

# Dealer Heterogeneity and Exchange Rates\*

Florent Gallien<sup>a</sup>, Sergei Glebkin<sup>b</sup>, Serge Kassibrakis<sup>a</sup>, Semyon Malamud<sup>c,d</sup>  
and Alberto Teguia<sup>e</sup>

<sup>a</sup>Swissquote

<sup>b</sup>INSEAD

<sup>c</sup>Swiss Finance Institute

<sup>d</sup>Ecole Polytechnique Fédérale de Lausanne

<sup>e</sup>Sauder School of Business, University of British Columbia

May 5, 2023

## Abstract

We show, both theoretically and empirically, that several statistics of dealer heterogeneity affect prices and liquidity in the foreign exchange (FX) market. A higher *cross-sectional covariance between dealers' risk aversions and inventories* is associated with higher FX returns. Although unobservable, this statistic can be proxied by the cross-sectional covariance between dealer-to-customer (D2C) prices and bid-ask spreads. A higher *cross-sectional dispersion of dealer risk aversions* is associated with higher liquidity in the dealer-to-dealer market and can be proxied by the cross-sectional dispersion of D2C spreads. These predictions are confirmed empirically using proprietary data on the largest FX dealers' D2C quotes.

**Keywords:** Liquidity, Foreign Exchange, OTC markets, Price Impact, Market Power

**JEL Classification Numbers:** F31, G12, G14, G21

---

\*Florent Gallien is at Swissquote, Sergei Glebkin ([glebkin@insead.edu](mailto:glebkin@insead.edu)) is at INSEAD, Serge Kassibrakis is at Swissquote, Semyon Malamud ([semyon.malamud@epfl.ch](mailto:semyon.malamud@epfl.ch)) is at EPFL and the Swiss Finance Institute, and Alberto Teguia ([alberto.mokakteguia@sauder.ubc.ca](mailto:alberto.mokakteguia@sauder.ubc.ca)) is at UBC Sauder. This paper was previously circulated under the title “Liquidity Provision in the Foreign Exchange Market.” We thank Lukas Altermatt, Alain Chaboud, Thomas Gilbert, Pietro Gottardi, Vrahm Kamara, Alexander Michaelides, Andreas Schrimpf, Vlad Sushko, Adrien Verdelhan, seminar participants at the University of Essex, conference participants at the 3rd Annual CEPR Symposium on Financial Economics, the 34th Annual Pacific Northwest Finance Conference and NES 30th Anniversary Conference for their helpful comments and remarks. Semyon Malamud gratefully acknowledges the financial support of the Swiss Finance Institute and the Swiss National Science Foundation, Grant 100018.192692. Sergei Glebkin, Semyon Malamud, and Alberto Teguia gratefully acknowledge the financial support of The Europlace Institute of Finance.

# 1 Introduction

Motivated by the failure of macroeconomic models to explain exchange rate dynamics, a growing literature emphasizes the role of dealers in price formation in foreign exchange (FX) markets.<sup>1</sup> In particular, it has been shown that given dealers' limited risk-bearing capacity, their aggregate inventory affects exchange rates. The previous literature has chiefly considered the dealer sector as a whole and abstracted away from dealer heterogeneity.<sup>2</sup> This paper shows theoretically and empirically that dealer heterogeneity is an essential determinant of exchange rates. In particular, we show that exchange rates are affected by how the inventory is allocated across different dealers and that the degree of heterogeneity in dealers' risk-bearing capacity affects the liquidity of the FX market.

The microstructure of the foreign exchange market is highly complex. Trading is decentralized and has a pronounced two-tier structure, characterized by two principal segments: the dealer-to-customer (D2C) and the dealer-to-dealer (D2D) segments. For pricing and liquidity in the D2D market, our model emphasizes the role of dealers' heterogeneity along the two dimensions: risk aversions and inventories.<sup>3</sup> However, neither of the two is empirically observable. Our key idea is that prices and bid-ask spreads quoted by dealers in the D2C market are informative about their unobservable characteristics. Our model of the D2C market helps to map the relevant summary statistics of unobservable dealer heterogeneity to the observable statistics of the cross-section of prices and spreads in the D2C market.

We develop a pure inventory-theoretic model that features D2C and D2D market seg-

---

<sup>1</sup>The fact that exchange rates are only weakly related to macroeconomic fundamentals is known as the [Meese and Rogoff \(1983\)](#) exchange rate disconnect puzzle.

<sup>2</sup>We expand on the literature review in Section 7.

<sup>3</sup>We interpret risk aversion broadly as a higher cost of holding more inventories for FX dealers. Heterogeneity in risk-aversions might also capture heterogeneity in customer bases of different dealers (order flow toxicity). The dealers are typically banks, and these costs usually come from banks' capital requirements, costs of obtaining funding, and regulatory constraints that may be binding at the bank level and transmitted into individual trading desk behavior. Moreover, the slackness of such constraints may change over time. Therefore, we see risk aversion as being time-varying. See [Cenedese et al. \(2021\)](#) for empirical evidence supporting this view.

ments. Dealers first trade with customers in the D2C market and then offload excess FX risk exposure in the centralized D2D market. Even though the models of both market segments are standard, jointly studying the D2D and D2C markets is challenging theoretically. We overcome this challenge by deriving the implications highlighting the first-order effects of dealer heterogeneity.<sup>4</sup> The novel empirical implications of our theory come from (i) jointly studying both segments and (ii) accounting for non-exclusive customer-dealer relationships in the D2C market.<sup>5</sup>

We model the D2D market as a standard uniform-price double auction.<sup>6</sup> Importantly, we assume that dealers are: (i) heterogeneous in their inventories and risk aversions and (ii) strategic; that is, they take their price impact into account. We say that a *liquidity mismatch* occurs when dealers with relatively higher risk aversion hold more inventories. Our measure of liquidity mismatch is a cross-sectional covariance between dealers' inventories and risk aversions.

Our D2D market model yields two main results: (a) higher liquidity mismatch is associated with higher FX returns, and (b) higher dispersion in dealers' risk aversions is associated with higher liquidity in the D2D market. Both results are intuitive. To see (a), note that when liquidity mismatch is higher, the inventory allocation to dealers is less efficient. Moreover, such inefficiency cannot be corrected instantly via inter-dealer trading because of price impact. As a result, the dealer sector as a whole is effectively more risk-averse, and the FX returns are higher. To see (b), consider the following example. Imagine that we have three dealers with a risk aversion of 1 each. They each provide 1 unit of

---

<sup>4</sup>That is, we can expand the equilibrium as the equilibrium in the homogeneous case, plus the corrections due to heterogeneity. We focus on the corrections that are first-order in the degree of heterogeneity.

<sup>5</sup>Non-exclusivity means that customers direct their orders to dealers offering the best prices instead of directing them to a preferred, exclusive dealer. As discussed in Section 2.1, such a non-exclusive relationship is realistic for the FX market. Other models of two-tiered markets that jointly study the two segments (most prominently, Babus and Parlato (2018)) feature exclusive customer-dealer relationships and imply the results that are opposite to our main empirical and theoretical results, as we show formally in Section G. See more discussion in Section 4.

<sup>6</sup>As in Kyle (1989), Vives (2011), Rostek and Weretka (2015) and Malamud and Rostek (2017).

liquidity, 3 in total. Now consider what happens if dealers' risk aversions are 0.5, 1, and 1.5 (so that average risk aversion is the same, but the dispersion is higher). Because the liquidity (price elasticity) is inversely proportional to risk aversion, the dealers will provide 2, 1, and  $1/1.5=0.66$  units of liquidity, 3.66 in total.<sup>7</sup>

We next use our D2C market model to help us map unobservable statistics of dealer heterogeneity to observable quantities.<sup>8</sup> First, we show that higher dispersion of dealer risk aversions is associated with higher cross-sectional dispersion of bid-ask spreads in the D2C market. This is intuitive because bid-ask spreads are positively related to risk aversions: More risk-averse dealers are less efficient at holding inventory and require higher compensation, resulting in wider spreads. Our first prediction follows: Bid-ask spreads in the D2D market are negatively related to the cross-sectional dispersion of D2C bid-ask spreads.

