The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXVIII, NO. 2 • APRIL 2013

The Cost of Short-Selling Liquid Securities

SNEHAL BANERJEE and JEREMY J. GRAVELINE*

ABSTRACT

Standard models of liquidity argue that the higher price for a liquid security reflects the future benefits that long investors expect to receive. We show that short-sellers can also pay a *net* liquidity premium if their cost to borrow the security is higher than the price premium they collect from selling it. We provide a model-free decomposition of the price premium for liquid securities into the *net* premiums paid by *both* long investors *and* short-sellers. Empirically, we find that short-sellers were responsible for a substantial fraction of the liquidity premium for on-the-run Treasuries from November 1995 through July 2009.

GIVEN TWO SECURITIES WITH similar cash flows, the more liquid security often trades at a higher price than its less liquid counterpart. This price premium is usually thought to reflect the future benefits that long investors attribute to securities that can be sold quickly and with little price impact (e.g., Amihud and Mendelson (1986)). The more liquid security also frequently costs more to borrow, or trades on special, in financing markets. Previous literature argues that this financing premium is a natural counterpart to the price premium: short-sellers readily pay more to borrow securities that can be sold at a premium (e.g., Duffie (1996), Krishnamurthy (2002)).

However, short-sellers themselves may also value a liquid security over and above the higher sale price they receive. When closing out a position, short-sellers are required to deliver the specific security that they initially borrowed and sold short. As such, they naturally prefer to use liquid securities that can be bought back easily. Indeed, short-sellers can pay a net premium for these future liquidity benefits if it costs them more to borrow the liquid security than they expect to recoup from selling it at a higher price. As the following example illustrates, we use this insight to decompose the price premium for a liquid security into the net premiums paid by long investors and short-sellers.

EXAMPLE: *Suppose that a liquid security trades for* \$100,000, *and an otherwise equivalent but less liquid security costs* \$99,850. *Prices are expected to converge at the end of the period so that the price premium for the liquid security is*

*Snehal Banerjee is with Northwestern University and Jeremy J. Graveline is with the University of Minnesota. We are grateful for comments from Philip Bond; Joost Driessen; Darrell Duffie; Joseph Engelberg; Michael Fleming; Nicolae Gârleanu; Arvind Krishnamurthy; Francis Longstaff; David Matsa; Lasse Pedersen; Raj Singh; Dimitri Vayanos; Pierre-Olivier Weill; and participants at the NBER AP Meeting (San Francisco, 2010), the Adam Smith Asset Pricing Meeting (Oxford, 2011), the SFS Cavalcade (Ann Arbor, 2011), and the Western Finance Association Meeting (Santa Fe, 2011).

$100,000 − $99,850 = $150 *relative to its illiquid counterpart. Suppose that it costs* $200 *more to borrow the liquid security for the period than it does to borrow the illiquid one. Finally, assume that each outstanding unit of the liquid security is borrowed and sold short once, so that the aggregate proportion of long positions relative to short positions is two to one.*

In this example, the net premium for a short position in the liquid security is $50. Short-sellers pay $200 more to borrow it but recoup $150 by selling it at the higher price. As a whole, long investors in the liquid security pay a net premium of $100. For each outstanding unit of the security, they pay the $150 price premium twice but recoup $200 by lending it once to short-sellers (2 × $150 − $200 = $100). In aggregate, the net liquidity premiums paid by longs and shorts (given by $100 and $50) account for two-thirds and one-third of the $150 price premium, respectively.

As the example illustrates, both long investors and short-sellers can simultaneously contribute to the price premium for a liquid security. Moreover, the magnitude of each side's contribution is characterized by the net premiums that they pay. Long investors pay a net liquidity premium when the price premium for the liquid security is higher than what they recover from lending out a portion of their position to short-sellers. Short-sellers pay a net liquidity premium when their incremental cost to borrow the liquid security is higher than the price premium they expect to recoup from selling it. In particular, note that:

- A higher borrowing fee does not imply that short-sellers pay a net liquidity premium. In fact, if the price and borrowing premiums are equal (as much of the earlier literature suggests), then short-sellers fully recoup their higher borrowing costs by selling the security at a price premium. In the example above, if the price premium and borrowing fees are both $150, then long investors actually pay for the entire liquidity premium because twice as many units of the security are held in long positions as in short positions (i.e., 2 × $150 − $150 = $150).
- Similarly, a positive price premium does not imply that long investors pay a net premium for the liquid security. In the example above, if the price premium is $150 but short-sellers pay $300 for each unit they borrow, then long investors do not pay a net premium since they recover all of the price premium by lending to short-sellers (2 × $150 − $300 = $0). Instead, short-sellers ultimately pay for the entire liquidity premium because they pay $300 − $150 = $150 more to borrow the security than they expect to recoup by selling it at the higher price.

To be clear, it is not our objective in this paper to provide a model that explains the level of the price and borrowing premiums for a liquid security, nor the relative proportion of long and short positions. Instead, we take these values as given and derive a decomposition of the price premium that explicitly quantifies how much long investors and short-sellers each pay for liquidity. Our decomposition is model-free and provides an important first step in understanding the economic determinants of the liquidity price premium by clarifying who ultimately pays for it. For instance, while earlier work argues that the price premium reflects the present value of future borrowing fees (e.g., Duffie (1996), Krishnamurthy (2002), Duffie, Gârleanu, and Pedersen (2002)) and liquidity benefits (e.g., Vayanos and Weill (2008)), it is silent on whether long investors or short-sellers ultimately pay for the price premium. Also, while earlier models emphasize the role of short-sellers in generating a liquidity premium (e.g., Vayanos and Weill (2008)), our decomposition is the first to explicitly quantify how much of the liquidity premium is actually paid by short-sellers.

We also show that this decomposition is empirically important. We decompose the liquidity premium for the 10-year on-the-run Treasury note and find that, from November 1995 through July 2009, short-sellers accounted for an average of 37% of the premium. In predictive regressions, we find that the liquidity premium paid by short-sellers is positively related to primary dealer transactions in Treasuries with similar maturities, which suggests that short-sellers are willing to pay more for positions in liquid notes when they anticipate having to trade more frequently. We also find a positive relation between the liquidity premium for shorting and the commercial paper and Treasury Bills (CP–TBill) spread over our entire sample, which suggests that the expected cost of shorting the liquid notes is higher during financial crises.

The remainder of this paper is organized as follows. In Section I, we discuss our marginal contribution relative to the existing literature. In Section II, we characterize the general decomposition of the price premium for liquid securities in terms of prices, borrowing fees, and the aggregate volume shorted, and discuss the implications of this decomposition. In Section III, we present the results from applying the decomposition to the liquidity premium for on-the-run Treasuries. In Section IV, we present a basic theoretical framework to illustrate how the presence of lending constraints for long investors can lead to a liquidity premium that is shared with short-sellers. Section V concludes.

I. Related Literature

Duffie (1996) is 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. Intuitively, short-sellers willingly pay more to borrow securities that they can sell at a price premium, while long investors willingly pay a higher price for securities that they can lend at a premium to short-sellers. Jordan and Jordan (1997), Krishnamurthy (2002), and Goldreich, Hanke, and Nath (2005), among others, provide empirical support for the relationship that higher prices and higher borrowing fees go hand in hand.

The working paper version of Barclay, Hendershott, and Kotz (2006) shows that, on a typical day, around 100% 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 a short position out (e.g., Duffie (1996), Krishnamurthy (2002), Vayanos and Weill (2008), Graveline and McBrady (2011)). However, earlier theoretical models by Duffie (1996) and Krishnamurthy (2002) imply that the price and borrowing premiums for securities must be equal. In this setting, short-sellers fully recoup their higher borrowing costs and thus do not actually pay a net premium or contribute to the price premium.

Vayanos and Weill (2008) develop a model with search frictions in the cash and financing markets. They show that short-sellers can endogenously concentrate their positions in the same security and long investors choose to follow suit. As a result, this security is more liquid and commands both price and borrowing premiums. Moreover, they show that neither premium would exist in the absence of short-sellers. However, as we highlighted in the introduction, the existence of price and borrowing premiums does not convey how much of the price premium (if any) is ultimately paid for by short-sellers. Our model-free decomposition, which we view as complementary to their work, quantifies the specific contributions of long investors and short-sellers to the price premium. Furthermore, as we discuss in the next section, the intuition behind our decomposition should extend beyond search-based models to other market structures, so long as both longs and shorts value positions in liquid securities (relative to their illiquid counterparts) but face constraints that prevent them from taking arbitrarily large positions.

In our empirical analysis, we calculate the historical premiums paid by short-sellers who borrow and sell on-the-run Treasuries and by long investors who buy these bonds and lend (finance) a portion of them in the repo market. With this integrated analysis of the cash and repo markets, we find that the average premium paid by short-sellers is a substantial fraction of the total liquidity premium for the on-the-run Treasuries. Consistent with earlier studies (e.g., Amihud and Mendelson (1991), Warga (1992)), we document that the average annualized liquidity premium on the 10-year on-the-run notes relative to less liquid off-the-run notes is 94 basis points during our sample. Our decomposition implies that the average annualized cost of short-selling these notes is 33 basis points, and varies substantially over time.

