Asset Heterogeneity, Market Fragmentation, and Quasi-Consolidated Trading [*]

Wei Li[†] Zhaogang Song[‡]

Last updated: December 7, 2022
(Latest version at: https://bit.ly/3B85vVZ)

Abstract

Asset heterogeneity is widely believed to restrict liquidity in many markets involving important fixed-income assets. We model the impact of quasi-consolidated (QC) trading—a design that allows sellers to deliver heterogeneous assets for identical payments—on over-the-counter (OTC) markets involving assets with varying values. We show that allowing for QC trading reduces market fragmentation but introduces a cheapest-to-deliver (CTD) effect. In consequence, although QC trading increases total trading volume and social welfare, it hurts liquidity for sellers who do not switch to QC trading and lowers profits for both these sellers and some other sellers who switch to QC trading. Consolidating multiple QC contracts increases (decreases) total trading volume and social welfare if the contracts cover assets with similar (distinct) values.

Keywords: Asset heterogeneity, Mortgage-backed Securities, Over-the-Counter Markets, Quasi-consolidated trading, TBA.

JEL Codes: G1, G11, G12, G21, D83, D53, D61

[*]We are grateful to Doug McManus at Freddie Mac and Marty Young at Goldman Sachs for helpful conversations about the MBS market. We appreciate the insightful comments of Hank Bessembinder, Vincent Maurin, Semih Üslü, seminar participants at Johns Hopkins University, and participants of the MFA 2021 meeting.

[†]Johns Hopkins Carey Business School, 100 International Drive, Baltimore, MD 21202. E-mail: wei.li@jhu.edu.

[‡]Johns Hopkins Carey Business School, 100 International Drive, Baltimore, MD 21202. E-mail: zsong8@jhu.edu.

Asset Heterogeneity, Market Fragmentation, and Quasi-Consolidated Trading

December 7, 2022

Abstract

Asset heterogeneity is widely believed to restrict liquidity in many markets involving important fixed-income assets. We model the impact of quasi-consolidated (QC) trading—a design that allows sellers to deliver heterogeneous assets for identical payments—on over-the-counter (OTC) markets involving assets with varying values. We show that allowing for QC trading reduces market fragmentation but introduces a cheapest-to-deliver (CTD) effect. In consequence, although QC trading increases total trading volume and social welfare, it hurts liquidity for sellers who do not switch to QC trading and lowers profits for both these sellers and some other sellers who switch to QC trading. Consolidating multiple QC contracts increases (decreases) total trading volume and social welfare if the contracts cover assets with similar (distinct) values.

1 Introduction

Asset heterogeneity is a salient feature of many over-the-counter (OTC) markets for fixed-income securities. In the corporate bond market, for example, "as of December 2017, the firms in the S&P 500 index had 11,990 outstanding bond CUSIPs, while firms in the Russell 1000 index had 13,083 outstanding bond CUSIPs"(Bessembinder, Spatt and Venkataraman, 2019). In the municipal bond market, some 50,000 issuers had issued more than 1.5 million bonds by the end of 2017. In the agency mortgage-backed securities (MBS) market, thousands of CUSIPs are issued every month with substantially different prepayment characteristics.[1]

Such high asset heterogeneity is believed to render these assets less liquid. Bessembinder, Spatt and Venkataraman (2019) argue, for example, that "one reason that individual corporate bonds trade less frequently than equities is that an issuer often has multiple bond issues outstanding. While equity shares issued at different points in time by a given firm are fully substitutable, each bond issue is a separate contract with differing promised payments, maturity dates, and priority in case of default, and is therefore traded separately." Indeed, round-trip costs, a standard measure of trading costs for over-the-counter (OTC) securities, are estimated to be above 100 basis points (bps) for corporate bonds and above 70 bps for municipal bonds (Bessembinder and Maxwell, 2008; Di Maggio, Kermani and Song, 2017; Li and Schürhoff, 2019). In comparison, the average trading cost of U.S. Treasury securities, which are much less heterogeneous, is about one to two basis points.

Nonetheless, the agency MBS market, which also features highly heterogeneous assets, presents an intriguing exception. It is actually "one of the most liquid fixed income markets in the world, with trading volumes typically in the trillions of dollars per year, involving many different types of investors."[2] Such great liquidity originates mainly in to-be-announced (TBA) contracts, which allow a cohort of similar (but nonidentical) MBSs to be sold at the same price. Indeed, as estimated by Gao, Schultz and Song (2017) and Bessembinder et al. (2019), the average round-trip cost in TBA trading is only about 2 bps, which is comparable to that for U.S. Treasuries. One might argue that the exceptional liquidity of TBA contracts arises mainly from the U.S. government's guarantee of default risks rather than trading design. Nevertheless, the average round-trip cost of standard trades that, like trades involving corporate and municipal bonds, fully specify the traded security and are known as the specified pool (SP), is about 20–60

[1] The cash flow of an MBS is generated by a collection of mortgages. Agency MBSs are guaranteed by Fannie Mae, Freddie Mac, or Ginnie Mae, such that investors are protected from losses if borrowers default on their mortgages. Agency MBSs are still risky because borrowers may prepay before maturity dates.

[2] https://www.alliancebernstein.com/sites/library/Instrumentation/MORT-MAG-GR-EN-0118-FINAL.PDF.

bps. The superior liquidity of TBA contracts over SP contracts suggests that the TBA trading design indeed helps improve market liquidity.

How does asset heterogeneity restrict market liquidity? How does TBA trading improve liquidity in the presence of asset heterogeneity? Do these effects vary across investors who trade different kinds of assets? Answering these questions is important not only for understanding the economic channels that facilitate OTC trading but also for regulating market designs. In fact, because of the extraordinary liquidity in the TBA market, it has been conjectured that introducing a TBA-like trading mechanism can enhance the liquidity of corporate bonds and municipal bonds. For example, Spatt (2004) mentions that "the analogy to the mortgage markets is instructive. Trading instruments based upon their main characteristics . . . may be helpful and narrow the spreads." Moreover, Gao et al. (2017) argue that "corporate and municipal bonds trade in relatively illiquid over-the-counter markets. Parallel trading in the securities themselves and a forward contract on a generic security may increase the liquidity of those markets." Bessembinder et al. (2019) also ask whether there is "scope for the trading of packages of corporate bonds based on a set of prescribed characteristics." In this paper, we build a theoretical framework to address these questions.

We first develop a baseline setup without TBA trading to demonstrate how asset heterogeneity restricts market liquidity. In the model, risk-neutral investors trade heterogeneously valued assets using standard contracts, which we call asset-specific (AS) contracts because such contracts fully specify the traded assets. We assume for simplicity that every seller owns one asset and every buyer can purchase no more than one asset. Naturally, a seller can sell only her own asset. A buyer, instead, faces no constraints regarding which asset she can purchase, consistent with the fact that in practice many investors do not prefer specific assets within a given asset class.[3] Nevertheless, buyers pay costs to participate in the trading of a given asset (Vayanos and Wang, 2013). For example, they need to collect data, predict future cash flows, and build pricing models to value assets. Because assets in our model differ in value, a buyer needs to pay such an analysis cost for every asset she trades. When this cost is higher, the buyer will likely participate in fewer trades. For simplicity, we consider a situation in which the participation cost is sufficiently high that a buyer participates in the trading of at most one asset. In consequence, in equilibrium the AS market is fragmented into multiple segmented AS submarkets, each of

[3] Spatt (2004), for example, points out that many corporate bonds and municipal bonds are rated AAA, among which "the identity of the individual issuer does not seem especially significant" and "typically, buyers are not focused upon particular securities but instead are interested in purchasing a security with certain characteristics." Our model can be extended to include buyers who prefer specific assets, e.g., those with low default or prepayment risks.

which consists of one seller and a distinct subset of buyers.

How does the resulting market fragmentation affect liquidity? Intuitively, fragmentation reduces the number of counterparties a trader faces, making it harder for her to trade. To capture this effect, we use a static search-and-matching framework to model the trading process. Specifically, on each trading venue, buyers and sellers are randomly chosen and matched; a trade then occurs between every matched buyer and seller pair. We thus measure the liquidity level a trader experiences with the probability that she trades. Importantly, we assume that the matching function features increasing returns to scale so that concentrating more traders in one trading venue improves liquidity; this effect reflects the positive network externality of market liquidity, often characterized as the"liquidity begets liquidity" feature in the literature (Vayanos and Weill, 2008; Weill, 2020). Because the AS market is fragmented, the liquidity benefit of concentrating traders is not fully realized. Thus, the AS market is less liquid than a market in which all agents trade homogeneous assets on the same venue.

In our main analysis, then, we introduce a TBA-like contract to the baseline market. Differing from a *fully* consolidated contract (e.g. an index ETF) that effectively combines a basket of assets into one tradable security, a TBA-like contract allows a seller to deliver *any* asset that meets certain eligibility requirements. We therefore call it a quasi-consolidated (QC) contract. While an AS contract sets a price that is specific to the traded asset, a QC contract sets a uniform price that we assume to equal the average value of assets traded through the QC contract.[4] A seller can sell her asset using either an AS contract or the QC contract. As in the baseline model, we assume that buyers need to pay a cost to participate in every trading venue—which can be either an AS submarket or the QC market—and consider the situation in which these participation costs are sufficiently high that a buyer participates in at most one trading venue. Multiple segmented trading venues, including one QC market and a number of AS submarkets, thus emerge.

The parallel-trading equilibrium with both AS and QC trading is as follows.[5] Because a QC contract prices all assets uniformly, lower-value assets are more likely to be delivered to such a contract, which is known in practice as a cheapest-to-deliver (CTD) contract. Therefore, in equilibrium, sellers choose AS trading only if their asset values exceed an endogenous thresh-

[4]Such pricing arises because buyers are risk-neutral. We investigate the effects of buyer risk-aversion in Appendix B.3. Because buyers in the QC market may receive heterogeneous assets at delivery, risk-averse buyers would bid a QC price that is lower than average value of QC assets, which then weakens sellers' incentives to choose QC trading; as a result, QC trading should attract fewer traders and generate lower liquidity benefits.

[5]We focus on the equilibrium in which both AS and QC trading are used. The degenerate equilibria in which only AS or QC trading is used are not only straightforward theoretically but also inconsistent with practice observed in, for example, markets for MBSs.

old. If QC trading is *strictly* less liquid than AS trading, then no seller would choose QC trading because of the CTD pricing; therefore, in equilibrium QC trading must be weakly more liquid than AS trading. Moreover, *buyers* on separate trading venues experience the same level of liquidity. In particular, because buyers are ex-ante homogeneous, every buyer earns the same expected profit in equilibrium; because a buyer's expected profit increases with the probability that she trades, in equilibrium a buyer on any trading venue must trade at the same probability.[6]

By comparing this parallel-trading equilibrium with the pure AS equilibrium, we reveal the following effects of introducing QC trading to the market.

We first consider the *overall* effects of introducing QC trading. We show that the introduction of QC trading increases both total trading volume and social welfare. Because in our model the total number of assets is given exogenously, the total trading volume is proportional to the *average* probability that an asset on any trading venue is sold; thus, the total trading volume measures overall liquidity. While in the pure AS equilibrium every asset is traded on a separate venue, in the parallel-trading equilibrium multiple assets are traded together on the QC market; thus, introducing QC trading partially "defragments" the market. Because of the liquidity benefit of pooling more traders together, trading frictions are reduced overall and the total trading volume increases. Social welfare in our model equals the total trading gain less the total participation costs buyers pay. Introducing QC trading increases the trading volume, reallocating more assets to buyers who value the traded assets more dearly than sellers; thus, the total trading gains increase. In addition, introducing QC trading reduces the probability that a buyer pays the participation costs but fails to trade; thus, although the participation cost per buyer is fixed, the average participation cost *per successful trade* is reduced. Therefore, social welfare also increases.

We then show that the effects of introducing QC trading vary across *individual* traders.

First, buyers on both QC and AS trading venues experience better liquidity and earn higher expected profits. Specifically, QC trading attracts traders from AS submarkets because it can improve liquidity by concentrating multiple traders. Note that, while switching to QC trading is free for buyers, it is costly for sellers who own high-value assets because of the CTD price discount. Thus, *disproportionately more* buyers than sellers migrate from AS submarkets to the QC market. Compared with buyers in the pure-AS equilibrium, those who migrate to QC trading enjoy better liquidity because of the liquidity externality of concentrating multiple sellers;

[6]When the participation cost per trading venue is sufficiently high, some (or all) buyers choose not to participate in any trading venue and earn zero profit; in this situation, a buyer who participates in trading also earns zero profit because her trading profit exactly offsets the cost of participating.

buyers who remain in AS submarkets also enjoy better liquidity because fewer buyers compete to purchase each asset. Because each buyer is more likely to trade, she is more likely to capture the trading gain and therefore earns a higher profit in expectation.

Second, sellers on AS submarkets experience worse liquidity and earn lower expected profits. Specifically, because disproportionately more buyers than sellers migrate from AS submarkets to the QC market, fewer buyers remain in AS submarkets. In consequence, a seller on an AS submarket is less likely to find a buyer with whom to trade and earns a lower profit in expectation. Importantly, this result shows that, although QC trading increases social welfare, it is not Pareto-improving; thus, QC trading is unlikely to emerge without intervention by regulators.

Third, on the QC market, while liquidity improves for all sellers, the expected profit rises for sellers of low-value assets but drops for sellers of high-value assets. In particular, all QC sellers enjoy better liquidity because of three effects. First, QC trading reduces trading frictions for any given buyer-to-seller ratio. Second, because disproportionately more buyers than sellers migrate to the QC submarket, the buyer-to-seller ratio on the QC market increases. Third, QC trading could attract additional buyers to participate in trading, further raising the buyer-to-seller ratio on the QC market. All these effects improve liquidity for sellers on the QC market.

Moreover, sellers whose asset values fall below the QC price earn higher expected profits because they benefit in terms of both liquidity and price. Nonetheless, sellers whose asset values exceed the QC price enjoy a liquidity benefit but suffer from the CTD pricing, so some earn lower expected profits. Consider the marginal seller who is indifferent between the QC market and an AS submarket. Her expected profit equals the expected profit she could earn by selling on an AS submarket, which, as explained above, is lower than her expected profit in the pure-AS equilibrium because of the worsened AS liquidity. Intuitively, this marginal seller suffers because QC trading reduces the value of her outside option to sell on the AS market. Overall, on the QC market, sellers of low-value assets earn higher expected profits and sellers of high-value assets earn lower expected profits than they would have earned in the pure-AS equilibrium.

In summary, we find that introducing QC trading improves liquidity for buyers in all markets and sellers in the QC market but restricts liquidity for sellers who remain in AS submarkets. In addition, all sellers who remain in AS submarkets and some sellers who switch to the QC market earn lower profits, while other traders earn higher profits. In particular, the QC contract does not benefit all traders because of the CTD issue associated with asset heterogeneity.

Finally, we analyze two issues related to QC contract design based on our model. First, although we take QC-eligibility requirements as given in the above analyses, in practice regulators can vary these requirements to change the set of QC-eligible assets. We show that the overall trading volume and social welfare are maximized when regulators choose QC-eligibility standards that maximize the (endogenous) proportion of assets traded through QC trading. In this situation, however, sellers who remain in AS submarkets would suffer the most severe losses in expected profits.

Second, we analyze the impact of consolidating multiple QC contracts. The analysis is motivated by the recent Single Security Initiative implemented by the Federal Housing Finance Agency (FHFA). Specifically, in June 2019, the FHFA consolidated TBA contracts for Fannie Mae MBS and TBA contracts for Freddie Mac MBSs into TBA contracts for Uniform Mortgage-Backed Securities (UMBSs), which include both Fannie Mae and Freddie Mac MBSs (Goodman and Parrott, 2018). We show that the effects of such a consolidation depend on the extent to which the two types of assets differ *in value*. In particular, the consolidation could decrease (increase) the total trading volume and social welfare if the two types of assets have very different (very similar) values. If, for example, the value ranges of two types of assets do not overlap, then consolidating two type-specific QC contracts for these assets could drive all assets of the high-value type to the AS market, lowering the total trading volume and social welfare. If, for example, all type-A assets are more valuable than all type-B assets, then consolidating type-specific QC contracts could drive all type-A assets to AS submarkets, reducing the total trading volume and social welfare. Our model hence provides a theoretical rationale for regulators' efforts to align the characteristics of Fannie Mae and Freddie Mac MBSs when implementing UMBSs (Liu, Song and Vickery, 2020).

