Finance and Economics Discussion Series
Divisions of Research & Statistics and Monetary Affairs
Federal Reserve Board, Washington, D.C.

Bond Market Intermediation and the Role of Repo

Yesol Huh and Sebastian Infante

2017-003

Please cite this paper as:
Huh, Yesol, and Sebastian Infante (2020). "Bond Market Intermediation and the Role of Repo," Finance and Economics Discussion Series 2017-003. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2017.003r1.

Bond Market Intermediation and the Role of Repo*

Yesol Huh Sebastian Infante

January 31, 2020

Abstract

We model the role that repurchase agreements (repos) play in bond market intermediation. Not only do repos allow dealers to finance their activities, but they also enable dealers to source assets without taking ownership. When the asset trades with *repo specialness*, i.e, when the associated repo rate is significantly below prevailing market rates, borrowing the asset is more expensive, resulting in higher bid-ask spreads. Limiting a single dealer's leverage decreases its market-making abilities, so the dealer charges a higher bid-ask spread. However, limiting all dealers' leverage reduces pressure on repo specialness, thus *decreasing* bid-ask spreads. More generally, this paper characterizes the importance of collateralized borrowing and lending for bond market intermediation, shows how frictions in repo markets can affect the underlying cash market liquidity, and underscores the importance of securities lending.

*This paper was previously circulated under the title "Bond Market Liquidity and the Role of Repo." We thank Giovanni Favara, Erik Heitfield, Stefan Nagel, Andreas Rapp, Jean-David Sigaux, Adrian Walton, participants at the 12th Annual Central Bank Conference on Market Microstructure of Financial Markets, 2018 Federal Reserve Research Scrum and seminar participants at the Bank of Canada, the Federal Reserve Board, and Vanderbilt University for their helpful feedback, and Joseph Shadel for excellent research assistance. The views of this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of any other person associated with the Federal Reserve System. Federal Reserve Board, 20th St. and Constitution Avenue, NW, Washington, DC, 20551. Corresponding author: Sebastian Infante, email: *sebastian.infantebilbao@frb.gov.*

1 Introduction

A well-functioning market for repurchase agreements (repos) is often cited as being crucial for U.S. Treasury bond markets, but the precise way in which these contracts facilitate market making is not well understood. This paper fills the gap by providing a theoretical framework that captures how dealers use repos to intermediate fixed income markets. Our model gives insight into how frictions in repo markets can affect the underlying cash market liquidity, how restrictions on dealers' balance sheets can affect their ability to intermediate the cash market, and the important role that securities lenders play.[1] The theoretical framework in this paper provides guidance on how to interpret data on dealers' cash and repo positions in the context of bond market intermediation.

We build a model in which dealers use repo markets to intermediate leveraged client order flow. Dealers access repo markets for three reasons: to finance their position and that of their counterparties, to take short positions, and, finally, to source assets for delivery without altering their cash portfolio position. In our model, all three functions of repos arise endogenously. While the first two roles of repo are well known, the third has not been captured by the existing literature, and we show that this role is a key ingredient for cash market intermediation.

In our model, as in practice, dealers heavily rely on repo markets to intermediate client order flow. We show how frictions in the repo market can spill over to the cash market. Specifically, *repo specialness*—the cost associated with sourcing a specific security rather than a broad class of securities—is correlated with bond market liquidity. We also characterize how balance sheet constraints can directly impact dealers' ability to make markets, highlighting the potential impact bank regulations can have on market liquidity. Specifically we show that from the perspective of an individual dealer, a size restriction limits intermediation capacity and, hence, the dealer's willingness to intermediate markets. However, somewhat surprisingly, we find that these types of restrictions can in fact *reduce* intermediation costs, as they alleviate pressures on repo specialness.

The model features a continuum of dealers that access three distinct markets—an interdealer cash market, a general collateral (GC) repo market, and a specific-issue (SI) repo market—to intermediate leveraged trades for their clients. The interdealer cash market allows dealers to buy

[1] Throughout the paper, the term *cash market* refers to the spot market for securities.

and sell securities outright. The GC repo market enables dealers to borrow or lend funds on a secured basis. Finally, the SI repo market allows dealers to borrow or lend specific securities using cash collateral. Scarcity of specific securities can lead to repo rates in the SI market trading below that of the GC market, an occurrence referred to as the SI market trading *special*. Repo specialness refers to the difference between the GC and SI repo rates. We assume that the securities lending sector is the net supplier of securities in the SI repo market and profits whenever repos trade special. Importantly, clients cannot access these markets directly, so they rely on dealers to intermediate their trades.

A key restriction that incentivizes dealers to use repo markets is the *box constraint*. Broadly speaking, the box constraint is a physical restriction that forces dealers to have access to securities, either by owning or borrowing them, to deliver to counterparties. This constraint can be interpreted as a budget constraint for securities, where the *box* refers to the net amount of securities owned and sourced. Because the box constraint forces dealers to access securities to make markets, the net supply of securities from securities lenders is just as important as the net supply of cash from cash lenders. In the model, we consider two types of box constraints. The first, the *global box constraint*, posits that the amount of securities sourced or delivered through the cash, GC, and SI repo market must be greater than or equal to zero. The second, the *SI box constraint*, is stricter than the global box constraint in two dimensions. First, SI box excludes the securities sourced and delivered through GC repo markets. Second, the SI box has to be no less than an amount proportional to clients' leveraged orders.

To satisfy the SI box constraint, dealers can either buy the securities from the interdealer cash market or borrow them from the SI repo market; however, the impact on dealers' portfolio payoff will be different. Namely, by purchasing a security, a dealer alters the risk of its portfolio, whereas when borrowing a security, his risk profile remains unchanged.[2] In the model, the dealer is able to post the additional amount of securities needed to satisfy the SI box in the GC market in order to finance them. But it entails a cost: the dealer will raise funding at the GC repo rate which is higher than the SI repo rate.

[2]The focus of the paper is to characterize how dealers intermediate markets, and, thus, we abstract from counterparty risk.

In equilibrium, dealers choose not to alter their optimal portfolio position, but instead heavily use repos to intermediate client order flow. This result implies that their inventory is an incomplete measure of their intermediation capacity and also that their flexibility to increase the size of the balance sheet is crucial. The model also suggests that repo specialness is an intermediation cost that dealers must bear and is thus correlated with the bid-ask spreads they charge their clients. This insight implies that, in the time series, repo specialness is an indirect measure of bond market liquidity.[3] This result is particularly useful in markets where cash market liquidity is difficult to measure directly. For example, data from the interdealer U.S. Treasury cash market suggest that liquidity has been fairly stable over the past few years, whereas specialness data suggest otherwise.

The SI box constraint is a key friction of the model and is motivated by a number of economic incentives. For example, failing on a promise to deliver a security entails a cost, which generates a motive to source in additional amount of securities.[4] In order to avoid the cost of failing to deliver, dealers have an incentive to hold or source additional securities in case a counterparty fails to deliver to them. Without any buffer, dealers would be forced to fail to deliver to their counterparties when others fail to deliver to them. Therefore, dealers have a precautionary motive to source in more securities than are needed—specifically, in proportion to the amount of assets it has to intermediate.[5,6]

Inspired by newly implemented regulations, we study how the equilibrium changes when limits to the size of dealers' balance sheet are introduced. Such restrictions impact cash market liquidity because the use of repos in cash market intermediation expands the size of dealers' balance sheets. In a partial equilibrium setting where a single dealer faces a size limit, all else equal, the affected dealer has less incentive to attract large trades from clients; hence, the dealer increases its bid-ask spread. However, in a general equilibrium setting where all dealers face the balance sheet restriction, bid-ask spreads decrease. This seemingly counterintuitive result arises because such a

[3]This result is seemingly at odds with previous literature which finds that repo specialness manifests itself in the most liquid securities. But those finding are about the relative liquidity across different types of bonds, for example, on-the-run vs off-the-run Treasury securities. See Subsection 7.2 for a more detailed discussion.

[4]The cost is proportional to the repo rate or a fails charge imposed by market convention. For more information on the history of the fails charge see Garbade et al. (2010).

[5]Subsection 7.1 provides suggestive empirical evidence of the securities buffer.

[6]The SI box constraint may also be justified by different intraday settlement timing between various repo and cash markets. Additionally, without such a constraint, a dealer may engage in an infinite number of leveraged orders.

restriction reduces dealers' demand to source assets in the SI repo market, alleviating pressures on repo specialness. As specialness is a cost of intermediation, lower costs will get passed on to clients in the form of lower bid-ask spreads.

Comparing our general equilibrium result with recent data suggests that the introduction of the Supplementary Leverage Ratio (SLR) cannot explain the increase in specialness in recent years. In our model, balance sheet restrictions *decrease* specialness, whereas it has increased modestly in recent years. This finding suggests that other factors are in play. We conjecture that changes in securities lenders' willingness to provide SI repos is a potential explanation, which is supported by our model.

Although the specific modeling ingredients of the paper are inspired by the U.S. Treasury market, the insights from the model can be applied more generally to securities that have dealer-intermediated cash markets and active and liquid repo markets. For example, many of the European government securities markets exhibit these characteristics.

The paper is structured as follows. Section 2 gives some background knowledge on U.S. Treasury market structure and empirical observations that motivate our framework, and Section 3 reviews the related literature. Section 4 describes the model setup, detailing the main agents and how they interact. Section 5 presents the symmetric equilibrium and its interpretation when dealer balance sheets are unrestricted. Section 6 incorporates restrictions on the size of dealers' balance sheet and shows how these restrictions affect their markup decision. Section 7 discusses the model assumptions and implications, and Section 8 concludes.

2 Institutional Setting

The cash Treasury market in the United States is segmented into interdealer markets and dealer-customer markets. Interdealer trades mostly takes place on fully electronic, anonymous limit order book platforms. Dealer-customer trades take place over-the-counter or on "request for quote" platforms such as TradeWeb or Bloomberg with each customer generally trading with one or a few dealers, and they account for about half of the trading volume.[7] The U.S. Treasury market has a core-periphery structure, with securities broker-dealers in the core and their customers in the periphery. According to recent analysis using the newly available Trade Reporting and Compliance Engine (TRACE) data, client trades account for only about 10% of volume on interdealer brokers, thus pointing to a stark segmentation between interdealer and dealer-customer markets.[8]

In a recent speech, Federal Reserve Governor Brainard noted that "dealers generally—and primary dealers particularly—still intermediate the majority of activity in all Treasury securities" (Brainard (2018)). These dealers also provide funding to their clients, so they are a vital part of the Treasury market intermediation. Thus, changes to their business models or regulations potentially impacts the Treasury market.

The U.S. Treasury repo market can be divided into the tri-party repo market and the bilateral repo market. The tri-party repo market is a wholesale funding market where broker-dealers raise short-term secured funding from cash investors such as money market funds.[9] In this repo market, within a certain collateral class, cash borrowers have the flexibility to choose from a wide range of assets to post as collateral. This flexibility is valued by cash borrowers, who can manage their collateral holdings by exchanging assets within a collateral class whenever different asset demands arise. This flexibility is not costly to borrowers because in this market cash lenders only value collateral as a backstop to a borrower default. Given an appropriately sized haircut, cash lenders are largely indifferent to which specific asset is posted as collateral. Thus, the tri-party market is a GC repo market.

In the bilateral repo markets, collateral can be specified—allowing participants to borrow a particular security via a reverse repo. This feature makes the bilateral repo market an SI repo market.[10] Because a large fraction of cash trading in the Treasury market happens on the "on-the-

[7] For example, Kruttli et al. (2018) finds that the median hedge fund has two prime brokers, and Klaus and Rzepkowski (2009) finds that 90% of hedge funds have one prime broker.

[8] "Unlocking the Treasury Market Through TRACE," `https://libertystreeteconomics.newyorkfed.org/2018/09/unlocking-the-treasury-market-through-trace.html`

[9] Tri-party repo markets can be further separated into two distinct segments: the general tri-party market and the GCF (general collateral financing) repo market. The general tri-party allows cash lenders, such as money market mutual funds, to enter into repo contracts with large broker dealers. The GCF repo market is a blind brokered interdealer market allowing broker dealers to manage their cash and collateral. For more details, see Copeland et al. (2012).

[10] Of note, dealers may also use the bilateral market to raise funding rather than to borrow specific securities.

run" securities (Barclay et al. (2006))—the most recently issued U.S. Treasury bond—an important fraction of SI repo market activity is for the on-the-run securities. When demand for borrowing a specific issue is high and borrowers compete by offering cash collateral at a reduced rate, the bilateral repo rate can be lower than the tri-party repo rate. This rate difference incentivizes the original collateral owners to lend their securities in the SI market, reinvest the cash proceeds into the GC repo market, and reap the difference between the GC and SI repo rates—i.e., repo specialness. Collateral owners who participate in this market are typically long-term investors who lend their securities through securities lenders to profit from repo specialness.

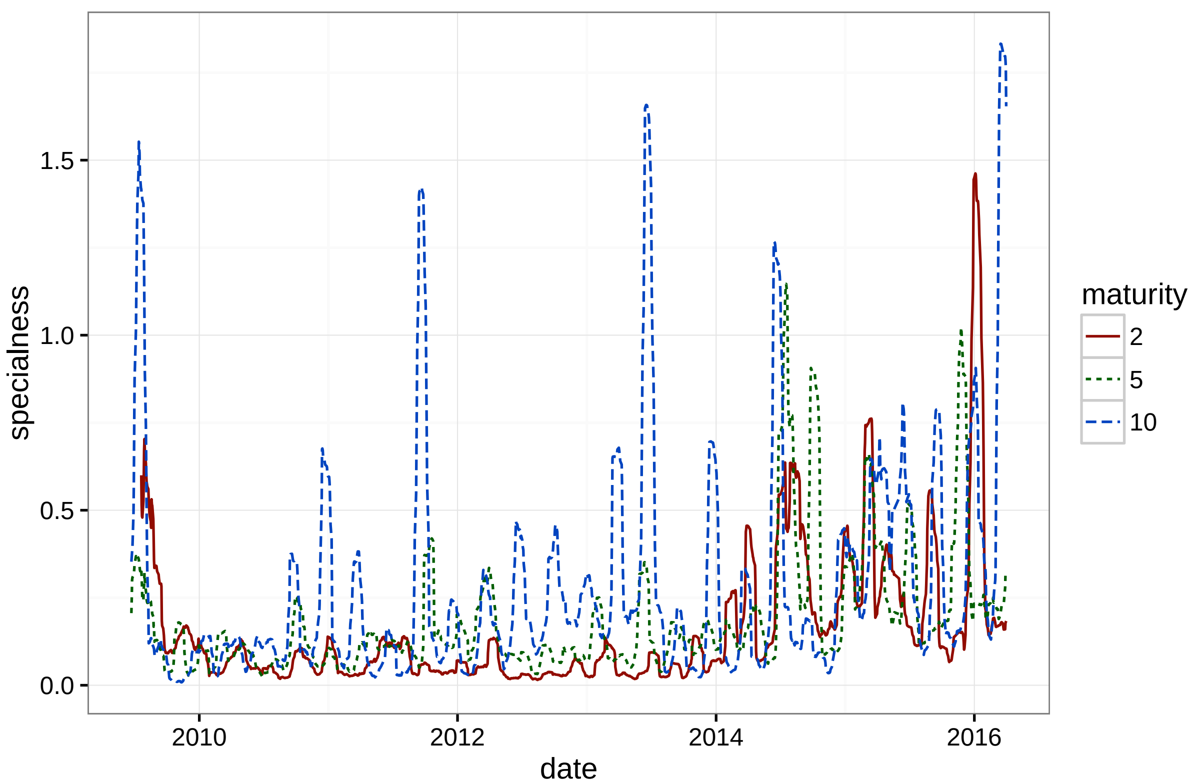


Figure 1: Repo specialness for on-the-run Treasuries
This figure presents the 20 day rolling average of repo specialness for the 2, 5, and 10 year on-the-run Treasury bonds. Repo specialness is calculated as the difference between the overnight general collateral repo rate on U.S. Treasuries and the volume-weighted average of the specific-issue repo rate. A more positive number refers to higher specialness. General collateral rates are from DTCC, and the specific issue repo rates are from the repo interdealer broker community.