Our empirical analysis is most closely related to Krishnamurthy (2002). Although he does not focus on the premium for 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. In contrast, we find that short-sellers account for a substantial fraction of the observed liquidity premium for on-the-run 5- and 10-year Treasury notes (relative to the second off-the-run notes). 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 5- and 10-year notes and use a data series from November 1995 to July 2009 that is three times as long. We also calculate the liquidity premium relative to the second most recently issued, or second off-the-run, Treasury note because the first off-the-run note is still frequently used for short-selling and is often expensive to borrow in the financing, or repo, market.[1] 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 rise and fall in unison. We statistically reject this hypothesis in our longer sample.

More generally, our paper relates to the broad literature that analyzes the "on-the-run" phenomenon. While differences in liquidity has been the most common explanation (e.g., Amihud and Mendelson (1986, 1991), Warga (1992)), other explanations for the differential pricing of comparable Treasury securities that have been proposed include asymmetric information or heterogeneous interpretation of public signals (e.g., Brandt and Kavajecz (2004), Green (2004), Li et al. (2009), Pasquariello and Vega (2009)), differential tax treatments (e.g., Kamara (1994), Strebulaev (2002)), and market squeezes (e.g., Cornell and Shapiro (1989), Nyborg and Sundaresan (1996), Nyborg and Strebulaev (2004)). In contrast to this literature, our main focus is to quantify who actually pays for the liquid premium, which is an important first step in understanding its economic determinants.

II. Decomposing the Price Premium

In this section, we provide a general framework to decompose the price premium for a liquid security into the portions that are paid by long investors and short-sellers. The decomposition that follows can be applied to the price premium for any liquid security relative to its illiquid counterpart, but as a concrete example, one can think of a recently issued on-the-run Treasury note versus a comparable, but less liquid, seasoned off-the-run Treasury.

Consider two securities with the same cash flows that differ only in their liquidity characteristics. Let C be the price premium for the liquid security in the cash market and R be the premium to borrow it in the financing market. That is, the liquid security costs C dollars more to buy and R dollars more to borrow than its illiquid counterpart. In the case of Treasury notes, C reflects the higher price or lower yield for on-the-run Treasuries as compared to off-the-run Treasuries with similar maturity, and R is the repo special (adjusted for haircuts) for borrowing on-the-run notes.[2] Finally, let δ denote the aggregate

[1] Barclay, Hendershott, and Kotz (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. 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 is financed at the general collateral repo rate.

[2] In practice, financing markets for Treasuries and most other securities often include haircuts. An H% haircut on a security means that long investors can use the security as collateral and borrow against $(1 - H)$% of its market value. In turn, short-sellers only pay any borrowing premium on $(1 - H)$% of the security's value. We discuss the issue of haircuts in more detail in Section III.E.

volume of the liquid security that is borrowed and sold short, expressed as a fraction of the total outstanding supply. Every security that is sold short must be held in a long position, so long investors hold a fraction $1+\delta$ of the outstanding supply of the liquid security.

Note that the price and financing premiums, C and R, reflect not only the future benefits that longs and shorts attribute to positions in the more liquid security, but also what they expect to recoup from each other. For instance, short-sellers are willing to pay a premium R to borrow the liquid security (rather than its illiquid counterpart), but they expect to recoup the price premium C by selling it at a higher price (again, relative to the illiquid counterpart). Therefore, the net premium they pay is $R-C$ per unit sold short. Since a fraction δ of the outstanding supply of the liquid security is sold short, the aggregate net premium paid by short-sellers is $\delta \times (R-C)$. Similarly, long investors pay a premium C each time they buy the liquid security, but expect to recoup the financing premium R each time they lend it to short-sellers. A fraction $1+\delta$ of the outstanding supply of the liquid security is held in long positions, but in aggregate long investors lend a fraction δ to short-sellers. As a result, the aggregate net premium paid by long investors in the liquid security is $(1+\delta)\times C-\delta\times R$. Together, the net premiums paid by long investors and short-sellers sum to the price premium on the liquid security, so that

$$C = \underbrace{(1+\delta)\times C-\delta\times R}_{\substack{\text{Longs'}\\ \text{Contribution}}} + \underbrace{\delta\times[R-C]}_{\substack{\text{Shorts'}\\ \text{Contribution}}}. \tag{1}$$

The decomposition in equation (1) highlights the importance of jointly analyzing the price premium, the borrowing or financing premium, and the fraction of the outstanding supply of the liquid security sold short. As we discussed in the introductory example, a financing premium (i.e., $R>0$) does not necessarily imply that short-sellers pay a premium for positions in the liquid security, since it is possible that they recover these higher borrowing costs completely (i.e., if $C=R$). Similarly, a positive price premium (i.e., $C>0$) does not imply that long investors pay a net premium since they may be able to fully recover these costs by lending out their positions (i.e., if $R=\frac{1+\delta}{\delta}C$).

It is important to note that the decomposition in equation (1) is at the aggregate level in that it measures the total liquidity premium paid by all longs and all short-sellers. However, while each short-seller must borrow the liquid security, there is likely to be significant variation across different long investors in the fraction of their positions they lend out. For instance, consider the spectrum of long investors in Treasury markets. At one extreme, hedge funds and dealers are often anxious to lend their positions to finance their trading activities. In fact, the decomposition in equation (1) implies that, if these active investors lend more than a fraction $\frac{\delta}{1+\delta}$ of their long position, they may actually get paid for increasing the supply of the liquid security that is available to be sold short. At the other extreme, many foreign central banks often forgo the specials they can earn by lending out their notes and so recover almost none of the price premium they pay for long positions in on-the-run Treasuries. In between, mutual funds and insurance companies can face institutional constraints on the amount they lend in repo markets.

In Section III, we apply the decomposition in equation (1) to the liquidity premium for on-the-run Treasury notes relative to their off-the-run counterparts. However, the decomposition also applies to the liquidity premium for other assets, and is especially important in understanding the liquidity premium for securities in which a substantial fraction of the outstanding supply is sold short. For example, a similar decomposition could be applied to the liquidity premium on Treasury notes relative to agency debt (e.g., Longstaff (2004), Krishnamurthy (2010)), the liquidity component of credit spreads (e.g., Longstaff, Mithal, and Neis (2005)), or the liquidity components of spreads on various securitized products (e.g., Gorton and Metrick (2012)). Measuring the liquidity premiums in these asset classes is difficult due to confounding factors like credit risk, counterparty risk, and differences in perceived safety. Even though empirically estimating the decomposition is more challenging for these securities, the insights from the decomposition in equation (1) are likely to be relevant.

A. Lending Constraints and the Relation to Previous Models

The decomposition in equation (1) is an identity that relates prices and quantities, and therefore it does not rely on any specific modeling assumptions. In this subsection, we discuss its implications for how the liquidity preferences and constraints faced by investors interact to determine equilibrium prices and quantities. In Section IV, we provide a simple theoretical framework to formally describe how the equilibrium price and borrowing premiums arise as a result of lending constraints faced by long investors.

The decomposition in equation (1) sheds some light on the constraints faced by long investors. Duffie (1996) shows that the premium to borrow the liquid security must be at least as large as the price premium (i.e., $R - C \geq 0$). Otherwise, if $R - C < 0$, there would be an arbitrage opportunity to short-sell the liquid security and hedge one's risk with an offsetting long position in the illiquid security. Our analysis suggests that the premium to borrow the liquid security may, in fact, be strictly larger than the price premium (i.e., $R - C > 0$), which implies that short-sellers pay a liquidity premium. It is instructive to examine the assumptions in earlier work that preclude this result.

The models in Duffie (1996) and Krishnamurthy (2002) assume that there is an unconstrained arbitrageur who can hold arbitrarily large positions in the liquid security and lend out his entire position to short-sellers (while hedging the risk with an offsetting position in the illiquid security). This assumption ensures that short-sellers do not pay a liquidity premium (i.e., $R = C$), since otherwise the unconstrained long investor could make arbitrarily large profits by lending out all of his long position in the liquid security at a premium and hedging with the illiquid security. Therefore, if short-sellers pay a liquidity premium (i.e., $R - C > 0$), then all long investors must be either constrained or reluctant to take arbitrarily large positions with this trade. In other words, all long investors are either unable or unwilling to lend out their entire position in the liquid security, or they find it difficult to create short positions in the illiquid security to hedge their risk.

Vayanos and Weill (2008) develop a model with search frictions that effectively cap the equilibrium fraction that each long investor expects to lend at $\delta/(1+\delta)$. They show that arbitrageurs stay out of the market in their model when[3]

$$C \leq R \quad \text{and} \quad \frac{\delta}{1+\delta} R \leq C. \tag{2}$$

These inequalities imply that both long investors and short-sellers can, but do not necessarily, pay a net premium in their model. Vayanos and Weill (2008) also decompose the price premium for the liquidity security as

$$C = L + \frac{\delta}{1+\delta} R, \tag{3}$$

where they refer to L as the liquidity premium. However, their decomposition in equation (3) is silent on whether long investors or short-sellers ultimately pay for the price premium.

