Hedge funds and the Treasury cash-futures disconnect*

Daniel Barth[1] and R. Jay Kahn[2]

[1]Board of Governors of the Federal Reserve System
[2]Office of Financial Research, U.S. Department of the Treasury

First version: April 1, 2021
This version: July 15, 2021

Abstract

We document the rise and fall of an arbitrage trade among hedge funds known as the Treasury cash-futures basis trade, which exploited a disconnect between cash and futures prices of Treasuries. This disconnect between cash and futures prices is highly persistent, widens during financial crises, and is correlated with repo spreads, margin requirements, and dealer inventories. Relying on a variety of regulatory and public data sets, we show that this trade involved over $2 trillion in daily gross exposures, half of all hedge fund Treasury positions, and a quarter of dealers' repo lending. We present a model of the trade as liquidity transformation, where hedge funds hold illiquid notes funded in liquid repo and futures markets, but have a limited ability to provide liquidity because of risks from margin and repo financing. In March 2020, these risks materialized, and we find hedge funds involved in the trade sold over $100 billion in Treasuries. Because of the limits to hedge funds' ability to engage in liquidity transformation, our results underscore the risks that the current reliance of Treasury market on non-bank financial actors could create going forward.

Keywords: Treasuries, repo, futures, basis trade, hedge fund, securities dealers, liquidity
JEL Codes: E43, G12, G13, G23

*Daniel Barth, Board of Governors of the Federal Reserve System (daniel.j.barth@frb.gov), and R. Jay Kahn, Office of Financial Research (robert.kahn@ofr.treasury.gov). Views and opinions are those of the authors and do not necessarily represent the views of the Office of Financial Research, the U.S. Department of the Treasury, or the Board of Governors of the Federal Reserve System. The authors thank Darrell Duffie, Prabhala Nagpurnanand, Zhaogang Song, Annette Vissing-Jorgenson, Haoxiang Zhu and seminar particpants at the Midwest Finance Assocation Conference, NBER Asset Pricing Seminar, University of Maryland Smith School of Business, Syracuse University Whitman School, and Federal Reserve Bank of Atlanta for their comments. Additionally, we thank Ron Alquist, Amanda Buckley, Katherine Gleason, Maryann Haggerty, Mathias Kruttli, Matthew McCormick, Sriram Rajan, Stacey Schreft, and Ram Yamarthy for their comments and assistance. Portions of this research circulated as OFR Brief no. 20-01.

1 Introduction

During the first week of March 2020, the market for U.S. Treasury securities — the safest and most liquid asset market in the world — began to experience stress. By March 11, Treasury markets faced unprecedented disruptions: bid-ask spreads, particularly on longer maturity bonds such as the 30-year, widened to unseen levels; repurchase agreement (repo) rates on Treasury collateral skyrocketed; and various arbitrage spreads diverged. It is difficult to overstate the importance of Treasury markets in the global financial system, and instability there would reverberate to every corner of financial markets. The Federal Reserve quickly stepped in, dramatically expanding purchases of Treasury securities from dealers and offering unlimited repo and reverse repo facilities on Treasury collateral.[1] By the end of March, turmoil in Treasury markets had largely subsided and market functioning returned to normal.

The focus soon turned to a post-mortem on how the U.S. Treasury market could have experienced such severe disruptions. One suspect is the exit of hedge funds from the Treasury cash-futures "basis trade."[2] Treasury cash and futures prices are related through an arbitrage relationship:

$$P_{t,\tau} = \sum_{s=t}^{T} B_{t,s} c_s + B_{t,T} F_{t,\tau,T}. \tag{1}$$

The price of the bond in the cash market ($P_{t,\tau}$) must be equal to the present values of the coupon payments ($B_{t,s}c_s$) plus the futures price ($B_{t,T}F_{t,\tau,T}$).[3] When the futures price is too high relative to the cash price, arbitrageurs can go "long the basis" by buying the cash bond and shorting the futures. In the absence of frictions, the basis trade would constitute pure riskless arbitrage, and the price of the bond should converge to the futures price at a rate equal to the yield on Treasury bills.

In this paper, we examine the disconnect between Treasury cash and futures prices from their no-arbitrage benchmark, including the large arbitrage activity this disconnect attracted in the lead-up to the March 2020 Treasury market episode, and the consequences of this activity for Treasury market stability. We show that the cash-futures disconnect is in part explained by limits to arbitrage, which arise because arbitrageurs are unable to borrow at the risk-free rate and must instead borrow in the repo market (using the long Treasury note position as collateral), and the short futures positions requires margin. When arbitrage activity is significant, as it was prior to March 2020, these limits to arbitrage create systemic vulnerabilities in Treasury markets. However, we

[1] While these weren't the only interventions, they were likely the most important. The Fed also excluded Treasuries from the Supplemental Leverage Ratio and Enhanced Supplementary Leverage Ratio.

[2] See, for example, Schrimpf et al. (2020) and Duffie (2020).

[3] Specifically, $P_{t,\tau}$ is the price of a government coupon security at time t that matures at time τ, $B_{t,s}$ is the price of a zero-coupon government security maturing at time s, c_s are coupon rates at time s, and $F_{t,\tau,T}$ is the invoice price for bond futures contract agreed to at time t and delivering at time T.

conclude that it was likely sales by real money investors, such as foreign official accounts and mutual funds, rather than hedge funds, that instigated the initial wave of disruptions in March 2020. Nonetheless, we also argue that hedge funds were likely poised to accelerate and worsen stress had the Federal Reserve not intervened.

The rise of hedge fund Treasury market participation that began in early 2018 constituted a major shift toward a heavier reliance on market-based finance. Hedge funds effectively took the role of warehouses of Treasuries, storing them on their balance sheets on behalf of holders of long Treasury futures positions and funding them in the repo market. Based on regulatory data collected on the U.S. Securities and Exchange Commission's (SEC) Form PF and the Office of Financial Research's Cleared Repo Collection, we show that this warehousing function was significant. At its peak in late 2019, we estimate the aggregate size of the hedge fund basis trade was between \$400 – \$500 billion, constituting more than 60% of total hedge fund Treasury exposure, more than 70% of hedge fund repo borrowing, and more than 25% of dealers' repo lending. By late 2019, hedge funds also constituted 41% of open interest in Treasury futures according to data from the Commodity Futures Trading Commission (CFTC). The rise of hedge fund basis trade positions reflected their expanded role as providers of traditional liquidity transformation, similar to that provided by banks, but which was exposed to a wider array of risks due to the use of more fragile market based financing. Our results therefore point to the ways in which a reliance on non-bank actors could pose risks to Treasury markets going forward.

We document that deviations from no-arbitrage between cash and futures markets are at times both large and persistent, and are correlated with previous episodes of stress in financial markets, suggesting the importance of limits to arbitrage. Both margin risk on the futures contract and financing risk on repo have the potential to disrupt arbitrage and drive a wedge between cash and futures prices. To understand the consequences of these frictions, we develop an equilibrium model similar in spirit to Brunnermeier and Pedersen (2008), Greenwood and Vayanos (2014), Gromb and Vayanos (2018) and Kondor and Vayanos (2019). It features a mixture of segmented markets and impediments to arbitrage from margin constraints and frictions in the repo market. Dealers and asset managers invest in segmented cash and futures markets. Hedge funds play the role of warehouses of Treasuries, holding illiquid notes on their balance sheet for delivery to asset managers in the futures market, and funding them through liquid repo. Through this warehousing role, hedge funds facilitate risk-sharing between dealers and asset managers. In the event of large sales by preferred-habitat investors, hedge funds are forced by their margin constraints to unwind their trades, exacerbating the direct effects of these sales. The risks of these margin-induced sales is compounded by rollover risk in the repo market. In combination with the difference in liquidity between repo and bills, these risks lead to a disconnect in equilibrium between cash and futures prices of Treasuries.

We provide empirical evidence that supports the model's predictions. First, borrowing costs for hedge funds proxied by the spread between the GCF repo rate and interest rate on excess reserves (IOER) are positively associated with the Treasury cash-futures disconnect. When funding costs are high, arbitrage is costly, and the futures price strays further from the cash price. Next, we show that the amount of Treasuries held on dealer balance sheets, which is inversely related to dealer's willingness to supply funding with Treasury collateral, is also associated with a higher arbitrage spread. Finally, conditional on dealer Treasury exposure and the VIX volatility index, maintenance margin on Treasury futures is positively associated with the basis for the 5-year, 10-year, and Treasury bond securities. The relationship is insignificant for the 2-year note, consistent with the much smaller initial margins on 2-year futures.

We next show that hedge fund activity in the basis trade rose dramatically beginning in early 2018 as the cash Treasury note became increasingly cheap relative to the futures. Between December 2017 and September 2019, total hedge fund Treasury exposure increased from $1.06 trillion to $2.02 trillion. Over the same period, the proportion of gross notional exposure allocated to Treasuries (cash and derivative securities combined) nearly doubled. Also consistent with large basis positions, repo borrowing and lending — which were roughly equal prior to 2018 — diverged, with borrowing exceeding lending by more than $580 billion by September 2019. Hedge funds trading the basis were also highly leveraged; funds we categorize as large basis traders had a mean leverage of 21-to-1, indicating $20 of borrowing for every $1 of investor capital. Large basis traders accounted for much of the Treasury and repo activity in the data. In 2019, large basis traders accounted for 60%–67% of total Treasury exposure and 73%–80% of total repo positions, despite constituting only 8.5% of total non-zero repo observations and 5.9% of total non-zero Treasury observations.

The model suggests that large dealer balance sheet Treasury exposure and the significant size of hedge fund activity in the basis trade left Treasury markets vulnerable to large, unanticipated sales. In the extreme case, the model indicates the potential for a liquidity spiral. We find that sales in Treasury markets by real money investors led to increases in margins on Treasury futures contracts and rising volatility in repo markets. In response, hedge funds appear to have partially unwound their basis positions, reducing short futures held in the 2-year, 5-year, and 10-year contracts from $659 billion to $554 billion between February 18 and March 17, 2020. The reduction in short futures was accompanied by sales of cash Treasuries held on the long side of the trade. We estimate that large basis traders sold between $91 – $105 billion between the end of February and end of March 2020.[4]

Sales of Treasuries by hedge funds likely reflected deteriorating conditions in the basis trade. Yet, these sales may not have yet been large enough to become systemic. We show that the cheapest-to-deliver Treasuries underlying the basis trade continued to trade at a premium, and

[4] These estimates are consistent with Schrimpf et al. (2020) and Barth and Kahn (2020), which also examine the role of the basis trade in March Treasury sales.

that dealers appear to have kept these Treasuries on their balance sheets rather than selling them to the Fed through the Fed's increased purchase program. Each is inconsistent with a fire sale of Treasuries from exits of the basis trade. We view this as evidence of timely intervention by the Federal Reserve. While hedge funds' unwinding of the basis trade had not yet resulted in further liquidity deterioration, the model suggests that had the stress continued, hedge funds would have soon become an important propagation mechanism for stress, accelerating the downturn and possibly instigating a full-blown liquidity spiral. The Fed intervention likely occurred just in time to prevent a much more severe crisis.[5]

Our results contribute to a number of different literatures. Our focus on the consequences of limits to arbitrage for a specific trade relates to the literature on deviations from covered-interest-parity, such as Du et al. (2018) and Avdjiev et al. (2019). Several papers have included the cash-futures basis trade as a component of broader measures of returns to near-arbitrage strategies, such as Boyarchenko et al. (2018) and Du et al. (2019), focusing on the basis as a measure of limits to arbitrage associated with dealers and banks. Fleckenstein and Longstaff (2018) focuses specifically on the cash-futures basis as a measure of dealer inventory constraints in Treasury markets. Our results also indicate that dealers' Treasury inventories are an important driver of returns on the basis trade. However, we are able to demonstrate the key role of non-bank actors and especially hedge funds in the trade. In contrast to Boyarchenko et al. (2018), who examine the effect of the SLR on funding of hedge fund arbitrage trades, we examine a rising disconnect between cash and futures prices and it's relationship to increasing hedge fund leverage and borrowing in repo markets.

Second, our paper highlights the increased importance of non-bank actors and repo markets in Treasury market functioning. While links between the Treasury market and repo market have been previously established, for instance by Singh and Stella (2012), D'Amico et al. (2018), Correa et al. (2020), Afonso et al. (2020), and Infante et al. (2020), we specifically explore the rise of hedge funds as participants in both repo and Treasury markets. We are the first paper to use regulatory DVP repo data, which allows us to examine hedge fund borrowing on a transacion by transaction basis.

Third, relative to the post-mortem on March Treasury illiquidity, we provide evidence on the contribution of hedge fund basis trades to this stress event. Previous studies have noted the importance of hedge fund Treasury sales in March, notably Schrimpf et al. (2020), Duffie (2020) and Barth and Kahn (2020). He et al. (2020) presents a model similar to ours in which hedge funds and dealers play a key role in episodes of illiquidity. However, in their model hedge funds participate in arbitrage between different Treasuries, with somewhat balanced short and long positions across maturities of cash Treasuries. We show that much of hedge fund Treasury exposure, and

[5] We do not claim that rescuing hedge funds was a motivation for the Fed's interventions, only that preventing further stress in the basis trade was one of many consequences of the Fed's policies.

around half of hedge funds sales during March 2020, can be attributed to the basis trade between cash Treasuries and Treasury futures. This highlights the importance of linkages across different markets in the March 2020 Treasury illiquidity episode.

Finally, we contribute to the literature on hedge funds and systemic risk. Since the failure of Long-Term Capital Management (LTCM) in 1994, regulators have recognized the potential for stress at a large hedge fund to have consequences for financial stability. For instance, Chan et al. (2006) examines the implications of hedge fund illiquidity for systemic risk; Ang et al. (2011) examines hedge fund leverage in the wake of the 2008 financial crisis; Boyson et al. (2010) documents contagion in hedge fund returns; Aragon and Strahan (2012) shows that shocks to traders' funding liquidity during the Lehman Brothers' bankruptcy reduced the liquidity of the assets they traded; and Brunnermeier and Nagel (2004) shows that hedge funds sold technology stock prior to the collapse of the dot come bubble. We demonstrate the financial stability vulnerabilities in Treasury markets that result from hedge funds trading the basis.

The paper is organized as follows. Section 2 establishes the basic arbitrage relationship between Treasury cash and futures prices and examines the convergence in prices by the delivery date. Section 3 discusses the frictions that impose limits to arbitrage in the Treasury cash and futures markets and highlights potential risks associated with the trade. Section 4 describes hedge funds' dramatically expanded participation in the basis trade. Section 5 develops a model of limited arbitrage that provides economic content to these empirical findings, and formalizes the risks to Treasury market functioning that arise from cash-futures arbitrage. Section 6 provides empirical evidence in support of the model. Section 7 explores the disruptions in Treasury markets in March and examine what, if any, role the hedge fund basis trade had. Section 8 concludes.

2 Data, Market Structure, and Construction of Prices

The Treasury futures market links the spot price and futures price of Treasuries. In this section, we first describe the structure of the Treasury futures market, with a focus on market details that are important for correctly calculating prices and arbitrage spreads. We then describe the multiple data sources used in our later analyses.

2.1 Structure of the futures market

The Treasury futures market is operated in the United States by the Chicago Board of Trade (CBOT). The details of this market are arcane in part because of the peculiar nature of delivery to settle a contract. Because the structure of this market is not widely discussed in economics, we briefly review some of these details, though we also refer the reader to more in-depth treatments such as Burghardt and Belton (2005).

In a futures contract, the short position agrees to sell (deliver) the security to the long position at a future date, at a price agreed upon today. Treasury futures contracts, unlike other interest rate futures, require physical delivery of an underlying Treasury. Delivery is permitted on any day during the delivery month, with delivery months corresponding to quarter-ends (March, June, September, and December).

The CBOT offers Treasury futures contracts at various maturity points. Not all Treasuries are eligible for delivery into a Treasury futures contract. To keep the contracts suitably liquid, any Treasury among a predefined set of maturities may be delivered to settle the contract. The maturities in the "deliverable set" are based both on the Treasury's original maturity and its residual maturity on the last day of the delivery month. Despite the breadth of deliverable maturities, futures contracts are commonly referred to as 2-year notes, 5-year notes, 10-year notes, 10-year ultra notes, bonds, and ultra bonds. Table 1 provides details on the deliverable sets and terms for each of these Treasury futures.

In an effort to make all Treasuries in the deliverable set for a particular contract comparable, the CBOT establishes "conversion factors". These conversion factors are applied to the Treasury futures price, and are determined by the coupons and remaining maturity of the Treasury security. The conversion factor is the approximate decimal price at which $1 par of a security would trade if it had a six percent yield-to-maturity. Conversion factors are not, however, directly based on market prices. As a result, depending on the futures price, conversion factor, and other details of the underlying bond, at any given time some deliverable Treasuries will be more desirable to settle a short futures position than others. The most desirable Treasury is referred to as the "cheapest-to-deliver" (CTD). As we show later, the cheapest-to-deliver is involved in much of the arbitrage activity in the basis trade, and will be the focus of our analysis in this paper.

While the option to deliver disciplines the prices of Treasury futures, actual delivery is rare.[6] The Chicago Mercantile Exchange reports that only 2.6% of Treasury futures open interest actually result in physical delivery. Instead, by the beginning of the delivery month, most Treasury futures contracts are "rolled" into the next contract by taking a new short position in the next-to-deliver contract, and an offsetting long position in the existing contract.[7] This roll occurs generally before the last two business days of the month prior to the delivery month. Figure 1 shows open interest in Treasury futures for 2019 by option maturity; very little volume remains in the current contract by the first week of the delivery month. In the discussion below, we will generally focus only on the contract with the highest volume for a particular day. However, the *option* to carry these contracts into delivery is ultimately what enforces the arbitrage relationship between cash and futures prices of Treasuries.

[6] Within the delivery month, several options are available for short positions regarding the timing of delivery and the exact Treasury delivered. These options are reviewed in Burghardt and Belton (2005).

[7] This offsetting long is then netted by CBOT and effectively settles the previous short futures position.

2.2 Data Sources and Construction of Prices

We assemble comprehensive data on returns to Treasuries and futures, taking into account the full features of the delivery process. We establish the deliverable set using the rules set by CBOT, and verify this set against Bloomberg.[8] For callable Treasuries, the CBOT treats their residual maturity as determined by their first callable date. We establish conversion factors for these Treasuries from the formula used by the Chicago Mercentile Exchange (CME).[9] For each deliverable Treasury, we use Chicago Research in Securities Prices (CRSP) mid-price data to measure $P_{t,\tau}$.

For comparability to the futures price, one must account for coupon payments on the cash security because the long futures position does not receive any of the intervening coupons. Where possible, we use the realized coupon schedule recorded in CRSP for the calculation of coupons.[10] We discount these coupons using zero-coupon Treasury bill prices. We use the exact bill price paying off on the date of a coupon payment or at the futures delivery whenever possible, and otherwise interpolate using existing Treasury bills.[11] We assume that a bill maturing today would trade at par, and use this as the lower value in interpolation when neccessary. As with the coupon security, we use the mid-price for bills.

To measure the futures invoice price, we use the last trade price for the futures contract in Bloomberg. We then apply the conversion factor for the particular Treasury to this futures contract. Finally, for the purposes of calculating accrued interest, we assume that the first delivery date is when the Treasury will be delivered. In reality, the exact delivery date is an option exercised by the seller. We have tried different delivery assumptions and found they do not have much effect on our results.

Finally, we form our estimate for the cheapest-to-deliver Treasury in the deliverable basket. We provide details of this procedure in the Appendix. In all, we are able to form daily matched futures and Treasury data going back to 1977 for the Treasury bond futures contract, to 1982 for the 10-year contract, to 1988 for the 5-year contract, and to 1990 for the 2-year contract. These dates match the introduction for these futures contracts on the CBOT.

[8] The rules for the deliverable set for Treasury bond futures changed in 2011 with the introduction of the ultra bond contract. Prior to 2011, Treasury bond futures included all Treasury securities with residual maturities greater than 15 years. After 2011, bonds with residual maturities greater than 25 years were removed from the deliverable set for bond futures and moved into the deliverable set for ultra bond futures.

[9] The conversion factor formula changed in February of 2000, and is reflected in our estimates.

[10] For Treasuries that pay coupons past the end date of our CRSP sample, we use the coupon schedule of the Treasury and account for the fact that coupons are generally paid on business days following holidays or weekends.

[11] It is not necessary to rely on more complicated methods such as spline interpolation, because there is always a bill of longer residual maturity than the delivery date on a traded contract.

3 The Treasury Cash-Futures Disconnect: The Frictionless Benchmark

In standard term-structure models, cash and futures prices are closely related; the value of a bond should be equal to the discounted value of coupons on that bond plus the value of a futures contract, as outlined in equation (1). Crucially, this equation does not depend on any assumptions about risk or preferences. This is in contrast to the expectations hypothesis, which would replace the futures price of the bond with the expected future price.[12] When the prices of cash securities and futures diverge, arbitrageurs can trade this difference at a profit, and push prices back into alignment. The only necessary assumptions for equation (1) are that agents are able to borrow and lend freely for any maturity at the risk-free rate, and that the futures contract does not require the posting of margin (we relax both of these assumptions in later sections).

3.1 Convergence of Cash and Futures Prices

We begin by examining the convergence of Treasury cash and futures prices as the delivery date nears. This is necessary condition for the existence of an arbitrage relationship. To do so, we construct a spot-equivalent price for the futures contract equal to the present value of the Treasury to be delivered into the futures contract. This allows us to compare cash and futures prices in present value terms. The synthetic spot price is constructed by subtracting from the price of the cash note the present value of the coupons to be paid prior to the futures delivery date:

$$\tilde{P}_{t,\tau,T} \equiv P_{t,\tau} - \sum_{s=t}^{T} B_{t,s} c_s.$$ [13]

Figure 3 plots $\tilde{P}_{t,\tau,T}$ and the futures price for the cheapest-to-deliver Treasury, averaged over all contracts from 1992 to 2020. We focus on the CTD security because it is most likely to be delivered into the futures contract. Both assets converge to their final values (at the futures delivery date) from below. The upward drift of the futures price represents a rejection of the expectations hypothesis because the futures price is in period (T) dollars, implying that under the expectations hypothesis the futures price should be a random walk around the final price. The cheapest-to-deliver price is always below the futures price in expectation, but rises faster. This is expected because the cash price is a present value.

The convergence of the CTD price to the note price is virtually guaranteed by the delivery date, because on that date the Treasury can be delivered directly into the short contract, receiving the futures price. As a result, cash and futures prices converge not only in expectation but also with near certainty. This is shown in Figure B.1. The equality between futures prices and note prices at

[12] The relationship between equation (1) and the expectations hypothesis is analogous to the relationship between covered and uncovered interest parity.

[13] Here, as above, $P_{t,\tau}$ is the dirty price of the Treasury note – the price after accounting for accrued interest.

delivery establishes the arbitrage relationship.

Convergence to the futures price at delivery does not hold for all deliverable Treasuries. The bottom-right panel of Figure 4 shows that the average price of non-CTD securities in the deliverable set diverges from the futures price as the delivery date nears. This is not simply the artifact of averaging; Figure 4 also shows that cash prices diverge from futures prices for the 2nd, 3rd, and 4th CTDs. This lack of convergence results because a long futures position is virtually guaranteed to *not* receive any bond other than the cheapest-to-deliver, meaning there is no upper bound on the futures contract from non-CTD maturities.

Importantly, the identify of the CTD is highly stable, suggesting the risk associated with incorrectly forecasting the identity of the CTD security at delivery is minimal. By 80 days prior to delivery, there is a more than an 80% chance that the cheapest-to-deliver does not change across contracts, and the 2- and 5-year contracts are mostly above 90%. More details are provided in the internet appendix.

3.2 The Frictionless Benchmark

In the previous section, we documented the convergence of cash and futures prices of Treasuries at the delivery date. However, in the absence of frictions, equation (1) necessitates the tighter restriction that the *rate* of convergence is equal to the risk-free rate.

To see this, Table 4 describes the strategy one would take if the relationship in equation (1) did not hold. We focus on the most common case where the futures contract is overvalued relative to the cash security. For simplicity, the note is assumed to have zero coupons. To ensure each column in Table 4 reflects the full payoff structure of the asset, we assume the arbitrageur is responsible for all cash flows through the note maturity date, τ. First, the arbitrageur buys the cash Treasury note and takes a short position in the futures contract, promising delivery at time T at a (fixed) price $F_{0,\tau,T}$. The note purchase is financed by selling $F_{0,\tau,T}$ of Treasury bills. At time T, the short position in Treasury bills is paid off using the proceeds from delivering the Treasury into the futures contract. Cash flows at time τ, the maturity of the note, are zero because the payoff from the long note position fully covers the short bill position.[14] This implies:

$$P_{0,\tau} = B_{0,T} F_{0,\tau,T},$$

which is the zero-coupon equivalent of equation (1).

