Auctions with Imperfectly Competitive Resale: Evidence from Central Bank Auctions*

James Chapman Eric Richert Yu Zhu

August 24, 2026

Abstract

Banks use central bank balances to settle customer payments. When supply is tight, additional balances are allocated as overnight loans by auction, then redistributed through interbank lending. Since auction winnings shape each bank's position, a bank's private information about customer payment needs and their expected rents from secondary market trading generate rents and allocative inefficiencies at the auction. We extend Kastl (2011) to estimate these distortions, relaxing independent private values to accommodate trading externalities from post-auction bilateral trading. The framework applies broadly to settings where auction winners trade in secondary markets. Using Canadian auction and trading data, we ask how resale opportunities alter the consequences of two primary market levers: who is allowed to bid, and how much is sold.

JEL: D44, D47, G12, G28

Keywords: Auction, Intermediaries, Secondary markets, Primary Markets, Regulation, Monetary Policy

*Preliminary and incomplete. The views presented are those of the authors, not necessarily of the Bank of Canada. All errors are our own. Thanks to Nathalie Gimenes and Howard Smith for excellent discussions. James Chapman, Bank of Canada, Email:jchapman@bank-banque-canada.ca Eric Richert, University of Chicago and NBER, Email: richert.eric@gmail.com Yu Zhu, Renmin University of China, Email: zhuyuzlf57@gmail.com.

1 Introduction

Central banks face a fundamental choice in monetary policy implementation: how many reserves to supply to the banking system. By supplying abundant reserves, the central bank takes on a large balance sheet and pins the overnight rate near the deposit rate with minimal interbank trading. When reserves are scarce, the central bank operates with a smaller balance sheet, and banks actively trade in interbank markets to meet their liquidity needs, with the overnight rate determined by market forces inside a corridor defined by standing facilities. A large central bank balance sheet reduces banks' incentives for monitoring each other, can lead to crowding out of private investment opportunities and introduces costs to the government from maturity mismatch (Ennis and Keister, 2008; Gravelle et al., 2023; Selgin, 2019; Baglioni, 2024; Duygan-Bump and Kahn, 2026). A scarce reserves system avoids these costs but creates its own: rents accrue to banks with stronger relationships and better liquidity management, raising the central bank's costs, weakening its control of the overnight rate, and tilting profits toward the best-connected banks.

When reserves are scarce they are often allocated through a primary mechanism and then traded. In Canada, the central bank allocated additional reserves to achieve their interest rate targets through multi-unit auctions where banks submit demand schedules. After the auction, banks trade bilaterally in over-the-counter interbank markets characterized by search frictions, repeated interactions, and heterogeneous credit concerns (Afonso et al., 2011) to fine-tune their positions before the end of the day as customer payment flows materialize. The auction allocation determines demand and supply in the secondary market, creating strategic interdependencies across banks' valuations that shape bidding, the central bank's cost of providing liquidity and the distribution of surplus among banks.

The resale opportunities can drive bidders' valuations away from the social marginal value, and can affect strategic bidding. Although there is some work on resale in single-unit auctions (e.g., Haile, 2001, 2003), much less is known about the multi-unit case with imperfectly competitive resale markets.[1] The workhorse structural multi-unit auction models (Kastl, 2011; Hortaçsu and McAdams, 2010) assume independent private values, which rules out resale externalities. While interdependencies driven by resale may be small relative to

[1] Budish and Bhave (2023) study ticket markets with many units, and resale on secondary platforms, but the secondary markets are competitive. Pagnozzi and Saral (2017) use an experiment to evaluate how resale affects bidding at multi-unit auctions, finding evidence of more aggressive bidding by stronger participants.

idiosyncratic shocks in Treasury auctions Hortaçsu and Kastl (2012), post-auction resale in imperfectly competitive markets is a common feature of settings with multi-unit auctions.

This paper develops a structural model of multi-unit auctions followed by trading in imperfectly competitive secondary markets, and applies it to measure information rents and to evaluate central-bank policy levers in Canada's overnight lending market. We extend the framework of Kastl (2011) to allow bidders' valuations to depend on the full vector of auction allocations through expected post-auction trading opportunities. In our model, banks bid at a central bank auction for overnight liquidity, then trade bilaterally in an interbank market characterized by heterogeneous trading relationships. A bank's value for winning liquidity depends on its own reserve needs, but also on its trading relationships: winning more liquidity increases its ability to lend profitably to banks with whom it has favorable relationships, while winning less increases its need to borrow at potentially less favorable rates from banks with whom it has poor relationships. This creates strategic interdependencies in bidding behaviour that are absent in existing models.

The key empirical challenge is to separately identify banks' private information (forecasts of their reserve needs), the distribution of post-auction unanticipated payment shocks, and the rents banks expect to earn from trading. Unlike the standard-approach we cannot directly invert a first-order-condition to recover unobserved values from observed bids because the game may not have a monotone pure strategy equilibrium, and because the necessary conditions depend on an unobserved distribution.[2] Instead, we develop an identification strategy that exploits partial invertibility on an identified set. Intuitively, we use variation in the variances of observed positions conditional on bids to identify the set of bids that are generated by unique types. Inside this set, conditioning on the bid is equivalent to conditioning on private type, so all residual variation in the post-auction customer orders can be attributed to shocks that the bidder did not know when bidding. This lets us pin down the distribution of shocks and the distribution of private types. With these in hand, we can recover preferences for trading from necessary conditions of optimal bidding.

We apply our framework empirically using detailed microdata from Canada's afternoon Receiver General (RG) auctions and the Large Value Transfer System (LVTS). The RG auctions were used by the Bank of Canada to manage end-of-day settlement balances as part of its corridor system for monetary policy implementation, in which the overnight rate

[2]As a result, we do not pursue a simple extension of arguments such as Somaini (2020) from single unit auction settings with interdependence.

was targeted through a combination of standing facilities (providing a 50 basis point corridor around the target rate) and active liquidity management through twice-daily auctions. The afternoon auction, which we study, allocated unsecured overnight liquidity through a discriminatory pay-as-bid format, with banks subsequently trading bilaterally in the LVTS interbank market before the end of the day. We observe bid-level data from the auctions, including quantities and prices for each step of banks' step-function bids, as well as detailed payment flows in LVTS that allow us to construct measures of each bank's pre-auction liquidity position and post-auction trading activity. We use a modified Furfine algorithm Furfine (1999) to identify bilateral overnight loans in the interbank market, giving us a comprehensive picture of both primary and secondary market activity.

We first document substantial heterogeneity in secondary market trading opportunities across banks. Using the rates at which each bank borrows from each potential lender, we classify trading relationships into "partners" (banks that regularly trade at favorable rates) and "non-partners" (banks that trade infrequently or at less favorable rates). We find large and persistent differences in the rates charged across these relationships, with partner banks often receiving discounts of several basis points relative to the overnight target rate, while non-partner trades occur at or above the target rate. This price dispersion indicates that the secondary market is far from perfectly competitive, and that banks face heterogeneous costs of accessing liquidity post-auction. Second, we show that auction allocations have causal effects on subsequent secondary market trading, highlighting the central mechanism in our model. Using a matching estimator that compares banks with similar pre-auction positions and similar bids but different allocation outcomes, we find that receiving more liquidity at the auction leads a bank to lend more to both partners and non-partners and to lower the rates it charges partners. Third, we find evidence that bidding behaviors respond to measures of expected post-auction liquidity needs constructed from customer payment flows, suggesting that bidders have private information about their post-auction demands.

Our estimated model lets us decompose banks' values into utility from end-of-day balances, expected secondary-market payments, and a bonus from bilateral trading relationships. Three findings stand out. First, the marginal-value curves we recover are much flatter than those an independent-private-values model would imply: because banks anticipate reshuffling their positions in the secondary market, their willingness to pay at the auction varies little with the quantity they win. Second, the secondary market is critical to system

functioning: shutting it down sharply raises the dispersion of welfare across banks and without the secondary market the historical supply of reserves would have been insufficient to keep the overnight rate off the penalty rate. Third, the flatness of demand means the primary auction has little leverage over the price of liquidity or over who ultimately holds reserves: restricting participation or changing supply moves prices and allocations only slightly.

There is ongoing debate about the optimal level of settlement balances and the size of the central bank balance sheet required for liquidity management and monetary policy implementation. A large literature weights the costs and benefits of a lean balance sheet against the risks of scarcity Ennis and Keister (2008), Gravelle et al. (2023), Selgin (2019), Acharya and Rajan (2024) and Duygan-Bump and Kahn (2026).[3] A recent literature quantifies the level required using estimates of reserve demand, i.e., Afonso et al. (2022), and López-Salido and Vissing-Jorgensen (2023). Our analysis contributes to this estimation of reserve demand, using granular auction bidding data and secondary market trading data from the period of scarce reserves. We also highlight a novel distortion. When scarce balances are allocated through discriminatory auctions and secondary markets are imperfectly competitive, strategic bidding generates allocative inefficiencies and information rents that themselves depend on the supply of balances.

A key policy question for central banks is how much liquidity to provide, and whether broad or narrow participation better achieves efficient allocations and interest rate stability. We address both by comparing outcomes under doubled and halved auction supply, and under restricted participation in which only the five largest bidders may bid using the structural estimates of demand. For tractability, we abstract from strategic bid shading throughout: each bank submits its marginal valuation schedule rather than a strategically optimized bid function. Equilibrium is still a fixed point in which banks' marginal valuations, auction allocations, and expected secondary-market payoffs are mutually consistent. Our results quantify the price and distributional consequences of these alternative designs.

The remainder of the paper proceeds as follows. Section 2 describes the institutional details of the Canadian overnight lending market, the LVTS payment system, and the RG auctions. Section 3 describes our data sources and presents summary statistics on auction bidding and secondary market trading patterns. Section 4 develops our structural model of bidding with secondary market trading. Section 5 presents our identification strategy.

[3] Allen et al. (2021) suggest the monitoring role may be less important in the Canadian context, based on evidence from morning auctions during the global financial crisis.

Section 6 presents our estimation strategy. Section 7 discusses our estimates. Section 8 presents counterfactuals and discusses policy implications. Section 9 concludes.

2 Institutional Setting

2.1 Settlement Balances and the Payment System

When one bank's customer makes a payment to a customer at another bank, the transaction is ultimately settled through a transfer of central bank deposits (settlement balances) between the two institutions. These assets are attractive for settlement since they have known values and are easy to transfer.[4] The central bank controls both the aggregate supply of these balances and the rates paid on them to conduct monetary policy.

Banks can hold positive or negative balances during the day as they process incoming and outgoing payments, but at the end of the day a bank with negative balances must borrow from the central bank's lending facility at a penalty rate, while a bank ending the day with positive balances earns only the deposit rate, which sits below the lending rate. The lending and deposit rates thus form a corridor: during our study period the Bank of Canada set the deposit rate 25 basis points below its target for the overnight rate and the lending rate 25 basis points above. Because banks prefer to avoid both the penalty of being short and the opportunity cost of holding excess reserves, they target balances near zero. Since banks arrive towards the end of the day with different positions, there are gains from trade: an interbank market for overnight loans lets them borrow and lend excess reserves to other banks at a negotiated rate between the deposit rate and the penalty rate. The trades that occur and the rates at which they occur depend on the relative scarcity of balances and the bargaining power of participants.

Traditionally most central banks operated with scarce reserves. The central bank set a wide corridor around the target and actively managed supply of balances to hit that target, keeping a small balance sheet and incentivizing banks to monitor each other since overnight loans were risky. When central banks engaged in Quantitative Easing (QE) in 2020 they bought Treasuries from banks and credited them with settlement balances. As a result,

[4]For example, if Bank A's customer transfers $5 million to Bank B's customer, this payment is executed by debiting Bank A's account at the central bank by $5 million and crediting Bank B's account by the same amount.

the supply of settlement balances exploded, increasing by more than 1500 times in Canada Chu et al. (2022) and with all banks holding excess deposits the overnight rate fell to the deposit rate. This is the abundant reserve or floor system. The choice between the two regimes is, at its core, a choice about who reallocates reserves. Under a floor system, the central bank supplies so much that no bank needs to trade, but with scarcity, banks must redistribute supply in the interbank market. As balance sheets normalize, central banks are again grappling with the appropriate level of settlement balances. The answer depends on the forces we measure in this paper: the information banks have about their liquidity needs, and the degree of competition in the secondary market. Better information reduces precautionary demand for reserves, and more competitive secondary markets allow banks to share liquidity more effectively after the auction. Both reduce the aggregate buffer stock needed to prevent settlement failures or spikes in the overnight rate.

The appropriate level of reserves and the mechanism used to supply them, is an active question. This is a market design problem: how to design the primary mechanism through which reserves are allocated, and how much reserve reallocation to leave to a frictional secondary market. Different countries have reached different answers. The Bank of England supplies reserves through two repo operations: a short-term repo that lends any quantity banks demand at a Bank Rate, and a long-term repo auction in which both the price and the quantity are set competitively. Access to the short-term operation is deliberately limited so that banks meet routine needs in private markets and turn to the central bank only at the margin (Bank of England, 2024; Saporta, 2025; Hauser, 2023). The European Central Bank has a spread between its main lending operation and its deposit facility of fifteen basis points, so that banks will find it nearly costless to obtain reserves on demand through that operation rather than through secondary market trading (European Central Bank, 2024; Schnabel, 2024). The Reserve Bank of Australia supplies as many reserves as banks demand through full-allotment repo auctions priced just above the cash-rate target (Kent, 2024, 2025). The Federal Reserve operates with ample reserves (Perli, 2025).[5] The Bank of Canada, recently committed to a floor, but set its deposit rate below the target policy rate partly to encourage settlement balances to circulate among banks as reserves decline (Bank of Canada, 2025b,a). Each of these is a choice of market design for reserve allocation

[5]Even when reserves were scarce the Federal reserve did not use an auction mechanism like Canada and the euro area. Instead, they adjusted the aggregate supply primarily through open-market operations conducted by the New York Fed (Perli, 2025).

which our two-stage model of a primary auction followed by frictional secondary trading is built to evaluate. Although the allocation mechanism differs across these systems, the economic fundamentals we study: private information about liquidity needs, bilateral trading relationships, and the distributional consequences of the reserve-supply regime, are common.

2.2 The Canadian Large Value Transfer System

The segment of the Canadian interbank market we study operated within the Large Value Transfer System (LVTS). During our study period (2016-2020), this system supported the Bank of Canada's monetary policy implementation as well as the Bank's role as fiscal agent for the Canadian federal government. Established in 1999, the LVTS was a real-time settlement system used for large-value or time-critical payments among direct participants providing real-time finality of payments. It served as a key conduit for monetary policy transmission and provided the infrastructure for settling interbank obligations with finality using central bank balances (Zhang, 2012). The LVTS was operated by Payments Canada and overseen by the Bank of Canada, to handle high-value transactions efficiently while managing settlement risk (Bech et al., 2013).

The Bank of Canada implemented monetary policy through the corridor system described in the previous section, and managed liquidity using twice daily Receiver General (RG) auctions. These auctions are the primary mechanism through which the Bank of Canada adjusts the supply of settlement balances, and they also manage the fiscal position of the Government of Canada, for which the Bank acts as fiscal agent (Chapman, 2006).

Trading in the interbank market is concentrated in the morning hours following an AM auction, as banks begin processing customer payments and adjusting their reserve positions. Activity drops off substantially in the afternoon as banks respond to the liquidity injections and withdrawals (Bech et al., 2013). During the afternoon, the Bank announces its target level of settlement balances prior to the PM auction. Given typical net inflows into government accounts throughout the day, this announcement effectively determines how much liquidity the PM auction will add to the system. After the PM auction, customer payments continue to arrive, and the banks trade with each other. The end of day position of the banks is the sum of the pre-PM auction position, quantity won at the auction, trades made with other banks and the customer orders.

The PM RG auction, which is the focus of our analysis, is an uncollateralized discrim-

inatory price multi-unit auction conducted among LVTS participants. Banks submit bids specifying both the quantity they wish to borrow and the interest rate they are willing to pay. The auction is typically oversubscribed. The Bank of Canada allocates the available supply by accepting bids starting with the highest rates and moving down until the total auction amount is exhausted. Crucially, this is a pay-as-bid auction: each winning bidder pays the rate they specified rather than a uniform clearing price. The discriminatory auction features strategic considerations in bidding, as banks must balance the probability of winning against the cost of bidding too aggressively. Unlike Treasury auctions, there is no opportunity to submit non-competitive bids.

When banks formulate their bidding strategies, they face substantial uncertainty. Each participant knows its own current position in settlement balances, and has private information about its expected payment obligations for the remainder of the day. However, banks do not observe their competitors' current balances or who else is participating in the auction. Based on their private information, each bank must determine how much liquidity to request and what rates to offer, trading off the cost of bidding aggressively against the risk of being rationed at the auction and needing to borrow later at potentially less favorable terms.

Following the auction, banks can turn to the interbank market to borrow or lend any additional amounts needed to adjust their positions. This secondary market provides an important margin of adjustment, allowing participants to respond to actual payment flows that may differ from expectations and to realize gains from trade based on their auction allocations and secondary market relationships. The interbank market also shapes the value of bidders at the auction. It may reduce their values by providing an outside opportunity to obtain liquidity, should they be short later (reducing precautionary motives for trading at the auction). However, it also provides banks with an opportunity to resell excess supply, reducing the risk of obtaining a large position.

While our analysis focuses on the Canadian LVTS during 2016-2020, the institutional features and economic trade-offs we study apply broadly to modern payment systems and central bank operating frameworks. In 2021, Canada replaced the LVTS with Lynx, which settles transfers immediately as long as liquidity is available; the need for liquidity to absorb customer payment shocks remains under the new system. Since we have been in an ample reserve regime since 2020, the RG PM auctions were suspended. However, as quantitative tightening drained the supply of reserves, overnight repo operations were used to relieve

upward pressure on the overnight rate by increasing the supply of settlement balances (Bank of Canada, 2024). Should Canada return to scarce reserves, afternoon auctions, or other mechanisms lead by the central bank could again be useful for targeting the policy rate.

2.3 Economic Distortions and Policy Concerns

The market structure of the settlement balances market has important implications for both allocative efficiency and the distribution of rents in the banking system. Central banks face a fundamental trade-off in setting the supply of settlement balances, with different regimes creating distinct economic distortions that affect both the real economy and the distribution of surplus between the government and banks. Several pieces of this trade-off are summarized by Duygan-Bump and Kahn (2026). The distortions of an ample reserve regime are largely balance-sheet phenomena, whereas the distortions in the scarce reserve system relate to primary and secondary market rents, that our model measures.

In regimes with ample reserves, the central bank supplies settlement balances far in excess of what banks need for settlement. This eliminates liquidity risk and pins overnight rates at the deposit rate, but it introduces inefficiencies. First, because banks differ in how profitably they can deploy funds, forcing them to hold large quantities of low-yielding central bank deposits misallocates resources and can crowd out private investment. Second, a large central bank balance sheet reduces the supply of Treasuries available to private investors. Because Treasuries are widely used as collateral, this shortage can make specific issues "special" and distort repo and related rates. Third, since large-scale asset purchases are typically funded by creating reserves, which are effectively overnight liabilities, they transfer duration risk from the private sector to the government, which bears it through central bank remittances.[6] Such transfers may be acceptable during financial crises, but raise questions about the appropriate size of the central bank balance sheet in normal times.

The alternative regime, operating with scarce reserves creates a different set of distortions. In this environment, the central bank supplies just enough reserves for the system to function smoothly, leading to active trading in the interbank market. When reserves are scarce, banks which better manage their liquidity, or with better access to interbank trading networks can

[6]Consider $200 billion in 10-year government bonds with a duration of about 7 years, funded by $200 billion in overnight reserves. A one-percentage-point rise in rates cuts the bond portfolio's value by roughly $15 billion, while the cost of the reserve liabilities rises immediately with the deposit rate.

extract substantial information rents through strategic bidding at the auction combined with advantageous post-auction trading.

In the absence of a secondary market, each bank would face a steep, bank-specific demand curve for reserves. A bank expecting to need additional liquidity later in the day would need to bid aggressively at the auction since it has no alternative source of funds. This creates strong demand and pushes up both auction prices and quantities. The secondary market changes this. A sophisticated bank with good trading relationships can afford to bid less aggressively at the auction, knowing it can access the interbank market later if needed. Banks can also use their private information about their own liquidity needs to extract rents at the auction. By bidding strategically, banks pay less on average for reserves than would be necessary under perfect information. Since the government (through the central bank) is effectively the seller of reserves, these information rents represent a transfer from the fiscal authority to the banking sector.

The magnitude and distribution of these rents depends on the structure of the interbank market. The interbank market exhibits substantial heterogeneity in trading relationships that interacts with these rent extraction opportunities. As we document below, some bank pairs trade frequently and on favorable terms while others rarely interact. Banks with many strong relationships obtain better terms both when borrowing (paying lower rates) and when lending (charging higher rates to non-partners), creating dispersion across banks in the rents they extract.

The distribution of rents has consequences for competition and efficiency. If certain banks systematically extract more rents depending on the level of balances, the regime effectively subsidizes them. This raises several questions: Is the distribution of rents roughly proportional to banks' contributions to payment system efficiency, or does it instead reflect market power and barriers to entry in interbank trading? How do changes in the supply of reserves affect this distribution, and does reduced supply concentrate rents among a smaller set of banks? Could alternative auction or allocation mechanisms achieve better distributional outcomes while maintaining allocative efficiency? Our empirical analysis speaks to these questions by measuring the precision of banks' private information about their liquidity needs, quantifying the role of secondary markets in improving allocative efficiency, and assessing how the distribution of rents varies with the supply of reserves and the set of banks admitted to the auction.

3 Data

We combine two distinct data sources in our empirical analysis. First, we have payment data among LVTS participants. Using this data, we have imputed unsecured overnight loans between LVTS participants using a modified Furfine algorithm Allen et al. (2016). Second, we use bidding data from the Receiver General's afternoon (PM) auctions.

3.1 LVTS Payment Flows and Secondary Market Trading

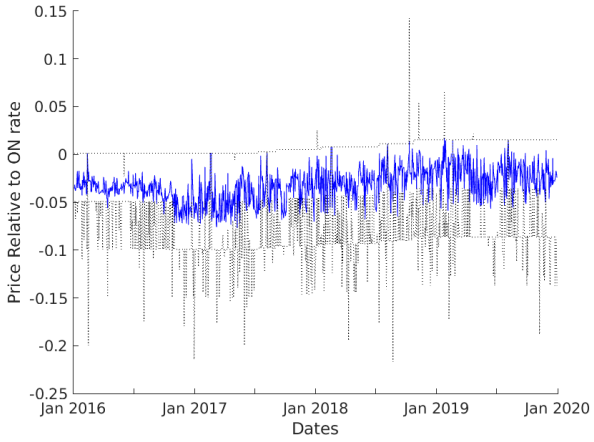
We use payment data from LVTS to calculate interbank loans and liquidity needs of LVTS participants. This dataset consists of daily bank transactions observed in LVTS. At the time of the study, there were 17 direct participants in LVTS. These are the Big 6 Canadian banks (Banque Nationale, Bank of Montreal, Bank of Nova Scotia, Canadian Imperial Bank of Commerce, Royal Bank of Canada, Toronto-Dominion Bank), HSBC, ING Canada (CHECK), Laurentian Bank, State Street Bank, Bank of America, BNP Paribas, Alberta Treasury Branches, Caisse Desjardins, and a credit union consortium (Central 1 Credit Union).

As in other wholesale payment systems, LVTS participants send and receive payments throughout the day. We observe who sends payments to whom, the amount of each payment, and the time of each payment. Using this information, we track the level of liquidity for each direct participant at three key points in time. First, we observe the balance for each participant ahead of the PM RG auction. Second, we calculate the balance immediately after the PM RG auction by including the amounts won by each participant. Third, we observe the balance at the end of the day. Since we use the Furfine algorithm to impute interbank loans, we use these balances to calculate the amount of interbank payments between direct participants from the end of the PM RG auction to the end of the day.

The distribution of secondary-market rates is stable across days. Figure 1 plots the average rate from all trades in the secondary market every day (blue solid line) along with the 10th and 90th percentiles (black dotted lines). The Bank of Canada's monetary policy framework operated with a 50 basis point corridor around the policy rate. On almost all days, the 90th percentile of secondary market trading occurs at the overnight rate, which represents the upper bound of this corridor. However, within day there is substantial variation between the 10th and 90th percentile of secondary market prices within the targeted range.

Figure 1: Secondary Market Rates

Figure 1 provides the secondary market average rate, and the 10th and 90th percentile of secondary market rates traded at each day during the sample.



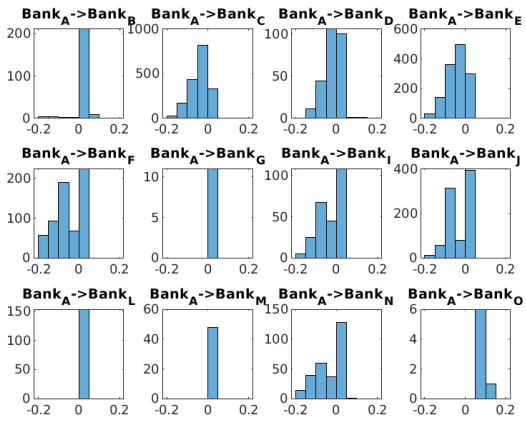
3.1.1 Heterogeneous Trading Relationships

To understand the patterns of trading in the secondary markets, we examine heterogeneity in the prices that banks charge when lending to different borrowers. Figure 2 plots the distribution of rates used by two banks when lending to each other bank. There is large heterogeneity in the frequency with which banks trade with each other. Looking at the left panel, Bank A lends frequently to some banks (e.g., Bank C) and rarely to others (e.g., Bank G). There is also large heterogeneity in the rates, with some banks often receiving discounts relative to the overnight rate, and others being charged at or slightly above that rate.