Second, we show that higher cross-sectional covariance between inventories and risk aversions (liquidity mismatch) is associated with higher cross-sectional covariance between prices and bid-ask spreads in the D2C market. We call the latter covariance the *price-based liquidity mismatch* measure. The relationship between the two mismatch measures holds because: (i) risk aversions and bid-ask spreads are positively related, and (ii) inventories (with which dealers start the D2D trading round) and D2C prices are positively related. Part (i) is discussed above. To see part (ii), note that dealers posting the highest (lowest) prices will attract a disproportionate share of aggregate customer sell (buy) volume. Thus, dealers with the lowest (highest) prices will decrease (increase) their inventories, implying a positive cross-sectional relationship between inventories and D2C prices. Our second

---

<sup>7</sup>The price elasticity is inversely proportional to risk aversion when heterogeneity in risk aversions is small. In the general case, this relationship is more complex. However, price elasticity is still a convex function of risk aversion in this case. Thus, the logic of our simple example still applies.

<sup>8</sup>We actually have two models of the D2C market. A simple model, presented in Section 2, abstracts from some of the features of the real FX market while delivering the same predictions as our full model. We provide an overview of the FX market in Section 2.1 and present the full model in Section C.1.

prediction follows: Prices in the D2D market are negatively related to the cross-sectional covariance between D2C prices and bid-ask spreads.<sup>9</sup>

To test these two novel predictions, one requires data on the cross-section of D2C quotes.<sup>10</sup> To our knowledge, such data is not publicly available and has never been studied in the literature. We use proprietary high-frequency data on quotes in the FX spot market for the EUR/USD currency pair by the major dealers in the D2C market segment provided by a major Swiss retail aggregator.<sup>11</sup> We merge these data with data on quotes and spreads for the same currency pair from Electronic Broking Services (EBS), one of the largest D2D FX platforms in the world.<sup>12</sup>

We find strong support for our predictions in the data. While we naturally expect that average price levels in D2C and D2D markets should be related, there is no apparent reason as to why the *dispersion* of prices and spreads in the D2C dataset should be related to the price and spread level in the D2D dataset. Yet the data strongly support the two key predictions of our model: First, the D2C price-based liquidity mismatch is non-trivial and exhibits a significant, positive predictive relationship with D2D prices. Second, the dispersion of D2C bid-ask spreads negatively predicts D2D spreads.

We also test and find strong support for our other predictions. First, we find that D2D

---

<sup>9</sup>The models that feature exclusive customer-dealer relationship, such as Babus and Parlato (2018), imply the opposite result.

<sup>10</sup>It is essential for our analysis to observe both bid and ask prices, which is a unique feature of our data. Indeed, to compute the liquidity mismatch at a particular time, one needs to observe mid-quote and bid-ask spread at that time. For example, the data for the corporate bond market is different. The bond market operates via a request-for-quote (RFQ) mechanism. In RFQ, dealers only provide one quote for a particular direction of the trade. In contrast, in the request-for-market (RFM) trading mechanism customer requests *two-sided* quotes from the dealers simultaneously. Such mechanisms are getting traction in the interest rate swaps market, as they allow traders to hide their trading intentions. Our empirical approach thus can be useful in markets with RFM trading mechanisms, such as the interest rate swap market.

<sup>11</sup>We view the retail aggregator as a customer in the framework of our model. Retail aggregators (RAs) are intermediaries between retail clients and large foreign exchange dealers. These dealers compete for RAs' order flow by providing high-frequency D2C quotes (bids and asks). Because we only consider the EUR/USD currency pair, it seems safe to assume that our RA does not have any informational advantage relative to the dealers, who are all large international investment banks. Thus, our assumption of pure inventory-driven trades is well justified.

<sup>12</sup>According to Mancini et al. (2013), EBS is the largest inter-dealer FX trading platform, accounting for 60% of total D2D volume.

prices are negatively related to the customer demand shocks that originate from the retail aggregator (RA) (our data provider). Given the relatively small size of the RA, this finding is surprising. We hypothesize that a common component in global customer order flow is highly correlated with the customer shocks from our RA. Second, dealers posting the highest prices among other dealers will likely be posting the lowest prices subsequently. In our model, this occurs because such dealers will attract a disproportionate share of sell volume and will end up with excess inventory. To decrease the inventory, they would try to attract buy order flow by posting the lowest prices. Such behavior is consistent with dealer ‘*quote shading*’ – shifting prices in the direction opposite to the inventory. While there is some mixed evidence for dealer quote shading in the D2D market, to the best of our knowledge, we are the first to document it in the D2C market.<sup>13</sup> Finally, such reversals in the cross-section of D2C prices imply that cross-sectional covariance between D2C prices and bid-ask spreads should exhibit negative auto-correlation. We find strong support for this prediction as well.

## 2 An Overview of FX Market Structure and a Simple Model

In this section, we present an overview of the FX market structure and a simple model of the FX market. The model here makes several strong assumptions and abstracts from some essential features of the real-world FX market to facilitate exposition. In Section C.1, we present our full model, which accounts for all FX market features. The implications of the model here and that in Section C.1 are essentially the same.

---

<sup>13</sup>While Lyons (1995) shows evidence consistent with quote shading, recent studies, such as Bjørnes and Rime (2005) and Osler et al. (2011), have found no evidence of quote shading in the D2D FX markets.

## 2.1 An Overview of FX Market Structure

The real-world FX markets are fragmented. Neither retail nor institutional traders (e.g., hedge funds, corporates, or smaller regional banks) can trade directly. To trade in the FX market, they are usually subscribed to continuous quotes from a set of major dealer banks (and, potentially, a set of non-bank electronic market-makers) through these dealer banks' single-bank platforms (SBP).<sup>14</sup> Most of the D2C trading happens on such SBPs. Since a customer typically subscribes to multiple SBPs, the dealer banks compete in the D2C markets through quotes.<sup>15</sup>

Thus, the market is naturally fragmented into two segments with quite different natures of competition:

- In the D2C segment, dealer banks compete in prices, whereby prices are quoted on both sides (bid and ask) as a function of order size. In our sample, separate quotes are provided for trades below USD 1 million and trades between USD 1 million and USD 5 million, with spreads naturally increasing in the trade size. These prices are often quoted on dealer-specific SBPs. Customers may split large orders (more than USD 1 million) across several dealers.
- In the D2D market segment, a few dealer banks trade and provide liquidity to each other to offload excess FX risk exposure (for example, inventory that they were not able to internalize through offsetting customer order flow).
- Given the relatively small number of major dealer banks, the competition is imperfect

---

<sup>14</sup>On such an SBP, a dealer bank provides continuous quotes to its customers and usually uses these quotes to manage the direction of customers' trading and to internalize customer order flow. Some smaller regional banks also have SBP, whereby they offer quotes to their customers that aggregate multiple quotes that they receive from major dealers and market-makers. Many of these smaller banks act as retail aggregators, trying to internalize customer order flow and offload excess inventory by trading with major banks at their quoted prices.

<sup>15</sup>Anecdotal evidence suggests that smaller banks usually subscribe for quotes from all major dealers, while a typical hedge fund subscribes to approximately three to five quote streams.

both in the D2C and the D2D segments, and dealers take their market power in both segments into account.

While this market structure is specific to FX markets, many other OTC markets have a similar architecture. For example, in CDS markets, customers would often trade with dealers using a request for quote protocol, whereby they would simultaneously request quotes from multiple dealers (see, e.g., [Collin-Dufresne et al. \(2019\)](#)). One unique feature of the FX market is the availability (for many customers) of both bid and ask quotes by multiple, non-anonymous dealers. This makes it different from (i) limit order markets, where one could see multiple quotes, but there is no way to know whether a given bid and a given ask come from the same dealer; (ii) one-sided RFQ markets, where it is possible to obtain quotes from multiple dealers, but only on one side of the market (either only the bid or only the ask).<sup>16</sup> Furthermore, we are unaware of any dataset containing such two-sided quotes. For our analysis, it is crucial to see the bid-ask spread quoted by every dealer because we use this information to back out important dealer characteristics. This is why the unique nature of our data (see Section 6 below) is ideally suited for our empirical analysis.

## 2.2 A simple model

There are three time periods,  $t = 0, 1, 2$ , and two tradable assets, a risk-free asset with a rate of return normalized to zero and a risky asset with a random payoff  $d$  at time  $t = 2$ . We assume that  $d$  is normally distributed with mean  $\bar{d}$  and variance  $\sigma_d^2$ .

