The Cost of Short-Selling Liquid Securities*

Snehal Banerjee† Jeremy J. Graveline‡

August 2010

Abstract

Since long investors typically pay higher prices for liquid securities, common intuition suggests that they are solely responsible for any liquidity premium. We argue that short-sellers may also pay a liquidity premium if their borrowing costs are higher than what they expect to recoup from future price declines. Market clearing implies that not every long investor can lend out her entire position to collect these higher borrowing rates. Therefore, both long investors and short-sellers can simultaneously contribute to the liquidity premium. We characterize this decomposition of the premium in terms of cash prices, borrowing fees, and the fraction of the outstanding security sold short. We use this decomposition to show that, from November 1995 through July 2009, short-sellers accounted for an average of 50% of the liquidity premium for on-the-run Treasuries.

*We are grateful for comments from Darrell Duffie, Nicolae Garleanu, Arvind Krishnamurthy, Lasse Pedersen, Dimitri Vayanos, and Pierre-Olivier Weill.

†Kellogg School of Management, Northwestern University, snehal-banerjee@kellogg.northwestern.edu

‡Carlson School of Management, University of Minnesota, jeremy@umn.edu

1 Introduction

Liquid securities that are in positive net supply typically trade at higher prices than their less liquid counterparts. Since there must be more long investors than short-sellers, traditional models of liquidity argue that these premiums are borne entirely by long investors who value securities they can easily sell in the future (see Amihud and Mendelson (1986)). Our paper makes a simple point: both long investors *and* short-sellers can *simultaneously* pay a premium for positions in a liquid security.

The Treasury market is an ideal setting for our empirical analysis as it provides us with very similar securities that differ primarily in how liquid they are. The most recently issued, on-the-run securities are extremely liquid. For instance, Barclay et al. (2006) report that daily trading volume in the 2-, 5-, and 10-year on-the-run Treasuries rivals that of all U.S. stocks combined. However, when new securities are issued and the existing bonds move off-the-run, this trading volume drops by 90%. Moreover, as evidenced by the events of the recent crisis, liquidity in the financing markets for these securities plays a critical role in the proper functioning of financial markets as a whole.[1]

It is well-documented that recently issued (on-the-run) Treasuries typically trade at higher prices than less liquid, seasoned (off-the-run) issues with similar coupons and maturities (see Amihud and Mendelson (1991) and Warga (1992)). Consistent with this earlier work, we estimate that the liquidity premium for on-the-run 10-year Treasuries relative to their off-the-run counterparts was \$195 per $\$100,000$ (\$171M per year) over our sample period from November 1995 through July 2009. However, to our knowledge, we are the first paper to document that, on average, short-sellers paid about 50% of this premium. Our results provide empirical evidence that liquid securities command a premium not just because they can be easily sold by long investors, but also because they can be easily borrowed and repurchased by short-sellers.

To illustrate our main point, suppose the Treasury issues one hundred, 10-year notes for $\$100,000$ each and otherwise equivalent, but less liquid, notes each cost $\$99,805$. Furthermore, suppose that these prices are expected to converge next period, so the liquidity premium is $\$100,000 - \$99,805 = \$195$ per note. Controlling for interest rate risk, the expected cost of short-selling the liquid security for the period is the cost of borrowing it, less the \$195 cash premium that a seller expects to recoup when the prices converge. Therefore, as Duffie (1996) and others have argued, a short-seller should be willing to pay at least \$195 to borrow the liquid Treasury. A short-seller may be willing to pay more if she values a position in the liquid security. For example, if she pays \$293 to borrow the liquid Treasury then she expects her cost of short-selling to be $\$293 - \$195 = \$98$.

In aggregate there must be more long positions than short positions, so not every security held long can be loaned to a short-seller. Therefore the net cost to long investors depends not only on the price premium and the lending fee, but also on the fraction of their position they expect to lend out. In our example, suppose there are one hundred total short positions

[1] See Fleming et al. (2010) for analysis of these events and the Federal Reserve's efforts to increase liquidity in the financing, or repo, markets using the Term Securities Lending Facility.

in the liquid Treasury, so that a total of two hundred notes must be held in long positions. Although each long may lend a different amount, half of the aggregate long positions must be loaned to short-sellers. For each outstanding liquid note, long investors as a whole pay the \$195 price premium twice, but recoup \$293 by lending it once to short-sellers. Thus, their net cost is $2 \times \$195 - \$293 = \$97$. In equilibrium, the \$195 liquidity premium is shared almost equally, with short-sellers paying \$98 and longs paying \$97.

Note that the mere existence of a higher price or higher borrowing cost for a liquid security does not reveal whether longs or shorts are ultimately responsible for the liquidity premium. For instance, it would be misleading to conclude that the short-sellers do not pay for liquidity based on the observation that long investors pay \$195 more for the liquid security in the cash market. In the previous example, if long investors collect \$390 for each security that they lend, then the liquidity premium is ultimately borne by short-sellers. Conversely, a positive borrowing fee in the financing market does not imply that short-sellers absorb any of the liquidity premium. If the borrowing fee is \$195 then short-selling is costless and longs are responsible for the entire premium. The cash and financing premiums must be analyzed together with the volume of short-selling. We use this approach to provide the first empirical estimates of the portion of the liquidity premium for on-the-run Treasuries that is paid for by longs and shorts.

Duffie (1996) was the first paper to demonstrate a relationship between the price premium for on-the-run Treasuries in the cash market and the premium to borrow them in the financing, or repurchase (repo), market. As the previous example illustrates, if long investors pay a higher price for the more liquid Treasuries, then short-sellers should also pay a higher borrowing fee since they expect to recoup the price premium when they close out their position. Conversely, if short-sellers pay a premium to borrow on-the-run Treasuries, then the cash price should be higher since long investors can expect to benefit from lending their securities to short-sellers. Jordan and Jordan (1997), Krishnamurthy (2002), and Goldreich et al. (2005), among others, provide empirical support for the relationship that higher prices and higher borrowing fees go hand in hand. The subsequent empirical literature has tried to disentangle whether the observed premiums in the cash and repo markets are driven by demand from long investors or demand from short-sellers. Krishnamurthy (2002) finds that the price premium in the cash market fluctuates with empirical proxies for investors' demand to own liquid securities. On the other hand, Graveline and McBrady (2008) find that the cost of borrowing on-the-run Treasuries fluctuates with proxies for short-selling demand from financial intermediaries who wish to hedge their interest rate exposure.

Our empirical analysis provides a novel approach to this question. We calculate the historical liquidity costs of trading on-the-run Treasuries (relative to their duration-matched off-the-run counterparts) for short-sellers who borrow and sell these bonds and for long investors who buy and lend (finance) a portion of these bonds in the repo market. With this integrated analysis of the cash and repo markets, we find that the average liquidity costs to long investors and short-sellers are roughly equal. That is, on average, the longs and shorts each account for approximately half of the observed on-the-run liquidity premium.

We also document substantial variation in the expected net cost of shorting on-the-run Treasuries. Using predictive regressions, we show that the expected cost of shorting is positively related to primary dealer transactions in Treasuries with similar maturities, which suggests that investors are willing to pay more for short positions when they anticipate having to trade more frequently. We also find a positive relation between the cost of shorting and the CP - TBill spread, which suggests that the expected cost of shorting is higher during financial crises.

Our empirical analysis is most closely related to Krishnamurthy (2002). Although he does not focus on the cost of short-selling, his empirical analysis implies that from June 1995 to November 1999, short-sellers of the on-the-run 30-year Treasury bond did not pay a liquidity premium relative to the next most recently issued, or first off-the-run, 30-year Treasury bond. We find that short-sellers account for roughly half of the observed liquidity premium for on-the-run 10-year Treasury notes (relative to the second off-the-run). Our results differ for a number of reasons. Since the Treasury did not issue 30-year bonds between August 2001 and February 2006, we instead focus attention on the 10-year note and use a data series from November 1995 to July 2009 that is three times as long. We also calculate the liquidity cost relative to the second most recently issued, or second off-the-run, Treasury because the first off-the-run is still frequently used for short-selling and is often still expensive to borrow in the financing, or repo, market. Finally, as Duffie (1996) and Krishnamurthy (2002) argue and as our earlier example illustrates, if long investors are responsible for the entire liquidity premium then the cash and repo market premiums should be equal on average and should tend to rise and fall together. We statistically reject this hypothesis in our longer sample.

Barclay et al. (2006) show that, on a typical day, 150% of the outstanding on-the-run 10-year Treasuries are borrowed and this amount declines significantly once there are two newer issues with the same initial maturity. Previous papers acknowledge that on-the-run Treasuries are appealing securities for short-sellers because they can be easily borrowed and sold when initiating a short position and, perhaps more importantly, they can be easily purchased when closing one out. However, while short-sellers may prefer liquid securities, earlier theoretical models (e.g., Duffie (1996) and Krishnamurthy (2002)) generally imply that they do not pay a premium for positions in these securities. Our empirical results are consistent with recent search-based models that focus on the over-the-counter nature of the financing market and allow for liquidity premiums to be shared by both longs and shorts (e.g., Duffie et al. (2002) and Vayanos and Weill (2008)).

We develop a simple theoretical framework to formally describe how longs and shorts can simultaneously pay for the liquidity premium and to highlight the key assumptions of earlier models that lead to their specific predictions. In order to describe the equilibrium, a model must identify the premiums in the cash and repo markets, as well as the fraction of long positions that are loaned to short-sellers. As Duffie (1996) and Krishnamurthy (2002) have emphasized, the market clearing conditions in the cash and financing markets are not sufficient to pin down these three variables in equilibrium. We then characterize the additional conditions that can be used to pin down an equilibrium, and argue that generically, these lead to equilibria in which both longs and shorts pay for the liquidity premium. We also highlight the specific assumptions in earlier papers that are used to complete the equilibrium. For instance, Duffie (1996) and Krishnamurthy (2002) assume that while some long investors face lending constraints, others are unconstrained and can always lend their entire long position to short-sellers. This assumption implies that in equilibrium, short-sellers do not pay a liquidity premium, since otherwise the unconstrained long investors would profit from the opposite side of the trade. In Vayanos and Weill (2008), longs and shorts search for each other and bargain over the gains to trade when they meet. The search costs determine the equilibrium fraction of long positions that are loaned to short-sellers. Moreover if short-sellers value the liquid security and do not have all of the bargaining power, they pay a borrowing fee that is larger than the expected cash premium, and hence pay for part of the liquidity premium.

