Inventory, Market Making, and Liquidity: Theory and Application to the Corporate Bond Market*

Assa Cohen† Mahyar Kargar‡ Benjamin Lester§ Pierre-Olivier Weill¶

August 30, 2022

Preliminary & Incomplete — Please Do Not Circulate

Abstract

We develop a search-theoretic model of over-the-counter markets in which customers with arbitrary preferences and asset holdings trade through dealers. Importantly, we assume that when a customer and a dealer meet, dealers can only sell assets that they already own. Within this environment, we derive the equilibrium relationship between dealers' cost of holding assets as inventory and various measures of liquidity, including dealers' inventory holdings (or "capital commitment"), bid-ask spreads, trade size, volume, and turnover. Using transaction-level data from the corporate bond market, we calibrate the model to quantitatively assess the impact of post-crisis regulations on dealers' inventory costs, liquidity, and welfare. We also exploit our structural framework to study the effects of other developments in the corporate bond market, including entry by non-regulated banks, the rise of electronic trading platforms, and the shift towards passive investment vehicles.

Keywords: Over-the-counter markets, Intermediation, Liquidity, Dealer inventory, Financial Regulation

JEL Classification: G11, G12, G21.

*The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. We thank Jack Bao, Briana Chang, Tim Johnson, Charlie Kahn, Lucie Lebeau, Dejanir Silva and seminar participants at the Federal Reserve Bank of Atlanta, the Federal Reserve Bank of Philadelphia, SED 2021, SFS Cavalcade 2022, Search and Matching in Macro and Finance Virtual Seminar Series, UCI, UCLA, UCSB, UIUC, the Virtual Finance Theory Seminar, Wisconsin, and Wharton Macro Lunch.

†University of Pennsylvania. Email: assac@sas.upenn.edu

‡University of Illinois at Urbana-Champaign. Email: kargar@illinois.edu

§Federal Reserve Bank of Philadelphia. Email: benjamin.lester@phil.frb.org

¶UCLA, NBER and CEPR. Email: poweill@econ.ucla.edu

1 Introduction

In many over-the-counter (OTC) markets, large dealer-banks provide liquidity through their willingness to hold *inventory*: they absorb assets onto their balance sheets when investors need to sell quickly, and they use these assets to fulfill investors' buy orders without delay. After the Great Financial Crisis (GFC) of 2007-2008, several regulations were introduced that increased the cost to dealers of holding inventory.[1] Not surprisingly, dealers responded by reducing their inventory holdings; for example, according to data from the Flow of Funds, the share of outstanding corporate bonds and non-agency mortgage-backed securities held by broker-dealers fell from approximately 3% in 2006 to less that 1% in 2018.[2] At the same time, both market participants and academics alike have argued that post-crisis regulations pose a threat to market liquidity.[3] Since maintaining liquid financial markets is crucial for a well-functioning economy, understanding and quantifying the effects of post-GFC regulations on market liquidity and welfare has emerged as a central challenge.

In this paper, we develop a structural model of dealer-intermediated OTC markets in order to meet this challenge. Our starting point is the benchmark search-theoretic framework developed by Duffie, Gârleanu, and Pedersen (2005) and extended to allow for arbitrary preferences and asset holdings by Lagos and Rocheteau (2009) and Gârleanu (2009). However, a key abstraction in these papers is that inventories do not play any economic role for marketmaking. Indeed, in these models, dealers never hold inventory—they merely enable customers to access a frictionless market, for which they charge a fee. Our innovation is to "put the inventory back in marketmaking" by introducing a simple and, arguably, quite natural constraint: we assume that a dealer can only

[1] These regulations include the 2010 Basel III framework, which introduced enhanced capital and liquidity requirements, along with the so-called "Volcker rule," which reduced implicit government guarantees (thus increasing banks' funding costs) and began monitoring banks' inventory holdings in concert with the regulation's ban on proprietary trading.

[2] Source: Table L.213 of the Federal Reserve's Flow of Funds. See Figure 4.

[3] See the extensive discussions by Thakor (2012), Duffie (2018), Bessembinder, Jacobsen, Maxwell, and Venkataraman (2018), and the many references therein.

sell to customers the assets that they currently hold in inventory.[4] As a result, a central feature of our model is that dealers choose an optimal amount of inventory in order to provide liquidity to their customers. We characterize the equilibrium, and study how dealers' inventory choice depends on the flow (dis)utility they receive from holding assets, which captures any regulatory costs imposed, and the subsequent effect on various measures of liquidity, including the bid-ask spread, the distribution of trade size, turnover, and volume.

Then, using transaction-level data from the corporate bond market, we calibrate the model to quantitatively assess the impact of the rising inventory costs associated with post-crisis regulation on liquidity, prices, allocations, and welfare. To start, we identify parameter values to match several target moments constructed from the corporate bond market data before the GFC. Given the parameter values implied by our calibration, we find that constraining dealers to hold assets before selling them had a relatively modest effect on equilibrium outcomes, relative to an environment where dealers did not face such a constraint. For example, relative to an environment without inventory constraints (but with search and bargaining frictions), we find that welfare decreases by approximately 0.5%. Intuitively, the implied cost of holding inventory before the GFC (and the reforms that followed) were relatively small, and hence dealers held sufficient inventory to fulfill most customer-buy orders in full.

Next, holding all other structural parameters fixed, we increase the dealers' cost of holding assets to levels consistent with the aggregate decline in dealers' inventory. We find that such a change has significant consequences for customers' equilibrium asset positions, trading costs, prices, and welfare. More specifically, we find that trading costs rise by approximately 50% relative to the pre-GFC benchmark, while the welfare costs of frictions increase by a factor of two.

Lastly, we use our model to study the effects of *other* important changes that have taken place in the corporate bond market since the GFC. These changes include the reduction in trading frictions

[4] Put differently, one could interpret the benchmark models of Duffie et al. (2005) and Lagos and Rocheteau (2009), and many others, as environments in which dealers can short sell (or borrow) assets instantaneously in a frictionless market. Under this interpretation, we relax this assumption by assuming instead that dealers cannot short sell the asset.

associated with the development of better electronic trading platforms, along with the general shift towards passive investment vehicles, which has arguably decreased the frequency with which the average investor re-balances her portfolio.

1.1 Related literature

The theoretical portion of our paper draws heavily from the literature that uses search-theoretic models of trade to study OTC markets. Many of these papers build off of the basic framework developed in Duffie et al. (2005), including the important contributions by Lagos and Rocheteau (2009) and Gârleanu (2009), who extend the basic framework to accommodate arbitrary preferences and asset holdings. Importantly, in most papers within this literature, dealers are assumed to have unfettered access to a frictionless, inter-dealer market, which obviates the need for any dealer to hold inventory.[5] Of course, while this assumption certainly makes the models more tractable—highlighting the important role of search, bargaining, and information frictions in the determination of prices and allocations—it also makes them ill-suited to study how recent developments have changed dealers' inventory holdings, and the consequences for trading activity, liquidity, and welfare.

In the literature on search-based OTC markets, several papers have proposed models of dealers' inventory management. In Weill (2007) and Lagos, Rocheteau, and Weill (2008), for example, dealers find it optimal to hold inventories in anticipation of aggregate fluctuations in customers' demand. However, in both environments, optimal inventory holdings are always zero in the "long run" non-stochastic steady state. In our model, inventories play a non-trivial economic role even in the non-stochastic steady state, which we believe is an important feature for studying the long-run decline in inventories between 2008 and 2018. Our work is also related to An (2018), who shows that, despite the presence of holding costs, imperfectly competitive dealers have incentive to hold

[5] Such papers include, but are not limited to, Feldhütter (2012), Lester, Rocheteau, and Weill (2015), Milbradt (2017), Pagnotta and Philippon (2018), and Lagos and Zhang (2020). See Weill (2020) for a thorough review of the literature.

inventories in order to gain market power with their customers. Tse and Xu (2018) also develop a model where dealers (with different trading capacity) carry inventory in order to rationalize empirical observations about inter-dealer trades OTC markets.

There are also a number of papers in which in all agents, including those who play the role of dealers, trade in decentralized markets.[6] In these models, since agents face a short-selling constraint, dealers must hold inventories. Though these models have proven very useful in studying the determinants of market structure and inter-dealer trading patterns, we chose instead to make the inter-dealer market centralized. This simplification allows us to focus our analysis more squarely on the issue at hand; to derive new, testable implications regarding, e.g, the relationship between dealers' inventory costs and the distribution of trade size; and to make the model more complex in other dimensions of economic interest, such as heterogeneity in dealers inventory cost or in customers' demand, to confront the ever-expanding set of facts emerging from dealer-intermediated OTC markets.