Our analysis adds two additional insights to their work. First, we highlight that, for each short position, short-sellers pay the borrowing premium R but recoup the price premium C from equation (3). Therefore, the net premium per unit that short-sellers in their model pay is

$$R - \underbrace{\left[\frac{\delta}{1+\delta} R + L\right]}_{C} = \frac{R}{1+\delta} - L. \tag{4}$$

Second, our decomposition emphasizes that the contributions of long investors and short-sellers also depend on the proportion of the aggregate supply that each holds. Long investors pay the liquidity premium L for each unit of their position. That is, they pay the price premium C but expect to recoup the borrowing premium R on the fraction $\delta/(1+\delta)$ of their position that is lent to short-sellers. However, the contribution of all long investors to the price premium scales the liquidity premium L by the aggregate proportion of the outstanding supply of the security that is held in long positions, that is, $1+\delta$. Similarly, the total contribution of all short-sellers to the price premium multiplies the net premium they pay per unit, $R - C$, by the fraction δ of the aggregate supply of the liquid security that is sold short. Thus, our decomposition of the price premium in equation (1), as applied to the model in equation (3)

[3] See equations (12) and (13) in Section III.D of their paper.

from Vayanos and Weill (2008), is

$$C = \underbrace{(1+\delta)L}_{\text{Longs}} + \underbrace{\delta\left[\frac{R}{1+\delta} - L\right]}_{\text{Shorts}}. \quad (5)$$

III. Empirical Analysis

As an empirical application of the decomposition in equation (1), we estimate the fraction of the liquidity premium for on-the-run Treasury notes that is paid for by short-sellers. The Treasury market is an ideal setting for our empirical analysis as it provides securities with very similar cash flows that differ primarily in how liquid they are Barclay, Hendershott, and Kotz (2006) report that the average daily trading volume in the on-the-run 2-, 5-, and 10-year maturities rivals the volume in all U.S. stocks combined. However, when new notes are issued and the existing ones move off-the-run, trading volume drops by 90% and these notes become relatively less liquid. Moreover, around 100% of the outstanding on-the-run 10-year notes are typically sold short, which suggests that the liquidity premium paid by short-sellers is likely to be an important component of the total premium on these notes.

Our earlier theoretical analysis assumed the existence of two securities with identical future cash flows that differ only in their liquidity. In practice, on-the-run and seasoned off-the-run Treasuries have similar, but not identical, cash flows. To address this issue, our empirical analysis compares the cash and financing market returns for duration-matched positions in these Treasuries with virtually identical exposure to interest rates. A related practical issue is that we do not directly observe the ex ante expected cash premium. As such, our subsequent empirical analysis assumes that the average ex post realized premium reflects the market's ex ante expectations.

As we describe in more detail below, our empirical estimate of the liquidity premium paid by short-sellers is based on the cost of the following "on-the-run versus off-the-run" trading strategy:

(i) short-sell \$1 of the on-the-run Treasury note, and
(ii) hedge the interest rate risk with a (duration-adjusted) long position in the second off-the-run.

Given that on-the-run Treasuries have historically traded at a price premium relative to their off-the-run counterparts, one might expect this strategy to be profitable (i.e., the cost of the strategy to be negative). However, we show that this strategy was costly on average over our sample period, which implies that short-sellers paid a liquidity premium.

It is important to emphasize that the cost of the above trading strategy reflects the incremental cost of a short position in the liquid on-the-run note relative to a short position in the (relatively) illiquid off-the-run note. It is not our objective to explain why market participants want to have short positions in Treasury notes. Rather, we want to understand the premium they pay for choosing to hold a short position in the more liquid note.

A. Data Description and Estimation Procedure

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.[4] 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. We focus on overnight repo rates since Barclay, Hendershott, and Kotz (2006) report that 94% of repos in their sample are overnight agreements. In addition, due to the settlement differences between the cash and financing markets (which we discuss below), a long investor can lend a security overnight and is still free to sell the security that same day. In contrast, a term repo agreement (i.e., longer than overnight) would exclude this activity, which is not consistent with the extremely high turnover rate that Barclay, Hendershott, and Kotz (2006) document for on-the-run Treasuries.

To measure the price and borrowing premiums, we begin by computing the ex post cost of short-selling a Treasury for a day. In so doing, we need to correctly account for the fact that the cash market for Treasuries is typically next-day settlement, whereas the repo, or financing, market is same-day settlement. Therefore, if one short-sells a Treasury at time t, she receives the sale price P_t at time $t+1$ and must borrow and deliver the security on that date. To borrow the security at time $t+1$, the short-seller lends the price of the security P_{t+1} to an owner of that security and receives the security as collateral. The interest rate r_{t+1} on the loan is referred to as the repo rate for that security. At the same time, the short-seller repurchases the Treasury and at time $t+2$ she receives the Treasury in exchange for the purchase price. She returns the Treasury to the owner that she originally borrowed it from and receives $1+r_{t+1}$ for every dollar that she lent against the security. Note that, at time $t+1$, a short-seller receives the sale price P_t but lends P_{t+1}. We assume that the difference, $P_{t+1}-P_t$, which may be either positive or negative, is financed at the general collateral repo rate r^{gc}_{t+1} (the highest available interest rate for lending against Treasury collateral). Thus, the ex post cost of short-selling \$1 of the on-the-run Treasury note for a day is

$$\begin{aligned} &\frac{P^{on}_{t+1} - P^{on}_{t+1}\left(1+r^{on}_{t+1}\right) + \left[P^{on}_{t+1} - P^{on}_{t}\right]\left(1+r^{gc}_{t+1}\right)}{P^{on}_{t}} \\ &= \frac{P^{on}_{t+1}}{P^{on}_{t}} - \left(1+r^{on}_{t+1}\right) + \left[\frac{P^{on}_{t+1}}{P^{on}_{t}} - 1\right]\left(r^{gc}_{t+1} - r^{on}_{t+1}\right), \end{aligned} \tag{6}$$

[4] Bloomberg follows industry convention and quotes (clean) prices without accrued interest. To compute the true (dirty) prices of the notes, we add the accrued interest to the clean prices that Bloomberg provides.

where $P_{t+1}^{\text{on}}/P_t^{\text{on}}$ is the return from time t to $t+1$ for the on-the-run note, and r_{t+1}^{on} is its repo rate from time $t+1$ to $t+2$.

To isolate the price and borrowing premiums, we compare the raw short-selling cost in equation (6) to the cost of a short position with similar interest rate exposure in the second off-the-run Treasury. That is, we compare the cost of short-selling \$1 of the on-the-run with the cost of short-selling $\$(\text{DUR}_t^{\text{on}}/\text{DUR}_t^{\text{off2}})$ in the second off-the-run, where DUR_t^{on} and $\text{DUR}_t^{\text{off2}}$ are the duration of the on-the-run and second off-the-run securities, respectively. We assume that the second off-the-run can be financed at the general collateral repo rate (i.e., $r_{t+1}^{\text{off2}} = r_{t+1}^{\text{gc}}$). Therefore, the ex post cost of short-selling $(\text{DUR}_t^{\text{on}}/\text{DUR}_t^{\text{off2}})$ of the second off-the-run note is

$$\frac{\text{DUR}_t^{\text{on}}}{\text{DUR}_t^{\text{off2}}} \underbrace{\left\{ \frac{P_{t+1}^{\text{off2}}}{P_t^{\text{off2}}} - (1 + r_{t+1}^{\text{off2}}) + \left[\frac{P_{t+1}^{\text{off2}}}{P_t^{\text{off2}}} - 1 \right] (r_{t+1}^{\text{gc}} - r_{t+1}^{\text{off2}}) \right\}}_{P_{t+1}^{\text{off2}}/P_t^{\text{off2}} - \left(1 + r_{t+1}^{\text{gc}}\right)}. \tag{7}$$

The liquidity premium paid by short-sellers is then just the difference in the cost of short-selling the on-the-run Treasury relative to the cost of selling the duration-adjusted position in the second off-the-run Treasury, which can be rewritten as[5]

$$\underbrace{\frac{\text{DUR}_t^{\text{on}}}{\text{DUR}_t^{\text{off2}}} r_{t+1}^{\text{gc}} - r_{t+1}^{\text{on}} + \left[\frac{P_{t+1}^{\text{on}}}{P_t^{\text{on}}} - 1 \right] (r_{t+1}^{\text{gc}} - r_{t+1}^{\text{on}})}_{R_{\text{on},t} = \text{Borrowing Premium}}$$
$$- \underbrace{\left[\frac{\text{DUR}_t^{\text{on}}}{\text{DUR}_t^{\text{off2}}} \left(\frac{P_{t+1}^{\text{off2}}}{P_t^{\text{off2}}} - 1 \right) - \left(\frac{P_{t+1}^{\text{on}}}{P_t^{\text{on}}} - 1 \right) \right]}_{C_{\text{on},t} = \text{Cash Price Premium}}. \tag{8}$$

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 $C_{\text{on},t}$ will be positive on average. Similarly, on-the-run Treasuries are frequently on special in the repo market (i.e., $r_t^{\text{gc}} > r_t^{\text{on}}$) and the durations are usually close so we expect that $R_{\text{on},t}$ will be positive on average. We construct the weekly counterpart to equation (8) by initiating the above trading strategy each Wednesday

[5] One could instead use the yield spread between the on- and off-the-run notes to construct a measure of the ex ante price premium, but this approach would require some additional assumptions. For example, one could adopt a longer horizon and assume that the (duration-adjusted) yields will be equal when the on-the-run note becomes the second off-the-run. However, there is no natural counterpart for a long horizon measure of the ex ante borrowing premium, as the vast majority of repurchase agreements are short-term contracts. Alternatively, a short horizon measure of the ex ante price premium would require a model that describes how the yield (price) difference is amortized over time (i.e., how quickly the difference is expected to converge). We opted to use the simple approach in equation (8) because it requires no additional modeling assumptions.