The remainder of this paper is organized as follows. In the next section, we describe the empirical framework, market mechanics, and data. In Section 3, we present the main empirical results of the paper. Section 3.1 presents the summary statistics for the trading strategies which we use to estimate the liquidity costs of shorting. Section 3.2 provides our estimates for the total liquidity premium and the fraction of the premium paid by the short sellers. Section 3.3 presents a discussion of the time variation in the cost of shorting. In Section 4 we present a basic theoretical model that formalizes our empirical framework and describes how cash and repo premiums for on-the-run Treasuries can arise in equilibrium and how the total liquidity premium is shared by longs and shorts. Finally, Section 5 concludes.

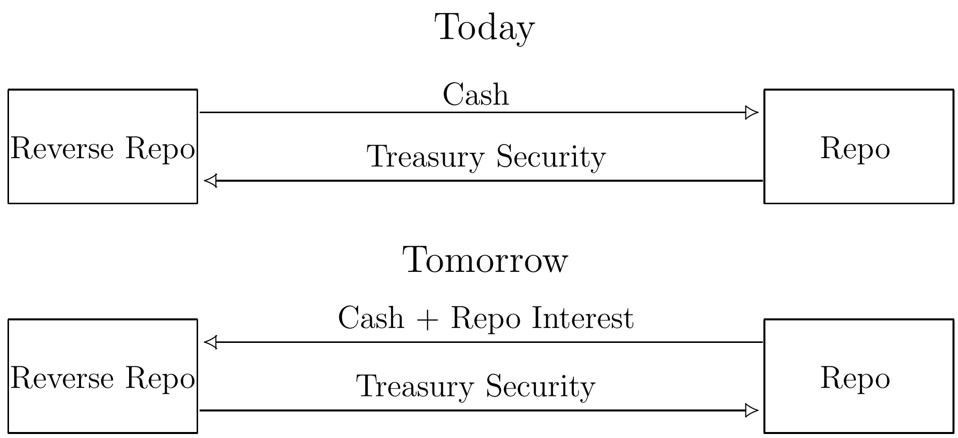
2 Market Mechanics and Data Description

2.1 Empirical Framework

In this section, we present a general framework to pin down ideas for the empirical analysis in the rest of the paper. For each issued on-the-run Treasury, denote the fraction (possibly greater than one) that is borrowed and sold short by δ so that the aggregate long position is $1+\delta$. Let C be the cash premium paid by long investors in the form of a higher cash price relative to an otherwise equivalent and less liquid off-the-run Treasury. Similarly, denote the financing premium paid by short-sellers in the form of higher borrowing costs by R. The average liquidity premium paid by long investors in the cash market is $(1+\delta)\,C$. However, they expect to earn a premium R on the fraction δ that they lend out, and so the net liquidity premium paid by longs is $(1+\delta)\,C-\delta R=C+\delta\,(C-R)$. Similarly, the shorts pay the financing premium R on the fraction δ that they borrow but expect to receive the cash premium C when they sell it. Thus, the net liquidity premium paid by shorts is $\delta\,(R-C)$. Note that for each unit of the on-the-run, the total liquidity premium paid by long investors and short-sellers is $C+\delta\,(C-R)+\delta\,(R-C)=C$, which is the amount that is initially collected by the Treasury when it issues the on-the-run security. The theoretical models in Duffie (1996) and Krishnamurthy (2002) require that $R=C$ so that the entire liquidity premium is paid for by long investors and none of it is recouped from short-sellers. In Section 4 we present a basic theoretical model that formalizes this empirical framework and illustrates how C, R, and δ can be determined in equilibrium.

Our empirical analysis provides estimates of R and C and we find that there is a liquidity premium for short-selling on-the-run Treasuries (i.e., $R-C>0$). We then use estimates for

Figure 1: An Overnight Repurchase Agreement



the fraction δ that is sold short to estimate what fraction of the total liquidity premium is paid for by short-sellers (i.e., what the value of $\delta\frac{R-C}{C}$ is). Finally, we we plot and examine the time-series variation in the cost of short-selling.

2.2 Market Mechanics and Trading Strategies

Before proceeding, we provide a brief review of the financing, or repurchase, market for Treasuries. A more extensive discussion can be found in Duffie (1996) and Fisher (2002). A repurchase agreement, or repo, is a contract for borrowing and lending fixed income securities. At inception, the owner sells the security and simultaneously agrees to repurchase it at a later date for an agreed upon price. The owner of the security essentially borrows money at its repo interest rate and pledges the security as collateral. The counter-party who lends money and receives the security as collateral is said to have entered into a reverse repo. The repo rate for the transaction is the difference between the sale price and the repurchase price expressed as an interest rate. Figure 1 illustrates the details of a repo and reverse repo agreements.

Repo agreements are typically considered safe investments because the investor receives a marketable security as collateral. For this reason, repo agreements are often the cheapest source of financing for long positions. Most Treasuries serve equally well as collateral in reverse repos and therefore they have the same repo rate which is referred to as the general collateral (or GC) rate. Reverse repos are also used by short-sellers who need to borrow securities that they have sold short. When there is a lot of demand to borrow a specific Treasury for short-selling, the repo rate for that Treasury may fall below the general collateral rate in order to equilibrate supply and demand. A security whose repo rate is below the general collateral rate is said to be "on special." It is more costly to borrow a Treasury that is "on special" because you receive a lower interest rate on the money that you lend against it. Conversely, the owner of a Treasury that is "on special" can benefit by using the security as collateral to borrow money at below-market interest rates.

We cannot directly observe the cash premium C that is paid by long investors since equivalent Treasury securities with the same maturity and coupon payments that differ only in liquidity do not trade simultaneously. Instead, as an empirical estimate of this cash premium we compare the ex-post cash return for an on-the-run Treasury relative to the return for the second most recently issued, or second off-the-run, Treasury with same initial time to maturity. We adjust our position in the second off-the-run to account for the fact that these two securities have different duration, or price sensitivity to changes in interest rates. That is, we compare the return from time $t - \Delta t$ to t of investing \$1 in the on-the-run with the return from investing \$ $\left(\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}/\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}\right)$ in the second off-the-run, where $\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}$ and $\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}$ are the duration of the on-the-run and second off-the-run securities respectively. The ex-post difference in these returns, expressed as a rate of return, is given by

$$C_{\mathrm{on},t} = \left[\frac{\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}}{\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}}\left(\frac{P^{\mathrm{off2}}_{t}}{P^{\mathrm{off2}}_{t-\Delta t}} - 1\right) - \left(\frac{P^{\mathrm{on}}_{t}}{P^{\mathrm{on}}_{t-\Delta t}} - 1\right)\right] / \Delta t, \tag{1}$$

where $P^{\mathrm{on}}_{t}/P^{\mathrm{on}}_{t-\Delta t}$ and $P^{\mathrm{off2}}_{t}/P^{\mathrm{off2}}_{t-\Delta t}$ are the returns (including coupons and accrued interest) to the on-the-run and second off-the-run respectively. $C_{\mathrm{on},t}$ can be viewed as the rate of return, ignoring financing costs, from selling \$1 of the on-the-run and buying \$ $\left(\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}/\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}\right)$ of the off-the-run. The on- and second off-the-run Treasuries have similar future payoffs, but the on-the-run is typically priced higher. Therefore, we expect that $C_{\mathrm{on},t}$ will be positive on average.

Similarly, we estimate the financing premium R as the difference in financing costs for this strategy, accounting for the difference in the duration of the two bonds,

$$R_{\mathrm{on},t} = \frac{\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}}{\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}} r^{\mathrm{gc}}_{t} - r^{\mathrm{on}}_{t}, \tag{2}$$

where r^{on}_{t} is the repo rate for the on-the-run and we have assumed that the second off-the-run is loaned at the general collateral repo rate r^{gc}_{t}. On-the-run Treasuries are frequently on special in the repo market and the durations are usually close so we expect that $R_{\mathrm{on},t}$ will be positive on average. When we combine the cash and financing premium, the total liquidity premium for shorting the on-the-run is given by

$$R_{\mathrm{on},t} - C_{\mathrm{on},t} = \left(\frac{P^{\mathrm{on}}_{t}}{P^{\mathrm{on}}_{t-\Delta t}} - 1\right) / \Delta t - r^{\mathrm{on}}_{t} - \frac{\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}}{\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}}\left[\left(\frac{P^{\mathrm{off2}}_{t}}{P^{\mathrm{off2}}_{t-\Delta t}} - 1\right) / \Delta t - r^{\mathrm{gc}}_{t}\right], \tag{3}$$

Equation (3) can be viewed as the cost (or negative profit) of borrowing and short-selling \$1 of the on-the-run and hedging the interest rate exposure with a long position in \$ $\left(\frac{\mathrm{DUR}^{\mathrm{on}}_{t-\Delta t}}{\mathrm{DUR}^{\mathrm{off2}}_{t-\Delta t}}\right)$ of the second off-the-run financed at the general collateral repo rate.

Note that the repo rates in equation (3), apply from time t to $t + \Delta t$, while the cash returns are calculated from time $t - \Delta t$ to t. This difference is due to the fact that the cash market for Treasuries is typically next-day settlement, while the repo market is same-day

Figure 2: Trading Time-line

	Date $t-\Delta t$		Date t		Date $t+\Delta t$
On-The-Run:	Short Sell \$1	$\Longrightarrow$	Buy Deliver Sold Borrow	$\Longrightarrow$	Receive Bought Deliver Borrowed
Off-The-Run:	Buy $\$\frac{DUR^{on}_{t-\Delta t}}{DUR^{off}_{t-\Delta t}}$	$\Longrightarrow$	Sell Receive Bought Lend	$\Longrightarrow$	Deliver Sold Receive Lent

settlement. Figure 2 shows a time-line that corresponds to the above trading strategy and accounts for the settlement differences between the cash and repo markets.

