Bond Market Liquidity and the Role of Repo*

Yesol Huh   Sebastian Infante

Federal Reserve Board

September 1, 2016

Abstract

This paper models the important role that repurchase agreements (repos) play in bond market intermediation. Not only do repos allow intermediaries (i.e., dealers) to finance their activities, but they also increase dealers' ability to satisfy levered client demands without having to compromise their own optimal portfolio choice. In effect, the ability to borrow specific assets for delivery allows dealers to source large quantity of assets without taking ownership of them. Larger levered client orders imply larger asset borrowing demands, thus increasing the borrowing cost for the asset (i.e., repo specialness). Dealers pass on the higher intermediation cost to their clients in the form of higher bid-ask spreads. Although this method of intermediation is optimal, the use of repos significantly increases dealers' balance sheets. Thus, limiting the size of dealers' balance sheets reduces their market making ability and hence market liquidity.

*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. Please send comments to: *sebastian.infantebilbao@frb.gov*.

1 Introduction

The perceived reduction in Treasury bond liquidity has become a highly debated topic among market participants and policy makers. Many market participants have argued that regulations such as the Supplementary Leverage Ratio (SLR) have made repurchase agreements (repos) more expensive for dealers, and that this in turn has hurt liquidity in the cash Treasury market. While it is well known that repos are widely used in cash market intermediation, especially for shorting, it is not clear how limiting dealer leverage would translate to lower liquidity. Most market commentary give matched book repos as an example of how leverage constraint would hurt the repo market. However, given that funding is usually the driver of matched book repos, it is not clear whether it should relate to cash market liquidity. In this paper, we aim to build a model that directly links repo markets with cash market liquidity by modeling how dealers use repo to intermediate in the cash market.

Repos give market makers the ability to finance and source assets without having to take ownership of them. It has long been understood that dealers rely on repo as an important source of funding, not only to finance their own positions, but also to finance counterparties' activities. In addition, repos, and more specifically, securities lending, play a critical role in enabling investors to take short positions. In effect, a short sale requires a borrowed asset. This paper accounts for these two roles that repo play, and also presents a third: to enable dealers to increase their intermediation capacity without affecting their optimal portfolio choice. Figure 1 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.

To show the above intuition, we present a two period model in which dealers service clients' trading and financing needs by purchasing and borrowing a single risky asset. The model features a continuum of dealers that intermediate for other investors, called clients, which seek to execute trades. Each dealer is a monopolist to their clients, which submit levered long or short sale orders to their dealer for execution. Client trade sizes are stochastic but depend on the bid/ask imposed by the dealer: a larger markup lowers demand, reducing the probability of a larger order size. Dealers trade in a centralized interdealer market to fill client orders, internalizing how trading affects their optimal portfolio choice.

In addition, dealers can access two types of repo markets to borrow and lend cash and assets: the *general collateral repo* (GC) and the *specific issue repo* (SI) market. The GC market allows dealers to raise funds from cash investors who are indifferent to the specific asset used as collateral. The SI market enables dealers to borrow and deliver specific assets, which are either used to cover their own short positions or to service client trades, from each other or from securities lenders. The rate at which the SI repo market clears is called the repo *specific issue rate*, and depends on the securities lenders' willingness to lend and dealers' demand for SI collateral. If the specific issue rate is lower than the prevailing GC interest rate, SI collateral is said to be trading on *special*. Repo special rates are a cost borne by all counterparties who choose to borrow though in the SI market.[1] By using these two repo markets simultaneously dealers can finance, hedge, and intermediate assets between themselves and their clients.

To satisfy client long or short demand, a dealer has to have the actual securities, either through outright ownership or sourcing it via reverse repo. Hence, dealers cannot write naked shorts or derivative-like positions. This *SI box constraint* is an important friction in our model. In our model, clients do fully levered long or short trades, thus, dealer will either sell the asset to the client and reverse repo it in, or repo out the asset to the client and buy it back. Thus, the number of actual assets that the dealer has in the box does not change after trading with the client. However, the *SI box constraint* dictates that the dealer still has to have the asset in order to do this trade. Hence, dealers may have to source in assets via the expensive SI repo market. The SI repo market allows dealers to increase the amount of assets they can source—the amount of assets purchased and borrowed—without compromising their risk adjusted return decision. That is, repos

[1] Even though the model considers just one risky asset, we assume that dealers cannot use assets received through GC repos in order to fill client trades, since a priori counterparties in the GC market do not know what asset they will receive.

allow dealers to maximize their market making ability, increasing market liquidity, while allowing them to maintain an optimal risky asset position.

Figure 2 shows the differences between primary dealers' aggregate box constraints for the 2-year on-the-run U.S. Treasury note. It indicates that the SI box constraint is mostly positive. This implies that assets which can be delivered into the SI market are either posted in the GC market or not used at all. This behavior seems puzzling, since whenever the bond trades on special this allocation implies an 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 the nature of dealers' SI box constraint results in an intermediation cost that dealers must bear.

The main results of the baseline model is that dealers' asset positions are driven solely by their risk-return considerations, and intermediation is facilitated through repo. The model shows that the fees dealers charge to their clients, which are modeled as a bid/ask spread, largely depend on the degree of repo specialness. In effect, if the cost of sourcing assets increases, dealers will pass on the cost to their clients. In addition, the model shows that filling client short orders involves sourcing more assets than filling client longs. This is because dealers must first source assets to deliver to their clients, and then source assets to hedge their position. Clients' demand is crucial to determine the degree of asset specialness, the cost of intermediation, and ultimately market liquidity. We also show that an increase in securities lenders' willingness to lend decreases specialness and bid-ask spreads.

Although this method of intermediation is optimal, it has its drawbacks. Specifically, it involves engaging in various repos and reverse repos, which can substantially increase the size of a dealer's balance sheet. Motivated by the recent regulatory initiatives to limit the amount of leverage a large bank holding company (BHC) can take, we impose an additional restriction that limits dealers' balance sheet size. We show that this size limit reduces dealers' incentives to intermediate large trades, and dealers increase bid/ask spreads, making the market more illiquid. In our model, dealers cap the amount of client orders they fill, thus limiting market depth when there is a balance sheet limit. Additionally, if the size limit becomes tighter for a given dealer, it will increase the bid-ask spread that it quotes.

The paper is structured as follows. Section 2 reviews related literature, and Section 3 gives some institutional background on the different repo markets in the United States and the role dealers play. The following section describes the model setup, detailing the main agents and how they interact. Section 5 presents the symmetric equilibrium and interpretation when dealer balance sheets are unrestricted. Section 6 incorporates restrictions on the size of dealers' balance sheet and shows how this affects their markup decision.

2 Literature Review

Our paper is related to the literature on Treasury market liquidity. Using high frequency data from the interdealer market, Fleming and Remolona (1999) studies how price and bid-ask spreads change around macroeconomic announcements. Fleming (2003) compares different liquidity measures, and finds that bid-ask spread and price impact measures are more informative than other liquidity measures such as quote size and on-the-run/off-the-run yield spreads. Mizrach and Neely (2006) studies how liquidity changed when the Treasury interdealer market transitioned from voice to electronic markets, and Fleming (1997) documents the intraday patterns of Treasury market liquidity. Goldreich et al. (2005) relates the on-the-run/off-the-run price difference to the auction cycle. These papers study the cash market without considering repo markets and focus on the interdealer markets, as there is very little data on the over-the-counter dealer-customer markets.

More directly related to our paper are the studies on repos and their specialness, and the asset pricing implications. First documented by Duffie (1996), repo specialness refers to a significantly low repo rate paid to borrow a security that is in high demand. 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. Jordan and Jordan (1997) document this pricing relationship empirically. Krishnamurthy (2002) studies the spread between newly issued U.S. Treasuries (termed *on-the-run*) and existing securities of similar characteristics (termed *off-the-run*). He shows that trading profits from shorting on-the-run securities and buying existing securities are close to zero due to different repo financing costs: on-the-run securities can trade on special implying a significant shorting cost.

The on-the-run/off-the-run price or yield differences that are studied in these papers are often used as a measure of liquidity in Treasury markets, as the price difference may be due to the liquidity premium embedded in the more-liquid on-the-run security. To that extent, these papers can be thought of as studying how repo markets affect cash market liquidity and vice versa. However, as Krishnamurthy (2002) points out, the price difference may simply due to cost of arbitrage stemming from repo specialness. Our main object of interest is the bid-ask spread, which is a more direct measure of liquidity. We also allow price to change due to change in repo specialness, similar to Krishnamurthy (2002); however, we only have one asset, so there is no arbitrage relationship, and we do not model liquidity premium.

In terms of the model setup, our model is closest to the literature on how market makers' inventory management affects their liquidity provision. In a setting where dealers are risk averse or have constraints on the size of their inventory, dealers adjust their quotes to change order flow so that they can bring the inventory size back to their optimal position. Amihud and Mendelson (1986) model a monopolistic, risk-neutral dealer with inventory size constraints. Stoll (1978) model a risk-averse monopolistic dealer, while Ho and Stoll (1983) model a market with multiple dealers. In most of these models, while the prices that dealers quote will change the order arrival rate of buy and sell trades, it is still probabilistic, and they have to wait until it arrives. In some models, dealers may trade in the interdealer market to offset the inventory, as in Ho and Stoll (1983). Since there is a finite number of dealers in that model, they behave strategically and may sometimes opt to bear inventory risk instead of laying it off in the interdealer market.

Our model has some similarities and differences with other models in this literature. As in the above papers, dealers are risk-averse and have to be compensated for deviating from their optimal portfolio. Also, the bid and ask spreads set by the dealers affect customer demand, but in our setup, dealer quotes alter the order size rather than the order arrival rate. Because our focus is on the intersection of repo markets and market making, we simplify other dimensions relative to the existing literature. We have a continuous mass of dealers and a frictionless interdealer market, thus dealers can easily lay off inventory risk through the interdealer market. Frictionless interdealer markets, and the ability to source assets through repo, allows dealers to intermediate trades without compromising their optimal portfolio. For simplicity we assume each dealer has a separate set of customers, therefore they do not compete to attract customer order flows. Another crucial difference is that in our model, dealers first need to source the securities in order to facilitate trade.