Figure 1 plots the 20 day moving average of repo specialness for on-the-run Treasury securities, showing a clear upward trend in repo specialness. Because our model predicts that repo specialness and bond illiquidity are correlated, this evidence is consistent with market commentary that Treasury markets have become more illiquid. Given that bid-ask spread data is generally only available for the interdealer segment of the U.S. Treasury market, it is not surprising that illiquidity has been hard to detect. This paper shows that the overall market can be illiquid while exhibiting limited evidence of illiquidity in the interdealer market.

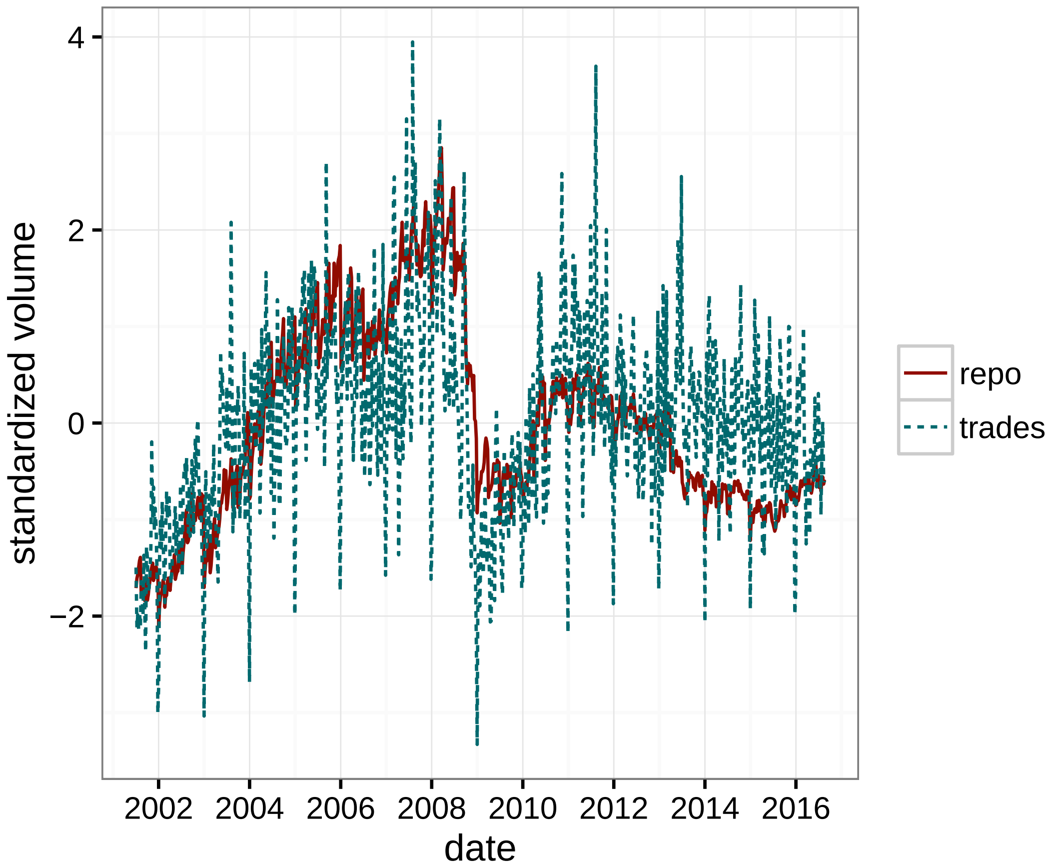


Figure 2: Trading and repo volume for Treasuries
This graph plots the standardized trading volume and repo volume for Treasury securities over time. Both are calculated using the primary dealer statistics data on the NY Fed website. We take the sum across all Treasury notes and bonds, and for repo volume, we include both repos and reverse repos.

In practice, dealers engage in repo transactions across markets not only to finance their positions (and those of their clients), but also to distribute collateral throughout the financial system. Figure 2 shows primary dealers' aggregate repo and trading volumes in the U.S. Treasury market. The high correlation between these two series is suggestive of how important the repo market is for trading activity. Adrian et al. (2013) highlights the importance of repo for bond market intermediation, and Infante et al. (2020) provides evidence of the high degree of collateral use and reuse in the US Treasury market through repo.

3 Literature Review

Our paper is related to the literature on repo specialness in the U.S. Treasury market and its effect on pricing. Duffie (1996) shows that the degree of repo specialness depends on the demand for short positions and the supply of loanable collateral and provides a theoretical relationship between bond prices and repo specialness. Krishnamurthy (2002) studies the spread between on-the-run and off-the-run U.S. Treasuries and shows that trading profits from shorting on-the-run securities and buying off-the-run securities are close to zero because of the on-the-run securities' specialness. The model in that paper incorporates frictions on agents' liquidity needs and their ability to lend securities in order to disentangle changes in bond prices from changes in their special repo rates. In our model, bond prices and special repo rates are determined uniquely because of the participation of a third party that supplies collateral in the SI repo market: securities lenders. In our model, securities lenders only participate in the GC and SI repo markets, allowing us to uniquely determine the bond price and its repo specialness.[11]

Related to Duffie (1996), Graveline and McBrady (2011) show that demand for hedging interest rate risk affects the degree of specialness, and Banerjee and Graveline (2013) decompose the impact of short sellers versus long investors on liquidity premium. Keane (1996) finds that repo specialness increase with the time since the last auction and conjectures that this phenomenon is due to a greater fraction of on-the-run Treasuries being locked up in buy-and-hold investors' portfolios. Overall, these papers tend to focus on end investors' impact; in contrast, our focus is on the impact of dealers' balance sheet and securities lenders willingness to lend.

The on-the-run/off-the-run yield difference is often used as a measure of liquidity in Treasury

[11] Relatedly, DAmico et al. (2018) studies the supply and demand factors that drive repo specialness, focusing on the Federal Reserve's purchase programs.

markets, as the price difference may be due to the liquidity premium embedded in the more liquid on-the-run security. Vayanos and Weill (2008) argues that repo specialness of the on-the-run securities arises naturally to alleviate search frictions that short sellers face when they need to return a borrowed asset in the future. This perspective implies that on-the-run specialness is a result of the security's higher liquidity. The framework developed in our paper adds to this insight by noting that even though specialness reflects liquidity *across assets*, it contributes to illiquidity *across time* because it exposes dealers to intermediation costs. This insight implies that repo specialness is a proxy for market liquidity.

In terms of the model setup, our model is closest to the literature on how market makers' inventory management affects their liquidity provision. Amihud and Mendelson (1980) models a monopolistic, risk-neutral dealer with inventory size constraints, Stoll (1978) models a risk-averse monopolistic dealer, Ho and Stoll (1983) models a market with multiple dealers, and Ho and Stoll (1983) incorporates an interdealer market to offset the inventory. The main difference between this literature and our model is how dealers source securities. In our model, dealers can access securities via two distinct markets, by either buying or borrowing them, which means that dealers' inventory positions are not representative of their intermediation capacity. Their ability to expand their balance sheet without altering the risk of their portfolio is more relevant in determining their intermediation capacity.

This paper is also related to the large literature on trading frictions in securities markets. In particular, Bottazzi et al. (2012) show how asset and repo markets coexist in a stylized general equilibrium framework. In that paper, the authors underscore a particularly relevant restriction that securities dealers must satisfy: the box constraint. This constraint forces intermediaries to borrow a security whenever they want to short. Our paper focuses on a market structure where clients pay dealers to service trades, allowing us to gauge the amount of liquidity dealers provide. This structure give us a framework to understand how dealers intermediate markets and to interpret dealers balance sheet data.

Number of papers study how the SLR rule has impacted the repo market. Duffie (2017) argues that SLR would disincentivize dealers from providing liquidity in the repo market, especially for repos with low-risk collateral such as the U.S. Treasury securities. Allahrakha et al. (2018) and Kotidis and Van Horen (2018) empirically show that dealers that are affected by the SLR decreases repo volume. The aim of our paper is to study how this constraint on dealers' repo activities would then spill over to cash market liquidity.

Relatedly, there is a growing literature aiming to understand how newly implemented bank regulation has affected liquidity in securities markets. For example, Trebbi and Xiao (2017), Bao et al. (2018), and Choi and Huh (2017) study whether corporate bond liquidity has deteriorated since regulations were adopted. Cimon and Garriott (2019) build a model to study the impact of liquidity, capital, and position constraints on banks' ability to make markets. Their focus is to study how regulations promote the entry of new non-regulated dealers. Our focus is to explicitly characterize the link between the cash market and different repo markets and to study how the leverage ratio, which is particularly onerous on repo, affects the cash market.

Lastly, securities lenders also play an important role in the repo market. Foley-Fisher et al. (2015) argue that securities lenders are not merely responding to the demand to borrow securities, but they also use securities lending as a way to finance higher-yielding, less liquid securities. To this end, we model the supply of assets from securities lenders as not only responding to changes in repo specialness, but also as having a free parameter that reflects their willingness to supply assets. We look at how the equilibrium changes with their willingness parameter.

4 Model Setup

The model consists of three periods $t \in \{0, 1, 2\}$. In $t = 0$, dealers post their bid and ask to clients. Each dealer faces one client, and each client submits either a levered buy or a levered sell order with equal probabilities.[12] That is, a client order will be a simultaneous cash and repo trade with the dealer, establishing either a levered long or a short position.[13] The size of clients' orders are

[12] Alternatively, we can assume that each client is infinitesimally small and that dealers face order flows from a mass of clients.

[13] In practice, there are various contracts investors use to take on leverage for their positions. For example, through securities lending, margin loans, or prime brokerage accounts. Although the regulatory treatment of these contracts may differ, they are economically equivalent. Infante et al. (2020) shows that repos are the most prevalent secured funding contract for US Treasuries.

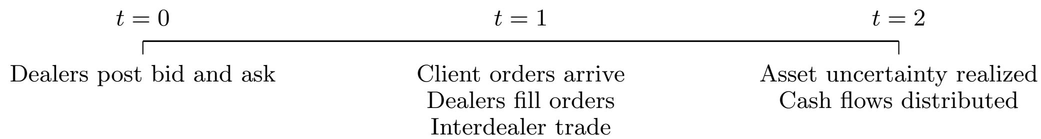


Figure 3: Model timeline

stochastic, with larger orders being more likely for smaller bid and ask spreads. At $t = 1$, client orders are realized. Throughout the model, we will assume that SI repos trade special, which in the model will imply that dealers have to bear a cost to source a specific asset.

4.1 Assets & Contracts

There is one type of risky asset, with an exogenous supply C available to dealers in $t = 1$, and final payoff given by $\tilde{v} \sim N(\mu, \sigma)$. We assume that there is an unrestricted secured lending market for GC repos with an exogenously specified one-period risk-free rate R. This means there is an abundance of cash and collateral that could be used in GC repos from outside investors. Only secured debt is allowed (i.e., repo) to raise funding.

The interdealer price of the risky asset p and the specific issue repo rate R^{SI} will be determined by market clearing. We will consider settings where in equilibrium, the repo rate on SI repos is below the risk-free rate: $R^{SI} < R$, and the difference between rates (i.e., repo specialness) is denoted by $s = R - R^{SI}$. The main difference between SI and GC repos is that dealers can use SI repos to establish short positions and deliver them to their clients.[14]

For simplicity, we assume that repos do not have haircuts, yet are risk-free. More specifically, we assume contract enforceability and no limited liability. These assumptions simplify the analysis greatly and keep the focus on dealers' use of repo to borrow assets in order to deliver them to clients. We can relax the assumption of risk-free repos somewhat by incorporating haircuts to make GC repos "virtually risk free" and by allowing dealers to charge different haircuts to their

[14]While in the model GC and SI repos appear similar because there is only one risky asset, in reality, the main difference between them is that the collateral cannot be specified in a GC repo, as elaborated in Section 2. To capture this difference, we assume that dealers cannot use GC repos to source assets to deliver to counterparties, nor to satisfy the SI box constraint.

clients, depending on whether they submit a levered long or short sale.[15,16] Adding these features complicates the analysis without significantly altering the results.[17]

In terms of notation, a repo using Q assets as collateral implies that the cash borrower receives pQ in funds and distributes Q in risky assets to the cash lender. On the closing leg of the repo, the cash lender of a repo receives RpQ or $R^{SI}pQ$ in cash, depending on whether it was a GC or SI repo, and returns the asset back to the cash borrower.

4.2 Agents

There are three types of agents in the economy: dealers, securities lenders, and clients. Each client interacts with only one dealer, and each dealer interacts with only one client. Hence, there is no outright competition between dealers.[18] Dealers service their client orders, which are levered long or short positions. Each client submits a levered long or short order, and same dealer services both legs (the outright buy/sell and the reverse repo or repo) of the order. Securities lenders are long-term investors who already hold their optimal portfolio in $t = 0$ and have the ability to lend their securities to dealers.

4.2.1 Clients

Each client submits either a levered long order of size $\tilde{q}_L$ or levered short order of size $\tilde{q}_H$ with equal probabilities. Hence, each client order involves two transactions for the dealer—for example, a levered long order of size $\tilde{q}_L$ involves an asset sale and reverse repo, both of size $\tilde{q}_L$, for the dealer. Client order sizes are independent, stochastic, and are drawn from an exponential distribution with

[15]The case of virtually risk free can be interpreted as repos with U.S. Treasuries that have a 2% haircut.

[16]Infante (2019) shows that SI repo haircuts can be negative whenever investors need to source an asset.

[17]Ignoring counterparty risk implies that the current version of the model is not well suited to study systemic risks that come from rehypothecation.