Table I
Summary Statistics for Costs of Short-Selling 10-Year Treasuries

This table reports the mean, standard deviation, and autocorrelation (AC) for the cost per dollar of shorting (i.e., $R - C$) more liquid Treasuries for various trading strategies, expressed as annualized weekly returns. The 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 versus second off-the-run	0.0033	0.0070	0.0430	0.0038	0.0065	0.0003	0.0094
$R - C$ on versus first off-the-run	0.0007	0.0068	0.1109	0.0002	0.0053	0.0035	0.0118
$R - C$ first versus second off-the-run	0.0024	0.0061	0.0306	0.0036	0.0051	−0.0036	0.0097
C on versus second off-the-run	0.0094	0.0069	0.0304	0.0099	0.0063	0.0061	0.0095
C first versus second off-the-run	0.0029	0.0059	0.0025	0.0025	0.0049	0.0049	0.0097
R on versus second off-the-run	0.0127	0.0015	0.7528	0.0138	0.0015	0.0064	0.0009
R first versus second off-the-run	0.0053	0.0010	0.7935	0.0061	0.0011	0.0013	0.0003
YTM on on-the-run	0.0492	0.0100	0.9896	0.0514	0.0088	0.0364	0.0062
YTM on first off-the-run	0.0494	0.0102	0.9898	0.0516	0.0090	0.0363	0.0063
YTM on second off-the-run	0.0493	0.0104	0.9901	0.0516	0.0092	0.0361	0.0064
Repo rate on on-the-run	0.0260	0.0192	0.8570	0.0284	0.0186	0.0128	0.0170
Repo rate on first off-the-run	0.0312	0.0199	0.9379	0.0339	0.0194	0.0172	0.0165
Repo rate on general collateral	0.0370	0.0195	0.9853	0.0403	0.0180	0.0182	0.0171

and financing the daily profits or losses at the federal funds rate. Our empirical estimates of the ex ante cash and financing premiums, C_{on} and R_{on}, are computed as the average of their ex post weekly counterparts in equation (8).

B. Summary Statistics

Tables I and II present summary statistics for the 10- and 5-year maturity trading strategies, respectively. On average, the repo rate for the on-the-run 10-year maturity note is about 110 basis points lower than the general collateral repo rate (since it is 260 basis points for the on-the-run vs. 370 basis points for general collateral). Over the whole sample, the weekly cost of shorting the on-the-run is 33 basis points (annualized return). For comparison, the cost of short-selling the on-the-run relative to the first off-the-run is seven basis points. Table II suggests that these results also extend to the 5-year maturity. The average repo rate is about 75 basis points lower than the general collateral repo rate and the cost of short-selling the on-the-run relative to the second off-the-run is 28 basis points (annualized return), which is lower than the same strategy for the 10-year maturity.

Table II

Summary Statistics for Costs of Short-Selling 5-Year Treasuries

This table reports the mean, standard deviation, and autocorrelation (AC) for the cost per dollar of shorting (i.e., $R - C$) more liquid Treasuries for various trading strategies, expressed as annualized weekly returns. The 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 versus second off-the-run	0.0028	0.0051	−0.1184	0.0018	0.0049	0.0091	0.0058
$R - C$ on versus first off-the-run	0.0014	0.0038	−0.0389	0.0001	0.0037	0.0082	0.0042
$R - C$ first versus second off-the-run	0.0008	0.0034	−0.1378	0.0008	0.0032	0.0010	0.0042
C on versus second off-the-run	0.0075	0.0051	−0.0849	0.0093	0.0049	−0.0026	0.0059
C first versus second off-the-run	0.0037	0.0034	−0.1441	0.0040	0.0032	0.0022	0.0042
R on versus second off-the-run	0.0104	0.0013	0.7083	0.0110	0.0013	0.0064	0.0009
R first versus second off-the-run	0.0045	0.0006	0.6020	0.0048	0.0006	0.0032	0.0007
YTM on on-the-run	0.0449	0.0132	0.9922	0.0478	0.0116	0.0283	0.0083
YTM on first off-the-run	0.0450	0.0134	0.9924	0.0479	0.0118	0.0282	0.0085
YTM on second off-the-run	0.0447	0.0136	0.9925	0.0476	0.0121	0.0280	0.0086
Repo rate on on-the-run	0.0295	0.0205	0.8784	0.0325	0.0198	0.0126	0.0163
Repo rate on first off-the-run	0.0329	0.0202	0.9626	0.0362	0.0191	0.0155	0.0163
Repo rate on general collateral	0.0370	0.0195	0.9853	0.0403	0.0180	0.0182	0.0171

C. Do Short-Sellers Pay a Liquidity Premium?

If long investors bear the entire liquidity premium, then the cash and repo premiums should be equal (i.e., $R = C$). A standard test of this null hypothesis is whether the sample average of $R - C$ is statistically different from zero. In our sample, although the estimates of $R - C$ from Tables I and II are generally positive, these estimates are extremely noisy and one cannot statistically reject the null hypothesis that the average cost of short-selling (i.e., average $R - C$) is zero.

However, as a more powerful test of the null hypothesis, 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 equal to one. Table III provides the results of this regression (both with and without a constant). For the on-the-run 5- and 10-year maturities, the regression coefficients are 0.66 and 0.49, respectively, but for both maturities we can reject the null hypothesis that the regression coefficients are equal to one at the 5% level.

Our regression results are consistent with earlier results documented in the literature. For instance, using a cross section of Treasury notes from

Table III

Regression of Cash Returns on Repo Returns

This table reports the results from the regression

$$C_t = \alpha + \beta R_t + \varepsilon_t,$$

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 five lags. The t-statistic is calculated based on the null hypothesis of $b = 1$. The autocorrelation (AC) in the regression residuals and the number of observations are also reported.

	Maturity: 10 years		Maturity: 5 years	
Intercept		0.0074		0.0015
		(0.0031)		(0.0020)
		[0.0030]		[0.0021]
REPO	0.4949	0.1547	0.6617	0.5841
	(0.1346)	(0.2168)	(0.1310)	(0.1927)
	[0.1454]	[0.2190]	[0.1151]	[0.1853]
t-statistic ($b = 1$)	(−3.752)	(−3.899)	(−2.582)	(−2.158)
	−[3.475]	[−3.860]	[−2.939]	[−2.244]
R^2	0.026	0.036	0.060	0.061
Adj. R^2	−0.008	−0.000	0.022	0.022
AC(ε)	0.028	0.028	−0.119	−0.118
N. obs	714	714	714	714

September 1991 to December 1992, Jordan and Jordan (1997) find that the coefficient from a regression of the difference in actual and reference prices for a bond on its total future specialness is significantly lower than one for some specifications (e.g., their benchmark linear interpolation specification in Table V of their paper). Similarly, Goldreich, Hanke, and Nath (2005) use a longer sample of 2-year notes from November 1995 to November 2000 but find that, when regressing the yield difference between on-the-run and off-the-run notes on future specialness and measures of future liquidity, "the coefficient on future specialness is *statistically indistinguishable from zero*" (p. 24, the emphasis is ours).[6] Our results, which are based on a much longer sample, confirm this earlier evidence and suggest that, in contrast to the null hypothesis, variation in the repo premium does not completely account for variation in the price premium.

[6] Goldreich, Hanke, and Nath (2005) argue that their finding may be due in part to measurement error, since the ex post realized repo rates are noisy observations of their ex ante expected values. In our specification, we regress the ex post realized cash premium C on the ex post realized repo premium R. The realizations of both variables are noisy observations of their ex ante expected values, but repo specials are much less variable than the cash premiums, so the concern of a downward bias in our estimates is muted.

Table IV

Liquidity Costs of Short-Selling

This table reports our estimates for the cost of short-selling Treasuries over our sample from November 1995 through July 2009. The first four rows repeat the full-sample estimates from Tables I and II for the cost of short-selling (i.e., $R - C$) and the cash component (i.e., C). The bottom portion of the table uses these estimates to compute the fraction of the total annual liquidity premium that is paid for by short-sellers as

$$\text{Frac. of cost to short sellers} = \delta_{\text{on}} \times \frac{R_{\text{on}} - C_{\text{on}}}{C_{\text{on}} + C_{\text{off 1}}} + \delta_{\text{off 1}} \times \frac{R_{\text{off 1}} - C_{\text{off 1}}}{C_{\text{on}} + C_{\text{off 1}}}.$$

We compute this value using three different scenarios for the fraction of each on-the-run and first off-the-run that are sold short (i.e., δ_{on} and δ_{off1}, respectively) that are based on the estimates in a working paper version of Barclay, Hendershott, and Kotz (2006). Scenario II uses the average estimates from that paper, whereas Scenario I uses more conservative estimates and Scenario III uses more aggressive estimates.

	Maturity: 10 year			Maturity: 5 year		
Estimate for $R_{\text{on}} - C_{\text{on}}$		33 b.p.			28 b.p.	
Estimate for $R_{\text{off1}} - C_{\text{off1}}$		24 b.p.			8 b.p.	
Estimate for C_{on}		94 b.p.			75 b.p.	
Estimate for C_{off1}		29 b.p.			37 b.p.	
Scenario	I	II	III	I	II	III
Frac. of on-the-run shorted δ_{on}	0.50	1.00	1.50	0.50	0.75	1.00
Frac. of off-the-run shorted δ_{off}	0.25	0.50	0.75	0.20	0.40	0.60
Frac. of cost to short sellers	18%	37%	55%	14%	22%	30%

D. Decomposition of the Liquidity Premium

Using the framework developed in Section II, Table IV provides estimates of the total annualized liquidity premium for on-the-run 5- and 10-year Treasuries and the fraction of this amount that is paid by short-sellers. To estimate this fraction, we need to estimate the proportion δ of each security that is borrowed and sold short. A working paper version of Barclay, Hendershott, and Kotz (2006) provides a plot of daily repo volume for on-the-run and first off-the-run Treasuries.[7] From that plot, we estimate that roughly 100% of the outstanding on-the-run 10-year notes are lent into repo agreements, as are around 75% 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 the 10- and 5-year maturities are roughly 50% and 40%, respectively.[8] In Table IV, we calculate the fraction

[7] Barclay, Hendershott, and Kotz (2006) collect daily settlement data from the Fixed Income Clearing Corporation (FICC) over the period January 2001 through November 2002. Their sample consists of transactions from all 127 FICC members including all primary dealers and other financial institutions.