The 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 particular market structure, where clients pay dealers to service trades, which allows us to gauge the degree of market liquidity through dealer markups. We also consider two distinct repo markets for the same underlying collateral, SI and GC repo markets, which are used either to source or to finance assets and are subject to different box constraints.

Lastly, securities lenders are 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 long-term asset positions. 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 which reflects their willingness to supply assets. We look at how the equilibrium changes with their willingness parameter.

3 Institutional Setting

In the United States, there are two distinct repo markets. The most commonly known is the U.S. tri-party repo market. This market is considered a wholesale funding market, where financial firms, typically large broker-dealers, can raise short term secured debt from cash investors. In this repo market, given a certain collateral class, borrowers have the flexibility to choose from a range of assets that they may post as collateral. In effect, cash lenders in this market only value collateral as a back-stop to a borrower default, implying a wide variety of collateral is acceptable. This makes the tri-party market a GC repo market. The tri-party market's collateral flexibility is also valued by cash borrowers, who can manage their collateral holdings by exchanging assets within a collateral class whenever different asset demands arise. In this market, a clearing agent stands between cash borrowers and lenders, providing clearing services and extending intraday credit to borrowers who may need to access their asset during the day.[2]

The second repo market is known as the bilateral repo market, where trades are made bilaterally and can be tailored to accommodate specific needs. Importantly, trades in the bilateral market can be against specific collateral, allowing investors to borrow a particular asset in order to short sell or deliver to another counterparty. This makes the bilateral market an SI repo market. When demand for borrowing a specific issue is high, the bilateral repo rate can be lower than the tri-party repo rate. This incentivizes the original collateral owners to lend their securities to reap the difference between GC and SI repo rates. It is important to note that bilateral repos are also an important means of funding for investors who do not qualify to participate in the tri-party market. These lower quality firms typically issue bilateral repos to large dealers, who in turn access the tri-party market to raise funds for the initial repo; a process known as rehypothecation.

Dealers are the primary market participants who access both repo markets. They use the tri-party market to raise funds and manage collateral, and use the bilateral market to source assets and

[2]In recent years the tri-party market has been reformed to reduce the amount of intraday credit needed for dealers to access their assets. More information on tri-party reform can be found in `http://www.newyorkfed.org/banking/tpr_infr_reform.html`.

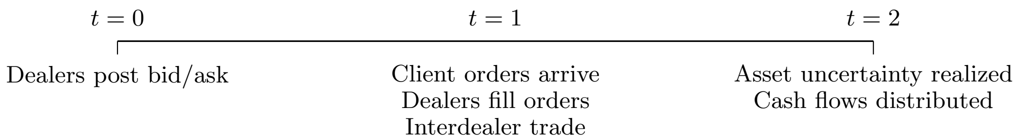


Figure 3: Model Timeline

fill client demands. Dealer counterparties in the bilateral market are their client base who demand trading and financing services, and securities lenders who effectively lend their securities to receive cash at the special rate. Cash investors in the tri-party market are largely money market funds and securities lenders who reinvest the cash received in their securities lending activities. Adrian et al. (2013) have a more detailed overview on how these markets are related and the role securities dealers play.

4 Model Setup

The model consists of 3 periods $t \in \{0, 1, 2\}$. In $t = 0$, dealers post their bid and ask to clients. At $t = 1$, client orders are realized, and order sizes are smaller for larger bid/ask spreads. Dealers receive both buy and sell orders, a fraction of which will be levered. That is, a portion of client orders will be a cash and repo trade with the dealer simultaneously, establishing either levered long or short positions. Throughout the model, we will assume SI repos will be trading special, which implies that dealers have to bear a cost to source a specific asset.

4.1 Assets & Contracts

The risky asset will have a 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 GC assets from outside investors. Only secured debt is allowed (i.e., repo) to raise funding.

The price of the risky asset p and the specific issue repo rate R^S will be determined by market clearing. We will consider settings where in equilibrium, the repo rate on SI repos are below the risk free rate: $R^S < R$. 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.

For simplicity, we assume that repos do not have haircuts, yet are risk free, which happens since we assume contract enforceability and no limited liability. This assumption simplifies the analysis greatly and keeps the focus on dealers' use of repo to borrow assets in order to deliver them to clients. We can relax this assumption somewhat by incorporating haircuts to make GC repos "virtually risk free"[3] and by allowing dealers to charge different haircuts to their clients, depending on whether they submit a levered long or short sale.[4] Adding these features complicates the analysis without significantly altering the results.[5]

The model's state variables in $t = 0$ are dealers' initial asset position D and dealers' initial cash holdings W. Although these initial variables are not crucial to the theoretical model, they allow us to better interpret the data. 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 $R \times pQ$ or $R^S \times pQ$ in cash, depending on whether it was a GC or SI repo, and returns the asset back to the cash borrower. It is useful to denote repo contracts by the amount of assets delivered as collateral rather than total amount borrowed or lent since it underscores the impact of box constraint.

4.2 Agents

There are three types of agents in the economy: dealers, securities lenders, and clients. Each client con only interact with one dealer; thus, there is no outright competition between dealers. Dealers are endowed with an initial portfolio in $t = 0$, but will have the ability to rebalance their position once a client order is received. Dealers service their client orders, which may be be buys or sales,

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

[4] Infante (2015) shows that SI repo haircuts can be negative whenever investors need to source an asset.

[5] The assumption on repos' riskiness implies that the current version of the model is not well suited to study systemic risks that come from rehypothecation.

or levered long or short positions. Securities lenders are long-term investors who already hold their optimal portfolio in $t = 0$, but have the ability to lend their securities to dealers.

4.2.1 Dealers

There is a continuum of dealers with CARA utility and risk aversion γ which consume the payoff from their portfolio in the final period. Specifically, their final payoff consists of cash flows from their risky investments, payoff from their repos, and fees in the form of bid/ask from clients. Therefore, a dealer's final wealth takes the following form:

$$\begin{aligned}\tilde{W}_2 &= (\tilde{v}-p)Q_D + (R^S-1)pQ_R^S + (R-1)pQ_R^G \\ &\quad -(\tilde{v}-\phi_L R^S p)\tilde{Q}_L + (\tilde{v}-\phi_S R^S p)\tilde{Q}_S + (1-\phi_L)p\tilde{Q}_L - (1-\phi_S)p\tilde{Q}_S + ap\tilde{Q}_L + bp\tilde{Q}_S \\ &\quad +\tilde{v}D + W\end{aligned}$$

where Q_D is the dealer's portfolio rebalancing in the cash market, Q_R^S is the amount of assets sourced through SI repos, and Q_R^G the amount of assets received as collateral through GC repos (i.e. cash lending); all are chosen in $t = 1$. Transactions with clients are not included in Q_D, Q_R^S, and Q_R^G. The second line in expression (1) highlights the effect of clients' orders on the dealer's portfolio. The size of a client's buy and sale orders are given by $\tilde{Q}_L$ and $\tilde{Q}_S$, of which a fraction ϕ_L and ϕ_S are accompanied by repo orders, respectively. Note that a client buy order implies a negative position for the dealer, thus it is associated with a $-(\tilde{v} - p)$. D is the dealer's initial portfolio position and W is the amount of cash the dealer has initially. D and W characterize the dealer's initial portfolio endowment, and we assume $W + pD > 0$. Finally, $ap\tilde{Q}_L$ and $bp\tilde{Q}_S$ are the profits from the bid (b) and ask (a) spreads. The dealer chooses a and b at $t = 0$, and the tradeoff is between spreads and 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 constraint. 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 orders: $g(\phi\tilde{Q}_L, \phi\tilde{Q}_S) > 0$. This assumption captures the intuition that dealers need to, at least in part, access assets in order to establish client positions.[6] Specifically, given client order sizes $\tilde{Q}_L = Q_L$ and $\tilde{Q}_S = Q_S$, the dealer's SI box constraint is given by,

$$D + Q_D + Q_R^S - (1 - \phi_L)Q_L + (1 - \phi_S)Q_S \geq g(\phi Q_L, \phi_S Q_S). \tag{1}$$

That is, the total amount of SI assets the dealer can access—its initial holdings, its assets bought in the interdealer market, assets source via SI repos, and unlevered buy/sell orders which alters the dealer's inventory—must be enough to deliver to its levered clients in the form of g. Note that with levered long and 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.[7] Although the presence of g implies a cost that dealers must bear to intermediate levered trades, there is suggestive evidence that the friction exists. In effect, Figure 2 shows the right hand side of U.S primary dealers SI and GC box constraint for 2 and 10 year on-the-run Treasury note. The difference between the two lines is the empirical counterpart of g, which is almost always positive.[8] Table 1 shows summary statistics on the size of the SI and GC box constraint, and the fraction of the time where the former is greater than zero.

[6] In effect, if $g = 0$, dealers would be able to write "naked" shorts and longs with their clients.

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

[8] Specifically, the difference shows the amount of on-the-run assets which are posted as collateral in the GC market, which as we will show, constitutes an intermediation cost dealers must bear.

4.2.2 Securities Lenders

We model securities lenders in a reduce form to focus on dealer behavior. Securities lenders lend assets through SI repos based on a reduced form supply function $\mathcal{SL}(R-R^S;\eta)$ with $\frac{\partial \mathcal{SL}}{\partial(R-R^S)}>0$. That is, the more the repo trades on special, the more 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 $\frac{\partial \mathcal{SL}}{\partial \eta}>0$.

4.2.3 Clients

Client order sizes are stochastic and exogenously specified. Specifically, the size of client orders are independent and distributed $\tilde{Q}_L, \tilde{Q}_S \sim Exp(\lambda(x))$ ($x \in \{a,b\}$), where λ is a function of the dealer's bid/ask quote. We assume λ, λ', and $\lambda'' \geq c > 0$ where c is an arbitrarily small constant. Clients can be interpreted as liquidity traders, who can only access one dealer, and are price sensitive to the dealer's markup: their trade sizes depend on their dealer's bid/ask quote.