[18]See Section 2 and footnote 7.

intensity $\lambda(\cdot)$, which is a function of the dealer's bid spread b or ask spread a. That is,

Client submits long order	Client submits short order
$\tilde{q}_L \sim Exp(\lambda(a))$	$\tilde{q}_H \sim Exp(\lambda(b))$
$\tilde{q}_H = 0$	$\tilde{q}_L = 0.$

Client orders are realized at the start of $t = 1$ (i.e., in $t = 1$, $\tilde{q}_L = q_L$ and $\tilde{q}_H = q_H$), which dealers then intermediate. We assume λ, λ', and $\lambda'' \geq c > 0$ where c is an arbitrarily small constant. This setup implies that higher dealer markups make larger trade sizes less likely. Clients can be interpreted as liquidity traders who can only access one dealer and are price sensitive to the dealer's markup.[19]

4.2.2 Dealers

There is a continuum $[0, 1]$ of dealers with constant absolute risk aversion (CARA) utility and risk aversion γ that consume the payoff from their portfolio in the final period. For simplicity, we assume that dealers start with zero initial wealth and no assets. After dealers receive client orders, dealers transact in the interdealer cash and repo market, and C assets are available to the dealers.[20]

Dealers' final payoff consists of cash flows from their risky investments, payoffs from their repos, and fees in the form of bid-ask spreads from clients. Therefore, the dealer's final wealth takes the following form:

$$\begin{aligned} \tilde{W}_2 &= (\tilde{v} - p)Q_C + (R^{SI} - 1)pQ_{SI} + (R - 1)pQ_{GC} \\ &\quad -(\tilde{v} - R^{SI}p)q_L + (\tilde{v} - R^{SI}p)q_H + apq_L + bpq_H \end{aligned} \tag{1}$$

The first line in expression (1) highlights the impact of the dealer's rebalancing on its portfolio.

[19] Given the exogenous specification of clients' order intensity, we could directly model dealer competition by incorporating the impact of other dealers' markups on λ. This would unnecessarily complicate the model without providing new insights.

[20] One interpretation is that there is a primary issuance of C units in $t = 1$. Alternatively, we can assume that each dealer is endowed with C assets and that market clearing in the interdealer market is equal to zero. For the purposes of the model, these assumptions are equivalent.

Here Q_C is the dealer's portfolio rebalancing in the cash market, Q_{SI} is the amount of assets sourced through SI repos, and Q_{GC} the amount of assets received as collateral through GC repos (i.e., cash lending); all are chosen in $t = 1$.[21] Negative (positive) Q_{SI} implies that the dealer lent (borrowed) assets through the SI repo market, and negative (positive) Q_{GC} implies that the dealer borrowed (lent) cash in the GC repo market. Equation (1) highlights the main difference between sourcing assets through the cash market or repo markets: in $t = 2$, Q_C is exposed to the stochastic payoff $\tilde{v}$, while Q_{SI} and Q_{GC} pay known rates R^{SI} and R.

The second line in expression (1) highlights the effect of the client's order on the dealer's portfolio. The size of the client's buy and sale orders are given by q_L and q_H, which are accompanied by their respective repo orders. Since each dealer receives exactly one client order, we have either $q_L > 0$ and $q_H = 0$ or $q_L = 0$ and $q_H > 0$. Note that for the dealer a client buy order implies a negative cash position $-(\tilde{v} - p)$ and positive repo position $(R^{SI} - 1)p$, resulting in an aggregate position $-(\tilde{v} - R^{SI}p)$. Finally, apq_L and bpq_H are the profits from the bid b and ask a spreads. The dealer chooses a and b at $t = 0$, and, as we will show when describing dealers' clients, the tradeoff is between spreads and the intermediation quantity. Note that in this setup, dealers charge their clients in the form of bid-ask spreads, but the model is equivalent to a case where dealers use other forms of contracting arrangements to charge their clients, such as repo markups.

The main restrictions for dealers are their global and SI box constraints. The total amount of assets (SI and GC) must always be non-negative, and the amount of SI assets the dealer can access must be large enough to cover the clients' levered orders. Specifically, we assume dealers' SI box restriction is a function of levered client orders: $g(q_L, q_H) > 0$. This assumption captures the intuition that dealers need to, at least in part, access assets in order to establish client positions.[22] Specifically, given client order sizes q_L and q_H, the dealer's SI box constraint is given by

[21] As our model is a one-shot trading game, there is no rollover risk. If repo trading happens in multiple periods, dealers face the risk that the cash borrower may not want to roll over the overnight repo or roll over at unfavorable rates, which will induce the dealers to use a mix of overnight and term repos for precautionary reasons. However, in our model, because dealer demand for sourcing securities through SI repo markets is temporary (that is, they need to source in the assets for settlement but do not need continuous access to it), including rollover risk should not have a material impact on the model outcome.

[22] In effect, if $g = 0$, dealers would be able to write "naked" shorts and longs with their clients. See detailed discussion of this important modelling assumption in section 7.1.

$$Q_C + Q_{SI} \geq g(q_L, q_H). \tag{2}$$

That is, the total amount of SI assets the dealer can access—either through assets bought in the interdealer market or assets sourced via SI repos—must be enough to establish their levered clients position, which is a function of g. Note that with levered long (short) orders, the dealer will get back the asset it buys (sells) and lends to (borrows from) their client, but we assume that the dealer must have g assets to facilitate these levered trades. Without such an assumption, the dealer would be able to do infinitely many levered trades with its clients.[23] The presence of g implies a cost that dealers must bear to intermediate levered trades; in Section 7.1, we will present suggestive evidence that such friction exists.

4.2.3 Securities Lenders

We model securities lenders in a reduced form to focus on dealer behavior. Securities lenders are endowed with a large inventory of securities, which they lend out through the SI repo market. We assume that they do not participate in the cash market (they already have their optimal portfolio), but they reinvest the cash received from their securities lending activity in the GC market, profiting from the specialness s. To capture this dynamic, we assume securities lenders lend out assets in the SI market based on an exogenously specified supply function $\mathcal{SL}(s;\eta)$ with $\frac{\partial \mathcal{SL}}{\partial s} > 0$. That is, the more the repo trades on special, the more the securities lenders are willing to lend. We also assume that the supply from securities lenders depends on η, which governs their willingness to provide securities lending services. In effect, Foley-Fisher et al. (2015) show that securities lending programs use funds from their activities to finance long-dated assets, making their lending services depend on factors beyond repo specialness. We capture these incentives in reduced form through η and assume $\frac{\partial \mathcal{SL}}{\partial \eta} > 0$.

[23] This assumption incorporates a friction that will decouple the cash market from the SI repo market, similar to Krishnamurthy (2002).

4.3 Market Structure Summary

To summarize our setup, Figure 4 illustrates how dealers distribute collateral between all three markets—cash, SI repo, and GC repo—and the role securities lenders and clients play. Clients submit buy q_L and sell q_H orders to their dealers, accompanied by repos and reverse repos, respectively. After the clients' order flow is realized, dealers trade in these three markets. The cash market will serve to buy and sell securities for dealers to rebalance their risky asset position. The GC repo market will allow dealers to raise or invest funds at the GC repo rate. The SI repo market will allow dealers to either source assets to satisfy their SI box constraint or deliver assets to raise relatively cheap funding. Finally, securities lenders will provide assets to the SI repo market as a function of repo specialness and their willingness to lend assets.

5 Optimal Strategies and Symmetric Equilibrium in the Unrestricted Case

In this section, we first solve for dealers' optimal strategies in $t = 1$ assuming no balance sheet restrictions. We then characterize the optimal bid-ask spreads, which determine the order intensity.

5.1 Optimal Strategies and Equilibrium in Interdealer Market

Given an order size $\tilde{q}_L = q_L$ and $\tilde{q}_H = q_H$, the dealer's final payoff takes the expression in equation (1). Therefore, the dealer solves the following problem.

$$\max_{\{Q_C, Q_{SI}, Q_{GC}\}} \mathbb{E}(u(\tilde{W}_2)|\tilde{q}_L = q_L, \tilde{q}_H = q_H),$$

subject to

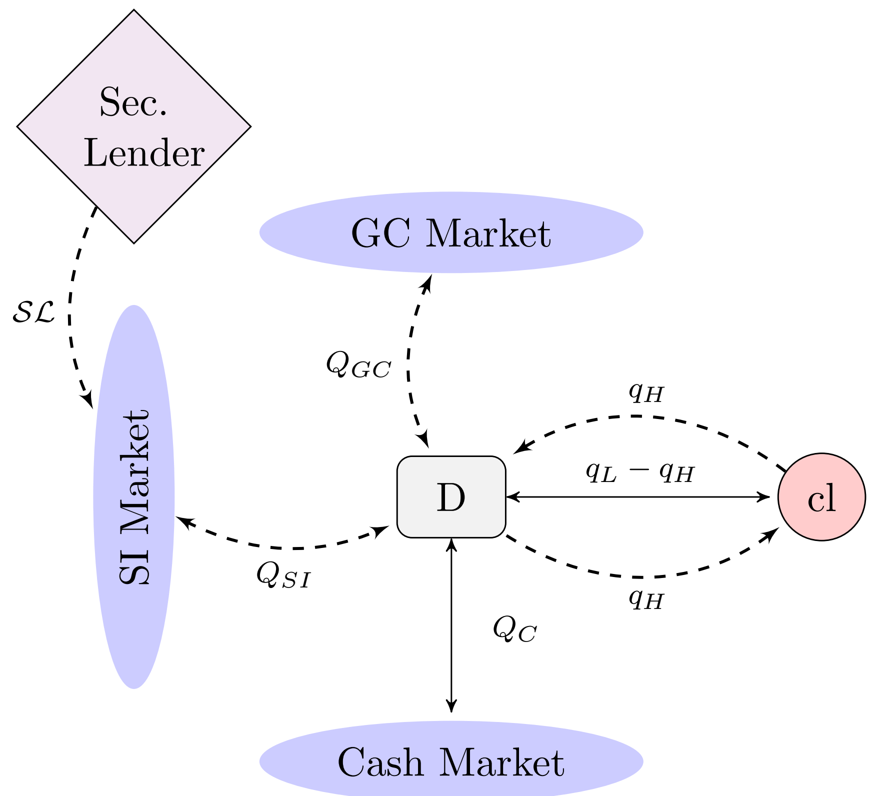
$$pQ_C + pQ_{SI} + pQ_{GC} \leq 0 \tag{3}$$

$$Q_C + Q_{SI} \geq g(q_L, q_H) \tag{4}$$

$$Q_C + Q_{SI} + Q_{GC} \geq 0. \tag{5}$$

Figure 4: Model market structure

D stands for dealer, cl stands for a dealer's client, and Sec. Lender stands for the securities lender. The dealer receives client long and sale orders q_L and q_H; which are accompanied by repos and reverse repos, respectively. *GC* Market stands for general collateral market, *SI* Market stands for specific-issue collateral market, and Cash Market stands for the underlying asset market. Upon receiving client orders, the dealer chooses how much collateral to source or deliver to the *GC* market Q_{GC} and the *SI* market Q_{SI} and how much to buy or sell in the cash market Q_C. Positive Q_{GC} or Q_{SI} means that the dealer is engaging in reverse repos. Positive Q_C means that the dealer is buying the asset. Securities lenders supply $\mathcal{SL}$ assets to the *SI* market. Only the asset movements (either through outright purchases and sales or as repo collateral) are drawn. Straight arrows indicate outright sales, and curved arrows indicate repo collateral movements.



where (3) is the dealer's budget constraint, (4) is the SI box constraint, and (5) is the global box constraint.[24] In (3), we assume that fees reaped from intermediation do not alter the dealer's budget. This simplifying assumption is made merely for tractability because we do not want client order sizes to affect the dealer's budget constraint. In reality, this effect, if any, is likely negligible. Given an asset price p, a GC repo rate R, and an SI repo rate R^{SI}, dealers will employ the following optimal strategies.

Lemma 1 (**Dealer's Optimal Strategy — Unrestricted Case**). *Given asset price p, SI repo rate R^{SI}, and secured funding rate $R > R^{SI}$, the dealer's optimal rebalancing strategy after receiving client orders $\tilde{q}_L = q_L$ and $\tilde{q}_H = q_H$ is*

$$\begin{aligned} Q^*_C &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} + q_L - q_H \\ Q^*_{SI} &= g(q_L, q_H) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} - q_L + q_H \\ Q^*_{GC} &= -g(q_L, q_H). \end{aligned}$$

Proof. See the appendix. □

Lemma 1 shows how dealers react to client order flow. One important observation from this result is that dealers' final asset positions are proportional to the asset's risk-adjusted return, which is the optimal solution for a regular CARA investor. That is,

$$q_H - q_L + Q^*_C = \frac{\mu - R^{SI}p}{\gamma\sigma^2}. \tag{6}$$

Because dealers have access to frictionless interdealer markets, it allows them to optimally adjust their portfolio to accommodate their clients' trades. Figure 5 shows how dealers rebalance their portfolios.

To better show the intuition, we consider a example where the dealer receives a short order of

[24] Because we assume that the dealer has no initial cash nor securities position, it seems mechanical that constraints 3 and 5 hold with equality. While formally true, a dealer's non-zero initial position would slacken these constraints. Specifically, a dealer with positive cash or asset position would alter the RHS of equation 3 or LHS of equation 5, respectively.

size q_H. Additionally, for expositional purposes, assume that $\mu = R^{SI}p$, that is, the asset's levered price is equal to its expected value.

Then, its optimal strategy will be

$$Q_C^* = -q_H$$
$$Q_{SI}^* = g(0, q_H) + q_H$$
$$Q_{GC}^* = -q_H.$$

This optimal strategy is shown in Panel (b) of Figure 5. To maintain an optimal portfolio, the dealer wants to sell q_H of the risky assets in the interdealer market. However, to do so, the dealer has to source an additional $g(0, q_H)$ assets from the SI repo market because of the SI box constraint; thus, he will source $g(0, q_H) + q_H$ assets. Note that in the client long scenario this does not happen, as the dealer buys q_L from the interdealer market.

Aggregating all interdealer cash market trades gives the following market clearing condition for the cash market. Here, we denote each dealer's demand with a subscript j to highlight that the equilibrium price is determined by aggregating across all dealers.

$$\int_j Q_{C,j} dj = C. \tag{7}$$

where C is an exogenous supply of assets in the cash market. The SI repo market, which also incorporates securities lenders' asset supply, clears through the following equation,

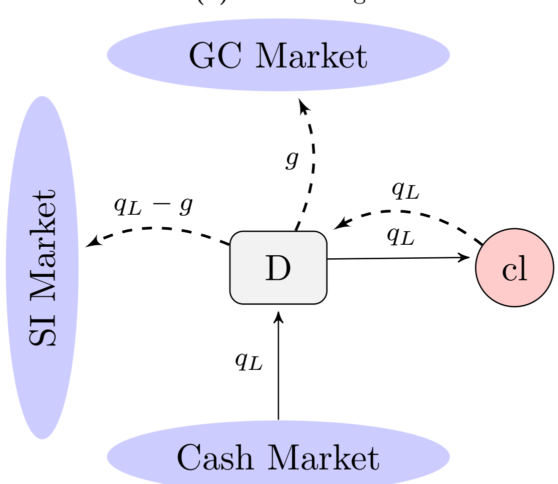
$$\int_j Q_{SI,j} dj = \mathcal{SL}(s; \eta). \tag{8}$$

Proposition 1 (Interdealer Equilibrium — Unrestricted Case). *Given a GC repo rate of R, securities lending function sufficiently small, and symmetric bid and ask spreads across dealers; then dealers' optimal strategies characterized in Lemma 1 result in an asset price p and SI repo*

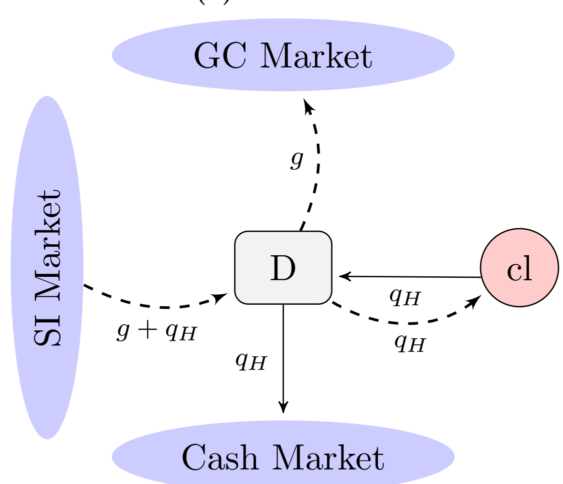
Figure 5: Simple example

These diagrams show the dealer's optimal solution when $\mu = R^{SI}p$. Clients submit either a levered long order (panel (a)) or a short order (panel (b)). D stands for dealer and cl stands for the dealer's client. GC Market stands for general collateral market, SI Market stands for specific-issue collateral market, and Cash Market stands for the underlying asset market. Dealers optimally choose how much collateral to source or deliver in the GC and SI markets and how much to buy or sell in the cash market. Only the asset movements (either through outright purchases and sales or as repo collateral) are drawn. Straight arrows indicate outright sales, and curved arrows indicate repo collateral movements. g stands for $g(q_L, 0)$ in panel (a) and $g(0, q_H)$ in panel (b).