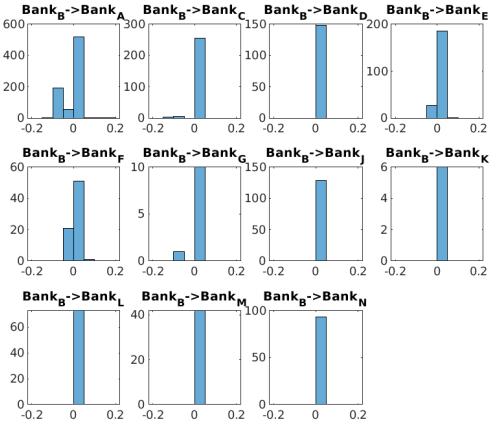
Given this pattern, we split each bank's trading counterparties into two categories: *partners* and *nonpartners*. Discount is defined as overnight rate minus observed traded rate. So lower rates mean larger (positive) discounts. Then a weighted average of discounts is calculated per pair of lenders and borrowers. Preferred counterparties are those that are in the top 75% of average discounts for the borrower. This bank-pair-specific definition controls for borrower-specific factors (such as counterparty risk) in pricing. For example, if Bank A typically lends to Bank B at favorable rates relative to Bank A's other borrowers, we classify B as a partner of A. Table 1 lists summary statistics of the resulting set of preferred lending and borrowing relationships. In many cases, partners lend and borrow at similar discounts relative to the overnight rate. Partners lend and borrow at discounted rates relative to the non-partner banks. The within-partner dispersion is also smaller than the gap between

Figure 2: Lending Rates

Figure 2 plots histograms of the lending rates in the secondary market. Panel (A) indicates the rate when Bank A lends to each other bank. Panel (B) indicates the rate at which Bank B lends to each other bank.



Panel (A)



Panel (B)

partner and non-partner average rates, suggesting that the simple classification captures an important margin of the pricing patterns. There is large heterogeneity in the number of partners and in the discounts that different banks are able to obtain.

This price dispersion means that there is substantial heterogeneity in the post-auction costs of accessing liquidity and the post-auction outside value for excess funds that they acquire. There could be many drivers of this heterogeneity, for instance credit risk of the borrowing bank or repeated relationships (from interactions both within this market and across other markets see Chiu et al. (2020), Afonso et al. (2011)). While understanding the sources of this price dispersion is interesting, we do not attempt to explain the sources of these prices. Instead, we ask how the existing secondary market structure impacts information rents, price volatility, and allocative inefficiencies in the primary auction.

3.2 Receiver General Auctions

We use bid-level data from the afternoon (PM) auctions conducted by the Receiver General for Canada. The dataset consists of bid-level data for the RG PM auctions from 2016 to 2020. For each bid, we observe the bid amount and bid yield as well as the amount allocated. If more than one bid is submitted at the market clearing price, bids are rationed pro-rata at the margin. The auctions were conducted among LVTS participants.

Table 1: Table 1: Partner Rates

Table 1 provides summary statistics on the partner relationships. Panel A reports the cross-bank distribution of trading volume and mean rates for the preferred-partner and non-partner overnight loans. Panel B reports the within-group rate dispersion.

Panel A: Rate Levels and Trading Volume

	Lending				Borrowing			
	Partner		Non-partner		Partner		Non-partner	
	N	Rate	N	Rate	N	Rate	N	Rate
Mean	3	-0.0232	7	0.0107	3	-0.0269	7	0.0239
p_{10}	1	-0.0681	2	-0.0038	1	-0.0643	1	-0.0033
p_{90}	6	0.0085	11	0.0492	3	0.0077	10	0.0338

Panel B: Within-group Rate Dispersion (σ)

	Lend Partner	Lend Non-partner	Borrow Partner	Borrow Non-partner
Mean	0.0134	0.0219	0.0154	0.0226
p_{10}	0.0000	0.0007	0.0000	0.0000
p_{90}	0.0302	0.0695	0.0355	0.0513

3.2.1 Auction Bidding

Table 2 describes the auctions. On average, 4 bidders participate in each auction. The average quantity supplied is \$750M, which is similar to the total volume of customer trades post-auction (\$690M) but is considerably smaller than the post-auction trades with partners (\$3B) or non-partners (\$1.3B). It is very common for bidders to submit steps at the floor rate (the rate they would earn for deposits at the central bank). This is both the minimum price and the modal price submitted. On average, the auction clearing price is below the secondary market average price. This suggests a potentially profitable strategy: buy more at the auction and then lend the additional supply for a profit later. However, the heterogeneity in post-auction trading opportunities may make this unprofitable.

Table 3 presents summary statistics for bidders' balances, quantities bid, and prices bid at the auctions. There is substantial heterogeneity in the rates bid across bidders. Bidders submit bids for a large fraction of the total supply. The average quantity bid is much larger than the quantity restrictions in other markets such as Treasuries, which are designed to prevent participants from cornering the market. Average balances pre-auction are close to zero, but participants tend to have slightly larger balances than non-participants. This may be surprising since banks with negative balances appear to have a greater need for cash. However, both the pre-auction balance and demand at auction may be driven by

Table 2: Auction Summary Statistics

Table 2 provides summary statistics on the auctions. The N-bidders is the number of bidders, total supply the volume issued at the auction, the total customers the total post-auction payments made by a bank excluding lending/borrowing, partner/non-partner trade volume the sum of absolute value of trades with partners/non-partners. The price statistics indicate the minimum price, maximum price modal price and standard deviation of all the prices bid at the auction, and the auction price-SM price the auction price minus the secondary market price on that day.

	mean	std	p10	p50	p90
N active	3.695	1.590	2	4	5
Total Q	0.751	0.407	0.360	0.684	1.215
Total Cust	-0.690	0.426	-1.202	-0.655	-0.158
Total partner trades	2.564	1.161	1.107	2.501	4.079
Total nonpartner trading	1.730	0.838	0.815	1.596	2.8222
Price					
Minimum	-0.238	0.035	-0.250	-0.250	-0.220
Maximum	-0.036	0.058	-0.120	0.000	0.010
Std	0.085	0.028	0.047	0.088	0.115
Clearing	-0.126	0.091	-0.250	-0.140	0.000
SM HHI lend	4507				
SM HHI borrow	4028				

anticipated post-auction needs for liquidity rather than current positions alone. Despite repeated interaction collusion in these auctions would be difficult as bidders do not learn the bids nor the allocations of other participants.

One striking feature of the bidding data is that many bidders use a small number of bid steps. However, there are a few participants that regularly submit multi-step bids. Figure 3 presents average prices per-steps, with the size of the dot reflecting the frequency with which the bidder uses that number of steps. The variation in bidding strategies across banks may be driven by heterogeneity in information about post-auction needs or opportunities in the secondary market.

3.2.2 Private Information About Post-Auction Needs

To evaluate whether auction participation and bidding decisions are impacted by post-auction customer orders and bidders' own balances, we regress participation and quantity purchased at the auction on post-auction customer orders, pre-auction balances, and a constant. Customer orders capture the incoming or outgoing payments from the bank's customers after the auction closes. Table 4 presents the results, which suggest that bidders have some information about their post-auction shocks. Customer orders are an important deter-

Table 3: Bidder Summary Statistics

Table 3 provides bidder level summary statistics. These summarize the bidding behavior for each auction participant. The prices are high/low prices, the share of demand bid is the sum of quantities bid by i divided by the total supply, the max bid-trade is preferred lending price minus the price bid by that bidder at the auction. The preauction balance and end of day balance summarizes the balance before the auction starts and at the end of the day respectively of all banks, while the same variables for only bidders, summarize these balances for the subset of banks that submit a bid at the auction stage.

	mean	std	p10	p50	p90
N bids	1.107	0.486	1	1	1
Max. bid (bps)	-11.503	9.124	-25.000	-10.000	0.000
Min. bid (bps)	-14.775	9.726	-25.000	-17.000	0.0000
Partner lend rate (bps)	-4.592	5.815	-9.923	-4.163	1.128
Non-partner lend rate (bps)	0.569	6.016	-4.283	0.281	1.536
Partner lend rate ($/night/$1M)	-1.258	1.593	-2.719	-1.140	0.309
Non-partner lend rate ($/night/$1M)	0.156	1.648	-1.173	0.077	0.421
Share of demand	0.496	0.380	0.047	0.392	1.000
Pre-auction balance ($B)	0.000	1.763	-1.379	0.000	1.192
EOD balance ($B)	-0.000	0.151	0.000	0.003	0.067
Pre-auction balance bidders ($B)	-0.711	1.725	-2.807	-0.408	0.788
EOD balance bidders ($B)	0.013	0.056	0.000	0.002	0.023
Max. bid minus pref rate (bps)	9.459	10.510	-5.171	11.365	20.896

Figure 3: Bids Across Steps

Figure 3 plots average bids within each step, with the size of the dot indicating the frequency with which that bidder submitted that many bids.

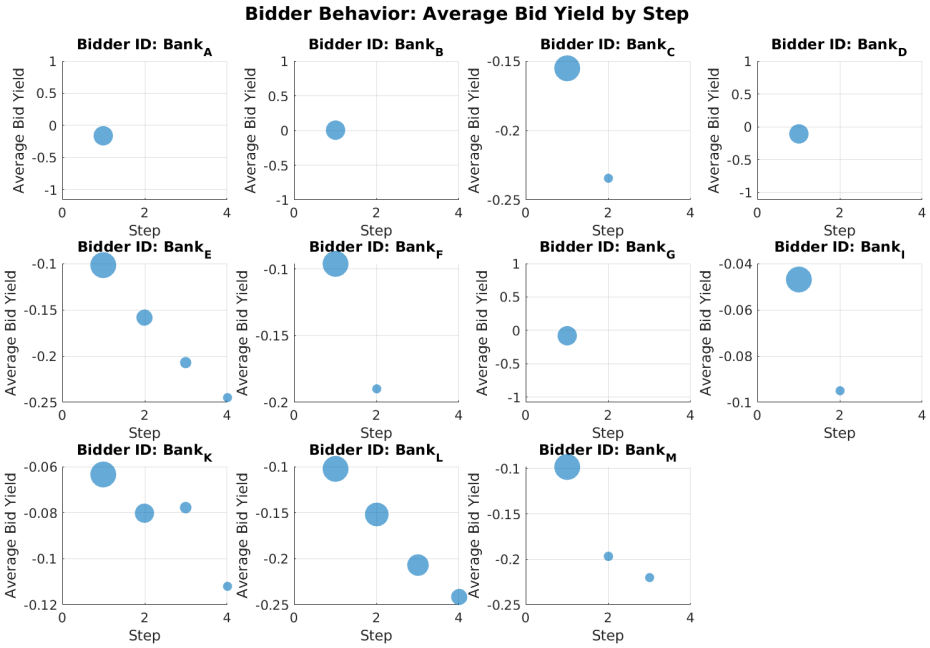


Table 4: Regression: Private Information

Table 4 provides a test for the role of private information. The first column shows estimates from regressing Participation on customer orders post-auction (observed orders not part of trading), and their pre-auction balance. There is one observation per LVTS member per day. The second column repeats this regression but replaces the binary dummy variable for participation by the Quantity won by the bidder at the auction.

	Participation	Q-won
Customer Orders	-0.007	-0.011
	(0.0049)	(0.0016)
Balance pre-auction	-0.057	-0.028
	(0.0059)	(0.0020)
Constant	0.214	0.040
	(0.0031)	(0.0010)

minant of quantity won at the auction, with bidders that anticipate receiving large incoming payments from customers bidding less. This suggests that bidders have some information on what these orders will be. This relationship may also decrease participation probability, but that effect is not statistically significant. In addition, bidders with large pre-auction balances are both less likely to participate and win less conditional on participation.

3.3 How Auction Allocations Impact Secondary Market Trading

We want to identify the impact of being allocated additional funds at the auction on a banks' lending and borrowing in the secondary market. To do this, we use a matching estimator. For every bidder that was allocated funds, we find the nearest neighbour based on Euclidean distance in standardized bid prices and pre-auction balances among bidders that were not allocated. Similarly, for every bidder not allocated, we find a nearest neighbour among those that were allocated. We then regress measures of post-auction trading behaviour-quantity traded with partners and non-partners, and average price traded at among partners and non-partners, on an indicator for being allocated (the treatment effect of interest) and a set of controls, including the balances of trading partners and the customer shocks faced by the bank post-auction.

The key identifying assumption is that a bidder with a similar pre-auction position that chose a similar bid expected similar market conditions post-auction, and whether the bid was allocated or not is an exogenous random event. Of course, allocations could be partially driven by market conditions (e.g., positions of partner banks) which also influence post-auction trading. We control for these effects separately in the regression to isolate the effect

Table 5: Impact of the Auction Allocation

Table 5 provides regression evidence to measure the impact of winning bonds at the auction on the amount that a bidder trades with their partners, with nonpartners and the price at which these trades occur. Partner/Nonpartner quantities are the total quantities traded with partners/nonpartners post-auction and partner/non-partner prices the average prices of those trades. There is one observation for each bidder that participated in an auction. In the regression, each bidder is matched to a similar bidder based on the nearest-neighbor match in euclidian distance of standardized bid prices and pre-auction balances that is (is not) allocated for bidders that were not (were) allocated and treated is an indicator for whether the bid was allocated. Customer shocks are the post-auction non-trading volumes for bank i, and partner balances the sum of the pre-auction balances of all i's partners.

	Partner Q	Nonpartner Q	Partner Price	Nonpartner Price
Bid Allocated	0.038	-0.005	-0.054	-0.000
	(0.0203)	(0.0058)	(0.0191)	(0.0195)
Customer Shocks	-0.089	-0.009	0.025	0.022
	(0.0037)	(0.0011)	(0.0030)	(0.0039)
Partner Balances	-0.071	0.004	0.003	-0.020
	(0.0029)	(0.0008)	(0.0025)	(0.0031)
Constant	0.557	0.118	1.066	1.188
	(0.0169)	(0.0048)	(0.0173)	(0.0157)

of the allocation.

Table 5 presents the results of this exercise. Being allocated more funds at the auction increases the amount that the bidder lends to both partners and non-partners, and decreases the average price of partner loans. This provides direct evidence of an important externality that is likely a significant force when auctions are followed by resale opportunities, but that is assumed away in conventional auction models. The coefficient on partner balances also provides direct evidence of interdependencies in values: if a bidder could learn its partners' balances before bidding, that would change the expected quantity it can access in the secondary market and hence its value for buying funds at the auction. The important role we find for auction allocations in shaping post-auction trade, motivates our extension of existing multi-unit auction models, to incorporate interactions from resale markets.

4 Model

There are N banks that interact in a two-stage game. In the first stage, bank i is endowed with a reserve holding Q_i for $i = 1, 2, \cdots, N$, which is accumulated throughout the day until the time when an auction is held. Bank i also knows that it needs to pay s_i in the second stage of the game, where s_i is negative if bank i expects to receive payments. Both Q_i and

s_i are banks' private information and we denote $\sigma_i = s_i + Q_i$. Bank i submits up to four price-quantity pairs in the auction. Denote the quantity alloted to bank i by Q_i^a. At the beginning of the second stage, bank i realizes that it needs to pay an additional amount of ϵ_i on behalf of their customers. Then banks borrow and lend reserves among themselves to ensure that the reserve holdings at the end of day to reach a value close to 0. Let χ_{pi}^b and χ_{ni}^b be quantity borrowed from partner and non-partner banks. Let χ_{pi}^l and χ_{ni}^l be quantity lent to partner and non-partner banks. Then at the end-of-the-day balances of bank i is $Q_i^e = \sigma_i + Q_i^a + \chi_{pi}^b + \chi_{ni}^b - \chi_{pi}^l - \chi_{ni}^l + \epsilon_i$. We assume that σ_i and ϵ_i are independent of each other, independent across banks, and follow distributions F_{σ_i} and F_{ϵ_i}, which can vary across banks. Throughout the paper, we use bold symbols to denote the vector of variables from all the banks and use subscript $-i$ to denote the vector of variables from banks other than i. For example, $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \cdots, \sigma_N)$ and $\boldsymbol{\sigma}_{-i} = (\cdots, \sigma_{i-1}, \sigma_{i+1}, \cdots)$.

Bank i aims to maximize its utility at the end of the second stage, which is

$$V_i^e = u(Q_i^e) + \sum_{\substack{j\in\{n,p\}\\ m\in\{b,l\}}} B_{ji}^m \chi_{ji}^m + \sum_{j\in\{n,p\}} i_{ji}^l \chi_{ji}^l - \sum_{j\in\{n,p\}} i_{ji}^b \chi_{ji}^b, \tag{1}$$

where i_{pi}^l and i_{pi}^b are lending and borrowing rates in trades with partner banks, and i_{ni}^l and i_{ni}^b the rates with non-partners. We assume that u is differentiable: if the underlying u is not differentiable at 0, we approximates it with a differentiable function that is "close enough". We also allow banks to value trades with partners and non-partners differently. The parameter $\boldsymbol{B}_i = \left(B_{ni}^b, B_{pi}^b, B_{ni}^l, B_{pi}^l\right)$ captures bank i's additional utility from borrowing and lending with non-partners and partners, respectively, which could be a result of dynamic considerations from the continued relationship benefits.

We are agnostic about how banks borrow and lend in the second stage except imposing the following weak assumption.

Assumption 1. *The joint distribution of χ_{ji}^m, $j \in \{n,p\}$ and $m \in \{b,l\}$ depends on $\boldsymbol{\sigma}$, $\boldsymbol{Q}^a$, $\boldsymbol{\epsilon}$ only through $\boldsymbol{\sigma} + \boldsymbol{Q}^a + \boldsymbol{\epsilon}$.*

This assumption requires the distributions of borrowing and lending quantities depend only on balances of banks at the end of the day excluding borrowing and lending. One may also model the second stage trading process following, e.g. Alfonso and Lagos (2015).[7] For

[7]The model propsed by Alfonso and Lagos (2015), for example, is consistent with this assumption.

our purpose, i.e., understanding how post-auction opportunities affect auction bidding, the simple formulation suffices.

In the first stage, bank i bids a schedule to maximize its expected payoff, accounting for trades in the second stage. To derive the payoff from bidding, first define the expected payoff of bidder i given all bidders' alloted quantities, $(Q_i^a, \boldsymbol{Q}_{-i}^a)$, and private information $(\sigma_i, \boldsymbol{\sigma}_{-i})$:

$$V_i\left(\sigma_i + Q_i^a, \boldsymbol{\sigma}_{-i} + \boldsymbol{Q}_{-i}^a\right) = \mathbb{E}\left[V_i^e \middle| \sigma_i + Q_i^a, \boldsymbol{\sigma}_{-i} + \boldsymbol{Q}_{-i}^a\right] = \mathbb{E}\left[V_i^e \middle| Q_i, Q_i^a, s_i, \boldsymbol{Q}_{-i}, \boldsymbol{Q}_{-i}^a, \boldsymbol{s}_{-i}\right], \tag{2}$$

where the expectation is taken over the unexpected post-auction trade flows of all participants ϵ_i and $\boldsymbol{\epsilon}_{-i}$, and the post-auction trade quantities χ_{ji}^m, $j \in \{n, p\}$ and $m \in \{b, l\}$. The second equality holds because V_i depends on Q_i, Q_i^a, s_i and $\boldsymbol{Q}_{-i}, \boldsymbol{Q}_{-i}^a, \boldsymbol{s}_{-i}$ only through $Q_i + Q_i^a + s_i$ and $\boldsymbol{Q}_{-i} + \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}$.

This payoff is different from the ones commonly used in multi-unit auctions in that it depends on other bidders' winning quantities. This is a feature introduced by the secondary market trading χ_{ji}^m, $j \in \{n, p\}$ and $m \in \{b, l\}$, and the fact that there are only a few participants in the secondary market. If bank i has private information σ_i and bids $y_i = \{(b_{ik}, q_{ik})\}_{k=1}^{K_i}$, where K_i is the total number of steps, b_{ik} decreases with k and q_{ik} increases with k, its expected profit is

$$\begin{aligned}
&\mathbb{E}[\pi_i(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i})|\sigma_i, y_i, \boldsymbol{y}_{-i}(\cdot|\cdot)] \\
= \mathbb{E} \int \Big\{ &\sum_{k=1}^{K_i} \left[\mathbf{1}(b_{ik} > p^c > b_{ik+1}) V_i(q_{ik} + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i}) - \mathbf{1}(b_{ik} > p^c) b_{ik}(q_{ik} - q_{ik-1})\right] \\
&+ \sum_{k=1}^{K_i} 1(b_{ik} = P^c)[V_i(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})]\Big\} d\boldsymbol{F}_{\boldsymbol{\sigma}_{-i}}(\boldsymbol{\sigma}_{-i}).
\end{aligned}$$

Here $\boldsymbol{y}_{-i}(\cdot|\boldsymbol{\sigma}_{-i})$ is the bidding schedule of bidders other than i given their private information $\boldsymbol{\sigma}_{-i}$, which specifies the quantities as a function of price; $\mathbb{E}$ in the first line denotes the expectation taken with respect to Q and $\boldsymbol{\sigma}_{-i}$ while $\mathbb{E}$ in the second line is taken with respect to Q; and p^c is the clearing price in the auction. Notice that $(Q_i^a, \boldsymbol{Q}_{-i}^a, p^c)$ is a vector determined by Q, y_i and $\boldsymbol{y}_{-i}(\cdot|\boldsymbol{\sigma}_{-i})$. Bidder i does not know the realizations of Q and $\boldsymbol{\sigma}_{-i}$ and therefore needs to form expectation about them. Unlike the standard multi-unit auctions without the secondary market, the expected profit of the bidder i depends on other bidders' private information and alloted quantities.

We follow Kastl (2011) to consider a local deviation from q_{ik} to $q'_{ik} = q_{ik} - \epsilon$ and characterize the necessary conditions for the equilibrium bidding. Let $V_{i1}(\cdot,\cdot)$ be the partial derivative of $V_i(\cdot,\cdot)$ with respect to the first argument and $V_{i2}(\cdot,\cdot)$ be the vector of partial derivatives of $V_i(\cdot,\cdot)$ with respect to every element in the second argument. Moreover, if $\boldsymbol{z}$ is a vector and x is a scalar, we use $\partial \boldsymbol{z}/\partial x$ to denote the vector of derivatives of each element of $\boldsymbol{z}$ with respect to x and use $\cdot$ to denote the inner product of two vectors. The following result holds, where all probabilities and expectations are implicitly conditioning on σ_i, y_i and $\boldsymbol{y}_{-i}$. In the Appendix, we prove the proposition for a payoff function that allows for a more general form of interdependence, but here we focus on the utility function from our model.

Proposition 1. *In an equilibrium, the following necessary conditions hold:*

1. *If $k < K_i$, then*

$$
\begin{aligned}
&\mathbf{Pr}(b_{ik} > p^c > b_{ik+1})\times \\
&\left\{\mathbb{E}\left[V_{i1}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}) + V_{i2}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} | b_{ik} > p^c > b_{ik+1}\right] - b_{ik}\right\} \\
&+\mathbf{Pr}(b_{ik} = p^c, tie)\times \\
&\mathbb{E}\left\{V_{i1}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\frac{\partial Q^a_i}{\partial q_{ik}} + V_{i2}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik}\frac{\partial Q^a_i}{\partial q_{ik}}\middle| b_{ik} = p^c, tie\right\} \\
&+\mathbf{Pr}(b_{ik+1} = p^c, tie)\times \\
&\mathbb{E}\left\{V_{i1}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\frac{\partial Q^a_i}{\partial q_{ik}} + V_{i2}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik+1}\frac{\partial Q^a_i}{\partial q_{ik}}\middle| b_{ik+1} = p^c, tie\right\} \\
&=\mathbf{Pr}(p^c \le b_{ik+1})\,(b_{ik} - b_{ik+1})
\end{aligned}
$$

2. *If $k = K_i$,*

$$
\begin{aligned}
0 =&\mathbf{Pr}(b_{ik} > p^c)\times \\
&\mathbb{E}\left[V_{i1}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}) + V_{i2}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik}\middle| b_{ik} > p^c\right] \\
&+\mathbf{Pr}(b_{ik} = p^c)\times \\
&\mathbb{E}\left[V_{i1}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\frac{\partial Q^a_i}{\partial q_{ik}} + V_{i2}(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i})\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik}\frac{\partial Q^a_i}{\partial q_{ik}}\middle| b_{ik} = p^c\right].
\end{aligned}
$$

Proof. See Appendix A. □

Proposition 1(1) describes the FOC with respect to the quantity for the non-terminal steps, while Proposition 1(2) describes the FOC for the last step. They differ because the trade-offs differ between the non-terminal steps and the last steps: reducing the quantity at a non-terminal step does not change the maximum quantity that a bidder could possibly win, but just moves the quantity from one step to the next. However, reducing the quantity at the last step reduces the maximum quantity that a bidder could win.

Secondary market trading complicates the auction stage in two ways. First, bidders understand that they have another opportunity to adjust their positions and to what extent they can do so depends on other bidders' positions. Therefore, a bidder takes into account the effect of its bid on other bidders' positions, which is captured by the term $V_{i2}(q_{ik} + \sigma_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{\sigma}_{-i}) \cdot \partial \boldsymbol{Q}^a_{-i} / \partial q_{ik}$. In principle, secondary market trading could increase or decrease the competitiveness in the bidding stage. On one hand, bidders can get liquidity in the secondary market even if they did not get enough in the auctions. This reduces bidders' incentives to bid aggressively. On the other hand, bidders can offload liquidity in the secondary market if they get too much in the auctions, which increases their incentives to bid aggressively. Which effect dominates depends on conditions in the secondary market.[8] Second, we cannot rule out ties at middle bidding steps as in Kastl (2011). Third, the first-order conditions also have a flavor of common value auctions: The clearing price affects a bidder's expected marginal value because it contains information on other bidders' signals, which affect trading in the secondary market.