Outside of search-based models, there is also, of course, a celebrated literature on inventory management by dealers, starting with Amihud and Mendelson (1980), Ho and Stoll (1981, 1983), and Mildenstein and Schleef (1983). Relative to this literature, our main contribution is to consider a model in which customers' supply and demand are derived from explicit, dynamic optimization problems. This enables us to quantify the gains from trade created by the inter-dealer market, analyze how customers' trading needs respond to other changes in the economic environment, and offer a welfare analysis of several important changes that have occurred in the corporate bond market since the GFC, including the effects of new regulations.

Because of its quantitative focus, our work is related to papers who structurally estimate models of OTC market, either search-based as in Feldhütter (2012) , Gavazza (2016) , Brancaccio, Li, and Schurhoff (2017), Hendershott, Li, Livdan, and Schürhoff (2020) and Liu (2020), or network-based

[6]See, e.g., Hugonnier, Lester, and Weill (2014, 2019), Shen, Wei, and Yan (2015), Üslü (2019), Farboodi, Jarosch, and Shimer (2018a), Farboodi, Jarosch, Menzio, and Wiriadinata (2018b), Bethune, Sultanum, and Trachter (2018), Yang and Zeng (2019), and Nosal, Wong, and Wright (2019).

as in Gofman (2014), Gofman (2017) and Eisfeldt, Herskovic, Rajan, and Siriwardane (2018). We contribute to this literature by developing a new model and focusing on a different market phenomena.

Given the focus of our application, our paper is related to several recent empirical studies that have attempted to identify the effect of post-crisis regulations on market liquidity, including Trebbi and Xiao (2017), Bao, O'Hara, and Zhou (2018), Bessembinder et al. (2018), Dick-Nielsen and Rossi (2019), and Choi and Huh (2018). By studying this issue within the context of a structural equilibrium model, our analysis complements these existing empirical exercises in several important ways. First, by calibrating our model to match moments before and after the introduction of new regulations, we are able to infer the implicit cost of these regulations on dealers; this cost is difficult to measure directly and, to the best of our knowledge, such an estimate is new to the literature. Second, while existing empirical studies based on difference-in-difference regressions identify "local" effects of new regulations on a particular measure of liquidity, such as price impact, our model allows us to explore the broader implications of policy for the behavior of customers and dealers, and the subsequent implications for a variety of outcomes, both observable (such as bid-ask spreads, trade size, or volume) and unobservable (such as the time customers wait to trade). Third, and perhaps most important, our structural equilibrium model provides natural measures of welfare, along with the opportunity to perform counterfactuals, which is crucial for evaluating the quantitative impact of policy. Lastly, since our model can accommodate alternative explanations for changes in market outcomes, we are able to study the potential effects of other recent developments in the corporate bond market, including the rise of electronic trading and the shift towards passive investment vehicles.

2 Benchmark Model

We consider a continuous time, infinite horizon model of an over-the-counter asset market in the spirit of Lagos and Rocheteau (2009). There are two types of infinitely-lived agents: a measure μ_c of customers and a measure μ_d of dealers. There is one asset that is durable, perfectly divisible, and in fixed supply, $s \in \mathbb{R}_+$.

We assume that customers have stochastically varying preferences defined over the quantity of asset they hold, and a numéraire consumption good. In particular, let $u(\delta, q) + c$ denote a customer's flow utility, where $q \in \mathbb{R}_+$ denotes the units of asset the customer holds, $\delta \in [\underline{\delta}, \bar{\delta}]$ denotes her current preferences for assets, and $c \in \mathbb{R}$ denotes her net consumption of the numéraire good. We assume that $u(\delta, q)$ is increasing and strictly concave in $q \in (0, \infty)$, and satisfies the Inada conditions $\lim_{q \to 0} u_q(\delta, q) = \infty$ and $\lim_{q \to \infty} u_q(\delta, q) = 0$. We also assume that $u_q(\delta, q)$ is increasing in δ, where u_q denotes the partial derivative with respect to q. Preference shocks arrive at rate γ, at which time a new δ' is drawn according to $F(\delta')$.[7] For simplicity, we assume that dealers have linear preferences that do not change over time: a dealer receives flow utility $v_d\, q + c$ from holding q units of the asset in inventory and consuming c (net) units of the numéraire good, where $v_d \in \mathbb{R}$. All agents discount the future at rate $r > 0$.

Dealers have continuous access to a frictionless, competitive market where they can buy or sell any amount of the asset at price p. Customers, do not meet each other and trade directly: instead, they meet dealers, only periodically, at which time they can rebalance their portfolio. In particular, let $M(\mu_c, \mu_d)$ denote the function that determines the (flow) number of matches formed between customers and dealers. We assume that M is increasing in both arguments, concave, and homogeneous of degree one. Normalizing $\mu_c = 1$, we denote by $\lambda = M(\mu_c, \mu_d)/\mu_c = M(1, \mu_d) \equiv m(\mu_d)$ the rate at which each customer meets a dealer.

[7] Micro-foundations for such a specification have been provided earlier in the literature. For example, under appropriate specification, $u(\delta, q)$ represents the flow certainty equivalent of holding q units of the asset. See Weill (2020) for a survey.

The key departure from the existing literature is an "asset-in-advance" constraint: when a dealer meets a customer, we assume that he can buy any quantity of assets from the customer, but he can only sell assets that he currently holds. Finally, we assume that the terms of trade between customers and dealers are determined by Nash bargaining, where we denote by θ the bargaining power of the dealer.

2.1 Customers

Let $V(\delta, q)$ denote the maximum expected discounted utility attainable by a customer with current asset holdings q and preferences δ. A standard dynamic programming arguments shows that $V(\delta, q)$ solves the Hamilton-Jacobi-Bellman (HJB) equation

$$rV(\delta, q) = u(\delta, q) + \gamma \int \left[V(\delta', q) - V(\delta, q)\right] \, dF(\delta') + \lambda \left[V(q + x, \delta) - V(\delta, q) - px - \tau\right], \quad (1)$$

where, abusing notation, x and τ denote the transfer of asset and numeraire, respectively, between a customer in state (δ, q) and a dealer, which we derive explicitly below. The interpretation of this HJB equation is straightforward: the customer enjoys flow utility $u(\delta, q)$ until one of two events occurs. First, at rate γ, a preference shock arrives, at which time a new δ' is drawn from $F(\delta')$. Second, at rate $\lambda = m(\mu_d)$, the customer has the opportunity to trade with a dealer. At this time, the dealer buys $x > 0$ (or sells $-x > 0$) units of the asset on behalf of the customer at the inter-dealer price p in exchange for a transfer $\tau > 0$ of numéraire good. Technically, we require that $x \geq -q$, but imposing Inada conditions on $u(\delta, q)$ ensures that customers never sell all of their asset holdings.

The terms of trade (x, τ) when a customer in state (δ, q) meets a dealer holding $n \in \mathbb{R}_+$ units of the asset in inventory are defined as the solution to the Nash bargaining problem

$$
\begin{aligned}
&\arg\max_{\tilde{x},\tilde{\tau}} \left[V(\delta, q+\tilde{x} - V(\delta, q) - p\,\tilde{x} - \tilde{\tau}\right]^{1-\theta}\,\tilde{\tau}^{\theta} \\
&\qquad\qquad \text{s.t. } \tilde{x} \leq n.
\end{aligned}
$$

As is standard, the solution to this bargaining problem is

$$
x^{\star}(\delta, q \mid p, n) = \arg\max_{\tilde{x}\leq n} \left[V(\delta, q+\tilde{x}) - V(\delta, q) - p\,\tilde{x}\right] \tag{2}
$$

$$
\tau^{\star}(\delta, q \mid p, n) = \theta\left[V(\delta, q + x(\delta, q \mid n, p)) - V(\delta, q) - p\,x(\delta, q \mid n, p)\right]. \tag{3}
$$

That is, the solution selects the value of $x \leq n$ that maximizes the bilateral surplus $V(q+x,\delta) - V(\delta, q) - p\,x$, and the transfer τ delivers a fraction θ of this surplus to the dealer.