Several studies have informally alluded to reduced market fragmentation when considering explanations for the liquidity of the TBA MBS market. For example, Vickery and Wright (2011) state that "the TBA trading convention allows trading to be concentrated in only a small number of liquid forward contracts" and Gao et al. (2017) state that "the TBA market takes thin markets for thousands of different MBS with different prepayment characteristics and trades them through a handful of thickly traded cheapest-to-deliver contracts. In this way, liquidity is increased for the MBS that are traded there instead of as SPs." To the best of our knowledge, we develop the first theoretical model that examines how TBA-like QC trading affects liquidity in markets for assets with heterogeneous values.[7] Our model not only explains the liquidity of the

[7] For empirical studies of the agency MBS market liquidity, see Bessembinder, Maxwell and Venkataraman (2013), Downing, Jaffee and Wallace (2009), Gao et al. (2017), Gao, Schultz and Song (2018), Schultz and Song (2019), and Huh and Kim (2019) among others. Related asset pricing studies include Gabaix, Krishnamurthy and Vigneron (2007), Duarte, Longstaff and Yu (2007), Chernov, Dunn and Longstaff (2018), Song and Zhu (2019), Boyarchenko, Fuster and Lucca (2019), Diep, Eisfeldt and Richardson (2021), and Carlin, Longstaff and Matoba (2014), Fusari, Li, Liu and Song (2022), and He and Song (2019), among others.

agency MBS market but also sheds light on the liquidity of various other OTC markets involving Treasury securities, corporate bonds, and municipal bonds.

A large body of literature studies market liquidity in OTC markets; see Weill (2020) for a recent comprehensive survey. Most studies in this literature study markets involving homogeneous assets. Our paper studies markets involving heterogeneous assets and is most closely related to Vayanos and Wang (2007), Vayanos and Weill (2008), Weill (2008), Milbradt (2018), An (2019) and Üslü and Velioglu (2019), which also feature multiple assets. In particular, Vayanos and Wang (2007) and Vayanos and Weill (2008) also investigate the impact of market fragmentation and focus on explaining why assets with the *same* value can trade at differing prices. In their models, the market fragments because of various portfolio constraints imposed on traders. In contrast, in our model the market fragments because assets differ in *value* and it is costly for buyers to analyze assets traded on a certain venue. We can therefore examine how the CTD issue embedded in QC contracts prevents certain traders from switching to QC trading, limiting the overall benefits and hurting certain traders.

Security issuers can, of course, mitigate the frictions caused by asset heterogeneity by designing less heterogeneous securities. Gorton and Pennacchi (1990), Subrahmanyam (1991), and DeMarzo (2005), for example, study how consolidating assets can improve market liquidity by diversifying risks and reducing information sensitivity. We differ by studying how a particular trading mechanism can help improve liquidity without changing the securities being issued. It is also worth pointing out that, although we focus on how QC trading improves market liquidity by mitigating frictions resulting from asset heterogeneity, we do not claim that this is the only benefit of QC trading. TBA contracts in the agency MBS market, for example, also enable issuers to hedge interest rate risks and enable outside investors to access the MBS market.

2 Institutional Background and Motivation

In this section, we briefly introduce the institutional background on the structure and liquidity of OTC markets involving major U.S. fixed-income assets, including Treasury securities, agency MBS, corporate bonds, and municipal bonds; see Bessembinder et al. (2019) for a comprehensive survey. We focus on the prevalence of asset heterogeneity across most of these markets. We also motivate the potential for QC trading to improve market liquidity.

In particular, fixed-income markets feature massive outstanding balances and new issuances. Countries, municipalities, corporate firms, and households raise debt capital in these markets. As Table 1 shows, based on data from the Securities Industry and Financial Markets Associa-

Table 1. Summary of U.S. Fixed-Income Markets

Markets	Outstanding Balance ($ trillion)	Asset heterogeneity		Trading cost (bps)	
		# of securities	Extent of heterogeneity	AS	QC
US Treasury	17	200	Small	2-3	1-2
Agency MBS	7	30,000	Medium	60-80	2-3
Corporate bond	8	11,990	Large	80-100	N/A
Municipal bond	6	1.5 million	Large	80-100	N/A

In this table, we report aggregate summaries of U.S. fixed-income markets for Treasury securities, agency MBS, corporate bonds, and municipal bonds, based on SIFMA reports, FINRA reports, and estimates in Gao et al. (2017), Di Maggio et al. (2017), and Li and Schürhoff (2019).

tion (SIFMA), as of the third quarter of 2019 the outstanding balance was about $16 trillion (tn) for Treasuries, $8.6 tn for agency MBS, $9.6 tn for corporate bonds, and $3.8 tn for municipal bonds.[8] Moreover, the aggregate total value of new issuances of all U.S. fixed-income assets was $7.4 tn in 2018, which was substantially greater than the equity issuance of $221.2 billion.[9]

Differing from equity markets, in which most assets are traded anonymously through all-to-all electronic limit order books, fixed-income assets are traded mainly bilaterally in OTC markets. In the U.S. Treasury market, for example, only the inter-dealer segment of on-the-run securities trades through an all-to-all trading platform run by inter-dealer brokers. Both the dealer-client segment of on-the-runs and all off-the-run securities, which account for over 95% of outstanding Treasuries, are traded off the all-to-all platform (Clark, Cameron and Mann (2016)).[10] These trades are done either through traditional voice/screen bilaterally or electronically facilitated requests for quotations (RFQs) where a client usually asks several dealers for indicative quotes. Either way, transactions involve standard OTC frictions, such as the need to search for counter-parties, negotiations, and bilateral settlements, which are usually modeled through search-theoretical frameworks, as reviewed in Duffie (2012) and Weill (2020). All-to-all electronic trading platforms are also used for agency MBS, but less often than for Treasuries, and used even less often for corporate and municipal bonds (O'Hara and Zhou (2020)).

A salient feature of fixed-income markets is substantial asset heterogeneity. As Table 1 shows, as of December 2017, there were 11,990 corporate outstanding bonds issued by firms in the S&P 500 index, over 1.5 million municipal bond issued by 50,000 issuers, and more than

[8]Other important but smaller fixed-income markets include the non-agency MBS markets ($1.7 tn), the federal agency securities market ($1.8 tn), and the ABS ($1.9 tn), among others.

[9]Of course, part of the large new issuance amount is because many fixed-income securities have short tenors and are rolled over at maturity. See https://www.sifma.org/resources/research/fact-book/ for details.

[10]Principal trading firms that specialize in electronic and automated intermediation participate in the inter-dealer segment. The summary of the trading volume of Treasuries can be found at https://nyfed.org/37HquPq.

30,000 outstanding agency MBS. In contrast, the U.S. Treasury market is much homogeneous: there are only up to 200 outstanding securities issued by a single entity—the U.S. government.

Asset heterogeneity has been conjectured to reduce market liquidity since Demsetz (1968). A comparison of the extent of asset heterogeneity and the magnitude of the trading costs for standard asset-specific trading indeed suggests so. In particular, the trading costs fall within the range of 80-100 bps for corporate and municipal bonds, which feature the greatest asset heterogeneity, and the range of 40-60 bps for agency MBS, which feature medium level of asset heterogeneity because of the credit risk guarantee. In contrast, transaction costs are only about 2 bps for Treasuries.

In addition to the standard AS trading mechanism, the Treasury and agency MBS markets also allow for alternative trading mechanisms: Treasury futures and MBS to-be-announced forward (TBA) contracts, respectively. The former are traded on the Chicago Mercantile Exchange (CME), while the latter are traded OTC. These two contracts are both standardized contracts that accept any eligible securities within a cohort for delivery. For example, the classic "30-year Treasury bond futures" accept the delivery of any bond with remaining term-to-maturity of 15 years or more, which includes a significant number of securities ranging widely in terms of coupon and maturity. Similarly, Fannie Mae 30-year 5% coupon TBA contracts accept any TBA-eligible MBS with these features and remaining term-to-maturity of at least 15 years.[11] Both are thus priced in the knowledge that sellers have incentives to deliver the cheapest eligible securities at settlement.

The "cohort-based" feature of these quasi-consolidated contracts is widely conjectured to have great potential for mitigating frictions associated with asset heterogeneity and improving market liquidity (Spatt (2004), Vickery and Wright (2011), Gao et al. (2017), and Bessembinder et al. (2019)). As summarized in Table 1, QC trading indeed incurs very low transaction costs. In particular, the QC trading cost is only 2 bps for agency MBS, substantially lower than that for AS trading. Moreover, the difference in trading costs between QC and AS contracts is substantially wider for agency MBS than for Treasuries, suggesting that QC trading is particularly helpful in improving liquidity in markets that feature considerable asset heterogeneity.

Not only is QC trading recommended for corporate and municipal bonds as a new mechanism, but also the design of QC trading has been reformed to further improve market liquidity in the agency MBS market. In particular, in June 2019, the Single Security Initiative by the Federal Housing Finance Agency (FHFA) was formally implemented, under which Fannie Mae and

[11] More specifically, Treasury futures contracts standardize more features, including the size of each contract, than TBA MBS contracts, which might contribute to its greater liquidity over TBA MBS contracts.

Freddie Mac MBS are consolidated into "Uniform" MBS (UMBS). The single TBA contract for UMBS replaces the two TBA contracts for Fannie Mae and Freddie Mac MBS.

3 Baseline Model of OTC Trading with Heterogeneous Assets

In this section, we develop a baseline model in which traders use only standard AS contracts that fully specify the deliverable assets. Using this model, we demonstrate how asset heterogeneity leads to market fragmentation and hurts liquidity. In the next section, we extend the baseline model and allow traders to use QC contracts that partially specify deliverable assets.

3.1 Setup

Assets are traded bilaterally between b buyers and s sellers, who are all risk-neutral.[12] Every seller intends to sell one asset, and every buyer can buy up to one asset. An asset in our model can represent a set of similar assets in practice. Let $j \in \{1, 2, \cdots, s\}$ index the assets. We assume that asset j is worth v_j to sellers and $v_j + \delta$ to buyers, so every trade would result in a gain of $\delta > 0$. Asset value $v_j \overset{\text{iid}}{\sim} F$ with support $\mathcal{V} = [v_{\min}, v_{\max}]$.

Social welfare is reduced because (1) buyers need to pay costs to participate in trading and (2) trading frictions exist on all trading venues. Importantly, buyers' participation costs result in market fragmentation, which hurts liquidity because trading frictions are more severe on smaller trading venues.

Naturally, a seller can use only the AS contract that is specific for her asset. Although a buyer could purchase any asset, we assume that a buyer needs to pay c to participate in the trading of every asset. In particular, because assets differ in values, to accurately value an asset, a buyer needs to conduct analysis on the asset that may involve, for example, collecting data, predicting future cash flows, and building pricing models (Eisfeldt, Lustig and Zhang, 2019). When c increases, each buyer participates in the trading of fewer assets, which results in more severe market fragmentation because every seller faces fewer buyers.

Moreover, trading frictions exist and are more severe on smaller trading venues. There are various ways to model trading frictions; we use a simple search-and-matching framework to capture how trading frictions depend on the number of traders choosing a trading venue. Specifically, we assume that (1) a trade may occur *only* after a buyer and a seller are matched

[12] In Appendix B.3, we allow buyers to be risk-averse and show that, despite the inherent adverse selection issue, the QC market can still attract sellers because of its liquidity advantage. Thus, risk-neutrality does not qualitatively change our main results.

and (2) when s_j sellers and b_j buyers participate in trading venue j, the expected measure of buyer-seller matches on this venue equals

$$m(s_j, b_j) = \lambda \cdot (s_j b_j)^{\frac{1+\theta}{2}}, \tag{1}$$

where exogenous parameters λ measures the matching efficiency and $\theta \in (0,1)$ represents the liquidity benefit obtained by pooling traders.[13] A trader experiences greater liquidity if she is more likely to be matched with a counter-party. We measure the liquidity levels for a seller and a buyer on trading venue j by, respectively,

$$\pi_j^s = \frac{m(s_j, b_j)}{s_j} = \lambda \left(\frac{b_j}{s_j}\right)^{\frac{1-\theta}{2}} b_j^\theta \quad \text{and} \quad \pi_j^b = \frac{m(s_j, b_j)}{b_j} = \lambda \left(\frac{s_j}{b_j}\right)^{\frac{1-\theta}{2}} s_j^\theta \tag{2}$$

because, according to Eq. (1), π_j^s and π_j^b equal, respectively, the probabilities that a seller and a buyer are matched on trading venue j. To focus on the situation in which no trading venue is perfectly liquid, we assume that λ is so low that π_j^s and π_b^b never exceeds 1.

Eq. (2) reveals two properties regarding liquidity levels π_j^s and π_j^b. First, buyers and sellers in a trading venue generally experience different levels of liquidity (i.e., $\pi_j^s \neq \pi_j^b$). Increasing the number of buyers b_j, for example, would enhance liquidity for sellers π_j^s but hurt liquidity for buyers π_j^b. Intuitively, buyer entries add counter-parties for sellers while adding competitors for buyers. Second, we assume that $\theta > 0$ to capture the positive liquidity externality when more traders choose the same submarket. If, for example, both b_j and s_j double, then every trader on this trading venue would enjoy higher liquidity (i.e., π_j^s and π_j^b both increase).

Consistent with reality, buyers in our models face capacity constraints: although a buyer may participate in multiple submarkets, she can buy no more than one asset.[14] We assume that, if a buyer is matched on multiple trading venues, she randomly chooses one venue and forgoes trading opportunities on the other venues. Nature then chooses the buyer with probability ρ and the matched seller with probability $1-\rho$ to make a take-it-or-leave-it trading proposal to the other side, where $\rho \in (0,1)$ denotes buyers' bargaining power. A chosen trader then proposes the most profitable price that is acceptable to the other side. Thus, when AS trading is used, the

[13]We normalize the mass of every asset at 1, so $s_j = 1$ in every AS submarket. We assume for simplicity that buyers are divisible, so b_j is a non-negative real number.

[14]Without capacity constraints, buyers would participate in either all or none of the trading venues, which is unrealistic.

transaction price for asset j equals

$$P_{\text{as}}(v_j) = \begin{cases} v_j + \delta & \text{with probability } 1-\rho, \\ v_j & \text{with probability } \rho. \end{cases} \tag{3}$$

Consequently, in expectation a trade generates a profit of $\rho\delta$ for the involved buyer and a profit of $(1-\rho)\delta$ for the involved seller.

3.2 Equilibrium

Next, we describe traders' choices and the equilibrium. A seller has no choice but to sell on the AS submarket specific for her asset. Every buyer takes other traders' venue choices as given and chooses her trading venues to maximize expected profit. For simplicity, We impose the following assumption regarding c to obtain a benchmark with maximal market fragmentation,

Assumption 1. The cost to analyze an asset $c > \rho\delta/4$.

This assumption requires c to exceed 25% of a buyer's expected gain from a trade $\rho\delta$. We show n the appendix that Assumption 1 implies that the buying side is maximally fragmented: every buyer participates in *at most* one trading venue and some buyers may choose not to participate in any trading venue.[15] Given Assumption 1, the equilibrium is as follows.

Lemma 1 (**Equilibrium with only AS trading**)**.** *Suppose that Assumption 1 holds. Every AS submarket is chosen by one seller and b_0^*/s buyers, where*

$$b_0^* = \min\{b, \bar{b}_0\} \tag{4}$$

equals the total number of participating buyers and

$$\bar{b}_0 := s\left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}. \tag{5}$$

equals the market capacity for buyers. The number of buyers who do not participate in any AS submarket equals $\max\{b - \bar{b}_0, 0\}$.

[15]Specifically, in the appendix we let $c_{\min}$ denote the minimum cost of analysis at any venue. We show (1) in Lemma A.1 that if $c_{\min} > \rho\delta(s-1)^{s-1}/s^s$, then not every buyer participates in all venues, and (2) in Lemma A.2 that, if $c_{\min} > \rho\delta/4$, a buyer participates in at most one venue. In addition, if $c > \rho\delta$, then no buyer participates because buyers can never break even. Note that Assumption 1 is sufficient but not necessary for maximal buyer fragmentation. In particular, even if $c \le \rho\delta/4$, buyers can still be maximally fragmented when λ, s, b and θ take certain values.