5 Identification

Our data contain all bidders' balances at each point in time, all the information on auctions including bids, quantity allotted, the clearing price, and all transfers after the auction. Transfers after the auction have two parts: payments on behalf of customers, which is $\omega_i = s_i + \epsilon_i$ in our model, and borrowing and lending for liquidity management, which we identify using the method proposed by Furfine (1999). For the transfers that are identified as lending and borrowing, we also obtain lending rates. Therefore, we observe the distribution

[8]This is related to Wright et al. (2020) and Cui et al. (2025) who investigate the effect of used capital market on investment in new capital. They find that if sellers have more market power in used capital market, investment in new capital is higher compared to the case without the used capital market. The reverse is true if buyers have more market power in the used capital market.

of the vector

$$\boldsymbol{D} = \left(Q, \left\{Q_i, y_i, Q_i^a, \omega_i, \chi_{ni}^b, \chi_{pi}^b, \chi_{ni}^l, \chi_{pi}^l, i_{ni}^b, i_{pi}^b, i_{ni}^l, i_{pi}^l\right\}_{i=1}^N\right),$$

where y_i is set to 0 if bidder i does not participate in the auction.

Since the utility from balances at the end of the day is known from institutions of this market, the unknown parameters are $\left(\{\boldsymbol{B}_i, F_{\sigma_i}, F_{\epsilon_i}\}_{i=1}^N\right)$. Our problem is how to identify these unknowns from the distribution of $\boldsymbol{D}$. We approach this problem in two steps. In the first step, we identify $\{F_{\sigma_i}, F_{\epsilon_i}\}_{i=1}^N$ using data from secondary market trading. In the second step, we identify $\{\boldsymbol{B}_i\}_{i=1}^N$ using the first-order conditions of bidding. Throughout the paper, we assume that the density function of F_{σ_i} andF_{ϵ_i}, denoted respectively by f_{σ_i} andf_{ϵ_i}, are bounded and continuous on their support.

We now show how to identify F_{ϵ_i} and F_{σ_i}. From the data, we observe Q_i and ω_i, which implies that the distribution of $\hat{q}_i = \sigma_i + \epsilon_i = Q_i + \omega_i$ is observed. We only need to show that we can disentangle σ_i and ϵ_i. To achieve this, we exploit the fact that σ_i is known to bank i at the bidding stage, while ϵ_i is not. This implies that the bids depend on σ_i and are independent of ϵ_i. Therefore, bids and $\hat{q}_i$ can be regarded as repeated measurements of σ_i, which can be used to identify F_{ϵ_i} and F_{σ_i}.[9] Although $\hat{q}_i$ is linear in σ_i, the bids are typically nonlinear functions of σ_i. Therefore, the methods from the classic measurement error literature do not apply. It is natural to consider using the methods for non-classic measurement error literature following, e.g., Hu and Schennach (2008), but it is not clear whether their crucial invertibility condition is satisfied in this setting.

In light of this challenge, we exploit the structure of our problem to develop a novel identification method.Let ϕ_i be a functional that maps y_i to a vector.

Assumption 2. *There exists a known functional ϕ_i such that bidder i's strategy is ϕ_i-partially invertible: There exists a set $\mathcal{Q}_i$ in the range of ϕ_i that satisfies $\forall\tilde{\phi} \in \mathcal{Q}_i$, there exists a unique value of σ_i such that $\phi_i(y_i) = \tilde{\phi}$.*

This assumption implies that $\phi_i(y_i)$ uniquely determines σ_i if it lies in $\mathcal{Q}_i$. One example is $\phi_i(y_i) = (b_{iK_i}, q_{iK_i}, q_{i,K_i-1})$. The FOC for the last step, shown in Proposition 1(2), defines an equation in σ_i given $(b_{iK_i}, q_{iK_i}, q_{i,K_i-1})$. Then the bidding strategy is ϕ_i-partially invertible

[9]Notice that the bids are also potentially noisy measurements of σ_i because it is possible that multiple values of σ_i can generate the same a bid profile.

if there exist values of $(b_{iK_i}, q_{iK_i}, q_{i,K_i-1})$ at which there is a unique solution to the equation defined by the FOC for the last step.

By Assumption 2, the distribution of σ_i conditional on $\phi_i(y_i)$ is degenerate if $\phi_i(y_i) \in \mathcal{Q}_i$. First, suppose $\mathcal{Q}_i$ is known. Recall that $\hat{q}_i = \sigma_i + \epsilon_i$ where ϵ_i is mean 0 and independent of y_i. Therefore, $\mathbb{E}\left[\hat{q}_i \middle| \phi_i(y_i)\right] = \sigma_i$ if $\phi_i(y_i) \in \mathcal{Q}_i$, which implies that $\epsilon_i = \hat{q}_i - \mathbb{E}\left[\hat{q}_i \middle| \phi_i(y_i)\right]$ can be recovered and then F_{ϵ_i} is identified.

The remaining challenge is to identify $\mathcal{Q}_i$, which is unknown in practice. Notice that the variance of $\hat{q}_i$ conditional on $\phi_i(y_i)$, denoted by $\mathbf{var}[\hat{q}_i|\phi_i(y_i)]$, is the sum of the conditional variance of σ_i, $\mathbf{var}[\sigma_i|\phi_i(y_i)]$, and the variance of ϵ_i. Notice that $\mathbf{var}[\sigma_i|\phi_i(y_i)]$ is 0 if $\phi_i(y_i) \in \mathcal{Q}_i$ and positive otherwise. Therefore, $\mathbf{var}[\hat{q}_i|\phi_i(y_i)]$ is minimized if and only if $\phi_i(y_i) \in \mathcal{Q}_i$, which allows us to identify $\mathcal{Q}_i$ by minimizing $\mathbf{var}[\hat{q}_i|\phi_i(y_i)]$. Once F_{ϵ_i} is identified, F_{σ_i} can be identified by applying a deconvolution argument to $\hat{q}_i$. Because $\hat{q}_i$ is observed even if bank i does not participate in the auction, this method is valid even though bank i endogenously decides whether to participate in an auction based on σ_i, as shown in Table 4.

Proposition 2. *If bidder i's bid is ϕ_i-partially invertible with positive probability, i.e., $\mathbf{Pr}\left(\phi_i(y_i) \in \mathcal{Q}_i\right) > 0$, then F_{ϵ_i} is identified. Moreover, if the characteristic function of F_{ϵ_i} takes the value 0 at most on a set of isolated points, F_{σ_i} is identified.*

Once F_{σ_i} and F_{ϵ_i} are identified, we can identify $\boldsymbol{B}_i$ from the first-order condition. To do so, we combine (1), (2) and Proposition 1(2) to obtain an equation in $\boldsymbol{B}_i$. In the appendix, we show that if $(\sigma_i, y_i) = (\tilde{\sigma}, \tilde{y})$ with $\tilde{y} = \left\{\left(\tilde{b}_{ik}, \tilde{q}_{ik}\right)\right\}_{k=1}^{K_i}$, the equation takes the form

$$\boldsymbol{B}_i \eta\left(\tilde{\sigma}, \tilde{y}\right)' = C_i\left(\tilde{\sigma}, \tilde{y}\right).$$

Here

$$\eta\left(\tilde{\sigma}, \tilde{y}\right) = \left(\eta^b_{ni}\left(\tilde{\sigma}, \tilde{y}\right), \eta^b_{pi}\left(\tilde{\sigma}, \tilde{y}\right), \eta^l_{ni}\left(\tilde{\sigma}, \tilde{y}\right), \eta^l_{pi}\left(\tilde{\sigma}, \tilde{y}\right)\right), \tag{3}$$

and

$$\begin{aligned}\eta_{ji}^{m}\left(\tilde{\sigma},\tilde{y}\right)=&P_{i1}\left(\tilde{y}\right)\mathbb{E}\Bigg[E_{1,ji}^{m}\left(\tilde{q}_{1}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\\&+E_{2,ji}^{m}\left(\tilde{q}_{1}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial\boldsymbol{Q}_{-i}^{a}}{\partial q_{iK_{i}}}\Bigg|\tilde{b}_{iK_{i}}>p^{c},y_{i}=\tilde{y}\Bigg]\\&+P_{i2}\left(\tilde{y}\right)\mathbb{E}\Bigg[E_{1,ji}^{m}\left(Q_{i}^{a}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\\&+E_{2,ji}^{m}\left(Q_{i}^{a}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial\boldsymbol{Q}_{-i}^{a}}{\partial q_{iK_{i}}}\Bigg|\tilde{b}_{iK_{i}}=p^{c},y_{i}=\tilde{y}\Bigg],\end{aligned}\tag{4}$$

where

$$P_{i1}\left(\tilde{y}\right)=\mathbf{Pr}\left(\tilde{b}>p^{c}\middle|\,y_{i}=\tilde{y}\right)\tag{5}$$

$$P_{i2}\left(\tilde{y}\right)=\mathbf{Pr}\left(\tilde{b}=p^{c}\middle|\,y_{i}=\tilde{y}\right)\tag{6}$$

$$E_{ji}^{m}\left(x,z\right)=\mathbb{E}\left(\chi_{ji}^{m}\middle|\,Q_{i}^{a}+\sigma_{i}=x,\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}=z\right),\tag{7}$$

with its derivative with respect to the first and the second argument denoted as $E_{1,ji}^{m}\left(x,z\right)$ and $E_{2,ji}^{m}\left(x,z\right)$, respectively. And

$$\begin{aligned}&C_{i}\left(\tilde{\sigma},\tilde{y}\right)\\=&P_{i1}\left(\tilde{y}\right)\tilde{b}+P_{i2}\left(\tilde{y}\right)\tilde{b}\mathbb{E}\left(\frac{\partial Q_{i}^{a}}{\partial q_{iK_{i}}}\middle|\,\tilde{b}_{iK_{i}}=p^{c},y_{i}=\tilde{y}\right)\\&-P_{i1}\left(\tilde{y}\right)\mathbb{E}\left[A_{1i}\left(\tilde{q}_{1}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i},\boldsymbol{\epsilon}_{-i}\right)\middle|\,\tilde{b}_{iK_{i}}>p^{c},y_{i}=\tilde{y}\right]\\&-P_{i1}\left(\tilde{y}\right)\mathbb{E}\left[A_{2i}\left(\tilde{q}_{1}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial\boldsymbol{Q}_{-i}^{a}}{\partial q_{iK_{i}}}\middle|\,\tilde{b}_{iK_{i}}>p^{c},y_{i}=\tilde{y}\right]\\&-P_{i2}\left(\tilde{y}\right)\mathbb{E}\left[A_{1i}\left(Q_{i}^{a}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i},\boldsymbol{\epsilon}_{-i}\right)\frac{\partial Q_{i}^{a}}{\partial q_{iK_{i}}}\middle|\,\tilde{b}_{iK_{i}}=p^{c},y_{i}=\tilde{y}\right]\\&-P_{i2}\left(\tilde{y}\right)\mathbb{E}\left[A_{2i}\left(Q_{i}^{a}+\tilde{\sigma},\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial\boldsymbol{Q}_{-i}^{a}}{\partial q_{iK_{i}}}\middle|\,\tilde{b}_{iK_{i}}=p^{c},y_{i}=\tilde{y}\right],\end{aligned}$$

where

$$A_{i}\left(x,z\right)=\mathbb{E}\left[u\left(Q_{i}^{e}\right)+\sum_{j\in\{n,p\}}\left(i_{ji}^{l}\chi_{ji}^{l}-i_{ji}^{b}\chi_{ji}^{b}\right)\middle|\,Q_{i}^{a}+\sigma_{i}=x,\boldsymbol{Q}_{-i}^{a}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}=z\right]$$

and A_{1i} and A_{2i} are its derivatives with respect to the first and the second argument, respectively.

If we can find $\tilde{y}_t$ for $t = 1, \cdots, 4$ such that $\phi_i(\tilde{y}_t) \in \mathcal{Q}_i$, we can compute the corresponding $\tilde{\sigma}_t = \mathbb{E}[\hat{q}_i | \phi_i(y_i) = \phi_i(\tilde{y}_t)]$. Then we can compute $\eta(\tilde{\sigma}_t, \tilde{y}_t)$ and $C_i(\tilde{\sigma}_t, \tilde{y}_t)$, denoted by η_t and C_{it} for simplicity. This leads to a system of equations $\boldsymbol{B}_i \boldsymbol{\eta} = \boldsymbol{C}_i$, where $\boldsymbol{\eta} = (\eta_1', \eta_2', \eta_3', \eta_4')$ is a 4×4 matrix and

$$\boldsymbol{C}_i = (C_{i1}, C_{i2}, C_{i3}, C_{i4}).$$

Because $\boldsymbol{C}_i$ and $\boldsymbol{\eta}$ are identified, $\boldsymbol{B}_i$ is point-identified as $\boldsymbol{B}_i = \boldsymbol{C}_i \boldsymbol{\eta}^{-1}$ if $\boldsymbol{\eta}$ has full rank.

Proposition 3. *Suppose there exist $\tilde{y}_t$, $t = 1, \cdots, 4$ in the support of y_i such that $\phi_i(\tilde{y}_t) \in \mathcal{Q}_i$ and the corresponding 4×4 matrix $\boldsymbol{\eta}$ has full rank. Then $\boldsymbol{B}_i$ is point-identified.*

6 Estimation

From data, we observe L replicates of D, one for each day. Denote it by

$$D_\ell = \left(Q_\ell, \left\{Q_{i\ell}, y_{i\ell}, Q^a_{i\ell}, \omega_{i\ell}, \chi^b_{ni\ell}, \chi^b_{pi\ell}, \chi^l_{ni\ell}, \chi^l_{pi\ell}, i^b_{ni\ell}, i^b_{pi\ell}, i^l_{ni\ell}, i^l_{pi\ell}\right\}_{i=1}^N\right), \ \ell = 1, \cdots, L.$$

Also define $\hat{q}_{i\ell} = Q_{i\ell} + \omega_{i\ell}$. Our estimation method closely follows the identification strategy, which consists of three steps. Throughout this section, we assume that $\phi_i(y_i)$ is an M-dimensional vector. For example, if $\phi_i(y_i) = (q_{iK_i}, b_{iK_i})$, $M = 2$. Let $\boldsymbol{D}_L = \{D_\ell\}_{\ell=1}^L$ denote the data.

Step 1: Estimating F_{ϵ_i} This step involves estimating conditional expectations nonparametrically. One can use standard methods in the literature to estimate the conditional expectations such as the Nadaraya-Watson (NW) Estimator, a local polynomial estimator or a sieve estimator. Let $\hat{\mathbb{E}}\left(\hat{q}_i | \phi_i(y_i) = \tilde{\boldsymbol{\phi}}\right)$ be an estimator of $\mathbb{E}\left(\hat{q}_i | \phi_i(y_i) = \tilde{\boldsymbol{\phi}}\right)$. If $y_{i\ell} \neq 0$, the residual can be estimated by $\hat{\epsilon}_{i\ell} = \hat{q}_{i\ell} - \hat{\mathbb{E}}_{-\ell}(\hat{q}_{i\ell} | \phi_i(y_{i\ell}))$ where $\hat{\mathbb{E}}_{-\ell}(\hat{q}_{i\ell} | \phi_i(y_{i\ell}))$ is the leave-one-out conditional expectation estimator, i.e., it use observations other than ℓ. This allows us to estimate the conditional variance of $\hat{q}_i$ by nonparametrically regressing $\hat{\epsilon}^2_{i\ell}$ on $\phi_i(y_{i\ell})$. Again, we can use the NW, a local polynomial estimator or a sieve estimator. Denote the estimator of $\mathbf{var}\left(\hat{q}_i | \phi_i(y_i) = \tilde{\boldsymbol{\phi}}\right)$ by $\widehat{\mathbf{var}}\left(\hat{q}_i | \phi_i(y_i) = \tilde{\boldsymbol{\phi}}\right)$. We can estimate the set

of ℓ such that $\phi_i(y_{i\ell}) \in \mathcal{Q}_i$ by

$$\hat{\Lambda}_i = \left\{ \ell : \widehat{\mathbf{var}} \left(\hat{q}_i | \, \phi_i(y_i) = \phi_i(y_{i\ell}) \right) - \min_{\ell'} \widehat{\mathbf{var}} \left(\hat{q}_i | \, \phi_i(y_i) = \phi_i(y_{i\ell'}) \right) \leq \delta_L, y_{i\ell} \neq 0 \right\}.$$

Let $|\hat{\Lambda}_i|$ denote the cardinality of $\hat{\Lambda}_i$. Then F_{ϵ_i} can be estimated by

$$\hat{F}_{\epsilon_i}(x) = \sum_{\ell=1}^{L} \frac{1}{|\hat{\Lambda}_i|} \mathbf{1}(y_{i\ell} \neq 0, \ell \in \hat{\Lambda}_i, \hat{\epsilon}_{i\ell} \leq x). \tag{8}$$

Step 2: Estimating F_{σ_i} We parametrize the density of F_{σ_i} by $f(\cdot;\theta)$ where θ is an unknown finite-dimensional column vector which lies in a known compact space Θ. It can be a parametric family or sieves such as Berstein polynomials or shape-constrained splines. Define

$$g_i \left(\hat{q}_{i\ell}; \theta \right) = \int f \left(\hat{q}_{i\ell} - z; \theta \right) dF_{\epsilon_i} \left(z \right).$$

Then $\hat{q}_{i\ell}$ has a density $g_i \left(\hat{q}_{i\ell}; \theta_{i0} \right)$ where θ_{i0} is the true value. An MLE based on g_i is infeasible because F_{ϵ_i} is unknown. To obtain a feasible estimator, we replace F_{ϵ_i} by its estimator from the previous step. Let

$$\hat{g}_i \left(\hat{q}_{i\ell}; \theta \right) = \frac{\sum_{j \in \hat{\Lambda}_i} f \left(\hat{q}_{i\ell} - \hat{\epsilon}_{ij}; \theta \right)}{|\hat{\Lambda}_i|}.$$

We can then estimate θ_{i0} by

$$\hat{\theta}_i = \arg\max_{\theta \in \Theta} \sum_{\ell=1}^{L} \log \hat{g}_i \left(\hat{q}_{i\ell}; \theta \right).$$

Step 3: Estimating $\boldsymbol{B}_i$ Let

$$U_{i\ell} = u \left(Q^e_{i\ell} \right) + \sum_{j \in \{n,p\}} i^l_{ji\ell} \chi^l_{ji\ell} - \sum_{j \in \{n,p\}.} i^b_{ji\ell} \chi^b_{ji\ell}$$

In principle, $U_{i\ell}$ and the borrowing and lending rates could depend on the vector $\boldsymbol{\sigma}_\ell + \boldsymbol{Q}^a_\ell + \boldsymbol{\epsilon}_\ell$. To reduce the dimension of the problem, we assume they depend on the vector through $(X_{i\ell}, X^p_{i\ell}, X^n_{i\ell})$ where $X_{i\ell} = \hat{q}_{i\ell} + Q^a_{i\ell}$, $X^p_{i\ell} = \sum_{j \in \mathcal{P}_i} (\hat{q}_{j\ell} + Q^a_{j\ell})$ and $X^n_{i\ell} = \sum_{j \in \mathcal{N}_i} (\hat{q}_{j\ell} + Q^a_{j\ell})$, i.e., bank i's own position and its partner and nonpartner's positions excluding borrowing and lending. We also define $\boldsymbol{X}_{-i\ell} = (X^p_{i\ell}, X^n_{i\ell})$. Let $\hat{\mu}(x, z)$ be an estimator of $\mathbb{E}(U_{i\ell} | \, X_{i\ell} = x, \boldsymbol{X}_{-i\ell} = z)$. We can again use a Nadaraya-Watson estimator for simplicity.

We can then estimate A_i by

$$\hat{A}_i(x,z) = \int \hat{\mu}\left(x+\epsilon, z\right) d\hat{F}_{\epsilon_i} = \frac{1}{|\hat{\Lambda}_i|}\sum_{\ell\in\hat{\Lambda}_i} \hat{\mu}\left(x+\hat{\epsilon}_{i\ell}, z\right).$$

Similarly, let $\hat{\xi}^m_{ji}(x,z)$ be an estimator of $\mathbb{E}\left(\chi^m_{ji\ell}\middle|\, X_{i\ell}=x, \boldsymbol{X}_{-i\ell}=z\right)$. We can the estimate E^m_{ji} by

$$\hat{E}^m_{ji}(x,z) = \int \hat{\xi}^m_{ji}\left(x+\epsilon, z\right) d\hat{F}_{\epsilon_i} = \frac{1}{|\hat{\Lambda}_i|}\sum_{\ell\in\hat{\Lambda}_i} \hat{\xi}^m_{ji}\left(x+\hat{\epsilon}_{i\ell}, z\right).$$

We next estimate the two functions C and η evaluated at $(\sigma_{i\ell}, y_{i\ell})$ for all $\ell\in\hat{\Lambda}_i$, where $\sigma_{i\ell}$ is the signal corresponding to $\phi_i\left(y_{i\ell}\right)$. For such an ℓ, first estimate $\sigma_{i\ell}$ by

$$\hat{\sigma}_{i\ell} \;=\; \hat{\mathbb{E}}\left[\,\hat{q}_{i\ell}\middle|\,\phi_i(y_{i\ell})\right].$$

Then resample the total auctioned quantity (Q), the bids (y) and corresponding post-auction payment on behalf of customers $(\hat{q})$ from other auctions with replacement. Denote the resampled vector of bids and after auction customer payment as $\left(Q^*_{\ell t}, \boldsymbol{y}^*_{-i\ell t}, \hat{\boldsymbol{q}}^*_{-i\ell t}\right)$ for $t=1,2,\cdots,T$ where T is the total number of resampled auctions. Combine this with bank i's bidding strategy $y_{i\ell}$ to obtain the clearing price and quantities alloted to every bidder: $\left(p^{c*}_{\ell t}, Q^{a*}_{i\ell t}, \boldsymbol{Q}^{a*}_{-i\ell t}\right)$. This allows us to calculate $X^*_{i\ell t} = Q^{a*}_{i\ell t} + \hat{q}_{i\ell}$, $X^{p*}_{i\ell t} = \sum_{j\in\mathcal{P}_i}(\hat{q}^*_{j\ell t} + Q^{a*}_{j\ell t})$ and $X^{n*}_{i\ell t} = \sum_{j\in\mathcal{N}_i}(\hat{q}^*_{j\ell t} + Q^{a*}_{j\ell t})$. Similarly, we denote $\boldsymbol{X}^*_{-i\ell t} = (X^{p*}_{i\ell t}, X^{n*}_{i\ell t})$. Next, we change bidder i's last step from $q_{iK_i\ell}$ to $q_{iK_i\ell}-\varepsilon$ for some small ε and recalculate the allocation of bidders other than i. We can then calculate $\partial Q^{a*}_{j\ell t}/\partial q_{iK_i\ell}$ for $j=1,2,\cdots,N$, using first differences. Define

$$\frac{\partial \boldsymbol{Q}^{a*}_{-i\ell t}}{\partial q_{iK_i\ell}} = \left(\sum_{j\in\mathcal{P}_i}\frac{\partial Q^{a*}_{j\ell t}}{\partial q_{iK_i\ell}}, \sum_{j\in\mathcal{N}_i}\frac{\partial Q^{a*}_{j\ell t}}{\partial q_{iK_i\ell}}\right).$$

This allows us to estimate $\eta^m_{ji\ell} = \eta^m_{ji}\left(\sigma_{i\ell}, q_{iK_i\ell}, b_{iK_i\ell}\right)$ and $C_{i\ell} = C_i\left(\sigma_{i\ell}, q_{iK_i\ell}, b_{iK_i\ell}\right)$ by

$$\begin{aligned}\hat{\eta}^m_{ji\ell} = &\sum_{t=1}^{T}\mathbf{1}\left(b_{iK_i\ell} > p^*_{\ell t}\right)\left(\hat{E}^m_{1,ji}\left(q_{iK_i\ell}+\hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right) + \hat{E}^m_{2,ji}\left(q_{iK_i\ell}+\hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right)\cdot\frac{\partial \boldsymbol{Q}^{a*}_{-i\ell t}}{\partial q_{iK_i\ell}}\right)\\ &+\sum_{t=1}^{T}\mathbf{1}\left(b_{iK_i\ell} = p^*_{\ell t}\right)\left(\hat{E}^m_{1,ji}\left(Q^{a*}_{iK_i\ell}+\hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right)\frac{\partial Q^{a*}_{iK_i\ell}}{\partial q_{iK_i\ell}} + \hat{E}^m_{2,ji}\left(q_{iK_i\ell}+\hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right)\cdot\frac{\partial \boldsymbol{Q}^{a*}_{-i\ell t}}{\partial q_{iK_i\ell}}\right),\end{aligned}$$

$$\hat{C}_{i\ell} = \sum_{t=1}^{T} \mathbf{1}\left(b_{iK_i\ell} > p^*_{\ell t}\right) b_{iK_i\ell} + \sum_{t=1}^{T} \mathbf{1}\left(b_{iK_i\ell} = p^*_{\ell t}\right) b_{iK_i\ell} \frac{\partial Q^{a*}_{iK_i\ell}}{\partial q_{iK_i\ell}}$$
$$- \sum_{t=1}^{T} \mathbf{1}\left(b_{iK_i\ell} > p^*_{\ell t}\right) \left(\hat{A}_{1i}\left(q_{iK_i\ell} + \hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right) + \hat{A}_{2i}\left(q_{iK_i\ell} + \hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right) \cdot \frac{\partial \boldsymbol{Q}^{a*}_{-i\ell t}}{\partial q_{iK_i\ell}} \right)$$
$$- \sum_{t=1}^{T} \mathbf{1}\left(b_{iK_i\ell} = p^*_{\ell t}\right) \left(\hat{A}_{1i}\left(Q^{a*}_{iK_i\ell} + \hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right) \frac{\partial Q^{a*}_{iK_i\ell}}{\partial q_{iK_i\ell}} + \hat{A}_{2i}\left(Q^{a*}_{iK_i\ell} + \hat{\sigma}_{i\ell}, \boldsymbol{X}^*_{-i\ell t}\right) \cdot \frac{\partial \boldsymbol{Q}^{a*}_{-i\ell t}}{\partial q_{iK_i\ell}} \right).$$