In what follows, we will adopt the usual convention of using the lower case n to denote an individual dealer's inventory, and using the upper case N to denote the choice of all dealers. Therefore, in a symmetric, steady-state equilibrium in which $n = N$, substituting (2) and (3) into the HJB equation and applying informally the Envelope Theorem yields

$$
\begin{aligned}
rV_q(\delta, q) = u_q(\delta, q) + \gamma \int \left[V_q(\delta, q') - V_q(\delta, q)\right] dF(\delta') \\
+ \lambda(1-\theta)\left[\max\left\{p, V_q(q+N, \delta)\right\} - V_q(\delta, q)\right].
\end{aligned}
$$

Let $\Sigma(\delta, q) \equiv V_q(\delta, q) - p$ denote the marginal surplus of a customer-to-dealer trade, i.e., the marginal value to a customer of an additional unit of asset, net of the inter-dealer price. We can rewrite the expression above as

$$
\begin{aligned}
\left[r + \gamma + \lambda(1-\theta)\right] \Sigma(\delta, q) = u_q(\delta, q) - rp + \gamma \int \Sigma(\delta, q')\, dF(\delta') \\
+ \lambda(1-\theta) \max\left\{\Sigma(q+N, \delta), 0\right\}.
\end{aligned} \tag{4}
$$

Equation (4) characterizes $\Sigma(\delta, q)$ given the pair (p, N), so at times it will be helpful to make this dependence explicit by writing $\Sigma(\delta, q \mid p, N)$.[8]

Proposition 1 *Equation (4) admits a unique continuous solution $\Sigma(\cdot)$, with the following properties*

- *It is strictly increasing in δ, strictly decreasing in q, p, and weakly decreasing in N;*
- *Given any $p > 0$, it is strictly positive for small q, and strictly negative for large q;*
- *It is the basis of a solution to the HJB equation (1).*

The function $\Sigma(\cdot)$ entirely characterizes a customer's optimal trading behavior. Indeed, let $q^\star(\delta \,|\, p, N)$ denote the solution of $\Sigma(\delta, q \mid p, N) = 0$, interpreted as the "target" portfolio of a customer with current preference shock δ.[9] Then, in an equilibrium in which (almost all) dealers hold inventory N and the inter-dealer price is p, when a customer in state (δ, q) meets an individual dealer who has chosen to hold some arbitrary inventory n, the optimal trade and transfers are:

$$x^\star(\delta, q \mid p, N, n) = \min\{q^\star(\delta \mid p, N) - q, n\}$$
$$\tau^\star(\delta, q \mid p, N, n) = \theta \int_q^{q+\min\{q^\star(\delta|p,N), q+n\}} \Sigma(\delta, x \,|\, p, N, 0)\, dx.$$

Notice that the optimal trading behavior is asymmetric: a customer who currently hold $q < q^\star(\delta)$ may need to make several purchases in order to reach its target. A customer who holds $q > q^\star(\delta)$, on the other hand, will be able to reach its target in one sale. Taken together, this observation implies that the size of a typical trade will be smaller, on average, for customer purchases than for sales.[10] Such an asymmetry in trade size is a well known empirical observation in major OTC

[8]Notice that the environment of Lagos and Rocheteau (2009), where there is no asset-in-advance constraint, corresponds to the case where $N \to \infty$ and the final term in (4) disappears.

[9]That the properties established in Proposition 1 ensures that this equation has indeed a unique solution.

[10]Since, on the aggregate, the total quantity of assets purchased and sold are equal, this means that our model predicts that the number of customer purchases will be larger than that of sales.

markets (see for example Green, Hollifield, and Schürhoff, 2007). It is a unique implication of our inventory-in-advance constraint: indeed, one can show that he asymmetry would disappear in an otherwise identical model without this constraint.

The properties of $\Sigma(\cdot)$ reported in Proposition 1 reveal some intuitive partial equilibrium relationships between an individual customer's current state, the aggregate state, and the customer's target asset holdings. In particular, as one would expect, the customer's target asset position, $q^\star$, is increasing in his idiosyncratic valuation δ, decreasing in his current asset holdings q, and decreasing in the price p. In addition, the presence of inventory constraints induces customers to acquire additional assets out of precautionary motives: since an additional unit of the asset is more valuable when dealers hold less inventory, ceteris paribus, $q^\star$ will be larger when N is closer to zero.

2.2 Dealers

Let $\Psi(\delta, q \mid p, N)$ denote the joint distribution of asset holdings and preference types across customers (which we characterize below). Using the Nash bargaining solution, along with the definition of $x^\star(\delta, q)$, we can write the dealer's (flow) profit function as

$$r\Pi(n) = (v_d - rp)n + \frac{\lambda}{\mu_d}\theta \int \max_{x \leq n} \left\{ V(\delta', q' + x) - V(\delta', q') - px \right\} \, d\Psi(\delta', q').$$

Hence, the dealer's objective function has two components: the flow payoff from owning n units of the asset, $(v_d - rp)n$; and the expected capital gains from trading with a randomly selected customer.

Lemma 1 *For any $\Psi(\delta, q \mid p, N)$, p, and N, the profit function $\Pi(n)$ is concave and continuously differentiable in n, with derivative:*

$$\frac{d\Pi}{dn}(n) = v_d - rp + \frac{\lambda}{\mu_d}\theta \int \max \left\{ \Sigma(\delta', q' + n'), 0 \right\} \, d\Psi(q', \delta').$$

The expression for the derivative of the profit function is intuitive. The first term is the direct flow utility that a dealer enjoys by holding a marginal unit of the asset. The second term is the user cost: what the dealer has to pay per unit of time to hold a marginal unit of the asset. The third term is the marginal impact of increasing inventory on intermediation profits. Indeed, the dealer meets customers with intensity λ/μ_d and appropriate a fraction θ of the marginal trading surplus created by increasing inventories, which is equal to $\max\{\Sigma(\delta', q' + n), 0\}$ if a dealer meets a customer of type (δ', q'). Notice in particular that this marginal surplus is strictly positive if and only if $q' + n < q^\star(\delta)$, that is, if and only if it relaxes a binding inventory-in-advance constraint and helps the customer to trade closer to its target.

The first-order condition for an optimal inventory is simply

$$\Pi'(n) \leq 0 \text{ with equality if } n > 0. \tag{5}$$

Note that a solution to (5) requires $rp \geq v_d$; if $v_d > rp$, then dealers would have incentive to acquire infinite inventory. Hence, in equilibrium, the price will always adjust to incorporate the dealers' flow value from holding the asset *and* the marginal benefit of increasing inventory on intermediation profts.

Moreover, as in our analysis of the customer's optimal asset position, the properties of $\Sigma(\cdot)$ allow for some natural, partial equilibrium comparative statics with respect to an individual dealer's optimal inventory holdings. In particular, given the behavior of all other agents (and, hence, aggregate variables), one can easily show that an individual dealer's optimal n is increasing in the rate at which he meets customers, λ/μ_d, and the fraction of the trading surplus he extracts through bargaining, θ.

2.3 The steady-state distribution and market clearing

We now characterize the joint distribution $\Psi(\delta, q)$. To start, we fix an upper bound $\bar{q}$ for asset holdings such that $\bar{q}$ is strictly larger than the highest possible target $q^\star(\bar{\delta} \,|\, p, N)$. We then construct the transition probability function, which dictates the probability that a customer in state (δ, q) transitions to a new state that lies in some Borel set B of $[\underline{\delta}, \bar{\delta}] \times [0, \bar{q}]$, conditional on receiving a shock or a trading opportunity:

$$\mathbb{P}\left((\delta, q), B\right) = \frac{\gamma}{\lambda + \gamma} \int \mathbb{I}_{(\delta', q) \in B} \, dF(\delta') + \frac{\lambda}{\lambda + \gamma} \mathbb{I}_{(\min\{q^\star(\delta), q+N\}, \delta) \in B}, \tag{6}$$

where we've suppressed the dependence of $q^\star(\delta)$ on p and N for notational convenience. One can show that the transition probability function is monotone, has the Feller property, and satisfies appropriate mixing conditions, which allows us to apply results in Chapter 12 of Stokey and Lucas (1989) to establish the next result.

Lemma 2 *There exists a unique steady-state distribution $\Psi(\delta, q)$. This distribution is weakly continuous and decreasing, in the sense of first-order stochastic dominance, in the inter-dealer price p.*

Finally, given the steady-state distribution, market clearing requires

$$\int q \, d\Psi(\delta, q \mid p, N) + \mu_d N = s. \tag{7}$$

Equation (7) simply equates the total measure of assets held by customers and dealers to the total supply.

2.4 Equilibrium

Characterizing a symmetric steady-state equilibrium requires solving a fixed point problem over the function $\Sigma^\star(\delta, q)$, the optimal inventory holdings of each dealer $N^\star$, the joint distribution of asset holdings and preferences $\Psi^\star(\delta, q)$, and the inter-dealer price $p^\star$ such that: (i) $\Sigma^\star(\delta, q)$ solves (4) given $N^\star$ and $p^\star$; (ii) $n = N^\star$ solves the dealer's optimality condition (5) given $\Psi^\star$ and $p^\star$; (iii) $\Psi^\star$ is the invariant distribution defined by the transition function in (6); and (iv) $\Psi^\star$ satisfies the market clearing condition (7).