The D2D market operates at  $t = 1$ . The market is populated by  $M > 2$  heterogeneous dealers, indexed by  $l = 1, \dots, M$ , and having linear-quadratic utilities. Dealer  $l$  that begins

---

<sup>16</sup>For example, this market structure is common in bond and CDS markets (see [Collin-Dufresne et al. \(2019\)](#)). While it is formally possible to request both bids and ask quotes from a given group of dealers, anecdotal evidence suggests that it is highly uncommon to do so because dealers in those markets want to know the actual direction of the customer's desired trade before providing the quote.



$t = 1$  with  $\chi_l$  units of risky asset and trades  $Q_l$  units at price  $\mathcal{P}^{D2D}$  derives expected utility

$$U_l^{D2D} = \bar{d}\chi_l + (\bar{d} - \mathcal{P}^{D2D})Q_l - \frac{\Gamma_l}{2}(\chi_l + Q_l)^2.$$

We refer to the coefficient  $\Gamma_l$  as dealer  $l$ 's risk aversion. We assume that the dealer knows his initial inventory  $\chi_l$  but does not know that of other dealers. The D2D market is structured as a uniform-price double auction. Dealer  $l$  submits a (net) demand schedule  $Q_l(\mathcal{P}^{D2D}) : \mathbb{R} \rightarrow \mathbb{R}$ , which specifies demanded quantity of the asset given its price  $\mathcal{P}^{D2D}$  in the interdealer market. The price  $\mathcal{P}^{D2D}$  is such that the market clears, that is,  $\sum_{l=1}^M Q_l(\mathcal{P}^{D2D}) = 0$ . All dealers are strategic, and there are no noise traders. The D2D market is the same as that of the full model, so the equilibrium characterization in that market (see Section 3) applies to both models.

The D2C market operates at  $t = 0$ . We do not model customers explicitly but rather assume that dealers compete for *exogenous* customer order flow  $\tilde{q}$  by providing bid and ask prices  $p_l^b$  and  $p_l^a$ , respectively. We assume that  $\tilde{q}$  is drawn from an arbitrary non-degenerate distribution with finite first two moments and that  $\tilde{q}$  is independent of all other random variables in the model. The distribution of  $\tilde{q}$  is public information. Positive (negative) realizations of  $\tilde{q}$  correspond to customer sell (buy) volume. The prices  $p_l^b$  and  $p_l^a$  are provided to customers and are not observed by competitor dealers. We denote the mid-price by  $\alpha_l \equiv \frac{p_l^b + p_l^a}{2}$  and the bid-ask spread by  $b_l \equiv p_l^a - p_l^b$ . Customers observe the prices of all dealers. We assume that dealers are *myopic*: They put zero weight on the utility derived in the D2D trading round. Dealer  $l$  begins  $t = 0$  with inventory  $x_l$ . We assume that customers are rational in that they route their orders to the dealer providing the best prices (smallest  $p_l^a$  for  $\tilde{q} < 0$  and largest  $p_l^b$  for  $\tilde{q} > 0$ ). Define the *reservation ask (bid) price*  $r_l^a$  ( $r_l^b$ ) as the

smallest ask price (largest bid price) acceptable to a dealer  $l$ . It is straightforward to derive<sup>17</sup>

$$r_l^a = \bar{d} - \Gamma_l \left( x_l + \frac{1}{2} \frac{E[\tilde{q}^2 | \tilde{q} < 0]}{E[\tilde{q} | \tilde{q} < 0]} \right) \text{ and } r_l^b = \bar{d} - \Gamma_l \left( x_l - \frac{1}{2} \frac{E[\tilde{q}^2 | \tilde{q} > 0]}{E[\tilde{q} | \tilde{q} > 0]} \right). \quad (1)$$

We assume that dealer's prices are affine functions of their reservation values, as follows:

$$p_l^a = k_0^a + k \cdot r_l^a \text{ and } p_l^b = k_0^b + k \cdot r_l^b, \quad (2)$$

where  $k$ ,  $k_0^a$  and  $k_0^b$  are constants. We show below that (2) holds exactly when the D2C market is structured as a second-price auction (in which case bidding reservation value is an equilibrium strategy). When the D2C market is structured as a first-price auction, and other dealers' reservation values are i.i.d. uniformly distributed, (2) holds exactly. In the case of a first-price auction and general distribution, (2) holds approximately when the heterogeneity in dealer's risk aversions  $\Gamma_l$  is small, and risk aversions themselves are small. Such approximate relation is sufficient for our main predictions.

**Remark 1 (Simplifying assumptions and the full model)** First, we have assumed myopic dealers. This assumption is made for simplicity. This assumption does not drive our results as our full model does not assume dealers' myopia and delivers the same results. Second, we have assumed that dealers' prices in the D2C market cannot depend on the quantities demanded by customers. In other words, in our simple model, dealers have *no flexibility* in conditioning prices on quantities. As we discuss in Section 2.1, in the real FX market, dealers have *partial flexibility*: they provide separate quotes for orders below \$1m and orders between \$1m and \$5m. In our full model, dealers can post linear price schedules; that is, they have *full flexibility* in conditioning prices on quantities. We believe that not capturing the partial flexibility of real FX markets is not crucial for our results because the

---

<sup>17</sup>Indeed, the  $r_l^a$  solves  $\bar{d}(x_l + E[\tilde{q} | \tilde{q} < 0]) - r_l^a E[\tilde{q} | \tilde{q} < 0] - \frac{\Gamma_l}{2} E[(x_l + \tilde{q})^2 | \tilde{q} < 0] = \bar{d}x_l - \frac{\Gamma_l}{2} x_l^2$ . A similar calculation applies to  $r_l^b$ .

results in the model with no flexibility (simple model) and full flexibility (full model) are the same. Third, we have assumed exogenous, price-inelastic customer demand  $\tilde{q}$ . In our full model, customer demand is endogenous and adjusts to prices posted by dealers. Because the implications in the two models are the same, this simplification of customer demand is not crucial for our results. Finally, unlike in our simple model, customers can split their large orders across multiple dealers in the real world. The absence of such order splitting is not crucial for our results as our full model allows for it and derives the same implications.

### 3 Equilibrium in the D2D Market

This section uses the results in [Malamud and Rostek \(2017\)](#) to characterize the unique, robust linear Nash equilibrium in the D2D double auction game. Our D2D game is the same in both the simple and full models. Thus, the results here (the entire Section 3) apply to both models.

The *bid-ask spread* ( $BA$ ) is our key empirical measure of liquidity in the D2D market. Our theoretical model defines  $BA$  as the difference in equilibrium prices when aggregate dealer demand increases by one unit versus when it decreases by one unit. Since our model will feature a linear equilibrium, the bid-ask spread equals two divided by the price sensitivity of aggregate dealers' demand.

Given his asset holdings  $\chi_l$ , dealer  $l$ 's objective is to choose the trade size  $Q_l$  in the D2D market that maximizes his utility by choosing the optimal trade size  $Q_l$ . The equilibrium demand schedule  $Q_l = Q_l(\chi_l, \mathcal{P}^{D2D})$  of dealer  $l$  equalizes his marginal utility with his marginal payment for each price,

$$d - \Gamma_l(\chi_l + Q_l) = \mathcal{P}^{D2D} + \beta_l Q_l,$$

where  $\beta_l$  measures the *price impact of dealer  $l$  in the D2D market* ("Kyle's lambda"; see

Kyle (1985)). Formally,  $\beta_l$  is the derivative of the inverse residual supply of dealer  $l$ , which is defined by aggregation through the market clearing of the schedules submitted by other traders,  $\{Q_\ell(\chi_\ell, \mathcal{P}^{D2D})\}_{\ell \neq l}$ . It follows that the price elasticity of dealer  $l$ 's demand is  $1/(\beta_l + \Gamma_l)$  and the bid-ask spread  $BA$  is given by

$$BA = 2 / \sum_l \frac{1}{\beta_l + \Gamma_l}.$$

To pin down equilibrium price impacts, we note that the market clearing condition requires that the price impact assumed by dealer  $l$  be equal to the slope of his inverse residual supply, which results from the aggregation of the other dealers' submitted schedules. Proposition 2 (Proposition 1 in Malamud and Rostek (2017)) shows that the system for equilibrium price impacts can be solved explicitly.<sup>18</sup> Additionally, the proposition below derives some comparative statics.