Let $\hat{\eta}_\ell = \left(\hat{\eta}^b_{ni\ell}, \hat{\eta}^b_{pi\ell}, \hat{\eta}^l_{ni\ell}, \hat{\eta}^l_{ni\ell}\right)$ and $\hat{\boldsymbol{\eta}}$ be a $4 \times |\hat{\Lambda}_i|$ matrix with each column being $\hat{\boldsymbol{\eta}}_\ell$ for one $\ell \in \hat{\Lambda}_i$. Let $\hat{\boldsymbol{C}}_i$ be an $1 \times |\hat{\Lambda}_i|$ vector whose elements are $\hat{C}_{i\ell}$ with $\ell \in \hat{\Lambda}_i$. Then we can estimate $\boldsymbol{B}_i$ by

$$\hat{\boldsymbol{B}}_i = \arg\min_{\boldsymbol{x}} \left(\hat{\boldsymbol{C}}_i - \boldsymbol{x}\hat{\boldsymbol{\eta}}\right)\left(\hat{\boldsymbol{C}}_i - \boldsymbol{x}\hat{\boldsymbol{\eta}}\right)'.$$

Next, we investigate the properties of the estimators. We focus on step 1 because it is the crucial part of the estimator. The other steps involve relatively standard estimators. Therefore, one can invoke existing results in the literature to analyze them. In Step 1, we exploit the identification that $\sigma_{i\ell}$ is point-identified if $\phi_i(y_{i\ell}) \in \mathcal{Q}_i$. A natural way to proceed is to first prove that the set of $\phi_i(y_{i\ell})$ for all $\ell \in \hat{\Lambda}_i$ converges uniformly to $\mathcal{Q}_i$ asymptotically then, to derive the corresponding convergence rate and using this information, prove the convergence of $\hat{\epsilon}_{i\ell}$ to $\epsilon_{i\ell}$ for relevant ℓ. However, this approach is challenging because it requires knowledge on the bidding strategy, which is too complicated to analyze. Instead, we derive the convergence rate of $\hat{\epsilon}_{i\ell}$ without showing the convergence of the set of $\phi_i(y_{i\ell})$ for all $\ell \in \hat{\Lambda}_i$. Intuitively, if $\ell \in \hat{\Lambda}_i$, the distribution of $\sigma_{i\ell}$, the variance of $\sigma_{i\ell} - \hat{\sigma}_{i\ell}$ converges to 0 by definition if δ_L is selected properly. This means that $\hat{\sigma}_{i\ell}$ is close to $\sigma_{i\ell}$ even if $\phi_i(y_{i\ell}) \notin \mathcal{Q}_i$, which implies that $\hat{\epsilon}_{i\ell}$ is close to $\epsilon_{i\ell}$. Define

$$S_\ell = \left\{ \ell \in \hat{\Lambda}_i \text{ and } \phi_i\left(y_{i\ell}\right) \notin \mathcal{Q}_i \right\}$$

to be the event that observation ℓ is incorrectly classified as an invertible point. Then let $\bar{S}_L = \cup_{\ell \leq L} S_\ell$

To formalize the above discussion, we impose the following high level assumptions:

Assumption 3. *As $L \to \infty$,*

$$\sup_{\ell} \left| \hat{\mathbb{E}}\left[\hat{q}_i \middle| \phi_i(y_i) = \phi_i(y_{i\ell})\right] - \mathbb{E}\left[\hat{q}_i \middle| \phi_i(y_i) = \phi_i(y_{i\ell})\right] \right| = O_p\left(\zeta_L\right)$$

$$\sup_{\ell} \left| \widehat{\mathbf{var}} \left[\hat{q}_i \middle| \phi_i(y_i) = \phi_i(y_{i\ell}) \right] - \mathbf{var} \left[\hat{q}_i \middle| \phi_i(y_i) = \phi_i(y_{i\ell}) \right] \right| = O_p\left(\xi_L\right)$$

where $\zeta_L \to 0$ *and* $\xi_L \to 0$ *are two sequences of positive numbers such that* $\sqrt{L}\zeta_L \to \infty$ *and* $\sqrt{L}\xi_L \to \infty$*. Moreover,* $\delta_L \to 0$ *and* $\delta_L / \max\left\{\xi_L, \zeta_L^2\right\} \to \infty$*.*

This condtion requires uniform convergence of the conditional mean and conditional variance estimators. They can be shown to hold under mild conditions.

Proposition 4. *Suppose Assumption 3 holds.*

1. *For any fixed* ℓ*,*

$$\limsup_{\ell} \mathbf{Pr}\left(S_\ell\right) \leq \lim \sup_{L \to \infty} \mathbf{Pr}\left(\bar{S}_L\right) = 0$$

Therefore, there exists a positive sequence $\kappa_L \to 0$ *such that* $\mathbf{Pr}\left(S_\ell\right) = O\left(\kappa_L\right)$*.*

2. *Moreover,*

$$\begin{aligned}
\sup_{\phi_i(y_{i\ell}) \in \mathcal{Q}_i} \left| \hat{\epsilon}_{i\ell} - \epsilon_{i\ell} \right| &= O_p\left(\zeta_L\right) \\
\sup_{\phi_i(y_{i\ell}) \neq \mathcal{Q}_i, \ell \in \hat{\Lambda}_i} \mathbb{E}\left(\hat{\epsilon}_{i\ell} - \epsilon_{i\ell}\right)^2 &= O_p\left(\delta_L\right) \\
\sup_{x} \left| \hat{F}_{\epsilon_i}(x) - F_{\epsilon_i}(x) \right| &= O_p\left(\max\left\{\left(\delta_L \kappa_L\right)^{1/3}, \zeta_L\right\}\right).
\end{aligned}$$

The first claim show that the probability of $\hat{\Lambda}_i$ containing any points outside of $\mathcal{Q}_i$ converges to 0. We denote the convergence rate of this probability as κ_L. Because of the complicated bidding strategies, we in general cannot get an exact characterization of κ_L.

The second claim shows the convergence rate of the estimators of ϵ and its distribution. Because $\hat{\Lambda}_i$ may contain ℓ such that $\phi_i(y_{i\ell}) \notin \mathcal{Q}_i$, we cannot obtain a uniform convergence on $\hat{\Lambda}_o$. Instead, we need to distinguish points such that $\phi_i(y_{i\ell}) \in \mathcal{Q}_i$ and $\phi_i(y_{i\ell}) \notin \mathcal{Q}_i$ separately. For the former group, we can obtain uniform convergence in probability but for the later group, we can only show the L^2 norm converges uniformly.

This also affects $\hat{F}_{\epsilon_i}$, whose convergence rate is determined by both the points inside $\mathcal{Q}_i$ and those outside. It is worth noting that even if $\kappa_L = 1$, i.e., $\hat{\Lambda}_i$ always contains points outside of $\mathcal{Q}_i$, $\hat{F}_{\epsilon_i}$ is still consistent but may converge at a slower rate.

We next derive the convergence rate of $\hat{\theta}_i$. We first derive the convergence rate of $\hat{\theta}_i$ under a set of high-level conditions and then show that our parametrization in the empirical analysis satisfies these conditions.

Assumption 4. *The parameter space Θ is compact and the functions g_i and $\hat{g}_i$ satisfy:*

1. $\log g_i(x;\theta)$ *is continuous in θ for all x and there exists a function M such that*

$$\sup_{\theta\in\Theta} |\log g_i(q_{i\ell};\theta)| < M(q_{i\ell})$$

and $\mathbb{E}M(q_{i\ell}) < \infty$. Moreover,

$$\sup_{\theta\in\Theta} \frac{1}{L}\sum_{\ell=1}^{L} \left| \frac{\hat{g}_i(\hat{q}_{i\ell};\theta) - g_i(\hat{q}_{i\ell};\theta)}{g_i(\hat{q}_{i\ell};\theta)} \right| = o_p(1)$$

2. *There exists a neighborhood of θ_{i0}, denoted by $\mathcal{N}_i(\theta_{i0})$ such that g_i There exists a positive sequence $a_L \to 0$ and $\sqrt{L}a_L \to \infty$ such that*

$$\sup_{\theta\in\mathcal{N}_i(\theta_{i0})} \left\| \frac{1}{L}\sum_{\ell=1}^{L} \left(\frac{\partial_\theta \hat{g}_i(\hat{q}_{i\ell};\theta)}{\hat{g}_i(\hat{q}_{i\ell};\theta)} - \frac{\partial_\theta g_i(\hat{q}_{i\ell};\theta)}{g_i(\hat{q}_{i\ell};\theta)} \right) \right\| = O_p(a_L).$$

3. *The Hessian matrix*

$$H_{i0} = \partial_{\theta\theta'}\, \mathbb{E}\log g_i(\hat{q}_{i\ell};\theta)|_{\theta=\theta_{i0}}$$

is non-singular and

$$\mathbb{E}\sup_{\theta\in\mathcal{N}_i(\theta_{i0})} \|\partial_{\theta\theta'}\log g_i(\hat{q}_{i\ell};\theta)\| < \infty, \mathbb{E}\left\| \frac{\partial_\theta g_i(\hat{q}_{i\ell};\theta_{i0})}{g_i(\hat{q}_{i\ell};\theta_{i0})} \right\|^2 < \infty.$$

Assumption (1) requires that the true likelihood is bounded by an integrable function uniformly on Θ and that the estimated log-likelihood converges uniformly to the true log-likelihood. Together with the compactness of Θ, they ensure that $\hat{\theta}_i$ is consistent. Assumption (2) requires that the estimated score function converges to the true score function at rate $1/a_L$. Because the estimated score function involves non-parametrically estimated variables, this convergence is typically slower than $\sqrt{L}$. Assumption (3) ensures that we can use the standard delta-method based on the score function to derive the convergence rate. The following proposition shows that $\hat{\theta}_i$ converges at the same rate as the score function.

Proposition 5. *Under Assumption 6, $\hat{\theta}_i - \theta_{i0} = O_p(a_L)$.*

In the empirical study, f is a normal density and the unknown parameter is $\theta = (\theta_1, \theta_2)$,

where θ_1 is the mean and θ_2 is the standard deviation. We next show that it satisfies Assumption 6 with an assumption on F_{ϵ_i}.

Assumption 5. *The functions f and F_{ϵ_i} satisfy*

1. *f is a normal density with parameter θ such that $\theta \in \Theta = \left[\underline{\theta}_1, \bar{\theta}_1\right] \times \left[\underline{\theta}_2, \bar{\theta}_2\right]$, where Θ is compact and $\underline{\theta}_2 > 0$.*

2. *F_{ϵ_i} has a density supported on a bounded interval and is greater than $\underline{p} > 0$ on its support.*

6 is the nonparametrically estimated objects satisfy Assumption 3. The main challe

Proposition 6. *Assumptions 3 and 5 imply Assumption 6 with $a_L = \zeta_L + \kappa_L\sqrt{\delta_L}$. Therefore, $\hat{\theta}_i - \theta_{i0} = O_p\left(\zeta_L + \kappa_L\sqrt{\delta_L}\right)$.*

The main challenge is to verify Assumption 6(2). For the normal specification, we exploit two properties. First,

$$\sup_{y\in[a,b]} \frac{\|\partial_\theta f(x-y;\theta)\|}{\int_a^b f(x-y;\theta)\,dy} \text{ and } \sup_{y\in[a,b]} \frac{\|f(x-y;\theta)\|}{\int_a^b f(x-y;\theta)\,dy}$$

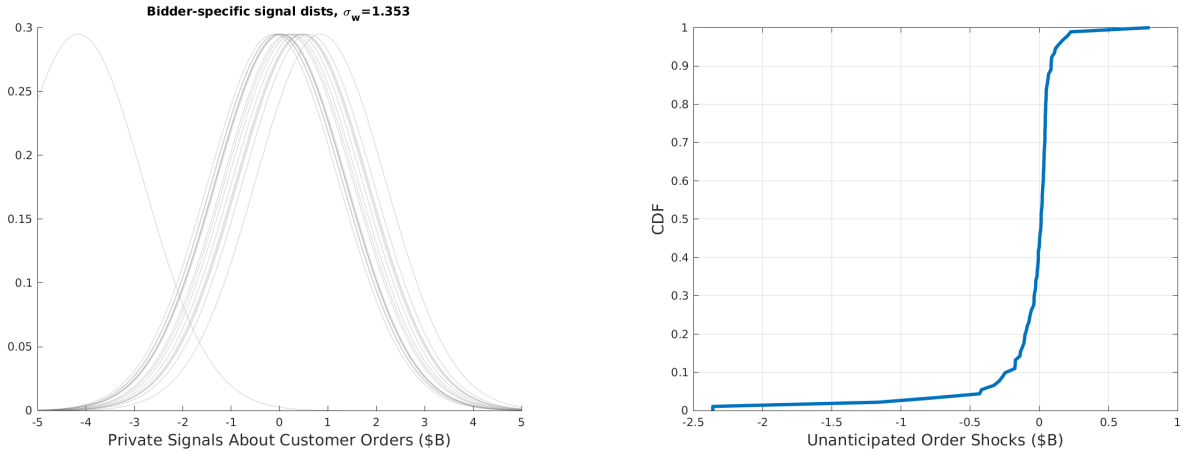
are bounded uniformly over $\theta \in \Theta$ by a polynomial function of x multiplied by a constant that depends only on the length of the interval $[a, b]$. This controls the effect of replacing ϵ_{ij} by their estimates. Second, the normal density has exponential tails, which ensures that the relevant polynomial envelopes are integrable and that sample maxima grow only logarithmically. Therefore, the same proof applies to other choices of f that satisfy both the relative-derivative polynomial bound and a sufficiently fast tail condition.

7 Results

Figure 4 plots the estimated distributions of post-auction client trading. Panel (A) provides the bidder-specific signal distributions $s_i \sim N(\hat{\mu}_i, \hat{\sigma}_w^2)$, while Panel (B) provides the estimated distribution of unanticipated client order shocks ε_i. Banks have good information about post-auction trading volumes, with most variance captured by the anticipated component s_i. The between-bidder heterogeneity in $\hat{\mu}_i$ reflects persistent differences in customer bases: some banks are systematically net receivers of funds while others are net senders. The

Figure 4: Estimated Distribution of Customer Trading

Figure 4 plots the empirical distributions of post-auction client trading. Panel (A) provides the bidder-specific signal distributions s_i. Panel (B) provides the estimated distribution of unanticipated client order shocks ε_i.



Panel (A) Panel (B)

Table 6: Estimated Preferences for Trading

Table 6 presents the $\hat{B}$ estimates obtained using the procedure described in the previous section. There are four parameters, that correspond to additional utility that the bank derives from trading volume, depending on who they trade with. These estimates can be compared with raw interest rate charges from these trades that have a similar magnitude (divide the price in bps by 100 to obtain the same units). The mean shift tells how that coefficient influences the marginal utility in the auction given the average quantity derivatives.

	Estimates	Mean Shift
Borrow Non-parter	0.033	-0.005
Borrow Partner	0.052	-0.007
Lend Non-partner	-0.249	-0.012
Lend Partner	-0.032	-0.004

distribution of ε_i exhibits a long lower tail, indicating that banks sometimes face much larger outflows than expected.

The estimated taste parameters for trading with partners and non-partners are presented in Table 6. Borrowing provides non-pecuniary benefits while lending provides additional non-pecuniary costs, possibly capturing the aversion of potential negative balances at the end of the day. Trading with partners, regardless of borrowing or lending, is preferred to trading with non-partners. In particular, lending to non-partners incurs a cost that is significantly larger than lending to a partner. This shows that banks value relationships, especially in lending.

Table 7: Marginal Values

Table 7 lists summary statistics of the marginal values (and relative marginal values) under a range of assumptions in basis points. The V_1 unconditional is the expectation of the V_1 term removing the conditioning on the event that the market clears between this and the next price. The Conditional on being allocated is the V_1 term isolated conditioning on the event that the market clears in that interval. The opponent spread moves partner and non-partner positions simultaneously from p10 to p90 while the partner and non-partner rows move the partner and nonpartner only, fixing the other at the median. V_2 displacement is the change in your value from winning more, reducing the quantity won by your opponents. Finally Marginal Value IPV presents the marginal values you would get using a model that imposed independent private values rather than our model of the secondary market.They are presented in basis points relative to the target rate.

Variable	Mean	Median	Std	p10	p90
Bid	-10.84	-10.00	9.49	-25.00	0.00
V_1 (unconditional)	-6.39	-3.97	6.53	-17.22	-2.03
V_1 (cond. allocated)	-6.53	-4.03	6.45	-18.09	-2.09
V_1 opponent spread	1.86	1.16	9.00	-5.24	9.75
Partner spread	5.37	0.36	55.29	-4.68	6.89
Non-partner spread	4.20	0.63	49.19	-3.93	7.77
V_2 displacement	-1.90	-1.12	2.59	-4.88	-0.23
Bid $-$ V_1	-4.31	-3.10	11.02	-19.90	6.90
Marginal Value IPV	-4.03	-10.00	41.16	-25.00	14.85

7.1 Secondary Market Effects on Marginal Values

To isolate the source of marginal values we quantify how a bank's marginal willingness to pay at the auction depends on the positions held by other banks, through the secondary market. Table 7 reports aggregate statistics. First we evaluate how V_1 is impacted by opponent positions. The unconditional expectation of V_1 is very close to the conditional expectation (given that a bidder acquires those units). This is the difference in valuations that play a central role in the *winners curse*, where the conditioning reveals that others valued the units less. This could arise either because bidders will make similar secondary market trades regardless of their opponents allocations or because the event that their bid is allocated reveals little information about the positions of opponents relevant for the secondary market trading. To distinguish between these explanations we present the "opponent spread" the difference in V_1 evaluated at the 10th and 90th percentiles of bootstrap opponent configurations, which measures how sensitive a bank's demand for reserves is to what its counterparties hold. This spread is large meaning that the allocations of opponents can meaningfully change trading opportunities. This is true of both partners and non-partners.

The displacement term V_2 captures the second distinct mechanism. When a bank wins an additional unit at the auction, it pushes other banks off the margin, changing their al-

locations and therefore the secondary market positions they bring to the trading session. Table 7 quantifies this feedback. V_2 is small relative to V_1 for most bidders, but it is non-zero and varies across bidders, confirming that strategic interaction operates through the auction clearing mechanism as well as through expectations about secondary market outcomes. Together, these findings suggest that valuations at the auction embed expectations about secondary market outcomes.

Finally, we ask how our implied marginal valuations relate to those that the researcher would recover if they estimated a model with independent private values, following Kastl (2011, 2012). Recall that these models impose a first order condition which relates the marginal value to bids with a similar structure to the equations in 1. At steps before the last one, if you drop terms involving ties (which may be relatively unlikely), then the marginal values from the IPV model should correspond to

$$\mathbb{E}\left[V_{i1}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}) + V_{i2}(q_{ik}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}) \cdot \frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} | b_{ik} > p^c > b_{ik+1}\right].$$

At the last step the IPV model and our model produce different predictions, as the IPV model imposes that bid is equal to marginal value at that step. Appendix Figure 5 shows the differences in the marginal values across steps. At steps 1 and 2 the marginal values implied by the two models are similar, but the full marginal value curve is substantially flatter once trading is accounted for. This flatness plays a central role in our counterfactual results.

7.2 The Role of the Secondary Market

There are three main components of bidder values: (i) expected payments for secondary market trading, (ii) utility from end-of-day balances, and (iii) a bonus utility from bilateral trading relationships. To understand how each shapes auction outcomes, we decompose the role of the secondary market by sequentially shutting down its components and recomputing equilibrium. Table 8 reports the results under both a one-step procedure, clearing the auction once at counterfactual marginal values, and an equilibrium version that iterates to a new equilibrium.

First, note that the allocation obtained with strategic bidding is quite different from the efficient allocation, with almost half the auctioned amount assigned to a different owner

Table 8: Secondary Market Decomposition: One-Step and Re-Equilibrated

Bids to MV shows the effect of reallocation from actual bids to allocations based on marginal values. This isolates the strategic bidding effect. This is presented for one-step adjustment only. For the other scenarios both a one-step adjustment, where the auction is cleared once at counterfactual marginal values fixing opponents allocations from the status-quo, and an equilibrium version that iteratively updates opponents' allocations until convergence. Reallocation is $\frac{1}{2}\sum_i |\Delta q_i|$ ($B) where the 1/2 avoids double counting the loss by one bank with the gain of another. Mean Δu is the average change in per-bank values ($B), for the MV column this is relative to bids, and for the remaining columns this is relative to marginal values. and the Std is the (across bank) standard deviation of the bank-specific average values. The columns No SM shuts down the secondary market, integrating over the customer order shocks, no secondary market ε known, allows the allocation to respond to the customer order shocks (i.e., based on values post announcement of shocks). No payment removes the interest payments from marginal values. No payment and No B removes both interest payments and utility bonus.

		Bids to MV	No SM	No SM, ε known	No Payments	No Payment, No B
One-step	Reallocation ($B)	0.2511	0.2341	0.2241	0.1138	0.1138
	Mean Δu ($10M)	-0.0101	-0.1086	-0.0992	-0.0065	0.0056
	Std u ($10M)	0.0241	0.1187	0.1173	0.0221	0.0010
Equilibrium	Reallocation ($B)	-	0.4108	0.4119	0.5423	0.5789
	Mean Δu ($10M)	-	-0.3447	-0.3447	0.0059	0.0062
	Std u ($10M)	-	0.1462	0.1463	0.0041	0.0004

in the case with and without strategic bidding. However, the welfare (end of day utility) produced by these two allocations is very similar.

We then evaluate how much the secondary market contributes to auction outcomes by recomputing bidders' valuations without post-auction trading. Without the secondary market, banks cannot adjust after the auction depending on the realized customer shocks, and marginal valuations become substantially steeper. Eliminating the secondary market entirely produces a substantial welfare decrease and causes a large increase in the standard deviation of expected outcomes at the bank level, making some banks much worse off than others. These declines are not driven by uncertainty about the customer shocks, even if these are known pre-auction welfare falls substantially. This is because the auction volume is small relative to the total exchanged in the secondary market, and therefore does not provide adequate opportunities for adjustment. Without the secondary market, the auctions tend to clear at the corridor ceiling, consistent with the role the secondary market plays in reallocating balances ex-post. The no-secondary-market counterfactual also speaks directly to reserve adequacy. We calculate the largest quantity needed across days so that each bank would reach zero position without the secondary market. We find that $8B of settlement balances would suffice. At the other extreme, a perfectly competitive secondary market would achieve the deposit rate with arbitrarily small supply: all banks could lend at the deposit rate, demand curves would be flat. The actual secondary market sits between

these extremes.

The remaining columns explore the source of the misallocation. A planner would like to allocate balances in the auction to the highest value users, but the private valuations diverge from the planners' target. First, the bidder is paid a price that may diverge from the surplus they provide in the secondary market and as a result may under or over purchase at the auction to increase their profits from payments. To measure how these between bidder transfers distort the allocation we recompute marginal values excluding the payment terms from the value function. We then re-run the allocation based on these new curves, and compute bidder welfare. Relative to the baseline allocation based on the marginal values in the status-quo, there is moderate re-allocation, but bidders obtain similar utility. In addition, while the cross-bank dispersion in expected welfare falls, the change is small. Second, the bidders earn utility from trading with partners and non-partners. In our model these directly shift the utility and so would be considered by the planner, however these terms may capture interbank transfers that a planner would not care about. Therefore, as a second exercise we shut down this force as well as the payments. Removing this force again leads to a moderate amount of reshuffling of allocations, and a small change in utility. However, there is noticeably less dispersion in expected utility across banks without neither payments or utility bonuses.

8 Counterfactual Exercises

We now use the estimated model to evaluate the interaction between the auction and the secondary market. We consider a sequence of policy levers that the central bank, as the market designer, could use. Because the policies operate through the secondary market as well as the auction, they could have different effects with secondary market interactions than for primary markets that exist in isolation. In particular, we consider a ban on participation of smaller bidders in the auction and changes to the auction supply. Both of these will change expected secondary market trading opportunities and hence reshape marginal values and allocations at the auction. Throughout, we maintain the truthful-bidding benchmark of Hortaçsu and McAdams (2010): bidders submit their marginal value curve as their demand function.[10] Our counterfactual holds fixed the trading structure of the secondary market,

[10] Although the method of Richert (2021) could potentially be deployed to solve for bounds on the outcomes with strategic bid-shading, the results would not yield precise predictions since we do not model how bidders

which implies that the conditional expectations of trading volumes and interest rates, given other bidders' positions, are also fixed. This assumption is reasonable for changes to the primary market, and is consistent with predictions of theoretical models including Afonso and Lagos (2015). To incorporate changes to the secondary market structure (e.g., changes in long-run relationship benefits shifting B or bans on participants in the secondary market), one could augment our framework with an explicit model of the secondary market. However, this is not necessary for our counterfactuals of interest.

The main take-away from the counterfactual results is that because resale flattens each banks' marginal value curve, the primary auction has remarkably little influence over the price of liquidity or who holds reserves.

8.1 Restricting Participation

The wedge between individual and planner marginal values in this setting operates through several channels. First, firms' V_1 includes expected secondary market payments which are pure transfers from the planner's perspective and carry no social value. Second, the B coefficients capturing bilateral partner rate advantages similarly reflect distributional gains from preferred relationships rather than aggregate liquidity value. These channels drive a wedge between what a bank is willing to pay and what the planner values. This motivates asking what happens when the auction is restricted to the most active participants. Can a smaller set of larger banks, that are more similar as a group, bring outcomes closer to the efficient allocation, can it reduce the government costs?

The exercise restricts participation to the five banks with the largest average auction shares in the data, with remaining banks as secondary-market-only participants. Table 9 reports aggregate outcomes. The average clearing price rises slightly, so the central bank pays marginally less. The effect is small, which again reflects the flat demand curves the secondary market produces, but its sign is informative: excluding banks removes counterparties who would otherwise have acquired reserves at the auction, leaving more of the system short, and that raises the marginal value of reserves to the banks still bidding. The small change in values would likely be accompanied by a bigger increase in strategic bid shading as par-

choose the number of steps they submit. In the data, many bidders submit a single step, and for these players, there would be lots of potential bids that are consistent with the necessary condition which would cause bounds on predictions to be wide.

ticipants are excluded (via the conventional auction competition channels), so the absence of meaningful gains from restricting participation should suggest that this is not a fruitful policy direction for this market.