Proposition 2 *There exists a $\underline{u}_d$ such that, for any $v_d \geq \underline{u}_d$, an equilibrium exists in which dealers actively intermediate.*

3 Quantitative Exercise

In this section, we use our model to quantitatively evaluate the effects of recent changes in the corporate bond market. We pay particular attention to understanding how changes in dealers balance sheet costs v_d, brought on by changes in post-crisis regulation, affected market liquidity, prices, and allocations. As a first step, we calibrate our model to match moments from the corporate bond market before the GFC. Then, we infer the change in v_d that is consistent with the observed decline in dealers' inventory holdings. Using this implicit change in dealers' inventory costs, we can evaluate the effects of post-crisis regulation on liquidity, prices, allocations, and welfare.

3.1 Data

We use the academic version of the Trade Reporting and Compliance Engine (TRACE) database of US corporate bond transactions, made available by the Finance Industry Regulatory Authority (FINRA). The raw TRACE data provides detailed information on all secondary market transactions self-reported by FINRA member dealers. These include bond's CUSIP, trade execution time and date, transaction price ($100 = par), the volume traded (in dollars of par), a buy/sell indicator, and flags for dealer-to-customer and inter-dealer trades. Unlike the public version, the academic TRACE does not censor trade volume at $5 million (for investment grade bonds) or $1 million (for high-yield bonds). The academic version also contains masked dealer identities as well as transactions in privately traded Rule 144A bonds that are not disseminated to the public.

Dealers are required to correct errors in previously reported trades with flags corresponding to trade cancellations, modifications, or reversals. We use the standard cleansing algorithm described in Dick-Nielsen (2009, 2014) and Dick-Nielsen and Poulsen (2019) to remove these self-reported errors. Our TRACE sample starts in July 2002 and covers transactions until the end of 2016.

We collect issue credit ratings and bond characteristics from Mergent Fixed Income Securities Database (FISD). We drop all bonds not contained in the FISD and only consider CUSIPs in TRACE identified by FISD as fixed-coupon US corporate debentures and US corporate bank notes with non-missing maturity dates and amounts outstanding. We also exclude bonds with equity-like and special features.[11] Finally, we exclude trades associated with new issuances and also remove transactions that happen within 90 days of the traded bond issuance.

Table 1 reports summary statistics for daily number and volume of interdealer, customer-bought, and customer-sold trades. It is interesting to highlight that while the total volume of customer-bought and -sold trades are very similar (approximately \$3.9 billion), consistent with our model, we observe, on average, more customer-bought trades than customer-sold trades.

3.2 Calibration to Pre-Crisis Corporate Bond Market

We set the discount factor, r, equal to 5% and assume that customers have an iso-elastic utility function of the form

$$u(q,\delta) = \delta \frac{q^{1-1/\eta}}{1-1/\eta}.$$

In addition, we assume that the preference shock, δ, is drawn from a quadrature-based discrete approximation of a log-normal distribution (see Tauchen and Hussey, 1991). We normalize the

[11] We drop all bonds that are convertible, puttable, exchangeable, preferred, asset-backed, secured lease obligations, unit deals, and Yankee. We also exclude bonds with variable coupons or sinking funds and are issued in a foreign currency or part of unit deals.

Variable	Mean	Std.dev	Q05	Q50	Q95
daily num. interdealer	6,512.16	3,851.31	71.95	5,720	12,118.80
daily num. customer	10,695.05	4,464.67	1,373.50	10,184.50	17,088.70
daily num. customer-bought	6,373.25	2,783.31	929.85	6,317	10,754.40
daily num. customer-sold	4,321.80	2,016.85	416.10	4,027	7,319.70
daily vol. interdealer ($m)	2,424.17	1,133.61	31.29	2,462.18	4,028.52
daily vol. customer ($m)	7,783.17	3,474.20	227.52	7,704.17	1,3319.89
daily vol. customer-bought ($m)	3,886.72	1,766.10	128.64	3,836.94	6,571.82
daily vol. customer-sold ($m)	3,896.44	1,796.46	94.34	3,852.99	6,839.93

Table 1. This table provides mean, standard deviation, median, 5th and 95th percentiles of the average daily number of trades and volume by counterparty type, all years. The "num" variables refer to number of trades and the "daily-vol" variables refer to the average total daily volume, in millions USD. "customer" trades refer to trades between a dealer and a customer which represent the sum of "customer-bought" and "customer-sold" trades. The sample is from the academic version of TRACE and runs from July 2002 to the end of December 2016. All Rule 144A bonds for which trades not disseminated to the public are excluded. We filter the sample as described in the main text.

mean of $F(\delta)$ so that the frictionless, Walrasian price of the asset is $1/r$. Given the iso-elastic utility function, one can show that the model is homogeneous of degree 1 in s, the per-capita supply of the asset.[12] Hence, we can normalize s. Finally, we assume the matching function $M(\mu_c, \mu_d) = \lambda\mu_c$, so that customers meet dealers at a constant rate λ. Following Hugonnier et al. (2019), we set λ so that a customer contacts a dealer every five days.[13]

Given the assumptions above, there are six parameters that we need to calibrate: the measure of active dealers, μ_d; the elasticity parameter for the customers' utility function, η; the variance of the preference shocks, σ_δ^2; the arrival rate of preference shocks, γ; the dealers' bargaining power, θ; and the dealers' utility parameter, v_d. In what follows, we describe the six target moments that we use to determine these parameters. Though many of the target moments and parameters are inter-related, we try to connect each moment to the parameter it affects most directly. All targets are matched exactly.

[12]In particular, for any $\tilde{s} = \kappa s$, scaling $\tilde{s} = \kappa^{1/\sigma}\delta$ renders the marginal utilities—and hence prices and allocations—unchanged.

[13]Since the time a customer spends searching is typically not observable, there is no consensus in the literature on an appropriate choice for λ. Our target of five days is consistent with Feldhütter (2012), and lies near the mid-point of existing studies, ranging from 1-2 days (Duffie, Gârleanu, and Pedersen, 2007; Pagnotta and Philippon, 2018) to as many as ten business days (He and Milbradt, 2014).

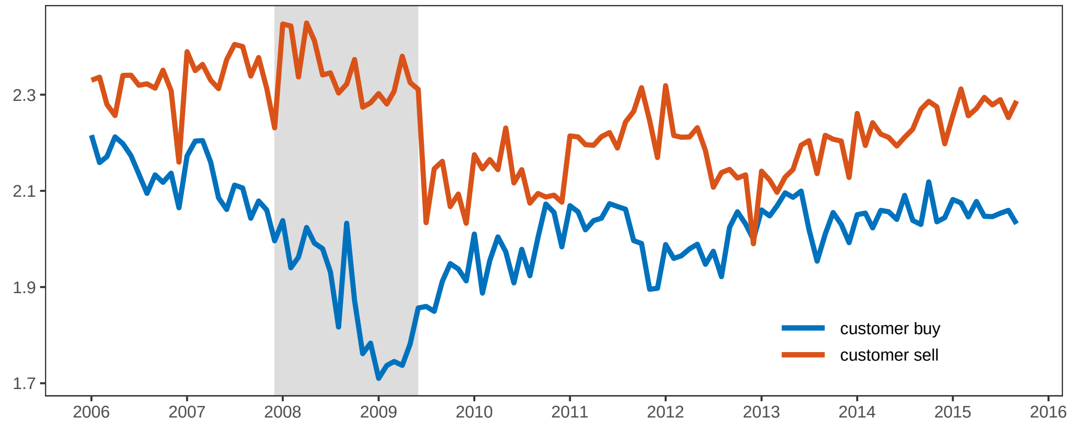

Figure 1. Monthly standard deviation of log trade size for customer-bought and customer-sold trades. Source: TRACE. The vertical shaded bars indicate NBER recessions.

1. The dispersion in preference shocks determines customers' equilibrium asset holdings, and the size of trades they execute when their asset holdings differ from their target portfolios. Hence, we target the standard deviation of log trade size, which is 2.1, to help identify σ_δ^2. Figure 1 plots monthly standard deviation of log trade size for customer-bought and -sold trades from TRACE.

2. Dealers' bargaining power determines the trading costs paid by customers. Hence, we target the value-weighted roundtrip trading cost proposed by Choi and Huh (2018) from before the GFC, which is 15 bps, to help identify θ. Figure 2 plots roundtrip transactions costs from Choi and Huh (2018) for investment grade (IG) bonds.[14]

3. The frequency of preference shocks is a primary determinant of how often customers want to buy or sell. Hence, we target the annual turnover of assets that customers buy, which is

[14] Choi and Huh (2018) construct their main measure of bid-ask spread as $2Q \times \dfrac{\text{traded price} - \text{reference price}}{\text{reference price}}$ where Q is equal to $+1(-1)$ for a customer buy (sell) trade. For each customer trade, a "reference price" is calculated as the volume-weighted average price of interdealer trades larger than \$100,000 in the same bond-day, excluding interdealer trades executed within 15 minutes. The measure is calculated at the trade level for all customer trades and is also calculated at the bond-day level by taking the volume-weighted average of trade level spreads.