On any AS submarket, the probabilities that a buyer and a seller trade equal, respectively,

$$\pi_0^b = \lambda\left(\frac{s}{b_0^*}\right)^{\frac{1-\theta}{2}} \quad \textit{and} \quad \pi_0^s = \lambda\left(\frac{b_0^*}{s}\right)^{\frac{1+\theta}{2}}. \tag{6}$$

The total expected trading volume equals

$$m_0 = \lambda s^{\frac{1-\theta}{2}}\left(b_0^*\right)^{\frac{1+\theta}{2}} \tag{7}$$

and the total expected profit for all traders equals

$$\Omega_0 = m_0\delta - cb_0^*. \tag{8}$$

The market is fragmented into multiple submarkets, each with a distinct asset and a disjoint set of buyers. Because in equilibrium every buyer earns the same expected profit, participating buyers must spread evenly across submarkets, so every submarket attracts the same number of buyers. In consequence, every seller experiences the same level of liquidity as well. In Fig. 1, for example, we illustrate an equilibrium in which the market is fragmented into 4 submarkets, each of which attracts 1 seller and 2 buyers.

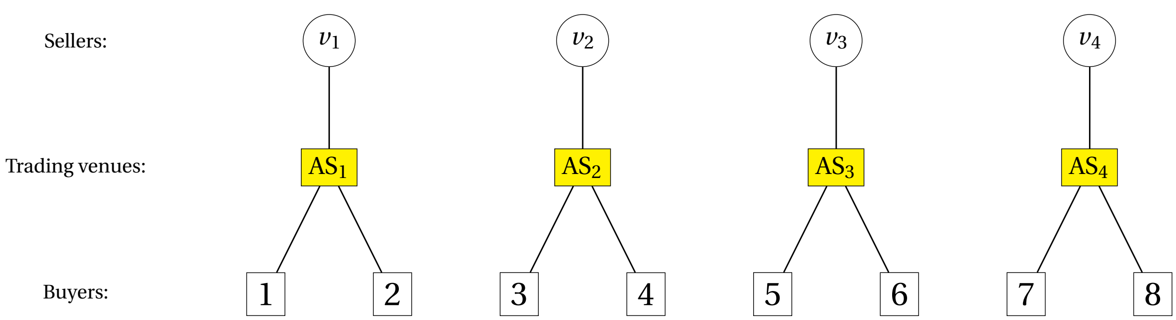


Figure 1. An Example Pure-AS Equilibrium with 4 AS Submarkets.

.

Note that, at most $\bar{b}_0$ buyers participate in trading. In particular, if more than $\bar{b}_0/s$ buyers choose an AS submarket, then these buyers would earn negative profits in expectation because the probability that a buyer trades on this AS submarket π_0^b would be less than $c/(\rho\delta)$. This cannot be an equilibrium because buyers can choose not to participate and earn zero profit. Thus, if the number of buyers b exceeds $\bar{b}_0$, then $b - \bar{b}_0$ buyers would choose not to participate in any AS submarket.

Crucially, Lemma 1 shows that, although pooling traders in the same venue would enhance liquidity (because $\theta > 0$), liquidity for traders in this pure-AS equilibrium may not improve as more traders join. Specifically, if the numbers of buyers and sellers b and s increase proportionally, then the buyer-to-seller ratio in every AS submarket b_0^* / s would remain constant because, according to Lemma 1,

$$\frac{b_0^*}{s} = \min\left\{\frac{b}{s}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\}. \tag{9}$$

Intuitively, if *proportionally* more buyers and sellers join, the market simply fragments into more AS submarkets, each of which still includes one seller and b^* / s buyers. Eq. (6) then implies that the liquidity levels π_0^b and π_0^s stay the same. Therefore, when assets are heterogeneous, a market with a higher number of traders is not necessarily more liquid because it can fragment into a higher number of submarkets.

3.3 Discussion on Model Setup

It is worth comparing our model to those developed in Vayanos and Wang (2007) and Vayanos and Weill (2008), who also endogenize market fragmentation. On the one hand, as in their models, the selling side in our model fragments because a seller can sell only her own asset and she cannot sell others' assets. On the other hand, unlike in their models, the buyer side in our model fragments because assets differ in values and it is more costly for a buyer to analyze more assets. This natural assumption enables us to examine how asset heterogeneity influences market liquidity, which lays the foundation for our investigation into the effects of QC trading in the following sections. In contrast, in the models developed by Vayanos and Wang (2007) and Vayanos and Weill (2008), such asset-specific analysis costs are absent because all assets share the same value; the buy side in their models fragments because of various portfolio constraints unrelated to asset values. This setup serves their purposes well because they focus on explaining why assets with identical values can trade at different prices across venues; nonetheless, it is not suited for examining the effects of asset heterogeneity.

4 Model with Both AS Trading and QC Trading

We now extend the baseline model to allow QC trading. We first describe how QC trading functions in our model and then derive the equilibrium when traders may choose between AS and QC trading.

4.1 Setup

Unlike standard AS contracts, QC contracts do not fully specify the deliverable assets. Instead, they specify some requirements the characteristics of deliverable assets must meet. For the sake of simplicity, we assume that such requirements translate to QC-eligibility threshold $\underline{v}$ such that assets are QC-eligible if and only if they are more valuable than $\underline{v}$.[16] The seller of a QC-eligible asset can choose between the QC market and an AS submarket, whereas the seller of a QC-ineligible asset must sell on an AS submarket.

The QC market is "quasi-consolidated" because although buyers receive heterogeneous assets, they agree to pay the same prices. Specifically, we assume that on the QC market the transaction price for a trade equals

$$P_{\text{qc}} = \begin{cases} \bar{v}_{\text{qc}} + \delta & \text{with probability } 1-\rho, \\ \bar{v}_{\text{qc}} & \text{with probability } \rho, \end{cases} \tag{10}$$

where

$$\bar{v}_{\text{qc}} = \mathbf{E}[v | v \in \text{QC}] \tag{11}$$

denote the expected (sellers') value of assets sold on the QC market. Consequently, as is the case in an AS submarket, a buyer expects to earn $\rho\delta$ and a seller expects to earn $(1-\rho)\delta$ from a trade.

While s and b are exogenous, traders choose their trading venues endogenously, which determines the total numbers of sellers s_{as} and buyers b_{as} on all AS submarket, and the numbers of sellers s_{qc} and buyers b_{qc} on the QC market. In total, there are $s_{\text{as}}+1$ trading venues that include s_{as} AS submarkets and one QC market. We assume that a buyer in the QC market also needs to spend c to estimate the average value of assets sold here $\bar{v}_{\text{qc}}$. We maintain Assumption 1 so that each buyer participates in at most one trading venue. In addition, because some buyers may choose not to participate in any trading venue, the total number of buyers who participate in all trading venues $b^* = b_{\text{as}} + b_{\text{qc}} \le b$. In contrast, because all sellers attempt to sell their assets, $s = s_{\text{as}} + s_{\text{qc}}$.

We assume that the QC market features the same search friction as in the AS market, represented by the matching function Eq. (1). Thus, the expected number of matches in the QC

[16] In Section 5 we examine regulators' choice of $\underline{v}$.

market equals

$$m_{\mathrm{qc}} = \lambda (s_{\mathrm{qc}} b_{\mathrm{qc}})^{\frac{1+\theta}{2}} \tag{12}$$

and the probabilities that a buyer and a seller trade, according to Eq. (2), equal, respectively,

$$\pi^b_{\mathrm{qc}} = \lambda \left(\frac{s_{\mathrm{qc}}}{b_{\mathrm{qc}}}\right)^{\frac{1-\theta}{2}} s^{\theta}_{\mathrm{qc}} \quad \text{and} \quad \pi^s_{\mathrm{qc}} = \lambda \left(\frac{b_{\mathrm{qc}}}{s_{\mathrm{qc}}}\right)^{\frac{1-\theta}{2}} b^{\theta}_{\mathrm{qc}}. \tag{13}$$

In contrast to the situation in the pure-AS equilibrium presented in Lemma 1, traders in the QC market do benefit from the positive network liquidity externality θ derived from pooling more traders in this trading venue. Specifically, Eq. (13) implies that proportionally increasing the numbers of buyers and sellers (s_{qc} and b_{qc}) would improve liquidity for all traders on the QC market (π^b_{qc} and π^s_{qc}).

4.2 Equilibrium

We make two simplifying assumptions to rule out degenerate equilibria in which only AS or only QC trading is employed. First, we assume that sellers prefer the QC market to an AS submarket when they are indifferent between the two. Thus, at least one asset is sold via QC trading ($s_{\mathrm{qc}} \geq 1$). Second, we assume that some assets are QC-ineligible ($\underline{v} > v_{\min}$) so that not all assets are sold on the QC market ($s_{\mathrm{qc}} < s$).

A seller takes equilibrium liquidity and price levels as given and chooses the venue to maximize her expected profits. Multiple equilibria may arise because prices in the QC market can be self-fulfilling. In particular, if sellers expect buyers to bid high prices in the QC market, sellers of higher-value assets would choose the QC market, which can in turn justify buyers' high bids. We focus on the equilibrium with the highest QC price because the total expected trading volume reaches the maximum in this equilibrium. For ease of exposition, we define a function

$$\mu(x) := x\left(x^{\frac{2\theta}{1-\theta}} - 1\right), \tag{14}$$

which is non-negative and increasing when $x \geq 1$. The equilibrium is as follows.

Theorem 1 (**Equilibrium with both AS and QC trading**). *Suppose that Assumption 1 holds. The* *equilibrium can be characterized by*

$$\bar{v} = \sigma(\underline{v}, F, s, \theta) := \max_{x \in \mathcal{V}} \left\{ x : x \geq \underline{v} \text{ and } x \leq \mathbf{E}[v | v \in [\underline{v}, x]] + \left(1 - \frac{1}{\left(s \Pr\left\{ v \in [\underline{v}, x] \right\} \right)^{\frac{2\theta}{1-\theta}}} \right) (1-\rho)\delta \right\}. \tag{15}$$

Denote the proportion of assets whose values fall in the interval $[\underline{v}, \bar{v}]$ *by*

$$q = \Pr\left\{ v \in [\underline{v}, \bar{v}] \right\}. \tag{16}$$

The following conditions obtain in equilibrium.

- *(**Buyers**) The number of buyers who participate in trading equals*

  $$b^* = \min\left\{ b, \bar{b} \right\}, \tag{17}$$

  where the market capacity for buyers equals

  $$\bar{b} = \left(s + \mu(sq) \right) \left(\frac{\lambda \rho \delta}{c} \right)^{\frac{2}{1-\theta}}. \tag{18}$$

  Each of the b^* *buyers participates in one trading venue. The other* $\max\left\{ b - \bar{b}, 0 \right\}$ *buyers do not participate in any trading venue. The number of buyers on the QC market and the total number of buyers on all AS submarkets equal, respectively,*

  $$b_{\text{qc}} = (sq)^{\frac{1+\theta}{1-\theta}} \gamma \quad \text{and} \quad b_{\text{as}} = s(1-q)\gamma, \tag{19}$$

  where

  $$\gamma = \frac{b^*}{s + \mu(sq)} = \min\left\{ \frac{b}{s + \mu(sq)}, \left(\frac{\lambda \rho \delta}{c} \right)^{\frac{2}{1-\theta}} \right\}. \tag{20}$$

  For a buyer who participates in any trading venue, the probability that she trades equals

  $$\pi^b = \frac{\lambda}{\gamma^{\frac{1-\theta}{2}}}. \tag{21}$$

  Buyers on the QC market bid $\bar{v}_{\text{qc}} = \mathbf{E}\left[v | v \in [\underline{v}, \bar{v}] \right]$ *and accept the ask of* $\bar{v}_{\text{qc}} + \delta$*. Buyers' total* *profit equals*

$$\Omega^b = b^* \left(\pi^b \rho\delta - c\right). \tag{22}$$

- *(**Sellers**) Sellers follow a threshold strategy in choosing the venue: a seller chooses the QC market if her asset's value $v \in [\underline{v}, \bar{v}]$ and an AS submarket otherwise. The number of sellers on the QC market and the total number of sellers on all AS submarkets equal, respectively,*

$$s_{\text{qc}} = sq \quad \textit{and} \quad s_{\text{as}} = s(1-q). \tag{23}$$

The probabilities that a seller trades successfully on the QC market and an AS submarket equal, respectively,

$$\pi^s_{\text{qc}} = \lambda\gamma^{\frac{1+\theta}{2}} s_{\text{qc}}^{\frac{2\theta}{1-\theta}} \quad \textit{and} \quad \pi^s_{\text{as}} = \lambda\gamma^{\frac{1+\theta}{2}}. \tag{24}$$

Sellers' total profit equals

$$\Omega^s = m(1-\rho)\delta, \tag{25}$$

where the total expected number of trades equals

$$m = \lambda(b^*)^{\frac{1+\theta}{2}} \left(s + \mu(sq)\right)^{\frac{1-\theta}{2}}. \tag{26}$$

- *(**Overall**) The total expected profit for all traders equals*

$$\Omega = \Omega^b + \Omega^s = m\left(\delta - \frac{c}{\pi^b}\right) = m\delta - cb^*. \tag{27}$$

A single endogenous parameter $\bar{v}$—the upper bound of the values of assets traded on the QC market—fully characterizes the equilibrium. Sellers choose the QC market only when they own medium-value assets ($v \in [\underline{v}, \bar{v}]$). Sellers of low-value assets ($v < \underline{v}$) choose AS trading because their assets are QC-ineligible. Sellers of high-value assets ($v > \bar{v}$) continue voluntarily to use AS trading despite its illiquidity because their assets are much more valuable than the uniform price in the QC market P_{qc}.[17] Fig. 2 illustrates an example equilibrium in which 4 sellers and 8

[17] In Guerrieri and Shimer (2014), because buyers do not observe asset quality, sellers accept a lower trading probability to signal the higher quality of their assets. In contrast, buyers in our model do observe asset quality in

buyers use both AS and QC contracts to trade. Note that, although traders are still fragmented, multiple assets are sold together on the QC market, so the degree of fragmentation is reduced relative to the pure-AS equilibrium in which every asset is sold on a separate AS submarket.

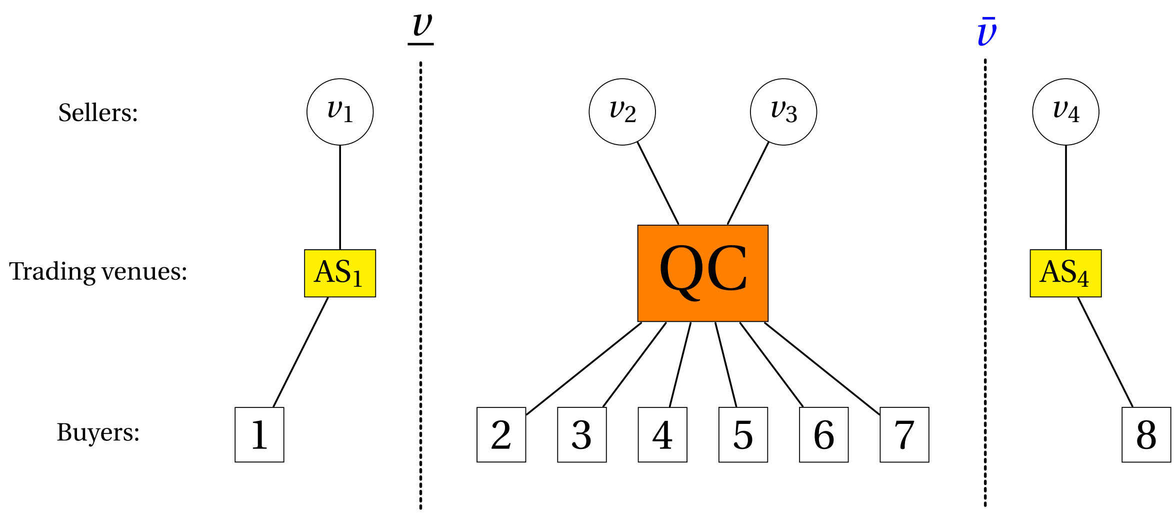


Figure 2. An Example Market with AS and QC Trading.

.

As discussed immediately before Theorem 1, multiple equilibria may exist. By imposing Eq. (15), we choose the equilibrium with the highest threshold $\bar{v}$, which results in the greatest q, the proportion of sellers choosing the QC market. Because both b^* and $\mu(sq)$ weakly increase with q, in this equilibrium the trading volume m reaches the maximum according to Eq. (26). Next, we discuss the effects of introducing QC trading to a pure-AS market based on Lemma 1 and Theorem 1.