We construct the weekly counterparts to equations (1) and (2) by initiating the above trading strategy each Wednesday and financing the daily profits or losses at the Federal Funds rate. Our empirical estimates of the cash and financing premiums, C_{on} and R_{on}, are computed as the average of these weekly counterparts.

Krishnamurthy (2002) estimates the profits on a similar trade which short-sells the on-the-run 30-year Treasury bond and hedges the interest rate exposure with the first off-the-run (the next most recently issued) 30-year bond. We use the 5- and 10-year Treasury notes for our empirical analysis because the Treasury stopped issuing 30-year bonds between August 2001 and February 2006, which is a large part of our sample. Also, we measure the cash and financing premiums relative to the second off-the-run rather than the more recently issued first off-the-run. Barclay et al. (2006) show that Treasuries still remain very liquid while they are the first off-the-run. Therefore, using the first off-the-run could lead us to underestimate the liquidity premium for the on-the-run Treasuries. Perhaps more importantly, the first off-the-run frequently trades on special in the repo market and, although all short-sellers must borrow a security at its repo rate, not all long positions can be financed at a special repo rate.[2] Since we do not have data on the daily volume of short-sales, we cannot compute the fraction of aggregate long positions that can be financed (lent) at a repo special and the fraction that must be financed at general collateral. The second off-the-runs do not often trade on special in the repo markets so it is more accurate to assume that the entire position

[2]Treasury securities are in positive net supply and the cash and financing markets have finite clearing capacity. In practical terms, this means that there are always more long positions than short positions. If we assume that all Treasuries serve equally well as investment collateral then only short-sellers will pay a financing premium and therefore not every Treasury can be financed at a special repo rate. Of course, some longs may be able to finance their entire position at a special repo rate. However, this will not be the case for the average long investor. In the context of our empirical framework, there are a finite number of short positions so that $\delta < \infty$. Therefore, a maximum fraction $\delta/(1+\delta) < 1$ of the average long position can be financed at a special repo rate and the remainder can only be financed at the general collateral rate.

is financed at the general collateral repo rate.

Our empirical analysis uses the trading strategy above which we denote by "On Less Off2." As robustness checks, we also calculate the cash and financing premiums for the following alternate strategies:

- On Less Off1: The cost of borrowing and short-selling the on-the-run and hedging the interest rate exposure with the first off-the-run which is financed at its repo rate. As mentioned above, this strategy may not be implementable in practice because the first off-the-run often has a special repo rate but not every long position can be financed at this rate.
- Off1 Less Off2: The cost of borrowing and short-selling the first off-the-run and hedging the interest rate exposure with the second off-the-run which is financed at the general collateral repo rate. We define $C_{\text{off1},t}$, $R_{\text{off1},t}$ and $R_{\text{off1},t} - C_{\text{off1},t}$ analogously to equations (1), (2) and (3) respectively, by replacing the on-the-run Treasury with the first off-the-run.

2.3 Data Description

Our sample spans over 13 years from November 1995 through July 2009. We use closing prices on 5- and 10-year Treasury notes from Bloomberg which takes the midpoint of the bid and ask quotes from a sample of dealers. Bloomberg provides clean prices which we augment with accrued interest and coupon payments in order to calculate returns. We use overnight repo rates for on-the-run and first off-the-run Treasuries from ICAP GovPX. GovPX also provides overnight general collateral rates for repurchase transactions in which any Treasury security can be provided as collateral. GovPX provides repo rates for a range of times each morning. We use the earliest quote for each security each day. The empirical results are quantitatively similar if we use the trade-weighted repo rates.

3 Discussion of the Results

3.1 Summary statistics of trading strategies

Tables 1 and 2 in Appendix A present summary statistics for the 5- and 10-year maturity trading strategies respectively. The mean yield-to-maturity of the on-the-run 10-year is about 1 basis point lower than that of the second off-the-run. On average, the repo rate for the on-the-run 10-year is about 110 basis points lower than the general collateral repo rate. The weekly cost of shorting the on-the-run, as measured by the return on the "On Less Off2" strategy, is over 30 basis points (annualized return). For comparison, the cost of short-selling the on-the-run relative to the first off-the-run is 7 basis points. Recall that the "On Less Off1" strategy is not always implementable since it involves lending one's entire long position in the first off-the-run at a special repo rate.

Table 2 suggests that these results also extend to the 5-year maturity. The average yield-to-maturity of the on-the-run 5-year is about 2 bps higher than the second off-the-run and the average repo rate is about 75 basis points lower than the general collateral repo rate, but note that these are not adjusted for differing duration. The cost of short-selling the on-the-run and hedging with the second off-the-run (the "On Less Off2" strategy) is 27 basis points (annualized return), which is lower than the same strategy for the 10-year.

3.2 The liquidity cost of short selling

Using the empirical framework that we developed in Section 2.1, Table 3 provides estimates of the total annual liquidity premium for on-the-run 5- and 10-year Treasuries and the fraction of this amount that is paid by short-sellers. In order to estimate the fraction of these liquidity premiums that is paid for by short-sellers, we need to estimate the fraction δ of each security that is borrowed and sold short. A working paper version of Barclay et al. (2006) provides a plot of daily repo volume for on-the-run and first off-the-run Treasuries. From that plot, we estimate that roughly 150% of the outstanding on-the-run 10-year notes are lent into repo agreements, as are around 100% of the outstanding on-the-run 5-year notes. When the notes become the first off-the-run, the fraction lent into repo agreements for both maturities are roughly 75%. In Table 3, we calculate the fraction of the total liquidity that is paid for by short sellers as

$$\text{short fraction} = \delta_{\text{on}} \times \frac{R_{\text{on}} - C_{\text{on}}}{C_{\text{on}} + C_{\text{off1}}} + \delta_{\text{off1}} \times \frac{R_{\text{off1}} - C_{\text{off1}}}{C_{\text{on}} + C_{\text{off1}}},$$

where δ_{on} and δ_{off1} are the fraction of each on-the-run and first off-the-run that are loaned, C_{on} and C_{off1} are the cash market returns as computed in equation (1) for the on-the-run and first off-the-run (both relative to the second off-the-run), and R_{on} and R_{off1} are the corresponding repo market costs as calculated in equation (2).

We estimate that the cash premium for on-the-run 10-year Treasuries relative to the second off-the-run is 94 bps, and the cash premium for the first off-the-run relative to the second off-the-run is 29 bps. The average issue size for 10-year Treasuries over our sample is \$13.9 billion, and each issue is initially the on-the-run and then the first off-the-run. This implies that on average, there is \$13.9 billion of Treasuries that earn a liquidity premium (relative to the second off-the-run) at any point in time. Therefore, we estimate that the average annual total liquidity premium for 10-year Treasuries is

$$\$13.9\text{B} \times \left\{ \left[(1 + 0.0094/52)^{52} - 1 \right] + \left[(1 + 0.0029/52)^{52} - 1 \right] \right\} \approx \$171\text{M}. \tag{4}$$

Over our sample period there are an average of 6.3 auctions per year in the 10-year note, so our estimate translates to a liquidity premium of \$195 per \$100,000 of the 10-year note, where

$$\left\{ \left[(1 + 0.0094/52)^{52} - 1 \right] + \left[(1 + 0.0029/52)^{52} - 1 \right] \right\} \times \$100,000/6.3 \approx \$195. \tag{5}$$

Similarly, the cash premium for the on-the-run 5-year is 76 bps and is 38 bps for the first off-the-run. The average issue size for 5-year Treasuries over our sample is \$17.4 billion so we estimate that the average annual total liquidity premium 5-year Treasuries is

$$\$17.4B \times \left\{ \left[(1+0.0076/52)^{52} - 1 \right] + \left[(1+0.0038/52)^{52} - 1 \right] \right\} \approx \$200\text{M}. \tag{6}$$

There are an average of 9.16 auctions per year, so the estimated liquidity premium is \$125 per \$100,000 5-year note.

We find that for the 10-year maturity, short-sellers pay an average of around 55% of liquidity premium, while for the 5-year maturity, they pay around 28%. Given the average annual liquidity premium, these percentages amount to about \$94 million per year for the 10-year and \$56 million for the 5-year. Since we do not directly observe repo volume, Table 3 also contains estimates with higher and lower values for the fraction δ that are borrowed and sold short. We find that even with conservative estimates of δ, short-sellers pay nearly 30% of the liquidity premium in the 10-year note and over 20% of the premium in the 5-year note.

One must keep in mind that these are estimates of the average cost to being long across all investors. However, while each short-seller must borrow the security in the repo market, there is likely significant variation across investors in the fraction of long positions in on-the-run securities that are loaned (financed) in the repo market. On one extreme, buy-and-hold investors, like foreign central banks and insurance firms, often forgo the specials they can earn by lending out the bonds and so recover almost none of the cash liquidity premium they pay for being long in the on-the-run Treasuries. At the other extreme, highly active institutions like hedge funds and dealers are likely to lend out most, if not all, of their long positions on repo and hence fully recover the premium they pay for being in the on-the-run.

We do not observe individual positions in either the cash or repo market, so unfortunately there is little more we can say about the composition of long investors who ultimately pay for the liquidity premiums that we observe. However, the Federal Reserve Bank of New York does collect data on the aggregate positions of the primary dealers in U.S. government securities with maturities ranging from 3 to 6 years and from 6 to 11 years (although it does not provide information on the positions of individual banks in individual securities). Figure 3 plots this data. One might expect primary dealers to be active in the repo market and lend out most of their long positions. Dealers may indeed be willing to lend out their long positions, but the Federal Reserve data indicates that, in aggregate, they are almost always net short Treasuries (the Russian default crisis of 1998 and after the sub-prime crisis in 2009 are the only exceptions in our sample). Thus, rather than lending long positions in the repo market, they are in aggregate borrowing and paying repo specials.

3.3 The time variation in the cost of shorting

If long investors bear the entire liquidity premium, then the cash and repo premiums should be equal and the return to the REPO less CASH strategy should be zero on average. In Section 3.2 we showed that the average return on this strategy is positive in our sample. As an additional statistical test, we regress the cash component of the trade on the repo component. If long investors pay for the entire liquidity premium, then the regression coefficient should be 1. Table 4 provides the results of this regression (both with and without a constant). For both the on-the-run 5- and 10-year maturities, the regression coefficients are positive, 0.67 and 0.49 respectively, but for both maturities we can reject the null hypothesis that the regression coefficients are 1 at the 5% level.