Figure 2. Monthly average trading costs proposed by Choi and Huh (2018) for investment grade bonds. Source: TRACE. The vertical shaded bars indicate NBER recessions.

approximately 0.2, to help determine γ.[15] Figure 3 plots the quarterly turnover for customer-bought and customer-sold trades form TRACE.

4 & 5. The measure of dealers and the flow utility they receive from holding assets determine how much inventory they hold in equilibrium. Moreover, their choice of inventory determines how many purchases customers need to make to reach their target portfolio. Hence, we propose two targets that help determine μ_d and v_d: the total asset holdings of the dealer sector before the GFC, as a share of outstanding assets; and the ratio of the number of customer-sell transactions to the number of customer-buy transactions. The value of the former is 0.03, while the value of the latter is 0.85. Figure 4 plots the share of corporate and foreign bond holding for security broker-dealers from the Flow of Funds. In Figure 5, we plot the ratio of the number of customer-sold to customer-bought trades for trade sizes less and greater

[15]The turnover of customer-bought trades is obtained by dividing total volume from dealers-to-customer trades by the average amount outstanding.

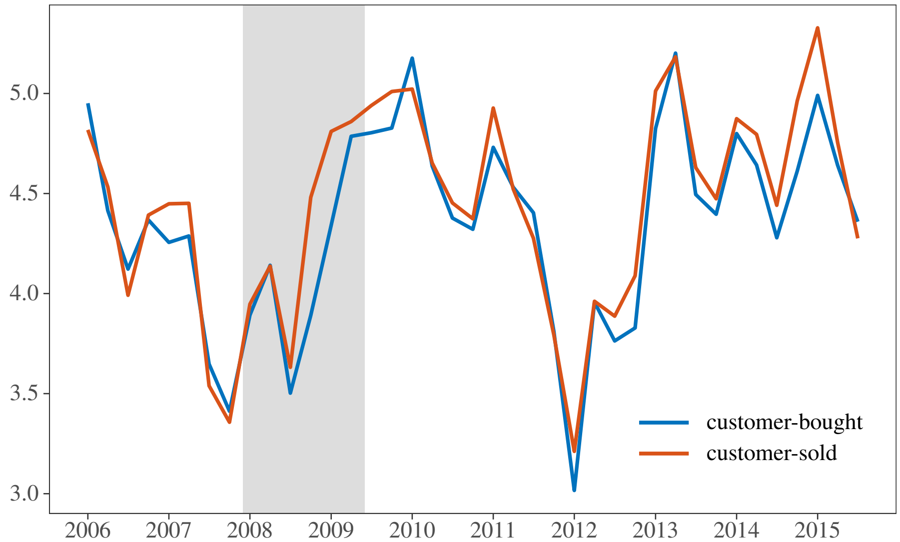

Figure 3. Quarterly turnover for customer-bought and customer-sold trades in percentage points. Source: TRACE. The vertical shaded bars indicate NBER recessions.

than $\$100,000$. Consistent withe the evidence from Table 1, we see there is an asymmetry between the number of customer-bought and -sold trades.

6. The elasticity of the customers' utility function is a key determinant of the response of asset prices to supply shocks. Hence, we use the estimate of price elasticity derived by Dick-Nielsen and Rossi (2019)—who find that a 0.9% increase in supply generates an 18 bps change in price—to determine η.

Table 2 reports the outcome of our calibration exercise. Since the Walrasian price of the asset would be $1/r$ and dealers are risk neutral, it is natural to interpret the flow utility that dealers receive from holding a unit of this bond as $v_d = 1 - \tau$, where τ denotes the (flow) inventory cost to dealers of holding the asset on their balance sheet. Hence, if we think of the asset as a consol bond, our calibration implies that inventory costs before the GFC were approximately 4.7% of the bond's dividend.

Figure 6 plots (log of) customers' target asset holdings, $q^{\star}(\delta)$, along with each dealers' equilibrium inventory holdings, N. To highlight the effects of the inventory constraint in our framework,

Table 2. Values of calibrated parameters.
This table reports parameter values used in calibrating the model with associated empirical targets in the academic TRACE data.

	Parameter	Value	Target (target value)
σ_δ^2	Dispersion of preference shock distribution	0.280	Std. dev. of log trade size (2.1)
θ	Dealers' bargaining power	0.587	Volume-weighted roundtrip trading cost from Choi and Huh (2018) (15 bps)
γ	Preference shock intensity	0.229	Annualized turnover for customer-bought trades (20%)
v_d	Flow utility of dealers	0.953	Dealer sector's pre-GFC share of corporate bonds outstanding (3%)
μ_d	Measure of dealers	0.003	Ratio of No. customer-sold to No. customer-bought (0.85)
η	Elasticity of customers' utility flow function	5	Price elasticity of supply in Dick-Nielsen and Rossi (2019)

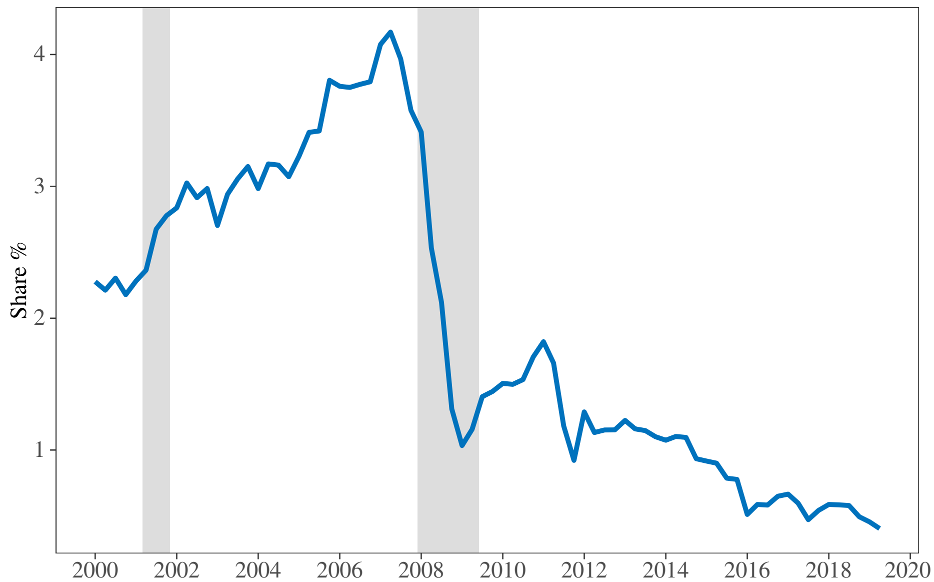

Figure 4. Share of outstanding corporate and foreign bonds held by security broker-dealers. Source: Table L.213 of the Federal Reserve's Flow of Funds. The vertical shaded bars indicate NBER recessions.

we also plot the target asset holdings in an environment without the inventory constraint (i.e., the environment of Lagos and Rocheteau, 2009). Note that the inventory constraint only binds for those customers who receive the largest preference shock. Moreover, note that the prospect of being constrained in the future introduces a precautionary motive, as investors who currently receive low utility from holding the asset target larger positions than they would in the no-constraint environment.

Introducing an inventory constraint increases the bid-ask spread charged by dealers: given the parameter values that emerge from our calibration, the trading costs in the no-constraint environment would be approximately 2 bps smaller. As trading costs rise, the customers' valuation for the asset declines, which puts downward pressure on the inter-dealer price. However, the presence of an inventory constraint also puts upward pressure on the price, for at least two reasons: first, precautionary incentives increase customers' demand for the asset; and second, since dealers are holding a fraction of the inventory in our environment, this mechanically reduces the supply of assets held by customers. In equilibrium, we find that the forces putting upward pressure on the

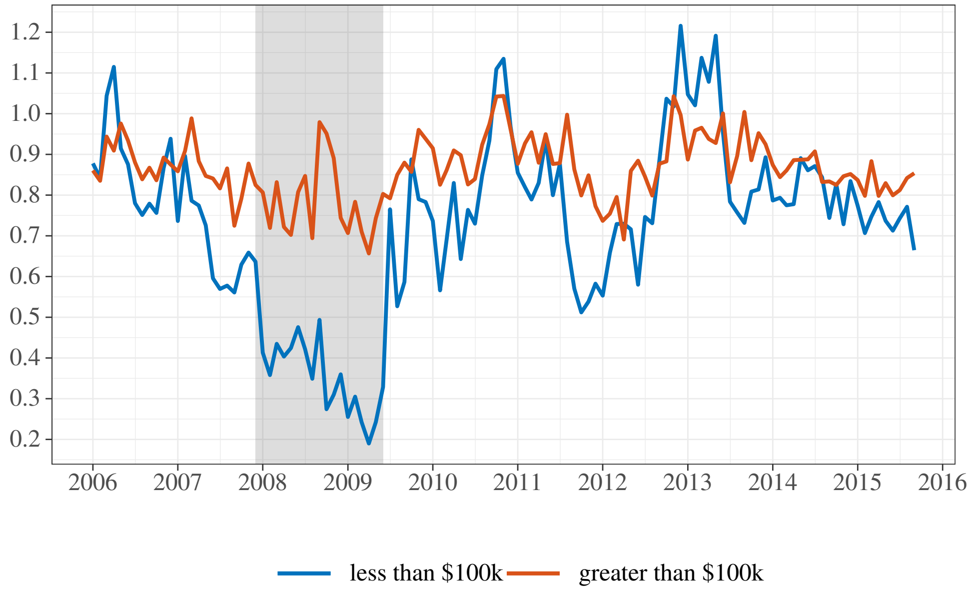

Figure 5. Ratio of the number of customer-sold to number of customer-bought trades. Source: TRACE. The vertical shaded bars indicate NBER recessions.

price dominate, as the inter-dealer price in our benchmark model is slightly higher (1 bps) than in the model without an inventory constraint.