4.3 Impact on Liquidity of Introducing QC Trading

Although QC trading reduces trading frictions, introducing QC trading results in varying effects on the level of liquidity a trader experiences. Specifically, all buyers enjoy better liquidity, whereas only some sellers do. In particular, sellers who remain on AS submarkets experience *worse* liquidity. The overall market liquidity improves because the total expected trading volume, which in our model is proportional to the *average* probability an asset is sold, increases.

We formally summarize the liquidity effects of introducing QC trading in Corollary 1 and then illustrate these effects in Fig. 3. Recall that, in the pure-AS equilibrium, at most $\bar{b}_0$ buy-

the AS market; sellers of high-value assets avoid the QC market because of its uniform pricing scheme.

ers participate in trading and b_0^* buyers actually participate in trading; π_0^s and π_0^b (defined in Eq. (6)) denote, respectively, the probabilities that a seller and a buyer trade; and the total expected trading volume equals m_0 (given in Eq. (7)). In the parallel-trading equilibrium, the corresponding variables are denoted by $\bar{b}$, b^*, π^s, π^b, and m.

Corollary 1 (Impact on liquidity). *In the parallel-trading equilibrium, the buyer-to-seller ratios satisfy*

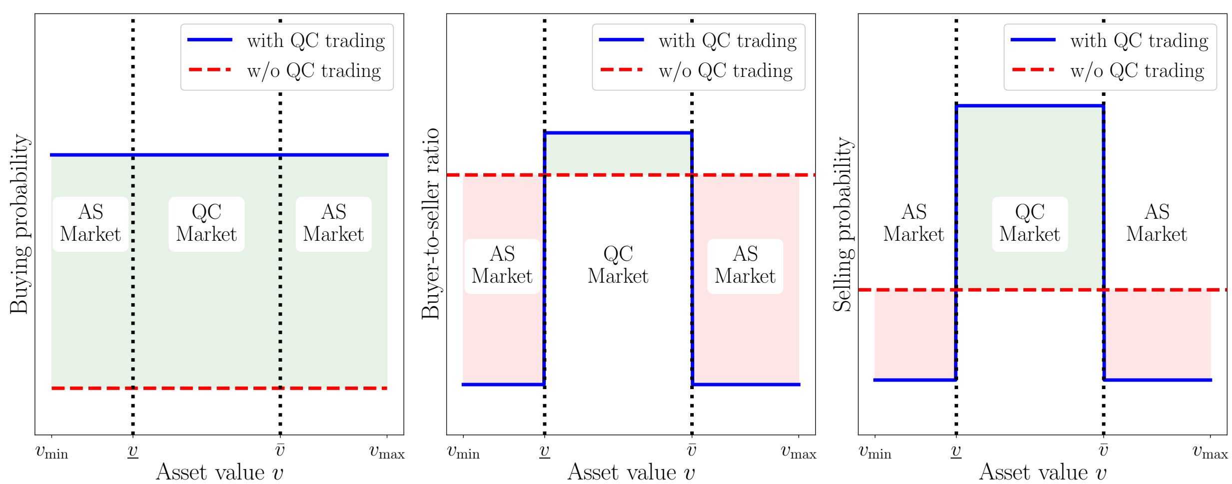
$$\frac{b_{\mathrm{as}}}{s_{\mathrm{as}}} \leq \frac{b_0^*}{s} \leq \frac{b_{\mathrm{qc}}}{s_{\mathrm{qc}}}. \tag{28}$$

and the liquidity levels satisfy

$$\pi_0^b \leq \pi^b \quad \text{and} \quad \pi_{\mathrm{as}}^s \leq \pi_0^s \leq \pi_{\mathrm{qc}}^s. \tag{29}$$

Introducing QC trading weakly expands the market capacity for buyers ($\bar{b} \geq \bar{b}_0$) and enables weakly more buyers to participate in trading ($b^* \geq b_0^*$). The total trading volume

$$m \geq m_0. \tag{30}$$



(a) Liquidity for buyers **(b)** Buyer-to-seller ratio **(c)** Liquidity for sellers

Figure 3. Impact of introducing QC trading when the QC market is non-trivial ($s_{\mathrm{qc}} > 1$).

.

As Fig. 3a illustrates, all buyers experience weakly better liquidity ($\pi^b \geq \pi_0^b$). Buyers who migrate to the QC market experience better liquidity because of reduced trading frictions. Buyers who remain on AS submarkets also experience better liquidity because QC trading attracts *disproportionately more* buyers than sellers to migrate from AS submarkets to the QC market, which, as Fig. 3b illustrates, raises the buyer-to-seller ratio on the QC market ($b_{\mathrm{qc}}/s_{\mathrm{qc}} \uparrow$) but lowers the ratio on AS submarkets ($b_{\mathrm{as}}/s_{\mathrm{as}} \downarrow$). In equilibrium, the higher buyer-to-seller ratio on the QC market exactly offsets the reduced search friction on this market, so buyers are indifferent between the QC market and AS submarkets. To understand why buyers respond more strongly than sellers to the introduction of QC trading, note that, migrating to the QC market is free for buyers, but is costly for sellers of high-value assets because all assets are priced uniformly on the QC market. Thus, disproportionately more buyers than sellers migrate to QC trading. Overall, although only some buyers choose QC trading, all buyers experience better liquidity because the migration of buyers effectively redistributes some liquidity gains from QC buyers to AS buyers.

Moreover, the enhanced liquidity for buyers could attract additional buyers to participate in trading, which would further improve liquidity for sellers but could partially reverse the liquidity gains for buyers. Specifically, in the pure-AS equilibrium the market reaches capacity for buyers when a buyer's expected profit $\pi_0^b\rho\delta - c$ equals zero. Introducing QC trading increases the probability that participating buyers trade ($\pi_0 > \pi_0^b$), enabling them to earn positive expected profits; thus, if additional buyers are present, they will enter and participate in trading. Because in equilibrium all participating buyers experience the same level of liquidity, the additional buyers would enter all trading venues, raising the buyer-to-seller ratios on all venues. Note that, buyers' entry would never make them experience *worse* liquidity than they do in the pure-AS equilibrium. The reason is that additional buyers could enter only if in the pure-AS equilibrium they earn zero expected profits and choose not to participate in trading, which implies that $\pi_0^b = c/(\rho\delta)$. In the parallel-trading equilibrium, a participating buyer's expected profit $\pi^b\rho\delta - c$ would never be negative, which implies that the probability that she trades π^b cannot be lower than $c/(\rho\delta) = \pi_0^b$.

Next, we discuss how liquidity effect of introducing QC trading on sellers. First, as Fig. 3c illustrates, liquidity improves for sellers who choose the QC market. These sellers benefit from three channels: (1) QC trading reduces trading frictions ($\theta > 0$); (2) migration of a disproportionately higher number of buyers to the QC market increases the buyer-to-seller ratio on this market $b_{\mathrm{qc}}/s_{\mathrm{qc}}$; and (3) the entry of additional buyers further increases the buyer-to-seller ratio $b_{\mathrm{qc}}/s_{\mathrm{qc}}$.

Second, as Fig. 3c illustrates, liquidity deteriorates for sellers who stay on AS submarket. These sellers continue to using AS trading because either their assets are not eligible for QC trading or prices on the QC market are far below their asset values. In particular, because buyers on AS submarkets experience better liquidity, the buyer-to-seller ratio on AS submarket must decline ($b_{as}/s_{as} \leq b_0^*/s$), which implies that AS sellers must experience worse liquidity. In other words, AS sellers experience worse liquidity because the positive effect of increased buyer participation cannot overcome the negative effect of buyers' migration.

Finally, introducing QC trading increases the total trading volume ($m \geq m_0$) because (1) more buyers participate ($b^* \geq b_0^*$), (2) every participating buyer is more likely to trade ($\pi^b \geq \pi_0^b$), and (3) the trading volume equals the product of the number of participating buyers and the probability that each participating buyer trades ($m = b^*\pi^b$ and $m_0 = b_0^*\pi_0^b$). Intuitively, the total trading volume increases because overall trading frictions are reduced.

Table 2. Liquidity effects of introducing QC trading on traders

	QC sellers	AS sellers	QC buyers	AS buyers
Lower frictions of QC trading	+	NA	+	NA
Buyer entry	+	+	−	−
Migration from AS market to QC market	+	−	−	+
Overall	+	−	+	+

In this table we summarize how introducing QC trading affects traders' liquidity levels. A "+" symbol means the effect weakly improves liquidity; a "−" symbol means the effect weakly hurts liquidity; "NA" means the effect has no liquidity impact.

We summarize liquidity effects of introducing QC trading in Table 2. Overall, introducing QC trading improves liquidity for all buyers and QC sellers, but hurts liquidity for AS sellers. Because buyers are ex ante identical and can move costlessly between all trading venues, the reduced trading frictions in the QC market not only directly benefit QC buyers, but also indirectly benefit AS buyers because the migration of buyers to the QC market leaves fewer competing buyers in every AS submarket. Thus, for buyers, liquidity spills over from the QC market to AS submarkets. In contrast, introducing QC trading does not improve liquidity for all sellers. In particular, it is more costly for sellers of higher-value assets to switch to the QC market because they would receive the same price for selling more valuable assets. In consequence, sellers of high-value QC-eligible assets remain in AS submarkets but experience worse liquidity. Thus, for sellers, the QC market siphons liquidity off AS submarkets. Without asset heterogeneity, all sellers would switch to the QC market and enjoy better liquidity. Overall, despite hurting AS sellers, QC trading always increases the total trading volume.

4.4 Impact on Trader Profits of Introducing QC Trading

Introducing QC trading also imposes varying effects on traders' profits, for two reasons. First, a trader's expected profit depends on liquidity she experiences and, as we show in Section 4.3, QC trading affects traders' liquidity non-uniformly. Second, in the QC market sellers receive the same prices for delivering heterogeneous assets, so they earn different profits. We show in this section that all buyers earn higher profits and all AS sellers earn lower profits, whereas only some QC sellers earn higher profits. Overall, social welfare, which equals the total profit of all traders, increase.

We denote by ψ_0^b and ψ_0^s, respectively, the expected profits a buyer and a seller earn in the pure-AS equilibrium. For the parallel-trading equilibrium, we denote by ψ^b, ψ_{as}^s, and $\psi_{\text{qc}}^s(v)$, respectively, the expected profits a buyer in any market earns, a seller in the AS market earns, and a seller in the QC market earns. The expected profit a QC seller earns $\psi_{\text{qc}}^s(v)$ is a function of her asset value v because in the QC market all assets are priced the same. In addition, social welfare, which equals all traders' expected profits, is denoted by Ω_0 (given in Eq. (8)) in the pure-AS equilibrium and Ω (given in Eq. (27)) in the parallel-trading equilibrium. In Corollary 2, we summarize the effects of introducing QC trading on traders' profits.

Corollary 2 (**Impact on trader profits**). *A buyer earns a higher profit ($\psi^b \geq \psi_0^b$); a seller in the AS market earns a lower profit ($\psi_{\text{as}}^s \leq \psi_0^s$); a seller in the QC market earns a higher profit if her asset's value $v < v^*$ and a lower profit if $v > v^*$, where*

$$v^* := \mathbf{E}[v | v \in [\underline{v}, \bar{v}]] + (1-\rho)\delta\left(1 - \frac{\pi_0^s}{\pi_{\text{qc}}^s}\right). \tag{31}$$

In particular, if $\bar{v} < v_{\max}$, then $v^* \leq \bar{v}$ and some QC sellers earn lower profits. Social welfare increases ($\Omega \geq \Omega_0$).

Every participating buyer's profit increases because she spends the same cost c to analyze an asset and, as Corollary 1 shows, is more likely to trade ($\pi^b \geq \pi_0^b$). Every seller who remains on an AS submarket earns a lower profit because she obtains the same selling prices (v_j or $v_j + \delta$) but is less likely to trade.

The effect on QC sellers' profits varies across sellers. One the QC market, all assets obtain the same prices $\bar{v}_{\text{qc}}$ or $\bar{v}_{\text{qc}} + \delta$, where $\bar{v}_{\text{qc}}$ equals the average sellers' value of all assets sold on the QC market. It follows that a QC seller's expected profit equals

$$\psi_{\text{qc}}^s(v) = \pi_{\text{qc}}^s\left(\rho(\bar{v}_{\text{qc}} - v) + (1-\rho)(\bar{v}_{\text{qc}} + \delta - v)\right) = \pi_{\text{qc}}^s\left((1-\rho)\delta + \bar{v}_{\text{qc}} - v\right), \tag{32}$$

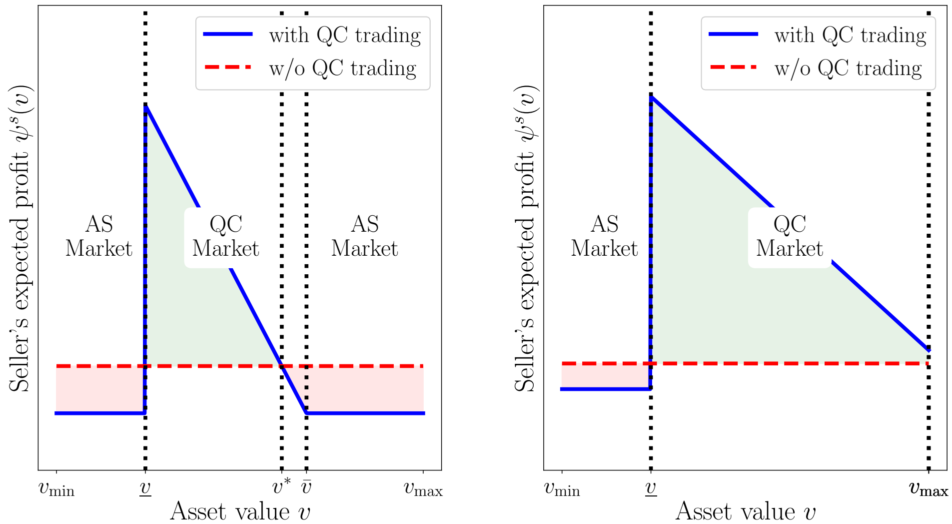
which decreases with asset value v.

Although QC sellers enjoy higher liquidity ($\pi^s_{\text{qc}} \geq \pi^s_0$), they suffer price-discount $v - \bar{v}_{\text{qc}}$ if their assets are more valuable than $\bar{v}_{\text{qc}}$. If the value of an asset v exceeds v^*, the price-discount cost dominates the liquidity benefit so that the seller earns a lower profit than what she would have earned in a pure-AS market ($\psi^s_{\text{qc}}(v) < \psi^s_0$).

If the most valuable asset is sold on the AS market (i.e., $\bar{v} < v_{\max}$), then, as Fig. 4a illustrates, QC sellers earn lower profits if their asset values fall in the interval $(v^*, \bar{v}]$. To understand this result, note that $\bar{v} < v_{\max}$ implies that the marginal seller, whose asset is worth $\bar{v}$, is indifferent between the AS and the QC market. Thus, her expected profit $\psi^s_{\text{qc}}(\bar{v})$ equals the expected profit she would have earned by selling on the AS market ψ^s_{as}. Because $\psi^s_{\text{as}} \leq \psi^s_0$, this seller earns a lower profit than she could haven earned in the pure-AS equilibrium. Intuitively, the marginal seller is worse off because introducing QC trading reduces the value of her outside option to sell on the AS market. Thus, if the most valuable asset is not sold through QC trading, then some sellers who choose QC trading would rather ban QC trading: although they experience better liquidity, their assets are so under-priced that they earn lower profits. It is possible, as Fig. 4b illustrates, that *all* QC sellers earn higher profits if even the most valuable asset is sold on the QC market (i.e., $\bar{v} = v_{\max} < v^*$). This situation, however, does not arise in markets involving MBS because, as documented by An, Li and Song (2022), the most valuable MBS are sold through SP contracts, which are not QC contracts but AS contracts. Thus, our results suggest that some issuers who sell through TBA contracts could actually benefit from banning the TBA market.