Note that a levered client order is in fact two transactions: an asset purchase (sale) with a client repo (reverse repo).

5 Optimal Strategies and Symmetric Equilibrium in Unrestricted Case

In this section, we solve for dealers' optimal strategies in $t=1$. Here we shall assume that dealers have no balance sheet restrictions. For simplicity, we characterize the resulting symmetric equilibrium. We then turn to the optimal bid/ask which determines the order intensity.

5.1 Optimal Strategies & Equilibrium in Interdealer Market

Given an initial total position D, W and order size $\tilde{Q}_L = Q_L$ and $\tilde{Q}_S = Q_S$, the dealer's final payoff takes the expression in equation (1). Therefore, the dealer solves the following problem,

$$\max_{\{Q_D, Q_R^S, Q_R^G\}} \mathbb{E}(u(\tilde{W}_2)|\tilde{Q}_L = Q_L, \tilde{Q}_S = Q_S),$$

subject to

$$\begin{aligned} pQ_D + pQ_R^S + pQ_R^G - p(1-\phi_L)Q_L + p(1-\phi_S)Q_S &\leq W \quad (2) \\ D + Q_D + Q_R^S - (1-\phi_L)Q_L + (1-\phi_S)Q_S &\geq g(\phi Q_L, \phi_S Q_S) \quad (3) \\ D + Q_D + Q_R^S + Q_R^G - (1-\phi_L)Q_L + (1-\phi_S)Q_S &\geq 0. \quad (4) \end{aligned}$$

(2) is the dealer's budget constraint, (3) is the SI box constraint, and (4) the global box constraint. In (2), we assume that fees reaped from intermediation do not alter the dealer's budget. This is a simplifying assumption merely for tractability: we do not want client order sizes to affect the dealer's budget constraint. In reality, this effect, if any, is likely to be negligible. Given an asset price p, a GC repo rate R, and a specials rate R^S, dealers will employ the following optimal strategies.

Lemma 1 (Dealer's Optimal Strategy — Unrestricted Case). *Given asset price p, SI repo rate R^S, and secured funding rate $R > R^S$; a dealer's optimal rebalancing strategy after receiving client orders $\tilde{Q}_L = Q_L$ and $\tilde{Q}_S = Q_S$ is:*

$$\begin{aligned} Q_D^* &= \frac{\mu - R^S p}{\gamma\sigma^2} - D + Q_L - Q_S \\ Q_R^{S*} &= g(\phi_L Q_L, \phi_S Q_S) - \frac{\mu - R^S p}{\gamma\sigma^2} - \phi_L Q_L + \phi_S Q_S \\ Q_R^{G*} &= \frac{W}{p} + D - g(\phi_L Q_L, \phi_S Q_S). \end{aligned}$$

Proof. See appendix. □

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

$$D + Q_S - Q_L + Q_D^* = \frac{\mu - R^S p}{\gamma \sigma^2}. \tag{5}$$

Because dealers have access to frictionless interdealer markets, it allows them to optimally adjust their portfolio to accommodate their clients' trades. Figure 4 shows how dealers rebalance portfolio. To better show the intuition, we consider a simple example where the dealer only receives a buy or sale order. That is, either $(Q_L, Q_S) = (Q, 0)$ or $(Q_L, Q_S) = (0, Q)$. Additionally, assume

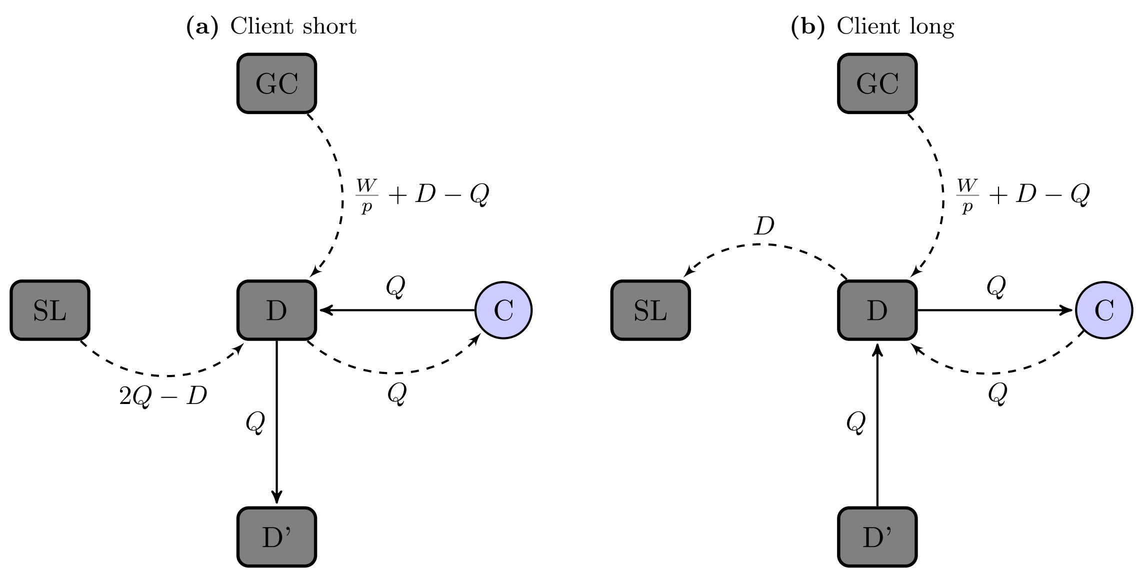
$$\begin{aligned} \phi_L = \phi_S &= 1 \\ D &= \frac{\mu - R^S p}{\gamma \sigma^2} \\ g(Q_L, Q_S) &= Q. \end{aligned}$$

Then, if the dealer receives a client short order, its optimal strategy will be

$$\begin{aligned} Q_D^* &= -Q \\ Q_R^{S*} &= 2Q - D \\ Q_R^{G*} &= \frac{W}{p} + D - Q. \end{aligned}$$

This is shown in Panel (a) of Figure 4. To maintain its optimal portfolio, the dealer wants to sell Q of the risky assets in the interdealer market. However, to do so, it has to source an additional Q assets from the SI repo market because of g; thus, it will source $2Q - D$ assets. Note that in the client long scenario, this does not happen as the dealer buys Q from the interdealer market.

Figure 4: Simple example
D and D' stands for dealer, GC stands for general collateral market, SL stands for securities lender, and C stands for client. 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. Direction of the arrows will flip if the values are negative.



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 i to highlight that the equilibrium price is determined by aggregating across all dealers.

$$\int_i Q_{Di} di = 0. \tag{6}$$

The SI repo market, which also incorporates securities lenders' asset supply, clears through the following equation,

$$\int_i Q^S_{Ri} di = \mathcal{SL}(R - R_S; \eta). \tag{7}$$

Denoting by $\int_i Ddi$ dealers' aggregate position at the beginning of period $t = 0$, and assuming symmetric strategies gives $\int_i Ddi = D$.

Proposition 1 (**Interdealer Equilibrium — Unrestricted Case**). *Given dealers' initial position W and D, with $W + pD > 0$, GC repo rate of R, securities lending function sufficiently small enough, and symmetric dealer bid and ask spreads a and b, dealers' optimal strategies characterized in Lemma 1 result in an asset price p and SI repo rate $R^S < R$ which solves the following system of equations:*

$$\frac{\mu - R^S p}{\gamma\sigma^2} = D - \frac{\mathbb{P}(CL)}{\lambda(a)} + \frac{\mathbb{P}(CS)}{\lambda(b)} \tag{8}$$

$$\mathcal{SL}(R - R^S; \eta) = \mathbb{E}(g(\phi_L \tilde{Q}_L, \phi_S \tilde{Q}_S)) - D + \frac{(1-\phi_L)\mathbb{P}(CL)}{\lambda(a)} - \frac{(1-\phi_S)\mathbb{P}(CS)}{\lambda(b)} \tag{9}$$

where $\mathbb{P}(CL)$ and $\mathbb{P}(CS)$ are the probability of a client long and short order, respectively.

Proof. The proof stems from considering dealers' strategies in Lemma 1, imposing market clearing conditions (6) and (7), and applying the law of large numbers for client orders. □

Equation (8) shows how the price responds to client order flow: An increase in client longs increases the price, whereas an increase in client shorts reduces the price. Equation (9) 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, then dealers need to source more assets in the SI market which puts upward pressure on repo specialness. In addition, changes in ϕ_L orϕ_S changes the amount of assets in the market due to unlevered client order flow; giving a predictable effect on repo specialness.

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

$$\mathcal{SL}(R - R^S; \eta) = \mathbb{E}(g(\phi_L \tilde{Q}_L, \phi_S \tilde{Q}_S)) - \left(\frac{\mu - R^S p}{\gamma \sigma^2}\right) - \frac{\phi_L \mathbb{P}(CL)}{\lambda(a)} + \frac{\phi_S \mathbb{P}(CS)}{\lambda(b)}.$$

The above partial equilibrium equation highlights what market participants often comment regards repo specialness: an increase in client short base increases repo specialness. In effect, if ϕ_S increases, $R - R^S$ would need to increase to clear the market. But this observation neglects the effect of repo specialness on the asset price. The general equilibrium solution in (9) 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.

Dealer's final wealth when using the optimal strategy is gven by

$$\begin{aligned}\tilde{W}_2^* | \{\tilde{Q}_L = Q_L, \tilde{Q}_S = Q_S\} &= (\tilde{v} - R^S p)\left(\frac{\mu - R^S p}{\gamma \sigma^2}\right) - (R - R^S) p g(\phi_L Q_L, \phi_S Q_S) \\ &\quad + RpD + RW + apQ_L + bpQ_S.\end{aligned}$$

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

$$\begin{aligned}\mathbb{E}(u(W^*) | \tilde{Q}_L = Q_L, \tilde{Q}_S = Q_S) &= -\exp\left\{-\gamma \left[\frac{1}{2}\left(\frac{(\mu - R^S p)^2}{\gamma \sigma^2}\right) - (R - R^S) p g(\phi_L Q_L, \phi_S Q_S)\right.\right. \\ &\quad \left.\left. + RpD + RW + apQ_L + bpQ_S\right]\right\}\end{aligned}$$

where p and R^S are given by Proportion 1.