**Proposition 2** *There exists a unique D2D market equilibrium. The equilibrium price is given by*

$$\mathcal{P}^{D2D} = \bar{d} - \mathbf{Q}^* \text{ with } \mathbf{Q}^* \equiv \mathcal{B}^{-1} \sum_{l=1}^M (\Gamma_l + \beta_l)^{-1} \Gamma_l \chi_l. \quad (3)$$

Trader  $l$ 's price impact  $\beta_l$  is given by

$$\beta_l = \frac{2\Gamma_l}{\Gamma_l \mathcal{B} - 2 + \sqrt{(\Gamma_l \mathcal{B})^2 + 4}},$$

where  $\mathcal{B} \in \mathbb{R}_+$  is the unique positive solution to  $\sum_l (\Gamma_l \mathcal{B} - 2 + \sqrt{(\Gamma_l \mathcal{B})^2 + 4})^{-1} = 1/2$ .

The bid-ask spread is given by  $BA = 2/\mathcal{B}$ . Moreover: (i) the price impact  $\beta_l$  is cross-

---

<sup>18</sup>For symmetric risk aversions, the equilibrium of Proposition 2 coincides with the equilibrium in Rostek and Weretka (2011), which in turn coincides with Kyle (1989), without nonstrategic traders and assuming independent values). The case of symmetric risk aversions has also been studied in Vayanos (1999), and Vives (2011).

sectionally monotone decreasing in  $\Gamma_l$ , i.e., if  $\Gamma_{l_1} > \Gamma_{l_2}$ , then  $\beta_{l_1} < \beta_{l_2}$ ; and (ii)  $\beta_l$  is monotone increasing  $\Gamma_\ell$  for any  $\ell \neq l$ : That is, an increase in risk aversion of any trader worsens liquidity for all other traders. The equilibrium post-D2D trade inventory of dealer  $l$  is

$$\tilde{\chi}_l = (\Gamma_l + \beta_l)^{-1} \mathbf{Q}^* + (\Gamma_l + \beta_l)^{-1} \beta_l \chi_l. \quad (4)$$

Dealers' equilibrium indirect utility (16) is given by

$$\mathcal{U}_l = E [\chi_l d - 0.5 \Gamma_l \chi_l^2 + (0.5 \Gamma_l + \beta_l)(\Gamma_l + \beta_l)^{-2} (\mathbf{Q}^* - \Gamma_l \chi_l)^2]. \quad (5)$$

It is well understood that the aggregate order flow from the D2C market affects D2D pricing (see, e.g., [Evans and Lyons \(2002b\)](#)). In other words, previous research has highlighted the importance of the first cross-sectional moment of inventories for exchange rates. The major consequence of the proposition above is that the second cross-sectional moment matter. We focus on two of them: (i) dispersion of inventories and (ii) cross-sectional covariance between inventories and dealers' risk aversions. In the following section, we derive the asset pricing implications of changes in (i) and (ii) and demonstrate how one can diagnose such changes by using D2C market prices.

### 3.1 The Liquidity Mismatch

We introduce some notation. For vectors  $\{x_l\}_l$  and  $\{y_l\}_l$ , we use  $E[x_l] = \frac{1}{M} \sum_l x_l$  and  $E[y_l] = \frac{1}{M} \sum_l y_l$  to denote their cross-sectional means, and  $\text{Cov}(x_l, y_l) = E[(x_l - E[x_l])(y_l - E[y_l])]$  to denote their cross-sectional covariance. We say that  $\{\hat{x}_l\}_l$  is a *mean-preserving spread* of  $\{x_l\}_l$  if  $\hat{x}_l = x_l + \epsilon_l$  and  $E[\epsilon_l] = 0$ . We use the shortcut  $x$  for a vector  $\{x_l\}_l$  and we denote  $\|x\| = (\sum_l x_l^2)^{1/2}$  the Euclidean norm of  $x$ . We also write  $a \stackrel{s}{=} b$  whenever  $\text{sign}(a) = \text{sign}(b)$ .

We say that a *liquidity mismatch* occurs when high-risk aversion dealers have large

inventories after the D2C trading round. Our first, inventory-based measure of liquidity mismatch  $Y_{mismatch}$ , mirrors this definition:

$$Y_{mismatch} \equiv \text{Cov}(\Gamma_l, \chi_l).$$

Note that neither risk aversions  $\Gamma_l$  nor post-trade inventories  $\chi_l$  are empirically observable. Thus, we introduce an alternative, price-based measure  $\mathcal{A}_{mismatch}$ , as follows:

$$\mathcal{A}_{mismatch} \equiv \text{Cov}(b_l, \alpha_l).$$

We show below that  $\mathcal{A}_{mismatch} \stackrel{s}{=} Y_{mismatch}$ , that is, when there is a positive (negative) liquidity mismatch, both measures are positive (negative). Thus,  $\mathcal{A}_{mismatch}$  is a valid, empirically observable alternative for  $Y_{mismatch}$ .

### 3.2 D2D Pricing and Dealer Characteristics

When there is a liquidity mismatch, the allocation of inventories across dealers is inefficient. Due to market power, such inefficiency will not be resolved after the D2D trade. As a result, dealers (on average) will require higher compensation for holding the inventories, and the prices will be lower. Consistent with this intuition, we show below that positive (negative) mismatch  $Y_{mismatch}$  introduces a downward (upward) distortion in the equilibrium D2D price.

Indeed, one can rewrite (3) as follows:

$$\mathcal{P}^{D2D} = \bar{d} - \mathcal{B}^{-1}M \left( E[(\Gamma_l + \beta_l)^{-1}\Gamma_l]E[\chi_l] + \text{Cov}((\Gamma_l + \beta_l)^{-1}\Gamma_l, \chi_l) \right). \quad (6)$$

The key term in (6) is  $\text{Cov}((\Gamma_l + \beta_l)^{-1}\Gamma_l, \chi_l)$ . We show below that  $\text{Cov}((\Gamma_l + \beta_l)^{-1}\Gamma_l, \chi_l) \stackrel{s}{=} Y_{mismatch}$ , and so, indeed, positive (negative) mismatch  $Y_{mismatch}$  introduces a downward (upward) distortion in the equilibrium D2D price.

To see that  $\text{Cov}((\Gamma_l + \beta_l)^{-1}\Gamma_l, \chi_l) \stackrel{s}{=} Y_{mismatch}$ , note that by Proposition 2, *less risk averse dealers have greater price impact*: If  $\Gamma_1 < \dots < \Gamma_M$ , then dealers' price impacts in the D2D market satisfy  $\beta_1 > \dots > \beta_M$ . This is intuitive: Less risk-averse dealers face a more risk-averse rest of the market and, therefore, a less elastic residual supply. By direct calculation, the weights  $(\Gamma_l + \beta_l)^{-1}\Gamma_l$  in Equation (3) are monotone increasing in  $\Gamma_l$ . Therefore, we can write

$$\text{Cov}((\Gamma_l + \beta_l)^{-1}\Gamma_l, \chi_l) \stackrel{s}{=} \text{Cov}(\Gamma_l, \chi_l) = Y_{mismatch}.$$

The following proposition follows immediately.

**Proposition 3** *Fix  $\{\Gamma_l\}_l$ . Consider three possibilities for distribution of inventories  $\{\chi_l\}_l$  across dealers:  $\{\chi_l\}_l$  is such that (a)  $\text{Cov}(\chi_l, \Gamma_l) > 0$ , (b)  $\text{Cov}(\chi_l, \Gamma_l) = 0$ , and (c)  $\text{Cov}(\chi_l, \Gamma_l) < 0$ . Suppose that  $E[\chi_l]$  is the same in all three cases. Then,  $\mathcal{P}_{(a)}^{D2D} < \mathcal{P}_{(b)}^{D2D} < \mathcal{P}_{(c)}^{D2D}$ .*

We also derive implications of changes in the dispersion of risk aversions.

**Proposition 4** *An increase in the dispersion of risk aversions  $\Gamma_l$  (defined as a mean-preserving spread) leads to a decrease in bid-ask spread  $BA$  in D2D market.*

To understand the intuition behind the proposition above, consider the following example. Imagine we have three dealers with a risk aversion of 1 each. They each provide 1 unit of liquidity, 3 in total. Now consider what happens if dealers' average risk aversion is 0.5, 1, and 1.5 (so that average risk aversion is the same, but we increased the dispersion). Since liquidity provided (price elasticity) is inversely related to risk aversion, the dealers will provide 2, 1, and  $1/1.5=0.66$  units of liquidity, 3.66 in total.<sup>19</sup> Thus, greater dispersion in risk aversions results in more liquidity.

