expectations of both sides,

\[ E(E_i^i(P^i(t) - P^i(T))) = K_t E(E_i^i(L^i(t))) \]

Now, appealing to the law of iterated expectations,

\[ E(E_i^i(P^i(t) - P^i(T))) \equiv \frac{1}{I} \sum_{i=1}^{I} (P^i(t) - P^i(T)) \]
\[ E(E_i^i(L^i(t))) \equiv \frac{1}{I} \sum_{i=1}^{I} (L(t)) \]
\[ K_t \equiv \frac{\sum_{i=1}^{I} (P^i(t) - P^i(T))}{\sum_{i=1}^{I} (L(t))} \]

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 \( E_i^i(P^i(t) - P^i(T)) \), and its realization \( P^i(t) - P^i(T) \), and \( \epsilon_{L_i}(t) \) as the difference between \( E_i^i(L^i(t)) \) and its realization \( L^i(t) \). Let \( E_i^i(P^i(t) - P^i(T)) \) be distributed with mean \( E(E_i^i(P^i(t) - P^i(T))) \), and variance \( \sigma_p^2 = E\left(E_i^i(P^i(t) - P^i(T))\right) - E(E_i^i(P^i(t) - P^i(T)))^2 \).
Then,

\[ \sigma_p^2 = E \left( E_i^t(P^i(t) - P^i(T))^2 \right) - 2E \left( E_i^t(P^i(t) - P^i(T))E \left( E_i^t(P^i(t) - P^i(T)) \right) \right) + E \left( E_i^t(P^i(t) - P^i(T))^2 \right) = E(\epsilon_i P(t)^2) \]  

(34)

Similarly let \( E_i^t(L^i(t)) \) be distributed with mean \( E(\epsilon_i(L^i(t))) \) and variance \( \sigma_l^2 \). A similar argument shows that

\[ \sigma_l^2 = E(\epsilon_i L(t)^2) \]  

(35)

and the covariance term is

\[ \sigma_{lp} = E(\epsilon_i L(t)\epsilon_i P(t)) \]  

(36)

The quantities on the right hand side of 34, 35 and 36 can be estimated from their sample moments. Appealing to the Delta method, the asymptotic variance of \( K_t \) is given by:

\[ \text{Asy.Var}(K_t) \equiv \frac{1}{\left( \frac{1}{I} \sum_{i=1}^{I}(L(t)) \right)^2} \sigma_l^2 
+ \frac{1}{\left( \frac{1}{I} \sum_{i=1}^{I}(P^i(t) - P^i(T)) \right)^4} \sigma_p^2 
+ 2\frac{1}{\left( \frac{1}{I} \sum_{i=1}^{I}(P^i(t) - P^i(T)) \right)^3} \sigma_{lp} \]
References


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*.