At the same time, price dispersion rises substantially when restricting to five bidders. The clearing price depends much more heavily on individual realized positions. Mean bidder expected value is essentially unchanged and its dispersion across banks falls slightly.

Table 9: Participation Restriction

Prices are spreads to the overnight target in percentage points. Allocations are in billions of dollars. V is expected private value at the counterfactual allocation, including expected secondary market interest flows and relationship rate advantages and is in units of \$10M, u is only the component of utility from end of day positions and excludes all trading benefits and payments. Bank groups are split at the median of partner volume as a share of secondary market volume, and of secondary market volume as a share of secondary market plus auction volume.

	Baseline		Top-5 only	
Aggregate outcomes – fixed distribution				
Avg. clearing price	-0.0983		-0.0956	
Std. clearing price	0.0076		0.0166	
Mean V (\$10M)	-0.0200		-0.0203	
Std. $\bar{V}$ across banks	0.0211		0.0194	
Mean u	-0.0037		-0.0039	
Std. u across banks	0.0004		0.0005	
Bank outcomes – fixed distribution				
	N	$\bar{V}$ (\$B)	$\bar{q}$ (\$B)	$\bar{V}$ (\$B)	$\bar{q}$ (\$B)
High partner volume	6	-0.0124	0.085	-0.0104	0.049
Low partner volume	5	-0.0115	0.034	-0.0203	0.077
High SM share	6	-0.0099	0.066	-0.0186	0.112
Low SM share	5	-0.0145	0.057	-0.0105	0.001
In top-5	5	-0.0124	0.085	-0.0137	0.136
Not in top-5	6	-0.0116	0.043	-0.0159	0.000

The lower panel shows who is impacted by the reallocation. Under both groupings, the group that receives more reserves is worse off. Banks with high secondary market shares roughly double their allocations, however their expected private value falls since with the additional quantity they are more likely to end up with excess supply of balances. Banks with low secondary market shares are pushed almost entirely out of the auction and their expected private value improves. The split by partner volume shows the same logic with the roles exchanged: the high-volume group gives up quantity and gains, the low-volume group absorbs it and loses. This is a consequence of the fact that reserves are costly to carry in

either direction. Both those in the top 5 and those outside obtain lower values.

Since the distribution of pre-auction balances, may respond in equilibrium, as forward looking bidders change their positions before the auction we compute outcomes under both a fixed distribution of initial positions and with initial positions that adjust to a new equilibrium. In particular, we imagine a simple model where banks choose a target pre-auction balance which is then perturbed by a bank-specific shock (reflecting things like frictions from trading opportunities). Carrying a higher target is costly, and we take that cost to be linear, with constant marginal cost α_i. Since the marginal values conditional on any balance are computable from the estimates, we recover α_i by requiring that the bank's observed average balance be optimal, $\alpha_i = \mathbb{E}[V_1^i \mid t_i = L]$. Intuitively, a bank that tends to hold large balances must find it cheap to do so. When we compute counterfactuals we hold fixed the stock of settlement balances, since that quantity is set by the central bank. Banks therefore reallocate positions among themselves, which in equilibrium requires the marginal benefit net of marginal cost be equalized across banks. The value at which that settles is the shadow price of settlement balances in the system. Details are provided in Appendix H. This provides a simple local approximation for how bidders' re-optimization of initial positions may affect outcomes. Allowing positions to adjust changes very little, so we report outcomes at the observed distribution of pre-auction positions throughout.

8.2 Varying Supply

A different way to shift allocations, and surplus in the settlement balance market is to change the supply available at the auction. Over time the Bank of Canada has experimented with substantially different levels of liquidity (see Allen et al., 2016 for details). This changes the expected trading opportunities in the secondary market, which in turn reshapes bidders' marginal values. It is unclear whether an expansion or reduction of supply would improve outcomes or reduce dispersion in expected payments across banks. Therefore, we consider both a halving and doubling of the current auction supply.

Table 10 presents the results. Larger quantities produce lower clearing prices and less price dispersion, but the magnitudes are small, again highlighting the flat marginal value curves that come from the active secondary market. With an active secondary market, the size of the primary auction has almost no influence over the price of liquidity. Banks have higher benefit from the market as supply expands, and the dispersion in values across banks

falls at the same time.

Table 10: Supply Variation

Prices are spreads to the overnight target in percentage points. Allocations are in billions of dollars. V is expected private value at the counterfactual allocation, including expected secondary market interest flows and relationship rate advantages and is in units of $10M. Bank groups are split at the median of partner volume as a share of secondary market volume, and of secondary market volume as a share of secondary market plus auction volume.

	Half Q	Baseline	Double Q
Aggregate outcomes – fixed distribution			
Avg. clearing price	-0.0906	-0.0983	-0.0991
Std. clearing price	0.0090	0.0076	0.0057
Mean V ($10M)	-0.0198	-0.0200	-0.0180
Std. V across banks	0.0206	0.0211	0.0190
Mean u	-0.0037	-0.0037	-0.0039
Std. u across banks	0.0004	0.0004	0.0004

Bank outcomes – fixed distribution							
	N	$\bar{V}$ ($B)	$\bar{q}$ ($B)	$\bar{V}$ ($B)	$\bar{q}$ ($B)	$\bar{V}$ ($B)	$\bar{q}$ ($B)
High partner volume	6	-0.0103	0.040	-0.0124	0.085	-0.0142	0.194
Low partner volume	5	-0.0138	0.020	-0.0115	0.034	-0.0095	0.039
High SM share	6	-0.0087	0.036	-0.0099	0.066	-0.0089	0.134
Low SM share	5	-0.0157	0.024	-0.0145	0.057	-0.0159	0.112

Banks with high partner volume absorb almost all of the additional supply, their allocations rise sharply across the range, and their value falls steadily as supply grows. Although these banks have higher marginal values than the others for the additional units, taking them on reduces their private value as they are more likely to end up long. Banks with low partner volume receive very little additional quantity at any supply level, however their values improve. Sorting instead by secondary market share, the high-share group has higher value than the low-share group at every supply level.

9 Conclusion

This paper develops and estimates a structural model of banks' bidding behavior at central bank liquidity auctions that accounts for a key institutional feature: imperfectly competitive bilateral trading in post-auction secondary markets. Using detailed data from Canada's Receiver General auctions and the Large Value Transfer System from 2016-2020, we estimate banks' private information about customer payment flows, their unanticipated liquidity

shocks, and their preferences for trading with different counterparties. This contributes to the literature on settlement balances by isolating the component of liquidity shocks that are known and unknown to banks. Using these estimates, we decompose utility and find substantial heterogeneity in the rents earned by different banks. The secondary market proves critical to system functioning. Without it, banks' utility variance would increase dramatically, and the historical supply of reserves would have been insufficient to prevent market clearing at the penalty rate.

From a methodological perspective, this paper makes important contributions to the empirical auction literature by extending multi-unit auction models with step-function bidding to incorporate interdependence in values and externalities from post-auction allocations. We develop novel identification arguments for settings where bidders' valuations depend on both opponents' private information and opponents' equilibrium allocations, leveraging the insight that first-order conditions can be partially invertible even when they are not globally monotone in private signals. Our estimation approach combines nonparametric methods to identify the set of invertible bids, maximum likelihood to recover the distributions of private signals and shocks, and resampling techniques to estimate preference parameters from equilibrium conditions. These methods extend naturally to other settings featuring post-auction trading in over-the-counter markets with relationship-specific frictions, such as Treasury auctions or spectrum allocation followed by secondary market resale.

The results also have important policy implications. First, they underscore the critical role of secondary markets in enabling central banks to operate with lower reserve supplies while maintaining interest rate stability. Our estimated marginal value curves are relatively flat in the own balances, so that the design of the primary auction has little effect on the price of liquidity or who ultimately holds reserves. Second, banks have good private information about their liquidity needs. Third, the scarce reserve system concentrates rents among banks with advantageous trading relationships, raising competition concerns. Policymakers seeking to reduce required reserve levels should consider both the efficiency gains from allowing secondary market reallocation against the potential for increased rent extraction and market power by well connected banks, along with the costs of running a larger central bank balance sheet.

References

Acharya, V. V. and Rajan, R. G. (2024). Liquidity, liquidity everywhere, not a drop to use: Why flooding banks with central bank reserves may not expand liquidity. *The Journal of Finance*, 79(5):2943–2991.

Afonso, G., Giannone, D., La Spada, G., and Williams, J. C. (2022). Scarce, abundant, or ample? A time-varying model of the reserve demand curve. Staff Reports 1019, Federal Reserve Bank of New York. Revised 2025.

Afonso, G., Kovner, A., and Schoar, A. (2011). Relationships matter: The importance of relationships in the fed funds market. *SSRN Electronic Journal*. Available at SSRN: https://ssrn.com/abstract=1787382.

Afonso, G. and Lagos, R. (2015). Trade dynamics in the market for federal funds. *Econometrica*, 83(1):263–313.

Allen, J., Chapman, J., Echenique, F., and Shum, M. (2016). Efficiency and bargaining power in the interbank loan market. *International Economic Review*, 57(2):691–716.

Allen, J., Hortaçsu, A., and Kastl, J. (2021). Crisis management in canada: Analyzing default risk and liquidity demand during financial stress. *American Economic Journal: Microeconomics*, 13(2):243–275.

Baglioni, A. (2024). Monetary policy implementation: Which "new normal"? *Journal of International Money and Finance*, 141:102998.

Bank of Canada (2024). Reintroduction of the morning auction of receiver general cash balances. https://www.bankofcanada.ca/2024/02/reintroduction-morning-auction-receiver-general-cash-balances/. Market notice, 2 February 2024.

Bank of Canada (2025a). Bank of canada announces an adjustment to the deposit rate and some changes to terms and conditions for overnight reverse repo operations. https://www.bankofcanada.ca/2025/01/bank-canada-adjustment-deposit-rate-terms-overnight-reverse-repo-operations/. Market notice, 29 January 2025.

Bank of Canada (2025b). Bank of canada reduces policy rate by 25 basis points to 3%, announces end of quantitative tightening. https://www.bankofcanada.ca/2025/01/fad-press-release-2025-01-29/. Press release, 29 January 2025.

Bank of England (2024). Transitioning to a repo-led operating framework. https://www.bankofengland.co.uk/paper/2024/dp/transitioning-to-a-repo-led-operating-framework. Discussion Paper.

Bech, M. L., Chapman, J., and Garratt, R. (2013). The impact of unconventional monetary policy on the overnight interbank market. Mimeo.

Budish, E. and Bhave, A. (2023). Primary-market auctions for event tickets: Eliminating the rents of "bob the broker"? *American Economic Journal: Microeconomics*, 15(1):142–70.

Chapman, James T. E. (2006). *The Economics of Payments: With Applications to the Canadian Wholesale Payment System*. PhD thesis, University of Iowa, Iowa City, Iowa, USA. Ph.D. thesis.

Chiu, J., Eisenschmidt, J., and Monnet, C. (2020). Relationships in the interbank market. *Review of Economic Dynamics*, 35:170–191.

Chu, P., Johnson, G., Kinnear, S., McGuinness, K., and McNeely, M. (2022). Settlement balances deconstructed. Staff Discussion Paper 2022-13, Bank of Canada.

Cui, W., Wright, R., and Zhu, Y. (2025). Endogenous liquidity and capital reallocation. *Journal of Political Economy*, 133(1):146–189.

Duygan-Bump, B. and Kahn, R. J. (2026). The central bank balance-sheet trilemma. Feds notes, Board of Governors of the Federal Reserve System.

Ennis, H. M. and Keister, T. (2008). Understanding monetary policy implementation. *Economic Quarterly*, 94(Sum):235–263.

European Central Bank (2024). ECB announces changes to the operational framework for implementing monetary policy. https://www.ecb.europa.eu/press/pr/date/2024/html/ecb.pr240313_1~a3a50a9add.en.html. Press release, 13 March 2024.

Furfine, C. H. (1999). The microstructure of the federal funds market. *Financial Markets, Institutions & Instruments*, 8(5):24–44.

Gravelle, T., Morrow, R., and Witmer, J. (2023). Reviewing canada's monetary policy implementation system: Does the evolving environment support maintaining a floor system? Staff Discussion Paper 2023-10, Bank of Canada, Ottawa, Ontario, Canada.

Haile, P. A. (2001). Auctions with resale markets: An application to u.s. forest service timber sales. *American Economic Review*, 91(3):399–427.

Haile, P. A. (2003). Auctions with private uncertainty and resale opportunities. *Journal of Economic Theory*, 108(1):72–110.

Hauser, A. (2023). 'less is more' or 'less is a bore'? re-calibrating the role of central bank reserves. https://www.bankofengland.co.uk/speech/2023/november/andrew-hauser-keynote-speech-bank-of-england-watchers-conference. Speech at the Bank of England Watchers' Conference.

Hortaçsu, A. and McAdams, D. (2010). Mechanism choice and strategic bidding in divisible good auctions: An empirical analysis of the turkish treasury auction market. *Journal of Political Economy*, 118(5):833–865.

Hortaçsu, A. and Kastl, J. (2012). Valuing dealers' informational advantage: A study of canadian treasury auctions. *Econometrica*, 80(6):2511–2542.

Hu, Y. and Schennach, S. M. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica*, 76(1):195–216.

Kastl, J. (2011). Discrete bids and empirical inference in divisible good auctions. *Review of Economic Studies*, 78(3):974–1014.

Kastl, J. (2012). On the properties of equilibria in private-value divisible good auctions with constrained bidding. *Journal of Mathematical Economics*, 48(6):339–352.

Kent, C. (2024). The future system for monetary policy implementation. https://www.rba.gov.au/speeches/2024/sp-ag-2024-04-02.html. Bloomberg Australia Briefing, Sydney, 2 April 2024.

Kent, C. (2025). The RBA's monetary policy implementation system—some important updates. https://www.rba.gov.au/speeches/2025/sp-ag-2025-04-02.html. KangaNews Debt Capital Markets Summit, Sydney, 2 April 2025.

López-Salido, D. and Vissing-Jorgensen, A. (2023). Reserve demand, interest rate control, and quantitative tightening. Working paper, Board of Governors of the Federal Reserve System.

Pagnozzi, M. and Saral, K. J. (2017). Demand reduction in multi-object auctions with resale: An experimental analysis. *The Economic Journal*, 127(607):2702–2729.

Perli, R. (2025). The evolution of the federal reserve's monetary policy implementation framework. https://www.newyorkfed.org/newsevents/speeches/2025/per250522. Speech, 22 May 2025.

Richert, E. (2021). Indirect inference techniques as an alternative to direct computation of equilibrium bid distributions. Working paper.

Saporta, V. (2025). Learning by doing. https://www.bankofengland.co.uk/speech/2025/june/victoria-saporta-speech-at-the-bank-of-finland-and-suerf-conference-monetary-pol Speech at the Bank of Finland and SUERF Conference, 11 June 2025; also SUERF Policy Note No. 371.

Schnabel, I. (2024). The eurosystem's operational framework. https://www.ecb.europa.eu/press/key/date/2024/html/ecb.sp240314~8b609de772.en.html. Speech, 14 March 2024.

Selgin, G. (2019). The fed's new operating framework: How we got here and why we shouldn't stay. *Cato Journal*, 39(2):317–326.

Somaini, P. (2020). Identification in auction models with interdependent costs. *Journal of Political Economy*, 128(10):3820–3871.

Wright, R., Xiao, S., and Zhu, Y. (2020). Frictional capital reallocation with ex post heterogeneity. *Review of Economic Dynamics*, 37(None):227–253.

Zhang, N. (2012). Estimating the demand for settlement balances in the canadian large value transfer system. Working Paper 2012-15, Bank of Canada. Funds Management and Banking Department.

A Proof of Proposition 1

Proof of Proposition 1(1) Let $\tilde{p}^c$ be the clearing price after bidder i deviating from q_{ik} to q'_{ik}. Then, if $p^c < b_{ik+1}$ then $p^c = \tilde{p}^c$ because

$$\{p^c < b_{ik+1}\} = \left\{ Q > \sum_{j \neq i} y_j(b_{ik+1}|s_j) + q_{ik+1} \right\}, \tag{9}$$

which is independent of q_{ik}. Moreover, the quantity that bidder i wins is also unchanged by the deviation. Similarly, if $p^c > b_{ik}$, $p^c = \tilde{p}^c$ because

$$\{p^c > b_{ik}\} = \left\{ Q < \lim_{p \downarrow b_{ik}} \sum_{j \neq i} y_j(p|s_j) + q_{ik-1} \right\}, \tag{10}$$

which is also independent of q_{ik}. And the quantity allocated to bidder i is not changes. Therefore, the deviation changes profit through two channels: It changes utility if $b_{ik} \geq p^c \geq b_{ik+1}$; and it changes payment if $b_{ik} \geq p^c$. Therefore, we only need to focus on thes two parts. Before deviation, the relevant profit is

$$\begin{aligned}
\pi_R =& \mathbb{E} \int \mathbf{1}(b_{ik} > p^c > b_{ik+1}) V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > p^c) b_{ik}(q_{ik} - q_{ik-1}) \\
& - \mathbf{1}(b_{ik+1} > p^c) b_{ik+1}(q_{ik+1} - q_{ik}) \\
& + \mathbf{1}(b_{ik} = p^c)[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(Q^a_i - q_{ik-1})] \\
& + \mathbf{1}(b_{ik+1} = p^c)[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - b_{ik+1}(Q^a_i - q_{ik})] d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}).
\end{aligned}$$

and after the deviation

$$\begin{aligned}
\pi'_R =& \mathbb{E} \int \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}) V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > \tilde{p}^c) b_{ik}(q'_{ik} - q_{ik-1}) \\
& + \mathbf{1}(b_{ik+1} > \tilde{p}^c) b_{ik+1}(q_{ik+1} - q'_{ik}) \\
& + \mathbf{1}(b_{ik} = \tilde{p}^c)[V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}^c_i - q_{ik-1})] \\
& + \mathbf{1}(b_{ik+1} = \tilde{p}^c)[V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}^c_i - q'_{ik})] d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}),
\end{aligned}$$

where $\tilde{Q}_i^c$ and $\tilde{\boldsymbol{Q}}_{-i}^c$ are the quantities allotted to bidder i and the rest of the bidders, and $\mathbb{E}$ is taken with respect to the random quantity Q. Then,

$$\pi_R - \pi_R' = \mathbb{E} \int (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4 + \Delta_5)\, d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}).$$

where

$$\begin{aligned}
\Delta_1 =& \mathbf{1}(b_{ik} > p^c > b_{ik+1}) V_i(q_{ik} + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}) V_i(q_{ik}' + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}),\\
\Delta_2 =& -\mathbf{1}(b_{ik} > p^c) b_{ik}(q_{ik} - q_{ik-1}) + \mathbf{1}(b_{ik} > \tilde{p}^c) b_{ik}(q_{ik}' - q_{ik-1}),\\
\Delta_3 =& \mathbf{1}(b_{ik} = p^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})]\\
& - \mathbf{1}(b_{ik} = \tilde{p}^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})]\\
\Delta_4 =& \mathbf{1}(b_{ik+1} = p^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik+1}(Q_i^a - q_{ik})]\\
& - \mathbf{1}(b_{ik+1} = \tilde{p}^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q_{ik}')]\\
\Delta_5 =& -\mathbf{1}(b_{ik+1} > p^c) b_{ik+1}(q_{ik+1} - q_{ik}) + \mathbf{1}(b_{ik+1} > \tilde{p}^c) b_{ik+1}(q_{ik+1} - q_{ik}')
\end{aligned}$$

Now we study each of the terms. First, notice

$$\begin{aligned}
\Delta_1 =& \mathbf{1}(b_{ik} > p^c > b_{ik+1}) \left[V_i(q_{ik} + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - V_i(q_{ik}' + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i})\right]\\
& + \left[\mathbf{1}(b_{ik} > p^c > b_{ik+1}) - \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1})\right] V_i(q_{ik}' + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}).
\end{aligned}$$

Moreover, $\tilde{p}^c \leq p^c$ and recall that if $p^c > b_{ik}$ or $p^c < b_{ik+1}$, then $\tilde{p}^c = p^c$, which implies $\tilde{p}^c > b_{ik}$ or $\tilde{p}^c < b_{ik+1}$, then $\tilde{p}^c = p^c$. Therefore,

$$\begin{aligned}
\mathbf{1}(b_{ik} > p^c > b_{ik+1}) =& \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik} > \tilde{p}^c > b_{ik+1})\\
& + \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik} = \tilde{p}^c)\\
& + \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} = \tilde{p}^c)\\
& + \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} > \tilde{p}^c)\\
& + \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik} < \tilde{p}^c).
\end{aligned}$$

The last two terms are 0 because if $\tilde{p}^c > b_{ik}$ or $\tilde{p}^c < b_{ik+1}$, then $\tilde{p}^c = p^c$. The second term is

0 because $\tilde{p}^c \leq p^c$. We can then write

$$\mathbf{1}(b_{ik} > p^c > b_{ik+1}) = \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik} > \tilde{p}^c > b_{ik+1}) + \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} = \tilde{p}^c).$$

Similarly,

$$\mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}) = \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}, b_{ik} > p^c > b_{ik+1}) + \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}, b_{ik} = p^c).$$

Combine the above two equalities to obtain

$$\begin{aligned}
&\mathbf{1}(b_{ik} > p^c > b_{ik+1}) - \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}) \\
=&\mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} = \tilde{p}^c) - \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}, b_{ik} = p^c).
\end{aligned}$$

We can then rewrite

$$\begin{aligned}
\Delta_1 =&\mathbf{1}(b_{ik} > p^c > b_{ik+1}) \left[V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right] \\
&+ \left[\mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} = \tilde{p}^c) - \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}, b_{ik} = p^c)\right] V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}).
\end{aligned}$$

Next notice that

$$\begin{aligned}
\Delta_2 =& -\mathbf{1}(b_{ik} > p^c) b_{ik}(q_{ik} - q_{ik-1}) + \mathbf{1}(b_{ik} > p^c) b_{ik}(q'_{ik} - q_{ik-1}) \\
&- \mathbf{1}(b_{ik} > p^c) b_{ik}(q'_{ik} - q_{ik-1}) + \mathbf{1}(b_{ik} > \tilde{p}^c) b_{ik}(q'_{ik} - q_{ik-1}) \\
=& -\mathbf{1}(b_{ik} > p^c) b_{ik} \left(q_{ik} - q'_{ik}\right) - b_{ik}(q'_{ik} - q_{ik-1}) \left[\mathbf{1}(b_{ik} > p^c) - \mathbf{1}(b_{ik} > \tilde{p}^c)\right].
\end{aligned}$$

Notice that

$$\begin{aligned}
\mathbf{1}(b_{ik} > p^c) =& \mathbf{1}(b_{ik} > p^c, b_{ik} > \tilde{p}^c) + \mathbf{1}(b_{ik} > p^c, b_{ik} = \tilde{p}^c) + \mathbf{1}(b_{ik} > p^c, b_{ik} < \tilde{p}^c) \\
=& \mathbf{1}(b_{ik} > p^c, b_{ik} > \tilde{p}^c) + \mathbf{1}(b_{ik} > p^c, b_{ik} = \tilde{p}^c) = \mathbf{1}(b_{ik} > p^c, b_{ik} > \tilde{p}^c),
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{1}(b_{ik} > \tilde{p}^c) =& \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} > p^c) + \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} = p^c) + \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} < p^c) \\
=& \mathbf{1}(b_{ik} > p^c, b_{ik} > \tilde{p}^c) + \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} = p^c).
\end{aligned}$$

This implies that

$$\Delta_2 = -\mathbf{1}(b_{ik} > p^c) b_{ik} \left(q_{ik} - q'_{ik}\right) + \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} = p^c) b_{ik} (q'_{ik} - q_{ik-1}).$$

Also it is worth noting that $\mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} = p^c)$ is equivalent to $\mathbf{1}(q_{ik} \geq Q^c > q_{ik} - \epsilon)$.