Having established that short-selling is costly on average, in this section we investigate whether the time-series variation is predictable. We focus on the per unit cost because the total aggregate volume of short sales is not observable. Our empirical analysis here is guided by two types of demand for the liquid security by short-sellers. First, market participants with frequent trading needs often prefer the agility afforded by liquid securities. For example, a dealer or intermediary may purchase a bond from their customer and expect to hold it in inventory for a short period until they can sell it. While the bond is in their inventory, the intermediary often short-sells an on-the-run Treasury with a similar maturity to hedge their temporary interest rate exposure. We label this type of demand as "transactional liquidity" and expect that its effects are more maturity-specific. Second, during times of financial crisis or higher aggregate uncertainty, agents often exhibit a "flight to liquidity" preference because they are uncertain about when they will need to close out their positions and what the market conditions will be at that time. We expect this "flight to liquidity" demand to have a similar effect on liquid Treasuries across all maturities.

As a proxy for maturity-specific "transactional liquidity" we use weekly data from the Federal Reserve Bank of New York on primary dealer transactions in U.S. Government securities. For the 5-year on-the-run Treasuries we use transactions in government securities with maturities ranging from 3 to 6 years and for the 10-year on-the-run Treasuries we use maturities ranging from 6 to 11 years. To measure "flight to liquidity" demand we follow Krishnamurthy (2002) and use the yield spread between 3 month Commercial paper and Treasury Bills (CP - TBill spread). We also focus special attention on the three main crises to affect fixed income markets during our sample period: the Asian crisis of 1998, the Russian default crisis in 1999, and the recent sub-prime crisis starting in 2007.

Figure 4 plots the returns on the REPO less CASH strategies for the 5- and 10-year maturities during our sample period, the 3-month CP - TBill spread, and the weekly primary dealer transactions in the 5- and 10-year bonds during this period. The three main crisis periods are indicated by red dotted lines. Since the primary dealer transaction data is only available from the beginning of 1997, we restrict the sample to this shorter period for the following analysis. While there is a lot of noise in the REPO less CASH return series at the weekly frequencies, the monthly returns series reveals systematic time series variation in the cost of shorting. Visually, the cost of shorting appears higher around periods of crises (delineated with vertical red dashed lines in the plots), which are also associated with higher CP - TBill spreads. However, there is substantial variation in the cost of shorting when there are no financial crises, which suggests that the demand for these securities is not driven solely by a "flight to quality" effect. Also, note that there is substantial time series variation in primary dealer transactions, and that the transactions at the two maturities are highly correlated (with a correlation of 72% for the full sample).

While the time-series plots are suggestive, we use predictive regressions to formally test whether the expected variation in the cost of shorting can be explained by our proxies for liquidity demand and report the results in Table 5. We regress the monthly return from the REPO less CASH strategy on (lagged) primary dealer transactions and the CP - TBill spread. We report the results from estimation for the full sample (including the sub-prime crisis and the period following it) and using the sample pre-2007, and for both the 5- and 10-year maturities.

Consistent with our interpretations, maturity specific dealer transactions and CP - TBill spread have incremental explanatory power and are positively related to the cost of shorting. Moreover, only maturity specific transactions are relevant — dealer transactions in the other maturity are not significantly related to the return on REPO less CASH and tend to decrease the adjusted R^2 when included in the regression. Finally, even though the adjusted R^2's are lower in the full sample (which includes the recent sub-prime crisis and hence is more noisy), CP - TBill spread and maturity specific transactions are still positively related to the REPO less CASH return for both maturities.

As a robustness check to the predictive regressions in Table 5, we report results from contemporaneous regressions with the same variables in Table 6. We regress the return on the REPO less CASH strategy on contemporaneous surprises in the CP - TBill spread and dealer transactions, where surprises are based on an AR(1) specification for each variable. By regressing the returns on contemporaneous surprises, we hope to reduce the effect of the noise in REPO less CASH. We find that the adjusted R^2's in the contemporaneous regressions are higher than the predictive regressions for many specifications. Moreover, maturity specific dealer transactions are still positively related, and often more statistically significant, even after controlling for transactions in the other maturity. We also find that the CP - TBill spread loses some of its statistical significance and has a negative coefficient for some specifications. One potential explanation for this drop in significance is that the level of CP - TBill spread is a noisy proxy for crisis episodes. In this case, changes (or surprises) in the CP - TBill spread need not explain variation in the cost of shorting.

To further explore how the cost of short selling is related to the CP - TBill spread, Table 7 introduces an indicator variable for whether the current period is in one of the crises during these periods. We date the Asian crisis of 1997 as occurring from July 1997 through December 1997, the Russian default crisis as occurring from August 1998 through January 1999, and the sub-prime crisis as occurring from August 2007 through January 2009. Introducing the indicator variable for the crises has additional explanatory power for all the specifications. Moreover, the coefficient on CP - TBill spread decreases in magnitude and statistical significance, which suggests that part of the positive relation between REPO less CASH and CP - TBill is driven by the fact that the CP - TBill spread is large during crises when "flight to liquidity" demand for the on-the-run security is high.

4 Theory

In this section we present a basic model that describes more formally how a liquidity premium can arise in equilibrium due to a demand for liquid securities that is driven by hedging motives, and is shared by long investors and short-sellers. Vayanos and Weill (2008) provide a much more detailed and complete search-based model of the cash and financing markets for on-the-run Treasuries. Our goal here is to provide a simple model that clearly illustrates how the liquidity premium is shared by long investors and short-sellers.

Suppose there is a risky asset with a payoff V in the next period, where

$$E[V] = m_V \text{ and } \mathrm{var}[V] = \sigma_V^2. \tag{7}$$

There are two types of investors indexed by $i = \{L, S\}$ with equal population weights, and an investor of type i has an initial wealth of W_0 and mean-variance preferences over next period's wealth. Finally, denote the quantity of the asset outstanding by Q.

We consider three cases to illustrate how a premium may arise in the cash and borrowing markets in the presence of liquidity demands. Case 0, which serves as the benchmark, considers the economy when there is no demand for liquidity and hence no premium. Case 1 characterizes the economy in which investors face endowment shocks which lead to a liquidity demand for the risky asset. However, in this case, we do not allow for a premium in the repo market in order to isolate the effect of liquidity demand on the cash premium. Finally, Case 2 considers the economy in which investors have a demand for liquidity and there can be a repo premium. This situation allows us to study how the cash and repo premiums depend on each other and on the preference parameters of the investors. The equilibrium price in Case n is denoted by P_n and the equilibrium quantity held by an investor of type i is denoted by $x_{i,n}$.

Case 0: No Liquidity shocks (benchmark case). To begin, consider the case when there is no liquidity, or hedging, demand for the asset. Investor i's optimal portfolio problem is given by

$$x_{i,0} = \arg\max_{x_i} E[W_i] - \tfrac{1}{2\tau_i}\mathrm{var}[W_i], \text{ where } W_i = x_i(V - P) + W_0, \tag{8}$$

where τ_i is a measure of investor i's risk tolerance. The first order condition implies that the optimal portfolio allocation is given by

$$x_{i,0} = \tfrac{\tau_i}{\sigma_V^2}(m_V - P_0). \tag{9}$$

The cash market clearing condition is given by

$$x_{L,0} + x_{S,0} = Q, \tag{10}$$

which implies that the cash price of the risky asset is given by

$$P_0 = m_V - \tfrac{\sigma_V^2}{\tau_S + \tau_L} Q. \tag{11}$$

The discount in price from the expected value m_V of the asset reflects the risk-premium associated with the uncertainty in the payoff. Since there are no hedging demands, there is no demand for shorting the asset. The equilibrium portfolio allocations depend on the risk-tolerance of each group of investors, and are given by

$$x_{L,0} = \tfrac{\tau_L}{\tau_S + \tau_L} Q \quad \text{and} \quad x_{S,0} = \tfrac{\tau_S}{\tau_S + \tau_L} Q. \tag{12}$$

Case 1: Liquidity shocks without differential pricing. Now suppose that investor L receives an endowment shock $\rho_L V$ in the next period and investor S receives an endowment shock $\rho_S V$, where $\rho_L < 0 < \rho_S$. Investors can use positions in the risky security to hedge these endowment shocks. However, the repo premium R is constrained to be zero so that the security can be sold short without a borrowing cost. This case illustrates the effect of a demand for liquidity on the price of the asset. Investor i's optimal portfolio problem is now given by

$$x_{i,1} = \arg\max_{x_i} E\left[W_i\right] - \tfrac{1}{2\tau_i}\mathrm{var}\left[W_i\right], \text{ where } W_i = x_i\left(V - P\right) + \rho_i V + W_0. \tag{13}$$

The first order condition for this optimization problem implies that the optimal portfolio allocation is given by

$$x_{i,1} = \tfrac{\tau_i}{\sigma_V^2}\left(m_V - P_1\right) - \rho_i. \tag{14}$$

The cash market clearing condition is given by

$$x_{L,1} + x_{S,1} = Q, \tag{15}$$

which implies that the cash price of the risky asset is given by

$$P_1 = m_V - \tfrac{\sigma_V^2}{\tau_S+\tau_L}\left(Q + \rho_L + \rho_S\right) = P_0 - \underbrace{\tfrac{\sigma_V^2}{\tau_S+\tau_L}(\rho_L + \rho_S)}_{\text{liquidity premium}} \tag{16}$$

and the equilibrium portfolio allocations are given by

$$x_{L,1} = x_{L,0} + \frac{\tau_L\rho_S - \tau_S\rho_L}{\tau_L + \tau_S} \quad \text{and} \quad x_{S,1} = x_{S,0} - \frac{\tau_L\rho_S - \tau_S\rho_L}{\tau_L + \tau_S}. \tag{17}$$

Comparing equation ((16)) to equation ((11)) we can see that the cash premium relative to the base case (i.e., $P_1 - P_0$) as a result of the liquidity demand can be positive or negative depending on whether the aggregate liquidity shock (i.e., $\rho_L + \rho_S$) is negative or positive, respectively. If the aggregate liquidity demand is positive (i.e., $\rho_L + \rho_S < 0$), then the cash premium is positive since investors are more willing to hold the asset in equilibrium. On the other hand, if the aggregate liquidity demand is negative (i.e., $\rho_L + \rho_S > 0$), then the cash premium is negative. However, the equilibrium portfolio allocation in this case relative to the base case is not ambiguous — the long investors hold more of the asset in equilibrium, while the short investors hold less.