Although some sellers earn lower profits, the welfare gain is greater because buyers' total profits Ω^b and sellers' total profits Ω^s both increase. On the one hand, buyers' total profits $\Omega^b = b^*(\pi^b \rho\delta - c)$. The profits increase because more buyers participate ($b^* \uparrow$), every participating buyer is more likely to trade ($\pi^b \uparrow$), and every participating buyer pays the same cost c for analysis. On the other hand, sellers' total profits increase because the total trading volume increases and sellers *on average* earn $(1-\rho)\delta$ from every trade. Specifically, every trade in the AS market generates the same profit of $(1-\rho)\delta$ for the involved seller. A trade in the QC market generates a profit of $(1-\rho)\delta + \bar{v}_{\text{qc}} - v$, where $\bar{v}_{\text{qc}} - v$ represents the cross-subsidy across sellers on the QC market. Because $\bar{v}_{\text{qc}} = \mathbf{E}[v | v \in [\underline{v}, \bar{v}]]$, the total cross-subsidy across QC sellers equals zero. As a result, the *average* sellers' profit from a trade in the QC market also equals $(1-\rho)\delta$. Therefore, sellers' total profits $\Omega^s = m(1-\rho)\delta$ increase proportionally with the trading volume m.

In practice, it would be hard for traders to organize QC markets voluntarily. Although introducing QC trading always increases the total profits for all traders, sellers would not be able



(a) $v^* < \bar{v} < v_{\max}$

(b) $\bar{v} = v_{\max} < v^*$

Figure 4. Impact of introducing QC trading on individual seller's profit.

to agree unanimously on a QC-eligible threshold $\underline{v}$ because QC trading forces some sellers to subsidize others. In particular, as Fig. 4 illustrates, at $v = \underline{v}$, the impact of QC trading on sellers' profits jump from negative to the highest positive value. Thus, every seller would want to set the QC-eligible threshold $\underline{v}$ at the value of her asset v. Sellers whose asset values fall just below $\underline{v}$, however, would strongly oppose such a threshold because their assets would be QC-ineligible and harder to sell.

4.5 Policy Implications of Introducing QC Trading

Next, we briefly summarize policy implications based on results we report in Corollaries 1 and 2. We list the effects of introducing one QC contract to a pure-AS market in Table 3.

Table 3. Effects of introducing QC trading on traders' liqudity and profits

	Overall	Buyers	AS sellers	QC Sellers	
				asset value $v < v^*$	asset value $v > v^*$
Liquidity	+	+	−	+	+
Profit	+	+	−	+	−

The table summarizes how introducing QC trading affect traders' liquidity and profits. A "+" symbol represents a weakly positive effect and a "−" symbol represents a weakly negative effect.

First, despite the CTD issue, QC trading always weakly increases overall liquidity (measured in total trading volume) and social welfare. In particular, because QC trading mitigates trading frictions, it improves liquidity for all participating buyers and attracts more buyers to participate. In consequence, total trading volume increases. Moreover, buyers' total profits and sellers' total profits both increase because the total trading volume increases and the total cross-subsidy across sellers equals zero.

Second, traders are unlikely to organize QC markets voluntarily. In particular, introducing QC trading is generally not Pareto-improving because it forces some traders to cross-subsidize others. Specifically, introducing QC trading generally (1) hurts liquidity for AS sellers and (2) reduces profits that AS sellers and some QC sellers earn. Therefore, regulators should consider the impact of QC trading on *all* traders rather than focusing only on traders in the QC market. The only exception occurs when the number of buyers b is so large that it exceeds capacity $\bar{b}$ even after QC trading is introduced. In this situation, additional buyers effectively enter only the QC market until QC buyers earn zero profit, so QC sellers benefit whereas AS traders are unaffected.

5 QC Contract Design

Thus far we have discussed the effects of introducing one exogenously specified QC contract to a pure-AS market. In this section, we explore how regulators can design QC contracts and explain the effects of introducing multiple QC contracts. We also discuss the policy implications of QC design.

5.1 Choice of QC-eligibility Threshold

In practice, regulators can specify various criteria for determining the characteristics of deliverable securities. To represent such choices parsimoniously, we assume that regulators can directly choose the QC-eligibility threshold $\underline{v}$ and explore the impact of changing the eligibility threshold $\underline{v}$.

The optimal choice of $\bar{v}$ in general has no closed-form solution because it depends on various parameters, including the distribution of asset value F. Nonetheless, we show in the following result that regulators can maximize social welfare by choosing the $\bar{v}$ that maximizes the proportion of assets traded in the QC market q.

Corollary 3 (**Impact of $\underline{v}$**). *Suppose that changing $\underline{v}$ expands the coverage of QC market to a larger proportion of assets ($q \uparrow$). Then, $\bar{b}$, b^*, π^b, m, and Ω weakly increase, whereas π^s_{as} weakly declines. In particular, if $b \geq \bar{b}$ after q increases, then π^b and π^s_{as} are unaffected.*

As Theorem 1 shows, varying $\underline{v}$ can influence the upper bound of QC assets' values $\bar{v}$, which can in turn affects the proportion of assets sold in the QC market q. Note that increasing $\underline{v}$ may increase or decrease the coverage of the QC market q. On the one hand, increasing $\underline{v}$ mitigates adverse selection because it excludes some low-value assets from the QC market. Hence, QC buyers would bid higher prices, motivating sellers of higher-value assets to choose the QC market, which increases q. On the other hand, increasing $\underline{v}$ could reduce the liquidity benefits for QC buyers because it may reduce the number of assets sold in the QC market. When, for example, $\bar{v} = v_{\max}$, then $\bar{v}$ cannot increase further and increasing $\underline{v}$ could only reduce q.

Corollary 3 shows that regulators can maximize social welfare Ω by choosing the threshold $\underline{v}$ that maximizes the coverage of the QC market q. In this situation, trading friction in the QC market is minimized, which maximizes liquidity for buyers and attracts the highest number of buyers to participate. In consequence, trading volume m and social welfare Ω are maximized.

Nonetheless, increasing q is generally not Pareto-improving because it hurts sellers who remain on AS submarkets. In fact, these sellers suffer the most severe losses when q is maximized. The only exception arises when there are so many buyers that the market capacity remains binding for buyers after QC trading is introduced ($b \geq \bar{b}$). In this situation, the entry of new buyers keeps the buyer-to-seller ratio on AS submarkets at the same level, so AS sellers experience the same level of liquidity.

5.2 Multiple QC Contracts

In Section 5.1, we investigate the trade-offs that occur when regulators choose the QC-eligibility threshold $\underline{v}$ of the only QC contract. In the agency MBS market, multiple TBA contracts with varying eligibility criteria are traded simultaneously. In this section, we examine the effects of introducing or consolidating multiple QC contracts.

We assume that assets can be categorized into T types indexed by $t \in \vec{T} := \{1, 2, \cdots, T\}$ and each QC contract accepts the delivery of only one type of asset. There are s_t type-t assets whose values follow the distribution F_t. Thus, there are $s = \sum_{t \in \vec{T}} s_t$ assets in total, whose values follow the mixed distribution $F = \frac{1}{s} \sum_{t \in \vec{T}} s_t F_t$.

Let $\underline{v}_{t,k}$ represent the eligibility threshold of the k-th QC contract for type-t assets, where $k \in \vec{K}_t := \{1, 2, \cdots, K_t\}$. This QC contract accepts the delivery of type-t assets whose values exceed

$\underline{v}_{t,k}$. Other types of assets, regardless how valuable they are, cannot be delivered to settle this contract. As in the main model, we assume that a buyer pays $c > \rho\delta/4$ to join a trading venue, so in equilibrium each buyer joins at most one trading venue and earns the same expected profit. Sellers choose trading venues to maximize their expected profits. In addition, if a seller could obtain the same expected profit by choosing one of several QC contracts, she prefers the contract with the highest eligibility threshold.

QC contracts can be "horizontally" or "vertically" differentiated. Specifically, QC contracts are horizontally differentiated if their deliverable assets differ in characteristics unrelated to asset value; QC contracts are vertically differentiated if their deliverable assets differ in value. To gain insights, we examine the impact of introducing *purely* horizontally and *purely* vertically differentiated QC contracts.

First, QC contracts are *purely* horizontally differentiated, if for every type of asset, the asset values follow the same value distribution ($F_t = F$) and one QC contract ($K_t = 1$) with the same eligibility threshold is available ($\underline{v}_{t,k} = \underline{v}$). In this situation, assets of distinct types only differ nominally—each type of assets have the same value distribution; nonetheless, they are never traded through the same QC contract. Thus, horizontal differentiation effectively separates one type of asset randomly into sub-types based on non-value characteristics and introduce a separate QC contract for each sub-type of asset. The effect of adding purely horizontally differentiated QC contracts is as follows.

Proposition 1 (**Impacts of horizontal differentiation**). *Suppose that $\underline{v}_{t,k} = \underline{v}$ and $F_t = F$ for any asset type $t \in \vec{T}$. If we separate type-i assets randomly into type-i' and type-i'' assets, then the coverage of QC trading shrinks, π^s_{as} increase, and π^b, m, and Ω all decrease.*

Proposition 1 shows that adding purely horizontally differentiated QC contracts *shrinks* the coverage of QC trading and *reduces* the total trading volume. Because a QC contract allows the delivery of only one type of asset, such horizontal differentiation fragments the QC market, which reduces the liquidity benefits of QC trading. In response, some buyers and sellers switch from QC trading to AS trading. In consequence, overall trading volume declines and every buyer experiences worse liquidity, which implies that AS sellers enjoy better liquidity. Similar effects are reported in the literature because in those models all assets share the same value and market fragmentation is purely horizontal.

Second, QC contracts are *purely* vertically differentiated if all assets belong to the same type ($T = 1$) and multiple QC contracts ($K > 1$) with varying eligibility thresholds are available. In this situation, assets traded through distinct QC contracts must differ in value.

Proposition 2 (**Impacts of vertical differentiation**). *Suppose that we add a QC contract for type-t assets with threshold $\underline{v}_{t,k+1}$ such that the coverage of this contract does not overlap with that of any existing QC contract. That is, $[\underline{v}_{t,k+1}, \sigma(\underline{v}_{t,k+1}, F_t, s_t, \theta)] \cap [\underline{v}_{t,k}, \bar{v}_{t,k}] = \emptyset$ for any $k \in \vec{K}_t$.*

Then, $\bar{v}_{t,k+1} = \sigma(\underline{v}_{t,k+1}, F_t, s_t, \theta)$. For any existing QC contract, $\bar{v}_{t,k}$ stays the same. In consequence, both π^s_{as} and $\pi^s_{\text{qc},t,k}$ decline, whereas π^b, m, and Ω all increase.

In the agency MBS market, for example, TBA contracts with differing interest rates are traded. Such differentiation is almost purely vertical because interest rates are strongly correlated with MBS values. Results reported in Proposition 2 show that adding purely vertically differentiated QC contracts expands the coverage of QC trading to additional assets.[18] In consequence, the total trading volume increases and buyers' liquidity improves. Nevertheless, because the new QC contract attracts buyers away from AS trading and existing QC contracts, all sellers, except those who use the new QC contract, experience worse liquidity.

Results reported in Propositions 1 and 2 can help regulators understand the effects of adding or consolidating QC contracts. To increase total trading volume and social welfare, they can, for example, keep adding purely vertically differentiated QC contracts to cover additional assets. Nevertheless, such interventions are in general not Pareto-improving because they (weakly) hurt sellers who remain in the AS market as well as sellers of existing QC contracts. Therefore, instead of focusing solely on trading volume or profits in QC markets, regulators need to quantify the effects on all traders.

Another practical concern can complicate the evaluation of a policy intervention: in general, QC contracts are neither *purely* horizontally nor *purely* vertically differentiated. We discuss the Uniform Mortgage Backed Security (UMBS) reform in the MBS market as an example.

Until recently, TBA contracts for MBS specify the agency that guarantees an MBS. Thus, MBS guaranteed by Freddie Mac, no matter how valuable they are, could not be delivered to TBA contracts for Fannie Mae-guaranteed MBS. In 2019, the FHFA replaced guarantor-specific TBA contracts with TBA contracts of the UMBS that do not limit the guarantor. The purpose of this reform, according to the FHFA, is to "reduce the costs to Freddie Mac and taxpayers that come from the persistent difference in the liquidity of Fannie Mae MBS and Freddie Mac MBS" and the FHFA projected that "use of the UMBS will save $400 million to $600 million per year."[19]

In the context of our model, this reform effectively consolidates two types of assets into one and integrates two type-specific QC contracts. Although Fannie Mae MBS are *on average* more

[18] If we add a QC contract whose coverage overlaps with some existing QC contracts, then the new contract is not "purely" vertically differentiated and the coverage of an existing QC contract may change. In practice, QC contracts for the same type of asset usually cover disjoint sets of assets.

[19] See the article at https://bit.ly/2Ir1jWx for details.

valuable than Freddie Mac MBS, some Fannie Mae MBS are less valuable than certain Freddie Mac MBS. Thus, this integration of TBA markets is neither *purely* horizontal nor *purely* vertical. The effects of the UMBS reform is therefore unclear ex ante. Overall market quality (measured by total trading volume and social welfare) would improve if Fannie Mae and Freddie Mac MBS have similar value distributions, but would deteriorate if Fannie Mae and Freddie Mac MBS differ substantially. Moreover, if the UMBS reform improves overall market quality, it generally hurts sellers who remain in the AS market. A comprehensive empirical analysis is therefore necessary to quantify these effects.

6 Conclusion

Substantial asset heterogeneity is a salient feature of markets for most fixed-income securities—including corporate bonds, municipal bonds, MBS, and asset-backed securities—and it is widely believed to hurt market liquidity. Motivated by the extraordinary liquidity of the TBA MBS market despite asset heterogeneity, some academics have recommended introducing similar cohort-based quasi-consolidated trading to other illiquid OTC markets for heterogeneous assets. Policymakers have further consolidated the TBA MBS market by integrating Fannie Mae and Freddie Mac TBA contracts.

We build a model that we use to study the impact of quasi-consolidated trading design—under which any eligible asset within a cohort can be delivered—in markets with search frictions for assets with heterogeneous fundamental values. We show that sellers of high-value assets voluntarily stay out of the QC market despite its superior liquidity. Thus, the extent to which QC trading can improve liquidity is limited by the degree of asset heterogeneity. Further, although QC trading always increases total trading volume and aggregate profit, it is not Pareto-improving because sellers who stay out of the QC market always suffer, so QC trading is unlikely to emerge without policy interventions. Regulators need to comprehensively evaluate the benefits and costs of QC trading for all traders.

Appendices

A Proofs

Lemma A.1 (**Conditions for buyer fragmentation**). *Consider a pure-AS market with s assets and let c_k denote a buyer's cost to participate in submarket k. It is an equilibrium for every buyer to participate in all s submarkets if and only if*

$$\max_{k\in\{1,\cdots,s\}}\left\{\frac{c_k}{\rho\delta}\right\} \le \frac{\lambda}{b^{\frac{1-\theta}{2}}}\left(1-\frac{\lambda}{b^{\frac{1-\theta}{2}}}\right)^{s-1}. \tag{A.33}$$

In particular, it is not an equilibrium for all buyers to participate in all submarkets if

$$\frac{c_k}{\rho\delta} > \frac{(s-1)^{s-1}}{s^s} \tag{A.34}$$

for any submarket k.

Proof of Lemma A.1. If every buyer participates in all s submarkets, then $\pi_k^b = \pi^b = \lambda(\frac{1}{b})^{\frac{1-\theta}{2}}$ in every trading venue k. Without loss of generality, suppose that all buyers but one participate in all submarkets. If the buyer participates in $I \le s-1$ submarkets indexed by $1,\cdots,I$, she would earn expected profits of

$$\phi = \rho\delta\left(1-(1-\pi^b)^I\right) - \sum_{i=1}^{I} c_i. \tag{A.35}$$

If the buyer also participates in submarket $I+1$, she would earn

$$\phi' = \rho\delta\left(1-(1-\pi^b)^{I+1}\right) - \sum_{i=1}^{I} c_i - c_{I+1}. \tag{A.36}$$

It follows that

$$\frac{\phi'-\phi}{\rho\delta} = \pi^b(1-\pi^b)^I - c_{I+1}. \tag{A.37}$$

On the one hand, if Eq. (A.33) holds, then

$$c_{I+1} \le \pi^b(1-\pi^b)^{s-1} \le \pi^b(1-\pi^b)^I, \tag{A.38}$$

so $\phi' \ge \phi$ and this buyer would always benefit from participating in an additional submarket. Thus, it is an equilibrium for the buyer to participate in all submarkets.