Next,

$$\begin{aligned}
\Delta_3 =& \mathbf{1}(b_{ik} = p^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})] \\
& - \mathbf{1}(b_{ik} = \tilde{p}^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{\boldsymbol{Q}}_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})] \\
=& \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})] \\
& - \mathbf{1}(b_{ik} = \tilde{p}^c, b_{ik} = p^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{\boldsymbol{Q}}_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})] \\
& - \mathbf{1}(b_{ik} = \tilde{p}^c, b_{ik} < p^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{\boldsymbol{Q}}_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})] \\
& + \mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})] \qquad (11)
\end{aligned}$$

where the second last term is 0 because the deviation can only reduce the clearing price. We can then decompose

$$\Delta_3 = \hat{\Delta}_3 + \Delta_{3a} + \Delta_{3b},$$

Here

$$\hat{\Delta}_3 = \mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})].$$

Notice that if $b_{ik} = p^c$ and $b_{ik} > \tilde{p}^c$, $q_{ik} - \epsilon < Q_i^a < q_{ik}$ and $q_{ik} - \epsilon < \tilde{Q}_i^c < q_{ik}$. Δ_{3a} is the first two terms of (11) when there are no ties before the deviation:

$$\begin{aligned}
\Delta_{3a} =& \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, no\ tie)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})] \\
& - \mathbf{1}(b_{ik} = \tilde{p}^c, b_{ik} = p^c, no\ tie)[V_i(\tilde{Q}_i^c + s_i, \tilde{\boldsymbol{Q}}_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})] = 0
\end{aligned}$$

where the second equality holds because if there is no tie and $b_{ik} = p^c$, $b_{ik} = \tilde{p}^c$, then $q_{ik-1} < Q_i^a \leq q_{ik} - \epsilon$. This implies $Q_i^a = \tilde{Q}_i^c$ and $\boldsymbol{Q}_{-i}^c = \tilde{\boldsymbol{Q}}_{-i}^c$. Δ_{3a} is the first two terms of

(11) when there are ties before the deviation:

$$\begin{aligned}\Delta_{3b} =& \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie)[V(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik}(Q_i^a - q_{ik-1})] \\ & - \mathbf{1}(b_{ik} = \tilde{p}^c, b_{ik} = p^c, tie)[V(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}_i^c - q_{ik-1})] \\ =& \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie)[V(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - V(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i})] \\ & - \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie) b_{ik}(Q_i^a - \tilde{Q}_i^c)\end{aligned}$$

Next, use a similar argument, we can obtain

$$\begin{aligned}\Delta_4 =& \mathbf{1}(b_{ik+1} = p^c, b_{ik+1} = \tilde{p}^c)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik+1}(Q_i^a - q_{ik})] \\ & - \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q'_{ik})] \\ & - \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q'_{ik})] \\ =& \Delta_{4a} + \Delta_{4b} + \hat{\Delta}_4.\end{aligned}$$

Here

$$\hat{\Delta}_4 = -\mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q'_{ik})]$$

And

$$\begin{aligned}\Delta_{4a} =& \mathbf{1}(b_{ik+1} = p^c, b_{ik+1} = \tilde{p}^c, no\ tie)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik+1}(Q_i^a - q_{ik})] \\ & - \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, no\ tie)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q'_{ik})] \qquad (12) \\ =& \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, no\ tie) b_{ik+1}(q_{ik} - q'_{ik}). \qquad (13)\end{aligned}$$

The second equality holds because if $b_{ik+1} = p^c$ and there is no tie, deviating to q'_{ik} cannot lead to ties if $b_{ik+1} = \tilde{p}^c$. On this event, the quantity alloted to any bidder is independent of q_{ik}, i.e., $Q_i^a = \tilde{Q}_i^c$ and $\boldsymbol{Q}_{-i}^c = \tilde{\boldsymbol{Q}}_{-i}^c$. Moreover,

$$\begin{aligned}\Delta_{4b} =& \mathbf{1}(b_{ik+1} = p^c, b_{ik+1} = \tilde{p}^c, tie)[V_i(Q_i^a + s_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{s}_{-i}) - b_{ik+1}(Q_i^a - q_{ik})] \\ & - \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, tie)[V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}_i^c - q'_{ik})]\end{aligned}$$

Lastly, if $\tilde{p}^c < b_{ik+1}$ or $p^c < b_{ik+1}$, both p^c and $\tilde{p}^c$ are not affected by the deviation and

hence $p^c = \tilde{p}^c$. Therefore,

$$\Delta_5 = \mathbf{1}(b_{ik+1} > p^c) b_{ik+1}(q_{ik} - q'_{ik})$$

Therefore, we can obtain

$$\pi_R - \pi'_R = \mathbb{E} \int \Delta_\pi d\boldsymbol{F}_{s_{-i}}(s_{-i})$$

where

$$\begin{aligned}
\Delta_\pi =& \mathbf{1}(b_{ik} > p^c > b_{ik+1}) \left[V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) \right] \\
&+ \mathbf{1}(b_{ik} > p^c > b_{ik+1}, b_{ik+1} = \tilde{p}^c) V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) \\
&- \mathbf{1}(b_{ik} > \tilde{p}^c > b_{ik+1}, b_{ik} = p^c) V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) && (\Delta_1) \\
&- \mathbf{1}(b_{ik} > p^c) b_{ik} (q_{ik} - q'_{ik}) + \mathbf{1}(b_{ik} > \tilde{p}^c, b_{ik} = p^c) b_{ik}(q'_{ik} - q_{ik-1}) && (\Delta_2) \\
&+ \mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c) V_i(Q^a_i + s_i, \boldsymbol{Q}^c_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c) b_{ik}(Q^a_i - q_{ik-1}) && (\hat{\Delta}_3) \\
&+ \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie)[V(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})] \\
&- \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie) b_{ik}(Q^a_i - \tilde{Q}^c_i) && (\Delta_{3b}) \\
&- \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c) V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) \\
&+ \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c) b_{ik+1}(\tilde{Q}^c_i - q'_{ik}) && (\hat{\Delta}_4) \\
&+ \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, no\ tie) b_{ik+1}(q_{ik} - q'_{ik}) && (\Delta_{4a}) \\
&+ \mathbf{1}(b_{ik+1} = p^c, b_{ik+1} = \tilde{p}^c, tie)[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - b_{ik+1}(Q^a_i - q_{ik})] \\
&- \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, tie)[V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - b_{ik+1}(\tilde{Q}^c_i - q'_{ik})] && (\Delta_{4b}) \\
&+ \mathbf{1}(b_{ik+1} > p^c) b_{ik+1}(q_{ik} - q'_{ik}). && (\Delta_5)
\end{aligned}$$

We next combine terms to simplify this expression. First notice that we can combine the two purple terms to obtain

$$\begin{aligned}
D_1 =& \mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c < b_{ik}\right) \left[V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) \right] \\
&- \mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik} = p^c\right) V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}).
\end{aligned}$$

Notice that $b_{ik+1} = \tilde{p}^c$ and $b_{ik+1} < p^c < b_{ik}$ only if

$$q_{ik} + \lim_{p \downarrow b_{ik+1}} \sum_{j \neq i} y_j(p|s_j) > Q, \; q_{ik} - \epsilon + \lim_{p \downarrow b_{ik+1}} \sum_{j \neq i} y_j(p|s_j) < Q. \tag{14}$$

Because Q has a continuous distribution, this has a probability that converges to 0 as $\epsilon \to 0$, i.e., $\mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c < b_{ik}\right) = o_p\left(1\right)$, where $o_p(1)$ means converges to 0 in probability. Moreover, $\tilde{p}^c = b_{ik+1}$ implies $q_{ik} - \epsilon < \tilde{Q}_i^c$, and $b_{ik+1} < p^c < b_{ik}$ implies $\tilde{Q}_i^c < q_{ik}$ because bidder i cannot win more by bidding less. Therefore, $\left|\tilde{Q}_i^c - q'_{ik}\right| < 2\epsilon$. Then by the smoothness of V_i

$$V_i(q'_{ik} + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) - V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) = O_p\left(\epsilon\right).$$

This implies

$$\begin{aligned} D_1 &= o_p\left(1\right) O_p\left(\epsilon\right) - \mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik} = p^c\right) V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}) \\ &= o_p\left(\epsilon\right) - \mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik} = p^c\right) V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}). \end{aligned}$$

Similarly, we can combine the two blue terms to obtain

$$D_2 = o_p\left(\epsilon\right) + \mathbf{1}\left(b_{ik+1} = \tilde{p}^c, b_{ik} = p^c\right) V_i(\tilde{Q}_i^c + s_i, \tilde{Q}_{-i}^c + \boldsymbol{s}_{-i}).$$

Therefore, $D_1 + D_2 = o_p\left(\epsilon\right)$. One can also show the two cyan terms together is $o_p\left(\epsilon\right)$. Moreover, the green term is also $o_p\left(\epsilon\right)$. To see this, notice that if $b_{ik+1} = \tilde{p}^c$ and $b_{ik+1} < p^c$, then $q_{ik} - \epsilon < \tilde{Q}_i^c < q_{ik}$ and $q_{ik} - \epsilon < Q_i^a < q_{ik}$. Therefore, $\tilde{Q}_i^c - q'_{ik} = o_p\left(\epsilon\right)$. Moreover, $b_{ik+1} = \tilde{p}^c$ and $b_{ik+1} < p^c$ only if (14) holds, which implies that $\mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} < p^c)$ is

$o_p(1)$. Therefore,

$$\begin{aligned}\Delta_\pi =&\mathbf{1}(b_{ik} > p^c > b_{ik+1}) \left[V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right] \\ &+ \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie)[V(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})] \\ &- \mathbf{1}(b_{ik} > p^c) b_{ik}\epsilon - \mathbf{1}(b_{ik} = p^c, b_{ik} = \tilde{p}^c, tie) b_{ik}(Q^a_i - \tilde{Q}^c_i) \\ &+ \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, no\ tie) b_{ik+1}\epsilon \\ &+ \mathbf{1}(b_{ik+1} = p^c, b_{ik+1} = \tilde{p}^c, tie)[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})] \\ &- \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, tie) b_{ik+1} \left(Q^a_i - \tilde{Q}^c_i\right) + \mathbf{1}(b_{ik+1} = \tilde{p}^c, b_{ik+1} = p^c, tie) b_{ik+1}\epsilon \\ &+ \mathbf{1}(b_{ik+1} > p^c) b_{ik+1}\epsilon + o_p(\epsilon)\end{aligned}$$

Therefore,

$$\begin{aligned}&\lim_{\epsilon\to 0} \frac{\mathbb{E}\int \Delta_\pi d\boldsymbol{F}_{s_{-i}}(s_{-i})}{\epsilon} \\ =&\mathbf{Pr}(b_{ik} > p^c > b_{ik+1})\times \\ &\left\{\mathbb{E}\left[V_{i1}(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) + V_{i2}(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) \cdot \frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} | b_{ik} > p^c > b_{ik+1}\right] - b_{ik}\right\} \\ &+ \mathbf{Pr}(b_{ik} = p^c, tie)\times \\ &\mathbb{E}\left\{V_{i1}(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) \frac{\partial Q^a_i}{\partial q_{ik}} + V_{i2}(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) \cdot \frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik}\frac{\partial Q^a_i}{\partial q_{ik}} \middle| b_{ik} = p^c, tie\right\} \\ &+ \mathbf{Pr}(b_{ik+1} = p^c, tie)\times \\ &\mathbb{E}\left\{V_{i1}(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) \frac{\partial Q^a_i}{\partial q_{ik}} + V_{i2}(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) \cdot \frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{ik}} - b_{ik+1}\frac{\partial Q^a_i}{\partial q_{ik}} \middle| b_{ik+1} = p^c, tie\right\} \\ &- \mathbf{Pr}(b_{ik+1} \geq p^c)(b_{ik} - b_{k+1})\end{aligned}$$

Equating this to 0 leads to Proposition 1(1).

Proof of Proposition 1(2)**.** Now suppose $k = K_i$. If we deviate from q_{ik}, only the price is equal to b_{ik} or less than b_{ik}. The deviation will not change anything if $b_{ik} < p^c$. So we only need to consider the terms in profit function with $b_{ik} \geq p^c$. Therefore,

$$\begin{aligned}\pi_R =&\mathbb{E}\int \mathbf{1}(b_{ik} > p^c) V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > p^c) b_{ik}(q_{ik} - q_{ik-1}) \\ &+ \mathbf{1}(b_{ik} = p^c)[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(Q^a_i - q_{ik-1})] d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}).\end{aligned}$$

And

$$\begin{aligned}\pi'_R =&\mathbb{E}\int \mathbf{1}(b_{ik} > \tilde{p}^c)V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > \tilde{p}^c)b_{ik}(q'_{ik} - q_{ik-1})\\ &+ \mathbf{1}(b_{ik} = \tilde{p}^c)[V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}) - b_{ik}(\tilde{Q}^c_i - q_{ik-1})]d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}).\end{aligned}$$

Then we can write

$$\pi_R - \pi'_R = \mathbb{E}\int (\Delta_1 + \Delta_2 + \Delta_3 + \Delta_4)\, d\boldsymbol{F}_{s_{-i}}(\boldsymbol{s_{-i}}),$$

where

$$\begin{aligned}\Delta_1 &= \mathbf{1}(b_{ik} > p^c)V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} > \tilde{p}^c)V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}),\\ \Delta_2 &= -\left[\mathbf{1}(b_{ik} > p^c)b_{ik}(q_{ik} - q_{ik-1}) - \mathbf{1}(b_{ik} > \tilde{p}^c)b_{ik}(q'_{ik} - q_{ik-1})\right],\\ \Delta_3 &= \mathbf{1}(b_{ik} = p^c)V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - \mathbf{1}(b_{ik} = \tilde{p}^c)V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}),\\ \Delta_4 &= -\left[\mathbf{1}(b_{ik} = p^c)b_{ik}(Q^a_i - q_{ik-1}) - \mathbf{1}(b_{ik} = \tilde{p}^c)b_{ik}(\tilde{Q}^c_i - q_{ik-1})\right].\end{aligned}$$

First notice that

$$\begin{aligned}\Delta_1 =&\mathbf{1}(b_{ik} > p^c)\left[V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right]\\ &+ \left[\mathbf{1}(b_{ik} > p^c) - \mathbf{1}(b_{ik} > \tilde{p}^c)\right]V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\\ =&\mathbf{1}(b_{ik} > p^c)\left[V_i(q_{ik} + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right]\\ &- \mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c)V_i(q'_{ik} + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}).\end{aligned}$$

Moreover,

$$\begin{aligned}\Delta_3 =&\mathbf{1}(b_{ik} = p^c)\left[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right]\\ &+ \left[\mathbf{1}(b_{ik} = p^c) - \mathbf{1}(b_{ik} = \tilde{p}^c)\right]V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\\ =&\mathbf{1}(b_{ik} = p^c)\left[V_i(Q^a_i + s_i, \boldsymbol{Q}^a_{-i} + \boldsymbol{s}_{-i}) - V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i})\right]\\ &+ \left[\mathbf{1}(b_{ik} = p^c, b_{ik} > \tilde{p}^c)\right]V_i(\tilde{Q}^c_i + s_i, \tilde{Q}^c_{-i} + \boldsymbol{s}_{-i}).\end{aligned}$$

Therefore,

$$\begin{aligned}\Delta_1+\Delta_3 =&\mathbf{1}(b_{ik}>p^c)\left[V_i(q_{ik}+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})-V_i(q'_{ik}+s_i,\tilde{Q}^c_{-i}+\boldsymbol{s}_{-i})\right]\\&+\mathbf{1}(b_{ik}=p^c)\left[V_i(Q^a_i+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})-V_i(\tilde{Q}^c_i+s_i,\tilde{Q}^c_{-i}+\boldsymbol{s}_{-i})\right]\\&+\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)\left[V_i(\tilde{Q}^c_i+s_i,\tilde{Q}^c_{-i}+\boldsymbol{s}_{-i})-V_i(q'_{ik}+s_i,\tilde{Q}^c_{-i}+\boldsymbol{s}_{-i})\right]. \quad (15)\end{aligned}$$

Notice that $b_{ik}=p^c$ and $b_{ik}>\tilde{p}^c$ only if

$$q_{ik}+\sum_{j\neq i}y_j(b_{ik}|s_j)>Q,\ q_{ik}-\epsilon+\sum_{j\neq i}y_j(b_{ik}|s_j)<Q.$$

This implies that $\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)=o_p(1)$ if $\epsilon\to 0$ and $q_{ik}-\epsilon=\tilde{Q}^c_i<Q^a_i<q_{ik}$. The latter implies that $\left|\tilde{Q}_i-q'_{ik}\right|<2\epsilon$. Therefore, the last term in (15) is $o_p(\epsilon)$. Similarly,

$$\Delta_2=-\left[\mathbf{1}(b_{ik}>p^c)b_{ik}(q_{ik}-q'_{ik})-\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)b_{ik}(q'_{ik}-q_{ik-1})\right],$$

and

$$\Delta_4=-\left[\mathbf{1}(b_{ik}=p^c)b_{ik}\left(Q^a_i-\tilde{Q}^c_i\right)-\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)b_{ik}(\tilde{Q}^c_i-q_{ik-1})\right].$$

Use a similar argument as the above, one can show both $\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)b_{ik}(q'_{ik}-q_{ik-1})$ and $\mathbf{1}(b_{ik}=p^c,b_{ik}>\tilde{p}^c)b_{ik}(\tilde{Q}^c_i-q_{ik-1})$ are $o_p(\epsilon)$. Then

$$\Delta_2+\Delta_4=-\mathbf{1}(b_{ik}>p^c)b_{ik}(q_{ik}-q'_{ik})-\mathbf{1}(b_{ik}=p^c)b_{ik}\left(Q^a_i-\tilde{Q}^c_i\right)+o_p(\epsilon).$$

Therefore, the FOC can be written as

$$\begin{aligned}0=&\mathbf{Pr}(b_{ik}>p^c)\times\\&\mathbb{E}\left[V_{1i}(q_{ik}+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})+V_{2i}(q_{ik}+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})\cdot\frac{\partial\boldsymbol{Q}^a_{-i}}{\partial q_{ik}}-b_{ik}\middle|b_{ik}>p^c\right]\\&+\mathbf{Pr}(b_{ik}=p^c)\times\\&\left[V_{1i}(Q^a_i+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})\frac{\partial Q^a_i}{\partial q_{ik}}+V_{2i}(Q^a_i+s_i,\boldsymbol{Q}^a_{-i}+\boldsymbol{s}_{-i})\cdot\frac{\partial\boldsymbol{Q}^a_{-i}}{\partial q_{ik}}-b_{ik}\frac{\partial Q^a_i}{\partial q_{ik}}\middle|b_{ik}=p^c\right]\end{aligned}$$

B Proofs for Propositions in Section 5

Proof of Proposition 2. By Assumption 2, if $\tilde{\phi} \in \mathcal{Q}_i$, there is a unique $\tilde{\sigma}$ in the support of σ_i such that $\sigma_i = \tilde{\sigma}$ if $\phi_i(y_i) = \tilde{\phi}$. Because ϵ_i is independent of σ_i and y_i

$$\mathbb{E}\left[\hat{q}_i | \phi_i(y_i) = \tilde{\phi}\right] = \tilde{\sigma}.$$

This implies that

$$\begin{aligned}
&\mathbf{Pr}\left(\hat{q}_i - \mathbb{E}\left(\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right) < x \middle|\, \phi_i(y_i) = \tilde{\phi}\right) \\
=&\mathbf{Pr}\left(\tilde{\sigma} + \epsilon_i - \tilde{\sigma} < x |\, \phi_i(y_i) = \tilde{\phi}\right) = \mathbf{Pr}\left(\epsilon_i < x\right) = F_{\epsilon_i}(x).
\end{aligned}$$

The second last equality follows by the independence between ϵ_i and y_i. Notice that $\mathbb{E}\left(\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right)$ and

$$\mathbf{Pr}\left(\hat{q}_i - \mathbb{E}\left(\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right) < x \middle|\, \phi_i(y_i) = \tilde{\phi}\right)$$

are both identified from data given $\tilde{\phi}$. Therefore, if we can find a pair $(b, q) \in \mathcal{Q}_i$, F_{ϵ_i} is identified. We next show how to find such a pair.

To this end, we show that $\mathbf{var}\left(\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right)$ is minimized if and only if $\tilde{\phi} \in \mathcal{Q}_i$, which allows us to minimize this conditional variance to find a such pair. To see this, notice that

$$\begin{aligned}
&\mathbf{var}\left[\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right] \\
=&\mathbb{E}\left\{\left[\hat{q}_i - \mathbb{E}\left(\hat{q}_i |\, \phi_i(y_i) = \tilde{\phi}\right)\right]^2 \middle|\, \phi_i(y_i) = \tilde{\phi}\right\} \\
=&\mathbb{E}\left\{\left[\sigma_i + \epsilon_i - \mathbb{E}\left(\sigma_i + \epsilon_i |\, \phi_i(y_i) = \tilde{\phi}\right)\right]^2 \middle|\, \phi_i(y_i) = \tilde{\phi}\right\} \\
=&\mathbb{E}\left\{\left[\sigma_i + \epsilon_i - \mathbb{E}\left(\sigma_i + \epsilon_i |\, \phi_i(y_i) = \tilde{\phi}\right)\right]^2 \middle|\, \phi_i(y_i) = \tilde{\phi}\right\} \\
=&\mathbb{E}\left\{\left[\sigma_i - \mathbb{E}\left(\sigma_i |\, \phi_i(y_i) = \tilde{\phi}\right)\right]^2 \middle|\, \phi_i(y_i) = \tilde{\phi}\right\} + \mathbb{E}\epsilon_i^2
\end{aligned}$$

The first term is 0 if and only if $\tilde{\phi} \in \mathcal{Q}_i$, which ensures the distribution of σ_i conditional on $\phi_i(y_i) = \tilde{\phi}$ is degenerate. This conditional variance is identified from data and we can first find all its minimizers to identify $\mathcal{Q}_i$.

Lastly, let ψ_{ϵ_i} be the characteristic function of F_{ϵ_i}. It is identified because F_{ϵ_i} is known. Let $\mathcal{Z}_0 = \{t \in \mathbb{R} : \psi_{\epsilon_i}(t) = 0\}$. The characteristic function of $\hat{q}_i$, denoted by $\psi_{\hat{q}_i}$ is identified because we observe $\hat{q}_i$ regardless whether bank i participated in the auction or not. If $t \in \mathbb{R}/\mathcal{Z}_0$, then $\psi_{\sigma_i}(t) = \phi_{\hat{q},i}(t)/\psi_{\epsilon_i}(t)$. Because characteristic functions are continuous, $\psi_{\sigma_i}(t)$ can be uniquely extended to $\mathbb{R}/\mathcal{Z}_0$. Then F_{σ_i} is identified. □

Proof of Proposition 3. Define

$$\begin{aligned}
\tilde{V}_i\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) =& \mathbb{E}\left[V_i^e \middle| Q_i^a, \sigma_i, \boldsymbol{Q}_{-i}^a, \boldsymbol{\sigma}_{-i}, \boldsymbol{\epsilon}_{-i}\right] \\
=& A_i\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) + \sum_{\substack{j\in\{n,p\}\\ m\in\{b,l\}}} B_{ji}^m E_{ji}^m\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) \\
=& A_i\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) + \boldsymbol{E}\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right)\boldsymbol{B}_i'. \qquad (16)
\end{aligned}$$

where $\boldsymbol{E} = \left(E_{ni}^b, E_{pi}^b, E_{ni}^l, E_{pi}^l\right)$ and

$$A_i\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) = \mathbb{E}\left[u\left(Q_i^e\right) + \sum_{j\in\{n,p\}}\left(i_{ji}^l\chi_{ji}^l - i_{ji}^b\chi_{ji}^b\right)\middle| \boldsymbol{Q}^a, \boldsymbol{\sigma}, \boldsymbol{\epsilon}_{-i}\right].$$

We now show that E_{ji}^m is identified $\forall m \in \{b,l\}$ and $j \in \{n,p\}$. By iterated expectation

$$E_{ji}^m\left(Q_i^a + \sigma_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right) = \mathbb{E}\left(\chi_{ji}^m \middle| \boldsymbol{Q}^a, \boldsymbol{\sigma}, \boldsymbol{\epsilon}_{-i}\right) = \mathbb{E}\left[\mathbb{E}\left(\chi_{ji}^m \middle| \boldsymbol{Q}^a, \boldsymbol{\sigma}, \boldsymbol{\epsilon}\right) \middle| \boldsymbol{Q}^a, \boldsymbol{\sigma}, \boldsymbol{\epsilon}_{-i}\right]. \qquad (17)$$

Because χ_{ji}^m and $\boldsymbol{Q}^a + \boldsymbol{\sigma} + \boldsymbol{\epsilon}$ are observed, we can identify

$$\mathbb{E}\left(\chi_{ji}^m \middle| \boldsymbol{Q}^a, \boldsymbol{\sigma}, \boldsymbol{\epsilon}\right) = \mathbb{E}\left(\chi_{ji}^m \middle| \boldsymbol{Q}^a + \boldsymbol{\sigma} + \boldsymbol{\epsilon}\right) = \tilde{E}_{ji}^m\left(Q_i^a + \sigma_i + \epsilon_i, \boldsymbol{Q}_{-i}^a + \boldsymbol{\sigma}_{-i} + \boldsymbol{\epsilon}_{-i}\right).$$

Because F_{ϵ_i} are identified by Proposition 2,

$$E_{ji}^m(x,y) = \int \tilde{E}_{ji}^m(x+\epsilon, y)\, dF_{\epsilon_i}(\epsilon)$$

is also identified. The function A_i can be identified in the same fashion.