Case 2: Liquidity shocks with differential pricing. Now suppose that investors have the same endowment shocks as in Case 1, but that an investor must pay a premium R to borrow a security and create a short position, and a long investor can lend a fraction δ of his position. Investor i's optimal portfolio problem is now given by

$$x_{i,2} = \arg\max_{x_i} E\left[W_i\right] - \tfrac{1}{2\tau_i}\mathrm{var}\left[W_i\right], \text{ where } W_i = x_i\left(V - P + \delta_i R\right) + \rho_i V + W_0. \tag{18}$$

We consider the region of parameter space where investor S short-sells the security (i.e. $\delta_S = 1$ and $\delta_L = \delta$). The first order conditions for the optimization problem imply that

$$x_{L,2} = \tfrac{\tau_L}{\sigma_V^2}(m_V - P_2 + \delta R) - \rho_L, \quad \text{and} \quad x_{S,2} = \tfrac{\tau_S}{\sigma_V^2}(m_V - P_2 + R) - \rho_S. \tag{19a}$$

The cash market clearing condition is given by

$$x_{L,2} + x_{S,2} = Q, \tag{20}$$

and the financing market clearing condition is given by

$$\delta x_{L,2} + x_{S,2} = 0. \tag{21}$$

If the repo premium is not constrained to be zero then there is no longer a unique equilibrium. Instead, we can only characterize the equilibrium relationships between prices (P_2 and R) and any pair of long and short positions, x_L and x_S, that clear the cash and repo markets by

$$P_2 = P_1 + \left[\frac{x_{L,2}}{Q}\frac{\sigma_V^2}{\tau_L}(x_{L,1} - x_{L,2}) - \frac{x_{S,2}}{Q}\frac{\sigma_V^2}{\tau_S}(x_{S,2} - x_{S,1})\right] \tag{22a}$$

$$R = \frac{x_{L,2}}{Q}\left[\frac{\sigma_V^2}{\tau_L}(x_{L,1} - x_{L,2}) + \frac{\sigma_V^2}{\tau_S}(x_{S,2} - x_{S,1})\right]. \tag{22b}$$

Therefore the liquidity premium in the cash price, $C = P_2 - P_0$, is given by

$$C = \underbrace{-\frac{\sigma_V^2}{\tau_L + \tau_S}(\rho_L + \rho_S)}_{\text{liquidity component}} + \underbrace{\left[\frac{x_{L,2}}{Q}\frac{\sigma_V^2}{\tau_L}(x_{L,1} - x_{L,2}) - \frac{x_{S,2}}{Q}\frac{\sigma_V^2}{\tau_S}(x_{S,2} - x_{S,1})\right]}_{\text{differential pricing component}}, \tag{23}$$

which is the sum of the standard liquidity component from Case 1 (i.e., $P_1 - P_0$) and an additional premium that arises from the repo market and differential pricing between longs and shorts.

The intuition for equation (22) is as follows. At the Case 1 equilibrium quantities, $x_{L,1}$ and $x_{S,1}$, the liquidity premium paid by the long must be exactly the opposite of the liquidity premium paid by the short. Now consider a perturbation, $x_L = x_{L,1} - \Delta x$ and $x_S = x_{S,1} + \Delta x$, with smaller long and short positions. For these positions to be an equilibrium allocation, prices must adjust so that markets clear. The inverted demand functions at these quantities, based on the first order conditions in (19), determine how prices adjust and are given by:

$$P_2 - \delta R = m_V - \frac{\sigma_V^2}{\tau_L}(x_L + \rho_L) = \underbrace{m_V - \frac{\sigma_V^2}{\tau_L}(x_{L,1} + \rho_L)}_{P_1} + \frac{\sigma_V^2}{\tau_L}\Delta x, \tag{24a}$$

$$R - P_2 = -m_V + \frac{\sigma_V^2}{\tau_S}(x_S + \rho_S) = -\underbrace{\left(m_V - \frac{\sigma_V^2}{\tau_S}(x_{S,1} + \rho_S)\right)}_{P_1} + \frac{\sigma_V^2}{\tau_S}\Delta x. \tag{24b}$$

That is, the long is willing to pay $\frac{\sigma_V^2}{\tau_L}\Delta x$ more to lower their long position by Δx and the short is willing to receive $\frac{\sigma_V^2}{\tau_S}\Delta x$ less to lower their short position by Δx. If the repo premium, R, is constrained to be zero, there is no price, P_2, that clears the market (i.e., satisfies 24). However, once we allow for a non-zero repo premium, this allocation is a viable equilibrium.

Both the long and short agree on a higher cash price if the short pays a positive repo premium to borrow each asset, but the long only collects the repo premium on a fraction $\delta = -x_S/x_L$ of the long position that is lent.

In this case, the amounts the long and short investors are willing to pay to change their positions by Δx are given by:

$$(P_2 - P_1) - \tfrac{x_S}{x_L} R = \tfrac{\sigma_V^2}{\tau_L} \Delta x \quad \text{and} \quad R - (P_2 - P_1) = \tfrac{\sigma_V^2}{\tau_S} \Delta x, \tag{25}$$

respectively. With market clearing, the cash premium is then given by the weighted average of premiums paid by the longs and the shorts:

$$P_2 - P_1 = \tfrac{x_L}{Q} \underbrace{\left((P_2 - P_1) - \tfrac{x_S}{x_L} R\right)}_{\text{premium from longs}} - \tfrac{x_S}{Q} \underbrace{(R - (P_2 - P_1))}_{\text{premium from shorts}} = \tfrac{x_L}{Q} \tfrac{\sigma_V^2}{\tau_L} (x_{L,1} - x_L) - \tfrac{x_S}{Q} \tfrac{\sigma_V^2}{\tau_S} (x_S - x_{S,1}).$$

The equilibrium can also be motivated using the first order conditions in equation (19). When the liquidity premium is higher for both the long and short, they each choose to hedge a smaller portion of their endowment shock.

If we consider the parameter space in which the S investors are short in equilibrium, i.e. $x_{S,1} < 0$, then the differential pricing component is positive if $x_L < x_{L,1}$ and $x_S > x_{S,1}$. In fact, the maximum possible cash price (and hence the maximum liquidity premium) is

$$P_2 = m_V - \tfrac{\sigma_V^2}{\tau_S + \tau_L} \left[Q + \rho_L + \rho_S - \tfrac{(\tau_L \rho_S - \tau_S \rho_L)^2}{4 Q \tau_L \tau_S} \right] > P_1, \tag{26}$$

and it is attained at

$$x_L^* \; = \; \tfrac{\tau_L}{\tau_S + \tau_L} Q + \tfrac{1}{2} \tfrac{\tau_L \rho_S - \tau_S \rho_L}{\tau_L + \tau_S} = \tfrac{x_{L,0} + x_{L,1}}{2}, \quad \text{and} \quad x_S^* = \tfrac{\tau_S}{\tau_S + \tau_L} Q - \tfrac{1}{2} \tfrac{\tau_L \rho_S - \tau_S \rho_L}{\tau_L + \tau_S} = \tfrac{x_{S,0} + x_{S,1}}{2}.$$

The equilibrium can also be characterized in terms of the cash premium, the repo premium and the fraction δ of the asset lent out in equilibrium. The cash and market clearing conditions imply that $x_{L,2} = \frac{Q}{1-\delta}$ and $x_{S,2} = -\frac{\delta}{1-\delta} Q$, which implies the following cash and repo premiums:

$$C \; = \; \frac{\delta \cdot \tau_L + \tau_S}{\tau_L + \tau_S} R\,, \tag{27a}$$

$$R \; = \; \frac{\sigma_V^2}{1-\delta} \left[\left(\frac{\rho_S}{\tau_S} - \frac{\rho_L}{\tau_L} \right) - \frac{Q}{1-\delta} \left(\frac{\delta}{\tau_S} + \frac{1}{\tau_L} \right) \right] \tag{27b}$$

Equation (27a) highlights the fact that the cash premium, C, is less than the repo premium, R, when long investors cannot lend out their entire position (i.e., $\delta < 1$), and as a result, both long and short investors simultaneously pay for the liquidity premium.

We model the cash liquidity premium as the difference in the price of the risky asset when it can and cannot be used to hedge endowment shocks, but one could instead model the liquidity premium as the difference in prices of two securities with identical payoffs that differ in their transactions costs or search frictions. The relevant features of the model are:

(i) some investors have a preference for a long position in the liquid security, while others prefer a short position, (ii) the liquid security is in positive net supply, and so equation (20) must hold and the aggregate long position must be larger than the aggregate short position by the quantity of the security outstanding, and (iii) a short position in the security must be borrowed (i.e., the financing market clearing condition in equation (21) holds).

The equilibrium described by (22) or (27) is pinned down only for a specific choice of δ. Put differently, for an arbitrary δ, the above equations characterize the equilibrium relationship between C and R. As Duffie (1996) and Krishnamurthy (2002) point out, this result is true in general — the market clearing conditions in the cash and repo markets are not sufficient to pin down three values in equilibrium, but instead only characterize the equilibrium relationship between the cash and repo premiums. The equilibrium fraction δ of the long positions lent pins down the cash and repo premiums, and may either be specified exogenously (e.g., as a fixed restriction on how much long investors can lend) or determined endogenously as a function of investor preferences and trading technology.

Risk-Neutral Shorts or Longs. One possible approach to determine the cash and repo premiums in equilibrium would be through assumptions on investor preferences. For instance, if we assume that the short-sellers investors are risk-neutral (i.e., $\tau_S = \infty$), then the equilibrium cash premium is given by

$$C = R. \tag{28}$$

Hence, the net cost to the short-seller is zero, even though there is a non-zero repo premium. Therefore the cost of the liquidity premium is borne completely by the long investor. Similarly, if we assume that the long investors are risk-neutral (i.e. $\tau_L = \infty$), then the equilibrium cash premium is given by

$$C = \delta R, \tag{29}$$

which implies that the net cost to the long investor is zero and the liquidity premium is completely borne by the short investor.