On the other hand, if Eq. (A.33) does not hold, then there exists submarket k such that $c_k > \pi^b(1-\pi^b)^{s-1}$. It cannot be an equilibrium for a buyer to participate in all submarkets because Eq. (A.37) implies that the buyer could earn strictly more profits by participating in all submarkets but submarket k. In particular, we know that

$$\pi^b(1-\pi^b)^{s-1} \leq \frac{(s-1)^{s-1}}{s^s} \tag{A.39}$$

and the inequality binds when $\pi^b = 1/s$. Thus, if $c_j > \frac{(s-1)^{s-1}}{s^s}\rho\delta$, then Eq. (A.33) does not hold and buyers fragment. □

When the number of assets s increases, buyers are more likely to fragment because the right-hand side of Eq. (A.33) $\left(1-\frac{\lambda}{b^{\frac{1-\theta}{2}}}\right)^{s-1}\frac{\lambda}{b^{\frac{1-\theta}{2}}}$ decreases with s.

Lemma A.2 (**Conditions for maximal buyer fragmentation**). *Let c_j denote a buyer's cost to participate in trading venue j. If $c_j > \rho\delta/4$ for any trading venue j, then a buyer participates in at most one trading venue.*

Proof of Lemma A.2. We prove by contradiction. Suppose, without loss of generality, that a buyer participates in multiple venues in the set $\mathcal{I} = \{1,\cdots,I\}$. The buyer earns expected profit of

$$\phi = \rho\delta\left(1-(1-\pi_1^b)\prod_{i=2}^{I}(1-\pi_i^b)\right) - c_1 - \sum_{i=2}^{I}c_i. \tag{A.40}$$

If the buyer quits trading venue 1, she would earn expected profit of

$$\phi' = \rho\delta\left(1-\prod_{i=2}^{I}(1-\pi_i^b)\right) - \sum_{i=2}^{I}c_i. \tag{A.41}$$

It follows that

$$\frac{\phi'-\phi}{\rho\delta} = \frac{c_1}{\rho\delta} - \pi_1^b(1-\pi_2^b)\prod_{i>2,i\in I}(1-\pi_i^b). \tag{A.42}$$

If $0 < \pi_1^b \leq \pi_i^b$ for $1 \leq i \leq I$, then

$$\frac{\phi'-\phi}{\rho\delta} \geq \frac{c_1}{\rho\delta} - \pi_2^b(1-\pi_2^b) \geq \frac{c_1}{\rho\delta} - \frac{1}{4} > 0. \tag{A.43}$$

Thus, the buyer could have earned strictly more profit by quitting the venue with the lowest matching probability, so it cannot be an equilibrium for a buyer to participate in more than one trading venue. □

Proof of Lemma 1. According to Lemma A.2, Assumption 1 implies that a buyer participates in at most one submarket. In submarket k, the expected profit of a buyer $\psi_k^b = \rho\delta\pi_k^b - c = \lambda\rho\delta/b_k^{\frac{1-\theta}{2}} - c$ decreases with b_k. For any two submarkets j and k, we must have that $b_j = b_k$; otherwise, buyers in the submarket with more buyers are not maximizing their profits.

If $b_j > \bar{b}/s$, a buyer's expected profit $\psi_j^b < 0$ and it cannot be an equilibrium. Thus, $b_j \leq \bar{b}/s$ and the total mass of buyers participating in all submarkets $\sum_{j=1}^{s} b_j = sb_j \leq \bar{b}$.

- If $b \geq \bar{b}$, then some buyers do not participate in any submarket and earn zero profit. It can be an equilibrium only if buyers who participate in trading also earn zero profits, which holds when $b_j = \bar{b}/s$ for all j.
- If $b < \bar{b}$, then every buyer earns positive profit when $b_j = b/s$. It is not an equilibrium for any buyer to earn zero profit.

Therefore, $b_j = \min\{\bar{b}, b\}/s = b^*/s$ for any submarket j. Other results then follow. □

Lemma A.3. *If buyers in the AS market and buyers in the QC market experience the same level of liquidity, then* $b_{\text{as}}/s_{\text{as}} = b_{\text{qc}}/s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}$ *and* $\pi_{\text{qc}}^s = \pi_{\text{as}}^s s_{\text{qc}}^{\frac{2\theta}{1-\theta}}$.

Proof of Lemma A.3. Every submarket in the AS market includes one seller and $b_{\text{as}}/s_{\text{as}}$ buyers, so $\pi_{\text{as}}^b = \lambda\left(s_{\text{as}}/b_{\text{as}}\right)^{\frac{1-\theta}{2}}$ and $\pi_{\text{as}}^s = \lambda\left(b_{\text{as}}/s_{\text{as}}\right)^{\frac{1+\theta}{2}}$. Together with Eq. (13), we have that

$$\frac{\pi_{\text{as}}^b}{\pi_{\text{qc}}^b} = \frac{s_{\text{as}}^{\frac{1-\theta}{2}}}{b_{\text{as}}^{\frac{1-\theta}{2}}}\frac{b_{\text{qc}}^{\frac{1-\theta}{2}}}{s_{\text{qc}}^{\frac{1+\theta}{2}}} = \left(\frac{s_{\text{as}} b_{\text{qc}}}{b_{\text{as}} s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}}\right)^{\frac{1-\theta}{2}} \tag{A.44}$$

and

$$\frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s} = \frac{b_{\text{as}}^{\frac{1+\theta}{2}}}{s_{\text{as}}^{\frac{1+\theta}{2}}}\frac{s_{\text{qc}}^{\frac{1-\theta}{2}}}{b_{\text{qc}}^{\frac{1+\theta}{2}}} = \left(\frac{\pi_{\text{qc}}^b}{\pi_{\text{as}}^b}\right)^{\frac{1+\theta}{1-\theta}} \frac{1}{s_{\text{qc}}^{\frac{2\theta}{1-\theta}}}. \tag{A.45}$$

If $\pi_{\text{as}}^b = \pi_{\text{qc}}^b$, then $b_{\text{as}}/s_{\text{as}} = b_{\text{qc}}/s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}$ and $\pi_{\text{qc}}^s = \pi_{\text{as}}^s s_{\text{qc}}^{\frac{2\theta}{1-\theta}}$. □

Lemma A.4. *Given* π_{qc}^s, π_{as}^s, *and* $\bar{v}_{\text{qc}}$, *a seller chooses the QC market if* $v \leq \bar{v}_{\text{qc}} + \left(1 - \frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s}\right)(1-\rho)\delta$.

Proof of Lemma A.4. If the seller chooses the AS market, her expected profit equals $\phi_{\text{as}}^s = \pi_{\text{as}}^s(1-\rho)\delta$; if she chooses the QC market, her expected profit equals

$$\phi_{\text{qc}}^s = \pi_{\text{qc}}^s\left[(1-\rho)(\bar{v}_{\text{qc}} + \delta - v) + \rho(\bar{v}_{\text{qc}} - v)\right] = \pi_{\text{qc}}^s\left[(\bar{v}_{\text{qc}} - v) + (1-\rho)\delta\right]. \tag{A.46}$$

It follows that

$$\phi_{\text{qc}}^s - \phi_{\text{as}}^s = \pi_{\text{qc}}^s(\bar{v}_{\text{qc}} - v) + (\pi_{\text{qc}}^s - \pi_{\text{as}}^s)(1-\rho)\delta, \tag{A.47}$$

which is positive only if $v > \bar{v}_{\text{qc}} + \left(1 - \frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s}\right)(1-\rho)\delta$. □

Proof of Theorem 1. First, we find buyers' venue choices. Lemma A.3 implies that $b_{\text{as}}/s_{\text{as}} = b_{\text{qc}}/s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}$. Thus,

$$\gamma = \frac{b^*}{s + \mu(sq)} = \frac{b_{\text{as}} + b_{\text{qc}}}{s_{\text{as}} + s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}} = \frac{b_{\text{as}}}{s_{\text{as}}} = \frac{b_{\text{qc}}}{s_{\text{qc}}^{\frac{1+\theta}{1-\theta}}} \tag{A.48}$$

It follows that

$$b_{\text{as}} = \frac{s_{\text{as}}}{s+\mu(sq)} b^* = \frac{s-sq}{s+\mu(sq)} \left(b_{\text{as}} + b_{\text{qc}}\right), \quad b_{\text{qc}} = b^* - b_{\text{as}} = \frac{sq+\mu(sq)}{s+\mu(sq)} \left(b_{\text{as}} + b_{\text{qc}}\right). \tag{A.49}$$

Eq. (2) then implies that

$$\pi^b = \pi^b_{\text{as}} = \pi^b_{\text{qc}} = \lambda \gamma^{\frac{1-\theta}{2}}. \tag{A.50}$$

Because a buyer's expected profit $\pi^b \rho\delta - c \geq 0$, we have that

$$b^* \leq (s+\mu(sq)) \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}} = \bar{b}. \tag{A.51}$$

In addition, $b^* = b_{\text{as}} + b_{\text{qc}} \leq b$. Thus, $b^* = \min\left\{\bar{b}, b\right\}$. We can then find π^s_{as}, π^s_{qc}, m_{as}, m_{qc}, and m by definition.

Second, we find sellers' venue choices. Define

$$\eta(v^*) := \mathbf{E}[v | v \in [\underline{v}, v^*]] + (1-\rho)\delta \left(1 - \frac{1}{\left(s \cdot \Pr\left\{v \in [\underline{v}, v^*]\right\}\right)^{\frac{2\theta}{1-\theta}}}\right), \tag{A.52}$$

which is an increasing function. If an asset's value $v' > \bar{v}$, then by definition of $\bar{v}$ in Eq. (15), $v' > \eta(v') \geq \eta(\bar{v})$. If an asset's value $v' \leq \bar{v}$, then $v' \leq \bar{v} \leq \eta(\bar{v})$. Lemmas A.3 and A.4 imply that a seller prefers the AS market if

$$v > \bar{v}_{\text{qc}} + \left(1 - \frac{\pi^s_{\text{as}}}{\pi^s_{\text{qc}}}\right)(1-\rho)\delta = \bar{v}_{\text{qc}} + (1-\rho)\delta \left(1 - \frac{1}{s_{\text{qc}}^{\frac{2\theta}{1-\theta}}}\right). \tag{A.53}$$

Thus, a seller prefers the AS market if $v > \bar{v}$ and the QC market otherwise. Sellers of QC-ineligible assets can only choose the AS market. Therefore, sellers choose the QC market if $v \in [\underline{v}, \bar{v}]$ and the AS market otherwise. It implies that $s_{\text{qc}} = s(1-q)$ and $s_{\text{as}} = sq$.

Third, trading volume in the AS market $m_{\text{as}} = \pi^b b_{\text{as}}$ and that in the QC market $m_{\text{qc}} = \pi^b b_{\text{qc}}$. Thus, $m = m_{\text{as}} + m_{\text{qc}} = \pi^b b^* = \lambda (b^*)^{\frac{1+\theta}{2}} (s+\mu(sq))^{\frac{1-\theta}{2}}$. Every participating buyer expects to earn $\rho\delta\pi^b - c$, so buyers' total expected profit equals $\Omega^b = b^* \left(\rho\delta\pi^b - c\right)$. AS sellers' total profit equals $\Omega^s_{\text{as}} = s_{\text{as}}\pi^s_{\text{as}}(1-\rho)\delta = m_{\text{as}}(1-\rho)\delta$. Eq. (32) implies that the total profit of QC sellers equals

$$\Omega^s_{\text{qc}} = \int_{v\in[\underline{v},\bar{v}]} s\pi^s_{\text{qc}} \left((1-\rho)\delta + \bar{v}_{\text{qc}} - v\right) dF_v = (1-\rho)\delta\pi^s_{\text{qc}} s \int_{v\in[\underline{v},\bar{v}]} dF_v = (1-\rho)\delta m_{\text{qc}}. \tag{A.54}$$

Thus, sellers' total profit equals

$$\Omega^s = \Omega^s_{\text{as}} + \Omega^s_{\text{qc}} = (1-\rho)\delta(m_{\text{as}} + m_{\text{qc}}) = (1-\rho)\delta m. \tag{A.55}$$

Therefore, $\Omega = \Omega^b + \Omega^s = m\delta - cb^*$. □

Lemma A.5. $1 \leq \frac{s+\mu(sq)}{s} \leq (sq)^{\frac{2\theta}{1-\theta}}$ *and inequalities bind when* $sq = 1$.

Proof of Lemma A.5. By definition, $q \leq 1$ and $\mu(1) = 0$. By assumption, the number of assets sold in the

QC market $sq \geq 1$. It follows that $\mu(sq) = sq\left((sq)^{\frac{2\theta}{1-\theta}} - 1\right) \leq s\left((sq)^{\frac{2\theta}{1-\theta}} - 1\right)$, which implies that $(sq)^{\frac{2\theta}{1-\theta}} \geq 1 + \frac{\mu(sq)}{s} = \frac{s+\mu(sq)}{s} \geq 1$. When $sq = 1$, $\mu(sq) = 0$ and $s + \mu(sq) = s$. $\square$

Proof of Corollary 1. Eqs. (5) and (18) implies that $\bar{b}/\bar{b}_0 = 1 + \mu(sq)/s \geq 1$. Eqs. (4) and (17) imply that

$$b^* - b_0^* = \min\{b, \bar{b}\} - \min\{b, \bar{b}_0\} = \begin{cases} 0 & \text{if } b < \bar{b}_0 \leq \bar{b}, \\ b - \bar{b}_0 & \text{if } \bar{b}_0 \leq b \leq \bar{b}, \\ \bar{b} - \bar{b}_0 & \text{if } \bar{b}_0 \leq \bar{b} \leq b. \end{cases} \tag{A.56}$$

Thus, $b^* - b_0^* \geq 0$. If $sq > 1$, then $\bar{b} > \bar{b}_0$; if $sq > 1$ and $b > \bar{b}_0$, then $b^* > b_0^*$.

Lemma 1 and Theorem 1 imply that $\frac{b_{\text{as}}}{s_{\text{as}}} = \gamma$, $\frac{b_0^*}{s} = \min\left\{\frac{b}{s}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\}$, and $\frac{b_{\text{qc}}}{s_{\text{qc}}} = \gamma s_{\text{qc}}^{\frac{2\theta}{1-\theta}}$. By assumption, $sq \geq 1$. Thus, $\mu'(\cdot) \geq 0$ and $\mu(sq) \geq 0$. It follows that

$$\frac{b_0^*}{s} = \min\left\{\frac{b}{s}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\} \geq \min\left\{\frac{b}{s+\mu(sq)}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\} = \gamma = \frac{b_{\text{as}}}{s_{\text{as}}}. \tag{A.57}$$

Lemma A.5 shows that $(sq)^{\frac{2\theta}{1-\theta}} \geq \frac{s+\mu(sq)}{s}$. Thus,

$$\frac{b_{\text{qc}}}{s_{\text{qc}}} = \gamma s_{\text{qc}}^{\frac{2\theta}{1-\theta}} \geq \min\left\{\frac{b}{s+\mu(sq)}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\} \times \frac{s+\mu(sq)}{s} \geq \min\left\{\frac{b}{s}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\} = \frac{b_0^*}{s}. \tag{A.58}$$

It follows that

$$\frac{\pi^b}{\pi_0^b} = \left(\frac{b_0^*/s}{b_{\text{as}}/s_{\text{as}}}\right)^{\frac{1-\theta}{2}} \geq 1 \quad \text{and} \quad \frac{\pi_{\text{as}}^s}{\pi_0^s} = \left(\frac{b_{\text{as}}/s_{\text{as}}}{b_0^*/s}\right)^{\frac{1+\theta}{2}} \leq 1. \tag{A.59}$$

Because $b_{\text{as}}/s_{\text{as}} = \gamma = b_{\text{qc}}/s_{\text{qc}} s_{\text{qc}}^{\frac{-2\theta}{1-\theta}}$, we have that

$$\frac{\pi_{\text{qc}}^s}{\pi_0^s} = \left(\frac{b_{\text{as}}/s_{\text{as}}}{b_0^*/s}\right)^{\frac{1+\theta}{2}} (sq)^{\frac{2\theta}{1-\theta}} = \underbrace{\left(\frac{b_{\text{qc}}/s_{\text{qc}}}{b_0^*/s}\right)^{\frac{1+\theta}{2}}}_{\text{impact of } b/s \text{ ratio}} \times \underbrace{(sq)^{\theta}}_{\text{QC benefits}} \geq 1. \tag{A.60}$$