Because $\boldsymbol{\epsilon}$ is independent of p^c, $\boldsymbol{Q}^a$, Q and $\boldsymbol{\sigma}$, the FOC for the last step can be written

as

$$
\begin{aligned}
0 =& \mathbf{Pr}(b_{iK_i} > p^c)\mathbb{E}\Bigg[\tilde{V}_{1i}\left(q_{iK_i}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right) \\
&+\tilde{V}_{2i}\left(q_{iK_i}+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{iK_i}} - b_{iK_i}\Bigg| b_{iK_i} > p^c\Bigg] \\
&+ \mathbf{Pr}(b_{iK_i} = p^c)\mathbb{E}\Bigg[\tilde{V}_{1i}\left(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\frac{\partial Q^a_i}{\partial q_{iK_i}} \\
&+\tilde{V}_{2i}\left(Q^a_i+\sigma_i, \boldsymbol{Q}^a_{-i}+\boldsymbol{\sigma}_{-i}+\boldsymbol{\epsilon}_{-i}\right)\cdot\frac{\partial \boldsymbol{Q}^a_{-i}}{\partial q_{iK_i}} - b_{iK_i}\frac{\partial Q^a_i}{\partial q_{iK_i}}\Bigg| b_{iK_i} = p^c\Bigg].
\end{aligned} \tag{18}
$$

We can then substitute the expression for $\tilde{V}_i$ into (18) and rearrange to obtain

$$
\boldsymbol{B}_i\eta(\sigma_i, y_i) = C_i(\sigma_i, y_i). \tag{19}
$$

Because A_i and E^m_{ji} are identified, both η and C_i are identified. Under the assumption of the proposition, $\tilde{\sigma}_t$ is idenfitied for $\tilde{y}_t$ for $t = 1, \cdots, 4$. As a result, the matrix $\boldsymbol{eta}_i$ and $\boldsymbol{C}_i$ are both identifed so is $\boldsymbol{B}_i$ because $\boldsymbol{eta}_i$ has full rank. □

C Proofs for Propositions in Section 6

Proof of Proposition 4. We start with Claim 1. Notice that $\bar{S}_L$ is equivalent to

$$
\hat{S}_L = \left\{\exists \ell \text{ such that } \phi_i(y_{i\ell}) \in \mathcal{Q}_i \text{ and } \widehat{\mathbf{var}}\left[\hat{q}_{i\ell}\middle|\, \phi_i(y_{i\ell})\right] - \min_{\ell'} \widehat{\mathbf{var}}\left[\hat{q}_{i\ell'}\middle|\, \phi_i(y_{i\ell'})\right] > \delta_L\right\}.
$$

Notice that as $L \to \infty$, there does not exist any ℓ such that $\phi_i(y_{i\ell}) \in \mathcal{Q}_i$ with probability $[1 - \mathbf{Pr}(\phi_i(y_{i\ell}) \in \mathcal{Q}_i)]^L$. Because

$$
\sum_{L=1}^{\infty}\left[1 - \mathbf{Pr}(\phi_i(y_{i\ell}) \in \mathcal{Q}_i)\right]^L = \frac{1 - \mathbf{Pr}(\phi_i(y_{i\ell}) \in \mathcal{Q}_i)}{\mathbf{Pr}(\phi_i(y_{i\ell}) \in \mathcal{Q}_i)} < \infty
$$

by the assumption that $\mathbf{Pr}(\phi_i(y_{i\ell}) \in \mathcal{Q}_i) > 0$, there exists at least one ℓ such that $\phi_i(y_{i\ell}) \in \mathcal{Q}_i$ almost surely as $L \to \infty$, and consequently, almost surely

$$
\min_{\ell} \mathbf{var}\left[\hat{q}_{i\ell}\middle|\, \phi_i(y_{i\ell})\right] = \sigma^2_\epsilon. \tag{20}
$$

By Assumption 3,

$$\begin{aligned}&\widehat{\mathbf{var}}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\min_{\ell'}\widehat{\mathbf{var}}\left[\hat{q}_{i\ell'}\middle|\,\phi_i\left(y_{i\ell'}\right)\right]\\=&\mathbf{var}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\min_{\ell'}\mathbf{var}\left[\hat{q}_{i\ell'}\middle|\,\phi_i\left(y_{i\ell'}\right)\right]+O_p\left(\xi_L\right).\end{aligned}$$

Define

$$\tilde{S}_L=\left\{\exists\ell\text{ such that }\phi_i\left(y_{i\ell}\right)\in\mathcal{Q}_i\text{ and }\mathbf{var}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\min_{\ell'}\mathbf{var}\left[\hat{q}_{i\ell'}\middle|\,\phi_i\left(y_{i\ell'}\right)\right]>\frac{\delta_L}{2}\right\}.$$

Then

$$\mathbf{Pr}\left(\bar{S}_L\right)=\mathbf{Pr}\left(\hat{S}_L\right)\leq\mathbf{Pr}\left(\tilde{S}_L\right)+\mathbf{Pr}\left(O_p\left(\xi_L\right)>\delta_L/2\right).$$

Because of (20) and $\mathbf{var}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]=\sigma_\epsilon^2$ if $\phi_i\left(y_{i\ell}\right)\in\mathcal{Q}_i$,

$$\lim_{L\to\infty}\mathbf{Pr}\left(\tilde{S}_L\right)=0$$

and

$$\lim_{L\to\infty}\mathbf{Pr}\left(O_p\left(\xi_L\right)>\delta_L/2\right)=0$$

by assumption. Therefore, for any given ℓ,

$$\lim_{L\to\infty}\mathbf{Pr}\left(\bar{S}_\ell\right)\leq\lim_{L\to\infty}\mathbf{Pr}\left(\bar{S}_L\right)=0.$$

Next, we prove the second claim. Notice that

$$\begin{aligned}\sup_{\phi_i(y_{i\ell})\in\mathcal{Q}_i}\left|\hat{\epsilon}_{i\ell}-\epsilon_{i\ell}\right|&=\sup_{\phi_i(y_{i\ell})\in\mathcal{Q}_i}\left|\hat{q}_{i\ell}-\hat{\mathbb{E}}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\epsilon_{i\ell}\right|\\&<\sup_{\ell}\left|\mathbb{E}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\hat{\mathbb{E}}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]\right|=O_p\left(\zeta_L\right).\end{aligned}$$

where the inequality uses $\hat{q}_{i\ell}=\sigma_{i\ell}+\epsilon_{i\ell}$ by definition and $\sigma_{i\ell}=\mathbb{E}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]$ if $\phi_i\left(y_{i\ell}\right)\in\mathcal{Q}_i$.

Lastly, by definition and Assumption 3$\sup_{\ell\in\hat{\Lambda}_i}\mathbf{var}\left[\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right]-\sigma_\epsilon^2=O_p\left(\delta_L\right)$. Because

$$\mathbb{E}\left\{\left[\hat{q}_{i\ell}-\mathbb{E}\left(\hat{q}_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right)\right]^2\middle|\,\phi_i\left(y_{i\ell}\right)\right\}=\mathbb{E}\left\{\left[\sigma_{i\ell}-\mathbb{E}\left(\sigma_{i\ell}\middle|\,\phi_i\left(y_{i\ell}\right)\right)\right]^2\middle|\,\phi_i\left(y_{i\ell}\right)\right\}+\sigma_\epsilon^2,$$

we have if $\ell \in \hat{\Lambda}_i$

$$\mathbf{var}\left[\hat{q}_{i\ell}|\,\phi_i\left(y_{i\ell}\right)\right] - \sigma_\epsilon^2 = \mathbb{E}\left\{\left[\sigma_{i\ell} - \mathbb{E}\left(\sigma_{i\ell}|\,\phi_i\left(y_{i\ell}\right)\right)\right]^2 \middle|\, \phi_i\left(y_{i\ell}\right)\right\} = O_p\left(\delta_L\right). \tag{21}$$

Therefore,

$$\begin{aligned}
\sup_{\phi_i(y_{i\ell})\notin Q_i, \ell\in\hat{\Lambda}_i} \mathbb{E}\left|\hat{\epsilon}_{i\ell} - \epsilon_{i\ell}\right|^2 = & \sup_{\phi_i(y_{i\ell})\notin Q_i, \ell\in\hat{\Lambda}_i} \mathbb{E}\left|\hat{q}_{i\ell} - \hat{\mathbb{E}}_{-\ell}\left[\hat{q}_{i\ell}|\,\phi_i\left(y_{i\ell}\right)\right] - \epsilon_{i\ell}\right|^2 \\
\leq & \sup_{\ell\in\hat{\Lambda}_i} \mathbb{E}\left|\sigma_{i\ell} - \hat{\mathbb{E}}_{-\ell}\left[\sigma_{i\ell}|\,\phi_i\left(y_{i\ell}\right)\right] + O_p\left(\zeta_L\right)\right|^2 \\
\leq & \sup_{\ell\in\hat{\Lambda}_i} \mathbb{E}\left(\sigma_{i\ell} - \hat{\mathbb{E}}_{-\ell}\left[\sigma_{i\ell}|\,\phi_i\left(y_{i\ell}\right)\right]\right)^2 + O_p\left(\zeta_L^2\right) \\
= & \, O_p\left(\delta_L\right) + O_p\left(\zeta_L^2\right) = O_p\left(\delta_L\right),
\end{aligned}$$

where the second last equality holds because of (21).

For the last claim, recall that if $\phi_i\left(y_{i\ell}\right) \in \mathcal{Q}_i$, $\ell \in \hat{\Lambda}_i$ with probability approaching 1. Therefore,

$$\lim\inf_{L\to\infty} \left|\hat{\Lambda}_i\right| / L \geq \mathbf{Pr}\left(\phi_i\left(y_{i\ell}\right) \in \mathcal{Q}_i\right) > 0.$$

Therefore, $\left|\hat{\Lambda}_i\right|^{-1/2} = O_p\left(L^{-1/2}\right)$. Define

$$F_{\epsilon_i}^o\left(x\right) = \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\ell\in\hat{\Lambda}_i} \mathbf{1}\left(\epsilon_{i\ell} \leq x\right).$$

$$\left|\mathbf{1}\left(\epsilon_{i\ell} \leq x\right) - \mathbf{1}\left(\hat{\epsilon}_{i\ell} \leq x\right)\right| \leq \mathbf{1}\left(\left|\epsilon_{i\ell} - x\right| < u\right) + \mathbf{1}\left(\left|\epsilon_{i\ell} - \hat{\epsilon}_{i\ell}\right| \geq u\right).$$

This implies that

$$\begin{aligned}
\left|F_{\epsilon_i}^o\left(x\right) - \hat{F}_{\epsilon_i}\left(x\right)\right| \leq & \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\ell\in\hat{\Lambda}_i} \mathbf{1}\left(\left|\epsilon_{i\ell} - x\right| < u\right) + \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\ell\in\hat{\Lambda}_i} \mathbf{1}\left(\left|\epsilon_{i\ell} - \hat{\epsilon}_{i\ell}\right| \geq u\right) \\
\leq & \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\ell\in\hat{\Lambda}_i} \mathbf{1}\left(\left|\epsilon_{i\ell} - x\right| < u\right) + \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\phi_i(y_{i\ell})\in\mathcal{Q}_i} \mathbf{1}\left(\left|\epsilon_{i\ell} - \hat{\epsilon}_{i\ell}\right| \geq u\right) \\
& + \frac{1}{\left|\hat{\Lambda}_i\right|} \sum_{\phi_i(y_{i\ell})\notin\mathcal{Q}_i, \ell\in\hat{\Lambda}_i} \mathbf{1}\left(\left|\epsilon_{i\ell} - \hat{\epsilon}_{i\ell}\right| \geq u\right).
\end{aligned}$$

First,

$$\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\ell\in\hat{\Lambda}_i}\left\{\mathbf{1}\left(\left|\epsilon_{i\ell}-x\right|<u\right)-\left[F_{\epsilon_i}\left(x+u\right)-F_{\epsilon_i}\left(x-u\right)\right]\right\}.$$

Since the indicator function $\mathbf{1}\left(\cdot\leq t\right)$ is a VC class,

$$\sup_x\left|\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\ell\in\hat{\Lambda}_i}\left\{\mathbf{1}\left(\left|\epsilon_{i\ell}-x\right|<u\right)-\left[F_{\epsilon_i}\left(x+u\right)-F_{\epsilon_i}\left(x-u\right)\right]\right\}\right|=O_p\left(L^{-1/2}\right).$$

Therefore,

$$\sup_x\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\ell\in\hat{\Lambda}_i}\mathbf{1}\left(\left|\epsilon_{i\ell}-x\right|<u\right)\leq O_p\left(L^{-1/2}\right)+2u\max_x f_{\epsilon_i}\left(x\right).$$

Next,

$$\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\phi_i(y_{i\ell})\in\mathcal{Q}_i}\mathbf{1}\left(\left|\epsilon_{i\ell}-\hat{\epsilon}_{i\ell}\right|\geq u\right)\leq\mathbf{1}\left(O_p\left(\zeta_L\right)\geq u\right).$$

Lastly,

$$\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\phi_i(y_{i\ell})\notin\mathcal{Q}_i,\ell\in\hat{\Lambda}_i}\mathbf{1}\left(\left|\epsilon_{i\ell}-\hat{\epsilon}_{i\ell}\right|\geq u\right)=\frac{\sum_{\ell=1}^L 1_{S_\ell}/L}{\left|\hat{\Lambda}_i\right|/L}O_p\left(\frac{\delta_L}{u^2}\right).$$

Because D_ℓ are iid draws, $\mathbb{E}1_{S_\ell}=\mathbf{Pr}\left(S_1\right)$ for all ℓ and

$$\left|\frac{\sum_{\ell=1}^L 1_{S_\ell}}{L}-\mathbf{Pr}\left(S_1\right)\right|\leq\frac{1}{L}\left[1-\mathbf{Pr}\left(S_1\right)\right]\mathbf{Pr}\left(S_1\right)=o\left(1/L\right)$$

because $\mathbf{Pr}\left(S_1\right)=O\left(\kappa_L\right)=o\left(1\right)$. Since $\kappa_L L\to\infty$, this implies that $\sum_{\ell=1}^L 1_{S_\ell}/L=O_p\left(\kappa_L\right)$. Therefore,

$$\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\phi_i(y_{i\ell})\notin\mathcal{Q}_i,\ell\in\hat{\Lambda}_i}\mathbf{1}\left(\left|\epsilon_{i\ell}-\hat{\epsilon}_{i\ell}\right|\geq u\right)=O_p\left(\frac{\delta_L\kappa_L}{u^2}\right).$$

Then if $\left(\delta_L\kappa_L\right)^{1/3}/\zeta_L\to\infty$,

$$\sup_x\left|F^o_{\epsilon_i}\left(x\right)-\hat{F}_{\epsilon_i}\left(x\right)\right|=O_p\left(\left(\delta_L\kappa_L\right)^{1/3}\right).$$

Because $\sup_x\left|F^o_{\epsilon_i}\left(x\right)-F_{\epsilon_i}\left(x\right)\right|=O_p\left(L^{-1/2}\right)$,

$$\sup_x\left|\hat{F}_{\epsilon_i}\left(x\right)-F_{\epsilon_i}\left(x\right)\right|=O_p\left(\left(\delta_L\kappa_L\right)^{1/3}\right).$$

Assumption 6(1) implies that $\{\log g_i(\cdot;\theta) : \theta \in \Theta\}$ is Glivenko-Cantelli, i.e.,

$$\sup_{\theta\in\Theta}\left|\frac{1}{L}\sum_{\ell=1}^{L}\log g_i(q_{i\ell};\theta) - \mathbb{E}\log g_i(q_{i\ell};\theta)\right| = o_p(1). \tag{22}$$

By the identification argument, $\mathbb{E}\log g_i(q_{i\ell};\theta)$ is uniquely maximized at $\theta = \theta_{i0}$. Let

$$\mathcal{L}(\theta) = L^{-1}\sum_{\ell=1}^{L}\log g_i(q_{i\ell};\theta),$$
$$\hat{\mathcal{L}}(\theta) = L^{-1}\sum_{\ell=1}^{L}\log \hat{g}_i(q_{i\ell};\theta).$$

Therefore, for any $\delta > 0$

$$\begin{aligned}
&\mathbf{Pr}\left(\left\|\hat{\theta}_i - \theta_{i0}\right\| > \delta\right) \\
&\leq \mathbf{Pr}\left(\sup_{\|\theta-\theta_{i0}\|>\delta}\hat{\mathcal{L}}(\theta) > \hat{\mathcal{L}}(\theta_{i0})\right) \\
&\leq \mathbf{Pr}\left(\sup_{\|\theta-\theta_{i0}\|>\delta}\mathcal{L}(\theta) + \sup_{\theta\in\Theta}\left|\mathcal{L}(\theta) - \hat{\mathcal{L}}(\theta)\right| > \hat{\mathcal{L}}(\theta_{i0}) - \mathcal{L}(\theta_{i0}) + \mathcal{L}(\theta_{i0})\right).
\end{aligned}$$

By Assumption 6(1)

$$\sup_{\theta\in\Theta}\left|\mathcal{L}(\theta) - \hat{\mathcal{L}}(\theta)\right| = o_p(1).$$

Therefore, as $L \to \infty$

$$\begin{aligned}
\mathbf{Pr}\left(\left\|\hat{\theta}_i - \theta_{i0}\right\| > \delta\right) &\leq \mathbf{Pr}\left(\sup_{\|\theta-\theta_{i0}\|>\delta}\mathcal{L}(\theta) > \mathcal{L}(\theta_{i0}) + o_p(1)\right) \\
&\leq \mathbf{Pr}\left(\sup_{\|\theta-\theta_{i0}\|>\delta}\mathbb{E}\mathcal{L}(\theta) > \mathbb{E}\mathcal{L}(\theta_{i0}) + o_p(1)\right) \to 0,
\end{aligned}$$

where the second inequality follows becasue of (22) and the convergence follows because $\sup_{\|\theta-\theta_{i0}\|>\delta}\mathbb{E}\mathcal{L}(\theta) < \mathbb{E}\mathcal{L}(\theta_{i0})$ by identification. This proves the consistency of $\hat{\theta}_i$.

Proof. To obtain the convergence rate, notice that $\hat{\theta}_i \in \mathcal{N}_i(\theta_{i0})$ with probability approaching

1. Therefore, with probability approaching 1,

$$\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial_\theta \hat{g}_i\left(\hat{q}_{i\ell};\hat{\theta}_i\right)}{\hat{g}_i\left(\hat{q}_{i\ell};\hat{\theta}_i\right)}=0.$$

By Assumption 6(2),

$$\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\hat{\theta}_i\right)}{g_i\left(\hat{q}_{i\ell};\hat{\theta}_i\right)}=O_p\left(a_L\right).$$

A Tylor expansion of the left-hand side implies

$$\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\theta_{i0}\right)}{g_i\left(\hat{q}_{i\ell};\theta_{i0}\right)}+\frac{1}{L}\sum_{\ell=1}^{L}\partial_{\theta\theta'}\log\left(\hat{q}_{i\ell};\theta_{i0}\right)\left(\hat{\theta}_i-\theta_{i0}\right)+o_p\left(\left\|\hat{\theta}_i-\theta_{i0}\right\|\right)=O_p\left(a_L\right).$$

By Assumption 6(3)

$$\frac{1}{L}\sum_{\ell=1}^{L}\frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\theta_{i0}\right)}{g_i\left(\hat{q}_{i\ell};\theta_{i0}\right)}=O_p\left(1/\sqrt{L}\right)$$

and

$$\frac{1}{L}\sum_{\ell=1}^{L}\partial_{\theta\theta'}\log\left(\hat{q}_{i\ell};\theta_{i0}\right)=H_i\left(\theta_{i0}\right)+o_p\left(1\right).$$

Therefore,

$$H_i\left(\theta_{i0}\right)\left(\hat{\theta}_i-\theta_{i0}\right)=O_p\left(a_L\right).$$

which implies that $\hat{\theta}_i-\theta_{i0}=O_p\left(a_L\right)$ because $H_i\left(\theta_{i0}\right)$ is non-singular. □

□

Proof of Proposition 6. Let

$$\tilde{g}_i\left(\cdot;\theta\right)=\frac{\sum_{j\in\hat{\Lambda}_i}f\left(\cdot-\epsilon_{ij};\theta\right)}{|\hat{\Lambda}_i|}.$$

Then

$$\left|\frac{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)-\tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(q_{i\ell};\theta\right)}\right|\leq\frac{\frac{1}{|\hat{\Lambda}_i|}\sum_{j\in\hat{\Lambda}_i}\left|f\left(\hat{q}_{i\ell}-\hat{\epsilon}_{ij};\theta\right)-f\left(\hat{q}_{i\ell}-\epsilon_{ij};\theta\right)\right|}{g_i\left(q_{i\ell};\theta\right)}.$$

Use the functional form of the normal density,

$$\left|f\left(\hat{q}_{i\ell}-\hat{\epsilon}_{ij};\theta\right)-f\left(\hat{q}_{i\ell}-\epsilon_{ij};\theta\right)\right| \leq \frac{1}{\sqrt{2\pi}\theta_2}\frac{\left|\hat{q}_{i\ell}-\epsilon_{ij}-\theta_1\right|}{\theta_2^2}\exp\left(-\frac{\left(\hat{q}_{i\ell}-\epsilon_{ij}-\theta_1\right)^2}{2\theta_2^2}\right)\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right|.$$

Moreover,

$$g_i\left(q_{i\ell};\theta\right)=\frac{1}{\sqrt{2\pi}\theta_2}\int_{\underline{\epsilon}_i}^{\bar{\epsilon}_i}\exp\left(-\frac{\left(\hat{q}_{i\ell}-x-\theta_1\right)^2}{2\theta_2^2}\right)dF_{\epsilon_i}\left(x\right)\geq\frac{1}{\underline{p}}\frac{1}{\sqrt{2\pi}\theta_2}\int_{\underline{\epsilon}_i}^{\bar{\epsilon}_i}\exp\left(-\frac{\left(\hat{q}_{i\ell}-x-\theta_1\right)^2}{2\theta_2^2}\right)dx.$$

Therefore,

$$\begin{aligned}\frac{\left|f\left(\hat{q}_{i\ell}-\hat{\epsilon}_{ij};\theta\right)-f\left(\hat{q}_{i\ell}-\epsilon_{ij};\theta\right)\right|}{g_i\left(q_{i\ell};\theta\right)} &\leq C_0\bar{p}\frac{\frac{\left|\hat{q}_{i\ell}-\epsilon_{ij}-\theta_1\right|}{\theta_2^2}\exp\left(-\frac{\left(\hat{q}_{i\ell}-\epsilon_{ij}-\theta_1\right)^2}{2\theta_2^2}\right)}{\int_{\underline{\epsilon}_i}^{\bar{\epsilon}_i}\exp\left(-\frac{\left(\hat{q}_{i\ell}-x-\theta_1\right)^2}{2\theta_2^2}\right)dx}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right| \\ &\leq C\left(1+\left|\hat{q}_{i\ell}\right|\right)^2\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right|.\end{aligned}$$

This holds independent of θ. This implies that

$$\left|\frac{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)-\tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right|\leq C\left(1+\left|\hat{q}_{i\ell}\right|\right)^2\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{j\in\hat{\Lambda}_i}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right|.$$

Notice that

$$\begin{aligned}\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{j\in\hat{\Lambda}_i}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right| &= \frac{1}{\left|\hat{\Lambda}_i\right|}\left[\sum_{\phi_i(y_{ij})\in\mathcal{Q}_i,j\in\left|\hat{\Lambda}_i\right|}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right|+\sum_{\phi_i(y_{ij})\notin\mathcal{Q}_i,j\in\left|\hat{\Lambda}_i\right|}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right|\right] \\ &= O_p\left(\zeta_L\right)+\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{\phi_i(y_{ij})\notin\mathcal{Q}_i,j\in\left|\hat{\Lambda}_i\right|}\left|\hat{\epsilon}_{ij}-\epsilon_{ij}\right| \\ &= O_p\left(\zeta_L\right)+O_p\left(\kappa_L\sqrt{\delta_L}\right).\end{aligned}$$

Therefore,

$$\begin{aligned}\frac{1}{L}\sum\left|\frac{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)-\tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(q_{i\ell};\theta\right)}\right| &\leq \left[O_p\left(\zeta_L\right)+O_p\left(\kappa_L\sqrt{\delta_L}\right)\right]\frac{1}{L}\sum_{\ell=1}^{L}C\left(1+\left|\hat{q}_{i\ell}\right|\right)^2 \\ &= \left[O_p\left(\zeta_L\right)+O_p\left(\kappa_L\sqrt{\delta_L}\right)\right]. \qquad (23)\end{aligned}$$

Notice that this bound is independent of θ.

Next, one can use the maximal inequality to obtain that given $\hat{q}_{i\ell}$

$$\begin{aligned}\mathbb{E}_{\epsilon}\sup_{\theta\in\Theta}\left|\frac{\tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)-g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right| &= \frac{\left|\frac{1}{\left|\hat{\Lambda}_i\right|}\sum_{j\in\hat{\Lambda}_i}\left[f\left(\hat{q}_{i\ell}-\epsilon_j;\theta\right)-\mathbb{E}_{\epsilon}f\left(\hat{q}_{i\ell}-\epsilon_j;\theta\right)\right]\right|}{g_i\left(\hat{q}_{i\ell};\theta\right)}. \\ &\leq \frac{C}{\sqrt{\left|\hat{\Lambda}_i\right|}}\left(1+\left|\hat{q}_{i\ell}\right|\right).\end{aligned}$$

This implies that

$$\sup_{\theta\in\theta}\frac{1}{L}\sum_{\ell=1}^{L}\left|\frac{\tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)-g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right| = O_p\left(\left|\hat{\Lambda}_i\right|^{-1/2}\right) = O_p\left(L^{-1/2}\right), \tag{24}$$

where the last equality holds because $\liminf_{L\to\infty}\left|\hat{\Lambda}_i\right|/L > 0$. Combine (23) and (24) to obtain

$$\sup_{\theta\in\Theta}\frac{1}{L}\sum_{\ell=1}^{L}\left|\frac{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)-g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right| = O_p\left(\zeta_L+\kappa_L\sqrt{\delta_L}+L^{-1/2}\right). \tag{25}$$

Next, notice that

$$\frac{\partial_\theta\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)}{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)}-\frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)} = \frac{v_\theta\left(\hat{q}_{i\ell}\right)-s_\theta\left(\hat{q}_{i\ell}\right)u_\theta\left(\hat{q}_{i\ell}\right)}{1+u_\theta\left(\hat{q}_{i\ell}\right)}$$

where

$$\begin{aligned}u_\theta\left(q\right) &= \frac{\hat{g}_i\left(q;\theta\right)-g_i\left(q;\theta\right)}{g_i\left(q;\theta\right)}, \\ v_\theta\left(q\right) &= \frac{\partial_\theta\hat{g}_i\left(q;\theta\right)-\partial_\theta g_i\left(q;\theta\right)}{g_i\left(q;\theta\right)}, \\ s_\theta\left(q\right) &= \frac{\partial_\theta g_i\left(q;\theta\right)}{g_i\left(q;\theta\right)}.\end{aligned}$$

Notice that by the previous argument,

$$\left|u_\theta\left(\hat{q}_{i\ell}\right)\right| \leq C\left(1+\left|\hat{q}_{i\ell}\right|\right)^2 O_p\left(\zeta_L+\kappa_L\sqrt{\delta_L}\right).$$

This implies that

$$\sup_{\theta\in\Theta,\ell\leq L}\left|u_\theta\left(\hat{q}_{i\ell}\right)\right| = O_p\left(\left[\zeta_L + \kappa_L\sqrt{\delta_L}\right]\log L\right).$$

This inequality holds because by $\hat{q}_{i\ell}$ has a tail of form $e^{-C_q q_{i\ell}^2}$ with $C_q > 0$. Therefore, by the maximal inequality,

$$\sup_{\theta\in\Theta,\ell\leq L}\left|\hat{q}_{i\ell}\right| = O_p\left(\sqrt{\log L}\right).$$

Because $\left(\zeta_L + \kappa_L\sqrt{\delta_L}\right)\log L \to 0$, then

$$\sup_{\theta\in\Theta,\ell\leq L}\left|u_\theta\left(\hat{q}_{i\ell}\right)\right| = o_p\left(1\right).$$

This implies that $1 + u_\theta\left(\hat{q}_{i\ell}\right) > 1/2$ uniformly in $\theta\in\Theta$ and $\hat{q}_{i\ell}$ with probability approaching 1. Then

$$\left\|\frac{\partial_\theta \hat{g}_i\left(\hat{q}_{i\ell};\theta\right)}{\hat{g}_i\left(\hat{q}_{i\ell};\theta\right)} - \frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right\| \leq 2\left\|v_\theta\left(\hat{q}_{i\ell}\right)\right\| + 2\left\|s_\theta\left(\hat{q}_{i\ell}\right)\right\|\left\|u_\theta\left(\hat{q}_{i\ell}\right)\right\|$$

with probability approaching 1. Using an argument similar to the one used to obtain (25), one can show that

$$\sup_{\theta\in\Theta}\left\|\frac{\partial_\theta \hat{g}_i\left(\hat{q}_{i\ell};\theta\right) - \partial_\theta \tilde{g}_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right\| \leq C\left(1 + \left|\hat{q}_{i\ell}\right|\right)^4 O_p\left(\zeta_L + \kappa_L\sqrt{\delta_L}\right),$$

and

$$\mathbb{E}_\epsilon \sup_{\theta\in\Theta}\left\|\frac{\partial_\theta \tilde{g}_i\left(\hat{q}_{i\ell};\theta\right) - \partial_\theta g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right\| \leq \frac{C}{\sqrt{\left|\hat{\Lambda}_i\right|}}\left(1 + \left|\hat{q}_{i\ell}\right|\right)^3.$$

They together implies that

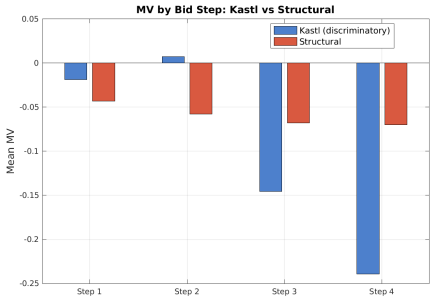
$$\left\|v_\theta\left(\hat{q}_{i\ell}\right)\right\| = O_p\left(\zeta_L + \kappa_L\sqrt{\delta_L} + L^{-1/2}\right). \tag{26}$$

Lastly,

$$\sup_{\theta\in\Theta}\left\|s_\theta\left(\hat{q}_{i\ell}\right)\right\| = \left\|\frac{\partial_\theta g_i\left(\hat{q}_{i\ell};\theta\right)}{g_i\left(\hat{q}_{i\ell};\theta\right)}\right\| \leq C\left(1 + \left\|\hat{q}_{i\ell}\right\|\right)^2.$$

Figure 5: Marginal Values Across Steps

Figure 5 plots average marginal values within each step comparing those of the Kastl model which assumes independent private values and our model.