(a) Client Long



(b) Client Short



rate $R^{SI} < R$, which solves the following system of equations:

$$\begin{aligned} \frac{\mu - R^{SI}p}{\gamma\sigma^2} &= C - \frac{1}{2\lambda(a)} + \frac{1}{2\lambda(b)} && (9) \\ \mathcal{SL}(s;\eta) &= \mathbb{E}(g(\tilde{q}_L, \tilde{q}_H)) - C. && (10) \end{aligned}$$

Proof. The proof stems from considering dealers' strategies in Lemma 1, imposing market clearing conditions (7) and (8), and applying the law of large numbers for client orders so that the probability of a long/short order (equal to $\frac{1}{2}$) is the aggregate order flow. □

Equation (9) shows how the price responds to the expected client order flow: A lower ask spread increases the expected order flow from client longs (higher $\lambda(a)$), increasing the price; whereas a lower bid spread increases the expected order flow from client shorts (higher $\lambda(b)$), reducing the price. Equation (10) highlights what drives repo specialness. Specifically, if there are larger frictions—namely if dealers need to hold more assets to service levered longs and shorts, captured through a higher g, relative to the total amount of assets available to dealers then they need to source more assets in the SI market which puts upward pressure on repo specialness.

It is interesting to note that considering the SI market clearing condition in isolation gives

$$\mathcal{SL}(s;\eta) = \mathbb{E}(g(\tilde{q}_L, \tilde{q}_H)) - \left(\frac{\mu - R^{SI}p}{\gamma\sigma^2}\right) - \frac{1}{2\lambda(a)} + \frac{1}{2\lambda(b)}.$$

The above partial equilibrium equation highlights what market participants often note about repo specialness: an increase in the client short base increases repo specialness. In effect, if for some reason b decreases, increasing short order flow, then s would need to increase to clear the market. However, this observation neglects the effect of repo specialness on the asset price. The general equilibrium solution in (10) shows that what matters is the total amount of assets added and subtracted from the interdealer market, along with any frictions associated with intermediating levered trades (i.e., g).

Given the realization of $\tilde{q}_L = q_L$ and $\tilde{q}_H = q_H$, the dealer's final wealth when using the optimal strategy is given by

$$\tilde{W}_2^*|\{q_L, q_H\} = (\tilde{v} - R^{SI}p)\left(\frac{\mu - R^{SI}p}{\gamma\sigma^2}\right) - spg(q_L, q_H) + apq_L + bpq_H.$$

The dealers' final wealth consists of the upside from taking a levered position in the asset, the fees charged to clients, and the cost of having to source SI collateral to intermediate client orders. The dealer's final utility is

$$\mathbb{E}(u(W^*)|q_L, q_H) = -\exp\left\{-\gamma\left[\frac{1}{2}\left(\frac{(\mu - R^{SI}p)^2}{\gamma\sigma^2}\right) - spg(q_L, q_H) + apq_L + bpq_H\right]\right\},$$

where p and R^{SI} are given by Proposition 1.

5.2 Posting bid-ask at $t = 0$

After obtaining the dealer's optimal strategy and final expected utility, we characterize the optimal bid-ask spread that dealers quote at $t = 0$. For simplicity, assume that

$$g(q_L, q_H) = g_0(q_L + q_H), \tag{11}$$

with a constant $g_0 < 1$. Then the dealer's expected payoff in $t = 0$ is given by

$$\mathbb{E}(u(W^*)) = -\Gamma\mathbb{E}\left(\exp\left\{-\gamma p[a\tilde{q}_L + b\tilde{q}_H - sg_0(\tilde{q}_L + \tilde{q}_H)]\right\}\right), \tag{12}$$

where $\Gamma = \exp\left\{-\gamma\left[\frac{1}{2}\left(\frac{(\mu - R^{SI}p)^2}{\gamma\sigma^2}\right)\right]\right\} > 0$ is a constant. Although p is determined by market clearing in $t = 1$, p is already known in $t = 0$. This is because in a symmetric equilibrium, dealers know all other dealers' strategies and the actual client order distribution due to the law of large numbers. Integrating over $\tilde{q}_L$ and $\tilde{q}_H$ gives

$$\mathbb{E}(u(W^*)) = -\frac{1}{2} \times \Gamma \times \left(\frac{\lambda(a)}{(\lambda(a) + \gamma pa - \gamma psg_0)} + \frac{\lambda(b)}{(\lambda(b) + \gamma pb - \gamma psg_0)}\right). \tag{13}$$

For the integral to exist, we need $\lambda(a) + \gamma pa - \gamma psg_0 > 0$ and $\lambda(b) + \gamma pb - \gamma psg_0 > 0$. We then get the following Lemma on dealers' optimal bid-ask spreads.

Lemma 2 (**Dealer's Optimal Bid-Ask — Unrestricted Case**). *Given a GC repo rate of R, a securities lending function sufficiently small enough, and g as in* (11)*; then a dealer's optimal bid and ask spreads solve the following equations:*

$$\begin{aligned}\lambda'(a^*)a^* - \lambda(a^*) - \lambda'(a^*)sg_0 &= 0\\ \lambda'(b^*)b^* - \lambda(b^*) - \lambda'(b^*)sg_0 &= 0,\end{aligned}$$

with $a^* > sg_0$ and $b^* > sg_0$.

Proof. Taking the first-order condition of expression (13), deduced from Proposition 1 with g as in equation (11), gives the Lemma's optimality conditions. In addition, note that $a^* > sg_0$ and $b^* > sg_0$ because

$$\frac{\partial \mathbb{E}(u(W^*))}{\partial a}\Big|_{a=sg_0} = \frac{1}{2} \times \Gamma \times \frac{\gamma p}{\lambda(sg_0)} > 0,$$

implying $a^* > sg_0$. The same argument holds for b^*. □

Intuitively, in the case of client longs, dealers' optimally balance receiving additional profits $a - sg_0$ from marginal client order flow with an overall decrease in client demand. Note that an individual dealer's bid and ask spread is increasing in repo specialness. Because dealers need to source g assets in order to intermediate, forcing them to bear the cost of repo specialness, they pass on those costs to their clients. That is, in a partial equilibrium setting, the costs that clients pay are increasing in repo specialness.

Note that the optimal bid and ask in Lemma 2 do not depend on the underlying asset's price. This feature will be useful when characterizing how liquidity changes with securities lending activity.

5.3 Sensitivity of Liquidity to Changes in Securities Lending

In this section, we can characterize how the equilibrium changes with securities lenders' willingness to provide assets η. From Lemma 2 we have that $a^* = b^*$ can constitute an equilibrium. Focusing on symmetric equilibria, we characterize an equilibrium where $a^* = b^*$. The symmetric equilibrium is given by

$$\begin{aligned} T_1 &:= \mathbb{E}(g_0(\tilde{q}_L, \tilde{q}_H)) - C - \mathcal{SL}(s;\eta) = 0 \\ T_2 &:= \lambda'(a^*)a^* - \lambda(a^*) - \lambda'(a^*)sg_0 = 0. \end{aligned}$$

Proposition 2 (**Sensitivity of Liquidity to** η **— Unrestricted Case**). *Given the same assumptions from Lemma 2, an increase in securities lenders' willingness to provide assets decreases repo specialness and dealers' optimal bid and ask spreads. Specifically,*

$$\begin{aligned} \frac{\partial s}{\partial \eta} &= \frac{1}{|J|}(\lambda''(a^*)(a^* - sg_0))\frac{\partial \mathcal{SL}}{\partial \eta} < 0 \\ \frac{\partial a^*}{\partial \eta} &= \frac{1}{|J|}\lambda'(a^*)g_0\frac{\partial \mathcal{SL}}{\partial \eta} < 0, \end{aligned}$$

where $|J| < 0$ *is the determinant of the Jacobian matrix of partial derivatives of* T_1, T_2 *with respect* s *and* a^**.*

Proof. Proof involves applying the implicit function theorem to equations T_1 and T_2. See appendix for details. □

Proposition 2 shows that as securities lenders provide more assets into the market through repos, repo specialness and dealers' markup decrease. Both of these changes are intuitive. More assets available to lend reduces the degree of repo specialness. And because repo specialness is a cost borne by dealers, they pass those savings onto their clients. This result highlights the tight link between repo specialness and market liquidity.

6 Optimal Strategies and Symmetric Equilibrium in the Restricted Balance Sheet Case

The analysis in Section 5 assumed that dealers had the liberty to alter the size of their balance sheets to accommodate arbitrarily large client orders. But since the financial crisis, broker-dealers affiliated with Bank Holding Companies (BHCs) are subject to a number of regulatory restrictions in an effort to make these BHCs more resilient. One of these regulatory initiatives, the Supplementary Leverage Ratio (SLR), restricts the amount of leverage a large BHC can take. Given that large broker dealers affiliated with BHCs intermediate a large fraction of client order flow, their regulatory restrictions are likely to have an effect on the underlying market.[25]

In the context of our model, the specific functional form of the leverage ratio restriction used in the SLR would be difficult to characterize. For example, our dealer has no initial cash nor asset holdings, consequently no initial equity. From this perspective, any amount of debt would imply infinite leverage. But, assuming that the dealer is a subsidiary of a regulated BHC, from the dealer's perspective, the amount of equity in the BHC is fixed. Thus, a leverage restriction can be translated into a size restriction on the dealer's balance sheet. Kotidis and Van Horen (2018) empirically show that the leverage ratio reduces their repo outstanding. That is, the SLR acts like a size restriction on dealers' repo activity.

In order to understand how this restriction may affect the dealer's intermediation activities in the model, we first have to translate the model's outcome onto a balance sheet. To do so, it is important to understand how each trade appears on a dealer's balance sheet for accounting purposes. Dealers' outright security positions are netted out, but their repo positions are not. Therefore, dealers' repo activities are much more "balance sheet intensive."[26]

The innovation of this section is that dealers have an upper bound K on the size of their balance sheets. Given that the dealer's balance sheet composition is different for long and short orders, a convenient form to express the constraint is to impose that total assets and liabilities must be smaller than $2K$. In addition, dealers have an additional choice variable: how much of the order to intermediate Q_I. In its general form, the balance sheet restriction can be written as

$$|Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L)| + |Q_{SI}| + |Q_{GC}| + Q_I \leq 2K. \tag{14}$$

Similar to the assumption adopted for dealers' budget constraints, we assume that dealers' markups do not affect the size of their balance sheets. These cash flows are likely to be negligible

[25] Section 2 provides more information on the structure of the U.S. Treasury market.

[26] Legally, repos transfer ownership to the counterparty receiving the collateral, but for accounting purposes repos are treated as secured loans. See Infante (2019) for more details.

relative to the total size of a BHC's balance sheet. In this notation, $\mathbb{1}_L$ is an indicator function that equals 1 if the client order is a levered long and zero otherwise. $\mathbb{1}_H$ is defined similarly for client shorts. The four components of equation (14) are a dealer's final asset position, its interdealer SI and GC repos, and, finally, the repo (or reverse repo) issued to its client.

6.1 Optimal Strategies & Equilibrium in the Interdealer Market

Given a levered long order $\tilde{q}_L = q\mathbb{1}_L$ or a levered short order $\tilde{q}_H = q\mathbb{1}_H$, the dealer's final payoff takes the expression in equation (1). Therefore, the dealer solves the following problem,

$$\max_{\{Q_C, Q_{SI}, Q_{GC}, Q_I\}} \mathbb{E}(u(\tilde{W}_2)|\tilde{q}_L = q\mathbb{1}_L, \tilde{q}_H = q\mathbb{1}_H)$$

subject to

$$\begin{aligned} pQ_C + pQ_{SI} + pQ_{GC} &\leq 0 \\ Q_C + Q_{SI} &\geq g(Q_I) \\ Q_C + Q_{SI} + Q_{GC} &\geq 0 \\ |Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L)| + |Q_{SI}| + |Q_{GC}| + Q_I &\leq 2K \\ Q_I &\leq q. \end{aligned}$$

That is, the dealer's problem is the same as in Section 5, except now dealers face a balance sheet restriction (equation (14)) and must also decide how much to intermediate ($Q_I \leq q$).

Without loss of generality, we will assume that $\mu > R^{SI}p$, so that the dealer's optimal asset position is positive. Having a positive (negative) creates a small asymmetry when intermediating long (short) orders. As we will show below, the dealer may decide to compromise it's optimal asset position to intermediate more trades. The analysis is identical when $\mu < R^{SI}p$. The above problem gives way to the following solution.

Lemma 3 (**Dealer's Optimal Strategy — Balance Sheet Restricted Case**). *Given an asset price p, SI repo rate R^{SI}, secured funding rate $R > R^{SI}$, and $\mu > R^{SI}p$; then, upon receiving a* *client long order $\tilde{q} = q\mathbb{1}_L$, the dealer's optimal portfolio is equal to the solution of Lemma 1 with $Q_I = q$ whenever $q < K - \frac{\mu - R^{SI}p}{\gamma\sigma^2} := \overline{q}_1^L$. If $q \geq \overline{q}_1^L$, then the dealer's optimal strategy is*

$$
\begin{aligned}
Q_C^* &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} + \overline{q}_1^L \\
Q_{SI}^* &= g(Q_I^*) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} - \overline{q}_1^L \\
Q_{GC}^* &= -g(Q_I^*) \\
Q_I^* &= \min\{q, \overline{q}_2^L\},
\end{aligned}
$$

where $\overline{q}_2^L$ solves $\overline{q}_2^L = \overline{q}_1^L + \frac{pa - psg'(\overline{q}_2^L)}{\gamma\sigma^2}$.