Equivalently, we can define the relationship in terms of yields. The *futures implied yield*, also

[14] Alternatively, one can understand cash flows in period τ being zero because the note has been delivered to the long futures position, and all of the arbitrageur's exposures have been effectively cleared.

known as the *implied repo rate* (IRR), is defined as:

$$y^F_{t,\tau,T} = \left(\frac{F_{t,\tau,T}}{P_{t,\tau} - B_{t,s} c_s} \right)^{\frac{252}{T-t}}. \quad (2)$$

The yield-equivalent no-arbitrage condition is therefore $y^F_{t,\tau,T} = r_{t,T}$, which states that the yield-to-delivery on a portfolio that is long the cash Treasury note and short the futures must be equal to the yield on a Treasury bill maturing on the delivery date, $r_{t,T}$. This no-arbitrage condition is empirically supported. Figure 2 plots the 30-day rolling average futures implied yield on the CTD along with the bill yield. The series move almost in lockstep. Variations in the Treasury bill yield closely match the variation in the disconnect between Treasury cash and futures prices.

Nonetheless, at various points throughout the sample, the futures implied yield and bill yield separate. Figure 5 shows that deviations from the no-arbitrage condition generally hover around zero, but that persistent deviations do exist. In particular, during two recent periods — mid 2015 through late 2017, and early 2018 through late 2019 — the cash note was substantially cheaper than implied by the futures price, especially for the two and five year contracts. In the internet appendix we establish the statistical significance and persistence of these deviations. Further, deviations are correlated across different Treasury contracts: the first principal component explains 65% of the daily variation across the first and second to deliver contracts for all maturities. This occurs even though the construction of each portfolio of futures and notes is essentially independent, and in principle these securities are linked only by having similar classes of assets and similar (although not identical) delivery dates. The persistence in deviations from arbitrage points towards the presence of sticky frictions or slow-moving economic fundamentals that give rise to limits to arbitrage.

3.3 The Deliverability Premium

Between January 2010 and January 2020, notional open interest in Treasury futures increased from $430 billion to $1.8 billiion, as detailed in Figure 8. The large volume of futures contracts traded at each maturity, as well as the convergence of the CTD to the futures price at the delivery date, endows the CTD with an additional source of demand compared to otherwise similar off-the-run securities. This additional demand leads to higher prices for the CTD which can be interpreted as a liquidity premium.

We calculate the deliverability premium as $\hat{y}_{t,\tau} - y_{t,\tau}$, where $y_{t,\tau}$ is the actual yield on the deliverable Treasury, and $\hat{y}_{t,\tau}$ is a fitted yield from the spline yield curve which follows Fisher et al. (1995) in constuction.[15] A positive value indicates that the Treasury is overvalued relative to the estimated benchmark.

[15] Our estimation includes on-the-run and deliverable Treasuries. As a result, the error from this exercise is likely to understate the true premium on deliverable Treasuries.

Table 2 reports the average premia on cheapest-to-deliver Treasuries. On average, all premia are larger than zero, and decline for longer maturities. For comparison, we also present the premia for on-the-run Treasuries. It is well known that on-the-run Treasuries receive a liquidity premium relative to off-the-run Treasuries (see for instance Krishnamurthy (2002) and Adrian et al. (2017)).[16] For 2-year deliverables, premia are similar in size to the on-the-run 2-year. At higher maturities, on-the-run Treasuries tend to receive significantly higher premia.

If the CTD price contains a liquidity premium we should expect it to comove with the premium on on-the-run Treasuries. Table 3 regresses the premia on the cheapest-to-deliver on the premia for on-the-run Treasuries. These regressions are conducted on a daily basis, and yields are winsorized at the 0.5% level to protect against general errors in the spline method.[17] In the full sample, from 1992-2020, the on-the-run premium is significant and positive for all but the bond cheapest-to-deliver. The premium on the on-the-run securities also explains a substantial amount of the variation of both the 2-year and 10-year cheapest-to-deliver premia on a daily basis.

The bottom panel of Table 3 restricts the regression to 2016 to 2020, when hedge funds were most active in the basis trade. In this case we find a significant and positive relationship across all contracts, and coefficient sizes are substantially larger. For the 2-year in particular, the estimated coefficient suggests the cheapest-to-deliver premium actually moves more than one-for-one with the on-the-run premium, and the on-the-run premium explains 83% of the variation in the cheapest-to-deliver premium. That these results are stronger following the rise of the basis trade suggests that arbitrage activity generates a substantial liquidity premium for these Treasuries.

The results in this section illustrate two important points about the basis trade. First, arbitrage activity in the basis trade confers a liquidity premium on the underlying Treasuries, indication that on average this trade is providing liquidity to the notes involved. Second, the covariance of the premium with the premium for on-the-run Treasuries suggests that this trade is exposed to the same sources of time-varying liquidty risk that underlie the spread between on-the-run and off-the-run Treasuries.

4 A Model of Limits to Arbitrage and the Cash-Futures Disconnect

The previous section showed that the disconnect between Treasury cash and futures prices is generally associated with the yield on Treasury bills. However, the association isn't perfect, and deviations can be substantial and persist through time. The frictionless benchmark misses two important costs associated with trading the disconnect between cash and futures: (1) arbitrageurs cannot borrow at the risk-free rate and instead must borrow in the repo market, and (2) there are

[16]The trade between on-the-run and off-the-run Treasuries is another popular relative value trade among hedge funds, making them a useful benchmark for comparison.

[17]We have experimented with different levels of winsorization and found little change in our result.

margin requirements on the futures contract. In this section, we develop a model that formalizes the effect of these frictions on the cash-futures disconnect, and which highlights the interconnections that arise from arbitrageur's trading activity. The arbitrage trade between Treasury cash and futures links Treasury cash and futures markets to repo markets, as well as connects arbitrageurs (hedge funds) to dealers and cash providers such as money market funds. The model combines aspects of liquidity spirals driven by margins, as in Brunnermeier and Pedersen (2008), with preferred habitats and limits to arbitrage on dealers, as in Greenwood and Vayanos (2014). The model has three purposes: first, to explain the existence of an equilibrium futures implied yield that is higher than the Treasury bill rate; second, to examine the determinants of hedge funds' role as warehouses of Treasuries; and finally, to assess how the basis trade may exacerbate financial instability.

The model delivers three predictions. First, as a result of constraints on dealers and limits to arbitrage, an equilibrium basis can emerge in which the return on holding a Treasury note to delivery in the futures market is higher than the bill rate. Second, arbitrage activity in the basis trade is larger when Treasuries are more costly to hold, and as demand for futures contracts increases. Third, arbitrageurs are exposed to margin constraints and repo market illiquidity, which in times of large Treasury sales can exacerbate pressure on dealers. This pressure can be directly counteracted through asset purchases by central banks. These features of the model also set the stage for an analysis of the major disruptions in Treasury markets in March 2020.

4.1 Structure of the "Basis Trade" and Relevant Frictions

When arbitrageurs are unable to borrow at the risk free rate, they must instead borrow in the repo market. The arbitrage trade between Treasury cash and futures that uses repo financing to fund the long position is called the "cash-futures basis trade." Figure 12 gives a stylized description of how the basis trade works. First, arbitrageurs (hedge funds) take a short position in the futures contract, agreeing to deliver the Treasury at the delivery date at a price agreed upon today. Hedge funds then purchase the cash Treasury through dealers in the Treasury market, using cash provided to them by money market funds — through dealers — in the repo market. These hedge funds then deliver the Treasury to the futures market on the delivery date, where the Treasury is passed on to asset managers, and repay their repo borrowing with cash from the long position. In this structure, hedge funds act as a temporary warehouse for Treasuries, holding them on their balance sheets for other agents and earning compensation for this service. Table 5 describes the trade in the familiar style of a payoff matrix.

Reliance of the basis trade on repo financing is important for the cash-futures disconnect because the rate on repo financing may differ from the yield on a Treasury bill. One reason is that repo financing has counterparty risk; the likelihood of a private entity such as a dealer or CCP defaulting is substantially higher than the risk of a U.S. government default. The second is that repo financing uses up balance sheet space of dealers who provide cash to the basis trader, as has been emphasized by He et al. (2020). Finally, Treasury bills are generally considered to be more liquid than repo, commanding their own premium (see Lenel et al. (2019)).

Much of the repo financing of the basis trade is likely to be of a shorter tenor. One reason term repo funding may be difficult to obtain is that dealers bear a higher regulatory burden for term repo funding than for overnight funding.[18] This means that over the lifetime of the futures contract, funding must be rolled over multiple times. This exposes the basis trade to financing risk. One potential offsetting benefit of shorter-term repo is that if futures prices and note prices increase in tandem, as would be expected, any increase in margin requirements may be offset by an increase in the value of the underlying collateral. In a term repo matched to the length of the futures contract, increases in the value of the collateral would not necessarily be reflected in the repo leg of the trade.

The second important limit to arbitrage is margin. The importance of margin requirements as a limit to arbitrage is underscored by Brunnermeier and Pedersen (2008). Traders in futures markets must post initial margin at the opening of a contract, and if necessary supply additional cash to stay above the maintenance margin requirement, as set by the CBOT. Whenever the value of an agent's margin account is less than the maintenance margin, a margin call is made, and agents must replenish their margin account or be in default. The CBOT sets initial margins to 110% of the maintenance margin.[19] With these limits to arbitrage in mind, below we develop an equilibrium model of the Treasury cash futures basis trade.

4.2 Environment

There are three periods, $t = 1, 2, 3$. There are two financial assets in positive net supply. The first are Treasury notes, which have fixed quantity, S, mature at time 3, and pay \$1 at maturity. The price of a note at time t is $P_{N,t}$. The second are Treasury bills, which have price $P_{C,t}$ and pay \$1 in period $t + 1$. Bill prices follow an exogenous stochastic process where:

$$\log(P_{C,t+1}) = \log(P_{C,t}) - \frac{\sigma}{2} + \sigma\epsilon_{t+1} \tag{3}$$

with $\epsilon_{t+1} \sim \mathcal{N}(0, 1)$. This process reflects underlying uncertainty about the short rate.

In addition, there are two financial assets in zero net supply: repo and futures. Futures contracts for the Treasury note struck at time $t = 1$ guarantee a price of F_t for the delivery of a Treasury note at time 2. These futures come with an exogenous margin requirement that a fraction, m_t, of

[18] In particular, since money market funds are limited in their ability to lend to dealers in term repo, and dealers have an incentive to match the term of their repo loans and their repo borrowings.

[19] Anecdotally, traders appear to keep excess cash in their margin accounts to avoid margin calls.

contract value must be deposited with the exchange clearing house. Repo markets are open in periods 1 and 2, where agents can borrow at a price $B_{R,t}$ to repay \$1 next period, provided they post Treasury notes as collateral.[20] A haircut of h is applied to these Treasuries.

There are four participants in these markets: *dealers*, *speculators*, *money market funds*, and *hedge funds*. Dealers represent the Treasury market dealers and cash investors from whom basis traders purchase cash Treasuries. Money market funds represent the source of repo loans to basis traders. Speculators represent the demand of asset managers for futures contracts. Finally, hedge funds in the model span these three markets and arbitrage differences between them. We now describe these participants in detail.

4.3 Dealers

Dealers follow a form similar to that in Greenwood and Vayanos (2014). Dealers are born at time $t \in 1, 2$ with an initial wealth W_t^D and an exogenous exposure to notes X_t, and live for one period. This exposure is meant to represent the Treasury exposure dealers and other cash investors in Treasuries carry associated with normal operations as market makers or existing Treasury positions. It can also be interpreted as representing Treasuries not deliverable into the futures market.

Dealers divide their initial wealth W_t^D between holdings of notes $q_{N,t}^D$ and bills $q_{C,t}^D$ in order to solve the mean-variance problem:

$$\max_{q_{N,t}^D,\, q_{C,t}^D} q_{C,t}^D + \mathrm{E}[P_{N,t+1}]q_{N,t}^D - \frac{\phi_D}{2}\mathrm{Var}(P_{N,t+1})(q_{N,t}^D + X_t)^2.$$

subject to

$$W_t^D \geq P_{C,t}q_{C,t}^D + P_{N,t}q_{N,t}^D$$

where ϕ_D is a preference parameter reflecting the disutility of risk. Solving the maximization problem gives dealer demand for Treasury notes:

$$q_{N,t}^D = \left[\frac{\mathrm{E}[P_{N,t+1}] - \frac{P_{N,t}}{P_{C,t}}}{\phi_D \mathrm{Var}(P_{N,t+1})} - X_t\right]_+$$

(provided $P_{N,t}q_{N,t}^D < W_t^D$). Modeling dealers as living for only one period keeps the analysis simple, by suppressing any precautionary motives that would emerge were they truly dynamic, while allowing Treasury prices to respond to high dealer demand. Dealer risk aversion means that the expected return on notes must exceed the return on bills for dealers to have a positive demand for them.

[20] Bills are not allowed to be posted as collateral for repo in the model, which matches their low usage as collateral in the data.

4.4 Speculators

Speculators also live for one period, and are born at time 1. A speculator takes a long position in futures in a quantity $q_{F,1}^S$ and then disposes of the position in the subsequent period. Like dealers, speculators are mean-variance optimizers, solving:

$$\max_{q_{F,1}^S} \left(\mathrm{E}[P_{N,2}] - F_1\right) q_{F,1}^S - \frac{\phi_S}{2}\mathrm{Var}(P_{N,2}) {q_{F,1}^S}^2. \tag{4}$$

Thus the long positions taken by speculators are

$$q_{F,1}^S = \left[\frac{\mathrm{E}[P_{N,1}] - F_1}{\phi_S \mathrm{Var}(P_{N,2})}\right]_+$$

Again, the expected return on a futures contract must be greater than the return on a bill in order for a speculator to hold a positive position. In the case of speculators, we can think of the parameter ϕ_S as representing in part the hedging demand asset managers may have for Treasury futures.

4.5 Money market funds

Money market funds trade off between investing in repo and in bills. This restriction follows the actual regulation of money market funds, which restricts them to investing in short-term instruments. Following a similar form to the literature on that in the pricing of short-term liquid assets (see for instance Krishnamurthy and Vissing-Jorgensen (2012) and Nagel (2016)), we assume that a spread exists between the price of repo and bills due to money market funds' liquidity demand:

$$\frac{P_{C,1}}{B_{RR,1}} - 1 = \psi_R(q_{N,1}^H) \tag{5}$$

Further, we assume that $\psi_R(q_{N,1}^H) = \omega q_{N,1}^H$. The linear form is for convenience, as it greatly simplifies the notation for the increasing price. In general we only need that $\psi_R > 0$ and $\psi_R' > 0$ and our results will follow. Note also that it is not crucial that money market funds maintain this premium, only that it is increasing in the amount of basis trades done by hedge funds.

4.6 Hedge funds

A hedge fund has access to notes which it buys in quantity $q_{N,t}^H$. These notes are deliverable into a futures contract at time 2, with price F_t. For simplicity, we will assume the hedge fund chooses to deliver every Treasury it holds into this futures contract, so that short futures positions are balanced with long cash positions. We assume that hedge funds cannot short bills, in which case they will not want to hold bills due to money market funds' liquidity demand. Instead, the hedge fund can borrow in the repo market at a price of $B_{R,t}$, promising to repay \$1 in the next period.

The cash flows to the hedge fund at time 2 are:

$$W_2 = F_1 q^H_{N,1} - q^H_{R,1} \tag{6}$$

The first term reflects the futures contract agreed to at time 1. The second term reflects the repayment of repo balances the hedge fund accumulated at time 1.

As in Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2008), at time 1 margin requirements for futures and haircuts for repo limit the extent to which wealth can be used to purchase Treasuries, imposing the constraint:

$$W_1 \geq (m+h) q^H_{N,1} \equiv \kappa \text{Var}(P_{N,1}) q^H_{N,1} \tag{7}$$

This constraint reflects that repo lenders and the long futures position will both require a portion of Treasuries held as equity by the fund against the possibility of the fund's default. The second equality is an assumption which encompasses, in a simple form, the general idea that margins are set to ensure the central counterparty against the risk of an adverse price movement.

4.7 Market clearing

Notes are in positive supply, so total quantities outstanding must equal quantities held by agents, while futures are in net zero supply, so quantity demanded must equal quantity supplied at time 1:

$$S = q^D_{N,t} + q^H_{N,t} \qquad \text{and} \qquad q^S_{F,1} = q^H_{N,1} \tag{8}$$

where the second equality reflects our assumption that all hedge fund Treasury note holdings are delivered into the futures market.

4.8 Properties of equilibrium

The equilibrium is derived by working backward. At time 3, notes and bills both pay off \$1 with certainty. Therefore, at time 2, according to the dealers' problem, $P_{C,2} = P_{N,2}$, in which case dealers are willing to hold all notes. With this constraint, all markets will then close at time 2.

Dealers and speculators both have mean-variance preferences. In equilibrium, all notes will either be held by dealers, or their risk will be borne by speculators. In combination with the market clearing constraint, risk sharing between the two agents leads to:

$$S + X_1 = \frac{1}{\text{Var}(P_{N,2})} \left[\frac{\psi_N}{\phi_D} + \frac{\psi_F}{\phi_S} \right] \tag{9}$$

where ψ_i is the relevant spread for notes and futures:

$$\psi_F \equiv \mathrm{E}[P_{N,2}] - F_1 \qquad \text{and} \qquad \psi_N \equiv \mathrm{E}[P_{N,2}] - P_{N,1}/P_{C,1} \tag{10}$$

These quantities represent the marginal compensation speculators and dealers require for an additional dollar of Treasury note exposure. Equation (9), which we call the "risk-sharing line," therefore describes how the marginal compensation for risk must shift as greater shares of notes are allocated to dealers or speculators. This is the downward sloping line in Figure 13. The downward slope is induced by the fact that, for a fixed supply of notes, allocating a greater share of those notes away from dealers and to speculators requires a higher compensation for speculators to take on more risk, and a lower compensation to dealers for their reduced risk.

The cash-futures disconnect in period 1 is exactly described by the difference between these two spreads:

$$F_1 - \frac{P_{N,1}}{P_{C,1}} = \psi_N - \psi_D. \tag{11}$$

Risk sharing between dealers and speculators is facilitated by hedge funds and money market funds, which set the marginal rate of substitution between $\psi_{S,t}$ and $\psi_{D,t}$ in equilibrium by arbitraging between cash and futures prices. We begin by discussing the behavior of prices when hedge funds are not constrained by margin and money market funds have no liquidity demand. In this case, hedge funds act as perfect warehouses for these trades, and marginal rates of substitution are effectively equalized. We then turn to consider the case where hedge funds face margin constraints and repo markets are subject to liquidity preference, and show this provides a wedge in risk-sharing.

4.8.1 Equilibrium without margin constraints or liquidity

In the absence of liquidity preference and margin constraints, dealers and speculators are able to perfectly share risk. For money market funds to be indifferent between bills and repo, $P_{C,1} = B_{RR,t}$. For hedge funds to be willing to hold Treasuries for delivery into the futures market from period 1 to period 2, it must also be the case that:

$$F_1 = \frac{P_{N,1}}{B_{RR,1}}$$

In the absence of any frictions, arbitrage then sets $\psi_F = \psi_N$, so that the marginal cost of Treasury holdings is equalized between dealers and speculators. This equilibrium is described graphically in Figure 13 by where the dashed line ($\psi_N = \psi_F$) intersects the risk-sharing line.

Solving for this intersection leads to:

$$\psi_F = \psi_N = \mathrm{Var}(P_{N,2}) \left(\frac{\phi_S \phi_D}{\phi_S + \phi_D} \right) (S + X_1)$$

This equation reflects the fact that as sales by preferred-habitat investors rise, Treasury prices must fall as dealers and speculators are forced to bear greater risk. At this equilibrium, for a given supply of notes, risk is shared optimally, with:

$$q^D_{N,1} = \frac{\phi_S}{\phi_S + \phi_D}(S + X_1) \qquad \text{and} \qquad q^S_{N,1} = \frac{\phi_D}{\phi_S + \phi_D}(S + X_1)$$

which reflects shares that are inversely proportional to the risk aversions of the speculators and dealers. This closes the frictionless model, and we now turn to a model with margin constraints and repo illiquidity.

4.8.2 Margin constraints and illiquidity

In the presence of margin constraints and illiquidity, a wedge appears in between the futures price and the note price:

$$F_1 - \frac{P_{N,1}}{P_{C,1}} \geq 0 \tag{12}$$

This wedge can be further decomposed:

$$F_1 - \frac{P_{N,1}}{P_{C,1}} = \underbrace{F_1 - \frac{P_{N,1}}{B_{RR,1}}}_{=\mu\kappa\,\mathrm{Var}(P_{N,2})} + \frac{P_{N,1}}{P_{C,1}} \underbrace{\left(\frac{P_{C,1}}{B_{RR,1}} - 1 \right)}_{=\psi_R(q^H_{N,1})} \geq 0 \tag{13}$$

where $\mu \geq 0$ is the shadow price of the margin constraint for the hedge fund. This drives a wedge in dealer and speculator risk-sharing in Equation (11): because hedge funds no longer serve as perfect warehouses for Treasuries, dealers will end up holding more Treasuries in equilibrium, and will have to be compensated more for a marginal dollar of holdings than speculators would have to be compensated.

Rearranging this equation, we can derive an "arbitrage capacity line" that describes how ψ_F and ψ_N are related, taking into account these limits to arbitrage. The margin constraint bounds ψ_F above by $\frac{\phi_S}{\kappa} W^H_{N,1}$, since this is the maximum volume hedge funds can sustain in the basis trade.

$$\psi_N = \left(\frac{1 + \omega\,\mathrm{E}[P_{N,2}]}{1 + \omega \psi_F} \right) \psi_F + \frac{\mu\kappa\,\mathrm{Var}(P_{N,2})}{1 + \omega\,\mathrm{E}[P_{N,2}]} \tag{14}$$

The first term reflects repo illiquidity, while the second term reflects the shadow price of the hedge funds' margin constraint. As long as the price of the futures contract is positive, this first term is greater than 1. The second term reflects the fact that above $\frac{\phi_S}{\kappa} W^H_{N,1}$, ψ_F is fixed, and any additional risk of the bill price changing must be born by dealers, leading to further increases in ψ_N. This line is shown in Figure 13.

Equilibrium in the model occurs where this arbitrage capacity line intersects the risk-sharing line. Note that in the equilibrium with frictions, dealer holdings are always larger than in the frictionless equilibrium, and the note price is always lower. This reflects the limited ability of hedge funds and money market funds to transfer risk to speculators, leaving dealers holding more risk than in the frictionless model. The compensation that money market funds and hedge funds require for the frictions they face results in a positive equilibrium basis, reflected in the fact that the equilibrium occurs above the line of equality between ψ_F and ψ_N.

This section demonstrates two important facts about hedge funds and the cash-futures disconnect. First, a positive cash-futures disconnect can result in a simple model with both repo illiquidity and margin constraints. These frictions prevent hedge funds from fully eliminating discrepancies between cash and futures prices of Treasuries. Second, as a result of these frictions, the warehousing function of hedge funds can be impaired. A positive cash-futures disconnect in the model represents the inability of dealers to offload enough of their interest-rate risk to asset-managers through hedge funds' balance sheets and the repo market.

4.8.3 Comparative statics

While the framework we have set up in this model is quite stylized, it does allow us to explore some of the risks the basis trade is exposed to, and how these risks affect the Treasury market. First, we examine how increases in dealer Treasury exposure, for instance caused by sales by noise traders, affect the market. Increases in X_1 induce a parallel shift in the risk-sharing line, as in the first panel in Figure 14. For small shifts, the basis may increase or decrease depending on the change in the liquidity premium. For larger shifts, the margin constraint on hedge funds binds, and dealers must bear a larger share of the increase in Treasury risk. As a result, the cash price of Treasuries can decrease dramatically. A spike in the cash-futures disconnect will then occur above the level explained by the increase in repo rates as margin requirements prevent hedge funds from arbitrage trades.

The framework can also be used to illustrate the effects of margin constraints on the basis trade and broader Treasury market. Higher margins shift the limit on ψ_F in, as hedge fund capacity to take on basis trades becomes more limited. For small shifts in margins, there is no effect, as the constraint may not be binding both before and after the shift. For larger shifts, margins may become binding. As a result, sales may occur from basis traders, shifting a greater supply to dealers. Cash prices for Treasuries will then fall and the basis will widen. This corresponds to a simple form of the fire-sale effects in standard models of margin constraints such as Brunnermeier and Pedersen (2008). This model is static, but in a dynamic context, the possibility of binding margin constraints in the future could generate a precautionary shift away from the basis trade, so that margin constraints need not be binding today to affect the returns on the trade and hedge funds' ability to pursue it.

Finally, the framework shows how repo market frictions can affect the basis trade. As ω increases, it shifts the arbitrage capacity line out from its initial position. The increase in the cost of repo leads to shifts from speculators bearing risk to dealers bearing risk as arbitrage becomes more costly. The cash-futures disconnect also rises to compensate hedge funds for this increased illiquidity. Again, while in this static model what matters is repo illiquidity today, in a dynamic model hedge funds would also consider the possibility of illiquidity in the future.