Propositions 3 and 4 establish the link between prices and liquidity in the D2D market

---

<sup>19</sup>The price elasticity is inversely proportional to risk aversion when heterogeneity in risk aversions is small. In the general case, this relationship is more complex. However, price elasticity is still a convex function of risk aversion in the general case. Thus the logic of our simple example still applies.

and some statistics of unobservable dealer characteristics. The next section demonstrates how to identify these statistics from prices and spreads in the D2C market.

## 4 Linking Dealer Characteristics to D2C Prices and Spreads

Because neither risk aversions  $\Gamma_l$  nor inventories  $\chi_l$  are observable, we cannot directly take the predictions of Propositions 3 and 4 to the data. The key idea of this paper is that mid-prices  $\alpha_l$  and bid-ask spreads  $b_l$  in the D2C market are informative about unobservable dealer characteristics  $\Gamma_l$  and  $\chi_l$ . Here, we use our simple model to map  $Y_{mismatch}$  and cross-sectional dispersion of risk aversions to observable statistics of prices and spreads in the D2C market. Section C.2 shows that the same mapping is applicable in the full model.

First, we note that D2C bid-ask spreads are informative about dealers' risk aversions. Indeed, it follows directly from (1) and (2) that

$$b_l = k_0^a - k_0^b + \frac{1}{2}k\Gamma_l \left( \frac{E[\tilde{q}^2|\tilde{q} > 0]}{E[\tilde{q}|\tilde{q} > 0]} - \frac{E[\tilde{q}^2|\tilde{q} < 0]}{E[\tilde{q}|\tilde{q} < 0]} \right).$$

Thus, changes in  $\Gamma_l$  are proportional to changes in  $b_l$ . This is intuitive: More risk-averse dealers are less efficient at holding inventory and require higher compensation for doing so, resulting in wider spreads. Directly from Proposition 4, we obtain the following.

**Proposition 5** *An increase in the dispersion of D2C bid-ask spreads  $b_l$  (defined as a mean-preserving spread) is associated with a decrease in bid-ask spread  $BA$  in D2D market.*

We now show that  $Y_{mismatch} \stackrel{s}{=} \mathcal{A}_{mismatch}$ . The equilibrium relationship between prices in the D2C market and dealers' inventories and risk aversions is generally complex. However, it is possible to derive analytical approximations when dealer heterogeneity is small. Such an approximation allows us to capture the first-order effects of heterogeneity on equilibrium



quantities while preserving analytical tractability. The case of small heterogeneity corresponds to small values of  $\|\Gamma - \Gamma^*\|$ , that is, when risk aversions  $\Gamma_l$  are close to some average level  $\Gamma^*$ . Under such an approximation, we can write

$$p_l^a \approx \text{const}^a - k\Gamma^*x_l, p_l^b \approx \text{const}^b - k\Gamma^*x_l, \text{ and } \alpha_l \approx \text{const} - k\Gamma^*x_l, \quad (7)$$

where  $\text{const}^a$ ,  $\text{const}^b$ , and  $\text{const}$  are some constants that are the same across dealers and  $\approx$  denotes approximate equality, up to terms of order  $O(\|\Gamma - \Gamma^*\|)$ .

It follows from (7) that the dealer posting the lowest ask (highest bid) price is also the dealer with the lowest (highest) mid-price. Because the customer order flow is directed to the dealer with the best price, the dealer with the smallest (highest) mid-price will decrease (increase) his inventory, while other dealers' inventories will be unchanged. Thus, we get a positive cross-sectional relationship between changes in inventories  $\chi_l - x_l$  and mid-prices  $\alpha_l$ . If total customer orderflow  $\tilde{q}$  is large compared to dealers' initial inventories  $x_l$ , dealers' final inventory  $\chi_l$  and change in inventories  $\chi_l - x_l$  have the same sign. Thus, we have a positive cross-sectional relationship between  $\chi_l$  and  $\alpha_l$ . Then we have  $\text{Cov}(\chi_l, \Gamma_l) \stackrel{s}{=} \text{Cov}(\alpha_l, \Gamma_l) \stackrel{s}{=} \text{Cov}(\alpha_l, b_l)$ , where the last equality follows because  $b_l$  is related positively to  $\Gamma_l$ . This results in  $Y_{\text{mismatch}} \stackrel{s}{=} \mathcal{A}_{\text{mismatch}}$ , and the following proposition follows directly from Proposition 3.

**Proposition 6** *Suppose that  $\|\Gamma - \Gamma^*\|^2$  is sufficiently small and that the realization  $|\tilde{q}|$  is sufficiently large. Then  $Y_{\text{mismatch}} \stackrel{s}{=} \mathcal{A}_{\text{mismatch}}$ . Holding  $\{\Gamma_l\}_l$  fixed, consider three possibilities for the joint distribution of mid-prices and bid-ask spreads in the D2C market: (a)  $\mathcal{A}_{\text{mismatch}} > 0$ , (b)  $\mathcal{A}_{\text{mismatch}} > 0 = 0$ , and (c)  $\mathcal{A}_{\text{mismatch}} < 0$ . Suppose that in all three cases,  $E[\alpha_l]$  is the same. Then,  $\mathcal{P}_{(a)}^{D2D} < \mathcal{P}_{(b)}^{D2D} < \mathcal{P}_{(c)}^{D2D}$ .*

We now provide several microfoundations for the relation (1).

**Proposition 7** *Suppose that the D2C market is structured as a second-price auction. Then (1) holds with  $k_0^a = k_0^b = 0$  and  $k = 1$ . Suppose that the D2C market is structured as a first-*

price auction. Suppose that from the perspective of any dealer  $l$ , other dealers' reservation values  $r_k^a, r_k^b, k \neq l$  are i.i.d uniformly distributed, with  $r_k^a \sim U[\underline{r}^a, \bar{r}_a]$  and  $r_k^b \sim U[\underline{r}^b, \bar{r}_b]$ . Then (1) holds with  $k_0^a = \bar{r}_a/M$ ,  $k_0^b = \underline{r}^b/M$ , and  $k = \frac{M-1}{M}$ . Suppose that the D2C market is structured as a first-price auction. Suppose that  $\Gamma_k = \Gamma^* + \hat{\Gamma}_k$ , where  $\hat{\Gamma}_k$  are i.i.d. distributed with a continuously differentiable CDF on the support  $\mathcal{S}$  such that  $\mathcal{S} \subset [-\Gamma^*, \Gamma^*]$ . Suppose that initial inventories  $x_k$  are i.i.d. distributed with a continuous CDF on a bounded support. Then (1) holds approximately up to terms of order  $O((\Gamma^*)^2)$  and the statements of Propositions 5 and 6 hold for small enough  $\Gamma^*$ .

We conclude with several remarks.

**Remark 8** Our central result is that when (i) dealers have market power and (ii) are heterogeneous in their risk aversion, the price in the D2D market is affected by the *distribution* of inventory risk (i.e., post-D2C inventories  $\{\chi_l\}_l$ ) across dealers. In contrast, if the D2D market was perfectly competitive or if dealers were homogeneous, only the aggregate inventory  $\sum_l \chi_l$  matters for D2D pricing. Indeed, in the homogeneous case, this follows because  $\mathbf{Q}^*$  is proportional to  $\sum_l \chi_l$  (as follows from (3)). In the competitive case, all dealers can be aggregated and substituted with a representative dealer holding the aggregate inventory; hence, the prices are affected only by total supply and not by the inventory distribution across dealers.<sup>20</sup>

**Remark 9** The key to a negative relationship between  $\mathcal{P}^{D2D}$  and  $\mathcal{A}_{mismatch}$  in our model (consistent with our empirical evidence), is a *non-exclusive* relationship between dealers and customers. In contrast, the models in which the relationship between customers and dealers is exclusive (see, e.g., Babus and Kondor (2018) and Babus and Parlatore (2018)) will predict the opposite.<sup>21</sup> To see why, suppose that in our simple model each dealer has its customer base, so each dealer is getting exogenous order flow  $\tilde{q}/M$ . In such a model,

<sup>20</sup>See, e.g. Proposition IA.3 in Glebkin, Malamud and Tegui (2022) for a formal treatment.