Upon receiving a client short order $\tilde{q} = q\mathbb{1}_H$, the dealer's optimal portfolio is equal to the solution of Lemma 1 with $Q_I = q$ whenever $q < \overline{q}^H$ which solves $g(\overline{q}^H) + \overline{q}^H = K$. If $q \geq \overline{q}^H$, then the dealer's optimal strategy is

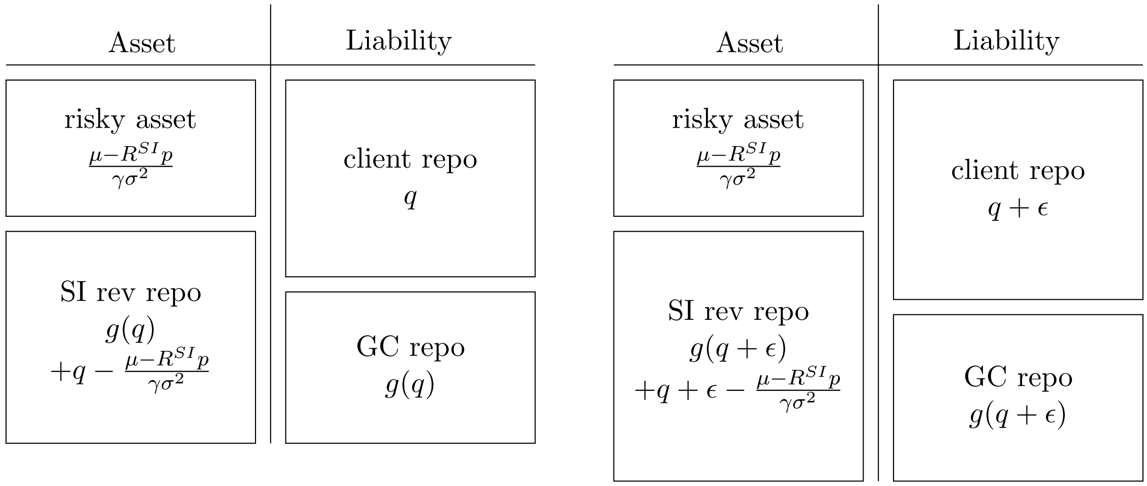
$$
\begin{aligned}
Q_C^* &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} - \overline{q}^H \\
Q_{SI}^* &= g(Q_I^*) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} + \overline{q}^H \\
Q_{GC}^* &= -g(Q_I^*) \\
Q_I^* &= \overline{q}^H.
\end{aligned}
$$

Proof. See the appendix. □

The intuition for the dealer's optimal response to a client short order is illustrated in Figure 6. For a relatively small order ($q < \overline{q}^H$), the dealer intermediates the trade as in the unrestricted case. If the client order size increases by ϵ as in Panel (b), the size of the dealer's balance sheet has to increase in order to accommodate it. This expansion happens because of the increase in repo the dealer issues to the client ϵ and the increase in the amount needed to intermediate the trade $g(q+\epsilon) - g(q)$. The dealer can do this until the balance sheet size reaches the limit K, which happens when $g(q) + q = K$; that is, when $q = \overline{q}^H$. Once the balance sheet limit is reached, the dealer will only intermediate $\overline{q}^H$.

Figure 6: Client short with restricted balance sheet

(a) $q < \overline{q}^H$ **(b)** $q + \epsilon \leq \overline{q}^H$



The difference between dealers' intermediation of long orders and those of short orders stem from the assumption that the unrestricted optimal portfolio is nonnegative, i.e., $\mu > R^{SI}p$. This assumption implies that dealers may accommodate larger client orders without increasing their balance sheets by compromising their risky asset positions. The dealers are willing to alter their optimal positions because they receive payments through a for intermediating large orders. They will stop intermediating more when the benefit from doing so is equal to the cost of altering their portfolios. That is, when q solves

$$(\mu - R^{SI}p) - \gamma\sigma^2(K - q) = pa - psg'(q)$$

which defines $\overline{q}_2^L$.

For both client long and short orders, a constraint on dealers' balance sheets limits the amount of orders they can intermediate. For the remainder of Section 6, for tractability, we take a simplified view of how regulatory restrictions limit the size of dealers' balance sheets. Specifically, instead of assuming that there is an overall restriction on the size of the dealer's balance sheet as in equation (14), we assume that dealers have an upper bound K on the total amount of securities they can intermediate for both client long and short orders. This assumption simplifies the analysis because dealers' optimal rebalancing strategies become symmetric: dealers do not have incentives to alter their optimal portfolio holdings to intermediate more client order flow. This reduces the number of equilibrium conditions.[27] Alternatively, we could have assumed that $\mu = R^{SI}p$, implying that the optimal position is zero, again making the problem of intermediating long or shorts symmetric. In the following subsections, we characterize this equilibrium and study how it changes as constraints become tighter and securities lenders' willingness to intermediate changes.

6.2 Optimal Strategies & Equilibrium in the Interdealer Market—Simplified Case

In this subsection we modify the dealers' balance sheet constraints and simply assume that each dealer will only be able to intermediate up to K securities of its clients' order flow, irrespective of the direction.

This assumption can be interpreted as internal controls imposed on traders to limit their contribution to the overall size of BHCs' balance sheets. As mentioned earlier, the benefit of this modification is that now dealers' strategies to intermediate client order flow are symmetric for both long and short order. Note that this restriction is mathematically equivalent to how the overall balance sheet restriction affects dealers' intermediation of client short positions when $\mu \geq R^{SI}p$. It then follows that, conditional on a client's order of size q, dealer's optimal strategies when intermediating a client long and short order take the following form, respectively,

$$
\begin{array}{ll}
\text{client long} & \text{client short} \\
Q_C^* = \dfrac{\mu - R^{SI}p}{\gamma\sigma^2} + Q_I^* & Q_C^* = \dfrac{\mu - R^{SI}p}{\gamma\sigma^2} - Q_I^* \\
Q_{SI}^* = g(Q_I^*) - \dfrac{\mu - R^{SI}p}{\gamma\sigma^2} - Q_I^* & Q_{SI}^* = g(Q_I^*) - \dfrac{\mu - R^{SI}p}{\gamma\sigma^2} + Q_I^* \\
Q_{GC}^* = -g(Q_I^*) & Q_{GC}^* = -g(Q_I^*) \\
Q_I^* = \min\{q, K\}, & Q_I^* = \min\{q, K\}
\end{array}
$$

[27] This assumption allows us to characterize closed-form solutions without affecting the main impact of the restriction.

That is, in the modified restricted balance sheet model, for a relatively small order flow, dealers intermediate as in the unrestricted balance sheet case. For a relatively large order flow, dealers intermediate up to K. Importantly, dealers' optimal responses to long and short orders are now symmetric.

As in the unrestricted balance sheet case, both the cash and SI repo markets clear. That is, integrating over all dealers i, market clearing is given by

$$\int_j Q_{C,j} dj = C, \quad \int_j Q_{SI,j} dj = \mathcal{SL}(s;\eta), \tag{15}$$

which results in the following equilibrium.

Proposition 3 (**Interdealer Equilibrium — Modified Restricted Balance Sheet Case**). *Given a GC repo rate of R and a sufficiently small securities lending function, dealers' optimal strategies characterized above result in an asset price p and specialness spread $s = R - R^{SI} > 0$, which solve the following system of equations:*

$$\begin{aligned} \frac{\mu - R^{SI}p}{\gamma\sigma^2} &= C - \frac{1}{2\lambda(a)}[1 - e^{-\lambda(a)K}] + \frac{1}{2\lambda(b)}[1 - e^{-\lambda(b)K}] \qquad (16) \\ \mathcal{SL}(s;\eta) &= \frac{1}{2}\left[\int_0^K g(q)\lambda(a)e^{-\lambda(a)q}dq + g(K)e^{-\lambda(a)K}\right] + \\ &\quad \frac{1}{2}\left[\int_0^K g(q)\lambda(b)e^{-\lambda(b)q}dq + g(K)e^{-\lambda(b)K}\right] - C \qquad (17) \end{aligned}$$

Proof. The proof stems from considering dealers' strategies, imposing market clearing conditions (15), and applying the law of large numbers for client orders. □

From Proposition 3, it can be appreciated how constraints on dealers' balance sheets can have a direct impact on the underlying asset price and repo specialness. Specifically, for a given bid and ask, reducing dealers' balance sheets decreases the demand for interdealer repos, thereby limiting repo specialness. The latter effect depends on frictions in dealer intermediation given by g.

6.3 Posting Bid & Ask at $t = 0$

As before, having characterized a dealer's optimal strategy given clients' order flow, we can express his expected payoff in terms of its optimal portfolio and optimal markup. As in the unrestricted balance sheet case, the expected payoff from intermediating a client long or short position take the same functional form. The expression for a client long order of size q is

$$\mathbb{E}(u(W^*)|\tilde{q}_L = q) = \begin{cases} -\Gamma \exp\{-\gamma p\,[aq - sg(q)]\} & \text{if } q < K \\ -\Gamma \exp\{-\gamma p\,[aK - sg(K)]\} & \text{if } q \geq K, \end{cases}$$

where $\Gamma = \exp\left\{-\gamma \frac{1}{2}\left(\frac{(\mu - R^{SI}p)^2}{\gamma\sigma^2}\right)\right\} > 0$ in this case. That is, the payoff is as in the unrestricted balance sheet case, but capped at $\tilde{q}_L = K$. As in previous sections, for tractability we will assume that $g(q) = g_0 q$, with g_0 a constant between 0 and 1.[28] Integrating over all possible client long orders, $\tilde{q}_L$, gives

$$\mathbb{E}(u(W^*)|\mathbb{1}_L) = -\Gamma\left(\frac{\lambda(a)}{\lambda(a) + \gamma p(a - g_0 s)} + \frac{\gamma p(a - g_0 s)}{\lambda(a) + \gamma p(a - g_0 s)} e^{-[\lambda(a) + \gamma p(a - g_0 s)]K}\right). \tag{18}$$

The expression of dealers' expected payoff after a client short is identical. Defining $w(a, s, K) := e^{-[\lambda(a) + \gamma p(a - g_0 s)]K}$ which is between 0 and 1, we have the following Lemma for the restricted dealer's optimal ask.

Lemma 4 (**Dealer's Optimal Ask — Modified Restricted Balance Sheet Case**). *Given the assumptions of Proposition 3, if K is sufficiently large and $g(q) = g_0 q$; dealer's optimal ask solves the following equation:*

$$\begin{aligned} 0 \;=\; & \left(\lambda'(a^*)a^* - \lambda'(a^*)g_0 s - \lambda(a^*)\right)[1 - w(a^*, s, K)] \\ & -(\lambda(a^*) + \gamma p(a^* - g_0 s))(a^* - g_0 s)(\lambda'(a^*) + \gamma p)K w(a^*, s, K)\,. \end{aligned} \tag{19}$$

Proof. See the Appendix. □

[28] Assumption $g' \leq 1$ is to take into account that the dealer may not need the entire asset to intermediate a client's levered order.

Lemma 4 shows how the balance sheet constraint affects the dealer's decision when setting his optimal ask. Equation (19) can be interpreted as a weighted average of two considerations. The first is identical to the optimality condition of Lemma (2), which takes into account the tradeoff between larger fees, adjusted for the cost of specialness, with a smaller expected order flow. The second highlights the limit on how much a dealer can intermediate. Note that as K increases, the second term disappears, reducing the optimality condition to the unrestricted case.

6.4 Sensitivity of Liquidity to Changes in Balance Sheet Constraints

To understand the aggregate effects of balance sheet restrictions, we would like to see how the general equilibrium changes with tighter balance sheet constraints. This analysis involves taking into account how dealers' optimal bid and ask spreads change, as well as p and R^{SI}, which implies incorporating the changes in market clearing conditions (16) and (17). To better understand how the balance sheet constraint works, we first characterize how an individual dealer's balance sheet constraint affects that dealer's own bid decision.

6.4.1 Balance Sheet Constraints on an Individual Dealer

In this subsection we ask how an individual dealer's balance sheet restriction affect the bid-ask spread offered to his clients. That is, we only explore a partial equilibrium change, taking as given the p and R^{SI}:

Proposition 4 (**Sensitivity of an Individual Dealer's Optimal Ask to Changes in** K)**.** *Given the assumptions of Proposition 3, if K is sufficiently large and $g(q) = g_0 q$, a tightening of an individual dealer's balance sheet constraint leads to an increase in his optimal ask. That is,*

$$\frac{\partial a^*}{\partial K} < 0.$$

Proof. See the Appendix. □

The result from Proposition 4 shows that—at least in a partial equilibrium setting—tighter balance sheet constraints induce dealers to increase their markups, effectively reducing liquidity for their clients. Intuitively, given that dealers are restricted from filling large client orders, they opt to increase the revenue from filling smaller ones, which increases clients' intermediation costs, making the underlying cash market less liquid.

This result is consistent with market commentary which suggests that size constraints on dealers' balance sheets have limited their ability to intermediate markets and, hence, reduced bond market liquidity. Similar to the intuition borne out of the model, market participants suggest that allocating balance sheet space to their market-making restricts them from accommodating larger trades. Therefore, dealers will charge their clients more for their market-making services, reducing the overall market's liquidity. But this intuition is from the narrow view of a single dealer's constraint. To explore the validity of this channel, in the following subsection, we study how aggregate constraints affect overall dealer liquidity provision.

6.4.2 Balance Sheet Constraints on All Dealers

Now we turn to analyze the equilibrium's sensitivity when all dealers face balance sheet restrictions. In this case, restricting the size of trades that dealers can intermediate has an effect on both the price and specialness.

Given the symmetry of the modified restricted balance sheet model, from Lemma 4 we know that the first-order condition for the optimal ask and bid spreads are the same. We focus on symmetric equilibria in which $a^* = b^*$ and all dealers quote the same bid-ask spreads. The symmetric equilibrium is given by the dealer's optimal bid-ask spreads and the equilibrium conditions in Proposition 3; that is,

$$\begin{aligned}
\hat{T}_1 &= \mathcal{SL}(s;\eta) - \frac{g_0}{\lambda(a^*)}[1 - e^{-\lambda(a^*)K}] + C = 0 \\
\hat{T}_2 &= \big(\lambda'(a^*)a^* - \lambda'(a^*)g_0 s - \lambda(a^*)\big)\left[1 - w\left(a^*, s, K\right)\right] \\
&\quad -(\lambda(a^*) + \gamma p(a^* - g_0 s))(a^* - g_0 s)(\lambda'(a^*) + \gamma p)Kw\left(a^*, s, K\right) = 0.
\end{aligned}$$

Note that in this formulation, the equilibrium conditions do depend on p. We implicitly use the cash market clearing equation, which in this case implies $pR^{SI} = \mu$.

Proposition 5 (Sensitivity of All Dealers' Optimal Ask to Changes K and η). *Given the assumptions of Proposition 3, if K is sufficiently large and $g(Q) = g_0 Q$; a loosening of all dealers balance sheet constraints leads to an increase in the optimal ask and repo specialness. That is,*

$$\frac{\partial a^*}{\partial K} > 0, \quad \frac{\partial s}{\partial K} > 0.$$

In addition, an increase in securities lenders' willingness to provide securities leads to a decrease in the optimal ask and repo specialness. That is,

$$\frac{\partial a^*}{\partial \eta} < 0, \quad \frac{\partial s}{\partial \eta} < 0.$$

Proof. See the Appendix. □

The result from Proposition 5 shows that the partial equilibrium intuition of Proposition 4 is reversed once the effect on prices is considered. Intuitively, as the balance sheet constraint becomes tighter, dealers' aggregate demand for specific-issue securities goes down, reducing repo specialness. As the cost of intermediation goes down, so does the dealer's markup. From an individual dealer's perspective, the partial equilibrium incentive is still present: as the dealer's ability to intermediate large order flow is restricted, he opts to increase revenue from smaller client orders. But the effect on specialness reduces dealers' marginal intermediation costs for all trades, incentivizing them to attract a larger share of smaller client order flow by lowering their markup.