In all three of these cases, limits to arbitrage faced by hedge funds either amplify the effects of noise trader sales or have a direct effect on cash prices of Treasuries. During March 2020, as we will show, all three of these risks materialized: real money investors sold Treasuries, margins on futures contracts increased, and repo market illiquidity drove repo rates up. The model therefore illustrates how the imperfect nature of hedge fund warehousing can create or amplify stress in Treasury markets. We will argue below that the Federal Reserve effectively short-circuited what would otherwise have been an amplifying role through its direct purchases of Treasuries and its interventions into the repo market.

5 Empirical Evidence of Limits to Arbitrage in the Treasury Cash-Futures Basis

In this section we provide empirical evidence of the equilibrium relationships predicted by our model. We show that (1) funding costs of arbitrageurs are correlated with the cash-futures disconnect; (2) the quantity of Treasuries held on dealer balance sheets is associated with the cash-futures disconnect; and (3) measures of Treasury volatility are associated with larger deviations from arbitrage.

5.1 Repo Financing and Dealer Exposures

In the model, one primary source of the cash-futures disconnect is the difference between the bill yield and repo rate. We test this by examining the arbitrage spread along with the spread between the GCF Treasury Index and the Interest Rate on Excess Reserves (IOER), which, while few agents have direct access to it, is a safe overnight and unsecured rate, and is therefore comparable to GCF rates.[21] Figure 15 shows that the relationship between repo rates and futures-implied yields is

[21] The GCF Treasury index is not ideal for this exercise because it is an inter-dealer rate: most hedge funds either borrow in DVP sponsored repo or in uncleared bilateral repo markets. While we use data from DVP repo for some sections of this paper, our sample is limited to the beginning of the OFR's collection in 2019. GCF is therefore the

strong. The 2-year and 5-year implied yields display the clearest correlation, particularly during the period 2015 through 2020, with both repo rates and futures-implied yields being relatively high between 2015 and 2017, low from 2017 to 2018, and rising again after 2019, peaking in March 2020. Similar but noisier patterns can be seen for the 10-year note contract and the bond contract. We examine the statistical properties of this relationship in Table 6. The first row shows results from a regression of the arbitrage spread on the spread of repo rates over the IOER from 2010 to 2020. To control for any term premia in the arbitrage spread, we include fixed effects for the distance to delivery of the contract, allowing for an arbitrary pattern over the life of the contract. For all contracts, the GCF spread is a highly significant predictor of the arbitrage spread, with higher GCF rates predicting higher arbitrage spreads. These results are sensible: the yield on Treasury bills does not account for the full cost of carrying a Treasury note on the balance sheet, and as the model shows when the cost of carrying the Treasury to delivery is high, the required return will also be high.

The model also highlights the importance of dealer Treasury exposure. When dealers carry large amounts of Treasuries on their balance sheets, they demand extra compensation to carry more. In the second row of Table 6, we replace the GCF spread with the net coupon Treasury exposure of primary dealers, which we take from the Federal Reserve Bank of New York's Primary Dealer Statistics.[22] Table 6 shows a consistent effect across contracts of dealer Treasury exposure on the cash-futures disconnect. Higher dealer exposure leads to cash Treasuries being discounted relative to futures because rising dealer carry costs cause a greater spread between cash and futures prices. These results are similar to the results found in Fleckenstein and Longstaff (2018).

5.2 Margin

Margin requirements apply equally to long and short Treasury futures positions. This suggests that, unlike repo, high margins are likely to predict high absolute deviations from arbitrage, but not their sign.

It is important to be clear about how margin requirements are likely to affect basis traders. A trader long the basis has a long position in the underlying Treasury in addition to their short position in futures markets. In the absence of haircuts on repo, if Treasury prices move together with futures prices, a margin call on short futures contracts will occur at the same time that the value of long cash Treasuries are increasing. This means that the basis trader can borrow more against their Treasury collateral. It is therefore ambiguous how margin requirements may effect basis traders. In principal, the rise in Treasury prices could exactly offset the rise in the futures price. However, because the counterparty on the futures contract and the repo contract are generally two different agents, this may not always be the case.

closest available rate for which we have a sufficiently long time-series.

[22]These data are available weekly, so we have far fewer observations.

Table 7 shows the results of regressions of the deviations from arbitrage on the level of maintenance margins, the VIX index, and dealer Treasury exposures. Maintenance margins are significant for all contracts except the 2-year, where margins are generally much smaller, and across these contracts higher margins are associated with higher deviations from arbitrage. This is suggestive that margin requirements also play a role in deviations from arbitrage, either through a lower expected return on total capital deployed or through greater margin risk.

Independently from margins, the VIX index also appears to affect arbitrage spreads. Figure 16 shows a time-series plot of the absolute value of cash-futures disconnect along with two measures of financial market volatility. The top panel shows deviations along with the MOVE index, which is a volatility measure specific to Treasury markets. This index is based off the implied volatility of Treasury options. As can be seen, the correlation is quite close following the year 2000. However, because the Treasury option market is based off the price of Treasury futures, it is conceivable that this relationship is purely the result of mispricing in Treasury futures affecting the price of Treasury options. The bottom panel shows the same futures implied yield deviations along with the VIX index, which being based off equity options should not be subject to any direct relationship. Figure 16 shows that periods of high volatility coincide with large deviations from arbitrage. Table 7 shows that this relationship is highly statistically significant for all but the 5-year contract.

Further, Table 7 shows that the association between the VIX and the cash-futures arbitrage spread is statistically significant even after controlling for the level of margin on the futures contract. There are two possibilities for this association. First, the current level of margins is not a basis traders' primary concern, but rather margin risk. The VIX index may capture information about the volatility of future margins and the possibility of margin calls. Second, the VIX index may capture limits to arbitrage on other participants in cash Treasury and futures markets. In many models of limits to arbitrage, including our own, volatility also affects the ability of these agents to close spreads among these assets.

Table 7 presents a contrast with other studies of arbitrage deviations and the repo market. For instance, Du et al. (2018) highlights the effects of quarter-ends in increasing the level of deviations from covered interest rate parity (CIP), and He et al. (2020) focuses on the effects of quarter-ends on repo rates. We do not find similar effects for the cash-futures arbitrage spread. The fact that quarter-ends do not appear to have a consistent effect on arbitrage spreads across contracts highlights the roles of the specific arbitrageurs involved in a trade. In contrast to CIP deviations, we show that the arbitrageurs in the cash-futures basis are largely hedge funds, who are not likely to face the same quarter-end incentives as the banks involved in CIP arbitrage.

5.3 Previous Crises

Previous crises offer additional evidence in support of limits to arbitrage in the disconnect between Treasury cash and futures prices. The clearest evidence may be events in March 2020, which we discuss in Section 7. However, two other events, the demise of Long Term Capital Management(LTCM) and the Lehman Brothers bankruptcy, provide additional case studies.

LTCM had large Treasury bond positions.[23] The top panel of Figure 6 shows that immediately prior to the ruble devaluation, Treasury bond open interest (the gray-shaded area) had been rising. After the ruble devaluation, this open interest began to fall, and at the same time the difference between implied yields and bill rates spiked. The decline in open interest accelerated following LTCM's bankruptcy, and yields spiked even higher. While the available data for this crisis is not nearly as complete as the data avaiable for the March 2020 crisis, Figure 6 is suggestive of the role arbitrageurs such as LTCM may have played in the futures market even during previous bouts of instability. While Treasury futures positions may not have been the ultimate cause of LTCM's demise, large margin calls on their futures positions by Bear Stearns did directly precipitate their bankruptcy. The resolution and unwinding of positions from LTCM may well have destabilized Treasury cash and futures markets.[24]

The Lehman Brothers' bankruptcy provides another example of the relationship between arbitrageur activity and Treasury cash and futures spreads. Following Lehman Brothers' bankruptcy, across note contracts futures implied yields deviated sharply from bill yields. However, the Lehman Brothers' bankruptcy was followed by a significant *undervaluation* of futures prices relative to cash prices. This is shown in the bottom panel of Figure 6. This may have reflected a difference in the nature of the financial crisis. While the LTCM crisis and the March 2020 episode were characterized by a flight to liquidity in the form of cash and cash-like securities such as on-the-run Treasuries, the 2008 financial crisis was characterized by a flight to safety. Counterparty concerns in both the repo market and the futures market may have led investors to generally prefer cash Treasuries where the counterparty was the U.S. government to futures and repo where the counterparties were other financial institutions. The general decline in the availability of repo financing may also have contributed to the undervaluation of futures.

6 Hedge funds and the Treasury cash-futures basis trade

Due to time-variation in the limits to arbitrage, a wider basis may not necessarily signal greater profit opportunities. In this section, we provide evidence that in recent years the cash-futures disconnect represented a profit opportunity for arbitrageurs, based on evidence from hedge fund balance sheets. The amount of capital deployed to the trade was substantial and sufficient to

[23] See, for instace, The President's Working Group on Financial Markets (1999).

[24] For a discussion of the effect of LTCM's failure on the Treasury futures market see Cai (2003).

foster vulnerabilities in Treasury markets, in line with the financial stability risks described in the model, and which would prove important for the stress in March 2020 as well as the regulatory intervention that followed.

6.1 Treasury Market Developments and Shrinking Balance Sheet Capacity

At least two important developments affected Treasury markets beginning in 2018. First, in the wake of the Tax Cuts and Jobs Act passed at the end of 2017, Treasury issuance grew significantly, as shown in the left panel of Figure 7. Around this time, the amount of Treasury securities held on dealer balance sheets also rose significantly, as shown in the right panel of Figure 7. Also at the beginning of 2018, restrictions on bank leverage, specifically the Supplementary Leverage Ratio (SLR) and Enhanced Supplementary Leverage Ratio, came into effect. The SLR and eSLR set a minimum value for the ratio of bank equity capital to leverage exposure, and included Treasury securities and net repo exposures in the denominator without any risk-adjustments for those activities. These simultaneous developments may have affected the elasticity of repo rates with Treasury collateral.

These market developments coincided with an increased demand for long Treasury futures positions by traditional asset managers, such as pension funds and mutual funds, as a means of getting cheap duration exposure without holding Treasuries on balance sheet. Figure 8 shows the increase in aggregate long futures positions held by traditional asset managers in the 2-year, 5-year, and 10-year contracts from the CFTC's Traders in Financial Futures data. Between December 2017 and December 2019, long Treasury futures positions held by asset managers grew from $396 billion to $732 billion, an increase of 85%.

6.2 Hedge Funds Enter the Basis Trade

The increase in long futures demand was met almost entirely by hedge funds. Between December 2017 and December 2019, hedge fund short futures positions grew from $279 billion to $631 billion, an increase of 126% which more than covers the $336 billion increase in long futures held by traditional asset managers.

To determine if hedge funds' rapid increase in short futures positions were associated with the Treasury cash-futures basis trade, we consult confidential filings from the Securities and Exchange Commission's Form PF, the first systematic regulatory collection of private fund data in the United States. Form PF was adopted in 2011 as part of the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. Form PF is filed by investment advisers registered with the SEC who manage at least $150 million in private fund assets. Private fund advisers file annually and report items such as gross and net asset values, monthly returns, total borrowings, investment strategy, investor composition, and their largest counterparties. *Large hedge fund advisers*, those with at least $1.5 billion in assets managed in hedge funds, are required to report this information at a quarterly frequency as well as more detailed information regarding asset class exposures, measures of liquidity, collateral posted, risk metrics, and more, for each of their *qualifying hedge funds* — funds with at least $500 million in net assets.[25] A more thorough summary of Form PF data can be found in Kruttli et al. (2019) and Barth et al. (2020).

Form PF data show that, coincident with the increase in hedge fund short futures positions, between 2018 and 2020 total hedge fund Treasury exposure, which in Form PF includes both cash holdings and derivatives, also increased significantly.[26] The top panel of Figure 9 shows that in December 2014, total hedge fund Treasury exposure, which we define as the sum of short and long exposures, was $851 billion. By the end of 2017, this exposure was $1.06 trillion, and by September 2019 had grown to $2.02 trillion.

Figure 9 provides evidence that much of the growth in hedge fund Treasury positions was the direct result of the basis trade. The growth in short Treasury exposure reported in Form PF aligns closely with the increase in short futures positions from CFTC data. However, the increase in long Treasury exposures on Form PF appears to come almost exclusively from cash securities; while long Treasury exposures were growing quickly on Form PF, long futures positions from CFTC data remained relatively flat. Because the basis trade comprises a long cash position and a short futures position, the combination of Form PF and CFTC data suggest much of the Treasury exposure growth on hedge fund balance sheets in 2018 and 2019 is directly attributable to an increase in the basis trade. In fact, if we assume that every dollar of short futures has a corresponding long position in a cash Treasury note, then 73% of the $960 billion increase in hedge fund Treasury exposures between December 2017 and December 2019 would be attributable to the basis trade.[27]

Additional evidence of a large increase in hedge fund basis trading can be found in the repo positions reported on Form PF. The basis trade requires only the long note position be financed in the repo market, with no corresponding repo lending associated with the short leg of the trade (unlike the on-the-run/off-the-run trade, for instance, which would include both a repo and a reverse repo). This implies an increase in the basis trade would increase repo borrowing but not repo lending. The bottom panel of Figure 9 shows that prior to 2018, hedge fund repo borrow-

[25]Form PF data are confidential. The form itself is publicly available and can be downloaded here: https://www.sec.gov/rules/final/2011/ia-3308-formpf.pdf. For more detail on the history and structure of Form PF, see Flood et al. (2015) and Flood and Monin (2016).

[26]Form PF instructs the reporting of derivatives as follows: "for derivatives (other than options), 'value' means gross notional value; for options, 'value' means delta adjusted notional value; for all other investments and for all borrowings where the reporting fund is the creditor, 'value' means market value or, where there is not a readily available market value, fair value; for borrowings where the reporting fund is the debtor, 'value' means the value you report internally and to current and prospective investors; and Form PF: General Instructions Page 10 for questions 20, 21, 25, 28, and 35, the numerator you use to determine the percentage of net asset value should be measured on the same basis as gross asset value and may result in responses that total more than 100%."

[27]In fact, since conversion factors are generally below 1, a basis trader should have slightly larger than $1 invested in cash Treasuries for every dollar of futures exposure in order to have a fully covered position, meaning that this number is likely an understatement of total gross exposures attributable to the basis trade.

ing and lending were largely matched, and in late 2017 and early 2018, repo lending was actually larger than repo borrowing. Coincident with the increase in Treasury exposure beginning in 2018, repo borrowing increased sharply, from $637 billion in December 2017 to $1.19 trillion in September 2019, while repo lending actually fell slightly from $686 billion to $655 billion over the same period.

In fact, the cash-futures disconnect appears to be a strong predictor of the mismatch between hedge fund repo borrowing and lending. Figure 10 shows hedge funds' aggregate net repo borrowing (borrowing minus lending) closely tracks the futures implied yield. Beginning in 2018, the arbitrage spread for both the two-year and five-year contracts increased significantly, and net repo borrowing increased in tandem. From the middle of 2015 through early 2017, the cash-futures basis was also large, and this was the only other time in the data when net repo borrowing was positive. In a simple linear regression, the two-year futures implied yield explains 33% of the variation in hedge fund net repo borrowing.

To focus on the experience of hedge funds involved in the basis trade, in each period we define a hedge fund as a "large basis trader" if the fund is in the top 50 across all funds in both total Treasury exposure and net repo borrowing in that period. During 2018 and 2019, there were 44 unique funds that meet this criteria. If we assume that the entirety of the dollar difference in repo borrowing and lending for large basis traders in September 2019 results from the cash-futures basis trade, this would imply a total basis position of $505 billion. Of course, this ignores repo haircuts and the price differences between the cash note and the futures contract, which may complicate this estimate. Alternatively, if we assume the total *increase* in hedge fund short futures positions from December 2017 to September 2019 (combining the 2-year, 5-year, and 10-year contracts) is due to the basis trade, this would imply a total basis trade size of $409 billion.

Large basis traders are responsible for much of the aggregate Treasury exposure and repo activity of hedge funds in the Form PF data. In 2019, the fraction of total hedge fund Treasury exposure attributable to large basis traders ranged from 60%–67%. The fraction of total repo positions attributable to large basis traders ranged from 73%–80%. These figures are particularly striking given that large basis traders comprised only 8.5% of total non-zero repo observations in 2019, and only 5.9% of total non-zero Treasury observations. For a sense of the relative size of the trade, hedge funds made up 41% of open interest in Treasury futures at the end of 2019, and the repo positions of basis traders would be roughly a quarter of all repo lending by primary dealers.

The hedge fund basis trade constituted a significant portion of hedge fund Treasury activity, short futures positions, and dealer repo transactions. But the risks associated with the sheer scale of the trade are likely magnified by the leverage that supported it. Figure 3 shows that the cash-futures basis is not particularly large, with an average return of no more than 50 basis points over the course of 100 days. The profitability of the trade would be too low to generate much hedge fund interest unless the trade could be significantly leveraged.

We measure hedge fund financial leverage, also referred to as "balance sheet leverage," as the ratio of gross assets (balance sheet assets) to net assets (equity capital). This is roughly equal to one plus the ratio of total borrowing to total equity capital. Average hedge fund leverage across all funds in the Form PF data in 2019 was 1.95, suggesting $0.95 of borrowing per $1.00 of equity capital. However, the leverage of large basis traders is substantially higher. In 2019, large basis traders had a median financial leverage of 17.6 and mean leverage of 21. That is, for every $1.00 of investor equity capital, the median large basis trader borrowed an additional $16.60. The standard deviation of leverage for this group is 15.02, suggesting the upper tail is significantly higher than the mean. The significant leverage employed in the basis trade is possible because of low haircuts on repo borrowing with Treasury collateral. The maximum available leverage using a particular security as collateral is the inverse of the haircut on that security; a 10% haircut implies a maximum leverage of $10 for every $1 of collateral. Treasuries typically have very low haircuts, often around 3% or less, implying possible leverage ratios of greater than 30. Together, the mean and standard deviation of leverage suggest that some funds would indeed be leveraged more than 30:1.

6.3 Evidence from the repo market

The previous section examined hedge fund basis activity through regulatory data on hedge fund balance sheets. In this section, we use data from the OFR's collection of cleared repo transactions, which allows for a daily, security-specific view of hedge funds' trades. The basis trade involves heavy leverage through repo borrowing, so these data provide insight into hedge funds' Treasury positions and their borrowing costs that would otherwise be unavailable.

Hedge fund repo borrowing is highly concentrated in the two bilateral repo markets in the United States: FICC's DVP Service, which is a cleared bilateral market, and the uncleared bilateral repo market. Hedge funds are a primary source of collateral for dealers in these markets, and the large volumes involved in the basis trade make this supply crucial. Their participation in DVP occurs through the sponsorship service, which allows entities that are not clearing members of FICC to participate in DVP so long as they are sponsored by qualified direct clearing members. We provide more details about this service in the internet appendix.

If hedge funds are using the CTD to trade the cash-futures basis, we would expect them to disproportionately borrow using the cheapest-to-deliver Treasury notes as collateral. Figure 11 shows hedge fund positions in repo by security CUSIP prior to the December 1, 2019 futures delivery date and following that delivery date. The shaded windows in the top panel are the maturity dates of notes eligible for delivery into the December futures contract. Securities just inside this delivery window had significantly more hedge fund repo volume than securities just outside this window. The largest position prior to December 1 was in the 2-year window and was for the Treasury security that was cheapest-to-deliver for this contract. After December 1, the position in the cheapest-to-deliver for the 2-year December contract had diminished considerably, while positions had expanded for deliverables for the March contract, highlighted in gray in the bottom panel. This is consistent with hedge funds maintaining positions in the cheapest-to-deliver for contracts near to delivery.

One weakness of the results based on DVP repo data is that most hedge fund repo activity is in the bilateral uncleared market. However, hedge funds would likely only use the DVP sponsorship service if repo rates were, at a minimum, no worse than the costs in the bilateral market. Otherwise funds would continue to finance their trades solely in the bilateral uncleared market. A standard cost-minimization argument would then suggest that the marginal borrowing rate in the DVP market for hedge funds must be, in expectation, similar to their marginal borrowing costs in the bilateral uncleared market. Thus, while only a limited fraction of hedge fund repo activity occurs in DVP, that activity may be a fair representation of the marginal costs faced by hedge funds in other repo markets.

7 Cash-futures basis trades and March Treasury illiquidity

A key prediction of the model is that during times of stress, heavy activity in the basis trade can serve as an amplifier of the stress caused by sales of Treasuries by other investors. In March 2020, selling pressure by real money investors in the Treasury market brought on by the COVID-19 pandemic led to pressure on dealers' balance sheets. This pressure led to increased Treasury volatility and decreased prices, resulting in more volatile repo rates on Treasury collateral and increased margin requirements. Simultaneously, arbitrage deviations between cash and futures prices widened. Whether the result of lower expected returns on the basis trade, margin calls on the short futures position, or rising financing risks, sales of the basis trade ensued.

While there is little evidence hedge funds were the cause of this stress, our model suggests they were primed to accelerate it. We present suggestive evidence that well-timed actions by the Federal Reserve were essential to minimizing the amplification of Treasury market stress through sales of the basis. Purchases of Treasuries by the Federal Reserve and expansions of their repo facilities eased the impact of sales on dealers, and provided a guarantee of a buyer for cheapest-to-deliver securities. These actions are justified by our model, where repo facilities and purchases by the Federal Reserve can directly offset selling pressure from preferred-habitat investors and can thwart a liquidity spiral. Absent these actions, the impact of basis trade delevaraging on Treasury markets might have been much worse.

7.1 The Onset of Treasury market illiquidity

In early March 2020, Treasury market liquidity decreased. As yields fell, volatility spiked, according to multiple option-implied indexes (see Figure 17). At the same time, bid-ask spreads began to increase, concentrated in off-the-run Treasuries (see Figure 18). Standard spreads associated with liquidity risk, such as the on-the-run/off-the-run spread, also spiked. These indicators are consistent with a general flight to liquidity, with investors selling off-the-run Treasuries and either holding the proceeds as cash or purchasing more-liquid on-the-run Treasuries.

The illiquidity in Treasury markets seems to have been spurred by large sales from foreign and domestic real money investors, particularly foreign central banks and domestic mutual funds. Sales by domestic mutual funds have been highlighted by Pástor and Vorsatz (2020). Less attention has been paid to foreign official accounts. Treasury International Capital System data show that net decreases in foreign Treasury positions were around $257 billion in the month of March, with a decrease of $147 billion in foreign official accounts.[28] In the Internet Appendix, we present evidence that these sales by foreign official accounts, which were likely undertaken to build up dollar buffers for foreign central banks, coincided with the initial stress in Treasury markets.

7.2 Stress in the basis trade

As our model highlights, one effect of these sales by real money investors is to increase the Treasury exposure of dealers. In the model, even without hedge funds to amplify Treasury market stress, a sudden increase in Treasury sales will cause Treasury prices to decrease. As the right panel of Figure 7 shows, leading into March, primary dealers already had elevated exposure to Treasuries, an increase that began in late 2018. Sales in the Treasury market increased this exposure, especially to the shortest and longest residual maturities. Without immediate buyers, these Treasuries remained on dealers' balance sheets and made dealers hesitant to create markets in these off-the-run Treasuries.

Growing Treasury market illiquidity had a detrimental effect on the basis trade. During March, Treasury price movements and volatility led to increases in maintenance margins on Treasury futures. The top panel of Figure 19 examines these increases in maintenance margins around the March stress for a trader short the futures contract. The gray area represents the size of margins, while the blue line is the change in price of a $200,000 notional 2-year contract. When the blue line goes beyond the gray area below (above) zero, the increase in Treasury price exceeds the maintenance margin for the short (long) position. In either case, a trader who held no extra cash in their margin accounts would then face a margin call. The green area in this panel denotes the range of 95% of the daily price movements in 2019. Price movement in late February and early March

[28] Unlike other figures from TIC, these figures, which come from the Major Foreign Holders of Treasuries data, are likely to exclude hedge funds domiciled abroad.

were far larger than these bounds, and beached margins on two occasions. Following the first breach of the 2-year contract on February 28, the CBOT began increasing margins, which reached their peak on March 12, and remained high through early May. This pattern held across Treasury futures contracts. From February 28 to March 16, across Treasury note futures contracts margins rose by more than 30%, while margins on bond futures more than doubled, corresponding to their longer duration.

Further, the variation margin payments on the futures contract were not fully offset by the increase in prices in the long note position. Futures prices rose more quickly than cash prices and remained elevated relative to recent history. While margin breaches do not necessarily correspond to margin calls, increases in maintenance margins increase the effective cost of positions in Treasury futures, and increased margin *risk* reduces the attractivness of the trade. Each will make the cash-futures basis trade less desirable.

Additionally, limited dealer capacity led to rising (relative) repo rates on Treasury collateral, increasing the cost of financing the long position in the basis trade. The bottom panel of Figure 19 shows that the futures implied yield followed the Treasury bill rate until March. However, in early March, the DVP sponsored borrowing rate failed to keep pace with the bill rate. As highlighted in the model, this increased the cost of trading the basis, and the futures-implied yield increased relative to the risk-free rate. Further, as shown in the top panel of Figure 21, the *volatility* of repo rates increased during the first two weeks of March, implying greater risk associated with the basis trade. The disconnect between the bill rate, repo rate, and the futures-implied yield peaked around March 17, and then began to normalize once the Federal Reserve intervened.