<sup>21</sup>We show so formally in Appendix G, where we consider a model of Babus and Parlatore (2018).

each dealer's inventory is expected to change by the same amount, and so  $\text{Cov}(\chi_l, \Gamma_l) = \text{Cov}(x_l + \text{const}, \Gamma_l) = \text{Cov}(x_l, \Gamma_l)$ . Since  $x_l$  is *negatively* related to  $\alpha_l$  (according to (7), dealers with higher inventories bear more risk and require higher compensation for doing so), we have  $Y_{\text{mismatch}} = \text{Cov}(\chi_l, \Gamma_l) = \text{Cov}(x_l, \Gamma_l) \stackrel{s}{=} -\text{Cov}(\alpha_l, b_l) = -\mathcal{A}_{\text{mismatch}}$ . Here, we substituted  $b_l$  for  $\Gamma_l$ , similarly to the derivations underlying Proposition 5. Thus, with an exclusive customer-dealer relationship,  $\mathcal{P}^{D2D}$  is positively related to  $\mathcal{A}_{\text{mismatch}}$ .<sup>22</sup>

## 5 Empirical Predictions

The theoretical analysis in both our simple and full models yields the following empirical predictions.

**Prediction 1.**  $\mathcal{P}^{D2D}$  is negatively related to  $\mathcal{A}_{\text{mismatch}}$ .

This prediction follows from Proposition 6 (simple model) and Proposition 15 (full model).

**Prediction 2.** Bid-ask spreads in the D2D market are negatively related to the dispersion of D2C bid-ask spreads.

This prediction follows from Proposition 5 (simple model) and Proposition 16 (full model).

**Prediction 3.**  $\mathcal{P}^{D2D}$  is negatively related to customer demand shock.<sup>23</sup>

This prediction follows from equation (6), with higher customer demand shock associated with higher  $E[\chi_l]$ . While Prediction 3 is not linked to heterogeneity and would arise in most fragmented market models, it is still instructive to test this prediction with our data to see whether the particular market segment we look at is representative of the global order flow.

---

<sup>22</sup>The derivations here are particularly simple because we assumed price-inelastic customer demand. In Appendix G, we do not make such an assumption and show that our conclusions still hold. Intuitively, customers and dealers trade toward an efficient allocation, where their risks are perfectly shared, but they only do so imperfectly due to dealers' market power. Thus, dealers who start with relatively high pre-D2C inventory  $x_l$  will end up with relatively high post-D2C inventories  $\tilde{x}_l$ .

<sup>23</sup>Demand shock stands for  $\tilde{q}$  in the simple model and  $\Theta$  in the full model.

We derive two additional predictions that stem from the non-exclusive relationship between customers and dealers, the key feature of our model. Non-exclusivity implies that dealers posting the lowest (highest) prices attract a disproportionate share of customer buy (sell) volume. Thus, these dealers will end up with the lowest (highest) inventories  $\chi_l$  post-D2C. Now note that dealers posting the lowest (highest) prices are also dealers with the highest (lowest) inventories  $x_l$  (see (7)). Then with the lowest pre-D2C inventory  $x_l$  will end up with the highest inventory post-D2C,  $\chi_l$ . Because inventories and prices are negatively related (see (7)), dealers posting lowest prices  $\alpha_l$  are likely to post the highest prices in the next period.<sup>24</sup> To test this, we define the price competitiveness measure  $pc_l$  for a dealer  $l$  to be equal to 1 (resp.  $-1$ ) if he posts the highest (resp. lowest) price, and 0 otherwise. Formally,

$$pc_l = \begin{cases} 1, & \alpha_l = \max_k \alpha_k, \\ -1, & \alpha_l = \min_k \alpha_k, \text{ and} \\ 0, & \text{otherwise.} \end{cases}$$

Our next prediction follows.<sup>25</sup>

**Prediction 4.** There is a negative auto-correlation in price competitiveness for any dealer  $l$ .

Holding  $\Gamma_l$  fixed, such reversals in the cross-section of  $\alpha_l$  imply (for large enough total customer order flow) the negative auto-correlation in  $\mathcal{A}_{mismatch}$ . Our final prediction follows.

**Prediction 5.** There is a negative auto-correlation in  $\mathcal{A}_{mismatch}$ .

---

<sup>24</sup>To go from inventories to prices, we need to define mid-prices in D2C in the next trading period. We define them as the D2C mid-prices when dealers' initial inventories  $x_l$  equal their post-D2D inventories  $\tilde{\chi}_l$ . We know that a dealer with the lowest  $\alpha_l$  will end up with the smallest post-D2C inventory  $\chi_l$ . Then, it follows from (4) that  $\chi_l$  and  $\tilde{\chi}_l$ , for small  $\|\Gamma - \Gamma^*\|$ . The statement then follows.

<sup>25</sup>The formal result underlying Predictions 4 and 5 is the Proposition 13 stated and proved in the Appendix.

## 6 Empirical Analysis

In this section, we empirically test Predictions 1 through 5.

### 6.1 Empirical Specifications

To test the basic predictions of our model, we assume that over a sufficiently short period (such as a few seconds), no new fundamental information arrives; thus, prices in both the D2C and D2D markets are driven exclusively by inventory shocks. We assume that all other model parameters, including dealers' risk aversions  $\Gamma_l$ ,  $l = 1, \dots, M$  stay constant over that time horizon.<sup>26</sup> Under this assumption, our model predicts that the following time series relationship between D2C and D2D prices should hold over short time horizons:

$$\mathcal{P}_{t+1}^{D2D} = a_0 + a_1 \bar{\alpha}_t + a_2 \mathcal{A}_{mismatch,t} + a_3 \Theta_t.$$

Here  $\Theta_t$  denotes the customer demand shock. Furthermore, Predictions 1 and 3 imply  $a_2, a_3 < 0$ . It is also natural that mid-prices in D2D and D2C markets are positively related, thus,  $a_1 > 0$ . To summarize, we expect the following signs for our coefficients:

$$a_1 > 0 > a_2, a_3.$$

We test this relationship by regressing D2D prices on both  $\bar{\alpha}$  and  $\alpha_{mismatch}$  and on a set of controls. Following the extant literature, we replace dependent and independent variables with their first differences in our main empirical specification because price levels are highly

---

<sup>26</sup>We do not micro-found the origins of the shocks to dealers' risk aversions  $\Gamma$ . These shocks typically come from banks' capital requirements, shocks to dealers' costs of funding, and regulatory constraints that may be binding at the bank level and transmitted into individual trading desk behavior at the bank.

persistent and replace prices by log prices.<sup>27</sup> Thus, our main empirical specification is

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t, \quad (8)$$

where for a variable  $X_t$  we denote

$$\Delta X_{t, \ell} = X_t - X_{t-\ell} \quad \text{and} \quad p_t^{D2D} = \log \mathcal{P}_t^{D2D}.$$

We consider several lags  $\ell$  ranging from 1 to 60 seconds. We include four controls in our analysis. The first is the liquidity demand of customers, *ProxyLiqShock*, which is a proxy for effects of  $\Theta$  and  $\bar{\theta}$  on prices. The other three are measures of the realized order flow in the D2D market. We include these variables following [Evans and Lyons \(2002b\)](#), who showed that D2D order flow is correlated to price changes in the D2D market.

A second regression specification is designed to test Prediction 2 of our model: The dispersion of spreads in the D2C market negatively predicts the spread in the D2D market. To this end, we run the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 \overline{Spread\_D2C}_t + a_2 STD\_Spread\_D2C_t + controls_t, \quad (9)$$

where  $\overline{Spread\_D2C}$  is the mean spread in the D2C market and  $STD\_Spread\_D2C$  is the standard deviation of spreads in the D2C market. Similar to the specification for the D2D prices, the specification for D2D spreads also contains control variables based on order flows. Prediction 2 implies that  $a_2 < 0$ . Also, it is natural that bid-ask spreads in D2D and D2C markets are positively related. Thus, in this regression, we expect that

$$a_1 > 0 > a_2.$$

---

<sup>27</sup>See, for example, [Evans and Lyons \(2002b\)](#). The results are qualitatively similar when we use prices instead.