[8] Specifically, Figure 3 in the working paper version of Barclay, Hendershott, and Kotz (2006) suggests that daily repo trading volume for 10-year on-the-run notes ranges from 14 to 18 billion dollars and that for 5-year on-the-run notes ranges from 12 to 17 billion dollars. Daily repo trading

of the total liquidity premium that is paid for by short-sellers as

$$\text{short fraction} = \delta_{\text{on}} \times \frac{R_{\text{on}} - C_{\text{on}}}{C_{\text{on}} + C_{\text{off 1}}} + \delta_{\text{off 1}} \times \frac{R_{\text{off 1}} - C_{\text{off 1}}}{C_{\text{on}} + C_{\text{off 1}}}, \tag{9}$$

where δ_{on} and $\delta_{\text{off 1}}$ are the fractions of each on-the-run and first off-the-run that are loaned, C_{on} and $C_{\text{off 1}}$ are the average cash market returns as estimated from equation (8) for the on-the-run and first off-the-run (both relative to the second off-the-run), and R_{on} and $R_{\text{off 1}}$ are the corresponding repo market costs.

We estimate that the cash premium for on-the-run 10-year Treasuries relative to the second off-the-run is 94 basis points, and the cash premium for the first off-the-run relative to the second off-the-run is 29 basis points. As a result, equation (9) implies that short-sellers pay an average of around 37% of the liquidity premium. Similarly, the cash premium for the on-the-run 5-year is 75 basis points and that for the first off-the-run is 37 basis points, which implies that short-sellers pay around 22% of the premium. Since we do not directly observe repo volume, Table IV also contains estimates with higher and lower values for the fraction δ of outstanding notes that are borrowed and sold short. We find that, even with conservative estimates of δ, short-sellers pay nearly 20% of the liquidity premium in the 10-year note and over 14% of the premium in the 5-year note.

While these estimates of the liquidity premium might appear small at first glance, they are economically important given the outstanding supply of Treasury securities and the leverage available in these markets. Krishnamurthy (2010) reports that the average haircut on short- and long-term Treasury securities has historically been around 2% and 5% to 6%, respectively. Therefore, positions in these securities can be financed with leverage of 15 to 50 times, which can dramatically magnify any liquidity premium.

E. Accounting for Haircuts

Financing markets for most securities include haircuts. A haircut of $H\%$ on a security means that long investors can use the security as collateral to borrow against $(1 - H)\%$ of its value, while short-sellers lend against this amount. As Krishnamurthy (2010) documents, liquid securities tend to have smaller haircuts. In particular, the haircut on generic Treasury securities is typically about 5% and remained relatively constant through the recent subprime mortgage crisis. Unfortunately, we do not have data on haircuts for specific Treasuries. However, given that on-the-run and first off-the-run Treasuries frequently

volume in the first off-the-run for either maturity drops to around 5 billion dollars. We compare these trading volumes to the actual issuance of 5-year and 10-year Treasuries over their sample (adjusted for reissuance) to approximate the fraction of each outstanding on-the-run note that is lent into repo agreements. Moreover, these estimates seem reasonable given the average size of primary dealer positions in Treasuries (around −20 billion for the 10-year notes and around −35 billion for the 5-year notes) relative to their average issuance over our entire sample.

trade at lower repo rates than other Treasury securities, it is reasonable to expect that the haircuts on these securities will also be lower.

A haircut of $H\%$ for the on-the-run Treasury implies that the premium to borrow in equation (8) becomes

$$R_{\text{on},t} = \frac{\text{DUR}_t^{\text{on}}}{\text{DUR}_t^{\text{off2}}} r_{t+1}^{\text{gc}} - r_{t+1}^{\text{on}} + \left[\frac{P_{t+1}^{\text{on}}}{P_t^{\text{on}}} (1 - H) - 1 \right] \left(r_{t+1}^{\text{gc}} - r_{t+1}^{\text{on}} \right). \tag{10}$$

We can use equation (10) to adjust our estimate of the fraction of the liquidity premium paid by short-sellers in equation (9). If we assume that the on-the-run and first off-the-run Treasuries both have a 5% haircut, then in Table IV our estimate of the average portion of the liquidity premium for on-the-run 10-year Treasuries paid for by short-sellers decreases from 37% to 31%. Similarly, our estimate of the average portion of the liquidity premium for on-the-run 5-year Treasuries paid for by short-sellers decreases from 22% to 18%. In short, the haircuts for Treasuries tend to be very small, so they have a minimal effect on our estimate of the decomposition of the price premium.

F. Time Variation in the Premium Paid by Short-Sellers

Having rejected the null hypothesis that short-sellers do not pay a liquidity premium, in this section we investigate whether the time-series variation in the premium paid by short-sellers 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 his inventory, the intermediary often short-sells an on-the-run Treasury with a similar maturity to hedge his 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 the 3-month CP–TBill spread. We also focus special attention on the three main crises to affect fixed income markets during

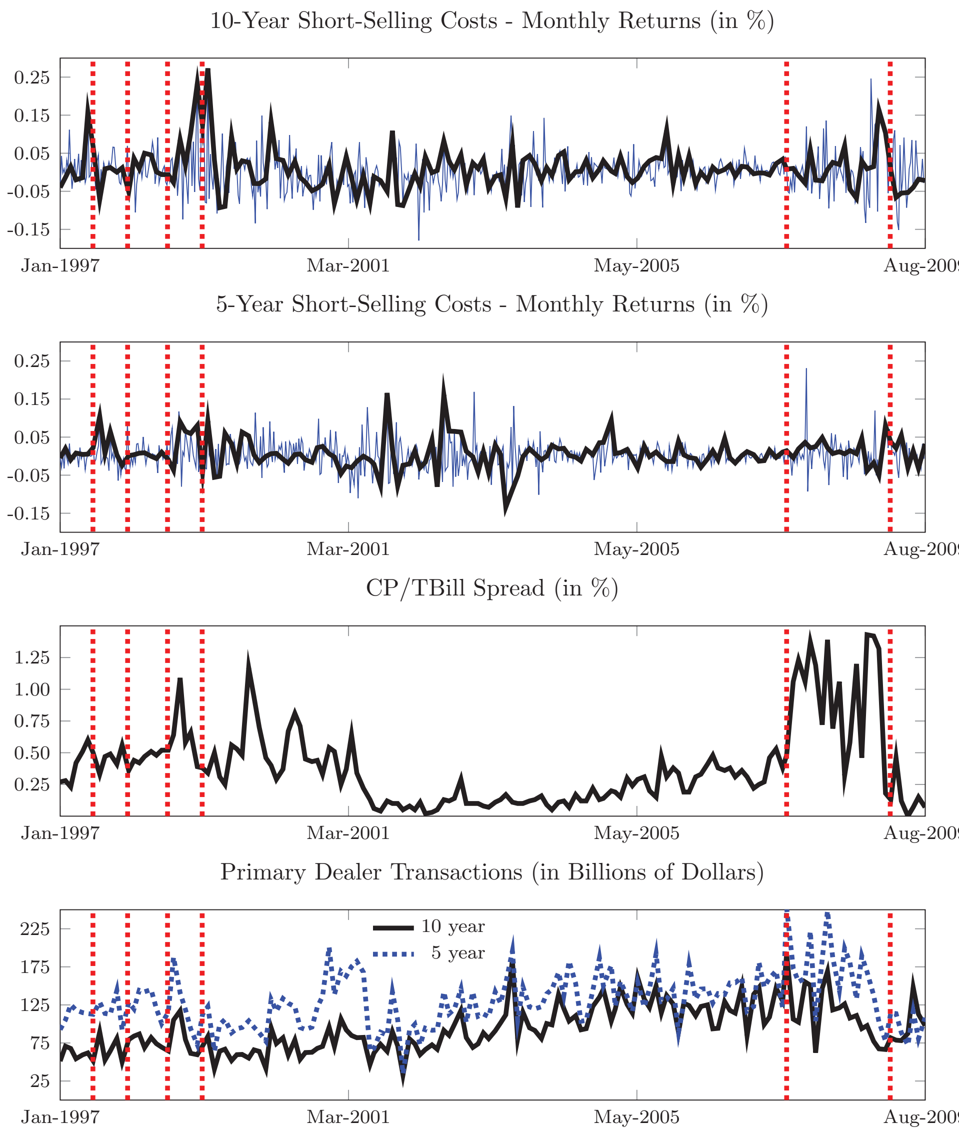

Figure 1. Short-selling costs, CP–TBill spread, and primary dealer positions. This figure plots the cost of shorting 5- and 10-year on-the-run Treasury securities (REPO less CASH) at the weekly (thin gray) and monthly (thick black) frequencies, the spread between 3-month commercial paper and Treasury bills (i.e., CP–T Bill Spread), and primary dealer transactions in the 5- and 10-year Treasuries.

our sample period: the Asian crisis of 1998, the Russian default crisis in 1999, and the recent subprime crisis starting in 2007.