5.2 Posting Bid/Ask at $t = 0$

Having the dealer's optimal strategy and final expected utility, we turn to characterizing the dealers optimal mark up before receiving client orders. For simplicity assume that,

$$g(\phi_L Q_L, \phi_S Q_S) = \phi_L Q_L + \phi_S Q_S. \tag{10}$$

This assumption implies that there is no internalization in dealer's levered operations. In a way, this stark assumption implies that dealers cannot offset client levered longs and shorts to intermediate assets, but we adopt it for tractability. Therefore, the dealer's expected payoff is given by,

$$\mathbb{E}(u(W^*)) = -\Gamma \mathbb{E}\left(\exp\left\{\gamma p[(R - R^S)(\phi_L \tilde{Q}_L + \phi_S \tilde{Q}_S) - (a\tilde{Q}_L + b\tilde{Q}_S)]\right\}\right) \tag{11}$$

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

$$\mathbb{E}(u(W^*)) = -\Gamma \frac{\lambda(a)}{(\lambda(a) + \gamma pa - \gamma p(R - R^S)\phi_L)} \frac{\lambda(b)}{\lambda(b) + \gamma pb - \gamma p(R - R^S)\phi_S)} \tag{12}$$

For the integral to exist, we need $\lambda(a) + \gamma pa - \gamma p(R - R^S)\phi_L > 0$ and $\lambda(b) + \gamma pb - \gamma p(R - R^S)\phi_S > 0$. Recall that dealers do not internalize the effect of their optimal strategy on the price or specials rate. This leads to the following Lemma on dealers' optimal bid/ask.

Lemma 2 (**Dealer's Optimal Bid/Ask — Unrestricted Case**). *Given a dealer's initial position W and D, with $W + pD > 0$, GC repo rate of R, securities lending function sufficiently small* *enough, and g as in* (10)*; the dealer's optimal bid and ask spread solve the following equations*

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

with $a^* > (R - R^S)\phi_L$ *and* $b^* > (R - R^S)\phi_S$.

Proof. Taking the first order condition of expression (12), deduced from Proposition 1 with g as in equation (10) gives the Lemma's optimality conditions. In addition, note that $a^* > (R - R^S)\phi_L$ and $b^* > (R - R^S)\phi_S$ because

$$
\begin{aligned}
\frac{\partial \mathbb{E}(u(W^*))}{\partial a}\Big|_{a=(R-R^S)\phi_L} &= -\Gamma \frac{\lambda(b)}{\lambda(b) + \gamma p b - \gamma p(R - R^S)\phi_S} \times \\
&\quad \frac{-\lambda((R - R^S)\phi_L)\gamma p}{(\lambda(a) + \gamma p a - \gamma p(R - R^S)\phi_L)^2} > 0,
\end{aligned}
$$

implying $a^* > (R - R^S)\phi_L$. The same argument holds for b^*. □

Note that each 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, from a partial equilibrium setting, client's market liquidity is decreasing 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 η, the willingness of securities lenders to provide assets. To do this, we fist further simplify the model by setting $\phi_L = \phi_S$. This simplifying assumption eliminates imbalances that affect repo specialness because of assets being added to, or drained from, the SI repo market mechanically through client trades.

If $\phi_L = \phi_S = \phi$, then from Lemma 2 we have that $a^* = b^*$, and therefore, the probability of a buy or sell order is equally likely—that is, $\mathbb{P}(CL) = \mathbb{P}(CL)$. The equilibrium level of specialness, $R - R^S$, incorporating dealers' optimal markup, is given by the following equations:

$$\begin{aligned} T_1 &:= \mathbb{E}(g(\phi\tilde{Q}_L, \phi\tilde{Q}_S)) - D - \mathcal{SL}(R - R^S; \eta) = 0 \\ T_2 &:= \lambda'(a^*)a^* - \lambda(a^*) - \lambda'(a^*)(R - R^S)\phi = 0 \end{aligned}$$

Proposition 2 (Sensitivity of Liquidity to η — Unrestricted Case). *Given the same assumptions from Lemma 2, with $\phi_L = \phi_S = \phi$, increases in securities lenders' willingness to provided assets decreases repo specialness and dealers' optimal bid/ask. Specifically,*

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

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

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

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

6 Optimal Strategies and Symmetric Equilibrium in 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 (BHSs) 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.

In the context of our model, the specific functional form of the leverage ratio restriction used in the SLR would be difficult to model. But, assuming the the BHC has a fixed level of equity, a leverage restriction can be translated into a size restriction on the dealer's balance sheet. In order to understand how this restriction may affect the dealer's intermediation activities, we first have to translate the model's outcome onto a balance sheet.

Given the additional complexity of incorporating a limit on dealers' balance sheet, this section will adopt a number of simplifying assumptions relative to the general model presented in Section 4. First, we assume that each dealer receives only one levered long or short order.[9] These orders can still be arbitrarily large, but the focus is on how each of them affect the dealer's ability to intermediate. Given that each dealer will receive one client order, g will be a function of that order's size.

Without loss of generality, we will assume that $\mu > R^S p$, which implies that dealers' unrestricted optimal asset position is positive. To simplify dealers' initial size, we assume all dealers' initial asset position is $D = 0$, and their initial wealth $W > 0$ is arbitrarily small.

The innovation of this section is that dealers have an upper bound C on the size of their balance sheet. 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 $2C$. In addition, dealers have an additional choice variable: how much of the order to

[9] This implies that $\phi_L = \phi_S = 1$.

intermediate Q_I. In its general form, the balance sheet restriction can be written as,

$$\frac{W}{p} + |Q_D + Q_I(\mathbb{1}_{CS} - \mathbb{1}_{CL})| + |Q_R^S| + |Q_R^G| + Q_I \leq 2C. \tag{13}$$

Similar to the the assumption adopted for dealers' budget constraint, we assume that dealers' markups do not affect the size of their balance sheet. These cash flows are likely to be negligible relative to the total size of a BHC's balance sheet. In this notation $\mathbb{1}_{CL}$ is an indicator function that equals 1 if the client order is a levered long and zero otherwise. $\mathbb{1}_{CS}$ is defined similarly for client shorts. The five components of equation (13) are a dealer's initial wealth, its 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 Interdealer Market

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

$$\max_{\{Q_D, Q_R^S, Q_R^G, Q_I\}} \mathbb{E}(u(W)|\tilde{Q}_L = \mathbb{1}_{CL}Q, \tilde{Q}_S = \mathbb{1}_{CS}Q)$$

subject to,

$$\begin{aligned} pQ_D + pQ_R^S + pQ_R^G &\leq W \\ Q_D + Q_R^S &\geq g(Q_I) \\ Q_D + Q_R^S + Q_R^G &\geq 0 \\ \frac{W}{p} + |Q_D + Q_I(\mathbb{1}_{CS} - \mathbb{1}_{CL})| + |Q_R^S| + |Q_R^G| + Q_I &\leq 2C \\ Q_I &\leq Q \end{aligned}$$

That is, the dealer's problem is the same as Section 5 except now dealers face a balance sheet restriction (equation (13)) and must also decide how much to intermediate ($Q_I \leq Q$). 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^S, secured funding rate $R > R^S$, and $\mu > R^S p$, then upon receiving a client long order $\tilde{Q} = Q\mathbb{1}_{CL}$ the dealer's optimal portfolio is equal to the solution of Lemma 1 with $Q_I = Q$ whenever $Q < C - \frac{\mu - R^S p}{\gamma\sigma^2} := \overline{Q}_1^L$. If $Q \geq \overline{Q}_1^L$ then the dealer's optimal strategy is*

$$\begin{aligned} Q_D^* &= \frac{\mu - R^S p}{\gamma\sigma^2} + \overline{Q}_1^L \\ Q_R^{S*} &= g(Q_I^*) - \frac{\mu - R^S p}{\gamma\sigma^2} - \overline{Q}_1^L \\ Q_R^{G*} &= \frac{W}{p} - g(Q_I^*) \\ Q_I^* &= \min\{Q, \overline{Q}_2^L\} \end{aligned}$$

where $\overline{Q}_2^L$ solves $\overline{Q}_2^L = \overline{Q}_1^L + \frac{ap - p(R - R^S)g'(\overline{Q}_2^L)}{\gamma\sigma^2}$.
Upon receiving a client short order $\tilde{Q} = Q\mathbb{1}_{CS}$, the dealer's optimal portfolio is equal to the solution of Lemma 1 with $Q_I = Q$ whenever $Q < \overline{Q}^S$ which solves $g(\overline{Q}^S) + \overline{Q}^S = C$. If $Q \geq \overline{Q}^S$, then the dealer's optimal strategy is,

$$\begin{aligned} Q_D^* &= \frac{\mu - R^S p}{\gamma\sigma^2} - \overline{Q}^S \\ Q_R^{S*} &= g(Q_I^*) - \frac{\mu - R^S p}{\gamma\sigma^2} + \overline{Q}^S \\ Q_R^{G*} &= \frac{W}{p} - g(Q_I^*) \\ Q_I^* &= \overline{Q}^S. \end{aligned}$$

Proof. See appendix. □

The intuition for the dealer's optimal response to a client short order is illustrated in Figure 5. For a relatively small order ($Q < \overline{Q}^S$), the dealer intermediates the trade as in the unrestricted case. If the client order size increases by ϵ as in Panel (b), dealer has to increase the size of

Figure 5: Client short with restricted balance sheet when $W = 0$

(a) $Q < \overline{Q}^S$

Asset	Liability
risky asset $\frac{\mu - R^S p}{\gamma\sigma^2}$	client repo Q
SI rev repo $g(Q) + Q - \frac{\mu - R^S p}{\gamma\sigma^2}$	GC repo $g(Q)$