This implies that $s_\theta(\hat{q}_{i\ell})$ is square integrable and combining with

$$\begin{aligned}
&\mathbb{E}\sup_{\theta\in\Theta}\left\|\frac{1}{L}\sum_{\ell=1}^{L}\left(\frac{\partial_\theta \hat{g}_i(\hat{q}_{i\ell};\theta)}{\hat{g}_i(\hat{q}_{i\ell};\theta)}-\frac{\partial_\theta g_i(\hat{q}_{i\ell};\theta)}{g_i(\hat{q}_{i\ell};\theta)}\right)\right\| \\
&\leq \mathbb{E}\sup_{\theta\in\Theta}\left\|\frac{\partial_\theta \tilde{g}_i(\hat{q}_{i\ell};\theta)-\partial_\theta g_i(\hat{q}_{i\ell};\theta)}{g_i(\hat{q}_{i\ell};\theta)}\right\| \\
&\leq 2\left[\mathbb{E}\sup_{\theta\in\Theta}\|v_\theta(\hat{q}_{i\ell})\|+\sup_{\theta\in\Theta}\sqrt{\mathbb{E}\|s_\theta(\hat{q}_{i\ell})\|^2}\sqrt{\mathbb{E}\|u_\theta(\hat{q}_{i\ell})\|^2}\right] \\
&= O_p\left(\zeta_L+\kappa_L\sqrt{\delta_L}+L^{-1/2}\right).
\end{aligned}$$

This concludes the proof. □

D Furfine

Report % of trade/payments volume captured by Furfine–and number of trades in 1M increments divided by all trades.

E Additional Tables and Figures

F Estimation Details

The estimation combines three raw datasets: auction-level bidding data with bid amounts, yields, allotments, pre-auction and end-of-day balances, the overnight rate, and secondary

market quantities and rates broken out by preferred versus non-preferred counterparty; bilateral interbank loan records at the date–counterparty–rate–amount level; and a pairwise partnership classification indicating whether each lender–receiver pair is preferred. Before anything else, we normalize all prices as spreads over the overnight policy rate, setting $\tilde{y}_{ik} = y_{ik} - r^{ON}$ and applying the same transformation to secondary market lending and borrowing rates. The end of day utility $u(x) = 0.25\,x \cdot \mathbf{1}(x < 0) - 0.25\,x \cdot \mathbf{1}(x > 0)$ is expressed in these units, with corridor bounds ± 0.25: banks pay 25bp above r^{ON} on shortfalls and earn 25bp below r^{ON} on excess reserves.

A central object is the total customer payments $\hat{c}_{it}$, which is backed out from the end-of-day accounting identity:

$$\hat{c}_{it} = L_{it}^{eod} - L_{it}^{pre} - q_{it}^{won} - q_{it}^{recv,p} + q_{it}^{lend,p} - q_{it}^{recv,n} + yes q_{it}^{lend,n}, \tag{27}$$

where L^{eod} and L^{pre} are the end-of-day and pre-auction balances, q^{won} is the bidder's total auction allocation, and the remaining terms collect secondary market receipts and payments with preferred and non-preferred counterparties.

To isolate the anticipated (signal) and unanticipated components of $\hat{c}_{it}$, we estimate the conditional mean $E[\hat{c}_i \mid q_{iK_i}, b_{iK_i}, q_{iK_i-1}]$ by Nadaraya–Watson regression with a product Gaussian kernel. We can then use the squared residuals to estimate a conditional variance $\hat{V}(X_i)$. We then estimate the invertible set $\hat{\Theta}$ as those observations at which $\hat{\Theta} = \{i : \hat{V}(X_i) - \min_j \hat{V}(X_j) \leq \delta_n\}$, with δ_n fixed at the gap between the 2.5th percentile and the minimum of $\hat{V}$. For observations in $\hat{\Theta}$ we set $\hat{s}_i = \hat{m}(X_i)$ and $\hat{\varepsilon}_i = \hat{c}_i - \hat{m}(X_i)$, and use the latter as our empirical draw from the distribution of ε.

The signal decomposition inside $\hat{\Theta}$ pins down the distribution of the unanticipated shock but not the bidder-specific signal distributions. For that, we use the fact that $\hat{c}_{it}$ and the empirical distribution of ε are observed for every bidder-auction pair (not just those in $\hat{\Theta}$). Writing $s_{it} \sim N(\mu_i, \sigma_w^2)$ with a common within-bidder variance, the log-likelihood for bidder i is

$$\ell_i(\mu_i, \sigma_w) = \sum_{t \in T_i} \log\left[\frac{1}{n_\varepsilon}\sum_{e=1}^{n_\varepsilon} \frac{1}{\sigma_w}\phi\left(\frac{\hat{c}_{it} - \hat{\varepsilon}_e - \mu_i}{\sigma_w}\right)\right], \tag{28}$$

where $\{\hat{\varepsilon}_e\}_{e=1}^{n_\varepsilon}$ are the ten decile quantiles of the ε distribution estimated from $\hat{\Theta}$ and T_i is the set of days in which bidder i appears in the data. We find the parameter vector $(\mu_1, \ldots, \mu_N, \sigma_w)$ to maximize the total log-likelihood $\sum_i \ell_i$ jointly under $\sigma_w > 0$.

With shocks and signals in hand, we estimate the value functions for secondary market trading conditional on own and opponent balances. For each observation in the K_i-set we construct the bidder's own post-auction position $X_i = (L_i^{pre} + \hat{c}_i + q_{iK_i})$ together with the summed positions of preferred and non-preferred opponents in the same auction, X_i^p and X_i^n. The outcome for the μ regression is expected utility of the final balance plus the four signed secondary market payment terms, while the outcome for the ξ regression is the vector of four secondary market quantities $(Q^{n,r}, Q^{p,r}, Q^{n,\ell}, Q^{p,\ell})$. Both conditional mean functions are estimated by three-dimensional Nadaraya–Watson with a product Gaussian kernel and per-dimension Silverman bandwidths. Partial derivatives $\hat{\mu}_k$ and $\hat{\xi}_k$ for $k \in \{1, 2, 3\}$ are obtained analytically. From the estimated $\hat{\mu}$ we further isolate the interest-payment component $\hat{\phi}(X, X^p, X^n)$ and the stripped reserve-utility function $\hat{\tilde{\mu}} = \hat{\mu} - \hat{\phi}$; both are used in the welfare decomposition.

Then, the secondary market coefficients $\hat{B}$ are estimated from the auction first-order condition. For each $i \in \hat{\Theta}$ we resample $B = 100$ similar auctions, kernel-weighted by the proximity of i's pre-auction balance to the balances observed in other auctions. In each bootstrap draw we combine i's actual bids with the opponents' bids from the resampled auction, sort by price, and clear via the discriminatory mechanism at the draw's total allotted quantity $Q_{t'_b}$, with proportional rationing at the marginal price. Displacement derivatives $\partial Q_j / \partial q_i$ are computed by finite difference: i's marginal step is perturbed by $\delta q = 0.01 \cdot Q_t$ and the clearing is redone. The clearing price P_b^* defines the events $I_b^< = \mathbf{1}(P_b^* < b_{iK_i})$ and $I_b^= = \mathbf{1}(P_b^* = b_{iK_i})$. Value functions are then evaluated at the post-auction positions, averaging over the ten-point empirical ε grid, so that for each draw $k(j, b)$ indexes whether j is i itself ($k = 1$), a partner ($k = 2$), or a non-partner ($k = 3$). Collecting terms yields, for example,

$$A_1^* = \frac{1}{B} \sum_b I_b^< \sum_j \bar{\mu}_{k(j,b)} \cdot \left. \frac{\partial Q_j}{\partial q_i} \right|_b, \tag{29}$$

with analogous expressions for the ξ averages. The marginal FOC then stacks as

$$\Big[\Pr^<(A_2^* + A_3^*) + \Pr^=(A_5^* + A_6^*) \Big] \cdot B = b_{iK_i} \big(\Pr^< + \Pr^= \cdot \overline{dQ_i^=} \big) - A_1^* \cdot \Pr^< - A_4^* \cdot \Pr^=, \tag{30}$$

giving a $|\hat{\Theta}| \times 4$ linear system once (30) is accumulated across observations; $\hat{B}$ is recovered by OLS.

Finally, for every observation in the full bidder list we recover the signal s_i by minimizing

the squared FOC residual over a grid $\{s_1, \ldots, s_{100}\}$ spanning the 1st–99th percentiles of customer shocks. For each candidate s we recompute the bootstrap FOC, integrating μ and ξ over the combined grid $\hat{q} = s + \varepsilon$ of signal and ε quantiles. Because more than one grid point can achieve the minimum, $\hat{s}_i$ is generally a set rather than a point.

G Decomposition Details

The decomposition asks what the estimated primitives imply about the sources of bidder value. For each recovered signal $\hat{s}_i$, we compute the continuation-value components by averaging over bootstrap draws and the ε grid:

$$\hat{\mu}_i = \frac{1}{B} \sum_b \frac{1}{|\varepsilon|} \sum_e \mu\big(L_i^{pre} + \hat{s}_i + Q_{i,b} + \varepsilon_e,\ (X_i^p)_b,\ (X_i^n)_b\big), \tag{31}$$

$$\hat{\xi}_i = \frac{1}{B} \sum_b \frac{1}{|\varepsilon|} \sum_e \xi\big(\,\cdot\,\big). \tag{32}$$

The total continuation value admits a three-way decomposition,

$$V_i = \underbrace{\hat{\tilde{\mu}}_i}_{\text{reserve utility}} + \underbrace{\hat{\phi}_i}_{\text{interest payments}} + \underbrace{\hat{\xi}_i \hat{B}}_{\text{SM relationship}}, \tag{33}$$

where $\hat{\mu}_i = \hat{\tilde{\mu}}_i + \hat{\phi}_i$: $\hat{\phi}_i$ captures the interest-payment component of secondary market payoffs and $\hat{\tilde{\mu}}_i$ is the stripped reserve-utility component. The reserve-utility term measures the value of i's position through the overnight corridor net of bilateral interest payments; the interest-payments term collects the net bilateral payments at prevailing SM rates; and the SM relationship term reflects the structural rate coefficients $\hat{B}$ weighted by expected quantities. Because $\hat{s}_i$ is set-valued, each component is reported as bounds $[\min, \max]$ over the identified signal set, and within-auction rankings across bidders are computed conservatively using those bounds.

To evaluate the contributes of each of these channels, we evaluate each banks willingness to pay on a quantity grid.From this we construct an aggregate demand curve and calculating counterfactual prices and allocations. The benchmark for this exercise is the marginal values under the status quo (i.e., removing bid shading incentives). As a first step we compare this allocation to those obtained under strategic bidding, and then we compare all other channels

to these baseline set of values. The four channels are:

(a) No SM, shocks unknown. Each bank's bid equals the expected marginal corridor utility:

$$MV_k^{(a)} = \mathbb{E}_\varepsilon[u'(L_i + \varepsilon_i + q_k)]\,.$$

No SM trading occurs; partner positions are set to zero. Welfare is $u(X_{alloc})$.

(b) No SM, shocks known. Same as (a) but the shock ε_i is realized at the time of bidding:

$$MV_k^{(b)} = u'(L_i + \varepsilon_{i,\text{realized}} + q_k)\,.$$

This full-information no-SM benchmark isolates the additional value of shock resolution. Welfare is $u(X_{alloc})$.

(c) Remove interest payments. Interest payments $\hat{\phi}$ are stripped from the value function; the partner structure and rate coefficients $\hat{B}$ are retained:

$$MV_k^{(c)} = \hat{\tilde{\mu}}_1(X_k, X_{pk}, X_{nk}) + \hat{\xi}_1(X_k, X_{pk}, X_{nk})'\hat{B}.$$

Welfare at the resulting allocation is $\hat{\tilde{\mu}}(X_{alloc}, X_p, X_n) + \hat{\xi}(X_{alloc}, X_p, X_n)'\hat{B}$.

(d) Remove interest payments and rate channel. In addition to stripping $\hat{\phi}$, the $\hat{B}$ coefficient is zeroed so the ξ term drops:

$$MV_k^{(d)} = \hat{\tilde{\mu}}_1(X_k, X_{pk}, X_{nk}).$$

Welfare is $\hat{\tilde{\mu}}(X_{alloc}, X_p, X_n)$.

For each channel $c \in \{a, b, c, d\}$, we report the per-bank welfare change $\Delta W_i^{(c)} = W_i^{(c)} - W_i^{base}$ and the quantity change $\Delta q_i^{(c)} = q_i^{(c)} - q_i^{base}$, together with aggregate reallocation volume $\frac{1}{2}\sum_i |\Delta q_i^{(c)}|$. We also report $\Delta q_i^{obs} = q_i^{base} - q_i^{obs}$, the reallocation attributable to bid frictions. We compute this both as a one-step solution (i.e., fixing opponent allocations) and an equilibrium which iteratively updates opponent allocations until a stable set is reached.

We also obtain a decomposition on the subsample of bidders drawn from their marginal bid steps with valid signals. We split the marginal value of a unit of additional auction

quantity into an own-effect

$$V_1 = \hat{\mu}_1(X_i, X_i^p, X_i^n) + \hat{\xi}_1(X_i, X_i^p, X_i^n)'\hat{B}, \tag{34}$$

and a displacement externality

$$V_2 = \sum_{j \neq i} \frac{\partial Q_j}{\partial q_i} \left[\hat{\mu}_{k(j)} + \hat{\xi}'_{k(j)} \hat{B} \right]. \tag{35}$$

For each observation we resample $B = 100$ opposing auction configurations, again kernel-weighted by pre-auction balance proximity. Within each draw we clear the auction, compute displacement derivatives by finite difference at $\delta q = 0.01Q$, read off post-auction positions and (X_i^p, X_i^n), and evaluate both components at $X_i = L_i^{pre} + \hat{s}_i + q_{iK_i} + \varepsilon_e$, averaging over the ε grid. V_1^{unc} averages over all draws; V_1^{alloc} conditions on draws in which i is actually allocated ($P^* \leq b_{iK_i}$). To quantify how marginal value varies with opponents' positions, we additionally evaluate V_1 at the 10th, 50th, and 90th percentiles of the bootstrap distribution of (X_i^p, X_i^n), reporting the opponent spread $V_1(p90) - V_1(p10)$, and decompose that spread by varying X_i^p with X_i^n fixed at its median (and vice versa). Observations whose median X_i across ε draws falls outside the 10th–90th percentile range of the NW training data are excluded, and bidders who are never allocated across any draw are dropped.

H Counterfactual Details

The counterfactual exercise has two conceptually distinct parts. First, we model banks' pre-auction positioning as the outcome of a target-choice problem and estimate the associated marginal cost of holding reserves, which supplies the primitive needed to let targets respond to counterfactual policies. Second, we iterate an auction-equilibrium algorithm in which bids are best responses to simulated opponent behavior and displacement externalities, clearing is done by discriminatory allocation with proportional rationing, and positions are updated to convergence.

We run two versions of all the counterfactuals, a *fixed-distribution* version in which targets are held at their data values and only bids and allocations adjust, and an *equilibrium* version in which targets themselves re-solve each iteration via a constrained Newton step that imposes the aggregate supply constraint. The estimates of marginal cost of holding re-

serves are only required for this second exercise, in order to provide a simple approximation for how pre-auction positions might adjust and how this impacts policy.

H.1 Modeling and Estimating Targets

Each bank i chooses a target pre-auction balance t_i at constant marginal cost α_i and then receives an idiosyncratic shock ν_i, so that its realized pre-auction balance is $L_i = t_i + \nu_i$. The central bank sets $\sum_{i=1}^{N} t_i = T$. To identify α_i we need the value of holding target L, namely $W_i(L) = \mathbb{E}[V_1 \mid \text{target} = L]$, which we estimate on a grid of 30 points spanning $\bar{L}_i \pm 3\hat{\sigma}_{L_i}$. The expectation is taken over the variation a bank faces conditional on having chosen target L: day-to-day gaps between target and realized position, opponent configurations, and unanticipated shocks. Concretely, for each auction day d in the data we define the composite deviation

$$\Delta_{i,d} = (L_{i,d}^{pre} - \bar{L}_i) + \hat{c}_{i,d} + q_{i,d}, \tag{36}$$

so that a bank choosing target L has total position approximately $L + \Delta_{i,d}$ on day d; we draw $D = 100$ such days. Separately, we draw $M = 50$ opponent configurations, each supplying an observed (X_i^p, X_i^n). These draws decouple the two sources of variation and let us evaluate V_1 efficiently on a fine grid.

In practice, we first pre-compute the opponent-averaged marginal value on a grid $\{\tilde{X}_1, \ldots, \tilde{X}_H\}$ of $H = 100$ points:

$$\bar{V}_1(\tilde{X}) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{n_\varepsilon} \sum_{e=1}^{n_\varepsilon} \Big[\hat{\mu}_1\big(\tilde{X} + \varepsilon_e,\, X_{i,m}^p,\, X_{i,m}^n\big) + \hat{\xi}_1(\cdot)' \hat{B} \Big]. \tag{37}$$

Throughout the counterfactual, all positional arguments to $\hat{\mu}$ and $\hat{\xi}$ are clipped to the 5th–95th percentile range of the NW training data and marginal-value outputs are clipped to the corridor $[-0.25, 0.25]$; we suppress this in the notation. We then form $W_i^{raw}(L_g) = D^{-1} \sum_d \bar{V}_1(L_g + \Delta_{i,d})$ by linear interpolation on the grid, and smooth across L with a Silverman-bandwidth Gaussian kernel: $W_i(L_g) = [\sum_{g'} K_{h_L}(L_g - L_{g'}) W_i^{raw}(L_{g'})] / [\sum_{g'} K_{h_L}(L_g - L_{g'})]$. Revealed preference then gives $t_i^* = \bar{L}_i$, the marginal cost is $\hat{\alpha}_i = W_i(\bar{L}_i)$, and the shock distribution is read off directly as $\hat{\nu}_{i,d} = L_{i,d}^{pre} - t_i^*$. The quadratic adjustment cost coefficient $\hat{\kappa}_i$ set to zero for all banks in the data.

H.2 Computing the Counterfactual

With $(\hat{\alpha}_i, \hat{\nu}_{i,d})$ in hand we can run the counterfactual algorithm. A counterfactual specification fixes a supply Q, an active bidder set $\mathcal{A}$, and whether the secondary market operates. Denote iteration-k targets by $\mathbf{t}^{(k)}$, initialized at $\mathbf{t}^{(0)} = (\bar{L}_1, \ldots, \bar{L}_N)$. Each iteration proceeds in four stages. In stage B.0 we draw simulated pre-auction balances as $L_{i,d}^{(k)} = t_i^{(k)} + \hat{\nu}_{i,\sigma(d)}$ for $d = 1, \ldots, n_{sim}$, with signal and ε draws held fixed across iterations.

Stage B.1 constructs bids. Each active bidder i builds step-function bids on a bidder-specific grid $\{q_0, \ldots, q_K\}$ with $q_K = \min(1.5 \max_d q_i^{data}, Q)$. The bid at step q_j for draw d is

$$b_i(q_j, d) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{n_\varepsilon^B} \sum_e \mathrm{FOC}_i\big(L_{i,d}^{(k)} + s_{i,d} + q_j + \varepsilon_e,\ X_{i,m}^p,\ X_{i,m}^n\big), \tag{38}$$

where

$$\mathrm{FOC}_i = \hat{\mu}_1 + \hat{\xi}_1' \hat{B} + \sum_{j \neq i} \frac{\partial Q_j}{\partial q_i} \Big[\hat{\mu}_{k(j)} + \hat{\xi}_{k(j)}' \hat{B}\Big] \tag{39}$$

collects the own-position and displacement contributions, all evaluated at the same opponent positions. On iteration 1, $(X_{i,m}^p, X_{i,m}^n)$ are sampled from observed data positions and the displacement derivatives are set to zero; on iteration $k \geq 2$, the opponent positions are replaced by simulated post-auction positions from the previous iteration's stage B.2, and the displacement derivatives are updated from stage B.3. Both the environment in which V_1 is evaluated and the externality term V_2 therefore evolve as the equilibrium converges.

Stage B.2 clears the auction. For each draw d, the step functions of all active bidders are pooled, sorted by price descending, and cleared at supply Q with proportional rationing at the marginal price. Post-auction positions follow as

$$X_{i,d} = L_{i,d}^{(k)} + s_{i,d} + \varepsilon_{i,d} + q_{i,d}^{won}, \quad X_{i,d}^p = \sum_{j \in P_i} X_{j,d}, \quad X_{i,d}^n = \sum_{j \notin P_i \cup \{i\}} X_{j,d}. \tag{40}$$

Stage B.3 then updates displacement derivatives: for each active bidder i, we perturb the last bid step by $\delta q = 0.01Q$, re-clear, and average the induced changes in j's allocation, $\widehat{\partial Q_j / \partial q_i} = n_{valid}^{-1} \sum_d [Q_j^{pert}(d) - Q_j^{orig}(d)]/\delta q$.

In the *fixed-distribution* version, targets are held at $\mathbf{t}^{(0)}$ and only stages B.1–B.3 iterate, with convergence judged by allocation stability, $\max_i |\bar{Q}_i^{(k+1)} - \bar{Q}_i^{(k)}|$. In the *equilibrium* version, stage C updates the targets by a constrained Newton step. The per-player FOC gap

is

$$g_i = \frac{1}{|D_{valid}|} \sum_{d \in D_{valid}} \left[\hat{\mu}_1(X_{i,d}, X^p_{i,d}, X^n_{i,d}) + \hat{\xi}_1(\cdot)' \hat{B} \right] - \hat{\alpha}_i, \tag{41}$$

and the Jacobian has diagonal $J_{ii} = \mathbb{E}[\hat{\xi}_1' \hat{B}] - \hat{\kappa}_i$ (equal to $\mathbb{E}[\hat{\xi}_1' \hat{B}]$ since $\hat{\kappa}_i = 0$ for all banks) and off-diagonals $J_{ij} = \mathbb{E}[\hat{\xi}_2' \hat{B}]$ if $P_{ij} = 1$ and $J_{ij} = \mathbb{E}[\hat{\xi}_3' \hat{B}]$ otherwise. Imposing the supply constraint $\sum_i t_i = T$ via a Lagrange multiplier gives the bordered system

$$\begin{pmatrix} J & \mathbf{1} \\ \mathbf{1}^\top & 0 \end{pmatrix} \begin{pmatrix} \Delta \mathbf{t} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{g} \\ \sum_i t_i^{(k)} - T \end{pmatrix}, \tag{42}$$

with damped update $\mathbf{t}^{(k+1)} = \mathbf{t}^{(k)} - \gamma \Delta \mathbf{t}$ at $\gamma = 0.15$, followed by additive re-projection onto the constraint: $t_i \leftarrow t_i + (T - \sum_j t_j)/N$.

H.3 Experiments

We run four policy experiments in both the fixed-distribution and equilibrium versions. The *baseline* sets $Q = Q_{median}$ with all bidders active and secondary markets open, and serves as a sanity check that the algorithm replicates the observed environment. *Double supply* and *half supply* take $Q = 2Q_{median}$ and $Q = Q_{median}/2$, testing absorption of excess liquidity and the response to scarcity. The *top-5 only* experiment keeps the five largest bidders by historical allocation share active and makes the rest passive, measuring concentration effects.