Figure 1 plots the time series estimate of $R - C$ 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 dotted lines. Since the primary dealer transaction data are 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 $R - C$ 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 crisis (delineated with vertical 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 Tables V and VI. We regress the monthly estimate of $R - C$ on (lagged) primary dealer transactions, CP–TBill spread, and an indicator variable for whether the current period is in one of the crises during the sample period. 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 subprime crisis as occurring from August 2007 through January 2009. Table V reports the results for the full sample and Table VI reports separate estimates for the subsamples before and after August 2007.

The empirical evidence is mixed. Consistent with our interpretations, Table V suggests that maturity-specific dealer transactions and CP–TBill spread have incremental explanatory power and are positively related to the cost of shorting. Maturity-specific transactions are statistically more important than transactions at the other maturity for the 10-year notes. Finally, the indicator variable for the crises is positively related to $R - C$, has additional explanatory power for both maturities, and attenuates the coefficient on CP–TBill spread in terms of both magnitude and statistical significance. This result suggests that part of the positive relation between $R - C$ 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.

However, the adjusted R^2s in the full-sample specifications are low, and most of the coefficients are not statistically significant. The estimates in Table VI suggest a partial explanation for these results. For instance, the sign of the coefficients is consistent with our interpretations during the pre-2007 sample, and the adjusted R^2s and the statistical significance of the coefficients are higher than in the full sample. In contrast, during the post-August 2007 subsample, the coefficient for maturity-specific transactions is negative (although not statistically significant) for both maturities and the coefficient for the CP–TBill spread is negative (although, again not statistically significant) for the 5-year maturity. The serial correlation in the residuals for both

Table V

Predicted Liquidity Cost of Short-Selling (Full Sample)

This table reports the results from predictive regressions

$$R_{t+1} - C_{t+1} = \alpha + \beta\,(\text{CP–TB})_t + \gamma_{10}\text{Tr}_t\,(10y) + \gamma_5\text{Tr}_t\,(5y) + \lambda I\,(Crisis)_{t+1} + \varepsilon_{t+1},$$

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–TBill spread, $\text{Tr}_t\,(10y)$ and $\text{Tr}_t\,(5y)$are the weekly primary dealer transactions in Treasury bonds with maturities comparable to the 10- and 5-year Treasuries, respectively, and $I\,(Crisis)_{t+1}$ is an indicator variable for whether the current period is during a 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 subprime crisis (August 2007 through January 2009). The coefficient on each regressor (except the intercept and the indicator) 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^2s for each regression, and OLS standard errors (in round brackets) and Newey–West standard errors with five lags (in square brackets) for each coefficient. The autocorrelation (AC) in the regression residuals and the number of observations are also reported. The single asterisk represents statistical significance at the 10% level.

	Maturity: 10 year					Maturity: 5 year				
Const	−0.0023	−0.0009	−0.0063	−0.0023	−0.0051	0.0010	−0.0019	−0.0020	−0.0023	−0.0005
	(0.0037)	(0.0076)	(0.0071)	(0.0090)	(0.0069)	(0.0025)	(0.0060)	(0.0061)	(0.0060)	(0.0063)
	[0.0039]	[0.0086]	[0.0078]	[0.0104]	[0.0077]	[0.0028]	[0.0062]	[0.0062]	[0.0061]	[0.0067]
CP–TB	0.0047		0.0047	0.0055	0.0023	0.0009		0.0006	0.0007	−0.0015
	(0.0034)		(0.0034)	(0.0039)	(0.0037)	(0.0017)		(0.0017)	(0.0019)	(0.0023)
	[0.0037]		[0.0038]	[0.0042]	[0.0040]	[0.0019]		[0.0019]	[0.0020]	[0.0029]
Tr 10y		0.0014	0.0013	0.0038*	0.0013				0.0006	0.0007
		(0.0021)	(0.0022)	(0.0023)	(0.0021)				(0.0020)	(0.0016)
		[0.0023]	[0.0024]	[0.0020]	[0.0023]				[0.0020]	[0.0016]
Tr 5y				−0.0034			0.0011	0.0009	0.0005	
				(0.0035)			(0.0015)	(0.0015)	(0.0024)	
				[0.0036]			[0.0015]	[0.0015]	[0.0025]	
Crisis					0.0096					0.0086
					(0.0074)					(0.0055)
					[0.0077]					[0.0061]
R^2	0.030	0.003	0.032	0.039	0.044	0.003	0.004	0.005	0.006	0.029
Adj. R^2	0.023	−0.004	0.019	0.019	0.024	−0.004	−0.003	−0.008	−0.015	0.009
AC(ε)	0.109	0.120	0.111	0.108	0.093	−0.096	−0.091	−0.095	−0.096	−0.092
N. obs	152	152	152	152	152	152	152	152	152	152

Table VI

Predicted Liquidity Cost of Short-Selling by Subsample

This table reports the results from predictive regressions

$$R_{t+1} - C_{t+1} = \alpha + \beta\,(\text{CP–TB})_t + \gamma \text{Trans}_t + \varepsilon_{t+1},$$

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–TBill spread, and Trans_t is the weekly primary dealer transactions in Treasury bonds with maturities comparable to on-the-run note considered. 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 2007. The table reports the adjusted R^2s for each regression, and OLS standard errors (in round brackets) and Newey–West standard errors with five lags (in square brackets) for each coefficient. The autocorrelation (AC) in the regression residuals and the number of observations are also reported. Single and double asterisks represent statistical significance at the 10% and 5% level, respectively.

	Maturity: 10 year			Maturity: 5 year		
	Pre-August 2007			Pre-August 2007		
Const	−0.0004	−0.0092	−0.0163**	−0.0027	−0.0066	−0.0107
	(0.0033)	(0.0065)	(0.0067)	(0.0030)	(0.0072)	(0.0081)
	[0.0037]	[0.0080]	[0.0080]	[0.0027]	[0.0080]	[0.0079]
CP–TB	0.0022		0.0032*	0.0027*		0.0027*
	(0.0021)		(0.0020)	(0.0016)		(0.0015)
	[0.0022]		[0.0019]	[0.0014]		[0.0013]
Trans		0.0039*	0.0047**		0.0019	0.0019
		(0.0018)	(0.0017)		(0.0016)	(0.0016)
		[0.0021]	[0.0020]		[0.0018]	[0.0017]
R^2	0.009	0.029	0.047	0.028	0.014	0.042
Adj. R^2	0.001	0.021	0.032	0.020	0.006	0.026
AC(ε)	0.046	0.049	0.046	−0.017	−0.004	−0.017
N. obs	127	127	127	127	127	127
	Post-August 2007			Post-August 2007		
Const	−0.0105	0.0416	0.0277	0.0135*	0.0143	0.0155
	(0.0111)	(0.0360)	(0.0293)	(0.0075)	(0.0157)	(0.0153)
	[0.0082]	[0.0352]	[0.0240]	[0.0068]	[0.0073]	[0.0096]
CP–TB	0.0116		0.0129	−0.0055		−0.0052
	(0.0087)		(0.0084)	(0.0043)		(0.0050)
	[0.0083]		[0.0081]	[0.0048]		[0.0055]
Trans		−0.0111	−0.0124		−0.0030	−0.0009
		(0.0092)	(0.0092)		(0.0043)	(0.0050)
		[0.0094]	[0.0093]		[0.0021]	[0.0032]
R^2	0.072	0.065	0.154	0.066	0.019	0.067
Adj. R^2	0.032	0.025	0.077	0.025	−0.023	−0.018
AC(ε)	0.209	0.200	0.161	−0.345	−0.370	−0.344
N. obs	25	25	25	25	25	25

maturities is higher in magnitude in the later subsample and of different signs, suggesting that not only did the subprime crisis affect the relations between $R - C$ and our explanatory variables, but it did so differently for the 5- and 10-year maturities.

Note that our empirical results are likely affected by the considerable uncertainty in financial markets after August 2007. For example, Krishnamurthy (2010) documents that the prices of many fixed income securities (even those unrelated to subprime mortgages) diverged from fundamental values during this period, and the premium for liquid securities increased dramatically. Our empirical analysis uses ex post realized cash and financing premiums, but the sample is small after August 2007 and it is unclear whether the ex post realizations over this period accurately reflect agents' ex ante expectations. There was also significant intervention by the Federal Reserve in repo markets during this period. As Fleming, Hrung, and Keane (2010) document, on March 11, 2008 the Federal Reserve introduced the Term Securities Lending Facility (TSLF), which allowed dealers to swap less liquid collateral (specifically, agency debt securities, agency mortgage-backed securities, and other investment grade debt securities) for Treasury collateral. The aim of the TSLF was to narrow the spread between the financing rates on Treasuries versus those on non-Treasuries so that dealers could more easily finance their positions in non-Treasury securities. Our empirical analysis compares the financing rates for on-the-run Treasuries versus off-the-run Treasuries, and it is unclear whether the TSLF affected the difference in financing rates between these securities.