(b) $Q + \epsilon \leq \overline{Q}^S$

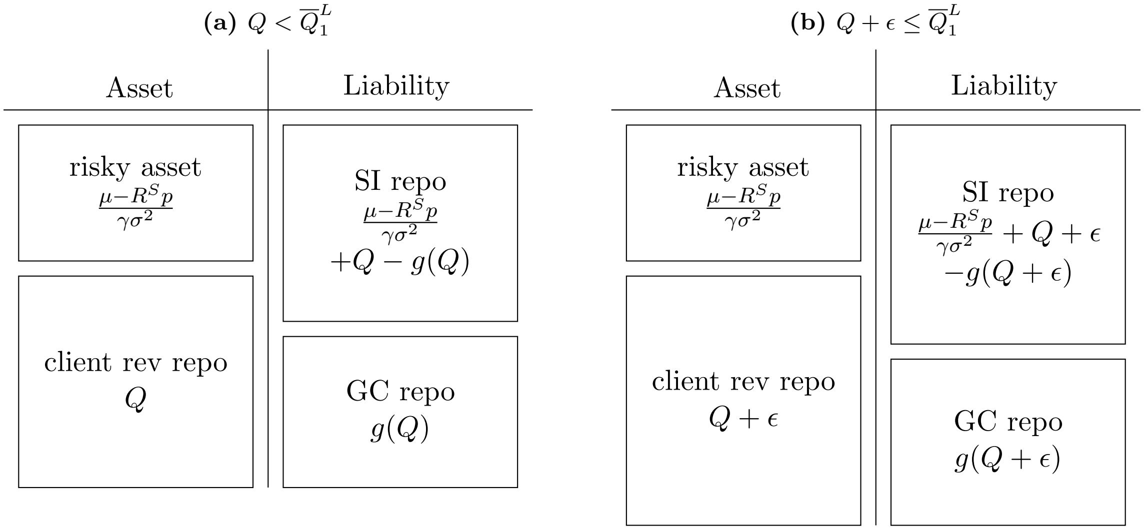
Asset	Liability
risky asset $\frac{\mu - R^S p}{\gamma\sigma^2}$	client repo $Q + \epsilon$
SI rev repo $g(Q+\epsilon) + Q + \epsilon - \frac{\mu - R^S p}{\gamma\sigma^2}$	GC repo $g(Q+\epsilon)$

his balance sheet in order to accommodate the increase. This expansion happens because of the increase in repo it issues to its client and the increase in the amount needed to intermediate the trade $g(Q)$. The dealer can do this until the balance sheet size reaches the limit C, which happens when $g(Q) + Q = C$, that is, when $Q = \overline{Q}^S$. Once the balance sheet limit is reached, the dealer will only intermediate $\overline{Q}^S$.

The difference in dealers' intermediation of long orders stem from the assumption that the unrestricted optimal portfolio is positive, i.e., $\mu > R^S p$. This implies that a dealer may accommodate larger client orders without increasing its balance sheet by compromising its risky asset position. As illustrated in Figure 6, if a client order is small ($Q \leq \overline{Q}_1^L$), the dealer will intermediate trades as in the unrestricted case. The dealer will increase its balance sheet size if client order size increases. However, when the balance sheet reaches the size limit C when $Q = \overline{Q}_1^L$, the dealer stops buying more assets in the interdealer market. This can be seen through the optimal rebalancing of risky assets in the interdealer market,

$$Q_D^* = \frac{\mu - R^S p}{\gamma\sigma^2} + \overline{Q}_1^L = C$$

Figure 6: Client long with restricted balance sheet when $Q < \overline{Q}_1^L$ and $W = 0$



For order sizes large than $\overline{Q}_L^1$ the dealer can still continue to intermediate client orders, but it cannot be done by expanding the balance sheet. Instead, as outlined in Figure 7, the dealer starts compromising its optimal asset position by decreasing its asset exposure to $\frac{\mu - R^S p}{\gamma \sigma^2} + \overline{Q}_1^L - Q$ for $Q \in (\overline{Q}_1^L, \overline{Q}_2^L)$. The dealer is willing to alter its optimal position because it receives payments through b for intermediating large orders. The dealer will stop intermediating more when the benefit from doing so is equal to the cost of altering its portfolio. That is, when Q solves

$$(\mu - R^S p) - \gamma \sigma^2 \underbrace{(Q_D^* - Q)}_{\text{Risky Asset Position}} = ap - p(R - R^S) g'(Q)$$

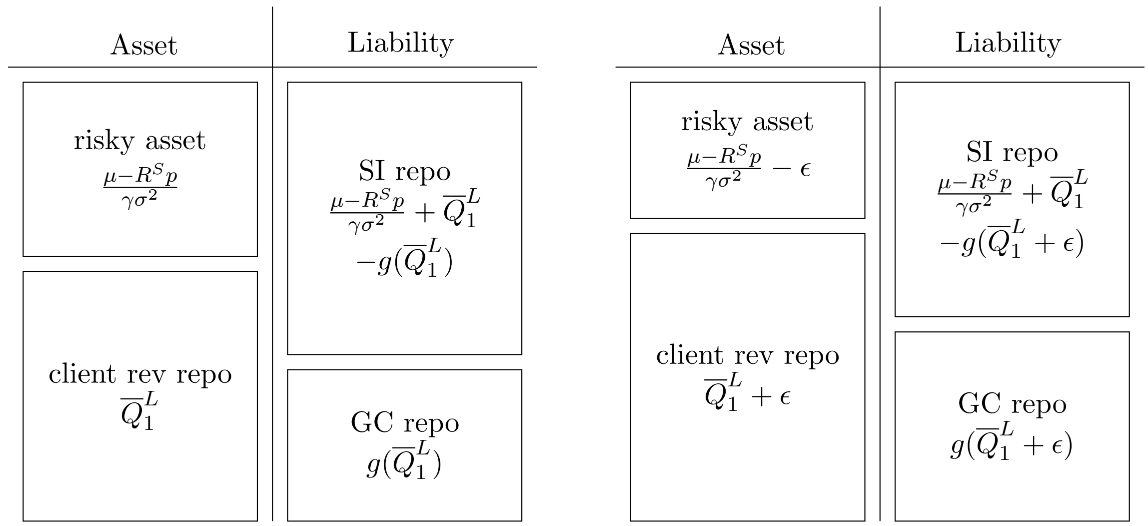
which defines $\overline{Q}_2^L$.

For both client long and short, a constraint on dealer's balance sheet limits the amount of orders it can intermediate, which we will later show leads to an increase in the markup. But first, we characterize the equilibrium in this setting.

As before, both the cash and *SI* repo market clears. To ensure that the $\mu > R^S p$, we will assume there is an exogenous supply of assets, $D > 0$, provided to the interdealer market. This

Figure 7: Client long with restricted balance sheet when $\overline{Q}_1^L = Q < \overline{Q}_2^L$ and $W = 0$

(a) $Q = \overline{Q}_1^L$ **(b)** $Q = \overline{Q}_1^L + \epsilon < \overline{Q}_2^L$



can be thought of as U.S. Treasury issuance. Thus, the cash market clearing condition is

$$\int_i Q_{Di} di = D. \tag{14}$$

And the SI repo market clears though equation (7), just as in the previous section. This gives the following interdealer equilibrium,

Proposition 3 (**Interdealer Equilibrium — Balance Sheet Restricted Case**). *Given dealers' initial position W with $W > 0$ arbitrarily small, GC repo rate of R, securities lending function sufficiently small enough, supply of assets in the cash market D sufficiently large enough, and symmetric dealer bid and ask spreads a and b; dealers' optimal strategies characterized in Lemma 3* *result in an asset price p and SI repo rate $R^S < R$ which solves the following system of equations:*

$$\frac{\mu - R^S p}{\gamma\sigma^2} = -\frac{\mathbb{P}(CL)}{\lambda(a)}[1 - e^{-\lambda(a)\overline{Q}_1^L}] + \frac{\mathbb{P}(CS)}{\lambda(b)}[1 - e^{-\lambda(b)\overline{Q}^S}] + D \tag{15}$$

$$\begin{aligned}\mathcal{SL}(R - R^S; \eta) + D &= \mathbb{P}(CL)\left[\int_0^{\overline{Q}_2^L} g(q)\lambda(a)e^{-\lambda(a)q}dq + g(\overline{Q}_2^L)e^{-\lambda(a)\overline{Q}_2^L}\right] + \\ &\quad \mathbb{P}(CS)\left[\int_0^{\overline{Q}^S} g(q)\lambda(b)e^{-\lambda(b)q}dq + g(\overline{Q}^S)e^{-\lambda(b)\overline{Q}^S}\right]\end{aligned} \tag{16}$$

where $\mathbb{P}(CL)$ and $\mathbb{P}(CS)$ are the probability of a client long and short order, respectively.

Proof. The proof stems from considering dealers' strategies in Lemma 3, imposing market clearing conditions (14) and (7), and applying the law of large numbers for client orders. □

From Proposition 3, it can be appreciated how constraints on dealers' balance sheet can have a direct impact on the underlying asset's price and repo specialness. On the one hand, given a fixed bid and ask, limited dealer intermediation skews prices depending on whether $\overline{Q}_1^L$ is larger or smaller than $\overline{Q}^S$. On the other hand, reducing dealers' balance sheet reduces the demand for interdealer repos, reducing repo specialness. This last effect depends on frictions in dealer intermediation given by g.

6.2 Posting Bid at $t = 0$

As before, having characterized a dealer's optimal strategy given clients' order flow, we get the expression for their final wealth. In this case, the calculation is slightly more involved since for relatively large client long orders, the dealer alters its portfolio position, effectively changing its expected payoff from its asset exposure.