The effect on changes in securities lenders' willingness to supply securities is intuitive. As their willingness to provide specific-issues securities increases (higher η), the cost of sourcing specific issue securities goes down, and so do dealers' ask spreads. In Section 7.2, we will discuss how these findings relate to recent evidence and current trends in bond market intermediation.

To illustrate the intuition, we provide a numerical simulation here. The exogenous parameters and functions of the model are specified as below, and we calculate the equilibrium for different values of K.

$$
\begin{aligned}
g = 1, &\quad R = 1.1 \\
\mu = 100, &\quad \sigma = 2 \\
\gamma_0 = 1, &\quad C = 0 \\
\mathcal{SL}(R - R^{SI}) &= SL1 \times (R - R^{SI}) \\
SL1 &= 100 \\
\lambda(a) &= 0.2 + 100 \times a^2
\end{aligned} \tag{20}
$$

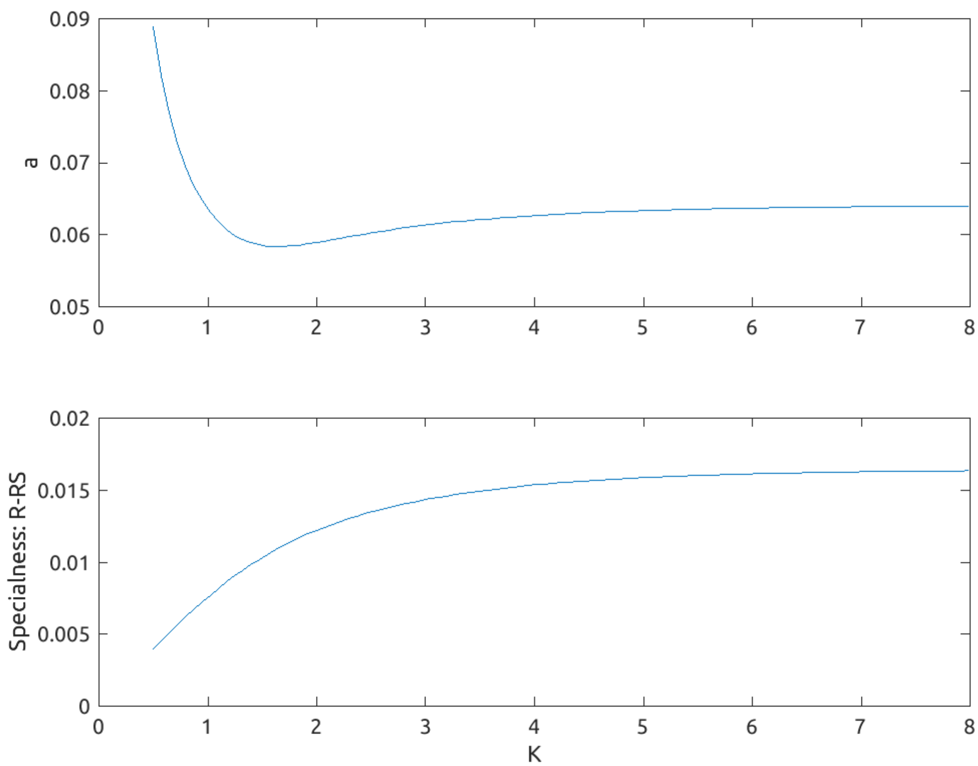


Figure 7: Simulation results for general equilibrium
We numerically calculate the general equilibrium for various values of K. The simulation specification is in (20). The top graph plots the equilibrium bid-ask spread, and the bottom graph plots the equilibrium specialness s.

Figure 7 plots the equilibrium bid-ask spread and specialness for various values of K. As a benchmark, in the unrestricted case (i.e., $K = \infty$), the equilibrium a is 0.064, which implies that the order arrival is $Exp(0.6099)$ and the average order size is 1.6395. Consistent with Proposition 5, for sufficiently large values of K, the equilibrium bid-ask spread and specialness decrease as the balance sheet constraint gets tighter (K decreases). At low levels of K, tighter balance sheet constraints increase the bid-ask spread as the partial equilibrium incentives outweigh the effects of lower specialness.

One of the crucial determinants of whether tighter balance sheets will decrease or increase the equilibrium a (for a given value of K) is the sensitivity of securities lending to specialness. To illustrate this relationship, we look how a and s change when K decreases from 1.1 to 1 for various values of $SL1$. Higher values of $SL1$ imply a more elastic supply curve. Figure 8 plots Δa and Δs across various values of $SL1$. When the securities lending function is inelastic ($SL1$ is low), the equilibrium a decreases when the balance sheet restriction gets stricter. Conversely, when the securities lending function is elastic ($SL1$ is higher), the equilibrium a increases when the balance sheet restriction gets stricter.

To see the intuition, consider the case where the securities lending function is perfectly elastic; that is, $SL1$ is infinity, and the specialness does not change with K. Hence, given that p and s remains the same, only the partial equilibrium channel remains, and equilibrium a will always increase with tighter balance sheet constraints. In contrast, if the securities lending function is fairly inelastic, the specialness would decrease by a large amount when K decreases, and thus, the impact of lower specialness would dominate, and equilibrium a will decrease.

7 Discussion

In this section, we discuss how the model's assumptions and setup capture key features of bond market intermediation in practice. This section also discusses how the model outcome and its implications can help interpret available data on dealers' securities books and explain observed patterns in the data.

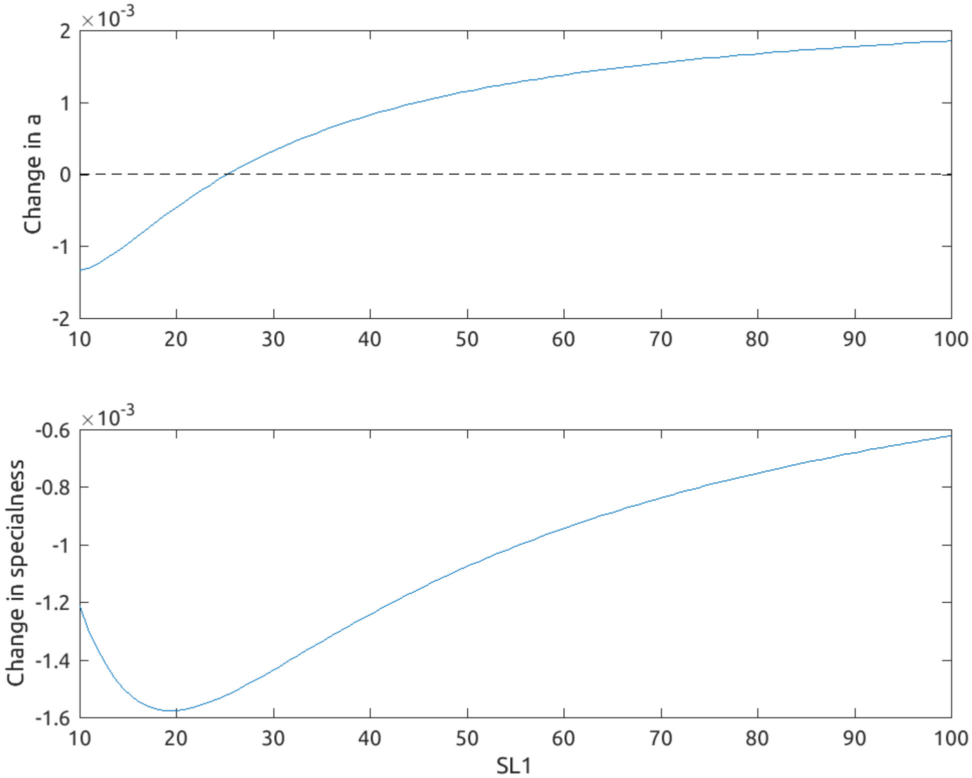


Figure 8: Securities lending elasticity and equilibrium
This figure plots the change in equilibrium a and s when K changes from 1.1 to 1, for various values of $SL1$. The simulation specification is in (20), but across various values of $SL1$. The top graph plots Δa, and the bottom graph plots Δs.

7.1 Discussion of Model Assumptions

The model considers a continuum of dealers that trade in three distinct markets to manage their portfolio and intermediate client order flow. This setup is in the spirit of many market making models, where a subset of agents are endowed with the ability to access markets, but some of their counterparties (in this case, clients) cannot. Clients cannot directly access interdealer cash or repo markets in our model, which is similar to current institutional arrangements of the U.S. Treasury market.[29]

The novelty of the model is to allow dealers to use all three markets—cash, GC repo, and SI

[29] See Section 2 for institutional details of the U.S. Treasury markets.

repo markets—to intermediate client order flow. The setup captures how dealers rely on the cash market to trade and how they depend on—and stand in the middle of—different repo markets to finance and source securities. The use of all three markets to intermediate client order flow arises naturally in the model because of dealers' box constraints. Without access to repo markets, dealers would not be able to intermediate leveraged client orders without deviating from their optimal portfolio.

An important feature of the model is the SI box constraint, which assumes that the dealer should have securities in an amount commensurate to clients' leveraged order flow. This constraint can be motivated by the fact that cash and repo trades have to be physically settled and that failing to deliver a security entails a cost. For example, consider a case where a client wants to short one unit of the asset and his dealer arranges for this by borrowing one unit of the asset from another dealer. If the dealer does not have the asset in the beginning, that dealer's SI box equals zero. If the dealer providing the initial asset fails to deliver it, the dealer filling the client order flow would fail to deliver to its customer. Because failing-to-deliver entails a cost (either an outright fails charge or not being able to intermediate clients' orders), the dealer *ex ante* would have precautionary motive to source in more assets than is needed. We do not model and solve for the optimal amount of precautionary sourcing, but instead assume a function $g(\cdot)$ that is proportional to the clients' leveraged order flow.

Empirical evidence suggests that the global box constraint and the SI box constraint are relevant. Figure 9 shows the average SI box and the global box (which includes assets sourced through GC repos) for the 2-year and 10-year on-the-run U.S. Treasury note. It indicates that the SI box is mostly positive, which implies that assets that can be delivered into the SI market are either posted in the GC market or not used at all. Absent an SI box constraint, this behavior seems puzzling, because whenever the bond trades on special, this allocation implies a loss. That is, assets are used to raise cash at a higher rate than they would otherwise if used in the SI market. This behavior is observed across U.S. Treasury securities of other tenors, providing evidence that dealers' SI box constraint results in an intermediation cost that they must bear. The summary statistics for the global and SI box, provided in Table 1, also show similar results.

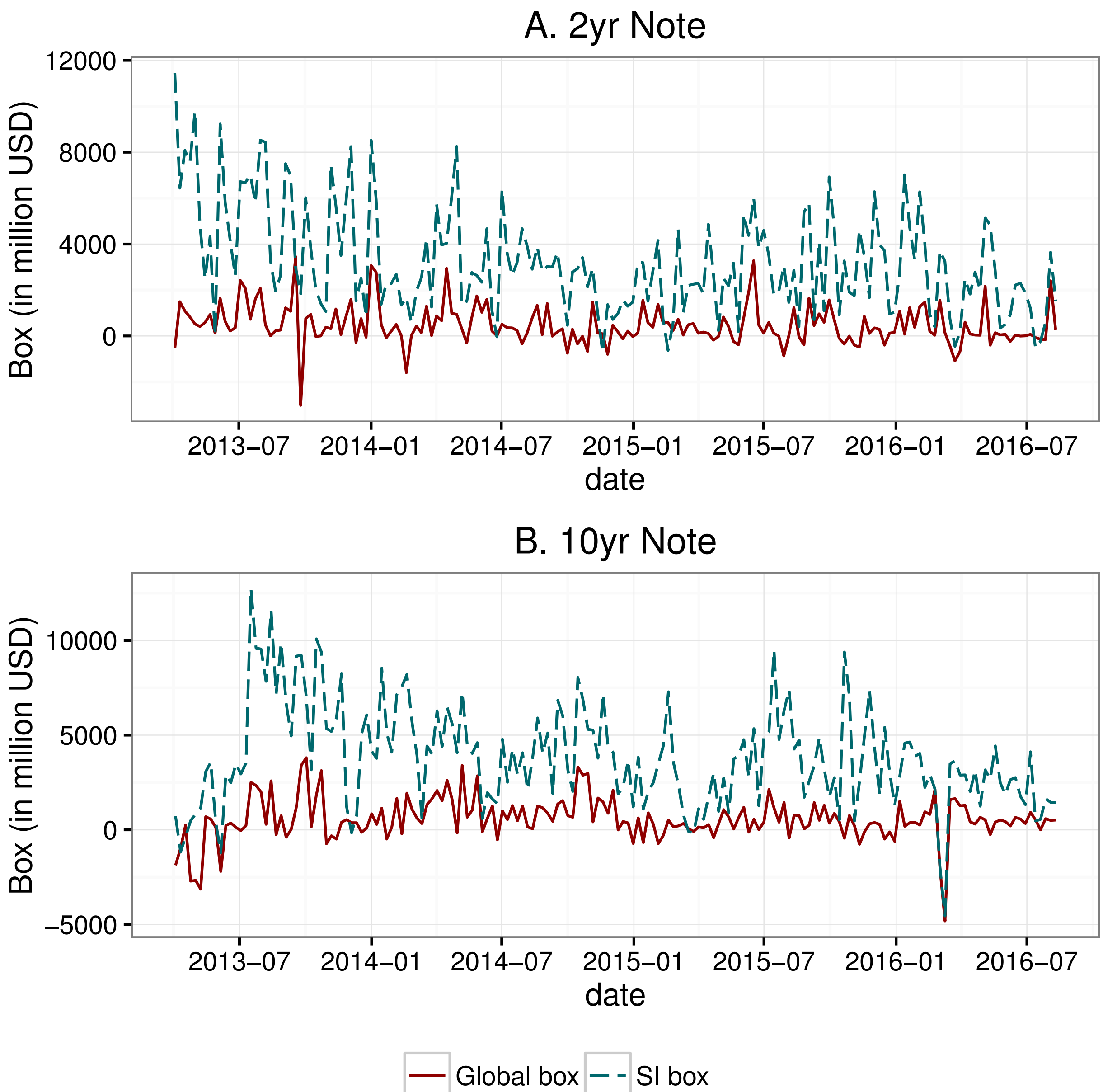


Figure 9: SI box and global box for primary dealers.
The SI box and global box are calculated from FR 2004 data downloaded from the New York Fed website. The SI box is defined as the net position plus the reverse repo on specific issue minus the repo on specific issue ("net settled position" plus "overnight/open, specific transactions, securities in" plus "term, specific transactions, securities in" minus "overnight/open, specific transactions, securities out" minus "term, specific transactions, securities out"). The global box is defined as the SI box plus reverse repo on general collateral minus the repo on general collateral (the SI box plus "overnight/open, general transactions, securities in" plus "term, general transactions, securities in" minus "overnight/open, general transactions, securities out" minus "term, general transactions, securities out"). Panel A calculates the SI box and global box for the 2 year Treasury note, and Panel B calculates for the 10 year Treasury note.

Table 1: Summary statistics of SI box
This table shows summary statistics for the SI box and the global box for each tenor, calculated from FR2004 data. The SI box is calculated as the net position plus the SI reverse repo position minus the SI repo position for the on-the-run Treasuries. The global box is calculated as the SI box plus the reverse repo position in the GC market that uses the on-the-run security for the underlying, minus the repo position in the GC market with the on-the-run as underlying collateral. Data are weekly, reported as of the end of Wednesday, from March 2007 to April 2015.