IV. Theory

In this section, we present a simple model that illustrates how the presence of binding lending constraints for long investors can lead to price and borrowing premiums in an asset. Our objective is to formalize the intuition we developed in Section II. We model the liquidity price premium as the difference in the price of a single risky asset when it can be used to hedge liquidity (endowment) shocks versus when it cannot. One could instead model the liquidity premium as the difference in prices of two securities with identical payoffs that differ in their transactions costs (e.g., Duffie (1996), Krishnamurthy (2002)) or search frictions (e.g., Vayanos and Weill (2008)). Our model takes a reduced-form approach to convey our basic intuition in a simple, tractable framework.

The relevant features of the model are: (i) some investors have a liquidity-based preference for a long position in the security, whereas others prefer a short position, (ii) the security is in positive net supply, and (iii) a short position in the security must be borrowed. We show that, when the lending constraint for long investors does not strictly bind, there is no borrowing premium and the net short liquidity demand for the asset decreases the price of the asset. In contrast, when the lending constraint for long investors strictly binds, short-sellers pay a borrowing premium and a higher liquidity demand from them leads to an increase in the price of the asset. An increased hedging/liquidity demand from shorts (longs) increases both the price premium and the borrowing premium, but leaves the net premium paid by longs (shorts, respectively) unaffected. Finally, in Section IV.A, we derive explicit expressions for the equilibrium price and borrowing cost in the case of mean-variance preferences, and show that the fraction of the price premium paid by long investors and short-sellers also depends on the relative risk tolerances of each group.

Suppose there is a security with outstanding quantity Q and uncertain payoff V in the next period. There are two types of agents indexed by $i = \{L, S\}$. Each agent has initial wealth W_0, receives an endowment shock $\rho_i V$ next period, and has a utility function U_i over next period's wealth. Agents can use the security to hedge their endowment risk. Without loss of generality, assume that the endowment shocks, $\rho_L \leq 0 \leq \rho_S$, are such that agent L chooses a long position in the security, whereas agent S short-sells it. The price of the security is P. A short-seller must pay $R \geq 0$ to borrow it, while a long investor can lend at most a fraction γ of his position.[9]

The optimization problem for agent L (the long investor) is given by

$$x^L(P - \gamma R, \rho_L) \equiv \arg\max_x \mathbb{E}[U_L(W_L)], \quad \text{where} \quad W_L = x[V - (P - \gamma R)] + \rho_L V + W_0, \tag{11}$$

and the optimization problem for agent S (the short-seller) is given by

$$x^S(R - P, \rho_S) \equiv \arg\max_x \mathbb{E}[U_S(W_S)], \quad \text{where} \quad W_S = -x[V - (R - P)] + \rho_S V + W_0. \tag{12}$$

We assume that the utility functions, U_i, and the distribution of the asset payoff, V, are such that the long investor's demand is downward sloping in his net cost, $P - \gamma R$, and the short-seller's demand is also downward sloping in her net cost, $R - P$, that is,

$$x_1^L \equiv \frac{\partial x^L}{\partial (P - \gamma R)} < 0 \quad \text{and} \quad x_1^S \equiv \frac{\partial x^S}{\partial (R - P)} < 0. \tag{13}$$

We also assume that, all else equal, an increase in ρ_L decreases the optimal long position (since $\rho_L \leq 0$), while an increase in ρ_S increases the optimal short position, that is,

$$x_2^L \equiv \frac{\partial x^L}{\partial \rho_L} < 0 \quad \text{and} \quad x_2^S \equiv \frac{\partial x^S}{\partial \rho_S} > 0. \tag{14}$$

Together, the long and short positions in the security must sum to the outstanding supply, so the cash-market clearing condition is

$$x^L(P - \gamma R, \rho_L) - x^S(R - P, \rho_S) = Q. \tag{15}$$

The market to borrow the security must also clear, which implies that

$$x^S(R - P, \rho_S) \leq \gamma x^L(P - \gamma R, \rho_L), \tag{16}$$

with an equality if $R > 0$. As a natural benchmark for the fundamental value of the asset, we consider the equilibrium price when there are no hedging or

[9] In the decomposition of the price premium presented in Section II, equation (1), the fraction of the outstanding security that is sold short is represented by δ, so that $1 + \delta$ is held in long positions. Therefore, the fraction of long investors' position that is lent to short-sellers is $\gamma = \delta/(1 + \delta)$, or equivalently, $\delta = \gamma/(1 - \gamma)$ and $1 + \delta = 1/(1 - \gamma)$.

liquidity motives for trade (i.e., $\rho_L = 0 = \rho_S$). In the analysis below, we compare the price of the asset when it offers liquidity benefits relative to this benchmark price P_0.

If there is no cost to borrow the security (i.e., $R = 0$), then the equilibrium price P is pinned down by the cash-market clearing condition (15), so that

$$x^L(P, \rho_L) - x^S(-P, \rho_S) = Q. \tag{17}$$

Differentiating both sides of equation (17) with respect to ρ_L and ρ_S implies that

$$-\frac{\partial P}{\partial \rho_L} = \frac{x_2^L}{x_1^L + x_1^S} > 0 \quad \text{and} \quad \frac{\partial P}{\partial \rho_S} = \frac{x_2^S}{x_1^L + x_1^S} < 0. \tag{18}$$

Therefore, a larger hedging/liquidity demand from long investors (i.e., a more negative ρ_L) increases the price of the security, P, whereas a larger demand from short-sellers (i.e., a more positive ρ_S) decreases it. Finally, note that, when there is no cost to borrow the security (i.e., $R = 0$), the net cost paid by long investors (i.e., P) is the exact opposite of the net cost paid by short-sellers (i.e., $-P$).

There is a positive cost to borrow the security (i.e., $R > 0$) if and only if the cash-market clearing price P in equation (17) fails to clear the borrowing market, that is,

$$x^S(-P, \rho_S) > \gamma x^L(P, \rho_L). \tag{19}$$

A positive borrowing premium relaxes the inequality in (19) since it leads short-sellers to demand less (i.e., $x^S(-P, \rho_S) > x^S(R - P, \rho_S)$).[10] Moreover, if there is a positive cost to borrow the security, then a long investor benefits from lending the maximum possible amount of his position and so the equality in equation (16) must bind, that is, $x^S(R - P, \rho_S) = \gamma x^L(P - \gamma R, \rho_L)$. Combining this observation with the cash-market clearing condition from equation (15) implies that the equilibrium positions of the long investor and short-seller are given by

$$x^L(P - \gamma R, \rho_L) = \frac{Q}{1-\gamma} \quad \text{and} \quad x^S(R - P, \rho_S) = \frac{\gamma Q}{1-\gamma}. \tag{20}$$

Denote the price premium, C, as the price P when the asset offers liquidity benefits minus its fundamental value P_0 (i.e., $C = P - P_0$) and note that R is the borrowing premium (since the borrowing cost is zero when there are no liquidity benefits from holding the asset). Recall that, by definition, P_0 does not depend on ρ_L or ρ_S. Therefore, the above expressions imply that the long investor's net premium, $C - \gamma R$, depends on his liquidity shock ρ_L, but not the

[10] In this case, the equilibrium price P and borrowing cost R also lead long investors to hold smaller positions since the net premium paid by them (i.e., $P - \gamma R$) in equilibrium is greater than the equilibrium price in (17) without a positive borrowing cost.

liquidity shock ρ_S of the short-seller, since

$$-\frac{\partial\,(C-\gamma R)}{\partial \rho_L} = -\frac{\partial\,(P-\gamma R)}{\partial \rho_L} = \frac{x_2^L}{x_1^L} > 0$$

and
$$\frac{\partial\,(C-\gamma R)}{\partial \rho_S} = \frac{\partial\,(P-\gamma R)}{\partial \rho_S} = 0\,. \tag{21}$$

Similarly, the short-seller's net premium $R - C$ depends on her liquidity shock ρ_S, but not ρ_L, since[11]

$$\frac{\partial\,(R-C)}{\partial \rho_S} = \frac{\partial\,(R-P)}{\partial \rho_S} = -\frac{x_2^S}{x_1^S} > 0 \quad \text{and} \quad \frac{\partial\,(R-C)}{\partial \rho_L} = \frac{\partial\,(R-P)}{\partial \rho_L} = 0. \tag{22}$$

In contrast to the case in which the cost to borrow is zero, the above results imply that an increase in liquidity demand from either short-sellers (i.e., more positive ρ_S) or long investors (i.e., more negative ρ_L) increases the price premium C, since one can always express $C = \frac{C-\gamma R}{1-\gamma} + \frac{\gamma[R-C]}{1-\gamma}$. Similarly, an increase in liquidity demand from either agent also increases the borrowing premium R, since $R = \frac{C-\gamma R}{1-\gamma} + \frac{R-C}{1-\gamma}$. As in our decomposition, in order to quantify the contribution of longs and shorts to the price premium, one must characterize the net premium paid by each group of investors.