Note that this does not occur for client shorts.[10] To simplify the analysis in this section, this version of the paper will calculate the expected payoff conditional on a short order. This implies

[10]Recall the reason behind the asymmetry is $\mu > R^S p$.

that we only have to be concerned with dealer costs and benefits to intermediating more assets, rather than the effect of its rebalancing on its portfolio. This gives,

$$\mathbb{E}(u(W^*)|\tilde{Q}_S = \mathbb{1}_{CS}Q_S) = \begin{cases} -\Gamma \exp\left\{-\gamma\left[bpQ_S - p(R-R^S)g(Q_S)\right]\right\} & \text{if } Q_S < \overline{Q}^S \\ -\Gamma \exp\left\{-\gamma\left[bp\overline{Q}^S - p(R-R^S)g(\overline{Q}^S)\right]\right\} & \text{if } Q_S \geq \overline{Q}^S \end{cases}$$

where in this case $\Gamma = \exp\left\{-\gamma\frac{1}{2}\left(\frac{(\mu - R^S p)^2}{\gamma\sigma^2}\right)\right\} > 0$ is as before, but with the section's simplifying assumption, and p and R^S are given by Proportion 3. That is, the payoff is as in the unrestricted balance sheet case, but it is capped at $\tilde{Q}_S = \overline{Q}^S$.

Similar to the previous section, for tractability we will assume that $g(Q) = Q$, with $g > 0$ and $g' \in (0, 1]$.[11] Note that in this case, from Lemma 3 we have $\overline{Q}^S = \frac{C}{2}$. Integrating over $\tilde{Q}_S$ gives,

$$\begin{aligned} \mathbb{E}(u(W^*)|CS) &= -\Gamma\left(\int_0^{\frac{C}{2}} \lambda(b)\exp\left\{-\gamma p[bq - (R-R^S)]q - \lambda(b)q\right\}dq + \right. \\ &\quad \left. \int_{\frac{C}{2}}^{\infty} \lambda(b)\exp\left\{-\gamma p[b\frac{C}{2} - (R-R^S)\frac{C}{2}] - \lambda(b)q\right\}dq\right) \\ &= -\Gamma\left(\frac{\lambda(b)}{\lambda(b) + \gamma p(b-(R-R^S))} + \frac{\gamma p(b-(R-R^S))}{\lambda(b)+\gamma p(b-(R-R^S))}e^{-[\lambda(b)+\gamma p(b-(R-R^S))]\frac{C}{2}}\right) \end{aligned}$$

Defining $w(b, Q) := e^{-[\lambda(b)+\gamma p(b-(R-R^S))]Q}$ which is between 0 and 1, we have the following Lemma for the restricted dealer's optimal mark up,

Lemma 4 (Dealer's Optimal Bid — Balance Sheet Restricted Case). *Given dealers' initial position W with $W > 0$ arbitrarily small, GC repo rate of R, securities lending function sufficiently small enough for $R^S < R$ (from Proposition 3), supply of assets in the cash market D sufficiently large enough for $\mu > pR^S$ (from Proposition 3), and $g(Q) = Q$; dealer's optimal bid solves the*

[11] 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.

following equation

$$
\begin{aligned}
0 &= \left(\lambda'(b^*)b^* - \lambda'(b^*)(R-R^S) - \lambda(b^*)\right)\left[1 - w\left(b^*, \frac{C}{2}\right)\right] \\
&\quad -(b^* - (R-R^S))(\lambda(b^*) + \gamma p(b^* - (R-R^S)))(\lambda'(b^*) + \gamma p)\frac{C}{2} w\left(b^*, \frac{C}{2}\right) \qquad (17)
\end{aligned}
$$

with $b^* > b^*_\infty$ *where* b^*_∞ *is the optimal bid with any unrestricted balance sheet given prices* (p, R^S).

Proof. See Appendix □

Lemma 4 shows how the balance sheet constraint affects a dealer's decision when setting its optimal bid. Equation 17 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 smaller expected order flow. The second highlights the limit on how much a dealer can intermediate. Note that that as C increases, the second term disappears[12], reducing the optimality condition to the unrestricted case. The Lemma also shows that for the same pair of prices (p, R^S), the optimal restricted markup is larger, which we interpret as a less liquid market.

To get an idea of 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 condition (17) changes, as well as p and R^S, which implies incorporating the changes in market clearing conditions (15) and (16).[13] In this version of the paper we will only characterize how changes in an individual dealer's balance sheet constraint affects its own bid decision. That is, we will only explore a partial equilibrium change. This gives way to the following Lemma

Lemma 5 (**Sensitivity of Dealer's Optimal Bid to Changes in Individual Restriction**). *Given the assumptions of Lemma 4 with C sufficiently large enogh, a tightening of an individual*

[12] $\lim_{C\longrightarrow\infty} \frac{C}{2} w\left(b^*, \frac{C}{2}\right) = 0$.

[13] And also, changes in the optimal ask.

dealers balance sheet constraint leads to in increase in its optimal bid. That is,

$$\frac{\partial b^*}{\partial C} < 0.$$

Proof. See Appendix. □

The result from Lemma 5 shows that—at least in a partial equilibrium setting—tighter balance sheet constraints induce dealers to increase their markups, effectively reducing liquidity for its clients. Intuitively, given that dealers are restricted from filling large client orders, they opt to increase the revenue from filling smaller ones. This increases clients' intermediation cost, making the underlying cash market less liquid.

7 Concluding Remarks

This paper presents a model of dealers' bond market making activities, specifically 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 special rate. This explains why the dealers are seemingly willing take a loss by lending cash at the specials rate (i.e. source assets) and borrow at the GC rate—satisfying the SI box constraint.

In addition, in a world where the size of a dealers' balance sheet is limited, dealers have reduced incentives to service large trade orders and increase the costs they pass onto their clients. Balance sheet limits reduces dealers' ability to intermediate large trades, reducing market depth; hence decreasing market liquidity.

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.

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

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.

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

Fleming, M. J. (1997), 'The round-the-clock market for us treasury securities', *Economic Policy Review* **3**(2).

Fleming, M. J. (2003), 'Measuring treasury market liquidity', *Economic Policy Review* (Sep), 83–108.

Fleming, M. J. and Remolona, E. M. (1999), 'Price formation and liquidity in the us treasury market: The response to public information', *The journal of Finance* **54**(5), 1901–1915.

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

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

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. (2015), 'Liquidity windfalls: The consequences of repo rehypothecation', *FEDS: Finance and Economics Discussion Series* .

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

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

Mizrach, B. and Neely, C. J. (2006), ‘The transition to electronic communications networks in the secondary treasury market’, *Review* **88**.

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

A Appendix

Proof of Lemma 1:

Given a realization $\tilde{Q}_L = Q_L$ and $\tilde{Q}_S = Q_S$, the dealer's optimization problem has the following Lagrangean,

$$\begin{aligned}
\mathcal{L} &= \gamma\Big[(\mu - p)Q_D + (R^S - 1)pQ_R^S + (R-1)pQ_R^G \\
&-(\mu - \phi_L R^S p)Q_L + (\mu - \phi_L R^S p)Q_S + (1-\phi_L)pQ_L - (1-\phi_S)pQ_S + apQ_L + bpQ_S \\
&+\mu D + W] \\
&-\frac{1}{2}\gamma^2\sigma^2(Q_D + D - Q_L + Q_S)^2 \\
&+\lambda\left[\frac{W}{p} - \{Q_D + Q_R^S + Q_R^G - (1-\phi_L)Q_L + (1-\phi_S)Q_S\}\right] \\
&+\xi_S[D + Q_D + Q_R^S - (1-\phi_L)Q_L + (1-\phi_S)Q_S - g(\phi_L Q_L, \phi_S Q_S)] \\
&+\xi_G[D + Q_D + Q_R^S - (1-\phi_L)Q_L + (1-\phi_S)Q_S + Q_R^G]
\end{aligned}$$

Giving the following FOC:

$$\begin{aligned}
Q_D^* &: \gamma(\mu - p) - \gamma^2\sigma^2(Q_D + D - Q_L + Q_S) - \lambda + \xi_S + \xi_G = 0 \\
Q_R^{S*} &: \gamma(R^S - 1)p - \lambda + \xi_S + \xi_G = 0 \\
Q_R^{G*} &: \gamma(R-1)p - \lambda + \xi_G = 0
\end{aligned}$$

Using the 3rd FOC in the 2nd gives, $\xi_S = \gamma(R - R^S)p > 0$, 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,

$$\begin{aligned}
Q_D^* &= \frac{\mu - R^S p}{\gamma\sigma^2} - D + Q_L - Q_S \\
Q_R^{S*} &= g(\phi_L Q_L, \phi_S Q_S) - \frac{\mu - R^S p}{\gamma\sigma^2} - \phi_L Q_L + \phi_S Q_S \\
Q_R^{G*} &= \frac{W}{p} + D - g(\phi_L Q_L, \phi_S Q_S)
\end{aligned}$$

■

Proof of Proposition 2: The result is derived from applying the implicit function theorem. Consider the two equilibrium equations,

$$\begin{aligned} T_1 &= \frac{1}{\lambda(a^*)} - D - \mathcal{SL}(R - R^S; \eta) = 0 \\ T_2 &= \lambda'(a^*)a^* - \lambda(a^*) - \lambda'(a^*)(R - R^S)\phi = 0 \end{aligned}$$

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

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

Therefore, the determinant of the Jacobian is,

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

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

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

Applying the implicit function theorem gives the result.