Duffie (1996) and Krishnamurthy (2002) make a hybrid assumption. They assume that while some long investors face lending constraints, others are risk-neutral and can always lend their entire long position to short-sellers. This implies that short-sellers do not pay a liquidity premium in equilibrium, since otherwise the unconstrained long investors would profit from the opposite side of the trade. In our setup, this is equivalent to the assumption that the short-sellers are risk-neutral or do not value liquidity and, as a result, the longs pay for the entire liquidity premium.

Search Frictions. In search-based models of over-the-counter markets (e.g., Vayanos and Weill (2008)), search frictions, investor preferences, and bargaining power jointly determine the equilibrium prices and quantities. The presence of search frictions imposes a bound on how much of his holdings each long investor can lend and thus effectively caps the equilibrium quantity i.e. $\delta \leq \bar{\delta}$. If such a condition is binding, the first order conditions in (19) need not hold and the net cost of being long (or short) may be lower than those given in (24). In particular, the following inequalities may be strict:

$$P_2 - \bar{\delta} R \leq P_1 + \frac{\sigma_V^2}{\tau_L}(x_{L,1} - x_{L,2}) \quad \text{and} \quad R - P_2 \leq -P_1 + \frac{\sigma_V^2}{\tau_S}(x_{S,2} - x_{S,1}) \tag{30}$$

The premiums can be lower (i.e. they no longer have to satisfy the equality) because longs and shorts may demand larger positions at lower premiums but the search technology restricts them from doing so. Investor preferences and bargaining power then determine the size of the liquidity premium and what fraction of this premium each side pays in equilibrium. Long investors and short-sellers both share the liquidity premium if they are risk-averse and do not have full bargaining power. For example, longs would happily pay any fraction $\beta_L \leq 1$ of the maximum amount they're willing to pay, and likewise, shorts would happily pay any fraction $\beta_S \leq 1$ of their maximum amount. Intuitively, one can think of $1-\beta_L$ as the long's bargaining power and $1-\beta_S$ as the short's bargaining power. The equilibrium cash price is given by

$$P_2 = P_1 + \beta_L \tfrac{x_{L,2}}{Q} \tfrac{\sigma_V^2}{\tau_L} \left(x_{L,1} - x_{L,2}\right) + \beta_S \tfrac{x_{S,2}}{Q} \tfrac{\sigma_V^2}{\tau_S} \left(x_{S,2} - x_{S,1}\right) \tag{31}$$

Endogenizing the fraction of securities lent. The fraction δ of their positions that long investors lend out can be endogenized in a number of ways. One possible approach would be to restrict how easily investors can trade the security. For example, suppose that along a investor incurs a per unit cost $c(\delta)$ for the fraction of shares he lends out. Then, the investor's wealth is given by

$$W_L = x_L \left(V - P + \delta R\right) + \rho_L V + W_0 - c(\delta)\, \delta x_L, \tag{32}$$

and the first order conditions for the optimal portfolio problem imply that

$$x_L = \tfrac{\tau_L}{\sigma_V^2}(m_V - P + \delta(R - c(\delta)) - \rho_L \quad \text{and} \quad \delta = \frac{R - c(\delta)}{c'(\delta)} \tag{33}$$

The cash and repo market clearing conditions imply that the equilibrium is given by the following:

$$C \;=\; \tfrac{\delta\tau_L(R - c(\delta)) + \tau_S R}{\tau_L + \tau_S}, \text{ and } R = \tfrac{\sigma_V^2}{1-\delta}\left[\left(\tfrac{\rho_S}{\tau_S} - \tfrac{\rho_L}{\tau_L}\right) - \tfrac{Q}{1-\delta}\left(\tfrac{\delta}{\tau_S} + \tfrac{1}{\tau_L}\right)\right] - \tfrac{\delta}{1-\delta} c(\delta)\,. \tag{34}$$

This implies that the equilibrium fraction δ lent out depends not only on the per unit cost function $c(\delta)$, but also on the preference parameters and endowment shocks in the model.

Another approach to endogenize the fraction of shares lent out is to endow the long investor with market power in the lending market. The long investor accounts for the price impact of how much he lends out and the repo rate he can charge. In our setup, this would imply that a long investor's wealth is given by

$$W_L = x_L \left(V - P + \delta_L R(\delta_L)\right) + \rho_L V + W_{0,} \tag{35}$$

and the first order conditions for the optimal portfolio problem imply that

$$x_L = \tfrac{\tau_L}{\sigma_V^2}(m_V - P_2 + \delta_L R(\delta_L)) \quad \text{and} \quad \delta_L = -\tfrac{R(\delta_L)}{R'(\delta_L)} \tag{36}$$

As before, the equilibrium conditions are given by (27), which implies that the fraction lent is

$$\delta = \tfrac{\rho_S\tau_S - \rho_L\tau_L - \tau_S Q}{\rho_S\tau_S - \rho_L\tau_L + Q(\tau_S + 2\tau_L)}. \tag{37}$$

The equilibrium in this case captures the notion that, in aggregate, the larger the fraction of their positions long investors lend out, the lower the borrowing cost to short-sellers. Given the preference parameters and endowment shocks, the equilibrium fraction δ lent maximizes the total borrowing fees long investors can collect from short-sellers.

5 Conclusions

In this paper we provide the first empirical evidence that short-sellers pay a portion of the liquidity premium for on-the-run Treasuries. Over our sample period from November 1995 through July 2009, we estimate that the liquidity premium for 10-year on-the-run notes was \$195 per \$100,000 issued and short-sellers accounted for about 50% of this amount. We show that short-sellers also accounted for a significant portion of the liquidity premium for on-the-run 5-year notes. Moreover, our results are robust to different methods for calculating the liquidity premium.

Our empirical results provide empirical support for more recent theoretical work such as Vayanos and Weill (2008) which argues that short-sellers also pay a portion of the liquidity premium because they value the ease with which on-the-run Treasuries can be purchased or re-borrowed in the future when closing out or rolling over a short position. Furthermore, we document that the cost of short-selling Treasuries exhibits substantial time-series variation, and is affected by both asset-specific "transactional liquidity" demand and aggregate "episodic liquidity" demand during times of financial crises.

References

Amihud, Y. and Mendelson, H., 1986, Asset pricing and the bid-ask apread, *Journal of Financial Economics*, 17(2), pp. 223–249.

Amihud, Y. and Mendelson, H., 1991, Liquidity, maturity, and the yields on u.s. treasury securities, *Journal of Finance*, 46(4), pp. 1411–1425.

Barclay, M. J., Hendershott, T., and Kotz, K., 2006, Automation versus intermediation: Evidence from treasuries going off the run, *Journal of Finance*, 61(5), pp. 2395–2414.

Duffie, D., 1996, Special repo rates, *Journal of Finance*, 51(2), pp. 493–526.

Duffie, D., Gârleanu, N., and Pedersen, L. H., 2002, Securities lending, shorting, and pricing, *Journal of Financial Economics*, 66(2-3), pp. 307–339.

Fisher, M., 2002, Special repo rates: An introduction, *Federal Reserve Bank of Atlanta Economic Review*, 87(2), pp. 27–43.

Fleming, M., Hrung, W., and Keane, F., 2010, Repo Market Effects of the Term Securities Lending Facility, *The American Economic Review*, 100(2), pp. 591–596.

Goldreich, D., Hankea, B., and Nath, P., 2005, The price of future liquidity: Time-varying liquidity in the u.s. treasury market, *Review of Finance*, 9(1), pp. 1–32.

Graveline, J. J. and McBrady, M. R., 2008, Who makes on-the-run treasuries special? Working Paper.

Jordan, B. D. and Jordan, S. D., 1997, Special repo rates: An empirical analysis, *Journal of Finance*, 52(5), pp. 2051–2072.

Krishnamurthy, A., 2002, The bond/old-bond spread, *Journal of Financial Economics*, 66(2-3), pp. 463–506.

Vayanos, D. and Weill, P.-O., 2008, A search-based theory of the on-the-run phenomenon, *Journal of Finance*, 63(3), pp. 1361–1398.

Warga, A., 1992, Bond returns, liquidity, and missing data, *Journal of Financial and Quantitative Analysis*, 27(4), pp. 605–617.

A Tables and Figures

Table 1: Summary Statistics for Costs of Short-Selling 10-year Treasuries

This table reports the mean, standard deviation, autocorrelation and number of observations for the cost per dollar of shorting more liquid Treasuries for various trading strategies, expressed as an annualized weekly returns. The three strategies considered are (i) shorting the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury, (ii) shorting the on-the-run Treasury and hedging with a duration adjusted long position in the first off-the-run Treasury, and (iii) shorting the first off-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury. The total cost of short-selling the more liquid Treasury in these strategies is given by $R - C$, and the cash and repo components of this cost are given by C and R respectively. The yield to maturity (YTM) and repo interest rates (REPO) for the on-the-run, first off-the-run, and second off-the-run Treasuries are also reported. The full sample is from November 1995 through July 2009, and summary statistics for the subsamples pre August 2007 and post August 2007 are also reported.