In March, the two primary risks of the basis trade were realized. Margins increased, at a minimum increasing the costs of maintaining basis positions, and possibly causing margin calls, while funding through the repo market became more volatile, increasing rollover risk for trades funded in the repo market. As a result, hedge funds began to unwind their basis trade positions.

7.3 The unwind of hedge fund basis trades

In response to the increased costs of trading the basis, hedge fund short futures positions declined significantly in early March, particularly for the 2-year note. Total hedge fund shorts in the 2-year, 5-year, and 10-year contracts declined from $659 billion in face value to $554 billion between February 18 and March 17, 2020, with a decline of more than $71 billion in the 2-year contract.[29] Some portion of this decline preceded March, with shorts declining by $21 billion between February 18 and March 3, which may have represented some foresight of the stress that the potential spread of COVID-19 to the United States could put on Treasury markets.

[29] There are two ways to reduce futures exposure: for contracts maturing in March, hedge funds may have simply not rolled over into the June contract. For contracts maturing after March, hedge funds would need to take on offsetting long positions.

Additional evidence of a partial unwind of the basis trade is found in Form PF data. Between the ends of February 2020 and March 2020, total hedge fund Treasury exposure declined from \$2.19 trillion to \$1.81 trillion. For hedge funds that we classify as large basis traders, long Treasury positions decreased from \$756 billion to \$652 billion, a decrease of \$104 billion and nearly identical to the \$105 billion reduction in short futures positions. Net repo borrowing decreased from \$515 billion to \$424 billion, which under our previous assumption suggests a slightly lower decrease in the basis trade of \$91 billion. Combined, this suggests hedge funds may have sold upwards of \$100 billion in cash Treasuries as a direct result of shrinking their basis trade positions.

While hedge funds sold Treasuries in response worsening conditions in the basis trade, the extent to which these sales impaired market liquidity is unclear. The prices of the CTDs, which hedge funds exiting the basis trade would have been selling, actually *rose* relative to similar non-CTD notes. The left panel of Figure 20 shows the yield premium of the June cheapest-to-deliver for all Treasury contracts over 2020. An increase in the spread indicates a lower yield (equivalently, a higher price) of the CTD relative to the predicted yield. The increase in the CTD's premium coincided with an increase in bid-ask spreads across Treasuries, and spreads were highest on the Treasuries most popular in the basis trade: the 2-year, 5-year, and to a lesser extent 10-year notes. The right panel of Figure 20 offers an alternative view of the size of the premium on the CTD at the onset of March instability. Deliverable maturities are highlighted in gray, while deliverable Treasuries are shown in green, and non-deliverable Treasuries in light blue. The dark blue line shows the fitted values from the spline model for each maturity, and the shaded blue area denotes the 95% confidence interval for the fitted curve. This figure shows that the premium for the 2-year and 5-year notes is unique even among other deliverable Treasuries.

Figure 20 demonstrates that the deliverability premium *increased* during March, inconsistent with the interpretation that the sales from the basis trade damaged market liquidity, but consistent with the liquidity premium carried by the CTD. If selling pressure from hedge funds exiting the basis trade had significantly impacted Treasury liquidity, we would expect the price of the cheapest-to-deliver securities to have fallen relative to comparable securities as dealers accumulated large net exposures to these specific Treasuries. Instead, the data suggest that any selling pressure was offset by the liquidity that the basis trade provides to these securities and the link it establishes to futures markets.[30]

[30] It is possible that in order to keep basis trades open while meeting margin calls, hedge funds may have sold Treasuries other than the cheapest-to-deliver, thus contributing to the lower price of other securities. It is difficult to reject this possibility without more detailed data on hedge funds' Treasury holdings. However, even in this case the willingness of hedge funds to sell other Treasuries to keep their basis trades open would indicate excess demand for the trade.

7.4 Effect of Federal Reserve actions

While hedge fund sales associated with exiting the basis trade may not have caused the initial instability in Treasury markets, prolonged disruptions could have led to even greater sales, eventually eclipsing a tipping point and instigating a liquidity spiral. This mechanism is highlighted in the model developed above.

Timely intervention by the Federal Reserve may have been crucial for limiting broader spillovers from hedge funds exiting the basis trade. Two actions by the Federal Reserve may have been particularly important. First, the the Federal Reserve expanded purchases, announcing on March 15, 2020 their intention to purchase $500 billion over the following months. The Fed also took the unusual action of including the cheapest-to-deliver Treasury across contracts in the set of securities they would purchase. Second, the Federal Reserve expanded their overnight repo facility, announcing on March 16 an afternoon operation with a $500 billion aggregate limit. On March 17, morning and afternoon operations of $500 billion were announced and extended through the week. These were then extended through the month on March 20.

The direct effect of purchases by the Federal Reserve may have been limited. Purchases of deliverables for longer duration securities picked up almost immediately after March 15 (see Figure 22). However, these longer-duration securities made up very little of hedge fund short futures positions. Alternatively, Fed purchases of the more popular 2-year and 5-year cheapest-to-deliver Treasuries were negligible until April. This is consistent with the cheapest-to-deliver being particularly valuable during March — dealers may not have been as eager to sell Treasuries to the Federal Reserve which had a natural source of demand from basis traders. Even so, the inclusion of these CTDs in the Fed's purchases may have made it easier for dealers to wait for the natural source of demand associated with futures to recover.

The repo market intervention was important because counterparty risk leads to an imperfect pass-through of the Federal Funds target rate to repo rates. The top panel of 21 presents the spread of DVP rates over the federal funds target rate. Rates are split into three segments: the sponsored lending segment, which is largely money market funds lending to banks and dealers; the intermember market, which is dealers and banks transacting with each other; and the sponsored borrowing market, which is largely hedge funds. The two black lines show the rates on the Federal Reserve's repo (RP) and overnight reverse-repo facilities (ON-RRP), which since September 2019 the Federal Reserve has used to control rates in the repo market and to enact monetary policy. Beginning in March, sponsored borrowing and lending rates began to rise and display bouts of volatility, and on the morning of March 16 rates skyrocketed to nearly unprecedented levels. Following the longer-term expansion of the repo facility on March 17, the sponsored lending rate fell to the zero lower bound defined by the ON-RRP facility (see Figure 21). The rate on sponsored borrowing largely fell in lockstep, reducing the cost of funding these Treasury positions for hedge funds.

In total, the large and timely intervention by the Fed into both Treasury and repo markets eased building pressure in the system. It is difficult to know exactly which of these actions was most important, in particular because they were mutually reinforcing. Nonetheless, without these steps the unwinding of the basis trade may have further destabilized markets and could have sparked a liquidity spiral. The vulnerabilities of the basis trade exposed during March 2020 highlight a larger issue associated with the migration of activity toward non-bank actors and market based finance. Prior to March, hedge funds hand taken on an unprecedented exposure to Treasury markets through their liquidity transformation role in the basis trade. In prior years, a similar role might have been played by more traditional participants in Treasury markets such as dealers and banks, who have access to larger and more stable pools of capital. The reliance of hedge funds' liquidity transformation on more fragile wholesale finance exposed their positions to the risks involved in this finance, and without Federal Reserve invtervention could well magnify stress during future episodes of instability.

8 Conclusion

The stress in Treasury markets in March 2020 has led to an evaluation of the structure of Treasury markets and their exposure to sudden bouts of illiquidity. Regardless of their direct impact during March, the involvement of hedge funds in the basis trade is a key feature of Treasury markets in recent years, with hedge funds involved playing a role both as a major holder of Treasuries and as a major supplier of collateral to repo markets. The sheer quantity of Treasury securities involved makes understanding the trade important to our picture of Treasury markets, especially as the trade drives a substantial proportion of holdings of Treasuries by hedge funds, a key non-bank actor in Treasury markets.

We show the basis trade became popular in part as a result of a fundamental disconnect between the prices of cash Treasuries and Treasury futures, one that has grown larger in recent years. In a frictionless market, the spreads we demonstrate would not exist. Yet we show these spreads in reality are both fairly large and relatively persistent. Further, we show the size of these spreads are associated with measures of volatility, disconnects between bills and repo, and the Treasury exposure of primary dealers. These facts suggest the importance of limits to arbitrage in cash and futures markets for Treasuries.

The popularity of this trade, as well as the persistent disconnect between cash Treasuries and a replicating portfolio of futures and bills, serves to illustrate more general issues affecting Treasury markets. The role that hedge funds involved in the trade played as a warehouse for Treasuries suggests substantial costs to other actors for holding Treasuries on their balance sheet. Our model illustrates the circumstances under which hedge funds end up playing this role as warehouses.

It also suggests the risks relying on hedge funds in this role could pose as the margin constraints and repo market frictions they face could amplify pressure on dealer balance sheets in a flight to liquidity.

In March 2020, sales by real money investors led to rising volatility in Treasury markets, and corresponding increases in margins and volatility in repo markets. Large sales from hedge funds trading the basis seem to have followed this event. We show some evidence that these sales may have had a smaller effect on dealer balance sheets than might otherwise have been expected, and in fact that dealers attached particular value to these Treasuries during the peak of March stress. However, these facts must be interpreted in the context of a timely and large intervention of the Federal Reserve into Treasury and repo markets. Without that intervention, our model suggests that the amplifying role of hedge fund sales could have exacerbated illiquidity in the Treasury market.

In the context of ongoing discussions of Treasury market reform, policy makers should therefore consider both the potential impact of the basis trade on Treasury market liquidity, and the broader context that allowed these trades to be profitable in the first place. While this broader research project is only at its beginning, our paper points to important links among repo markets, Treasury markets, and futures markets spanned by hedge funds.

References

Adrian, T., Fleming, M., and Vogt, E. (2017). An index of Treasury Market liquidity: 1991-2017. *FRB of NY Staff Report*, (827).

Afonso, G., Cipriani, M., Copeland, A., Kovner, A., Spada, G. L., and Martin, A. (2020). The Market Events of Mid-September 2019. Staff Reports 918, Federal Reserve Bank of New York.

Ang, A., Gorovyy, S., and van Inwegen, G. B. (2011). Hedge fund leverage. *Journal of Financial Economics*, 102(1):102–106.

Aragon, G. O. and Strahan, P. E. (2012). Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy. *Journal of Financial Economics*, 103(3):570–587.

Avdjiev, S., Du, W., Koch, C., and Shin, H. S. (2019). The Dollar, Bank Leverage, and Deviations from Covered Interest Parity. *American Economic Review: Insights*, 1(2).

Barth, D., Joenvaara, J., Kauppila, M., and Wermers, R. (2020). The Hedge Fund Industry is Bigger (and has Performed Better) Than You Think. Working Papers 20-01, Office of Financial Research, US Department of the Treasury.

Barth, D. and Kahn, J. (2020). Basis Trades and Treasury Market Illiquidity. Briefs 20-01, Office of Financial Research, US Department of the Treasury.

Boyarchenko, N., Eisenbach, T. M., Gupta, P., Shachar, O., and Tassel, P. V. (2018). Bank-intermediated arbitrage. Staff Reports 858, Federal Reserve Bank of New York.

Boyson, N. M., Stahel, C. W., and Stulz, R. M. (2010). Hedge fund contagion and liquidity shocks. *Journal of Finance*, 65:1789–1816.

Brunnermeier, M. K. and Nagel, S. (2004). Funds and the technology bubble. *Journal of Finance*, 59(5):2013–2040.

Brunnermeier, M. K. and Pedersen, L. H. (2008). Market Liquidity and Funding Liquidity. *The Review of Financial Studies*, 22(6):2201–2238.

Burghardt, G. and Belton, T. (2005). *The Treasury Bond Basis: An in-Depth Analysis for Hedgers, Speculators, and Arbitrageurs*. McGraw-Hill Library of Investment and Finance. McGraw-Hill Education.

Cai, F. (2003). Was there front running during the LTCM crisis. Technical report.

Chan, N., Getmansky, M., Haas, S. M., and Lo, A. W. (2006). Do hedge funds increase systemic risk? Economic Review 4, Federal Reserve Bank of Atlanta.

Correa, R., Du, W., and Liao, G. Y. (2020). U.S. Banks and Global Liquidity. NBER Working Papers 27491, National Bureau of Economic Research, Inc.

D'Amico, S., Fan, R., and Kitsul, Y. (2018). The scarcity value of treasury collateral: Repo-market effects of security-specific supply and demand factors. *Journal of Financial and Quantitative Analysis*, 53(5):2103–2129.

Du, W., Hébert, B., and Wang, A. (2019). Are Intermediary Constraints Priced? Research Papers 3770, Stanford University, Graduate School of Business.

Du, W., Tepper, A., and Verdelhan, A. (2018). Deviations from Covered Interest Rate Parity. *Journal of Finance*, 73(3).

Duffie, D. (2020). Still the World's Safe Haven? Redesigning the U.S. Treasury Market After the COVID-19 Crisis. *Hutchins Center Working Papers*, 48(62):409–432.

Fisher, M., Nychka, D., and Zervos, D. (1995). Fitting the term structure of interest rates with smoothing splines. Finance and Economics Discussion Series 95-1, Board of Governors of the Federal Reserve System (U.S.).

Fleckenstein, M. and Longstaff, F. A. (2018). Shadow Funding Costs: Measuring the Cost of Balance Sheet Constraints. NBER Working Papers 24224, National Bureau of Economic Research, Inc.

Flood, M. and Monin, P. (2016). Form PF and the Systemic Risk of Hedge Funds: Risk-measurement precision for option portfolios. *Journal of Alternative Investments*, 18(4):125–147.

Flood, M., Monin, P., and Bandyopadhyay, M. (2015). Gauging Form PF: Data tolerances in regulatory reporting on hedge fund risk exposures. Working Paper No. 15-13, Office of Financial Research.

Greenwood, R. and Vayanos, D. (2014). Bond Supply and Excess Bond Returns. *The Review of Financial Studies*, 27(3):663–713.

Gromb, D. and Vayanos, D. (2002). Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66(2-3):361–407.

Gromb, D. and Vayanos, D. (2018). The Dynamics of Financially Constrained Arbitrage. *Journal of Finance*, 73(4):1713–1750.

He, Z., Nagel, S., and Song, Z. (2020). Treasury Inconvenience Yields during the COVID-19 Crisis. Technical report, National Bureau of Economic Research, Cambridge, MA.

Infante, S., Press, C., and Saravay, Z. (2020). Understanding Collateral Re-use in the US Financial System. *AEA Papers and Proceedings*, 110:482–486.

Kondor, P. and Vayanos, D. (2019). Liquidity Risk and the Dynamics of Arbitrage Capital. *Journal of Finance*, 74(3):1139–1173.

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

Krishnamurthy, A. and Vissing-Jorgensen, A. (2012). The Aggregate Demand for Treasury Debt. *Journal of Political Economy*, 120(2):233–267.

Kruttli, M., Monin, P., and Watugala, S. (2019). The Life of the Counterparty: Shock Propagation in Hedge Fund-Prime Broker Credit Networks. Working Papers 19-03, Office of Financial Research, US Department of the Treasury.

Lenel, M., Piazzesi, M., and Schneider, M. (2019). The short rate disconnect in a monetary economy. *Journal of Monetary Economics*, 106.

Nagel, S. (2016). The Liquidity Premium of Near-Money Assets*. *The Quarterly Journal of Economics*, 131(4):1927–1971.

Pástor, L. and Vorsatz, M. B. (2020). Mutual Fund Performance and Flows during the COVID-19 Crisis. Technical Report 4.

Schrimpf, A., Shin, H. S., and Sushko, V. (2020). Leverage and margin spirals in fixed income markets during the Covid-19 crisis. BIS Bulletins 2, Bank for International Settlements.

Singh, M. and Stella, P. (2012). Money and Collateral. IMF Working Papers 2012/095, International Monetary Fund.

The President's Working Group on Financial Markets (1999). *Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management*. Number April. CreateSpace Independent Publishing Platform.

Figure 1: **Open interest in 5-year Treasury futures contract by delivery month.** The volumes suggest that most contracts tend to roll to the next delivery date just prior to the beginning of the current contract's delivery month.

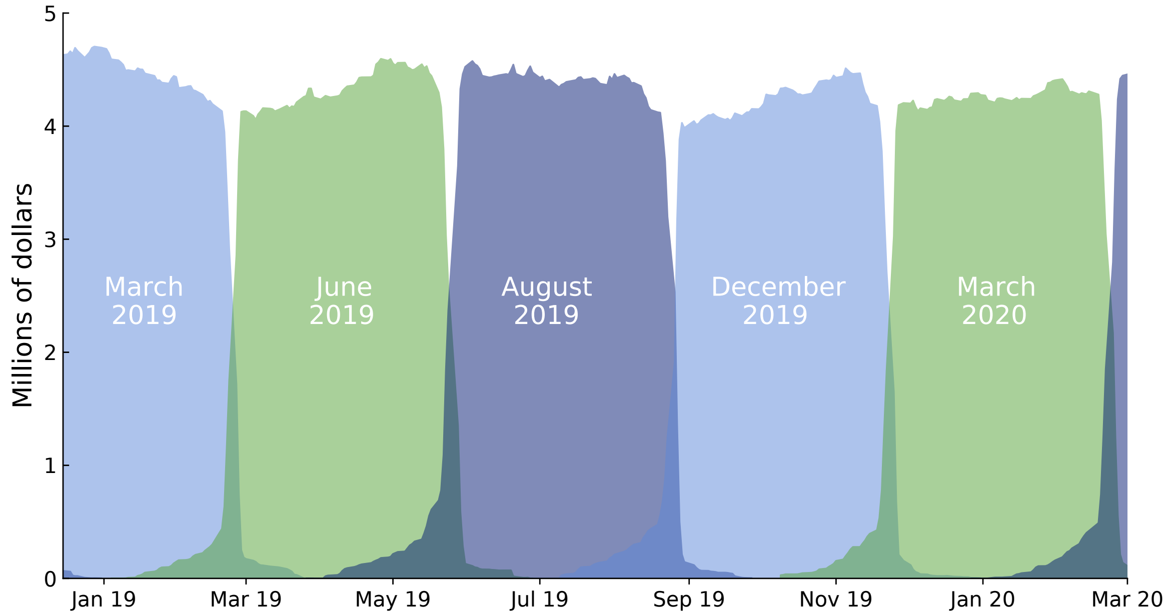


Figure 2: **Time-series of the bond futures implied yield and bill yield.** Long time series of the annualized futures implied yield and the yield for an equivalent maturity bill. Uses the second-to-deliver futures contract.

Figure 3: **Convergence of cash and Treasury prices for the cheapest-to-deliver (means).** Each series in this graph is the average deviation of the futures and cash Treasury from the last invoice price of the futures on the delivery date. The x-axis denotes days to delivery. The gray background denotes open interest in the contract, reported on the right axis.

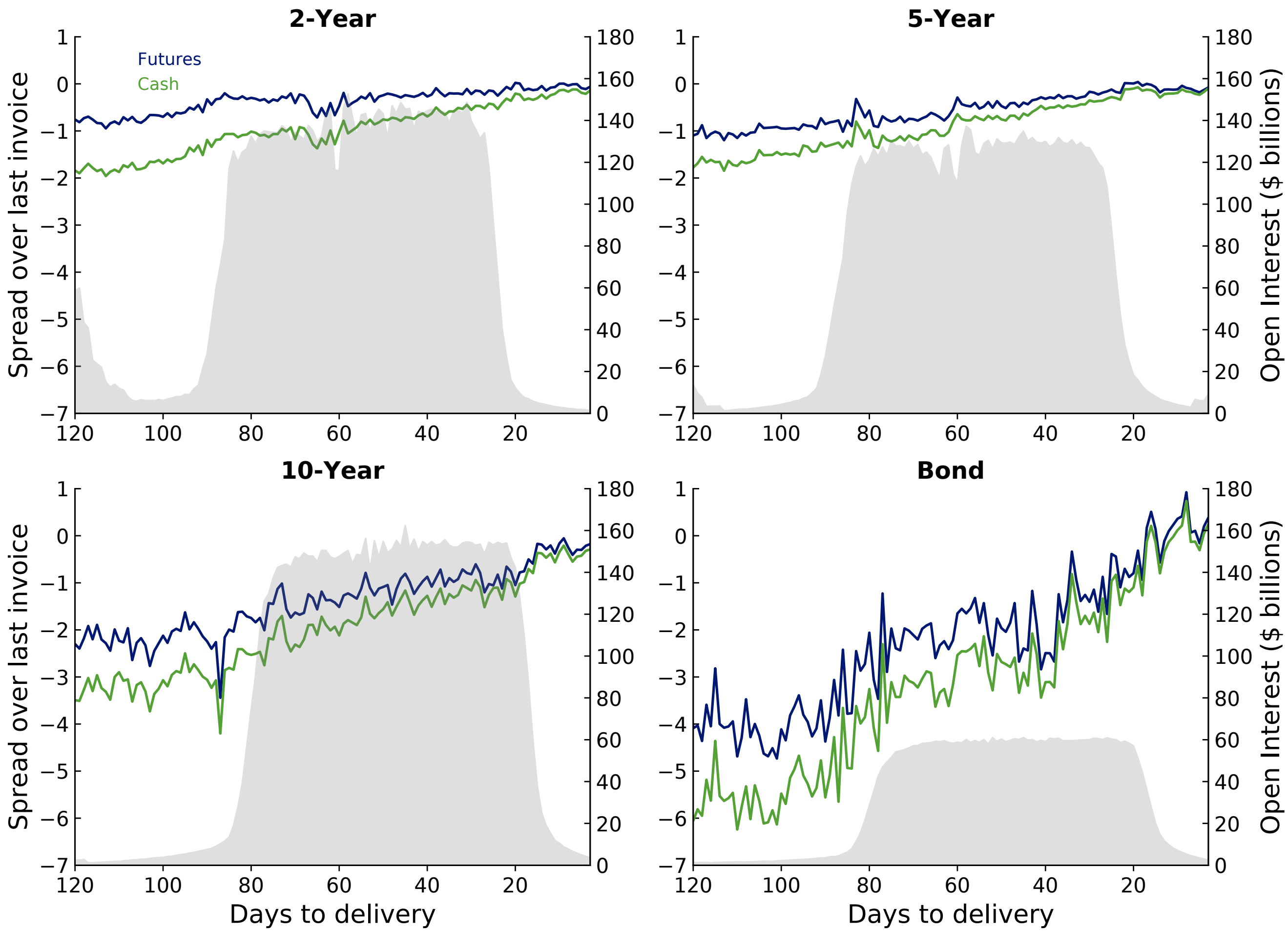


Figure 4: **Non-convergence of cash and futures prices for the non-cheapest-to-deliver.** Each series in this graph is the average deviation of the futures and cash Treasury from the last invoice price of the futures on the delivery date. All panels are for the 5-year futures, each showing a different deliverable. The bottom right panel shows the average across all deliverable Treasuries for the contract.

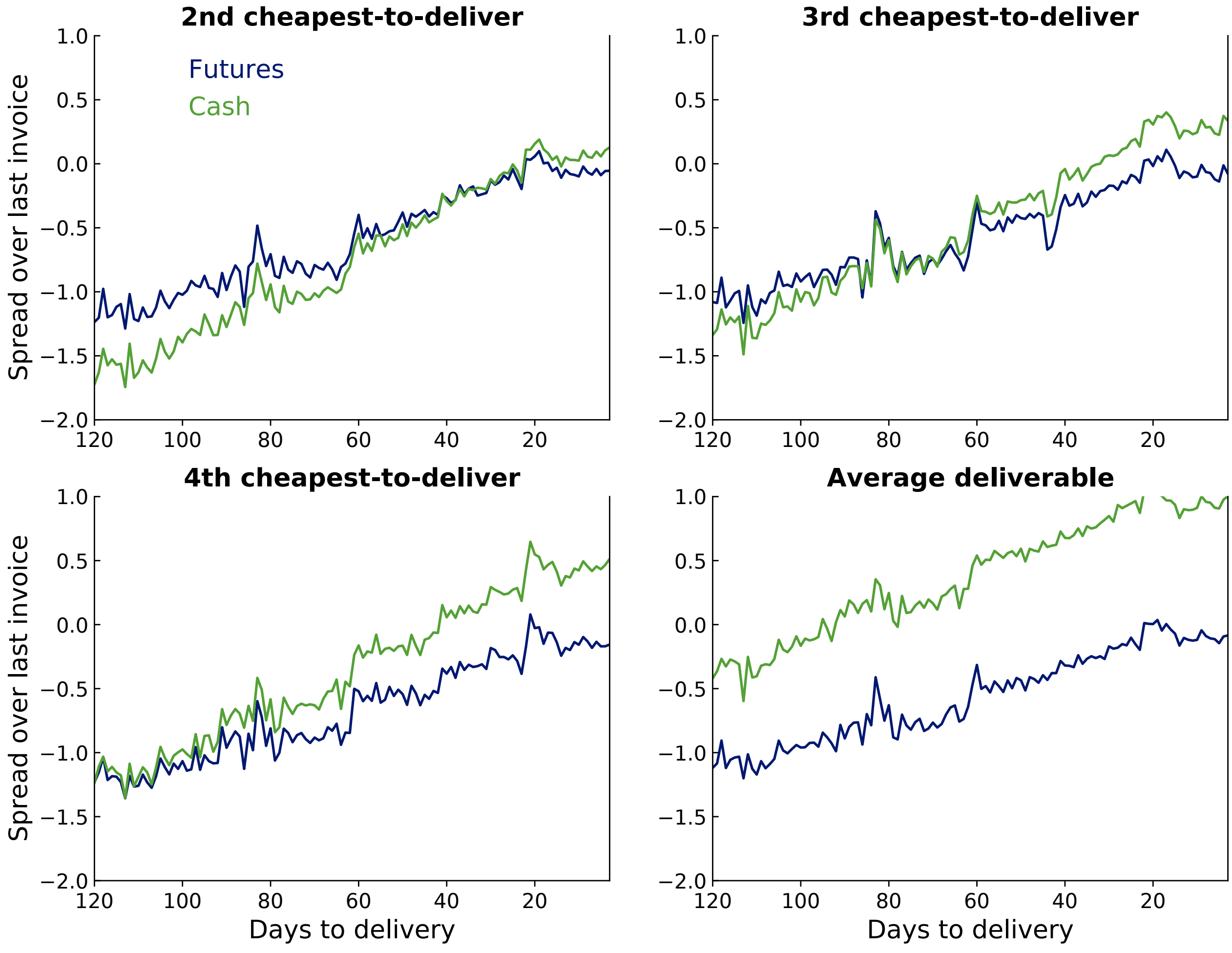


Figure 5: **Deviations of Treasury prices from the replicating portfolio.** Each series graph shows the deviation of the futures-implied yield from the yield on a similar maturity bill and the open interest in that contract. Values above zero imply the replicating portfolio is overvalued relative to the cheapest-to-deliver. These series use the second-to-deliver contract.