Overall, however, we find that the presence of an inventory constraint in the pre-GFC economy had quite mild effects on equilibrium prices and allocations, relative to an environment where dealers are not required to hold inventory in order to intermediate trade. To get a sense of the welfare cost of the inventory constraint, we calculate the gains from trade that are realized in equilibrium relative to the gains from trade in a frictionless environment.[16] As a point of reference, we find that the welfare loss in an environment without an inventory constraint—but with search and bargaining frictions—is approximately 0.7% trade relative to the frictionless environment. In our environment, where dealers must hold inventory in order to sell, the corresponding welfare loss is approximately 1.2%

[16]More precisely, we calculate $\dfrac{W_{eqm} - W_{aut}}{W_{fb} - W_{aut}}$, where W_{eqm}, W_{fb}, and W_{aut}, denote total welfare in equilibrium, in the (first best) frictionless environment, and in autarky, respectively.

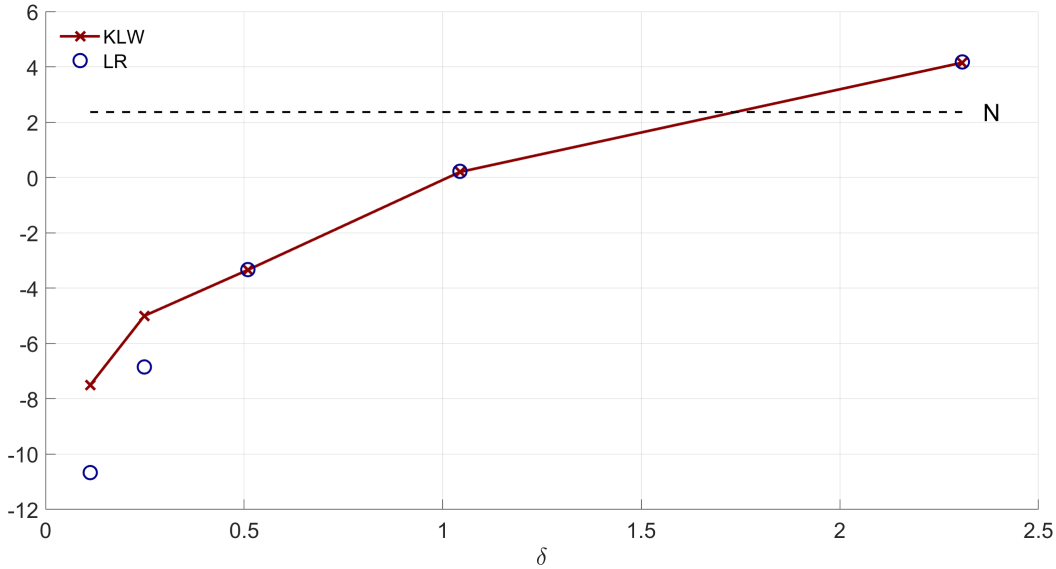

Figure 6. Log of target asset holdings, $q^{\star}(\delta)$, with inventory constraints (CKLW) and without (LR)

3.3 The Effects of Rising Inventory Costs

We now study what happens when we increase the dealers' cost of holding assets (by decreasing u_d) to a level that is consistent with their inventory holdings after the introduction of post-crisis regulations. Figure 7 plots dealers' aggregate inventories as a fraction of total supply, $\mu_d N/A$, as a function of the inventory cost τ. In order to engineer a decline in dealers' inventory holdings from 3% to 1%, we see that inventory costs need to increase from approximately 5% in the pre-crisis calibration to approximately 40%, holding all other parameters fixed.

As inventory costs increase, the inter-dealer price falls, which makes dealers less inclined to buy assets from customers. However, by reducing the flow utility dealers receive from holding the asset, rising inventory costs also make dealers more inclined to sell the asset. Hence, both the bid price and the ask price fall, on average. Figure 8 illustrates that the former effect is stronger than the latter, so that rising inventory costs ultimately lead to larger trading costs for customers. In fact, according to our calibration, the increase in τ consistent with dealers' aggregate inventory

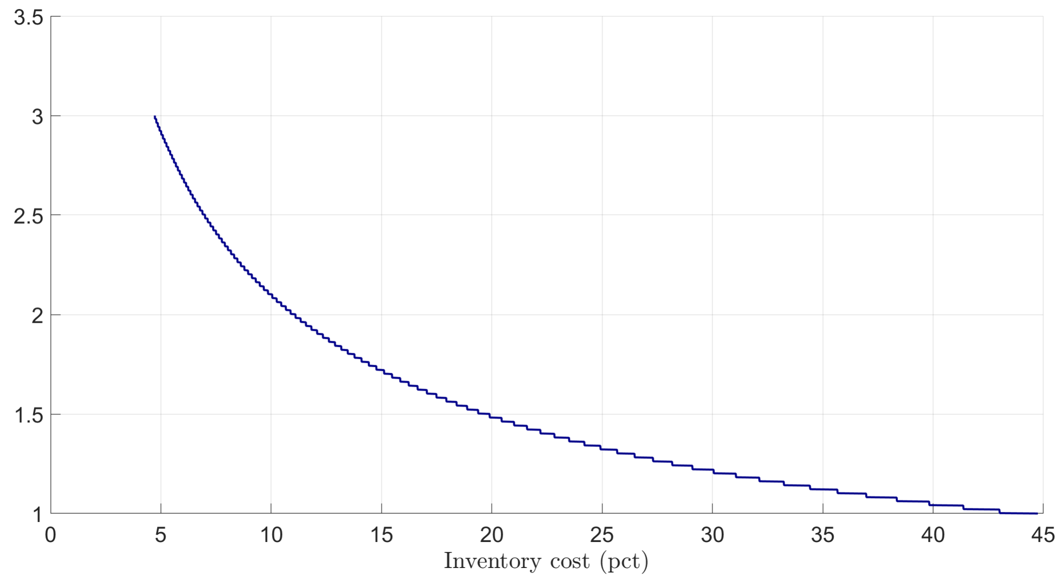

Figure 7. Dealers' aggregate inventory as a percentage of total supply

holdings corresponds to an increase in trading costs of approximately 7 bps; from figure 8, this represents about two-thirds of the increase in trading costs between before the GFC and after the introduction of post-crisis regulations.

Figure 9 illustrates the effect of trading costs on welfare. In particular, we plot the fraction of gains from trade (relative to the frictionless allocation) realized in equilibrium as the inventory cost rises. According to our calibration, an increase in τ from approximately 5% to 40% implies an additional 1% decline in the gains from trade that realized. To the best of our knowledge, this is the first estimate of the welfare cost of increasing balance sheet costs though, of course, we have abstracted here from the potential benefits (e.g., financial stability).