## 6.2 Data Description

We test the model’s predictions using three high-frequency datasets (time-stamped with 1-second precision). The first two datasets are provided to us by a large Swiss retail aggregator (the retail aggregator hereafter). The first dataset contains price schedules for the currency pair “EUR/USD”<sup>28</sup> submitted to the retail aggregator by the largest FOREX dealers. This dataset covers the period from June 1, 2016, to September 30, 2016. Price schedules are stepwise functions mapping volumes to bid-ask quotes: These are discretized versions of the price schedules  $p_t(q)$  in our full model (see [subsection C.1](#)). The retail aggregator does not have access to the D2D market and can manage its inventory only by trading with the dealers. Thus, from the point of view of our model, the retail aggregator is a customer trading with dealers in the D2C market.

The second dataset contains the orders submitted for execution by its clients for the same currency pair and period. The geographical composition of the retail aggregator’s clients who trade the EUR/USD currency pair is highly heterogeneous and is dominated by large groups of clients from Italy, Switzerland, China, and Spain. 95% of clients in this dataset are retail clients, while the rest are corporate clients. For 99% of the observations, the order size does not exceed USD 1 mln. Thus, for our empirical analysis, we filter the data to keep the quotes and the orders with an order size not exceeding 1 million. The negative of the aggregate demand of all the retail aggregator clients represents the inventory shock  $\Theta_t$ . The third dataset is the EBS dataset provided to us by NEX Data. EBS is one of the largest D2D FX platforms in the world. This dataset is a comprehensive account of FX’s best bids and asks aggregated within each second. It also indicates when a transaction occurs and its price and size. [Mancini et al. \(2013\)](#) compared the EBS dataset with other FX datasets and concluded that “EBS is effectively the only current data source for intraday data.” For the sake of consistency with the second dataset obtained from the retail aggregator, we restrain

---

<sup>28</sup>According to the Triennial Central Bank Survey (April 2016), this currency pair accounts for the most significant turnover on the FOREX market.

our final dataset to the subset of transactions with an order size of 1 mln or less.<sup>29</sup> All three of our datasets are time-stamped with 1-second precision.

We follow [Mancini et al. \(2013\)](#) to clean our two (high-frequency) quote datasets. In addition, for the retail aggregator datasets, we filter the data sample on quotes to retain price schedules from a subsample of dealers simultaneously providing quotes during a long continuous time interval within each trading day. We also use quotes associated with a volume of 1 million, the dominant volume submitted by dealers to the retail aggregator. After applying such a filtering procedure, we retain quotes from 10 large dealers. Given the trade size  $q = 1$  million, we observe the bid  $p_{l,t}(-q)$  and the ask  $p_{l,t}(q)$  for each dealer at each time instant  $t$ , which we then use to directly back out the mid-prices and slopes, as follows:

$$\alpha_{l,t} = \frac{p_{l,t}(q) + p_{l,t}(-q)}{2} \quad \text{and} \quad b_{l,t} = \frac{p_{l,t}(q) - p_{l,t}(-q)}{2q}.$$

The relevant aggregate quantities in the D2C market are as follows. The average mid-price is

$$\bar{\alpha}_t = \frac{1}{M} \sum_{l=1}^M \alpha_{l,t},$$

the average spread is

$$\overline{Spread\_D2C}_t = \frac{1}{M} \sum_{l=1}^M b_{l,t},$$

and the sample dispersion of spreads is

$$STD\_Spread\_D2C_t = \sqrt{\frac{1}{M-1} \sum (b_{l,t} - \overline{Spread\_D2C}_t)^2}$$

---

<sup>29</sup>Transactions of size one mln or less account for more than 50% of deals in the EBS data. We ran our analysis with all transaction data in an earlier paper version and obtained the same qualitative results.



We then construct our measure of liquidity mismatch, as follows:

$$\mathcal{A}_{mismatch,t} = \text{Cov}(\alpha_{l,t}, b_{l,t}) = \frac{1}{M} \sum_l \left( \alpha_{l,t} - \frac{1}{M} \sum \alpha_{l,t} \right) \left( b_{l,t} - \frac{1}{M} \sum b_{l,t} \right). \quad (10)$$

Formula (10) corresponds precisely to our theoretical liquidity mismatch.

Our proxy for the customers' liquidity demand is the net of all seller-initiated orders and buyer-initiated orders from the retail aggregator's clients, that is, the net selling pressure. This definition follows our assumption that any online broker acts as an intermediary between retail clients and large dealers, and such brokers are competitive. The net selling pressure can then be seen as *a scaled proxy* of the total customer shocks  $\Theta_t$ .<sup>30</sup> However, clients' orders arrive at a much lower frequency than the frequency at which dealers update their quotes, thus, the obtained time series of the proxy of  $\Theta_t$  is sparse. To mitigate this problem, we smooth the original time series of the aggregate clients' order flow by taking a five-minute moving average.

Finally, our proxy for the D2D price  $\mathcal{P}^{D2D}$  and bid-ask spreads are the mid-price and half-Bid-Ask spread in the inter-dealer market, which we compute using the EBS record for orders of size 1-mln, as follows:

$$\mathcal{P}_t^{D2D} = \frac{BestBid_t + BestAsk_t}{2} \quad \text{and} \quad Spread_{D2D}_t = \frac{(BestAsk_{l,t} - BestBid_{l,t})}{2}.$$

We construct proxies for order flow in the D2D market using the EBS dataset, aggregated at the second level, which provides volume and buy/sell information for each executed transaction. We use completed orders of size 1 million and construct the following three proxies for order flow in the D2D market:

---

<sup>30</sup>Or  $\tilde{q}$  in a simple model.

- *D2D\_Vol*: Aggregate daily (signed) net order flow (in mln) in the D2D market. This variable stands for **daily** net order flow (in mln) in the D2D market.
- *D2D\_Vol\_1h*: Aggregate (signed) net order flow (in mln) in the D2D market computed within a trailing window of 1 hour.
- *D2D\_Vol\_24h*: Aggregate (signed) net order flow (in mln) in the D2D market computed within a trailing window of 24 hours.

We report the summary statistics for the final merged dataset in Table 1. Mid-price distributions are essentially the same in both the D2C and the D2D markets, while the spreads differ.<sup>31</sup> This is because dealers have much larger trades relative to customers, hence, larger bid-ask spreads in a linear price impact model like ours. Moreover, dealers know that customers’ order flow (from the retail aggregator) is uninformative.

### 6.3 Empirical results

We begin with a test of Predictions 1 and 3. We test that our novel D2C price-based measure of liquidity mismatch positively predicts future FX exchange rates while customers’ order flow negatively predicts those rates. To this end, we estimate (8). We regress changes in D2D log prices on both lagged changes in D2C prices and lagged clients’ order flow proxies, with the lag  $\ell = 10$  seconds. We also standardize independent variables (subtracting the sample mean and dividing by the sample standard deviation). Table 2 reports our main estimation results for the 10-second lag. Heteroscedasticity and auto-correlation robust standard errors are shown in parentheses. The first column shows that our D2C price-based measure of liquidity mismatch,  $\mathcal{A}_{mismatch}$ , negatively and significantly forecasts price fluctuations in the D2D market. This is entirely consistent with Prediction 1. In terms of economic significance,

---

<sup>31</sup>Differences in prices are small because FX markets are very liquid and have very narrow bid-ask spreads. As one can see from Table 1, a typical bid-ask spread is about 0.3-0.5 basis points. So one can only see the difference in the fifth digit after the comma, whereas we only show the first four digits after the comma in Table 1 for  $\bar{\alpha}$  and  $\mathcal{P}^{D2D}$ .