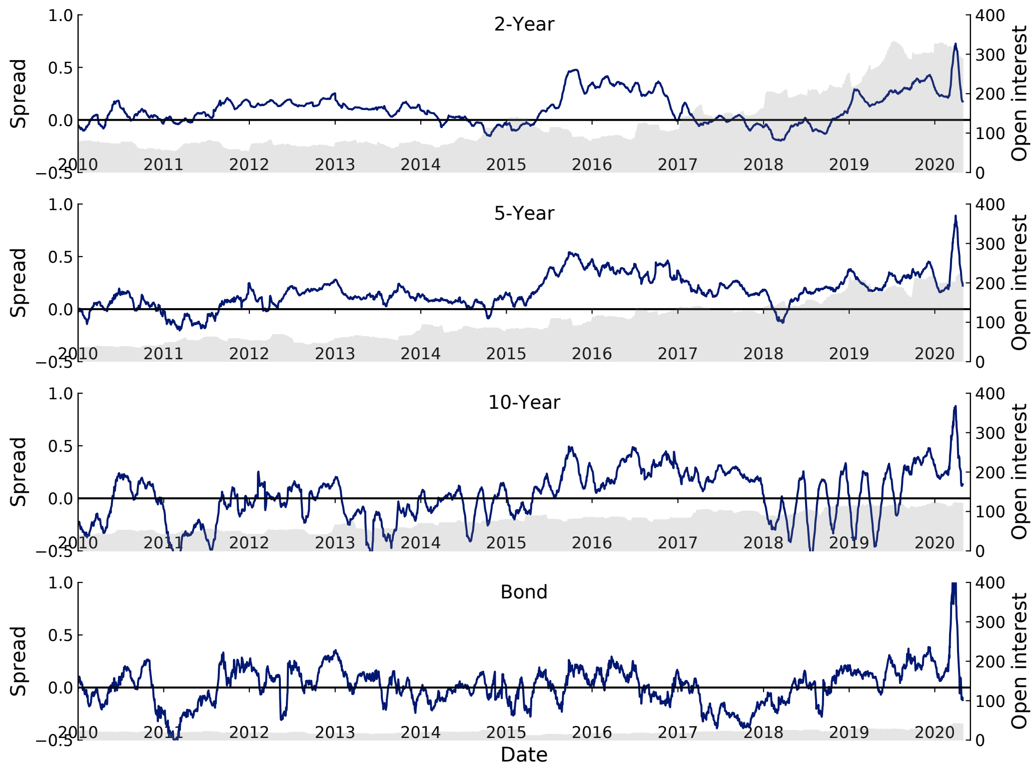


Figure 6: **Treasury bond and note deviations around the LTCM crisis and the Lehman Brothers' bankruptcy.** The top panel shows the deviation of the futures implied yield of the bond futures as a spread over the bill yield in the period surrounding the LTCM crisis. In gray, we show open interest in the futures contract, with values on the right axis. The bottom panel shows the deviations of the 2-year note yield from the bill yield around Lehman Brothers' bankruptcy. Again, in gray we show open interest in the futures contract, with values on the right axis.

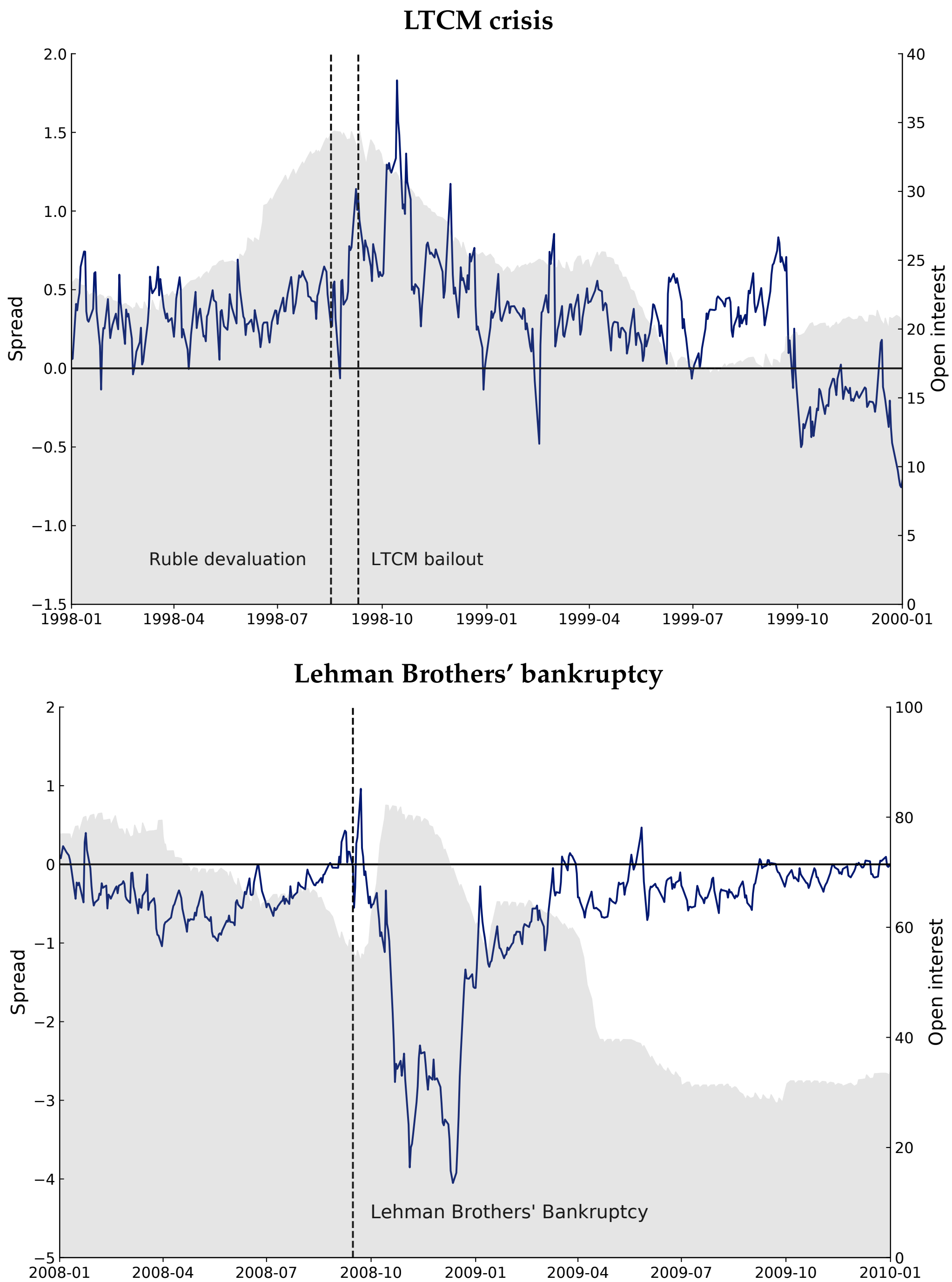


Figure 7: **Holdings of Treasuries over time ($ billions).** The left panel of this figure shows outstanding Treasury holdings by category of holder, specifically foreign holdings, Federal Reserve holdings, and holdings by all other investors. The right panel shows primary dealer net Treasury exposures broken out by maturity.

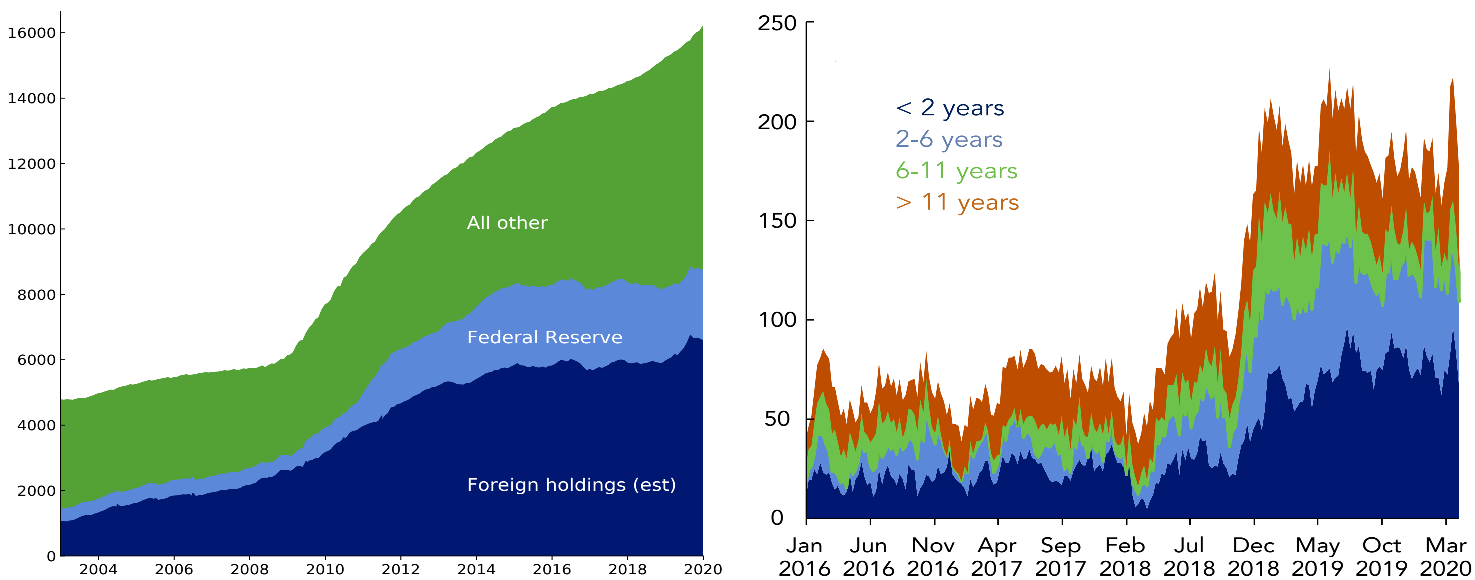


Figure 8: **Futures contracts of hedge funds and asset managers.** This figure shows notional dollars in long and short Treasury futures positions for asset managers, hedge funds and other traders across all Treasury futures contracts.

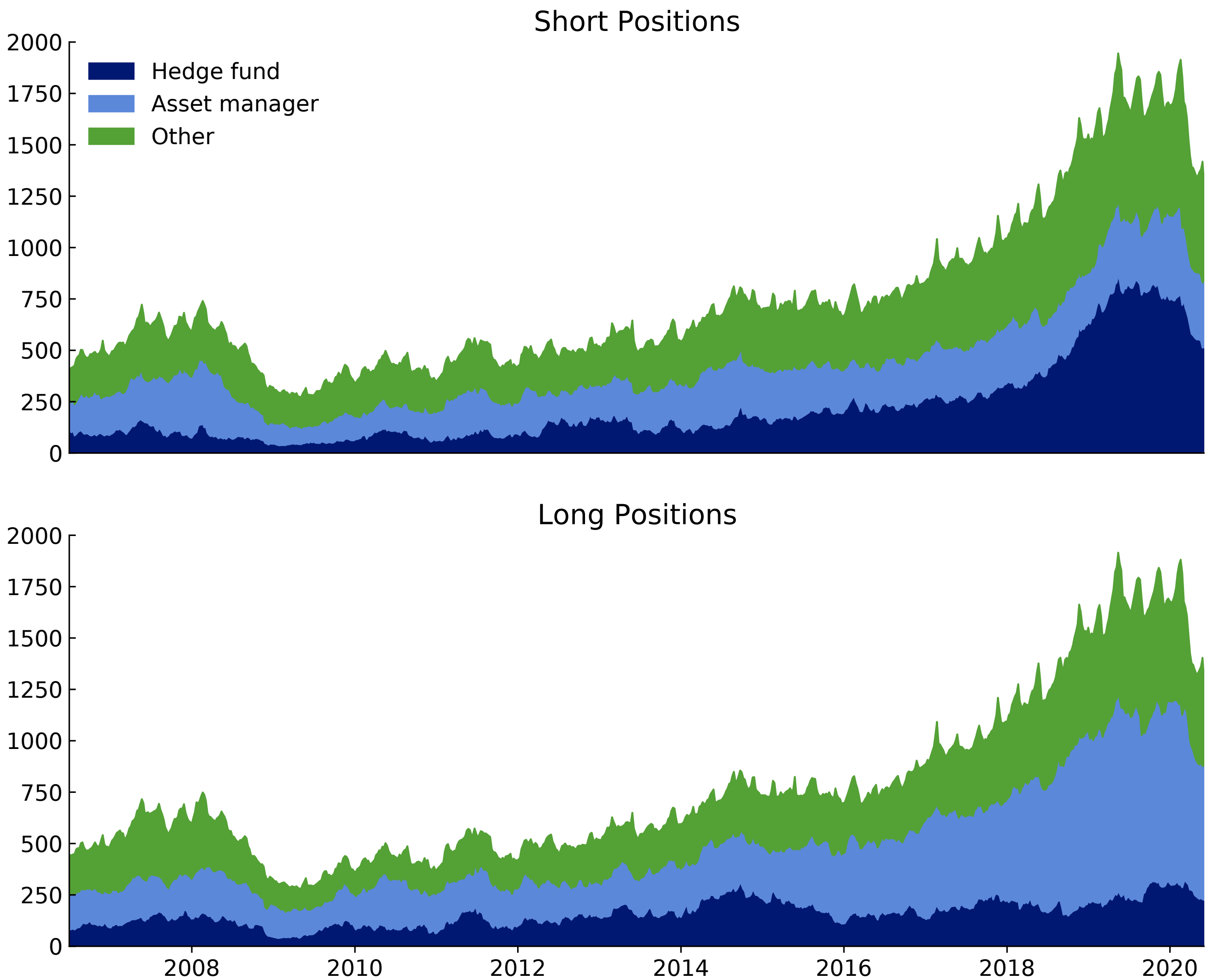


Figure 9: **Hedge fund Treasury exposures and repo ($ billions).** The top panel shows hedge fund long and short Treasury exposure along with their notional long and short futures exposure. The bottom panel shows hedge fund repo borrowing and lending as reported in the SEC's Private Fund Statistics.

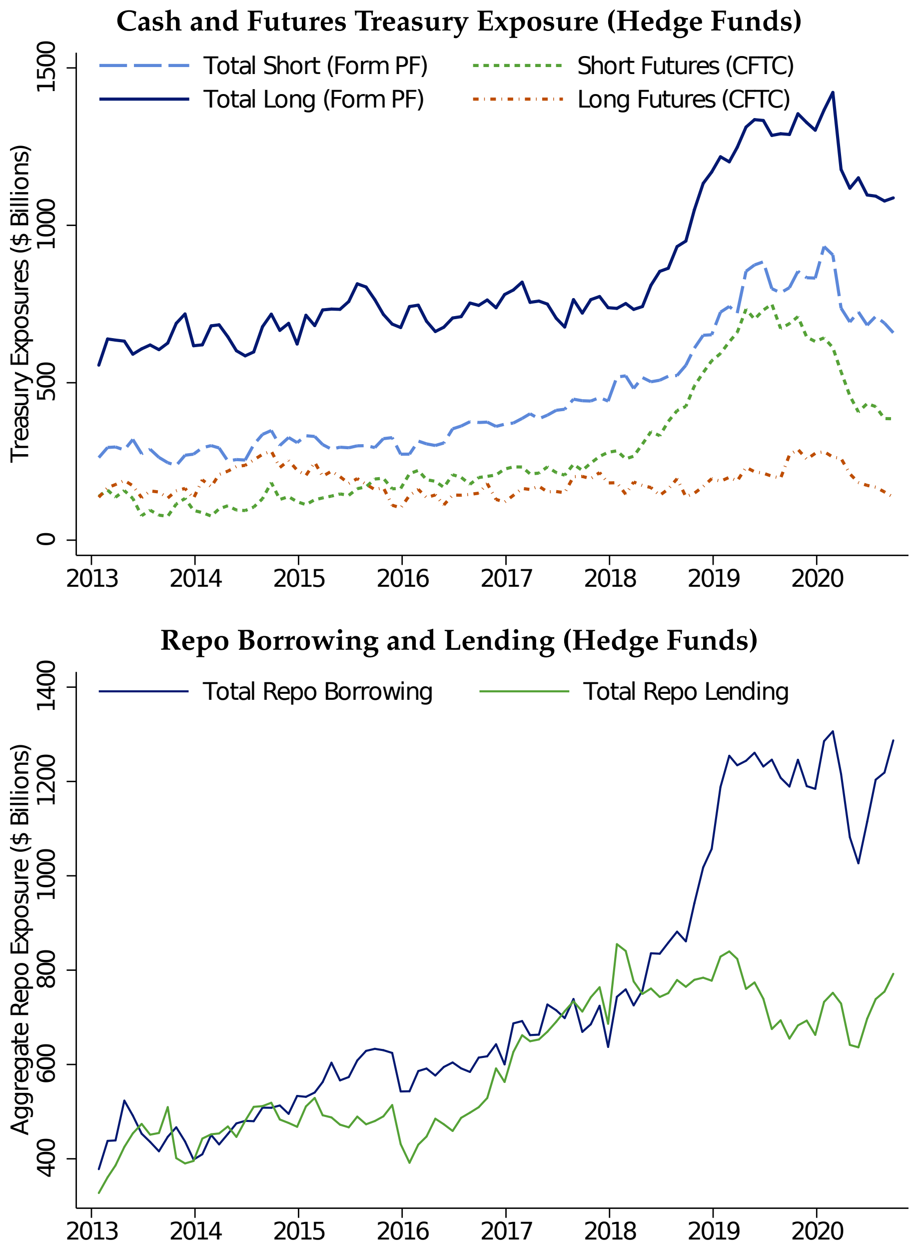


Figure 10: **Hedge fund net repo borrowing vs. the cash-futures disconnect.** This figure shows hedge fund aggregate net repo positions against the cash-futures disconnect for the 2-year note contract.

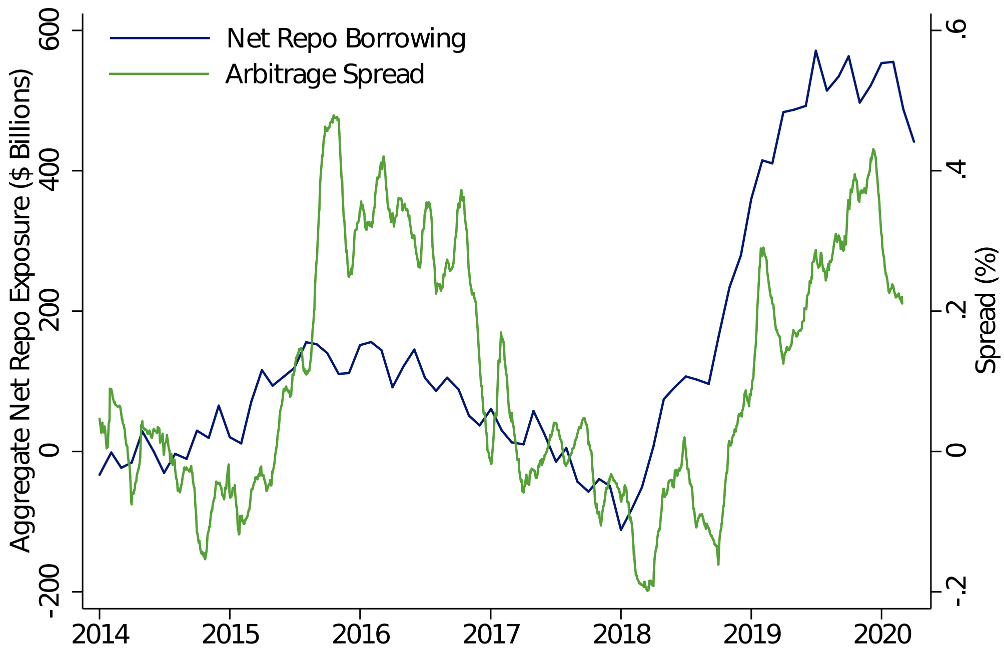


Figure 11: **Hedge fund DVP repo in Treasuries by maturity date ($ billions, average daily transaction value).** Above, gray areas are deliverables for December 2019, below for March 2020. The green dot denotes positions in the cheapest-to-delivery for the two-year December contract.

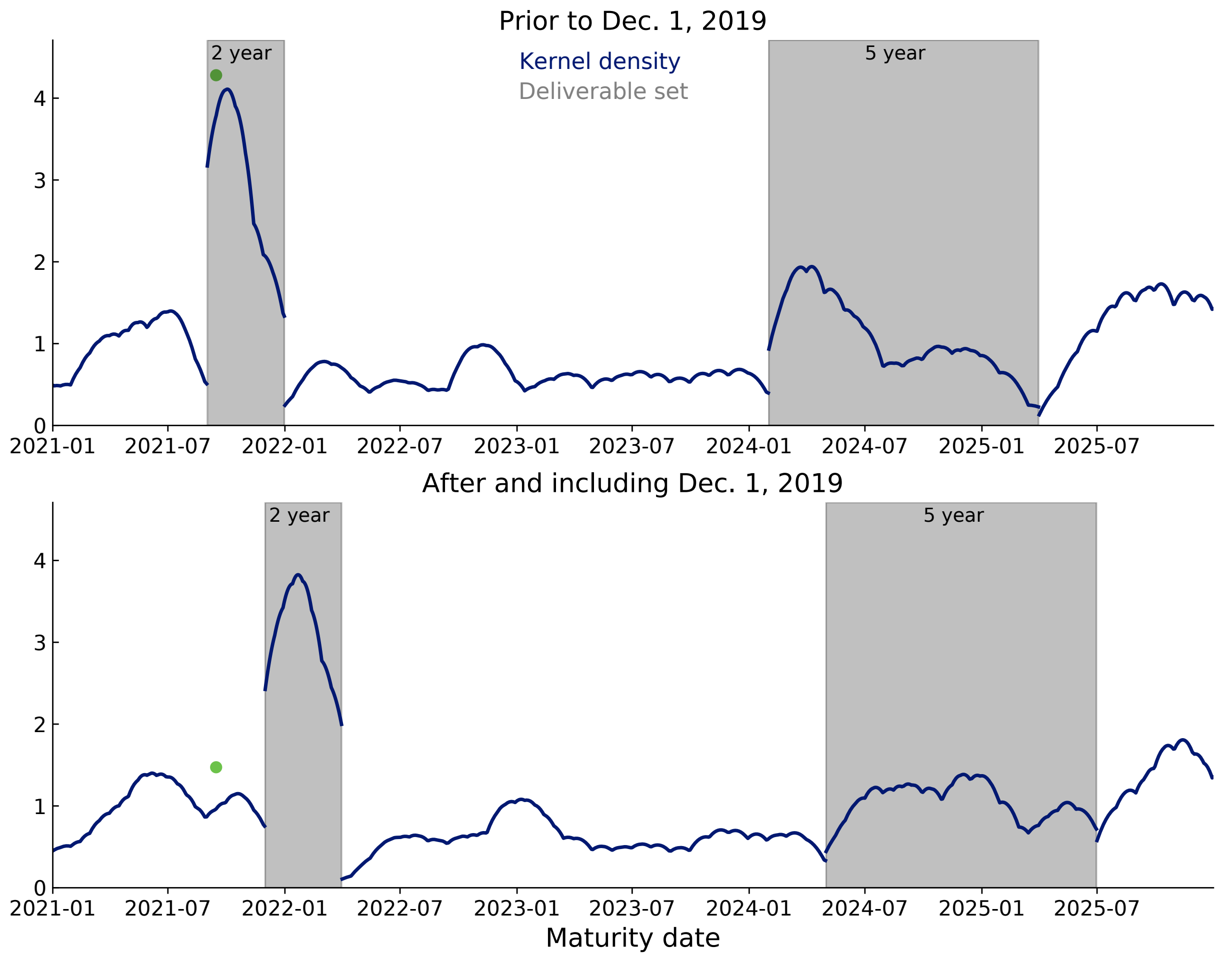


Figure 12: **Structure of the basis trade in recent years.**

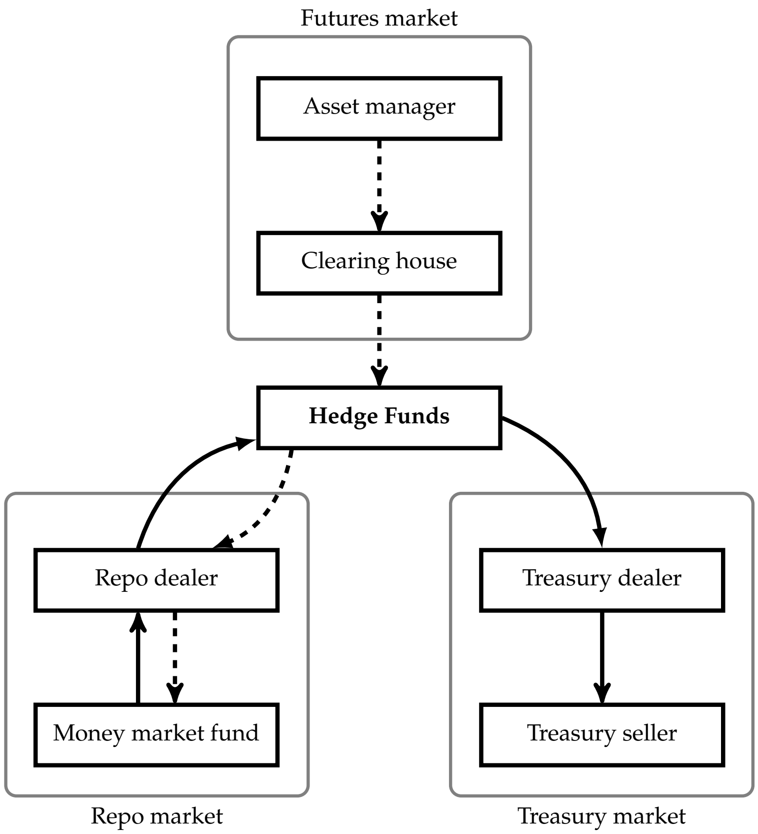


Figure 13: **Equilibrium in the limits to arbitrage model.** This graph displays the equilibrium determination of futures and cash Treasury prices in the limits to arbitrage model. Equilibrium occurs where the "arbitrage capacity" line determined by hedge funds' margin constraints and repo illiquidity intersects the "risk sharing" line, which determines futures and spot prices given an allocation of Treasuries to dealers and speculators.