4 Conclusion

We have extended the standard search-theoretic model of dealer-intermediated OTC markets, in which dealers never hold inventory, by introducing a simple and natural "inventory-in-advance"

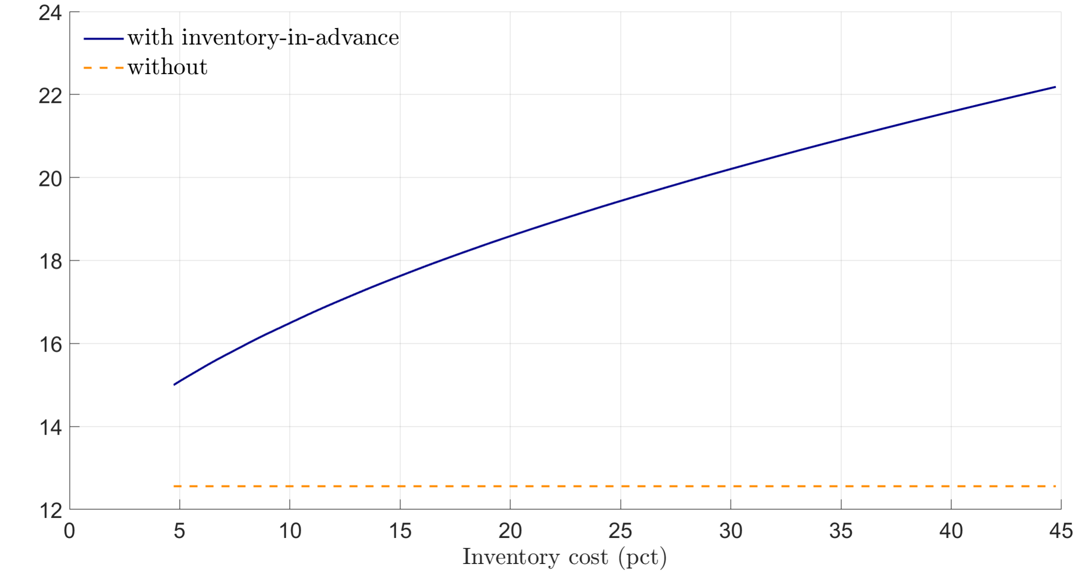

Figure 8. Value-weighted trading costs

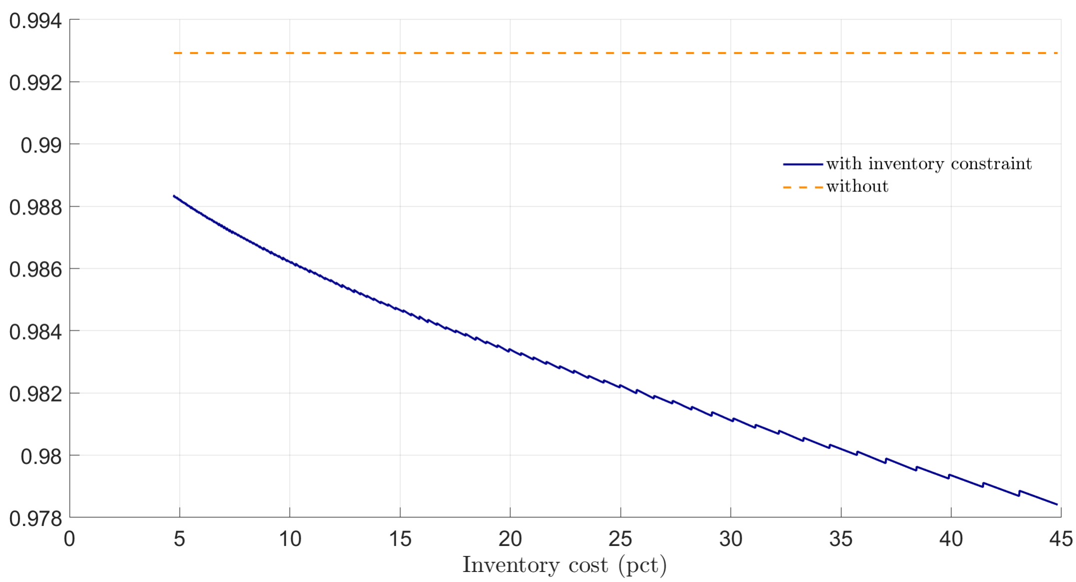

Figure 9. Equilibrium Welfare Relative to First Best Allocation

constraint, which makes inventory a necessary input to intermediation. We have characterized the equilibrium, and study how dealers' optimal inventory choice depends on inventory costs. We have calibrated the model to transaction-level data from the corporate bond market and analyzed the welfare impact of rising inventory costs associated with post-crisis regulation. We measured the welfare loss as the fraction of total gains from trade that the OTC market fails to generate. We have found that rising inventory costs increased the welfare loss substantially, from about 1.2% to about 2.2% of total gains from trade.

References

Yakov Amihud and Haim Mendelson. Dealership market: Marketmaking with inventory. *Journal of Financial Economics*, 8:31–53, 1980.

Yu An. Competing with inventory in dealership markets. Working paper, Stanford University, 2018.

Jack Bao, Maureen O'Hara, and Alex Zhou. The volcker rule and market making in times of stress. *Journal of Financial Economics*, 130:95–113, 2018.

Hendrik Bessembinder, Stacey Jacobsen, William Maxwell, and Kumar Venkataraman. Capital commitment and illiquidity in corporate bonds. *Journal of Finance*, 73:1615–1661, 2018.

Zachary Bethune, Bruno Sultanum, and Nicholas Trachter. An information-based theory of financial intermediation. Technical report, UVA and Federal Reserve Bank of Richmond, 2018.

Giulia Brancaccio, Dan Li, and Norman Schurhoff. Learning by trading: The case of the u.s. market for municipal bond. Working paper, Cornell University, 2017.

Jaewon Choi and Yesol Huh. Customer liquidity provision: Implications for corporate bond transaction costs. Working paper, UIUC and Federal Reserve Board, 2018.

Jens Dick-Nielsen. Liquidity biases in TRACE. *The Journal of Fixed Income*, 19(2):43–55, 2009.

Jens Dick-Nielsen. How to clean enhanced TRACE data. Working paper, CBS, 2014.

Jens Dick-Nielsen and Thomas Kjær Poulsen. How to clean academic TRACE data. Working paper, CBS and BI Norwegian, 2019.

Jens Dick-Nielsen and Marco Rossi. The cost of immediacy for corporate bonds. *Review of Financial Studies*, 32:1–41, 2019.

Darrell Duffie. Post-crisis bank regulations and financial market liquidity. *Lecture, Baffi*, 2018.

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Over-the-counter markets. *Econometrica*, 73:

1815–1847, 2005.

Darrell Duffie, Nicolae Gârleanu, and Lasse H. Pedersen. Valuation in over-the-counter markets. *Review of Financial Studies*, 20:1865–1900, 2007.

Andrea Eisfeldt, Bernard Herskovic, Srimam Rajan, and Emil N. Siriwardane. Otc intermediaries. 2018.

Marham Farboodi, Gregor Jarosch, and Robert Shimer. The emergence of market structure. Working paper, MIT, Princeton, University of Chicago, 2018a.

Maryam Farboodi, Gregor Jarosch, Guido Menzio, and Ursula Wiriadinata. Intermediation as rent extraction. Technical report, MIT, Princeton, NYU, and IMF, 2018b.

Peter Feldhütter. The same bond at different prices: Identifying search frictions and selling pressures. *Review of Financial Studies*, 25(4):1155–1206, 2012.

Nicolae Gârleanu. Portfolio choice and pricing in illiquid markets. *Journal of Economic Theory*, 144: 532–564, 2009.

Alessandro Gavazza. An empirical equilibrium model of a decentralized asset market. *Econometrica*, 84: 1755–1798, 2016.

Michael Gofman. A network-based analysis of over-the-counter markets. Working paper, University of Rochester, 2014.

Michael Gofman. Efficiency and stabilitgy of a financial architecture with too-interconnected-to-fail institutions. *Journal of Financial Economics*, 124:113–14, 2017.

Richard Green, Burton Hollifield, and Norman Schürhoff. Dealer intermediation and price behavior in the aftermarket for new bond issues. *Journal of Financial Economics*, (86):643–682, 2007.

Zhiguo He and Konstantin Milbradt. Endogenous liquidity and defaultable bonds. *Econometrica*, 82:1443–1508, 2014.

Terrence Hendershott, Dan Li, Dmitry Livdan, and Norman Schürhoff. Relationship Trading in OTC Markets. *Journal of Finance*, 75:683–734, 2020.

Thomas Ho and Hans R. Stoll. Optimal dealer pricing under trading transactions and return uncertainty. *Journal of Financial Economics*, 9:47–73, 1981.

Thomas Ho and Hans R. Stoll. The dynamics of dealer markets under competition. *Journal of Finance*, 38: 1053–1074, 1983.

Julien Hugonnier, Benjamin Lester, and Pierre-Olivier Weill. Heterogeneity in decentralized asset markets. Working paper, EPFL, Federal Reserve Bank of Philadelphia, and UCLA, 2014.

Julien Hugonnier, Ben Lester, and Pierre-Olivier Weill. Frictional intermediation in Over-the-Counter markets. *Review of Economic Studies*, Forthcoming, 2019. Working paper EPFL, Federal Reserve Bank of Philadelphia, and UCLA.

Ricardo Lagos and Guillaume Rocheteau. Liquidity in asset markets with search frictions. *Econometrica*, 77:403–426, 2009.

Ricardo Lagos and Shengxing Zhang. Turnover liquidity and the transmission of monetary policy. *American Economic Review*, 110:1635–1672, 2020.