When $sq = 1$, then $\mu(sq) = 0$ and all above inequalities bind. In addition, if $b \geq \bar{b}$, then $b_0^* = \bar{b}_0$ and $b^* = \bar{b}$. It follows that $b_{\text{as}}/s_{\text{as}} = b_0^*/s = (\lambda\rho\delta/c)^{\frac{2}{1-\theta}}$, $\pi^b = \pi_0^b = c/(\rho\delta)$ and $\pi_{\text{as}}^s = \pi_0^s = \lambda(\lambda\rho\delta/c)^{\frac{1+\theta}{1-\theta}}$. If $sq > 1$ and $b < \bar{b}_0$, then

$$\frac{b_{\text{qc}}}{s_{\text{qc}}} > \frac{b}{s} = \frac{b_0^*}{s} > \frac{b}{s+\mu(sq)} = \frac{b_{\text{as}}}{s_{\text{as}}} = \gamma \tag{A.61}$$

It follows that $\pi^b > \pi_0^b$, $\pi_{\text{as}}^s < \pi_0^s$, and $\pi_{\text{qc}}^s > \pi_0^s$. Eqs. (7) and (26) imply that

$$\frac{m}{m_0} = \left(\frac{b^*}{b_0^*}\right)^{\frac{1+\theta}{2}} \left(1 + \frac{\mu(sq)}{s}\right)^{\frac{1-\theta}{2}} \geq 1. \tag{A.62}$$

$\square$

Proof of Corollary 2. We have that $\psi_0^b = \pi_0^b \rho\delta - c$, $\psi^b = \pi^b \rho\delta - c$, $\psi_0^s = (1-\rho)\delta\pi_0^s$, and $\psi_{\text{as}}^s = (1-\rho)\delta\pi_{\text{as}}^s$. According to Corollary 1, $\pi^b \geq \pi_0^b$ and $\pi_{\text{as}}^s \leq \pi_0^s$. It follows that $\psi^b - \psi_0^b = \rho\delta\left(\pi^b - \pi_0^b\right) \geq 0$ and $\psi_{\text{as}}^s - \psi_0^s = (\pi_{\text{as}}^s - \pi_0^s)(1-\rho)\delta \leq 0$. In addition, given Eq. (32), we have that

$$\psi_{\text{qc}}^s(v) - \psi_0^s = (1-\rho)\delta\left(\pi_{\text{qc}}^s - \pi_0^s\right) + \pi_{\text{qc}}^s(\bar{v}_{\text{qc}} - v) = \pi_{\text{qc}}^s\left(v^* - v\right). \tag{A.63}$$

If $\bar{v} < v_{\max}$, then Eq. (15) implies that

$$v_{\max} > \mathbf{E}[v|v \in [\underline{v}, \bar{v}]] + (1-\rho)\delta\left(1 - \frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s}\right) \geq v^*. \tag{A.64}$$

Hence $v^* < v_{\max}$. If $v^* \in (\bar{v}, v_{\max})$, then Eq. (15) implies that

$$v^* > \mathbf{E}[v|v \in [\underline{v}, \bar{v}]] + (1-\rho)\delta\left(1 - \frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s}\right), \tag{A.65}$$

which implies that $\pi_{\text{as}}^s > \pi_0^s$. It contradicts Corollary 1. Thus, $v^* \leq \bar{v}$. Note that if $\bar{v} = v_{\max}$, then it is possible that $v_{\max} < v^* \leq \mathbf{E}[v|v \in [\underline{v}, \bar{v}]] + (1-\rho)\delta\left(1 - \frac{\pi_{\text{as}}^s}{\pi_{\text{qc}}^s}\right)$. Corollary 1 and Eq. (A.62) show that $b^* \geq b_0^*$, $\pi^b \geq \pi_0^b$, and $m \geq m_0$. Theorem 1 implies that

$$\Omega = \underbrace{b^*(\rho\delta\pi^b - c)}_{\text{buyers' profits } \Omega^b} + \underbrace{m(1-\rho)\delta}_{\text{sellers' profits } \Omega^s} \geq \underbrace{b_0^*(\rho\delta\pi_0^b - c)}_{\text{buyers' profits } \Omega_0^b} + \underbrace{m_0(1-\rho)\delta}_{\text{sellers' profits } \Omega_0^s} = \Omega_0. \tag{A.66}$$

$\square$

Lemma A.6. *Let $q_{t,k}$ represent the fraction of type-t asset sold in the k-th QC contract. Then, the market capacity for buyers and the number of participating buyers equal, respectively,*

$$\bar{b} = \left(s + \sum_{t\in\vec{T}, k\in\vec{K}_t} \mu(s_t q_{t,k})\right)\left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}} \tag{A.67}$$

$$b^* = \min\left\{b, \bar{b}\right\}. \tag{A.68}$$

The adjusted buyer-to-seller ratio equals

$$\gamma = \frac{b^*}{s + \sum_{t\in\vec{T}, k\in\vec{K}} \mu(s_t q_{t,k})} = \min\left\{\frac{b}{s + \sum_{t\in\vec{T}, k\in\vec{K}} \mu(s_t q_{t,k})}, \left(\frac{\lambda\rho\delta}{c}\right)^{\frac{2}{1-\theta}}\right\}, \tag{A.69}$$

The liquidity levels for any AS seller and any buyer equal, respectively,

$$\pi_{\text{as}}^s = \lambda\gamma^{\frac{1+\theta}{2}} \quad \textit{and} \quad \pi^b = \frac{\lambda}{\gamma^{\frac{1-\theta}{2}}}. \tag{A.70}$$

The expected total trading volume equals

$$m = \frac{\lambda b^*}{\gamma^{\frac{1-\theta}{2}}} \tag{A.71}$$

and the total welfare gain equals

$$\Omega = \underbrace{b^*(\rho\delta\pi^b - c)}_{\text{buyers' profits } \Omega^b} + \underbrace{m(1-\rho)\delta}_{\text{sellers' profits } \Omega^s} = m\delta - cb^*. \tag{A.72}$$

Proof of Lemma A.6. Lemma A.3 implies that $b_{\mathrm{as},t}/s_{\mathrm{as},t} = b_{\mathrm{qc},t,k}/s_{\mathrm{qc},t,k}^{\frac{1+\theta}{1-\theta}}$ for any $t \in \vec{T}$ and $k \in \vec{K}_t$. It implies that

$$\gamma = \frac{b^*}{s + \sum_{t\in\vec{T},k\in\vec{K}} \mu(s_t q_{t,k})} = \frac{\sum_{t\in\vec{T}} b_{\mathrm{as},t} + \sum_{t\in\vec{T},k\in\vec{K}} b_{\mathrm{qc},t,k}}{\sum_{t\in\vec{T}} s_{\mathrm{as},t} + \sum_{k\in\vec{K}} s_{\mathrm{qc},t,k}^{\frac{1+\theta}{1-\theta}}} = \frac{b_{\mathrm{as},t}}{s_{\mathrm{as},t}} = \frac{b_{\mathrm{qc},t,k}}{s_{\mathrm{qc},t,k}^{\frac{1+\theta}{1-\theta}}}. \tag{A.73}$$

We can then find π^b. All other results are proved similarly as Theorem 1. □

Proof of Corollary 3. When q increases, $s_{\mathrm{qc}} = sq$ increases and $s_{\mathrm{as}} = s - s_{\mathrm{qc}}$ decreases, which implies that $\mu(sq)$ increases. Eq. (18) then implies that $\bar{b}$ increases and $b^* = \min\{b, \bar{b}\}$ weakly increases. Eq. (20) implies that γ decreases. Then, Eqs. (21), (24) and (26) imply that π^b increases, π^s_{as} decreases, m increases, Ω^b increases, Ω^s increases, and $\Omega = \Omega^b + \Omega^s$ increases.

If $b > \bar{b}$ after q increases, then the market capacity for buyers $\bar{b}$ is binding and $b^* = \bar{b}$ increases. Eq. (20) implies that $\gamma = \left(\lambda\rho\delta/c\right)^{\frac{2}{1-\theta}}$ is unaffected. Then, π^b and π^s_{as} are unaffected. □

Proof of Proposition 1. First, when s increases, $s\Pr\{v \in [\underline{v}, x]\}$ in Eq. (15) increases for any v, so $\bar{\sigma}(\underline{v}, s, F, \theta)$ increases with s. Second, $s_i = s_{i'} + s_{i''} \ge \max\{s_{i'}, s_{i''}\}$. Hence, $\bar{v}_{i,k} \ge \max\{\bar{v}_{i',k}, \bar{v}_{i'',k}\}$ and $q_{i,k} \ge \max\{q_{i',k}, q_{i'',k}\}$. Third, when $x > 1$ and $y > 1$, $\mu(x)$ is increasing and $\mu(x+y) - \mu(x) - \mu(y) = (x+y)^\kappa - x^\kappa - y^\kappa \ge 0$, where $\kappa = (1+\theta)/(1-\theta) \ge 1$. Hence, $\mu(s_i q_{i,k}) \ge \mu(s_{i'} q_{i',k} + s_{i''} q_{i'',k}) \ge \mu(s_{i'} q_{i',k}) + \mu(s_{i''} q_{i'',k})$. Thus, $s + \sum \mu(s_t q_{t,k})$ decreases. Lemma A.6 then implies that $\bar{b}$ decreases, b^* decreases, γ increases, π^s_{as} increases, π^b decreases, m decreases, and Ω decreases. □

Proof of Proposition 2. Theorem 1 implies that sellers of type-t assets choose the new QC contract if $v \in [\underline{v}_{t,k+1}, \sigma(\underline{v}_{t,k+1}, F_t, s_t, \theta)]$. Other sellers' choices are unaffected. Lemma A.6 then implies that γ decreases, whereas π^b and m increase. In addition, $b_{\mathrm{as},t} = \gamma s_{\mathrm{as},t}$ and $b_{\mathrm{qc},t,k} = \gamma s_{\mathrm{qc},t}^{\frac{1+\theta}{1-\theta}}$ both decrease because $s_{\mathrm{as},t}$ and $s_{\mathrm{qc},t}$ are unchanged. □

B Comparative Statics

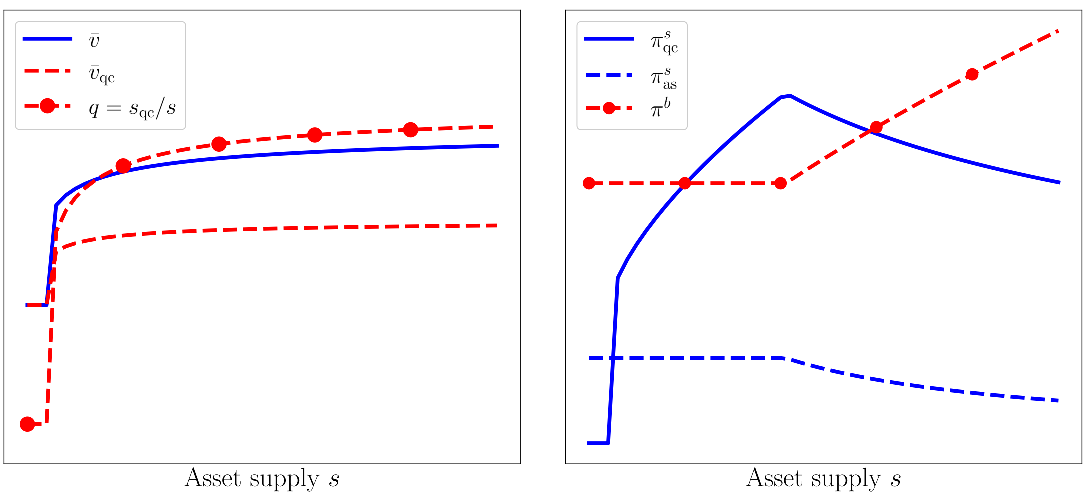
In this section, we briefly discuss the comparative statics concerning several exogenous factors.

B.1 Asset Supply

We examine, based on Theorem 1, the effect of increasing asset supply s while keeping the distribution of asset value F as fixed. In this situation, assets are "denser" because more assets are available within any given range of asset values.[20]

Corollary A.4 (Effects of asset supply s). *Suppose that s increases. Prices in the QC market increase ($\bar{v}_{\text{qc}}\uparrow$). Greater proportions of sellers and buyers choose the QC market (i.e., s_{qc}/s and b_{qc}/b both increase). All buyers experience better liquidity ($\pi^b\uparrow$) but AS sellers experience worse liquidity ($\pi^s_{\text{as}}\downarrow$). The total trading volume m and the welfare gain Ω both increase.*

Proof of Corollary A.4. When s increases, $s\Pr\{v \in [\underline{v}, x]\}$ in Eq. (15) increases for any v. Thus, $\bar{v}$ increases. It follows $q = \Pr\{v \in [\underline{v}, \bar{v}]\}$, $s_{\text{qc}} = sq$, $\bar{v}_{\text{qc}} = \mathbf{E}[v|v \in [\underline{v}, \bar{v}]]$, $s + \mu(sq)$, $\bar{b}$, and b^* all increase. Thus, γ and $b_{\text{as}} = s_{\text{as}}\gamma$ decrease, whereas $b_{\text{qc}} = b^* - b_{\text{as}}$ increases. It follows that $\pi^b = \frac{\lambda}{\gamma^{\frac{1-\theta}{2}}}$, $m = b^*\pi^b$, $\Omega^b = b^*(\rho\delta\pi^b - c)$, $\Omega^s = m(1-\rho)\delta$, and $\Omega = \Omega^b + \Omega^s$ all increase, whereas $\pi^s_{\text{as}} = \lambda\gamma^{\frac{1+\theta}{2}}$ decreases. Note that $\pi^s_{\text{qc}} = \lambda\gamma^{\frac{1+\theta}{2}} s_{\text{qc}}^{\frac{2\theta}{1-\theta}}$ is not monotonic in s because γ decreases and s_{qc} increases. □



(a) QC market properties **(b)** Liquidity

Figure A.5. Impact of asset supply s.

Fig. A.5 illustrates the results reported in Corollary A.4. First, contrary to the usual intuitions, increasing the supply of assets actually *increases* the price in the QC market because of a composition effect. In particular, if s increases, then the upper bound of QC asset values $\bar{v}$ must also increase. If $\bar{v}$ stayed the same, then the number of sellers in every trading venue would increase proportionally with s, which, because of the liquidity benefit of pooling assets, would increase the liquidity advantage of the QC market over the AS market for buyers. It cannot be an equilibrium because buyers can move freely from the AS market to the QC market. Once some buyers have migrated from the AS market to the QC market, the QC market will be more liquid than the AS market for sellers, which pushes sellers of higher-value assets to switch from the AS market to the QC market and increases $\bar{v}$. Therefore, in equilibrium the average value of assets traded in the QC market $\bar{v}_{\text{qc}}$ increases and buyers in the QC market bid higher prices.[21]

[20] In Appendix B.4, we examine the effects when s is fixed but the distribution F becomes more concentrated.

[21] A similar positive relationship between asset supply and asset price also arises in the models developed by

Second, because $\bar{v}$ increases, a larger proportion of assets are sold through the more efficient QC market ($q \uparrow$). Thus, every buyer is more likely to trade, which increases the total trading volume and welfare gains. Nonetheless, the usual caveat applies: sellers who remain in the AS market experience worse liquidity ($\pi^s_{\text{as}} \downarrow$) when s increases.

Third, the liquidity levels for QC sellers π^s_{qc} may increase or decrease with s. Intuitively, an increase in s attracts both buyers and sellers to the QC market, with the former effect improving but the latter hurting liquidity for QC sellers. When almost all buyers are already in the QC market, further increasing s mostly intensifies the competition between QC sellers and hurts their probabilities of trading.

In summary, the positive effects of introducing QC trading on trading volume and welfare gains will be more prominent when a higher number of assets with heterogeneous values are traded.

B.2 Costs to Analyze Assets

Next, we examine the effects of lowering the cost c to analyze an asset. A key difference between this paper and other papers in the literature is that in our paper assets differ in fundamental value, which requires buyers who participate in a trading venue to analyze the assets traded on that venue.