■

Reminder:

$$\begin{pmatrix} \frac{\partial (R-R^S)}{\partial \eta} \\ \frac{\partial a^*}{\partial \eta} \end{pmatrix} = -\underbrace{\begin{bmatrix} \frac{\partial T_1}{\partial (R-R^S)} & \frac{\partial T_1}{\partial a^*} \\ \frac{\partial T_2}{\partial (R-R^S)} & \frac{\partial T_2}{\partial a^*} \end{bmatrix}^{-1}}_{:=J^{-1}} \begin{pmatrix} \frac{\partial T_1}{\partial \eta} \\ \frac{\partial T_2}{\partial \eta} \end{pmatrix}$$

Proof of Lemma 3:

Given a realization $\tilde{Q}_L = \mathbb{1}_{CL}Q$ or $\tilde{Q}_S = \mathbb{1}_{CS}Q$, the dealer's optimization problem has the following Lagrangean,

$$\begin{aligned}
\mathcal{L} &= \gamma\left[(\mu-p)Q_D+(R^S-1)pQ_R^S+(R-1)pQ_R^G+W+(\mu-R^Sp)Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL})+apQ_I\right] \\
&-\frac{1}{2}\gamma^2\sigma^2(Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL}))^2 \\
&+\lambda\left[\frac{W}{p}-\{Q_D+Q_R^S+Q_R^G\}\right]+\xi_G[Q_D+Q_R^S+Q_R^G]+\xi_S[Q_D+Q_R^S-g(Q_I)] \\
&+\psi\left[2C-\left\{\frac{W}{p}+|Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL})|+|Q_R^S|+|Q_R^G|+Q_I\right\}\right]+\psi_m[\overline{Q}-Q_I]
\end{aligned}$$

Giving the following FOC:

$$Q_D \quad : \quad \gamma(\mu-p)-\gamma^2\sigma^2(Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL}))-\lambda+\xi_S+\xi_G-\psi\,\mathrm{sgn}(Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL}))=0 \tag{18}$$

$$Q_R^S \quad : \quad \gamma(R^S-1)p-\lambda+\xi_G+\xi_S-\psi\,\mathrm{sgn}(Q_R^S)=0 \tag{19}$$

$$Q_R^G \quad : \quad \gamma(R-1)p-\lambda+\xi_G-\psi\,\mathrm{sgn}(Q_R^G)=0 \tag{20}$$

$$\begin{aligned}
Q_I \quad : \quad & \gamma(\mu-R^Sp)(\mathbb{1}_{CS}-\mathbb{1}_{CL})-\gamma^2\sigma^2(Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL}))\,\mathrm{sgn}(\mathbb{1}_{CS}-\mathbb{1}_{CL})+\gamma ap \\
& -\xi_S g'(Q_I)-\psi\,\mathrm{sgn}(Q_D+Q_I(\mathbb{1}_{CS}-\mathbb{1}_{CL}))(\mathbb{1}_{CS}-\mathbb{1}_{CL})-\psi-\psi_m=0
\end{aligned} \tag{21}$$

where $\mathrm{sgn}(\cdot)$ 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 $a-(R-R^S)>0$, 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,

$$\begin{aligned}
Q_D &= \frac{\mu-R^Sp}{\gamma\sigma^2}-Q_I \\
Q_R^S &= Q_I+g(Q_I)-\frac{\mu-R^Sp}{\gamma\sigma^2} \\
Q_R^G &= \frac{W}{p}+D-g(Q_I) \\
Q_I &= Q
\end{aligned}$$

Because we expect Q_I to be relatively large whenever the balance sheet constraint binds, it is natural to assume that $\text{sgn}(Q_R^G) = -1, \text{sgn}(Q_R^S) = 1$. Also, because $\mu > R^S p$, we expect the dealer's final cash position to be positive, that is, $\text{sgn}(Q_D + Q_I) = 1$. We denote $\overline{Q}^S$ as the maximum amount intermediated which is to be determined. In that case, because $\text{sgn}(Q_D + Q_I) = \text{sgn}(Q_R^S) =$, from FOC (18) and (19) we have,

$$Q_D^* = \frac{\mu - R^S p}{\gamma\sigma^2} - \overline{Q}^S.$$

Because $\text{sgn}(Q_R^G) = -1$ FOC (20) implies that $\lambda > 0$, making the budget bind. In addition, because $\text{sgn}(Q_R^G) = -\text{sgn}(Q_R^S) = -1$, FOC (19) and (20) imply that $\xi = \gamma(R - R^S)p + 2\phi > 0$, making the SI box constraint bind. The above observations imply,

$$\begin{aligned} Q_R^{S*} &= g(Q_I^*) - \frac{\mu - R^S p}{\gamma\sigma^2} + \overline{Q}^S \\ Q_R^{G*} &= \frac{W}{p} - g(Q_I^*) \end{aligned}$$

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

$$\frac{W}{p} + Q_D^* + Q_I^* + Q_R^{S*} - Q_R^{G*} + \overline{Q}^S = 2(Q_I^* + g(Q_I^*)) = 2C$$

pining down $\overline{Q}^S$, characterizing the optimal position.

Client Long

In the unrestricted model, dealers' optimal strategies are,

$$\begin{aligned} Q_D &= \frac{\mu - R^S p}{\gamma\sigma^2} + Q_I \\ Q_R^S &= g(Q_I) - \frac{\mu - R^S p}{\gamma\sigma^2} - Q_I \\ Q_R^G &= \frac{W}{p} - g(Q_I). \end{aligned}$$

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

Using FOC (19) and (20) we have $\xi_S = \gamma(R - R^S)p$ and $\lambda > 0$, therefore both the SI box constraint and the budget constraint bind.

$$Q_R^{S*} = g(Q_I^*) - Q_D^*$$
$$Q_R^{G*} = \frac{W}{p} - g(Q_I^*)$$

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

$$\frac{W}{p} + Q_D^* - Q_I^* - Q_R^{S*} - Q_R^{G*} + Q_I^* = 2Q_D^* = 2C$$

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

From FOC (21), and expressions for Q_D^* and ξ we have,

$$-\gamma(\mu - R^S p) + \gamma^2\sigma^2(C - Q_I) + \gamma a p - \gamma(R - R^S)g'(Q_I) = \psi_m.$$

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

$$\gamma a p - \gamma(R - R^S)g'(Q_I) = \gamma(\mu - R^S p) - \gamma^2\sigma^2(C - Q_I)$$

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:

We first calculate the partial derivatives of $w(b, Q)$,

$$w_b = -(\lambda'(b) + \gamma p)Qw$$

Equation (17) simply stems from taking the FOC of $\mathbb{E}(u(W^*)|CS)$ with respect to b. To ensure that this equation has a solution, consider the solution to limit of $\frac{\partial \mathbb{E}(u(W^*)|CS)}{\partial b}$ when $Q \longrightarrow \infty$. That is, the solution of the unrestricted balance sheet case, b^*_∞.[14] $\frac{\partial \mathbb{E}(u(W^*)|CS)}{\partial b}$ evaluated in $b = b^*_\infty$ is negative, because the term associated with $1 - w$ is

[14] Note that $\lim_{b\longrightarrow\infty}(b-(R-R^S))(\lambda(b)+\gamma p(b-(R-R^S)))w_b(b,Q) = 0$ where $w_b(b,Q) = -(\lambda'(b)+\gamma p)Qw(b,Q)$ is

zero and the remaining expression is strictly less than zero ($b^*_\infty > (R - R^S)$).
Because $\lambda''(b) > c > 0$, $\lambda'(b)b - \lambda'(b)(R - R^S) - \lambda(b)$ strictly increases to infinity, implying there exists a $b^* > b^*_\infty$ which solves equation (17).

■

Proof of Lemma 5

We first calculate the partial derivatives of $w(b, Q)$,

$$\begin{aligned}
w_b &= -(\lambda'(b) + \gamma p)Qw \\
w_Q &= -(\lambda(b) + \gamma p(b - (R - R^S)))w \\
w_{bQ} &= (\lambda'(b) + \gamma p)w + (\lambda'(b) + \gamma p)(\lambda(b) + \gamma p(b - (R - R^S)))Qw \\
w_{bb} &= -\lambda''(b)Qw + (\lambda'(b) + \gamma p)^2 Q^2 w
\end{aligned}$$

Denoting $H := \left(\lambda'(b)(b - (R - R^S)) - \lambda(b)\right)[1 - w] + (b - (R - R^S))(\lambda(b) + \gamma p(b - (R - R^S)))w_b$, from the implicit function theorem we have that $\frac{\partial b^*}{\partial Q} = -\frac{\partial H}{\partial Q} / \frac{\partial H}{\partial b}$. Therefore, we must sign H's sensitivity to both Q and b. In effect,

$$\begin{aligned}
\frac{\partial H}{\partial Q} &= -\left(\lambda'(b)(b - (R - R^S)) - \lambda(b)\right) w_Q(b, Q) + \\
&\quad (b - (R - R^S))(\lambda(b) + \gamma p(b - (R - R^S)))w_{Qb}(b, Q)
\end{aligned}$$

where $w_Q(b, Q)$ is the partial derivative of w with respect to Q, and w_{Qb} is the partial derivative of w_a with respect to Q. Using the equilibrium expression for b^*, i.e., equation (17), we can solve for $\lambda'(b^*)(b^* - (R - R^S)) - \lambda(b^*) = -\frac{w_b}{1-w}[(b^* - (R - R^S))(\lambda(b^*) + \gamma p(b^* - (R - R^S)))]$, we have the following equality,

$$\frac{\partial H}{\partial Q} = (b^* - (R - R^S))(\lambda(b^*) + \gamma p(b^* - (R - R^S))) \left[\frac{w_b(b^*, Q)w_Q(b^*, Q)}{1 - w(b^*, Q)} + w_{Qb}(b^*, Q)\right]$$

therefore the sign of $\frac{\partial U}{\partial Q}$ is equal to the sign of $\frac{w_b w_Q}{1-w} + w_{Qb}$.[15] That is, we have to sign $w_b w_Q + w_{Qb}(1 - w)$ which is equal to

$$(\lambda'(b^*) + \gamma p)w(b^*, Q)\left[w(b^*, Q) + (\lambda(b^*) + \gamma p(b^* - (R - R^S)))Q - 1\right].$$

the partial derivative of w with respect to b, because the exponential term converges faster to zero than a polynomial to infinity.

[15] Because $b^* > b^*_\infty > (R - R^S)$.