For each week, we calculate the average SI box and average global box across primary dealers. We then report the summary statistics for this time series data. The third and fourth columns report the average SI box and global box, and the fifth column presents the fraction of dealer-weeks with a positive SI box.

Tenor	# of weeks	avg SI box	avg global box	SI box >0
2 year	424	188.14	25.74	89%
3 year	424	166.64	27.78	90%
5 year	424	140.45	26.67	80%
7 year	320	214.53	32.55	94%
10 year	424	191.04	57.57	88%
30 year	424	176.8	44.05	95%

7.2 Discussion of Model Outcome

One of the main implications of the model is that higher specialness leads to a more illiquid market. This finding seemingly goes against previous research that argues that a bond trades special *because* it is liquid. Vayanos and Weill (2008) argue that on-the-run securities are easier to find than off-the-run securities, increasing demand to source the on-the-run securities in SI repo markets in order to short and, consequently, increasing their specialness. That is, in the cross section, the more liquid securities will trade with higher specialness. The intuition that higher demand for SI repos increases specialness is consistent with our model; however, we do not model the cross-section of securities directly.

But the positive relationship between illiquidity and specialness predicted by the model is in the time series. The intuition is that because specialness is a cost dealers must bear to intermediate client order flow, a higher degree of specialness will translate into a higher transaction cost for the client and, hence, lower liquidity. This result is significant because it provides a different metric to gauge bond market liquidity. Measuring market liquidity is particularly difficult in the U.S.

Treasury market because there is little to no data in the OTC dealer to customer market. The model prescribes that specialness can gauge the relevant costs to intermediate bonds and hence their liquidity. Figure 1 shows that in recent years, specialness has been increasing, implying a deterioration of bond market liquidity that has not been measured elsewhere but has been commented on by market participants.

Much of the aforementioned market commentary has blamed regulation, in particular the SLR, for the deterioration in bond market liquidity. The results from our restricted balance sheet model suggests this intuition is incomplete. In a partial equilibrium setting, restricting an individual dealer's balance sheet would reduce its incentives to attract a large order flow, increasing the markup for smaller trades. But this conclusion ignores the effects that dealer balance sheet restrictions can have on specialness. Intuitively, if dealer demand for SI repos is capped, then the cost to source them should decrease. Our general equilibrium analysis shows that as balance sheet constraints for all dealers tighten, specialness decreases, and so do bid and ask spreads in the cash market.

This result suggests that something other than the SLR might be hindering liquidity. In the general equilibrium model, we show that a decrease in securities lenders' willingness to lend will increase both specialness and bid-ask spreads. That is, if the overall supply of securities is reduced, then dealers' intermediation costs to source them increase, which translates into higher dealer markups. This observation is consistent with anecdotal evidence that suggests that securities lenders have become conservative with their lending activity since the financial crisis. Specifically, securities lenders have imposed stricter counterparty limits and become more conservative with their reinvestment portfolios, reducing their incentives to lend securities. From the model, this trend is consistent with both the increase in specialness and market illiquidity.

Lastly, the model is also useful in interpreting data on dealers' securities books and in backing out information on dealers' market activity. One example of these data is the FR 2004 survey that documents dealers' positions, trading, and short-term financing (which includes both repos and securities lending activity) for U.S. Primary Dealers. The analysis in Section 6 illustrates that when dealers intermediate trades, repos have a much larger impact on the size of the dealers' balance sheets than the outright long and short positions, which can be netted. Therefore, restrictions on the size of dealers' balance sheets have a larger impact on their repo activity than on their outright positions. Moreover, Figures 6 show that an important fraction of repos and reverse repos are not just to raise and distribute funding, but also to source and deliver securities, a fact that is often overlooked. In particular, a contraction in a dealer's repo book may not be solely attributed to a reduction in funding, but also on a reduction in trading activity. This interpretation is consistent with Figure 2.

8 Concluding Remarks

This paper presents a model of dealers' bond market-making activities, taking into account the importance of repo markets, and shows how repo markets are closely linked to the underlying asset market. Repos allow dealers to source and finance assets in order to fill client orders. We show that filling client orders is balance sheet intensive. The fees dealers charge are proportional to the cost of sourcing specific assets, which is captured by the repo specialness.

In a world where the size or the leverage of dealers' balance sheets is limited, dealers have reduced incentives to service large orders and increase the costs they pass onto their clients. In effect, balance sheet limits reduce dealers' ability to intermediate large trades, reducing market depth. In the partial equilibrium setting with fixed prices, reducing dealers' balance sheet size increases the bid-ask spreads dealers charge to their clients, thus decreasing market liquidity. In the general equilibrium setting when specialness is endogenous, reducing dealers' balance sheet size decreases their demand for specific-issue repo, reducing repo specialness and putting downward pressure on bid-ask spreads.

The above observation puts into question recent criticism that new regulatory initiatives restricting the size of dealers' balance sheets have decreased bond market liquidity. In the context of the model, the criticism is natural from an individual firm's perspective: balance sheet restrictions limit order flow, incentivizing an increase in intermediation fees. But the general equilibrium model shows a more complex picture, as these types of restrictions can translate into lower repo specialness.

The general equilibrium results suggest that further research is needed to explain the upward trend in the main intermediation cost, the repo specialness, highlighted in this paper. In the paper we present one possible channel, which is consistent with an increase in specialness and a decrease in liquidity: a decrease in securities lenders' willingness to lend. Anecdotally, the securities lenders' business model has significantly changed over the past decade. Understanding securities lenders' activities, and how their incentives to lend securities may have changed, is an interesting topic for future research.

References

Adrian, T., Begalle, B., Copeland, A. and Martin, A. (2013), Repo and securities lending, *in* 'Risk Topography: Systemic Risk and Macro Modeling', University of Chicago Press.

Allahrakha, M., Cetina, J. and Munyan, B. (2018), 'Do higher capital standards always reduce bank risk? the impact of the basel leverage ratio on the us triparty repo market', *Journal of Financial Intermediation* **34**, 3–16.

Amihud, Y. and Mendelson, H. (1980), 'Dealership market: Market-making with inventory', *Journal of Financial Economics* **8**(1), 31–53.

Banerjee, S. and Graveline, J. J. (2013), 'The cost of short-selling liquid securities', *The Journal of Finance* **68**(2), 637–664.

Bao, J., OHara, M. and Zhou, X. A. (2018), 'The volcker rule and corporate bond market making in times of stress', *Journal of Financial Economics* .

Barclay, M. J., Hendershott, T. and Kotz, K. (2006), 'Automation versus intermediation: Evidence from treasuries going off the run', *The Journal of Finance* **61**(5), 2395–2414.

Bottazzi, J.-M., Luque, J. and Páscoa, M. R. (2012), 'Securities market theory: Possession, repo and rehypothecation', *Journal of Economic Theory* **147**(2), 477–500.

Brainard, L. (2018), 'The structure of the treasury market: What are we learning?'. Remarks by

Governor Lael Brainard at the Fourth Annual Conference on The Evolving Structure of the U.S. Treasury Market.

Choi, J. and Huh, Y. (2017), 'Customer liquidity provision: Implications for corporate bond transaction costs', *Working Paper* .

Cimon, D. A. and Garriott, C. (2019), 'Banking regulation and market making', *Journal of Banking & Finance* **109**, 105653.

Copeland, A., Duffie, D., Martin, A. and McLaughlin, S. (2012), 'Key mechanics of the US tri-party repo market', *Federal Reserve Bank of New York Economic Policy Review* **18**(3), 17–28.

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

Duffie, D. (2017), 'Financial regulatory reform after the crisis: An assessment', *Management Science* **64**(10), 4835–4857.

DAmico, 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.

Foley-Fisher, N., Narajabad, B. and Verani, S. (2015), 'Securities lending as wholesale funding: Evidence from the US life insurance industry', *Available at SSRN* .

Garbade, K. D., Keane, F. M., Logan, L., Stokes, A. and Wolgemuth, J. (2010), 'The introduction of the TMPG fails charge for US treasury securities', *Economic Policy Review* (Oct), 45–71.

Graveline, J. J. and McBrady, M. R. (2011), 'Who makes on-the-run treasuries special?', *Journal of Financial Intermediation* **20**(4), 620–632.

Ho, T. S. and Stoll, H. R. (1983), 'The dynamics of dealer markets under competition', *The Journal of Finance* **38**(4), 1053–1074.

Infante, S. (2019), 'Liquidity windfalls: The consequences of repo rehypothecation', *Journal of Financial Economics* **133**(1), 42–63.

Infante, S., Press, C. and Saravay, Z. (2020), ‘Understanding collateral re-use in the u.s. financial system’, *AEA Papers and Proceedings (forthcoming)* .

Keane, F. M. (1996), ‘Repo patterns for new treasury notes’, *Current Issues in Economics and Finance* **2**(10).

Klaus, B. and Rzepkowski, B. (2009), ‘Hedge funds and prime brokers: The role of funding risk’.

Kotidis, A. and Van Horen, N. (2018), ‘Repo market functioning: The role of capital regulation’.

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

Kruttli, M., Monin, P. and Watugala, S. W. (2018), ‘Prime broker lending and hedge fund exposures’, *Working Paper* .

Stoll, H. R. (1978), ‘The supply of dealer services in securities markets’, *The Journal of Finance* **33**(4), 1133–1151.

Trebbi, F. and Xiao, K. (2017), ‘Regulation and market liquidity’, *Management Science* **65**(5), 1949–1968.

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

A Appendix

Proof of Lemma 1:

Given a realization $\tilde{q}_L = q_L$ and $\tilde{q}_H = q_H$, the dealer's optimization problem has the following Lagrangean,

$$
\begin{aligned}
\mathcal{L} \;=\; & \gamma \Big[(\mu - p)Q_C + (R^{SI} - 1)pQ_{SI} + (R - 1)pQ_{GC} \\
& -(\mu - R^{SI}p)q_L + (\mu - R^{SI}p)q_H + apq_L + bpq_H\Big] - \frac{1}{2}\gamma^2\sigma^2(Q_C - q_L + q_H)^2 \\
& -\lambda\left[Q_C + Q_{SI} + Q_{GC}\right] + \xi_S[Q_C + Q_{SI} - g(q_L, q_H)] + \xi_G[Q_C + Q_{SI} + Q_{GC}]
\end{aligned}
$$

Giving the following FOC:

$$
\begin{aligned}
Q_C^* &: \gamma(\mu - p) - \gamma^2\sigma^2(Q_C - q_L + q_H) - \lambda + \xi_S + \xi_G = 0 \\
Q_{SI}^* &: \gamma(R^{SI} - 1)p - \lambda + \xi_S + \xi_G = 0 \\
Q_{GC}^* &: \gamma(R - 1)p - \lambda + \xi_G = 0
\end{aligned}
$$

Using the 3rd FOC in the 2nd gives, $\xi_S = \gamma(R - R^{SI})p > 0$, therefore the box constraint is binding. Directly from the 3rd FOC we can note that $\lambda > 0$, implying that the budget constraint is binding. Finally, using the 2nd FOC in the first gives an expression for the optimal portfolio. Therefore, the dealer has the following optimal strategies,

$$
\begin{aligned}
Q_C^* &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} + q_L - q_H \\
Q_{SI}^* &= g(q_L, q_H) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} - q_L + q_H \\
Q_{GC}^* &= -g(q_L, q_H)
\end{aligned}
$$

■

Proof of Proposition 2:

The result is derived from applying the implicit function theorem. Consider the two equilibrium equations,

$$
\begin{aligned}
T_1 &= \frac{g_0}{\lambda(a^*)} - C - \mathcal{SL}(s;\eta) = 0 \\
T_2 &= \lambda'(a^*)a^* - \lambda(a^*) - \lambda'(a^*)sg_0 = 0
\end{aligned}
$$

The sensitivities of T_1 and T_2 respect to equilibrium variables $R - R^{SI}$ and a^* are,

$$
\begin{aligned}
\frac{\partial T_1}{\partial s} &= -\frac{\partial \mathcal{SL}(R - R^{SI};\eta)}{\partial (R - R^{SI})} \\
\frac{\partial T_1}{\partial a^*} &= -\frac{\lambda'(a^*)g_0}{\lambda(a^*)^2} \\
\frac{\partial T_2}{\partial s} &= -\lambda'(a^*)g_0 \\
\frac{\partial T_2}{\partial a^*} &= \lambda''(a^*)(a^* - sg_0)
\end{aligned}
$$

Therefore, the determinant of the Jacobian is,

$$
\begin{aligned}
|J| &= \frac{\partial T_1}{\partial s}\frac{\partial T_2}{\partial a^*} - \frac{\partial T_1}{\partial a^*}\frac{\partial T_2}{\partial s} \\
&= -\frac{\partial \mathcal{SL}(s;\eta)}{\partial s}\lambda''(a^*)(a^* - sg_0) - \frac{(\lambda'(a^*))^2}{\lambda(a^*)^2}g_0^2 < 0
\end{aligned}
$$

And the partial derivatives of T_1 and T_2 respect η are,

$$
\begin{aligned}
\frac{\partial T_1}{\partial \eta} &= -\frac{\partial \mathcal{SL}(s;\eta)}{\partial \eta} \\
\frac{\partial T_2}{\partial \eta} &= 0
\end{aligned}
$$

Applying the implicit function theorem gives the result.

■

Proof of Lemma 3:

Given a realization $\tilde{q}_L = q\mathbb{1}_L$ or $\tilde{q}_H = q\mathbb{1}_H$, the dealer's optimization problem has the following Lagrangean,

$$
\begin{aligned}
\mathcal{L} = {} & \gamma\left[(\mu - p)Q_C + (R^{SI} - 1)pQ_{SI} + (R - 1)pQ_{GC} + (\mu - R^{SI}p)Q_I(\mathbb{1}_H - \mathbb{1}_L) + apQ_I\mathbb{1}_L + bpQ_I\mathbb{1}_H\right] \\
& -\frac{1}{2}\gamma^2\sigma^2(Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L))^2 \\
& -\lambda\left[Q_C + Q_{SI} + Q_{GC}\right] + \xi_G[Q_C + Q_{SI} + Q_{GC}] + \xi_S[Q_C + Q_{SI} - g(Q_I)] \\
& +\psi\left[2K - \{|Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L)| + |Q_{SI}| + |Q_{GC}| + Q_I\}\right] + \psi_m[q - Q_I]
\end{aligned}
$$

Giving the following FOC:

$$
\begin{aligned}
Q_C &: \quad \gamma(\mu - p) - \gamma^2\sigma^2(Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L)) - \lambda + \xi_S + \xi_G - \psi\,\text{sgn}(Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L)) = 0 && (21)\\
Q_{SI} &: \quad \gamma(R^{SI} - 1)p - \lambda + \xi_G + \xi_S - \psi\,\text{sgn}(Q_{SI}) = 0 && (22)\\
Q_{GC} &: \quad \gamma(R - 1)p - \lambda + \xi_G - \psi\,\text{sgn}(Q_{GC}) = 0 && (23)\\
Q_I &: \quad \gamma(\mu - R^{SI}p)(\mathbb{1}_H - \mathbb{1}_L) - \gamma^2\sigma^2(Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L))(\mathbb{1}_H - \mathbb{1}_L) + \gamma ap\mathbb{1}_L + \gamma bp\mathbb{1}_H \\
&\quad\; -\xi_S g'(Q_I) - \psi\,\text{sgn}(Q_C + Q_I(\mathbb{1}_H - \mathbb{1}_L))(\mathbb{1}_H - \mathbb{1}_L) - \psi - \psi_m = 0 && (24)
\end{aligned}
$$

where $\text{sgn}(\cdot)$ is the operator which is either 1 or -1 depending on whether the argument is positive or negative, respectively. Note that whenever $\psi = 0$, then we have the original FOC which are characterized in Lemma 1. Intuitively, given that $a - (R - R^{SI}) > 0$, the dealer would want to intermediate as much as possible, that is, from the final FOC we have $\psi_m > 0$.