A. Mean Variance Preferences

As a specific example that satisfies conditions (13) and (14) and yields a closed-form equilibrium solution, suppose that agent $i = \{L, S\}$ has mean-variance preferences over next period's wealth with risk tolerance τ_i, and that the payoff V on the security is such that

$$\mathbb{E}[V] = m_V \quad \text{and} \quad \text{var}[V] = \sigma_V^2. \tag{23}$$

These preferences imply that the optimal demands for the long and short investors are[12]

$$x^L = \frac{\tau_L}{\sigma_V^2}[m_V - (P-\gamma R)] - \rho_L \quad \text{and} \quad x^S = -\frac{\tau_S}{\sigma_V^2}[m_V + (R-P)] + \rho_S. \tag{24}$$

[11] The expressions in (21) and (22) follow from differentiating (20) with respect to ρ_L and ρ_S, since

$$\frac{\partial x^L}{\partial \rho_L} = x_2^L + x_1^L\frac{\partial(P-\gamma R)}{\partial \rho_L} = 0, \qquad \frac{\partial x^L}{\partial \rho_S} = x_1^L\frac{\partial(P-\gamma R)}{\partial \rho_S} = 0,$$
$$\frac{\partial x^S}{\partial \rho_S} = x_2^S + x_1^S\frac{\partial(R-P)}{\partial \rho_S} = 0, \quad \text{and} \quad \frac{\partial x^S}{\partial \rho_L} = x_1^S\frac{\partial(R-P)}{\partial \rho_S} = 0.$$

[12] In this specific example, the equilibrium positions of the long investor and short-seller are

$$x^L = \frac{\tau_L}{\tau_L+\tau_S}Q + \frac{\tau_L\rho_S - \tau_S\rho_L}{\tau_L+\tau_S} - \left[\frac{\tau_L\rho_S-\tau_S\rho_L}{\tau_L+\tau_S} - \frac{\gamma\tau_L+\tau_S}{\tau_L+\tau_S}\left(\frac{Q}{1-\gamma}\right)\right]^+, \quad \text{and}$$
$$x^S = -\frac{\tau_S}{\tau_L+\tau_S}Q + \frac{\tau_L\rho_S - \tau_S\rho_L}{\tau_L+\tau_S} - \left[\frac{\tau_L\rho_S-\tau_S\rho_L}{\tau_L+\tau_S} - \frac{\gamma\tau_L+\tau_S}{\tau_L+\tau_S}\left(\frac{Q}{1-\gamma}\right)\right]^+.$$

In this case, the equilibrium price is

$$\begin{aligned} P &= \frac{P-\gamma R}{1-\gamma} + \frac{\gamma(R-P)}{1-\gamma}, \\ &= \underbrace{m_V - \frac{\sigma_V^2}{\tau_L+\tau_S} Q}_{P_0} - \frac{\sigma_V^2}{\tau_L+\tau_S}(\rho_L+\rho_S) \\ &\quad + \frac{\sigma_V^2}{1-\gamma}\left(\frac{1}{\tau_L}+\frac{\gamma}{\tau_S}\right)\left[\frac{\tau_L\rho_S-\tau_S\rho_L}{\tau_L+\tau_S} - \frac{\gamma\tau_L+\tau_S}{\tau_L+\tau_S}\left(\frac{Q}{1-\gamma}\right)\right]^+, \end{aligned} \tag{25}$$

and the borrowing premium is

$$\begin{aligned} R &= \frac{P-\gamma R}{1-\gamma} + \frac{R-P}{1-\gamma} \\ &= \frac{\sigma_V^2}{1-\gamma}\left(\frac{1}{\tau_L}+\frac{1}{\tau_S}\right)\left[\frac{\tau_L\rho_S-\tau_S\rho_L}{\tau_L+\tau_S} - \frac{\gamma\tau_L+\tau_S}{\tau_L+\tau_S}\left(\frac{Q}{1-\gamma}\right)\right]^+. \end{aligned} \tag{26}$$

In particular, there is a strictly positive premium to borrow the security, that is, $R > 0$, if and only if

$$\gamma < \frac{\tau_L\rho_S - \tau_S(Q+\rho_L)}{\tau_L(Q+\rho_S) - \tau_S\rho_L} < 1. \tag{27}$$

Note that agents' risk aversion affects the proportion of the price premium that they pay. When either group of investors is risk-neutral, the benchmark price of the asset is given by the expected value, that is, $P_0 = m_V$. Inverting the short-seller's demand function in equation (24), we can see that, if the marginal short-seller is risk-neutral (i.e., $\tau_S = \infty$), then her net premium $R - C = R - (P - P_0)$ is zero. Similarly, if the marginal long investor is risk-neutral (i.e., $\tau_L = \infty$), then his inverse demand function implies that the net premium paid by longs is zero, that is, $C - \gamma R = 0$.

For simplicity, we exogenously impose the restriction that long investors can lend at most a fraction γ of their position to characterize the equilibrium. Existing models pin down the equilibrium either by exogenously specifying γ or by making additional assumptions about investor preferences or trading constraints. For instance, Duffie (1996) and Krishnamurthy (2002) assume the existence of some unconstrained risk-neutral investors who can lend their entire long position to short-sellers. This assumption implies that short-sellers do not pay a liquidity premium in equilibrium (i.e., $R - C = 0$), since otherwise the unconstrained long investors would profit from the opposite side of the trade. 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. One may also be able to endogenize γ by imposing a cost of lending for long investors, or by endowing longs with market power in the lending market.

V. Conclusions

In contrast to standard intuition, we argue that longs are not solely responsible for the price premium on liquid securities, as short-sellers may also contribute to this premium by paying more to borrow a liquid security than they expect to recover by selling it. Since they must deliver the same security they borrow when closing out their positions, short-sellers may be willing to pay a premium for positions in a liquid security because they value the ability to repurchase this security more easily in the future. We provide a decomposition of the price premium on a liquid security into the net premiums paid by long investors and short-sellers, and characterize this decomposition in terms of cash prices, borrowing fees, and the volume of the outstanding supply that is sold short. We also show that this decomposition is empirically relevant. Over our sample period from November 1995 through July 2009, we estimate that short-sellers were responsible for a substantial fraction of the liquidity premium for on-the-run Treasury notes relative to their off-the-run counterparts.

Our results highlight that, to understand what determines the liquidity premium in any security, it is not enough to focus on the price premium alone. Instead, one must explain how investor preferences and constraints jointly determine the cash premium, the borrowing or financing premium, and the volume of outstanding supply sold short in equilibrium.

Initial submission: July 7, 2010; Final version received: April 3, 2012
Editor: Campbell Harvey

REFERENCES

Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223–249.

Amihud, Yakov, and Haim Mendelson, 1991, Liquidity, maturity, and the yields on U.S. Treasury securities, *Journal of Finance* 46, 1411–1425.

Barclay, Michael J., Terrence Hendershott, and Kenneth Kotz, 2006, Automation versus intermediation: Evidence from treasuries going off the run, *Journal of Finance* 61, 2395–2414.

Brandt, Michael W., and Kenneth A. Kavajecz, 2004, Price discovery in the U.S. Treasury market: The impact of orderflow and liquidity on the yield curve, *Journal of Finance* 59, 2623–2654.

Cornell, Bradford, and Alan C. Shapiro, 1989, The mispricing of U.S. Treasury bonds: A case study, *Review of Financial Studies* 2, 297–310.

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

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

Fleming, Michael J., Warren B. Hrung, and Frank M. Keane, 2010, Repo market effects of the term securities lending facility, *American Economic Review* 100, 591–596.

Goldreich, David, Bernd Hanke, and Pernendu Nath, 2005, The price of future liquidity: Time-varying liquidity in the U.S. Treasury market, *Review of Finance* 9, 1–32.

Gorton, Gary, and Andrew Metrick, 2012, Securitized banking and the run on repo, *Journal of Financial Economics* 104, 435–451.

Graveline, Jeremy J., and Matthew R. McBrady, 2011, Who makes on-the-run treasuries special? *Journal of Financial Intermediation* 20, 620–632.

Green, T. Clifton, 2004, Economic news and the impact of trading on bond prices, *Journal of Finance* 59, 1201–1234.

Jordan, Bradford D., and Susan D. Jordan, 1997, Special repo rates: An empirical analysis, *Journal of Finance* 52, 2051–2072.

Kamara, Avraham, 1994, Liquidity, taxes, and short-term Treasury yields, *Journal of Financial and Quantitative Analysis* 29, 403–417.

Krishnamurthy, Arvind, 2002, The bond/old-bond spread, *Journal of Financial Economics* 66, 463–506.

Krishnamurthy, Arvind, 2010, How debt markets have malfunctioned in the crisis, *Journal of Economic Perspectives* 24, 3–28.

Li, Haitao, Junbo Wang, Chunchi Wu, and Yan He, 2009, Are liquidity and information risks priced in the Treasury bond market? *Journal of Finance* 64, 467–503.

Longstaff, Francis A., 2004, The flight-to-liquidity premium in U.S. Treasury bond prices, *Journal of Business* 77, 511-526.

Longstaff, Francis A., Sanjay Mithal, and Eric Neis, 2005, Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market, *Journal of Finance* 60, 2213–2253.

Nyborg, Kjell G., and Ilya A. Strebulaev, 2004, Multiple unit auctions and short squeezes, *Review of Financial Studies* 17, 545.

Nyborg, Kjell G., and Suresh Sundaresan, 1996, Discriminatory versus uniform Treasury auctions: Evidence from when-issued transactions, *Journal of Financial Economics* 42, 63–104.

Pasquariello, Paolo, and Clara Vega, 2009, The on-the-run liquidity phenomenon, *Journal of Financial Economics* 92, 1–24.

Strebulaev, Ilya A., 2002, Liquidity and asset pricing: Evidence from the U.S. Treasury securities market, Working paper, Stanford University.

Vayanos, Dimitri, and Pierre-Olivier Weill, 2008, A search-based theory of the on-the-run phenomenon, *Journal of Finance* 63, 1361–1398.

Warga, Arthur, 1992, Bond returns, liquidity, and missing data, *Journal of Financial and Quantitative Analysis* 27, 605–617.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web site:

Appendix S1: Internet Appendix