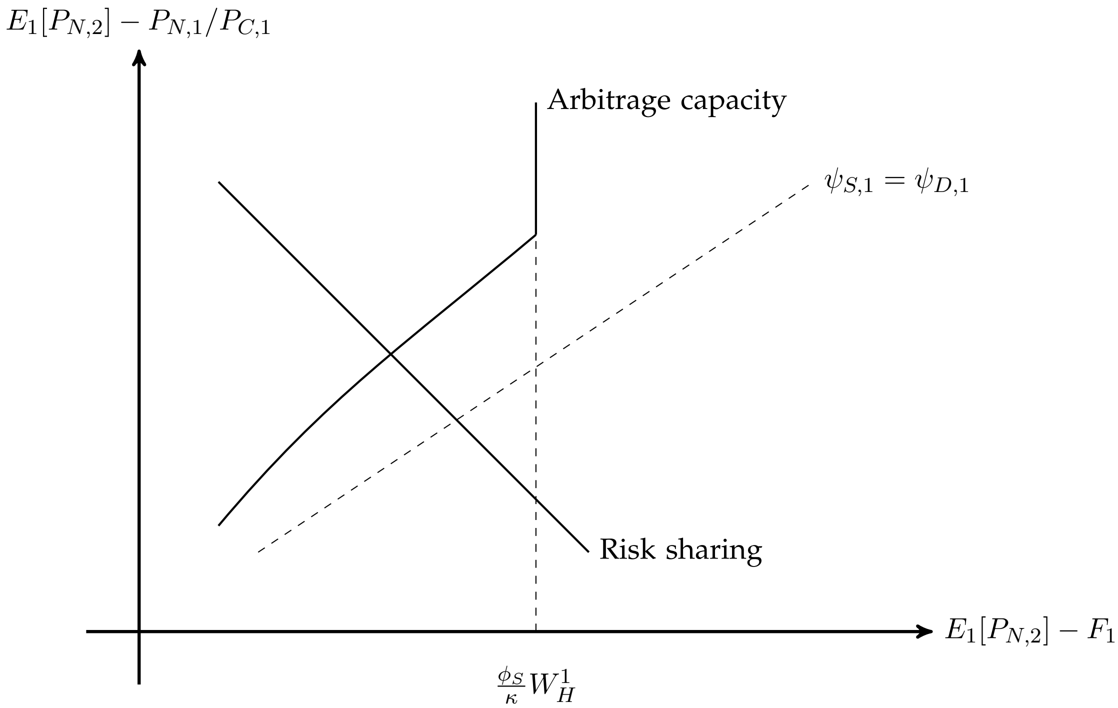


Figure 14: **Comparative statics in the limits to arbitrage model.** These graphs display how changing noise trader demand, increases in margins, and repo market illiquidity affect the equilibrium of our limits to arbitrage model. Darker lines correspond to higher sales (top panel), higher margins (middle panel), and greater illiquidity (lower panel).

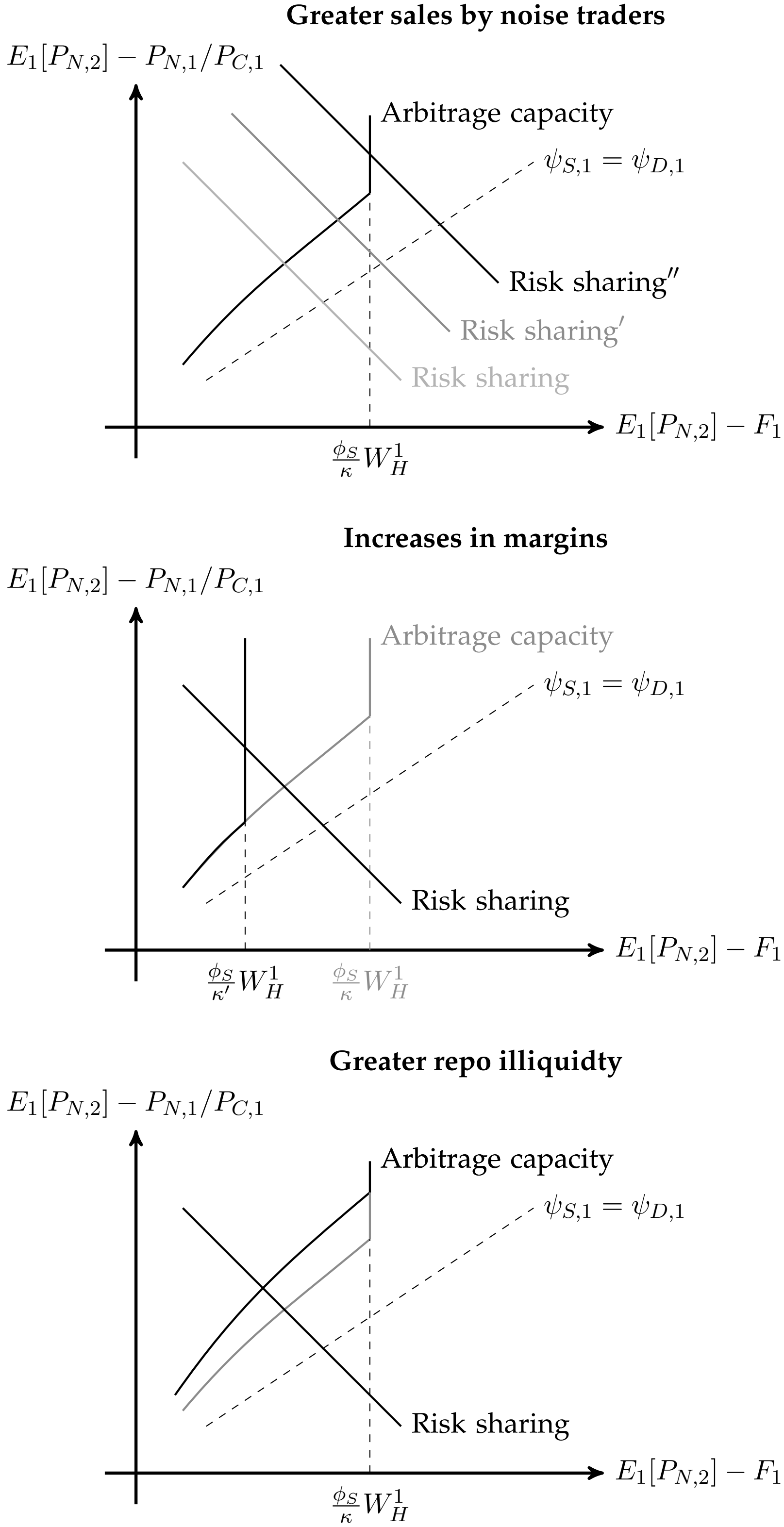


Figure 15: **Futures implied yields and the GCF repo spread.** Each panel in this figure shows the spread of futures implied yields over equivalent maturity bill yields in blue, and the spread of the GCF repo rate over the bill rate in green.

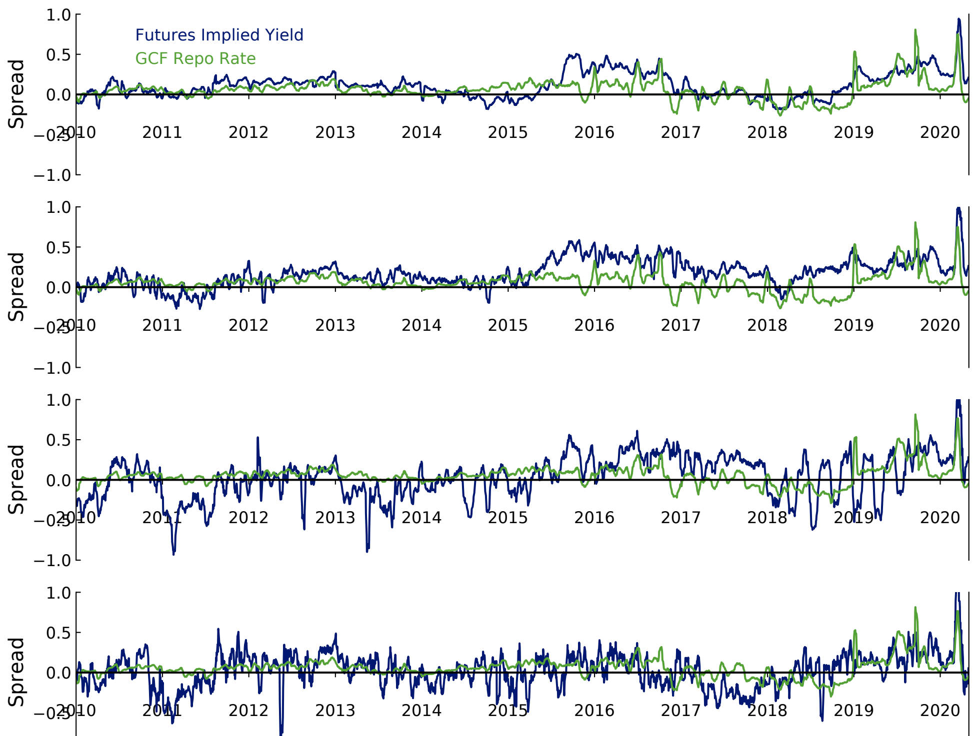


Figure 16: **Absolute value of 5-year note deviations and measures of financial market volatility.** In the top panel, we plot absolute values of the deviation between the 5-year note futures implied yield and the yield on an equivalent maturity bill as well as the MOVE index, with values in light blue recorded on the right axis. The bottom panel has the same values for the futures implied yield, but displays the VIX index in light blue with values on the right axis.

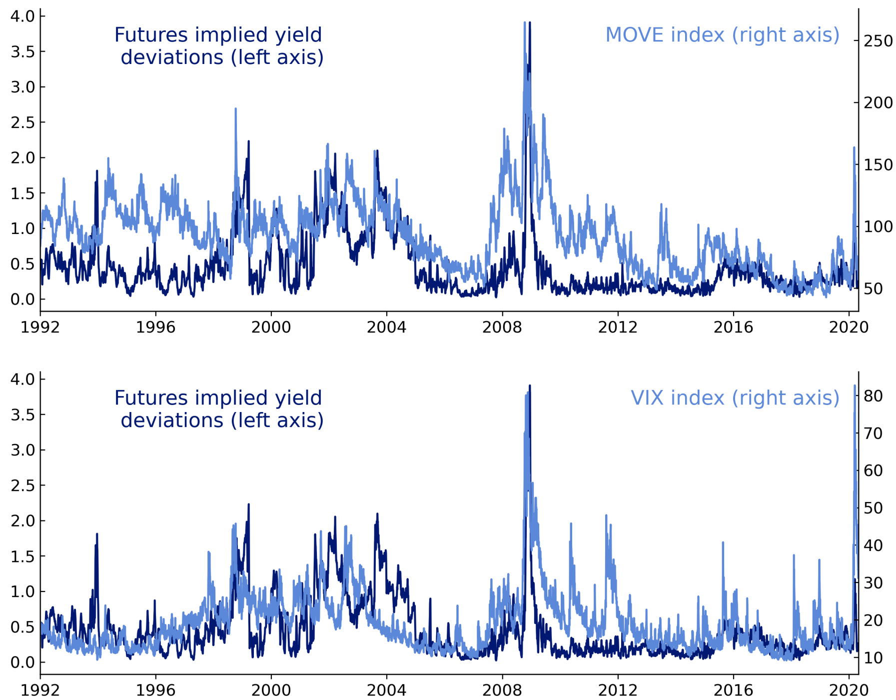


Figure 17: **Treasury volatility indexes.** CME 10-year Treasury VIX and the MOVE Index are option implied Treasury volatility indexes.

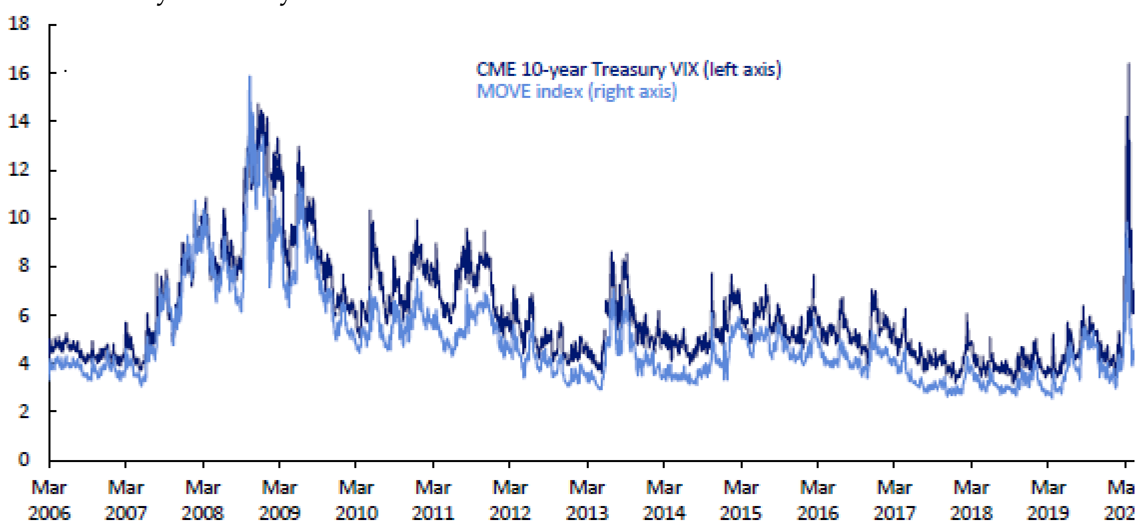


Figure 18: **Bid-ask spreads for off-the-run Treasuries ($).** March illiquidity was concentrated in off-the-run securities. Spreads are the difference between bid and ask prices for $100 notional in the fourth-from-most-recent Treasury issuance as of January 2020.

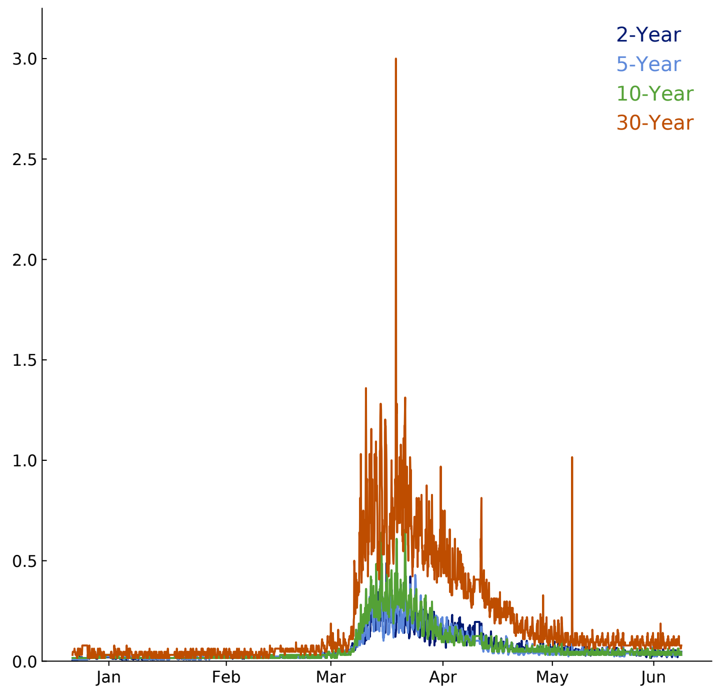


Figure 19: **March 2020 stress in the basis trade.** Top panel shows maintenance margins on $200,000 notional in two-year Treasury futures contracts, and price movements are normalized to changes in those notional values for January through April 2020. Bottom panel shows the futures-implied yield, bills rate, and DVP repo rate for the same period. DVP repo rate is the average overnight rate for sponsored borrowers with Treasury collateral. Futures-implied yields are for July contracts.

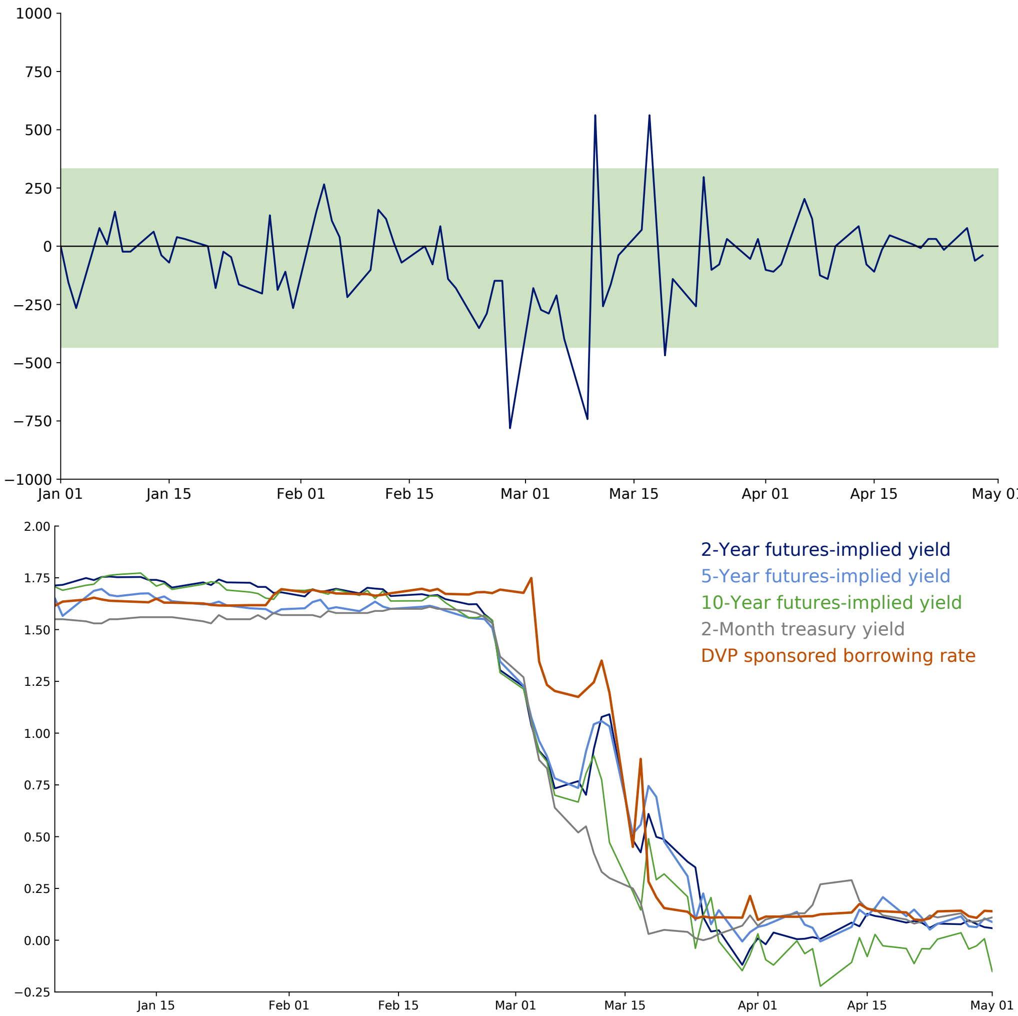


Figure 20: **Spread on the cheapest-to-deliver Treasury (percentage points).** Left panel shows the spread on the cheapest-to-deliver over time. Wider spreads show deliverable Treasuries were more valuable during March. Spread is the fitted spline yield minus the yield on the cheapest-to-deliver. Right panel shows the underlying spline fit and individual yields for March 11, 2020. The actual values are shown as dots, with blue dots denoting non-deliverable Treasuries, green dots denoting deliverable Treasuries, and red dots denoting CTDs. The fitted spline yield curve is the line in blue, and the blue area around it denotes the 95% confidence interval for these yields. The gray areas denote the bounds of deliverable Treasuries for the 2-year, 5-year, and 10-year contracts. The cheapest-to-deliver and deliverable Treasury windows are for June futures contracts.

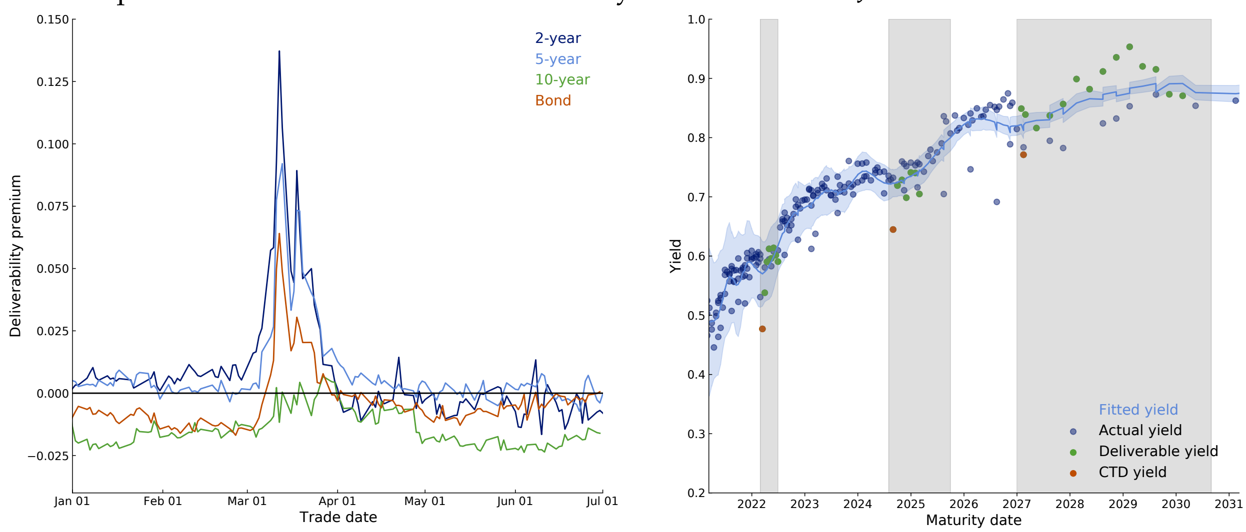


Figure 21: **DVP repo rates (top panel, percentage point spread over fed funds target midpoint) and Federal Reserve facility participation (bottom panel, $ billions).** In the top panel, we present DVP repo rates from January to May 2021 across different segments of the market. Repo rates are average overnight Treasury rates for each market segment. The two black lines represent the average rate offered by the Federal Reserve's Overnight Reverse-Repurchase Facility (ON-RRP) and Repo Facility (RP). In the bottom panel we present volumes in the Federal Reserve's RP and ON-RRP facilities in billions of dollars. Gray shaded areas represent days when the average rate in the RP facility was bid up beyond its minimum rate.