So far, we have taken c as given. In practice, regulators may influence c by, for example, varying the requirements for sellers' information disclosure. Moreover, introducing QC trading could indirectly reduce c because it narrows the range of asset values a buyer may encounter. First, QC trading can reduce the heterogeneity of asset values in the QC market. As An et al. (2022) show, because all TBA MBS obtain the same price, issuers package a large number o low-value mortgages together into a small number of TBA MBS, resulting in a value distribution of TBA MBS that is significantly more concentrated than the value distribution of underlying mortgages. Second, QC trading can narrow the range of asset values in the AS market. Specifically, introducing a QC market separates assets into three segments by their values: an AS segment for QC-ineligible assets, a QC market for low-value QC-eligible assets, and an AS segment for high-value QC-eligible assets. Thus, buyers on a submarket of the AS market can infer that the traded asset's value is either very low or very high, depending on whether it is QC-eligible. In summary, introducing QC trading could reduce buyers' costs of analysis on every trading venue.

We describe the effects of reducing c as follows, in Corollary A.5.

Corollary A.5 (**Effects of asset analysis cost** c)**.** *Suppose that Assumption 1 holds. Reducing c does not affect sellers' venue choice ($\bar{v}$) and weakly increases buyers' total profits Ω^b.*

- *If the capacity for buyers is not binding ($b \le \bar{b}$), then reducing c does not affect buyers' venue choices ($b^*, b_{\text{as}}, b_{\text{qc}}$), the liquidity levels ($\pi^b, \pi^s_{\text{as}}, \pi^s_{\text{qc}}, m$), and sellers' profits (Ω^s).*
- *If the market capacity for buyers is binding ($b > \bar{b}$), then reducing c hurts buyers' liquidity π^b, but increases buyer participation b^*, sellers' liquidity levels π^s_{as} and π^s_{qc}, total trading volume m, sellers' total profit Ω^s, and total welfare gain Ω.*

Vayanos and Wang (2007) and Weill (2008) because buyers in their models are willing to pay premia for more liquid assets. In contrast, buyers in our model always pay the expected values of assets regardless of the liquidity associated with those assets. Thus, in our paper the effect arises from a distinct channel: additional supply changes the composition of assets traded in the QC market. Our channel is absent from their models because in their papers assets share the same fundamental values.

Proof of Corollary A.5. Results follow from Theorem 1. Eq. (15) implies that c does not affect $\bar{v}$, so q, s_{as}, and s_{qc} stay the same. Further, lowering c reduces $\lambda\rho\delta/c$, so it increases $\bar{b}$.

If $b \le \bar{b}$, then $b^* = b$ and $\gamma = b/(s+\mu(sq)$ are unaffected by c. Thus, b_{as}, b_{qc}, π^b, π^s_{as}, π^s_{LLJ}, m, Ω^s stay the same. $\Omega^b = b^*(\pi^b\rho\delta - c)$ increases when c declines.

If $b > \bar{b}$, then lowering c increases b^*, γ, π^s_{qc}, π^s_{as}, Ω^s, and m, whereas reduces π^b. Because $\Omega^b = 0$ when $b > \bar{b}$ and Ω^b cannot be negative, it must weakly increase. □

First, although reducing c expands the market capacity for buyers $\bar{b}$, it does not affect sellers' venue choices. In particular, a buyer would still be indifferent between venues after additional buyers participate. As a result, for sellers, the relative liquidity difference between the QC market and the AS market would be unaffected, so they do not switch venues.

Second, reducing c directly lowers a buyer's cost and if it attracts more buyers to participate, in which case at least some benefits are passed through to sellers because of the increased competition between participating buyers. On the one hand, if the market capacity for buyers is not binding ($b \le \bar{b}$), then all buyers participate in the pure-AS equilibrium and no more buyers can participate. Thus, reducing c would have no impact other than increasing a buyer's profit mechanically by c. Second, if the capacity is binding ($b < \bar{b}$), then reducing c induces more buyers to participate, which hurts liquidity for participating buyers ($\pi^b \downarrow$) but improves liquidity for sellers (π^s_{as} and $\pi^s_{\text{qc}} \uparrow$). As a result, every asset is more likely to be sold and a seller earns a higher expected profit, so the trading volume m and sellers' total profit Ω^s both increase. If the market capacity is still binding *after* c declines, then buyers still earn zero profits and all the gains resulting from the reduction in c would be passed through to sellers.

B.3 Buyers' Risk Attitude

We assume in the main model that buyers are risk-neutral. In this section, we allow buyers to be risk-averse. We assume that buyers in the QC market bid $\widetilde{P}_{\text{qc}}$ and accept the ask price $\widetilde{P}_{\text{qc}} + \delta$, where

$$\widetilde{P}_{\text{qc}} = \zeta \min\left\{v : v \in QC\right\} + (1-\zeta)\,\mathbf{E}[v|v \in QC]. \tag{A.74}$$

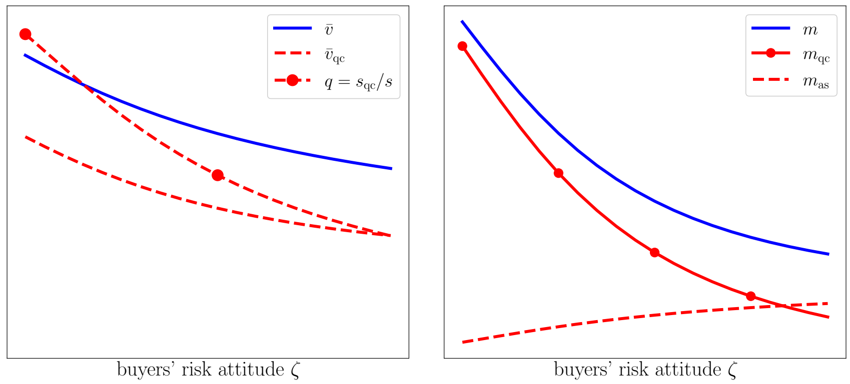
In particular, the parameter ζ reflects buyers' degree of risk aversion. When $\zeta = 0$, buyers are risk-neutral; when $\zeta = 1$, buyers are ambiguity-averse and bid the value of the least valuable QC asset $\underline{v}$.

Fig. A.6 shows that when buyers are more risk-averse ($\zeta \uparrow$), they bid lower prices, which drives some sellers out of the QC market ($\bar{v} \downarrow$). As a result, the QC market becomes less liquid for buyers, which drives some buyers to the AS market and improves liquidity for AS sellers. Overall, because fewer traders use QC trading, the total trading volume m falls despite an increase in AS trading volume m_{as}.

Notice that the QC market may not shut down even if buyers are infinitely risk-averse ($\zeta = 1$). In this scenario, sellers of assets that are slightly more valuable than $\underline{v}$ still sell on the QC market because the liquidity benefit of QC trading outweighs the associated price discount.

B.4 Impact of Asset Value Dispersion

The effects of QC trading also depends on asset heterogeneity. One measure of asset heterogeneity is the variance of asset values. Fig. A.7 illustrates, for example, the equilibria when $\underline{v} = 0.4$ and v follows an unimodal distribution Beta$(0.5x, 0.5x)$ for $x \in (5, 50)$ so that $\mathbf{E}[v]$ is fixed and $\mathbf{V}\text{ar}[v]$ varies. In this example, as asset values become more heterogeneous ($\mathbf{V}\text{ar}[v] \uparrow$), the coverage of QC trading shrinks ($\bar{v}$

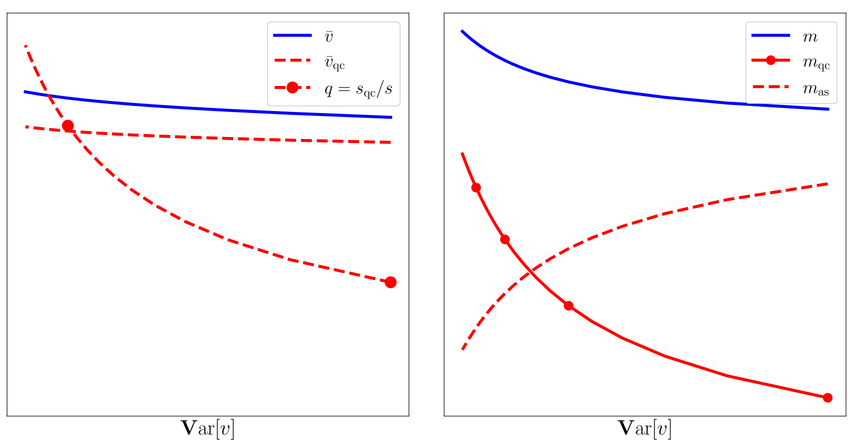


(a) QC market properties **(b)** Trading volume

Figure A.6. Impact of buyers' risk attitude ζ.

and $q \downarrow$) and the price in the QC market $\bar{v}_{\text{qc}}$ declines. In consequence, buyers experience worse liquidity and the overall trading volume m falls. Intuitively, increasing $\mathbf{V}\text{ar}[v]$ reduces the number of assets in the value interval $[\underline{v}, \bar{v}]$, which reduces the liquidity benefits of QC trading for buyers. Thus, buyers migrate to the AS market, which prompt some sellers to exit the QC market. Note that these effects depend on $\underline{v}$. If, for example, $\underline{v} = 0.8$, then increasing $\mathbf{V}\text{ar}[v]$ could *increase* the number of assets in the interval $[\underline{v}, \bar{v}]$.

In summary, results in Appendices B.1 and B.4 suggest that introducing QC trading is most likely to improve market liquidity when numerous assets with similar values (but different characteristics) are available right above the threshold $\underline{v}$. If too few assets are available or asset values differ substantially, then the market is too "thin" for QC trading to effect considerable impact because few traders would use it.



(a) QC market properties **(b)** Trading volume

Figure A.7. Impact of $\mathbf{V}\text{ar}[v]$ when $\underline{v} = 0.3$ and $v \sim \text{Beta}(0.5x, 0.5x)$ for $x \in (5, 50)$.

References

An, Yu, "Competing with Inventory in Dealership Markets," *Working Paper*, 2019, *Johns Hopkins Carey Business School.*

______, **Wei Li, and Zhaogang Song**, "TBA Trading and Security Issuance in the Agency MBS Market," *Available at SSRN 3674660*, 2022.

Bessembinder, Hendrik and William Maxwell, "Transparency and the corporate bond market," *Journal of Economic Perspectives*, 2008, *22*, 217–234.

______, **Chester Spatt, and Kumar Venkataraman**, "A Survey of the Microstructure of Fixed-Income Markets," *Journal of Financial and Quantitative Analysis*, 1 2019.

______, **William Maxwell, and Kumar Venkataraman**, "Trading activity and transaction costs in structured credit products," *Financial Analysts Journal*, 2013, *69(6)*, 55–68.

Boyarchenko, Nina, Andreas Fuster, and David O. Lucca, "Understanding Mortgage Spreads," *Review of Financial Studies*, 2019.

Carlin, Bruce, Francis Longstaff, and Kyle Matoba, "Disagreement and Asset Prices," *Journal of Financial Economics*, 2014, *114*, 226–238.

Chernov, Mikhail, Brett Dunn, and Francis Longstaff, "Macroeconomic-driven Prepayment Risk and the Valuation of Mortgage-Backed Securities," *Review of Financial Studies*, 2018, *31* (3), 1132–1183.

Clark, James, Chris Cameron, and Gabriel Mann, "Examining Liquidity in On-the-Run and Off-the-Run Treasury Securities," *Treasury Notes Blog*, 2016.

DeMarzo, Peter M., "The Pooling and Tranching of Securities: A Model of Informed Intermediation," *Review of Financial Studies*, 2005, *18* (1), 1–35.

Demsetz, Harold, "The Cost of Transacting," *Quarterly Journal of Economics*, 1968, *82* (1), 33–53.

Di Maggio, Marco, Amir Kermani, and Zhaogang Song, "The value of trading relations in turbulent times," *Journal of Financial Economics*, 2017, *124(2)*, 266–284.

Diep, Peter, Andrea Eisfeldt, and Scott Richardson, "The Cross Section of MBS Returns," *Journal of Finance*, 2021, *76* (5), 2093–2151.

Downing, Christopher, Dwight Jaffee, and Nancy Wallace, "Is the Market for Mortgage Backed Securities a Market for Lemons?," *Review of Financial Studies*, 2009, *22–7*, 2457–2494.

Duarte, Jefferspm, Francis Longstaff, and Fan Yu, "Risk and Return in Fixed Income Arbitrage: Nickels in Front of a Steamroller?," *Review of Financial Studies*, 2007, *20* (3), 769–811.

Duffie, Darrell, *Dark Markets: Asset Pricing and Information Transmission in Over-the-Counter Markets*, New Jersey: Princeton University Press, 2012.

Eisfeldt, Andrea, Hanno Lustig, and Lei Zhang, "Complex Asset Markets," *Working Paper*, 2019.

Fusari, Nicola, Wei Li, Haoyang Liu, and Zhaogang Song, “Asset Pricing with Cohort-Based Trading in MBS Markets,” *Journal of Finance (Forthcoming)*, 2022.

Gabaix, Xavier, Arvind Krishnamurthy, and Olivier Vigneron, “Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market,” *Journal of Finance*, 2007, *2*, 557–595.

Gao, Pengjie, Paul Schultz, and Zhaogang Song, “Liquidity in a Market for Unique Assets: Specified Pool and TBA Trading in the Mortgage Backed Securities Market,” *Journal of Finance*, 2017, *72-3*, 1119–1170.

______, ______, and ______, “Trading Methods and Trading Costs for Agency Mortgage Backed Securities,” *Journal of Investment Management*, 2018, *forthcoming.*

Goodman, Laurie and Jim Parrott, “A Progress Report on Fannie Mae and Freddie Mac’s Move to a Single Security,” *Urban Institute*, 2018.

Gorton, Gary and George Pennacchi, “Financial Intermediaries and Liquidity Creation,” *Journal of Finance*, 1990, *45* (1), 49–71.

Guerrieri, Veronica and Robert Shimer, “Dynamic Adverse Selection: A Theory of Illiquidity, Fire Sales, and Flight to Quality,” *American Economic Review*, 2014, *104* (7), 1875–1908.

He, Zhiguo and Zhaogang Song, “Agency MBS as Safe Assets,” *Working Paper*, 2019, *Johns Hopkins Carey Business School.*

Huh, Yesol and You Suk Kim, “The Real Effects of Secondary Market Trading Structure: Evidence from the Mortgage Market,” *Working Paper*, 2019, *Board of Governors of the Federal Reserve System.*

Li, Dan and Norman Schürhoff, “Dealer Networks,” *Journal of Finance*, 2019, *74* (1), 91–144.

Liu, Haoyang, Zhaogang Song, and James Vickery, “Fragmentation, Liquidity, and Pricing in MBS Markets,” *working paper*, 2020.

Milbradt, Konstantin, “Asset heterogeneity in Over-The-Counter markets,” *Working Paper*, 2018.

O’Hara, Maureen and Alex Zhou, “The Electronic Evolution of Corporate Bond Dealers,” *Journal of Financial Economics*, 2020, *Forthcoming.*

Schultz, Paul and Zhaogang Song, “Transparency and dealer networks: Evidence from the initiation of post-trade reporting in the mortgage backed security market,” *Journal of Financial Economics*, 2019, *133* (1), 113–133.

Song, Zhaogang and Haoxiang Zhu, “Mortgage Dollar Roll,” *Review of Financial Studies*, 2019, *32* (8), 2955–2996.

Spatt, Chester, “Frictions in the Bond Market,” *Keynote Speech: Second MTS Conference on Financial Markets*, 2004.

Subrahmanyam, Avanidhar, “A Theory of Trading in Stock Index Futures,” *Review of Financial Studies*, 1991, *4* (1), 17–51.

Üslü, Semih and Güner Velioglu, “Liquidity in the Cross Section of OTC Assets,” *Working Paper*, 2019.

Vayanos, Dimitri and Jiang Wang, “Chapter 19 - Market Liquidity—Theory and Empirical Evidence,” *Handbook of the Economics of Finance*, 2013, *2*, 1289 – 1361.

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

______ **and Tan Wang**, “Search and endogenous concentration of liquidity in asset markets,” *Journal of Economic Theory*, 2007, *136* (1), 66 – 104.

Vickery, James and Joshua Wright, “TBA Trading and Liquidity in the Agency MBS Market,” *Federal Reserve Bank of New York Economic Policy Review*, 2011, *19*.

Weill, Pierre-Olivier, “Liquidity premia in dynamic bargaining markets,” *Journal of Economic Theory*, 2008, *140* (1), 66–96.

______, “The Search Theory of Over-the-Counter Markets,” *Annual Review of Economics*, 2020, *12* (1), 747–773.