	Full Sample			Pre Aug 2007		Post Aug 2007	
	Mean	StDev	AC	Mean	StDev	Mean	StDev
$R - C$ on vs second off-the-run	0.0033	0.0504	0.043	0.0038	0.0466	0.0003	0.0681
$R - C$ on vs first off-the-run	0.0007	0.0488	0.111	0.0002	0.0386	0.0035	0.0848
$R - C$ first vs second off-the-run	0.0024	0.0436	0.031	0.0036	0.0366	-0.0036	0.0699
C on vs second off-the-run	0.0094	0.0496	0.030	0.0099	0.0456	0.0061	0.0682
R on vs second off-the-run	0.0127	0.0107	0.753	0.0138	0.0110	0.0064	0.0064
C first vs second off-the-run	0.0029	0.0426	0.003	0.0025	0.0353	0.0049	0.0699
R first vs second off-the-run	0.0053	0.0074	0.794	0.0061	0.0077	0.0013	0.0022
YTM on on-the-run	0.0492	0.0100	0.990	0.0514	0.0088	0.0364	0.0062
YTM on first off-the-run	0.0494	0.0102	0.990	0.0516	0.0090	0.0363	0.0063
YTM on second off-the-run	0.0493	0.0104	0.990	0.0516	0.0092	0.0361	0.0064
Repo rate on on-the-run	0.0260	0.0192	0.857	0.0284	0.0186	0.0128	0.0170
Repo rate on first off-the-run	0.0312	0.0199	0.938	0.0339	0.0194	0.0172	0.0165
Repo rate on General Collateral	0.0370	0.0195	0.985	0.0403	0.0180	0.0182	0.0171

Table 2: Summary Statistics for Costs of Short-Selling 5-year Treasuries

This table reports the mean, standard deviation, autocorrelation and number of observations for the cost per dollar of shorting more liquid Treasuries for various trading strategies, expressed as an annualized weekly returns. The three strategies considered are (i) shorting the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury, (ii) shorting the on-the-run Treasury and hedging with a duration adjusted long position in the first off-the-run Treasury, and (iii) shorting the first off-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury. The total cost of short-selling the more liquid Treasury in these strategies is given by $R - C$, and the cash and repo components of this cost are given by C and R respectively. The yield to maturity (YTM) and repo interest rates (REPO) for the on-the-run, first off-the-run, and second off-the-run Treasuries are also reported. The full sample is from November 1995 through July 2009, and summary statistics for the subsamples pre August 2007 and post August 2007 are also reported.

	Full Sample			Pre Aug 2007		Post Aug 2007	
	Mean	StDev	AC	Mean	StDev	Mean	StDev
$R - C$ on vs second off-the-run	0.0027	0.0368	-0.126	0.0016	0.0357	0.0091	0.0421
$R - C$ on vs first off-the-run	0.0015	0.0272	-0.045	0.0002	0.0264	0.0082	0.0302
$R - C$ first vs second off-the-run	0.0007	0.0246	-0.136	0.0007	0.0234	0.0010	0.0301
C on vs second off-the-run	0.0076	0.0370	-0.093	0.0095	0.0357	-0.0026	0.0426
R on vs second off-the-run	0.0104	0.0095	0.709	0.0111	0.0097	0.0064	0.0068
C first vs second off-the-run	0.0038	0.0243	-0.140	0.0041	0.0230	0.0022	0.0302
R first vs second off-the-run	0.0045	0.0044	0.600	0.0047	0.0042	0.0032	0.0052
YTM on on-the-run	0.0450	0.0132	0.992	0.0479	0.0116	0.0283	0.0083
YTM on first off-the-run	0.0450	0.0134	0.992	0.0480	0.0118	0.0282	0.0085
YTM on second off-the-run	0.0448	0.0136	0.992	0.0477	0.0121	0.0280	0.0086
Repo rate on on-the-run	0.0295	0.0205	0.878	0.0325	0.0198	0.0126	0.0163
Repo rate on first off-the-run	0.0329	0.0202	0.963	0.0362	0.0191	0.0155	0.0163
Repo rate on General Collateral	0.0370	0.0195	0.985	0.0403	0.0180	0.0182	0.0171

Table 3: Liquidity Costs of Short Selling

This table reports the our estimates for the liquidity costs per dollar of shorting (i.e. $R - C$), fraction of the total liquidity costs paid by the short sellers (i.e. $\delta\frac{R-C}{C}$) and the estimated dollar costs per year for shorting, for various levels of δ. The sample is from November 1995 through July 2009.

Maturity: 10 year	on-the-run	first off-the-run	
Estimate for $R - C$	33 b.p.	24 b.p.	
Estimate for C	94 b.p.	29 b.p.	
Average Issuance	\$13.9B		
Annual Liquidity Costs	\$171M		
Frac of on-the-run shorted δ_{on}	0.75	1.5	1.75
Frac of off-the-run shorted δ_{off}	0.5	0.75	1
Frac of cost to short sellers	30%	55%	66%
Annual dollar cost to short sellers	\$51M	\$94M	\$114M

Maturity: 5 year	on-the-run	first off-the-run	
Estimate for $R - C$	27 b.p.	7 b.p.	
Estimate for C	76 b.p.	38 b.p.	
Avg. Annual Issuance	\$17.4B		
Annual Liquidity Costs	\$199M		
Frac of on-the-run shorted δ_{on}	0.75	1	1.5
Frac of off-the-run shorted δ_{off}	0.5	0.75	1
Frac of cost to short sellers	21%	28%	42%
Annual dollar cost to short sellers	\$41M	\$56M	\$83M

Table 4: Regression of CASH on REPO

This table reports the results from the regression

$$C_t = a + bR_t + \varepsilon_t \tag{38}$$

where C is the return on the cash component and R is the return on the repo component of the cost of shorting (i.e., $R - C$) the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury. The sample is from November 1995 through July 2009. The standard errors and t-statistics in the round brackets are based on OLS standard errors and the standard errors and t-statistics in the square brackets are based on Newey West standard errors with 5 lags. The t-statistic is calculated based on the null hypothesis of $b = 1$.

Maturity: 10 years (714 observations)

Intercept	R	t-stat ($b = 1$)	Adj. R^2
	0.4949		2.65%
	(0.1346)	(-3.752)	
	[0.1454]	[-3.475]	
0.0074	0.1547		3.55%
(0.0031)	(0.2168)	(-3.899)	
[0.0030]	[0.2190]	[-3.860]	

Maturity: 5 years (707 observations)

Intercept	R	t-stat ($b = 1$)	Adj. R^2
	0.6682		6.16%
	(0.1316)	(-2.5215)	
	[0.1152]	[-2.8789]	
0.0016	0.5855		6.24%
(0.0020)	(0.1930)	(-2.1481)	
[0.0021]	[0.1855]	[-2.2339]	

Table 5: Predicted Liquidity Cost of Shorting Selling

This table reports the results from predictive regressions

$$R_{t+1} - C_{t+1} = a + b\,(\text{CP-TB})_t + c\text{Tr}_t\,(10y) + d\text{Tr}_t\,(5y) + \varepsilon_{t+1} \tag{39}$$

where $R - C$ is the cost of shorting the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury, $(\text{CP - TB})_t$ is the lagged 3-month CP - T Bill spread, and $\text{Tr}_t\,(10y)$ and $\text{Tr}_t\,(5y)$ are the weekly Primary dealer transactions in Treasury bonds with maturities comparable to the 10 year and 5 year Treasuries, respectively. The coefficient on each regressor is standardized by the standard deviation of the regressor. Observations are monthly and the sample ranges from January 1997 through July 2009. The table reports the adjusted R^2's for each regression, and OLS standard errors (in round brackets) and Newey West standard errors with 5 lags (in square brackets) for each coefficient.

Maturity: 10 years

Pre Aug 2007 (127 obs)					Full Sample (152 obs)				
Intercpt	CP-TB	Tr $(10y)$	Tr $(5y)$	AdjR^2	Intercpt	CP-TB	Tr $(10y)$	Tr $(5y)$	AdjR^2
-0.0099		0.0042**		2.57%	-0.0015		0.0017		0%
(0.0065)		(0.0018)			(0.0076)		(0.0021)		
[0.0080]		[0.0021]			[0.0086]		[0.0023]		
-0.0171**	0.0032*	0.0050**		3.64%	-0.0068	0.0046	0.0016		1.87%
(0.0067)	(0.0020)	(0.0017)			(0.0072)	(0.0034)	(0.0022)		
[0.0080]	[0.0019]	[0.0020]			[0.0079]	[0.0038]	[0.0024]		
-0.0196*	0.0030	0.0038	0.0016	3.09%	-0.0032	0.0053	0.0039*	-0.0032	1.77%
(0.0089)	(0.0020)	(0.0022)	(0.0028)		(0.0090)	(0.0038)	(0.0023)	(0.0034)	
[0.0105]	[0.0018]	[0.0026]	[0.0032]		[0.0105]	[0.0042]	[0.0020]	[0.0036]	

Maturity: 5 years

Pre Aug 2007 (127 obs)					Full Sample (151 obs)				
Intercpt	CP-TB	Tr $(10y)$	Tr $(5y)$	AdjR^2	Intercpt	CP-TB	Tr $(10y)$	Tr $(5y)$	AdjR^2
-0.0088			0.0024	1.41%	-0.0034			0.0014	0%
(0.0069)			(0.0016)		(0.0060)			(0.0015)	
[0.0076]			[0.0018]		[0.0060]			[0.0015]	
-0.0134*	0.0028*		0.0025	3.72%	-0.0036	0.0006		0.0013	0%
(0.0079)	(0.0016)		(0.0016)		(0.0060)	0.0017		(0.0015)	
[0.0073]	[0.0013]		[0.0016]		[0.0061]	0.0019		[0.0015]	
-0.0132*	0.0028*	-0.0003	0.0027	2.95%	-0.0036	0.0006	0.00003	0.0013	0%
(0.0078)	(0.0016)	(0.0022)	(0.0026)		(0.0060)	0.0018	0.0020	(0.0024)	
[0.0069]	[0.0013]	[0.0023]	[0.0028]		[0.0060]	0.0020	0.0021	[0.0025]	

Table 6: Robustness - Contemporaneous Regressions on Surprises

This table reports the results from predictive regressions

$$R_{t+1} - C_{t+1} = a + b\Delta\,(\text{CP-TB})_{t+1} + c\Delta\text{Tr}_{t+1}\,(10y) + d\Delta\text{Tr}_{t+1}\,(5y) + \varepsilon_{t+1} \tag{40}$$

where $R - C$ is the cost of shorting the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury, $\Delta\,(\text{CP - TB})_t$ is the surprise in 3-month CP - T Bill spread, and $\Delta\text{Tr}_t\,(10y)$ and $\Delta\text{Tr}_t\,(5y)$ are the surprise in the weekly Primary dealer transactions in Treasury bonds with maturities comparable to the 10 year and 5 year Treasuries, respectively. The surprises are calculated relative to an AR(1) model for each regressor. Observations are monthly and the sample ranges from January 1997 through July 2009. The coefficient on each regressor is standardized by the standard deviation of the regressor. The table reports the adjusted R^2's for each regression, and OLS standard errors (in round brackets) and Newey West standard errors with 5 lags (in square brackets) for each coefficient.