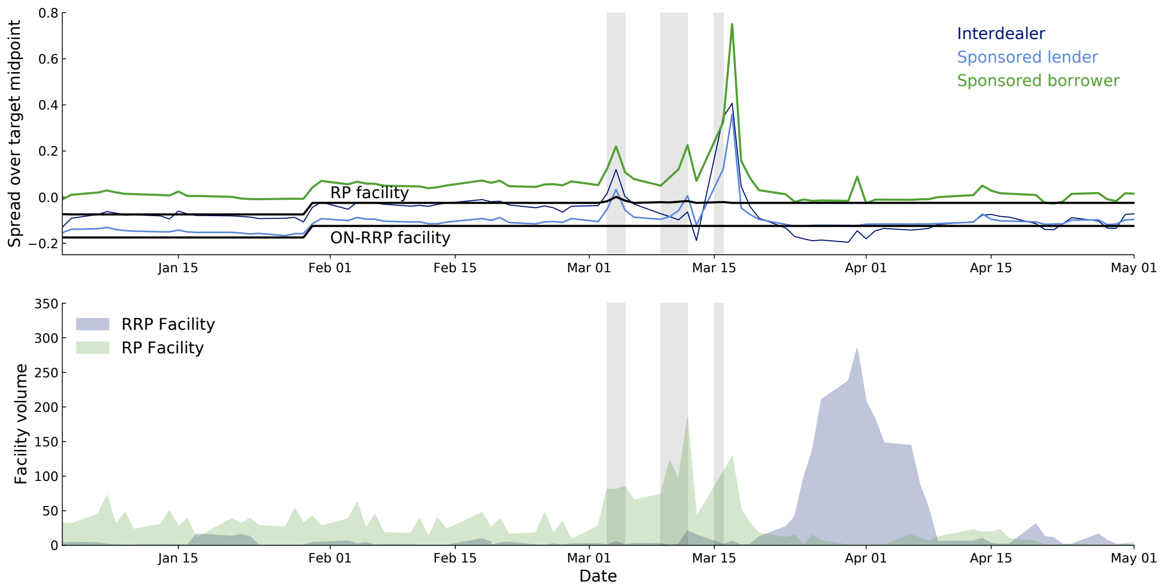


Figure 22: **Cumulative Federal Reserve purchases of the cheapest-to-deliver securities ($ billions).** Cumulative purchases of the cheapest-to-deliver (CTD) Treasuries for June delivery.

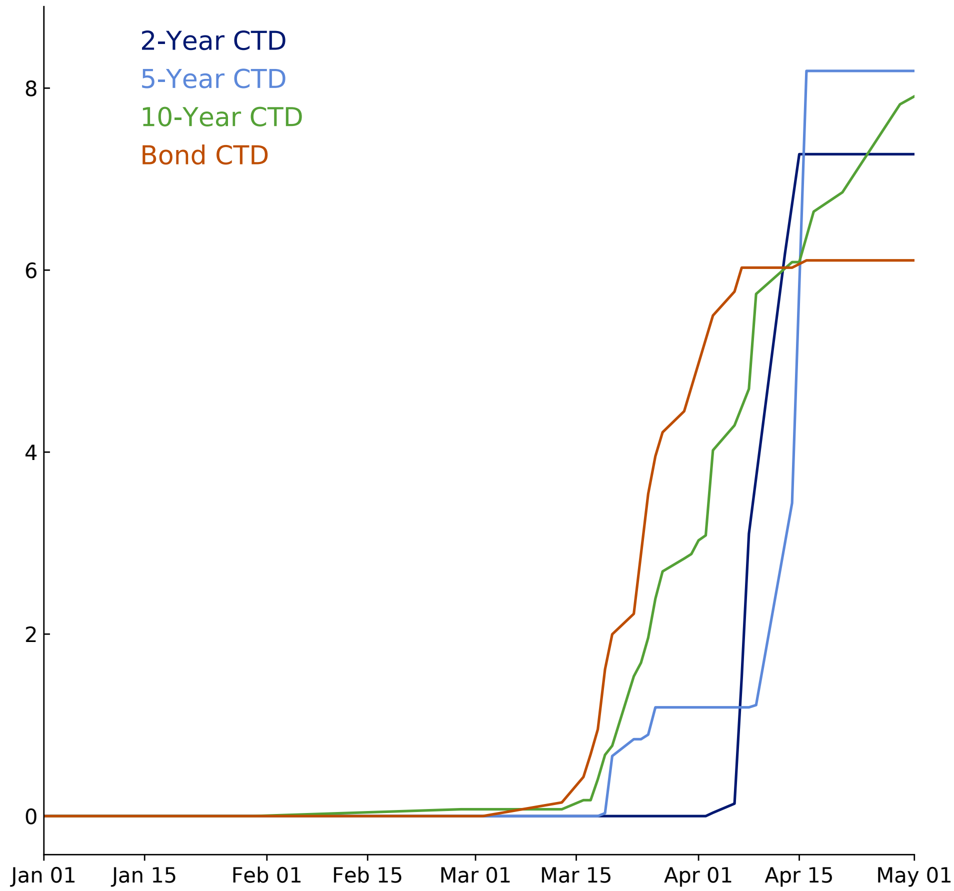


Table 1: **Details of terms for different Treasury futures contracts.** This table provides details on the contract terms for Treasury futures traded at the Chicago Board of Trade, including the original and residual maturity required for cash Treasuries to be deliverable into the contract and notional amounts for each contract.

Treasury futures contract name	Original maturity restrictions	Residual maturity restrictions at delivery	Notional amount
2-year note	$\leq$ 5 years, 3 months	$\geq$ 1 year, 9-months $\leq$ 2 years	200,000
5-year note	$\leq$ 5 years, 3 months	$\geq$ 4 years, 2 months	100,000
10-year note	$\leq$ 10 years	$\geq$ 6 years, 6 months $\leq$ 10 years	100,000
10-year ultra note	$\leq$ 10 years	$\geq$ 9 years, 5-months $\leq$ 10 years	100,000
Bond		$\geq$ 15 years $\leq$ 25 years	100,000
Ultra bond		$\geq$ 25 years	100,000

Table 2: **Fitted yield spreads for cheapest-to-deliver and on-the-run Treasuries.** For the right two columns, we show the average pricing error from our spline model as well as the standard deviation for Treasuries that are cheapest-to-deliver in each category. Averages are taken from the full sample between January 1, 1992 and May 1, 2020. For the left two columns, we show the same statistics for on-the-run Treasuries in that maturity category.

	Cheapest-to-deliver		On-the-run	
Maturity	Mean	Std. Dev	Mean	Std. Dev
2-year	1.09	6.70	1.17	5.86
5-year	0.85	2.49	2.09	4.38
10-year	0.53	2.65	6.08	6.10
30-year	0.34	2.24	2.28	2.87

Table 3: **Yield spread for cheapest-to-deliver regressed on the on-the-run spread.** This graph shows a regression of the cheapest-to-deliver premium, obtained as the residual from a spline curve, against a similar premium calculated for the on-the-run premium. These premia are winsorized at the 0.5% level. Regressions in the top panel use daily data from January 1, 1992 to May 1, 2020. Regressions in the bottom panel use daily data from January 1, 2016 to May 1, 2020. Standard errors are calculated using Newey-West with 22 business days of lags.

1992-2020

	Dependent variable: Cheapest-to-deliver premium			
	2-Year	5-Year	10-Year	Bond
On-the-run premium	0.727^{***}	0.135^{***}	0.187^{***}	-0.127^{*}
	(0.097)	(0.036)	(0.047)	(0.021)
R^2	0.403	0.052	0.175	0.021
N	7,080	7,080,	7,080	6,886
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

2016-2020

	Dependent variable: Cheapest-to-deliver premium			
	2-Year	5-Year	10-Year	Bond
On-the-run premium	1.167^{***}	0.492^{***}	0.115^{***}	0.631^{***}
	(0.134)	(0.064)	(0.037)	(0.219)
R^2	0.832	0.532	0.044	0.109
N	1,081	1,081	1,081	1,081
Note:				$^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 4: **Cash flows from arbitrage strategy with a short bills position.**

	Buy τ-maturity note	Sell $F_{t,\tau,T}$ of T-maturity bill	Short τ futures delivering at T	Net cash flow
Time 0	$-P_{0,\tau}$	$F_{0,\tau,T}B_{0,T}$	0	$F_{0,\tau,T}B_{0,T} - P_{0,\tau}$
Time T	0	$-F_{0,\tau,T}$	$F_{0,\tau,T}$	0
Time τ	1	0	-1	0

Table 5: **Cash flows from arbitrage strategy with repo.**

	Buy τ-maturity note	Borrow against note in repo market	Short τ futures delivering at T	Net cash flow
Time 0	$-P_{0,\tau}$	$P_{0,\tau}$	0	0
Time T	0	$-P_{0,\tau}(1+r)^T$	$F_{0,\tau,T}$	$F_{0,\tau,T} - P_{0,\tau}(1+r)^T$
Time τ	1	0	-1	0

Table 6: **Regression of arbitrage spreads on the repo rate.** This table shows results of a regression of the spread between futures implied yields and equivalent maturity bill yields on either the spread of the GCF repo index over the interest rate on excess reserves or dealer Treasury exposures. Fixed effects are included to account for the days to deliver of the futures contract. Regressions use data from January 1, 2010 through May 1, 2020. For the specifications using the GCF-IOER spread, regressions are daily, while the specifications using dealer Treasury exposure are weekly. Dealer exposure is net dealer exposure in billions of dollars. Standard errors are calculated using Newey-West with a 22-business-day lag for the specifications using the GCF-IOER spread, and with a three-week lag for the dealer exposure measure.

	Dependent variable: Arbitrage Spread							
	2-Year		5-Year		10-Year		Bond	
GCF - IOER	0.299***		0.355***		0.524***		0.296**	
	(0.092)		(0.122)		(0.196)		(0.122)	
Dealer exposure		1.127***		1.737***		2.233***		1.444***
		(0.298)		(0.361)		(0.374)		(0.332)
Observations	2,154	446	2,001	416	2,014	422	2,582	535
R^2	0.136	0.353	0.119	0.323	0.141	0.319	0.078	0.245
Adjusted R^2	0.081	0.158	0.071	0.156	0.089	0.118	0.031	0.067
Note:			*p<0.1; **p<0.05; ***p<0.01					

Table 7: **Arbitrage deviations, margins, and the VIX.** This table shows results of a regression of the absolute value of the spread of futures implied yields over equivalent maturity bill yields on maintenance margins for the contract as well as dealer Treasury exposure, the VIX, and a dummy for quarter ends. Fixed effects are included to control for the distance to delivery of the futures contract. Regressions use data from January 1, 2010 through May 1, 2020. Data are weekly, and dealer exposure are net exposures in billions of dollars. Dealer exposure is net dealer exposure in billions of dollars. Standard errors are calculated using Newey-West with a three-week lag.

	Dependent variable: |Arbitrage Spread |			
	2-Year	5-Year	10-Year	Bond
Maintenance margins	0.051	0.450***	0.214**	0.070***
	(0.075)	(0.127)	(0.103)	(0.024)
Dealer exposure	0.897***	1.192***	0.556***	0.733***
	(0.200)	(0.262)	(0.211)	(0.281)
VIX	0.665***	0.271	0.640***	1.555***
	(0.155)	(0.225)	(0.242)	(0.396)
Quarter end	−0.030	0.018	0.042	−0.014
	(0.038)	(0.051)	(0.063)	(0.067)
Observations	446	416	422	535
R^2	0.423	0.292	0.247	0.365
Adjusted R^2	0.243	0.109	0.015	0.210
Note:			*p<0.1; **p<0.05; ***p<0.01	

Internet Appendix

A Estimate of the CTD

Formally, the definition of the cheapest-to-deliver is the Treasury most profitable for delivery into the short position, taking account of the repo rate, Treasury prices, and cash prices. For our purposes we define the cheapest-to-deliver as:

$$\text{CTD}_{t,\tilde{\tau},T} = \underset{\tau \in \Omega(t,\tilde{\tau},T)}{\arg\min} \frac{F_{t,\tau,T}}{P_{t,\tau} - B_{t,s} c_s} \tag{15}$$

where $\Omega(t, T)$ includes all Treasuries in the delivery set $\tilde{\tau}$ futures delivering at time T that are available to traders as of time t. This approximates practice among many traders, who tend to form the cheapest-to-deliver using the futures price and cash price and assuming that coupons are reinvested at a constant rate. However, it is important to note that this formulation implicitly assumes that repo rates across the deliverable basket are approximately equal. In the case of special collateral repo this may not be true, as the repo rate on the cheapest-to-deliver may differ from other repo rates.

Because of its convergence to the futures price, we restrict much of our analyses to the cheapest-to-deliver. However, throughout the lifetime of a futures a contract, the cheapest-to-deliver may change. This can happen in two ways: either prices of one Treasury fall relative to the cheapest-to-deliver, or a new Treasury is issued that becomes the cheapest to deliver. While this matters for convergence, it is important to note that a short futures position with a Treasury that is no longer cheapest-to-deliver always preserves the option to hold until delivery and receive the fixed invoice price. As a result, arbitrage strategies are not directly affected by a change in the cheapest-to-deliver whenever a short futures position is advantageous.

Figure A.1 shows the probability within sample that the cheapest-to-deliver on a particular date will be the cheapest-to-deliver on the last trade date. The jump in the 2-Year reflects the fact that the 2-year cheapest-to-deliver on the delivery date has often been a Treasury issued while the contract was active. The jump thus reflects the delivery cycle for 2-year Treasuries. Outside of this departure, however, the cheapest-to-deliver is less likely to change as the underlying becomes lower duration. This reflects in part the increased exposure of long-duration futures to interest rate risk. In particular, a trader holding the cheapest-to-deliver for the bond future has a greater than one-in-four chance that the cheapest-to-deliver will change even in the week prior to delivery.

Figure A.2 shows the probability that the cheapest-to-deliver on a particular date will be the cheapest-to-deliver on the last trade date for the period 2010-2020, which contains our main sample. The stability in the cheapest-to-deliver in the last 10 years is somewhat higher than in previous periods. This increased stability may reflect lower variance in interest rates over this period. The increased stability of cheapest-to-deliver bonds is likely to have reduced risks for long futures positions, but for reasons discussed in the main text is less likely to be an issue for short Treasury futures positions.

B Convergence of the CTD with near certainty

Figure B.1 shows that the variation in the cash-futures disconnect falls to zero for the cheapest-to-deliver at the delivery date. Each line is the squared deviation of the futures and cash prices from

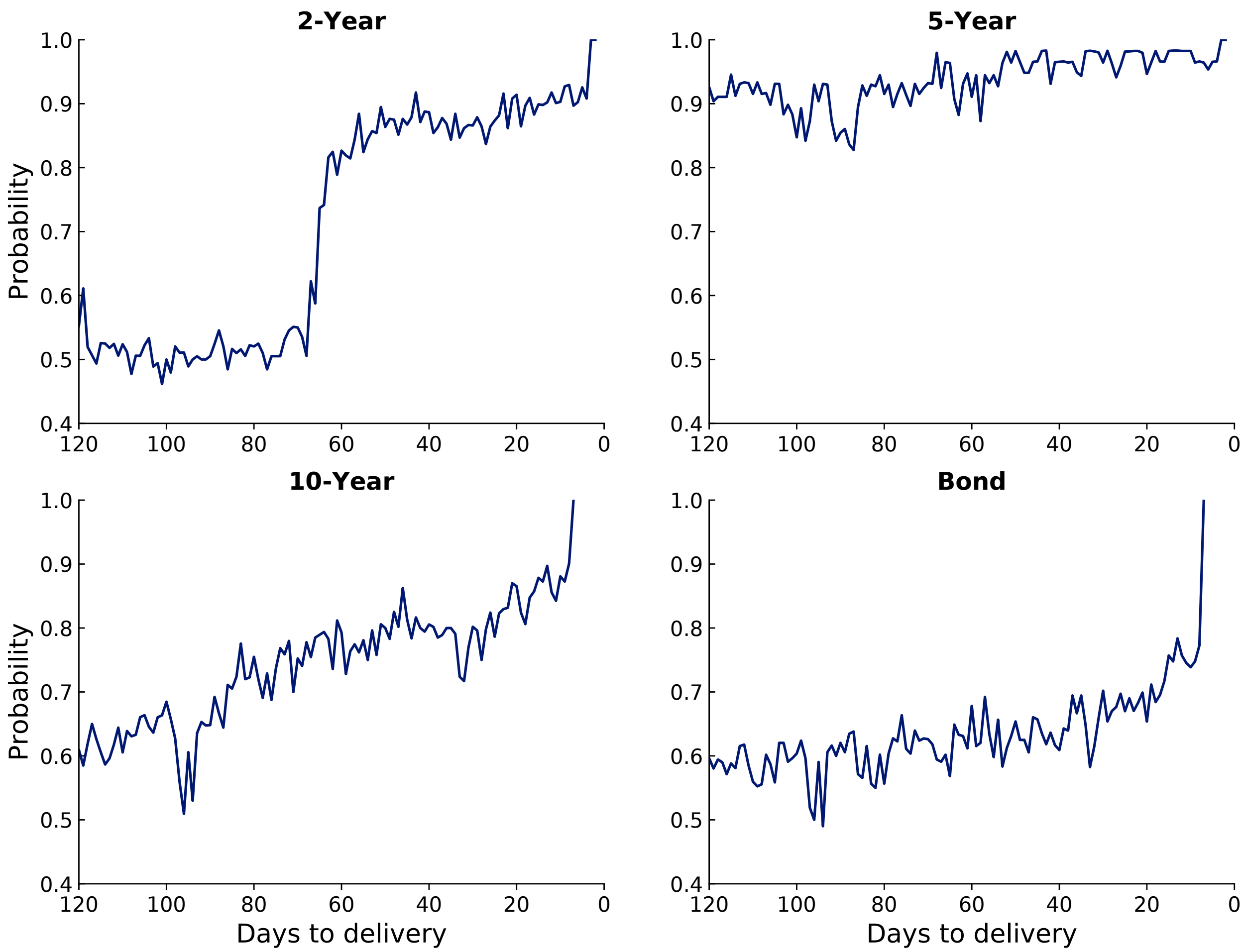


Figure A.1: **Stability of the cheapest-to-deliver for the full sample.** Each series in this graph is the deviation of the cheapest-to-deliver price from the price of a replicating portfolio of bills and futures. Values below zero imply the replicating portfolio is overvalued relative to the cheapest-to-deliver. These series use the second-to-deliver contract.

Figure A.2: **Stability of the cheapest-to-deliver, 2010-2020.** Each series in this graph is the sample probability that the Treasury that is the cheapest-to-deliver on that day is the same as the Treasury that is cheapest-to-deliver at the delivery date.

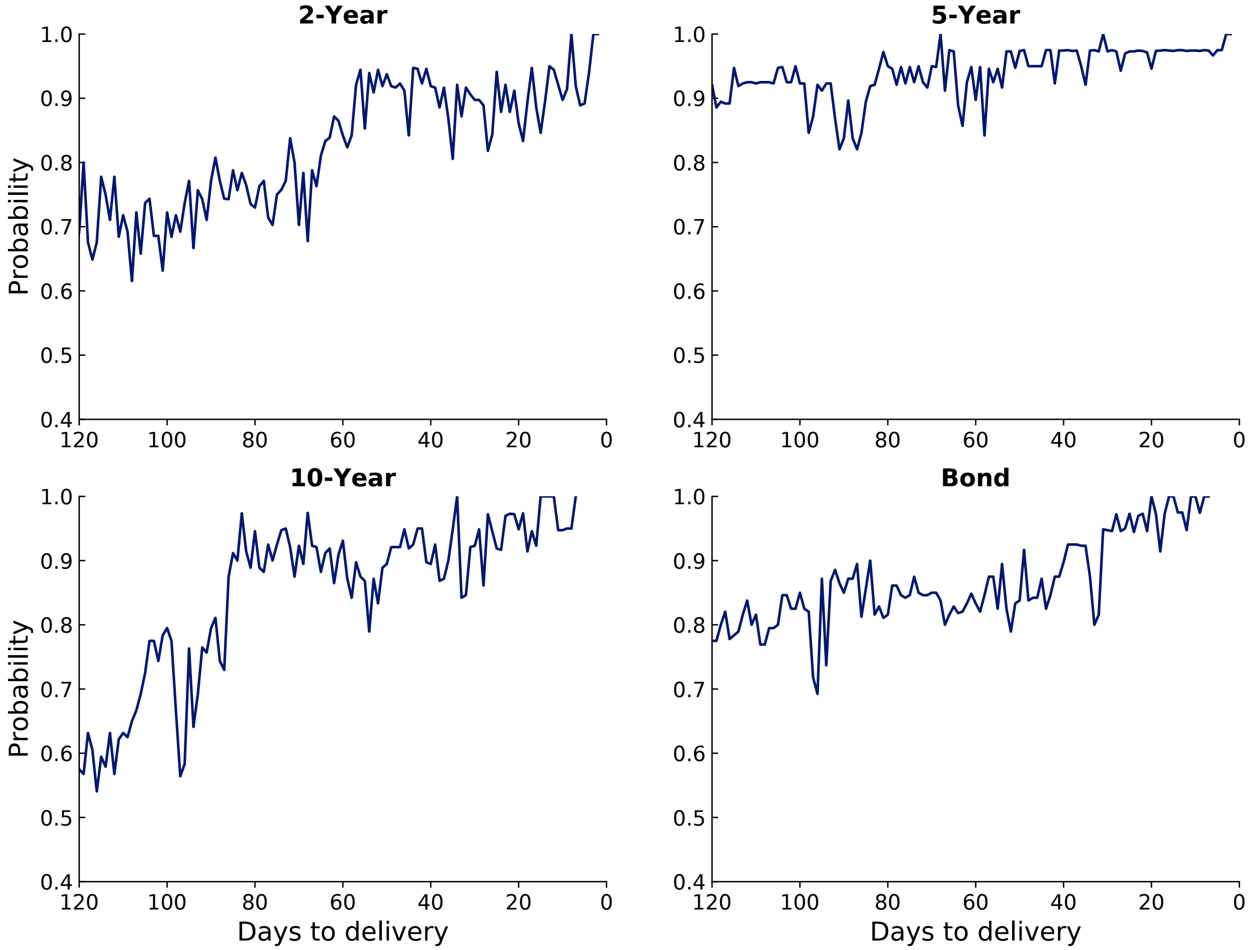
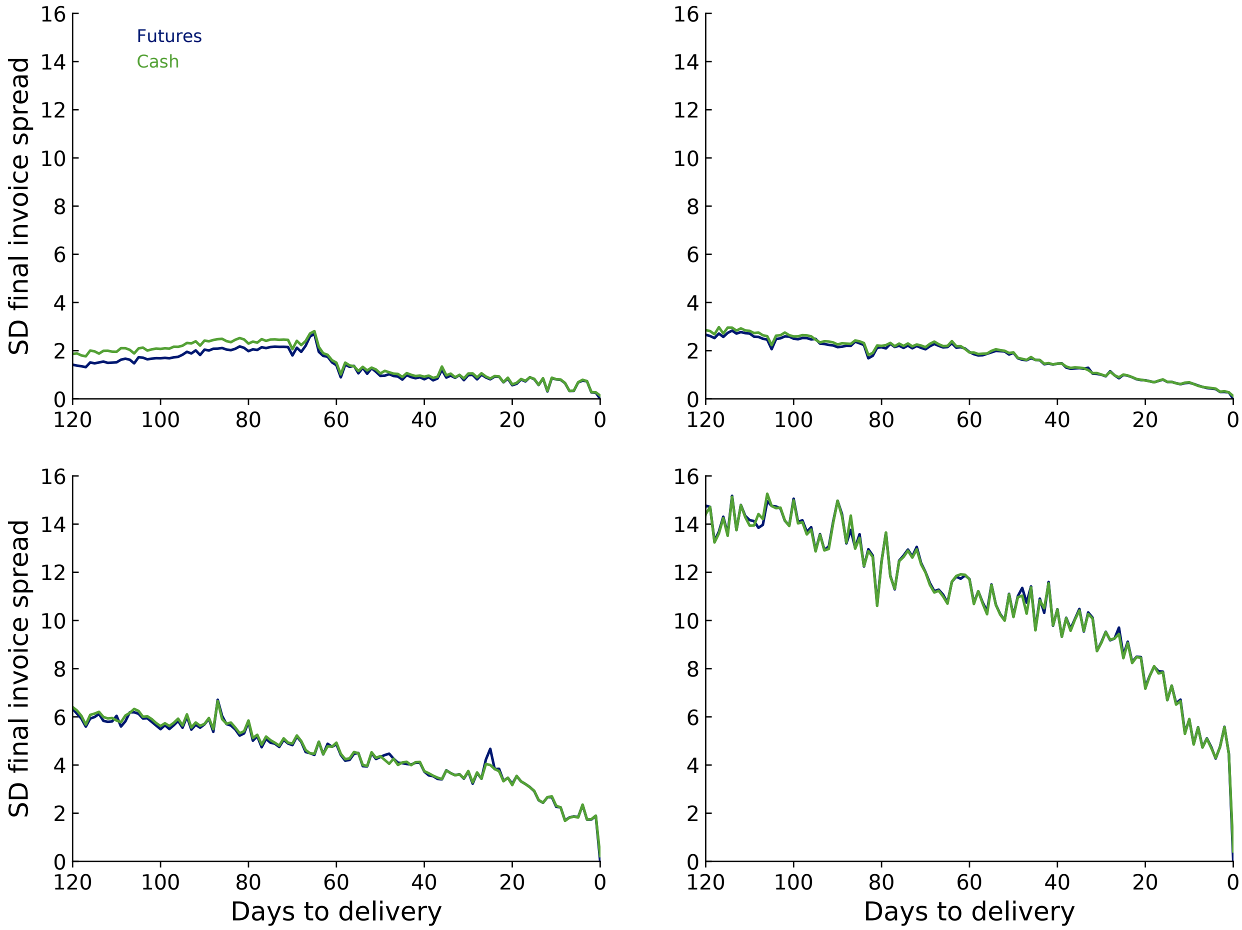


Figure B.1: **Convergence of cash and Treasury prices for the cheapest-to-deliver (variance).** Each series in this graph is the average squared deviation of the futures and cash Treasury from the last invoice price of the futures on the delivery date. The x-axis denotes days to delivery. The gray background denotes open interest in the contract, reported on the right axis.



the final invoice price of the futures on the deviation. The decline in these series indicates that not only do the average prices converge, but also the distribution converges. The equality between futures prices and note prices at delivery establishes the arbitrage relationship between cash and futures prices.