We separate the analysis between client longs and client shorts. In each case, the sign of dealers interdealer repos will give different representations of the balance sheet constraint. It is reasonable to assume that the restricted model will be some sort of "continuation" of the unrestricted case. Therefore, we assume that dealers will have the same type of interdealer repo trade and verify that in fact it is an equilibrium.

Client Short

In the unrestricted model, dealers' optimal strategies are,

$$
\begin{aligned}
Q_C &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} - Q_I\\
Q_{SI} &= Q_I + g(Q_I) - \frac{\mu - R^{SI}p}{\gamma\sigma^2}\\
Q_{GC} &= -g(Q_I)\\
Q_I &= q
\end{aligned}
$$

Because we expect Q_I to be relatively large whenever the balance sheet constraint binds, it is natural to assume that $\text{sgn}(Q_{GC}) = -1, \text{sgn}(Q_{SI}) = 1$. Also, because $\mu > R^{SI}p$, we expect the dealer's final cash position to be positive, that is, $\text{sgn}(Q_C + Q_I) = 1$. We denote $\overline{q}^H$ as the maximum amount intermediated which is to be determined. In that case, because $\text{sgn}(Q_C + Q_I) = \text{sgn}(Q_{SI}) = 1$, from FOC (21) and (22) we have,

$$Q_C^* = \frac{\mu - R^{SI}p}{\gamma\sigma^2} - \overline{q}^H.$$

Because $\text{sgn}(Q_{GC}) = -1$ FOC (23) implies that $\lambda > 0$, making the budget bind. In addition, because $\text{sgn}(Q_{GC}) = -\text{sgn}(Q_{SI}) = -1$, FOC (22) and (23) imply that $\xi = \gamma(R - R^{SI})p + 2\psi > 0$, making the SI box constraint bind.

The above observations imply,

$$\begin{aligned} Q_{SI}^* &= g(Q_I^*) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} + \overline{q}^H \\ Q_{GC}^* &= -g(Q_I^*) \end{aligned}$$

Finally, the total amount intermediated is $Q_I^* = \overline{q}^H$ and is determined by the balance sheet constrain, which in this case is,

$$Q_C^* + Q_I^* + Q_{SI}^* - Q_{GC}^* + \overline{q}^H = 2(\overline{q}^H + g(\overline{q}^H)) = 2K$$

pining down $\overline{q}^H$, characterizing the optimal position.

Client Long

In the unrestricted model, dealers' optimal strategies are,

$$\begin{aligned} Q_C &= \frac{\mu - R^{SI}p}{\gamma\sigma^2} + Q_I \\ Q_{SI} &= g(Q_I) - \frac{\mu - R^{SI}p}{\gamma\sigma^2} - Q_I \\ Q_{GC} &= -g(Q_I) \\ Q_I &= q. \end{aligned}$$

As before, because we expect Q_I to be relatively large whenever the balance sheet constraint binds. Because $g' \in (0,1]$ and $\mu > R^{SI}p$ it is natural to assume that $\text{sgn}(Q_{GC}) = -1, \text{sgn}(Q_{SI}) = -1$. And that the dealer's final cash position to be positive, that is, $\text{sgn}(Q_C - Q_I) = 1$

Using FOC (22) and (23) we have $\xi_S = \gamma(R - R^{SI})p$ and $\lambda > 0$, therefore both the SI box constraint and the budget constraint bind.

$$\begin{aligned} Q_{SI}^* &= g(Q_I^*) - Q_C^* \\ Q_{GC}^* &= -g(Q_I^*) \end{aligned}$$

In this case, the balance sheet restriction takes the following form,

$$Q_C^* - Q_I^* - Q_{SI}^* - Q_{GC}^* + Q_I^* = 2Q_C^* = 2K$$

that is, the total cash trades in the interdealer market is the total size of the balance sheet.

From FOC (24), and expressions for Q_C^* and ξ we have,

$$-\gamma(\mu - R^{SI}p) + \gamma^2\sigma^2(K - Q_I) + \gamma ap - \gamma sg'(Q_I) = \psi_m.$$

Therefore, the optimal solution has $Q_C^* = \frac{\mu - R^{SI}p}{\gamma\sigma^2} + q$, until $Q_C^* = K$ which defines $\overline{q}_1^L = K - \frac{\mu - R^{SI}p}{\gamma\sigma^2}$. For $q > \overline{q}_1^L$ the optimal interdealer cash purchase stays constant at K, but the dealer keeps on intermediateing client orders, altering SI and GC interdealer repos, until $\psi_m = 0$. That is,

$$(\mu - R^{SI}p) - \gamma\sigma^2(K - Q_I) = pa - psg'(Q_I)$$

which pins down $\overline{q}_2^L$. For any $q > \overline{q}_2^L$ the dealer just intermediates $\overline{q}_2^L$. Therefore,

$$Q_I^* = \min\{q, \overline{q}_2^L\}$$

characterizing the optimal position.

■

Proof of Lemma 4:

Given p and s and taking the derivative of expression (18) with respect to a gives the equation in shows in the Lemma. To ensure that this equation has a solution, consider the solution to limit of $\frac{\partial \mathbb{E}(u(W^*)|\mathbb{1}_L)}{\partial a}$ when $K \longrightarrow \infty$, that is, the solution of the unrestricted balance sheet case, a^*_∞.[30] $\frac{\partial \mathbb{E}(u(W^*)|\mathbb{1}_L)}{\partial a}$ evaluated in $a = a^*_\infty$ is negative, because the term associated with $1 - w$ is zero and the remaining expression is strictly less than zero ($a^*_\infty > s$). Because $\lambda''(a) > c > 0$, $\lambda'(a)a - \lambda'(a)s - \lambda(a)$ strictly increases to infinity, with K sufficiently large there exists a $a^* > a^*_\infty$ which solves equation (19).

■

Proof of Proposition 4:

Denote $\hat{T}$ the right-hand side of the optimality equation in (19). Taking the implicit derivatives of the dealers optimal choice of a while holding the price and specialness fixed implies,

$$\frac{\partial a^*}{\partial K} = -\frac{\partial \hat{T}}{\partial K}\Big/\frac{\partial \hat{T}}{\partial a}.$$

Note that

$$\begin{aligned}
\frac{\partial \hat{T}}{\partial K} &= (\lambda + \gamma p(a - g_0 s))(\lambda' a - \lambda' g_0 s - \lambda)w - (\lambda + \gamma p(a - g_0 s))(a - g_0 s)(\lambda' + \gamma p)w \\
&\quad + (\lambda + \gamma p(a - g_0 s))^2 (a - g_0 s)(\lambda' + \gamma p)Kw \\
&= (\lambda + \gamma p(a - g_0 s))^2 [(a - g_0 s)(\lambda' + \gamma p)K - 1]w
\end{aligned}$$

[30]Note that $\lim_{K\longrightarrow\infty}(a-s)(\lambda(a)+\gamma p(a-s))w_a(a,s,K) = 0$ where $w_a(a,s,K) = -(\lambda'(a)+\gamma p)Kw(a,s,K)$ is the partial derivative of w with respect to a, because the exponential term converges faster to zero than a polynomial to infinity.

and

$$\begin{aligned}\frac{\partial \hat{T}}{\partial a} &= (a-g_0 s)\lambda''(1-w) - (\lambda' a - \lambda' g_0 s - \lambda) w_a \\ &\quad -(a-g_0 s)(\lambda'+\gamma p)^2 K w - (\lambda + \gamma p(a-g_0 s))(\lambda'+\gamma p) K w \\ &\quad -(\lambda+\gamma p(a-g_0 s))(a-g_0 s)\lambda'' K w - (\lambda+\gamma p(a-g_0 s))(a-g_0 s)(\lambda'+\gamma p) K w_a \\ &= (a-g_0 s)\lambda''\{1-w-(\lambda+\gamma p(a-g_0 s))Kw\} \\ &\quad +(\lambda+\gamma p(a-g_0 s))(\lambda'+\gamma p)\{(a-g_0 s)(\lambda'+\gamma p)K - 2\} K w\end{aligned}$$

where $w_a = \frac{\partial w}{\partial a}$. Note that $1 - w - (\lambda + \gamma p(a - g_0 s))Kw > 0$ because $1 - e^{-x} - xe^{-x}$ for $x > 0$. From the dealer's first order condition we have that,

$$(a - g_0 s) > \frac{\lambda}{\lambda'}$$

Given that λ is increasing, if $\lambda(0)K > 2$, then $(a - g_0 s)(\lambda' + \gamma p)K > 2$ when evaluated in a^*. This condition implies that both $\frac{\partial \hat{T}}{\partial a}$ and $\frac{\partial \hat{T}}{\partial K}$ are positive, completing the proof.

■

Proof of Proposition 5:

The result is derived from applying the implicit function theorem on the system of equations $\hat{T}_1$ and $\hat{T}_2$. Consider the two equilibrium equations,

$$\frac{\partial a^*}{\partial x} = \frac{-1}{|\hat{J}|}\left[\frac{\partial \hat{T}_2}{\partial s}\frac{\partial \hat{T}_1}{\partial x} - \frac{\partial \hat{T}_1}{\partial s}\frac{\partial \hat{T}_2}{\partial x}\right], \quad \frac{\partial s}{\partial x} = \frac{-1}{|\hat{J}|}\left[\frac{\partial \hat{T}_1}{\partial a}\frac{\partial \hat{T}_2}{\partial x} - \frac{\partial \hat{T}_2}{\partial a}\frac{\partial \hat{T}_1}{\partial K}\right] \tag{25}$$

for $x \in \{K, \eta\}$. In the above formulation $\hat{J}$ is the Jacobean of the partial derivatives of $\hat{T}_1, \hat{T}_2$. The partial derivative for $\hat{T}_1$ are given by,

$$\begin{aligned}\frac{\partial \hat{T}_1}{\partial a} &= \frac{g_0 \lambda'(a)}{\lambda^2(a)}\left[1 - e^{-\lambda(a)K} - \lambda(a) K e^{-\lambda(a)K}\right] > 0 \\ \frac{\partial \hat{T}_1}{\partial s} &= \frac{\partial \mathcal{SL}(s;\eta)}{\partial s} > 0 \\ \frac{\partial \hat{T}_1}{\partial K} &= -g_0 e^{-\lambda K} < 0 \\ \frac{\partial \hat{T}_1}{\partial \eta} &= \frac{\partial \mathcal{SL}(s;\eta)}{\partial \eta} > 0\end{aligned}$$

where the sign of $\frac{\partial \hat{T}_1}{\partial a}$ is positive because $1 - e^{-x} - xe^{-x}$ for $x > 0$. Note that the partial derivatives of $\hat{T}_2$ are the same as that of $\hat{T}$ in the proof of Proposition 4. From that proof, we were able to determine that for a high enough K, $\frac{\partial \hat{T}}{\partial a}$ and $\frac{\partial \hat{T}}{\partial K}$ are positive. We are only left to characterize it's partial derivative with respect to s and η,

$$
\begin{aligned}
\frac{\partial \hat{T}_2}{\partial s} &= -\lambda' g_0(1-w) - \lambda'(a-g_0 s)\frac{\partial w}{\partial s} + \lambda\frac{\partial w}{\partial s} - \left(\gamma\frac{\partial p}{\partial s}(a-g_o s) - \gamma p g_0\right)(a-g_0 s)(\lambda'+\gamma p)Kw \\
&\quad +(\lambda+\gamma p(a-g_0 s))g_0(\lambda'+\gamma p)Kw - (\lambda+\gamma p(a-g_0 s))(a-g_0 s)\gamma\frac{\partial p}{\partial s}Kw \\
&\quad -(\lambda+\gamma p(a-g_0 s))(a-g_0 s)(\lambda'+\gamma p)K\frac{\partial w}{\partial s} \\
&= -\lambda' g_0\{1-w-(\lambda+\gamma p(a-g_0 s))Kw\} + (\lambda+\gamma p(a-g_0 s))\gamma p(a-g_0 R)\{(a-g_0 s)(\lambda'+\gamma p)K-2\}\frac{Kw}{R^{SI}} \\
\frac{\partial \hat{T}_2}{\partial \eta} &= 0
\end{aligned}
$$

Note that if K is large enough then $\frac{\partial \hat{T}_2}{\partial s} < 0$. In effect,

$$
\frac{\partial \hat{T}_2}{\partial s} = -\lambda'(a)g_0 + \Xi_0(a,s)w + \Xi_1(a,s)Kw + \Xi_2(a,s)K^2 w
$$

where $\Xi_0(a,s), \Xi_1(a,s)$, and $\Xi_2(a,s)$ are finite. As K increases, we have that the terms accompanying $\Xi_0(a,s), \Xi_1(a,s), \Xi_2(a,s)$ tend to zero and $-\lambda'(a)g_0$ is bounded below by $-\lambda(0)g_0$. Thus for a large enough K we have $\frac{\partial \hat{T}_2}{\partial s} < 0$. From this, it is direct to see that,

$$
|J| = \frac{\partial \hat{T}_1}{\partial a}\frac{\partial \hat{T}_2}{\partial s} - \frac{\partial \hat{T}_2}{\partial a}\frac{\partial \hat{T}_1}{\partial s} < 0.
$$

With the above analysis, turning to equations (25) we can quickly see that $\frac{\partial a^*}{\partial \eta}, \frac{\partial s}{\partial \eta} < 0$. Similarly, $\frac{\partial s}{\partial K} > 0$ can also be observed directly. Finally, for $\frac{\partial a^*}{\partial K} > 0$ we have,

$$
\begin{aligned}
\frac{\partial \hat{T}_2}{\partial s}\frac{\partial \hat{T}_1}{\partial K} &= \lambda'(a)g_0^2 e^{-\lambda K} \\
&\quad - \left[\Xi_0(a,s) + \Xi_1(a,s)K + \Xi_2(a,s)K^2\right] w g_0 e^{-\lambda K} \qquad (26) \\
\frac{\partial \hat{T}_1}{\partial s}\frac{\partial \hat{T}_2}{\partial K} &= \frac{\partial \mathcal{SL}(s;\eta)}{\partial(s)} \times \\
&\quad (\lambda+\gamma p(a-g_0 s))^2[(a-g_0 s)(\lambda'+\gamma p)K-1]w \qquad (27)
\end{aligned}
$$

Note that $w = e^{-(\lambda+\gamma p(a-g_0 s))K}$, and as K increases the expression in equation (27) converges to zero faster than the expression in (26). Thus, for K high enough, $\frac{\partial a}{\partial K} > 0$, completing the proof.

■