Ricardo Lagos, Guillaume Rocheteau, and Pierre-Olivier Weill. Crashes and recoveries in illiquid markets. Working paper, NYU, UCI, and UCLA, 2008.

Benjamin Lester, Guillaume Rocheteau, and Pierre-olivier Weill. Competing for order flow in otc markets. *Journal of Money, Credit and Banking*, 47(S2):77–126, 2015.

Shuo Liu. Dealer's search intensity in u.s. corporate bond markets. Working paper, UCLA, 2020.

Konstantin Milbradt. Asset heterogeneity in over-the-counter markets. Working paper, Kellogg School of Management, 2017.

Eckart Mildenstein and Harold Schleef. The optimal pricing policy of a monopolistic marketmaker in equity market. *Journal of Finance*, 38:218–231, 1983.

Ed Nosal, Yuet-Yee Wong, and Randall Wright. Intermediation in markets for goods and markets for assets. Working paper, Federal Reserve Bank of Atlanta, 2019.

Emiliano S. Pagnotta and Thomas Philippon. Competing on speed. *Econometrica*, 86(3):1067–1115, 2018.

Ji Shen, Bin Wei, and Hongjun Yan. Financial intermediation chains in an otc market. 2015.

Nancy L. Stokey and Robert E. Lucas. *Recursive Methods in Economic Dynamics*. Harvard University Press, Cambridge, 1989.

George Tauchen and Robert Hussey. Quadrature based methods for obtaining approximate solutions to nonlinear asset pricing models. *Econometrica*, 59(2):371–396, 1991.

Anjan V Thakor. The economic consequences of the volcker rule. *Report by the US chamber's center for capital market competitiveness*, 20120, 2012.

Francesco Trebbi and Kairong Xiao. Regulation and market liquidity. *Management Science*, 2017.

Chung-Yi Tse and Yujing Xu. Inter-dealer trades in otc markets – who buys and who sells? Working paper, University of Hong-Kong, 2018.

Semih Üslü. Pricing and liquidity in decentralized asset markets. *Econometrica*, Forthcoming, 2019.

Pierre-Olivier Weill. Leaning against the wind. *Review of Economic Studies*, 74:1329–1354, 2007.

Pierre-Olivier Weill. The search theory of over-the-counter markets. *Annual Review of Economics*, 12, 2020.

Ming Yang and Yao Zeng. The coordination of intermediation. Working paper, Duke University and University of Washington, 2019.

A Omitted Proofs

A.1 Proof of Proposition 1

Existence, uniqueness, and continuity. Let $x = (\delta, q, p, N)$ and $X = [\underline{\delta}, \overline{\delta}] \times (0, \infty) \times [0, \infty) \times [0, \infty)$. For any strictly positive q' and p', let $X' = [\underline{\delta}, \overline{\delta}] \times [q', \infty) \times [0, p'] \times [0, \infty)$. Now consider the set $C_b(X')$ bounded continuous function of $x \in X'$, endowed with the sup norm. Now, for any $h \in C_b(X')$, define the operator:

$$\begin{aligned} T[h](\delta, q, p, N) =& \frac{u_q(\delta, q) - rp}{r + \gamma + \lambda(1-\theta)} + \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \int h(\delta', q, p, N)\, dF(\delta') \\ &+ \frac{\lambda(1-\theta)}{r + \gamma + \lambda(1-\theta)} \max\{h(\delta, q + N, p, N), 0\}. \end{aligned}$$

Given that the domain is restricted to $q \geq q'$ and $p \leq p'$, $u_q(\delta, q) - rp$ is bounded, so $T[h]$ is bounded. One easily sees that it is continuous. Hence, the operator T maps $C_b(X')$ into itself, and to verify that it satisfies the Blackwell sufficient condition for a contraction (see Theorem 3.3 in Stokey and Lucas, 1989), with modulous of contraction $(\gamma + \lambda(1-\theta))/(r + \gamma + \lambda(1-\theta))$. This establishes uniqueness of a solution over any X'. Given uniqueness, by letting $q' \to 0$ and $p' \to \infty$, this solution can be extended over the entire set X.

Monotonicity. The operator T preserve the stated monotonicity properties: that is, if h is increasing in δ and decreasing in (q, p, N), then so is $T[h]$. Since monotonicity properties are preserved by passing to the limit, they are inherited by the fixed point. Note that the first term of $T[h]$, $u_q(\delta, q) - rp$ is strictly increasing in δ, strictly decreasing in q, and strictly decreasing in p. Hence, the fixed point inherit these strict monotonicity properties as well.

$\Sigma(\delta, q) > 0$ for q small enough, and $\Sigma(\delta, q) < 0$ for q large enough. Let $\underline{q}$ denote the solution to $u_q(\underline{\delta}, q) = rp$ and $\overline{q}$ the solution to $u_q(\overline{\delta}, q) = rp$. Evaluating $T[\Sigma]$ at $(\underline{\delta}, \underline{q})$ and $(\overline{\delta}, \overline{q})$, and keeping in mind that Σ is a fixed point, we obtain:

$$\begin{aligned} \Sigma(\underline{\delta}, \underline{q}) &= \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \int \Sigma(\underline{q}, \delta')\, dF(\delta') + \frac{\lambda(1-\theta)}{r + \gamma + \lambda(1-\theta)} \max\{\Sigma(\underline{\delta}, \underline{q} + N), 0\} \\ \Sigma(\overline{\delta}, \overline{q}) &= \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \int \Sigma(\overline{q}, \delta')\, dF(\delta') + \frac{\lambda(1-\theta)}{r + \gamma + \lambda(1-\theta)} \max\{\Sigma(\overline{\delta}, \overline{q} + N), 0\}, \end{aligned}$$

where we omitted the dependence of Σ on (p, N) for notational convenience. Now using that Σ is increasing in δ, that $\max\{\Sigma(\underline{\delta}, \underline{q} + N), 0\} \geq 0$, and that $\Sigma(\overline{\delta}, \overline{q} + N) \leq \Sigma(\overline{\delta}, \overline{q})$, we obtain

$$\begin{aligned} \Sigma(\underline{\delta}, \underline{q}) &\geq \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \Sigma(\underline{q}, \underline{\delta}) \\ \Sigma(\overline{\delta}, \overline{q}) &\leq \frac{\gamma}{r + \gamma + \lambda(1-\theta)} \Sigma(\overline{q}, \overline{\delta}) + \frac{\lambda(1-\theta)}{r + \gamma + \lambda(1-\theta)} \max\{\Sigma(\overline{\delta}, \overline{q}), 0\}, \end{aligned}$$

These two equations clearly imply that $\Sigma(\underline{\delta}, \underline{q}) \geq 0$ and $\Sigma(\overline{\delta}, \overline{q}) \leq 0$. Given that Σ is strictly increasing in δ and strictly decreasing in q the result follows.

Using $\Sigma(\delta, q)$ to construct the value function $V(\delta, q)$. As defined in the paragraph following Proposition 1, let us $q^\star(\delta)$ denote the target holding and $x^\star(\delta, q) = \min\{q^\star(\delta,) - q, N\}$ denote the optimal trade of an investor of type δ with current holding equal to q. Fix any $q_0 \in (0, \infty)$ and let $\delta \mapsto V(\delta, q_0)$ solve:

$$\begin{aligned}(r + \gamma + \lambda(1-\theta))V(\delta, q_0) =& u_q(\delta, q) \\ &+ \gamma \int \left[V(\delta', q_0) - V(\delta, q_0)\right] \, dF(\delta') + \lambda(1-\theta) \int_{q_0}^{q_0 + x^\star(\delta, q_0)} \Sigma(\delta, x) \, dx.\end{aligned}$$

The existence and uniqueness of such a function is guaranteed by standard contraction-mapping arguments. Our guess for the value function at any (δ, q) is:

$$V(\delta, q) = V(\delta, q_0) + \int_{q_0}^{q} \Sigma(\delta, x) \, dx + p(q - q_0).$$

One can then directly verify after some algebra that $V(\delta, q)$ constructed above solves the HJB equation (1).

A.2 Proof of Lemma 1

Concavity in Lemma 1 follows from the fact that the objective function inside the maximum is concave in the choice variable x (indeed, the derivative of $V(\delta', q' + x) - V(\delta', q') - px$ with respect to x is equal to $\Sigma(\delta', q' + x)$, which is decreasing according to Proposition 1), and the constraint set is defined by linear inequalities. To obtain the derivative, one easily verifies that, for any (δ', q')

$$\frac{d}{dn} \max_{x \leq n} \left\{V(\delta', q' + x) - V(\delta', q') - px\right\} = \max\{\Sigma(\delta', q' + n), 0\}.$$

The expression for the derivative follows.