C Persistence and Predictability of Arbitrage Deviations

Figure C.1 shows a longer time-series for the plot in Figure 5. The link between cash and futures values of Treasuries has changed over time. From 1992-2020 on average futures have been undervalued relative to the underlying Treasuries. However, in the last decade, futures have become overvalued.

Table C.1 shows that these price deviations are also highly statistically significant during this period. For both the second and third to deliver contracts and for the 2-year, 5-year, and 10-

year contracts, t-statistics often exceed 20 for the periods 2010-2015 and 2015-2020. Coefficient estimates often exceed 0.10, indicating a difference of 10 basis points between the bill yield and futures implied yield.

Table C.2 of the Appendix shows that the persistence of price deviations is statistically significant using a GARCH(1,1) with an auto-regressive term for the mean process. The model shows that the means are highly persistent, nearing a random walk, though predictability declines for longer tenors. The variance of the arbitrage spreads are also highly predictable across contracts over time.

The time-series of deviations in Figure 5 suggests that the link between cash and futures values of Treasuries has changed over time. Table C.1 shows that from 1992-2020, on average futures have been undervalued relative to the underlying Treasuries. However, in the last decade, futures have become overvalued. The most relevant line for each Treasury is the second roll date, where most basis trading activity seems to be concentrated, and where there is less interference from the delivery month. Here, we can see that across note contracts, futures have been significantly overpriced relative to underlying bonds, with t-statistics regularly well above significance. These t-statistics are also particularly large in 2-year and 5-year futures, due primarily to lower volatility in mispricings at shorter maturities, and are much smaller for bond futures.

While the size of these deviations has decreased since the financial crisis, the variation of these deviations has also decreased. This pattern becomes more clear when looking only at the data since the financial crisis, when Treasury market structure and the nature of intermediation changed dramatically due to changes in financial regulation. Examining figure 5 it is clear that for note futures the cheapest to deliver has been significantly overvalued in recent years relative to the volatility of the spread.

D Hedge funds in DVP repo

In the main text, we employ data from the OFR's collection of data from the DVP repo market, and in particular the sponsored borrowing segment of this market. In this market, lenders know the exact collateral they are promised by borrowers, and unlike tri-party markets, there is no custodian bank that locks up collateral in its own account. At present, hedge funds make up the vast majority of sponsored borrowing (see Figure D.1). Their participation in DVP has expanded dramatically since the Depository Trust & Clearing Corporation (DTCC) first allowed hedge fund participation in sponsored repo in 2017.

E Determinants of Treasury futures margins

The rules for margin setting by the CBOT are somewhat opaque, and give the central counterparty some leeway in the determination of margins. Here we argue margins are well correlated with volatility in futures markets as measured by the MOVE and VIX indexes. Figure E.1 shows the level of variation margins along with the MOVE index. As can be seen, the two are highly correlated, with margins increasing in high volatility times. The highest margin periods have been the 2008 financial crisis and March 2020. However, with Treasury bonds futures, March 2020 was the highest margin period, with the second highest in 2017 (possibly associated with fluctuations around the debt ceiling). These particularly high margins on Treasury bond futures may help to explain why the bond futures contract has been relatively unpopular in recent years. In interpreting the relationship between margin on Treasury futures and the MOVE index, it once again

Figure C.1: **Deviations of Treasury prices from the replicating portfolio.** Each series graph shows the deviation of the futures-implied yield from the yield on a similar maturity bill and the open interest in that contract. Values above zero imply the replicating portfolio is overvalued relative to the cheapest-to-deliver. These series use the second-to-deliver contract.

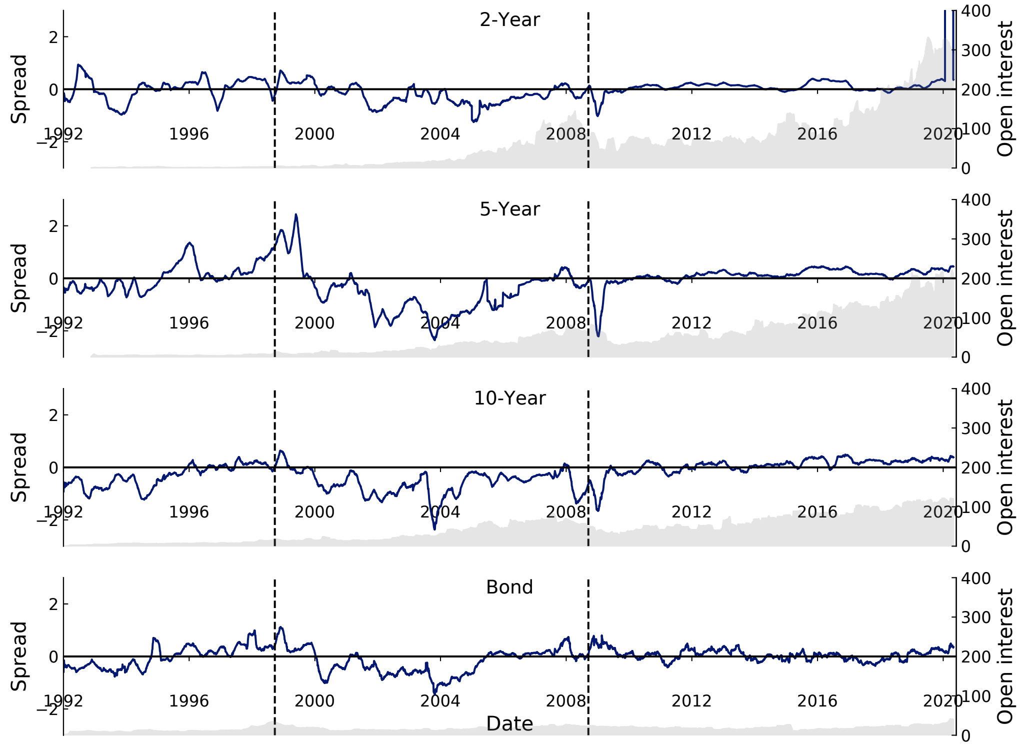


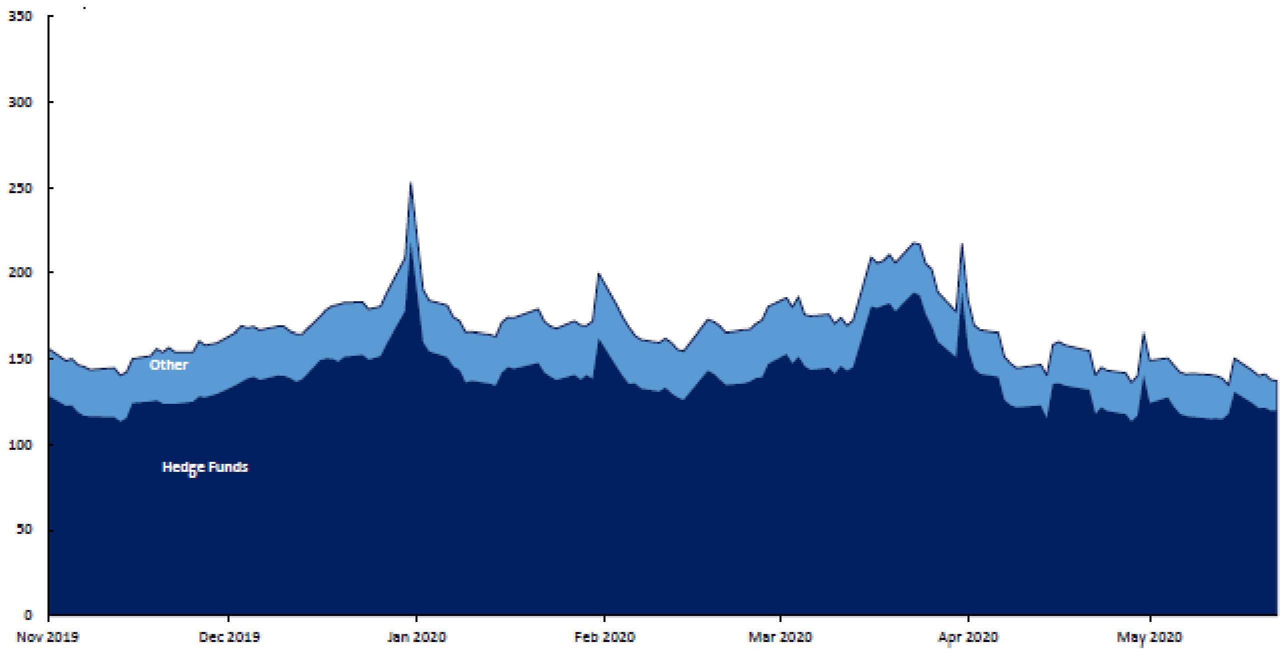
Table C.1: **Difference between futures implied yields and bill yields.** This table displays the difference between the cheapest-to-deliver price and the replicating portfolio of Treasury bills and futures by sample. In parentheses, t-statistics test the hypothesis that the average difference between these yields is zero.

Contract	Roll	1992-1995	1995-2000	2000-2005	2005-2010	2010-2015	2015-2020	1992-2020
2-Year	**1st**	-0.097	0.318	-0.024	0.193	0.211	0.246	0.138
		(-2.6)	(9.87)	(-0.84)	(7.19)	(20.65)	(18.83)	(13.01)
	2nd	-0.194	0.111	-0.612	-0.51	0.07	0.13	-0.206
		(-20.76)	(9.6)	(-26.07)	(-36.96)	(19.84)	(23.61)	(-33.16)
	3rd	-0.116	0.255	-0.398	-0.305	0.042	0.227	-0.13
		(-0.68)	(2.97)	(-3.78)	(-22.54)	(4.5)	(42.11)	(-12.14)
5-Year	**1st**	-0.264	0.753	-1.324	-0.152	0.122	0.327	-0.165
		(-7.61)	(14.91)	(-29.15)	(-3.34)	(5.21)	(14.96)	(-9.18)
	2nd	-0.508	0.295	-0.932	-0.326	0.072	0.254	-0.257
		(-48.11)	(19.22)	(-65.2)	(-18.09)	(13.34)	(44.61)	(-34.07)
	3rd	-0.567	0.144	-1.066	-0.074	0.212	0.281	-0.097
		(-30.25)	(8.76)	(-12.0)	(-4.85)	(12.03)	(29.35)	(-9.18)
10-Year	**1st**	-0.536	0.093	-0.729	-0.235	0.098	0.209	-0.239
		(-11.84)	(1.82)	(-19.68)	(-5.05)	(2.61)	(6.6)	(-13.48)
	2nd	-0.721	-0.084	-0.845	-0.631	-0.095	0.128	-0.454
		(-58.93)	(-6.65)	(-54.47)	(-35.57)	(-9.04)	(13.65)	(-63.48)
	3rd	-0.749	-0.044	-0.959	-0.511	0.207	0.098	-0.368
		(-60.8)	(-2.7)	(-25.84)	(-13.71)	(11.89)	(8.53)	(-31.48)
Bond	**1st**	-0.243	0.346	-0.655	0.209	0.054	0.04	-0.043
		(-3.69)	(4.86)	(-12.79)	(3.23)	(0.92)	(0.73)	(-1.71)
	2nd	-0.425	0.182	-0.57	0.012	0.004	0.006	-0.135
		(-30.12)	(15.38)	(-38.17)	(0.81)	(0.35)	(0.48)	(-20.98)
	3rd	-0.465	0.039	-0.555	-0.15	0.137	0.341	-0.176
		(-54.16)	(4.26)	(-48.62)	(-4.11)	(3.32)	(22.29)	(-15.61)

Table C.2: **GARCH estimates of the cash-futures arbitrage spread.** This table shows the results of a GARCH(1,1) process estimated on the arbitrage spread: the difference between the futures implied yield and the yield on a similar maturity bill. The regression uses daily observations for the second-to-deliver contract from 1992 to May 2020. For each regression, days where futures prices are unchanged from the previous day are dropped. In the top panel, we show our estimates for the AR-1 process for means, while in the bottom panel we show our GARCH process estimates for variances.

	Dependent variable: Arbitrage Spread			
	2-Year	5-Year	10-Year	Bond
		Mean model		
Constant	0.001^{***}	-0.011^{***}	-0.035^{***}	-0.03^{**}
	(0.002)	(0.003)	(0.005)	(0.012)
Lagged arbitrage spread	0.943^{***}	0.93^{***}	0.884^{***}	0.527^{***}
	(0.02)	(0.013)	(0.01)	(0.071)
		Variance model		
Constant	0.002^{***}	0.005^{***}	0.031^{***}	0.006^{**}
	(0.001)	(0.002)	(0.007)	(0.003)
Auto-regressive term	0.419^{***}	0.36^{***}	0.34^{***}	0.182^{***}
	(0.061)	(0.078)	(0.07)	(0.03)
Moving average term	0.581^{***}	0.631^{***}	0.434^{***}	0.818^{***}
	(0.07)	(0.072)	(0.075)	(0.033)
Observations	6,061	6,826	6,857	6,873
R^2	0.774	0.815	0.623	0.17
Adjusted R^2	0.774	0.815	0.622	0.17
Note:				*p<0.1; **p<0.05; ***p<0.01

Figure D.1: **DVP sponsored reverse repo by participant type ($ billions).** Data are aggregate daily transaction volumes.

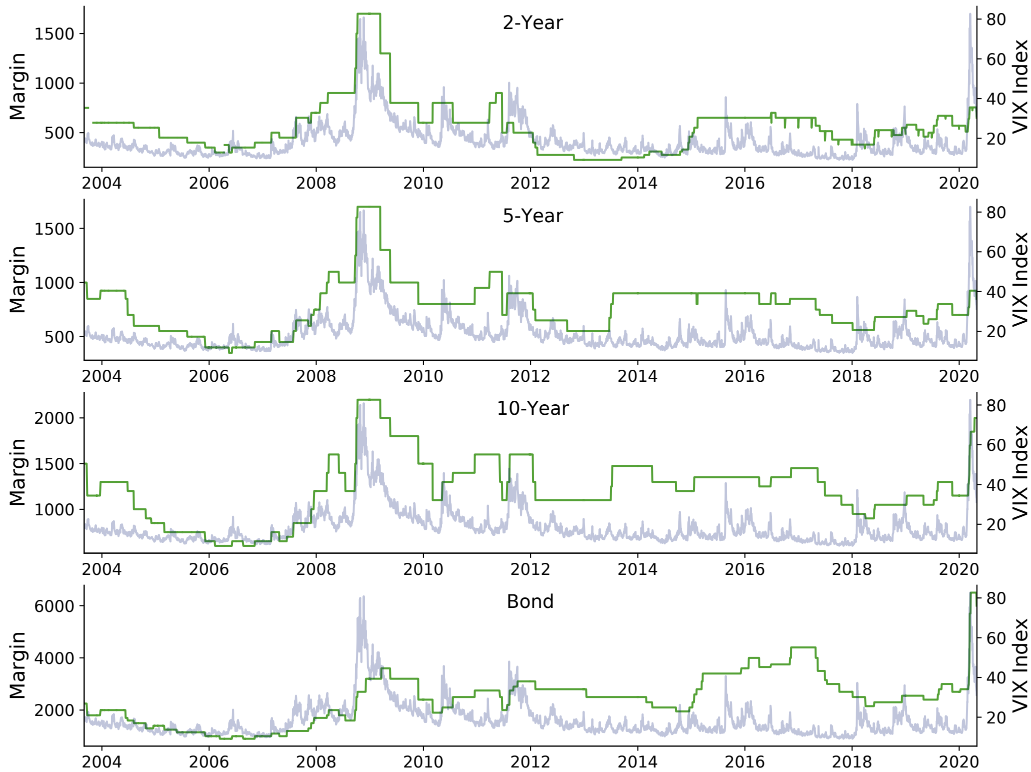


becomes important to recognize that the MOVE index is ultimately based on the Treasury futures to which the margins apply. In Table E.1, we show regressions of the maintenance margin on the MOVE index as well as the VIX index. While the MOVE index explains more of the movements in maintenance margins, the VIX index is also significant for all but the bond contract, and the explanatory power of the MOVE and VIX models are roughly similar.

Table E.1: **Maintenance margins and volatility indexes.** This table reports results for a regression of margins on the second-to-deliver Treasury futures contract on the values of the VIX and MOVE indexes.

	Dependent variable: Maintenance Margin							
Index	2-Year		5-Year		10-Year		Bond	
MOVE	6.446***		5.084***		6.609***		0.680	
	(1.119)		(0.977)		(1.167)		(2.826)	
VIX		20.745***		16.936***		24.895***		28.149**
		(4.947)		(4.015)		(3.358)		(12.864)
Observations	3,464	3,464	3,268	3,268	3,681	3,681	4,272	4,272
R^2	0.540	0.485	0.435	0.420	0.332	0.418	0.001	0.076
Note:								*p<0.1; **p<0.05; ***p<0.01

Figure E.1: **Margins on futures contracts and the MOVE index.** In green, for each contract, we show the initial margin for the second-to-deliver futures contract, which are recorded on the left axis. In light blue we show the value of the VIX, which are recorded on the right axis.



F Treasury holdings and hedge fund leverage

igure F.1 shows the relationship between hedge fund leverage and the fraction of the hedge fund's notional investment portfolio made up of Treasury securities.[31] There is a strong, positive relationship between Treasury investments and leverage. Funds with limited exposure to Treasuries have average leverage ratios near or below two, similar to the unconditional average leverage. However, as the fraction of investments held as Treasuries increases, average leverage increases substantially, nearing 10 to 1 for funds with more than 60% of their portfolio allocated to Treasuries.

G Foreign official sales

Data from the Federal Reserve's Factors Affecting Reserves, which provide a higher-frequency view of foreign official custody holdings with the Federal Reserve, suggest these sales began in the last weeks of February, as shown in Figure G.1. These sales were likely made in order to build up dollar buffers for foreign central banks for currency interventions as well as spending. The later addition of swap lines by the Federal Reserve allowed foreign official accounts to build these buffers without the need for greater selling of Treasuries.

Sales from these foreign official accounts may have had particular importance for Treasury market illiquidity. Primary dealers are required to make "reasonable" markets for sales of Treasuries by these accounts. Additionally, the funds from these sales seem to have been invested in significant part in the Federal Reserve's foreign repo pool, which would have affected the availability of reserves. When a domestic agent sells Treasuries to a dealer and invests the proceeds in a domestic bank account, these funds can still be made available to the dealer to fund the Treasury purchase through the repo market. When a foreign seller invests the proceeds of a sale into the foreign repo pool, reserves are effectively removed from the system, potentially making repo financing of Treasuries more expensive. While large increases in reserves provided by the Federal Reserve may have blunted the impact of the pool on the availability of funding, on the margin the foreign repo pool may still have had a deleterious impact on Treasury liquidity by making repo balances more expensive.

[31] Notional values are calculated as market values, except for equity derivatives which are delta-adjusted, and interest rate derivatives, which are reported as 10-year bond equivalents.

Figure F.1: **Hedge fund balance sheet leverage vs. Treasury exposure.** This figure shows a binned scatter of hedge fund gross notional Treasury exposure as a percentage of total gross notional exposure against hedge fund leverage (defined as gross asset value over net asset value). Data are pooled from 2013 to 2020. Each point represents a percentile group of Treasury exposure, for which averages are calculated for both leverage and percentage gross Treasury exposure. The dark blue line represents a linear fit.

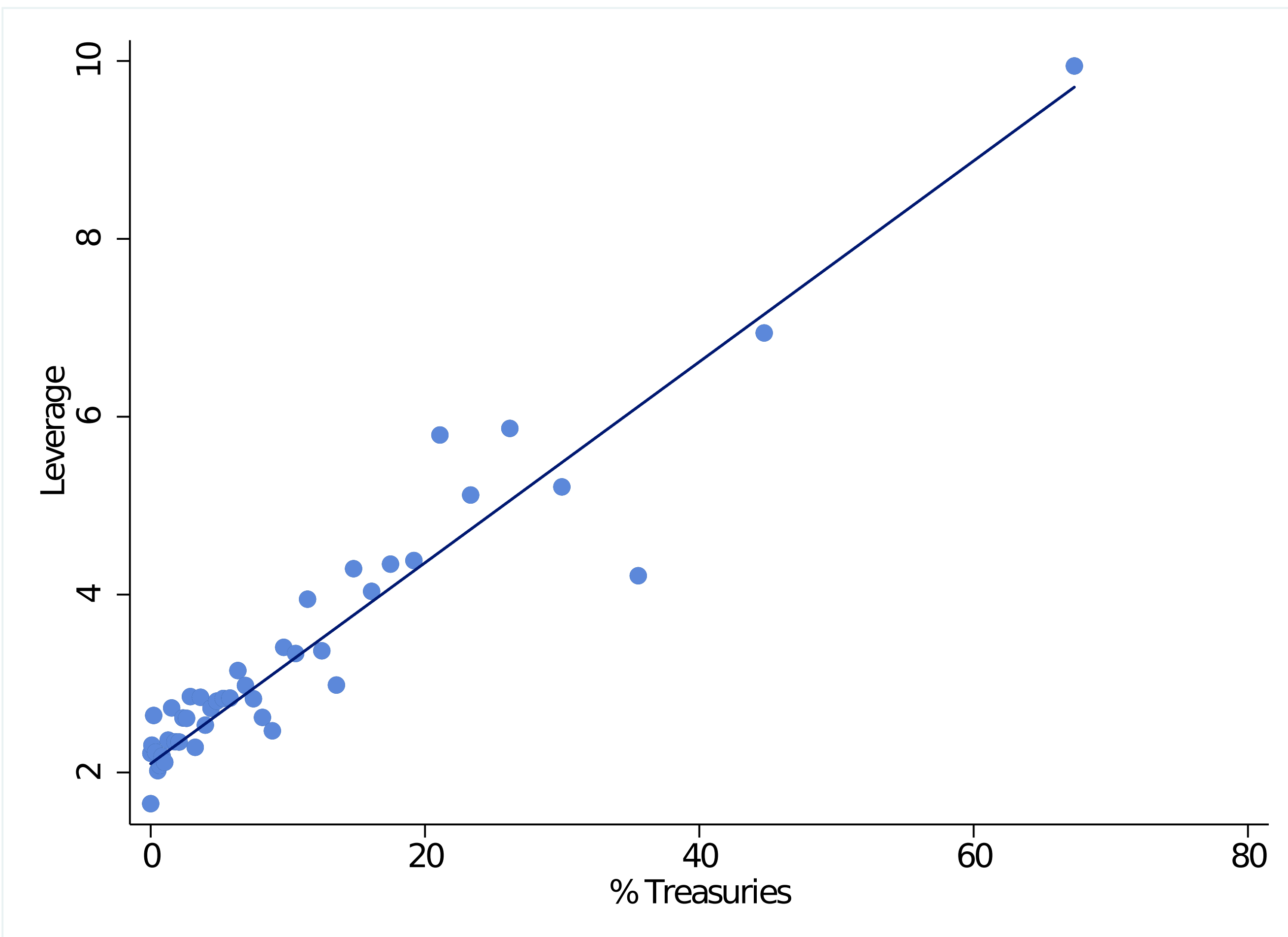


Figure G.1: **Foreign official sales and dollar liquidity.** This figure shows foreign official Treasury holdings, swap lines, and investments into the foreign repo pool as reported in the Federal Reserve's Factors Affecting Reserves release. All values are differences from their values as of March 1st, 2020.