Maturity: 10 years

Pre Aug 2007 (127 obs)					Full Sample (152 obs)				
Intercpt	Δ (CP-TB)	ΔTr (10y)	ΔTr (5y)	AdjR^2	Intercpt	Δ (CP-TB)	ΔTr (10y)	ΔTr (5y)	AdjR^2
0.0034*		0.0058**		5.46%	0.0036		0.0042**		1.71%
(0.0020)		(0.0018)			(0.0022)		(0.0021)		
[0.0022]		[0.0019]			[0.0024]		[0.0021]		
0.0037*	0.0025	0.0060**		5.90%	0.0036	0.0010	0.0042**		1.18%
(0.0020)	(0.0020)	(0.0018)			(0.0022)	(0.0024)	(0.0021)		
[0.0021]	[0.0018]	[0.0019]			[0.0024]	[0.0023]	[0.0021]		
0.0037*	0.0025	0.0052**	0.0011	5.26%	0.0036	0.0012	0.0065**	-0.0032	1.19%
(0.0020)	(0.0020)	(0.0019)	(0.0024)		(0.0022)	(0.0023)	(0.0020)	(0.0028)	
[0.0021]	[0.0017]	[0.0019]	[0.0025]		[0.0024]	[0.0022]	[0.0017]	[0.0029]	

Maturity: 5 years

Pre Aug 2007 (127 obs)					Full Sample (151 obs)				
Intercpt	Δ (CP-TB)	ΔTr (10y)	ΔTr (5y)	AdjR^2	Intercpt	Δ (CP-TB)	ΔTr (10y)	ΔTr (5y)	AdjR^2
0.0014			0.0023	1.17%	0.0020			0.0013	0%
(0.0014)			(0.0014)		(0.0014)			(0.0014)	
[0.0016]			[0.0015]		[0.0014]			[0.0013]	
0.0014	0.0001		0.0023	0.4%	0.0020	-0.0032*		0.0015	2.59%
(0.0014)	(0.0013)		(0.0014)		(0.0014)	(0.0016)		(0.0014)	
[0.0016]	[0.0011]		[0.0015]		[0.0015]	[0.0018]		[0.0013]	
0.0014	0.00001	-0.0009	0.0029	0%	0.0020	-0.0032*	-0.0004	0.0018	1.96%
(0.0014)	(0.0013)	(0.0018)	(0.0022)		(0.0014)	(0.0016)	(0.0019)	(0.0020)	
[0.0016]	[0.0010]	[0.0019]	[0.0023]		[0.0015]	[0.0018]	[0.0017]	[0.0021]	

Table 7: Robustness 2 - Crisis

This table reports the results from predictive regressions

$$R_{t+1} - C_{t+1} = a + b\,(\text{CP-TB})_t + c\text{Tr}_t + dI\,(Crisis)_t + \varepsilon_{t+1} \qquad (41)$$

where $R - C$ is the cost of shorting the on-the-run Treasury and hedging with a duration adjusted long position in the second off-the-run Treasury, $(\text{CP - TB})_t$ is the lagged 3-month CP - T Bill spread, $\text{Tr}_t\,(10y)$ is the weekly Primary dealer transactions in Treasury bonds with comparable maturities, $I\,(Crisis)_t$ is an indicator variable for whether the current period is in a period of financial crisis. The three crises in the sample are: the Asian crisis in 1998 (July 1997 through December 1997), the Russian default crisis (August 1998 through January 1999) and the recent sub-prime crisis (August 2007 through January 2009). The coefficient on each regressor is standardized by the standard deviation of the regressor. Observations are monthly and the sample ranges from January 1997 through July 2009. The table reports the adjusted R^2's for each regression, and OLS standard errors (in round brackets) and Newey West standard errors with 5 lags (in square brackets) for each coefficient.

Maturity: 10 years

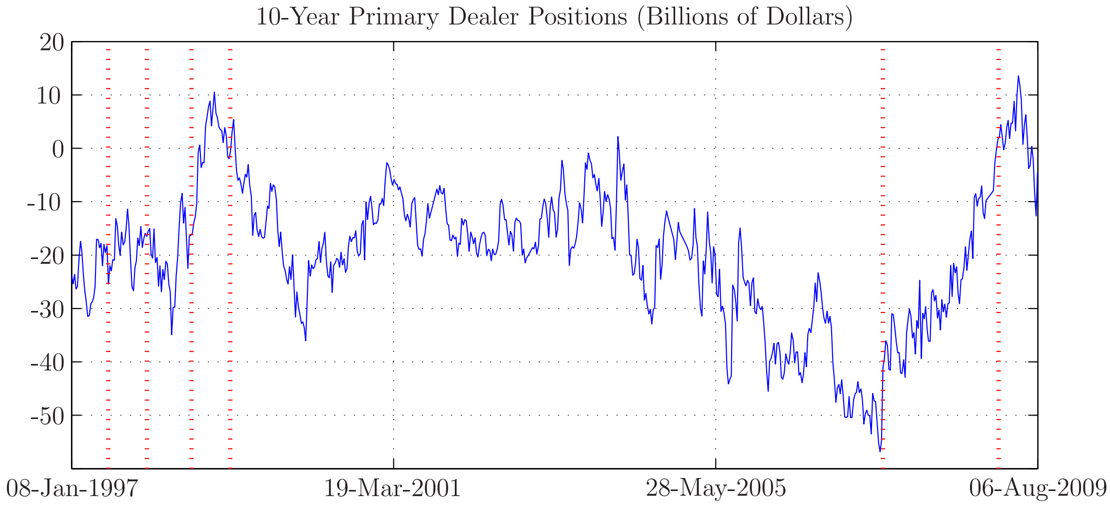
Pre Aug 2007 (127 obs)					Full Sample (152 obs)				
Intercpt	CP-TB	Tr	$I\,(Crisis)$	AdjR^2	Intercpt	CP-TB	Tr	$I\,(Crisis)$	AdjR^2
-0.0171**	0.0032*	0.0050**		3.64%	-0.0068	0.0046	0.0016		1.87%
(0.0067)	(0.0020)	(0.0017)			(0.0072)	(0.0034)	(0.0022)		
[0.0080]	[0.0019]	[0.0020]			[0.0079]	[0.0038]	[0.0024]		
-0.0179	0.0018	0.0055	0.0043	5.92%	-0.0057	0.0022	0.0016	0.0037	2.36%
(0.0066)	(0.0021)	(0.0017)	(0.0026)		(0.0070)	(0.0037)	(0.0021)	(0.0029)	
[0.0081]	[0.0019]	[0.0019]	[0.0021]		[0.0078]	[0.0040]	[0.0023]	[0.0030]	

Maturity: 5 years

Pre Aug 2007 (127 obs)					Full Sample (151 obs)				
Intercpt	CP-TB	Tr	$I\,(Crisis)$	AdjR^2	Intercpt	CP-TB	Tr	$I\,(Crisis)$	AdjR^2
-0.0134*	0.0028*		0.0025	3.72%	-0.0036	0.0006	0.0013		0%
(0.0079)	(0.0016)		(0.0016)		(0.0060)	(0.0017)	(0.0015)		
[0.0073]	[0.0013]		[0.0016]		[0.0061]	[0.0019]	[0.0015]		
-0.0135	0.0022	0.0026	0.0018	4.04%	-0.0021	-0.0015	0.0011	0.0034	1.22%
(0.0078)	(0.0017)	(0.0016)	(0.0015)		(0.0062)	(0.0023)	(0.0015)	(0.0022)	
[0.0072]	[0.0016]	[0.0016]	[0.0017]		[0.0066]	[0.0030]	[0.0016]	[0.0024]	

Figure 3: Primary Dealer Positions

This figure plots the aggregate primary dealer positions (in Billions of Dollars) in 5 year and 10 year Treasury bonds over our sample.



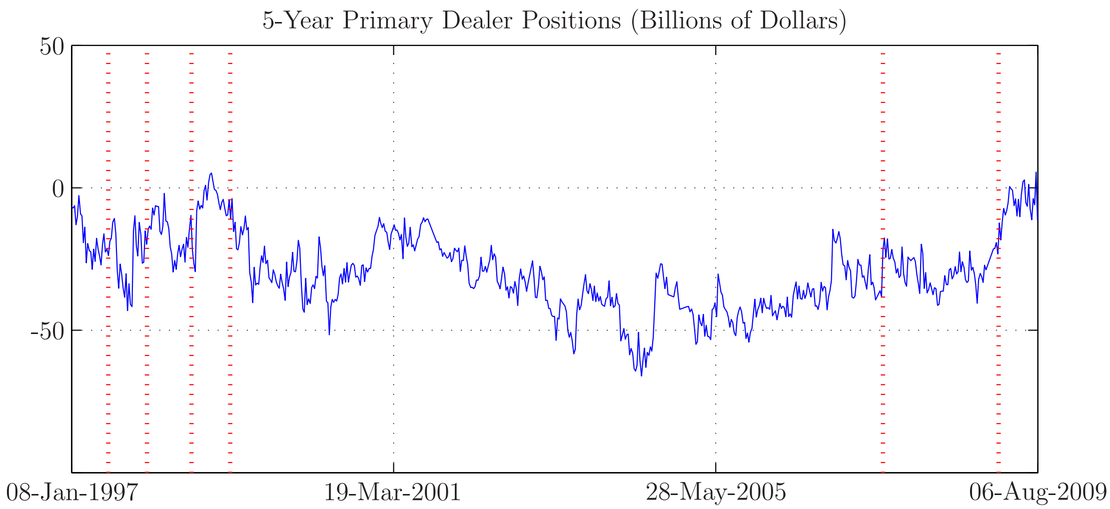


Figure 4: REPO less CASH, CP-TBill spread and Primary Dealer positions

This figure plots the cost of shorting 5-year and 10-year on-the-run Treasury securities (REPO less CASH) at the weekly (thin blue) and monthly (thick black) frequencies, the spread between 3 month commercial paper and Treasury bills (i.e., CP - T Bill Spread), and primary dealer transactions in the 5-year and 10-year Treasuries.