The above expression is positive since the bracket term takes the form $e^{-x} + x - 1$ which is positive whenever $x > 0$. [16]

Turning to H's sensitivity to b give,

$$\begin{aligned}\frac{\partial H}{\partial b} &= \lambda''(b)(b-(R-R^S))(1-w(b,Q)) + \\ &(\lambda(b)+\gamma p(b-(R-R^S)))[2w_b(b,Q)+(b-(R-R^S))w_{bb}(b,Q)]\end{aligned}$$

where w_{bb} is the partial derivative of w_b with respect to b. Note that $w_{bb} = -\lambda''(b)Qw + (\lambda'(b)+\gamma p)^2Q^2w$. Grouping terms accompanying λ'' gives

$$\lambda''(b)(b-(R-R^S))(1-w(b,Q)-(\lambda(b)+\gamma p(b-(R-R^S)))Qw(b,Q))$$

That is, the exponential term is $1 - e^{-x} - xe^{-x}$, which is positive for $x > 0$. Grouping the remaining terms gives,

$$(\lambda(b)+\gamma p(b-(R-R^S)))(\lambda'(b)+\gamma p)Qw(b,Q)[(b-(R-R^S))(\lambda'(b)+\gamma p)Q-2]$$

which holds for Q sufficiently large, by assumption.[17]

Therefore, $\frac{\partial H}{\partial Q}$ and $\frac{\partial H}{\partial b} > 0$ implying that $\frac{\partial b^*}{\partial Q} < 0$. Finally, note that $Q = \overline{Q}^S = \frac{C}{2}$, thus $\frac{\partial b^*}{\partial C} = \frac{\partial b^*}{\partial Q}\frac{1}{2}$, completing the proof.

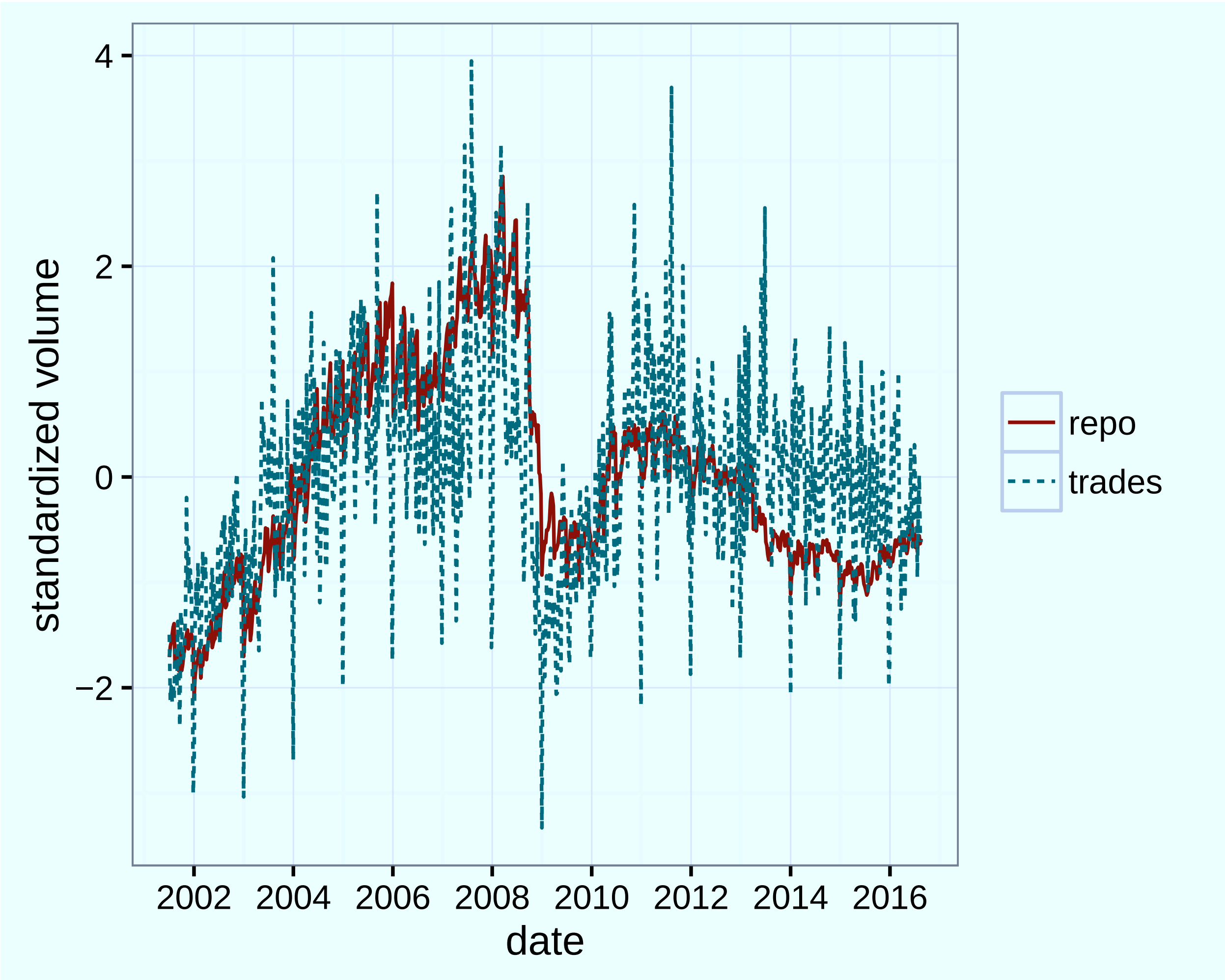
■

[16] The expression is equal to zero for $x = 0$ and is strictly increasing for $x > 0$.

[17] The intuition behind this condition is that if the restriction is small, by increasing the ask, more orders are concentrated on the short end, increasing profitability. Mathematically, it depends on when the

Figure 1: 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 repo and reverse repo.



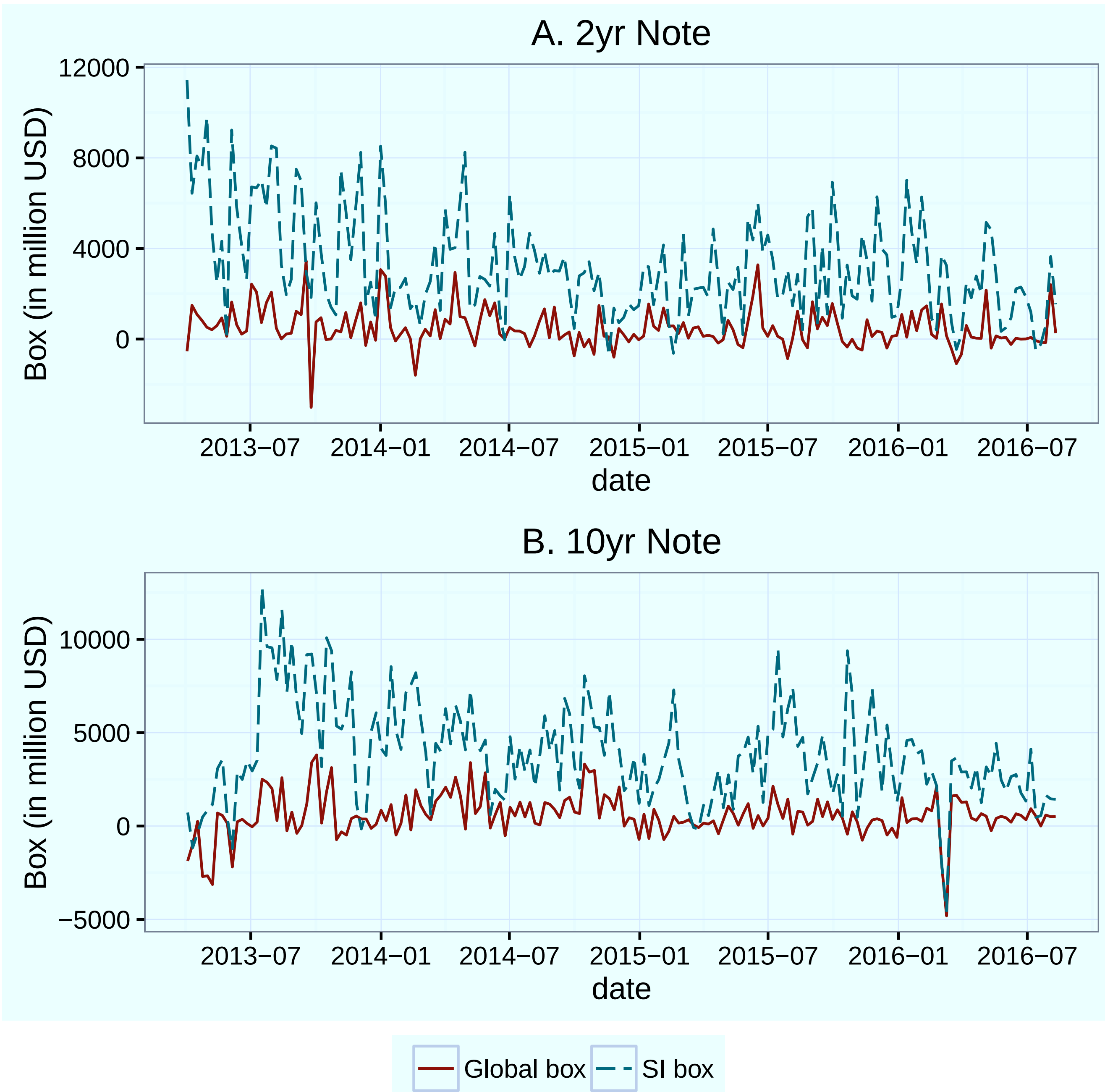


Figure 2: SI box and global box for primary dealers.
SI box and global box is calculated from FR 2004 data downloaded from NY Fed website. SI box is defined as net position plus the reverse repo on specific issue minus the repo on specific issue. Global box is defined as SI box plus reverse repo on general collateral minus the repo on general collateral. 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 SI box and global box for each tenor. SI box is calculated as net position plus SI reverse repo position minus the SI repo position for the on-the-run treasury. Global box is calculated as 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 is weekly, reported as of 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. Third and fourth column report the average SI box and global box, and the fifth column calculates what fraction of the time SI box is positive.

Tenor	N	avg SI box	avg global box	% pos SI box
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%