**Table 1:** Summary statistics

Variable	mean	q50	min	max	q05	q95	sd
$\bar{\alpha}$	1.1187	1.1186	1.0952	1.1416	1.1012	1.1349	0.0096
$\mathcal{P}^{D2D}$	1.1187	1.1186	1.0952	1.1416	1.1012	1.1349	0.0096
$10^{-8} \cdot \mathcal{A}_{mismatch}$	-0.6718	0.0003	-5605.62	20711.5	-0.0671	0.0721	96.585
<i>ProxyLiqShock</i>	79.92	13.63	-1003498	940646.9	-11103.46	11560.01	14447.76
<i>D2D_Vol</i> (mln)	65.0821	49	-280	343	-202	322	148.37
<i>D2D_Vol_1h</i> (mln)	2.8236	2	-141	142	-29	41	21.68
<i>D2D_Vol_24h</i> (mln)	54.0072	40	-464	388	-185	299	138.04
$10^{-5} \cdot \overline{Spread\_D2C}$	4.7849	4.1111	1.8750	667	2.9375	8.3571	6.7528
$10^{-5} \cdot STD\_Spread\_D2C$	2.8242	2.1380	0.3780	1786.72	1.0501	4.8990	17.8299
$10^{-5} \cdot Spread\_D2D$	4.9185	5.0000	0.0091	150	2.5000	10.0000	3.6401

an increase in the cross-sectional liquidity-risk mismatch in the D2C market by one sample standard deviation results in a 19 basis points decrease in the FX exchange rate. This corresponds to an aggregate daily mispricing of \$259 mln.<sup>32</sup>

Column 2 shows that customers’ order flow in the D2C market negatively forecasts price changes in the D2D market, consistent with Prediction 3. The negative sign means that an increase in the customers’ liquidity demand results in lower future prices. Given that a significant fraction of the retail aggregator’s customers are based in Switzerland, it is astonishing that their order flow predicts changes in the EUR/USD exchange rates quoted on EBS, one of the largest international inter-dealer platforms in the world. Note, however, that the economic significance of customer order flow is much lower than that of the liquidity mismatch. We hypothesize that there is a common component in global customer order flow that our proxy for liquidity shock picks up. The remaining columns of the table show that controlling for lagged D2D order flow keeps our main results the same.

We repeat the same empirical test for lags  $\ell = 1, 3, 5, 15, 20, 30, 40, 45, 50$ , and 60 seconds and present the results in Section E.1. There are two main observations following from that section. First, the impact of liquidity mismatch on future D2D prices is significant for small lags  $\ell$  but loses significance as we increase  $\ell$ . This result is intuitive: Liquidity mismatch matters because price impact precludes heterogeneously risk-averse dealers from immediately trading toward the efficient allocation of risk. Instead, they are forced to trade toward their risk target over multiple rounds. It takes dealers several seconds to rebalance their portfolios in an ultra-fast market such as foreign exchange. Second, the predictive power of the D2C customers’ liquidity demand (*ProxyLiqShock*) is significant over longer horizons relative to both liquidity mismatch and the average mid-price in the D2C market. This suggests that customers’ liquidity shocks induce (relatively) long-term changes in dealers’ inventories.

---

<sup>32</sup>There are 74 days in our sample, so an increase in mismatch by one daily standard deviation results in a  $19/\sqrt{74} = 2.2$  basis points decrease in the FX exchange rate (recall that we standardize the variables by dividing them by *sample* standard deviation). Multiplying the daily EUR/USD volume of \$1,172 trillion (Bank for International Settlements (2016)) by 2.2 basis points results in the stated number.

[Include Table 2 here.]

We turn to test Prediction 2 of our model: The dispersion of spreads in the D2C market negatively predicts the spreads in the D2D market. To this end, we estimate (9).<sup>33</sup> Table 3 presents the results of this estimation, with heteroscedasticity and auto-correlation robust standard errors shown in parentheses. It shows that the average D2C spread positively predicts D2D spreads, which is intuitive. The table also shows that the standard deviation of D2C spreads also predicts D2D spreads, but negatively so. The economic magnitude of this effect is quite large: A sample standard deviation increase in the spread dispersion leads to a \$0.61 decrease in spreads. This corresponds to the aggregate daily decrease in trading costs of \$74 bln when spread dispersion increases by one daily standard deviation.<sup>34</sup> This result provides strong empirical support for our model’s prediction that spread dispersion in the D2C market negatively predicts future D2D spreads. Controlling for D2D order flow and customers’ liquidity shocks does not impact the results.

We repeat this empirical test for lags  $\ell = 1, 3, 5, 15, 20, 30, 40, 45, 50$ , and 60 and present the results in Section E.2. As one can see, regression results strongly support our empirical predictions at all horizons. The predictive power of the standard deviation of D2C spreads decreases as the lags increase but remains statistically significant for longer lags.

[Include Table 3 here.]

Finally, we test Predictions 4 and 5. To this end, we run two regressions,

$$\Delta pc_{t+\ell, \ell} = a_0 + a_1 \Delta pc_{t, \ell}, \text{ and } \Delta \mathcal{A}_{mismatch, t+\ell, \ell} = a_0 + a_1 \Delta \mathcal{A}_{mismatch, t, \ell},$$

---

<sup>33</sup>As in (8), we standardize independent variables (subtract sample mean and divide by sample standard deviation).

<sup>34</sup>Table 3 implies that 1 sample standard deviation (74 days) increase in the spread dispersion leads to a \$0.61 decrease in spreads. Then, the daily change in spreads, relative to the average EUR/USD rate of 1.1187 (see Table 1), is  $0.61/\sqrt{74} \cdot 1.1187$ . Multiplying this by the daily EUR/USD volume of \$1.172 trillion (Bank for International Settlements (2016)) yields the stated number.

with a lag  $\ell = 10s$ .<sup>35</sup> We expect  $a_1 < 0$  in both. Tables 5 and 4 present the results. In line with our predictions, there is negative auto-correlation in both  $pc$  and  $\mathcal{A}_{mismatch}$ . In Section E.4, we repeat these empirical tests for lags  $\ell = 1, 3, 5, 15, 20, 30, 40, 45, 50$ , and 60. The negative auto-correlation is present and is statistically significant for all these lags.

[Include Tables 5 and 4 here.]

## 7 Literature Review

A growing body of literature shows how order flow and dealer inventories serve as essential determinants of exchange rates. Lyons (1995) was one of the first to provide strong empirical evidence that dealers actively control their inventories and study how this inventory control creates a link between order flow and exchange rates. In particular, consistent with classical theories (see, e.g., Ho and Stoll (1981)), Lyons (1995) shows that dealers “shade prices”—that is, they shift prices in the direction opposite to their inventory. While such price shading has been documented in other markets,<sup>36</sup> recent studies, such as Bjørnnes and Rime (2005) and Osler et al. (2011), have not found evidence of price shading in the FX markets. In contrast to these papers, and in agreement with Lyons (1995), our empirical results provide strong evidence of price shading in the D2C market.

In a seminal contribution, Evans and Lyons (2002b) develop the first theoretical model to account for the two-tier structure of the FX market and derive an endogenous link between order flow and exchange rates. Consistent with the constraints that the real-world FX dealers face, Evans and Lyons (2002b) assumes that dealers need to hold zero inventory overnight. As a result, dealers optimally shade their prices to achieve the zero inventory target. This price shading leads to a contemporaneous relationship between order flow and exchange rates.

---

<sup>35</sup>We report the estimates of the price competitiveness regressions for one particular dealer. The results for other dealers are similar and are available upon request.

<sup>36</sup>See, e.g., (Madhavan and Smidt, 1993) and (Dunne et al., 2010).

Previous research has mostly focused on interdealer trading. An incomplete list of papers focusing on inter-dealer trading includes [Reiss and Werner \(1998\)](#), [Reiss and Werner \(2005\)](#) and, more recently, [Li and Schürhoff \(2019\)](#). A recent paper of [Eisfeldt, Herskovic and Liu \(2022\)](#) (circulated after this one), considers a corporate bond market and shows that when interdealer price dispersion is high, bond prices are low. Similarly, we show that exchange rates are low when liquidity mismatch is high. Our analysis (available upon request) found that D2C price dispersion is not statistically significant in predicting D2D prices. At least in the FX market, the liquidity mismatch appears to be a better measure than price dispersion. We note, however, that liquidity mismatch cannot be computed in all markets as one needs to see both bid and ask prices at all times.

Although many papers study (both theoretically and empirically) prices and liquidity in the D2D segment of the FX market, to the best of our knowledge, there are no papers that study the joint price formation in the D2C and D2D segments. This is an important gap in our understanding of FX markets, especially given that the D2C trading volume is higher than the D2D volume (see, [Moore et al. \(2016\)](#)). Our paper seeks to fill this gap. Similar to [Evans and Lyons \(2002b\)](#), ours is a pure inventory-theoretic model.<sup>37</sup> However, it differs from that of [Evans and Lyons \(2002b\)](#) in several important dimensions. First, we assume that dealers are heterogeneous in their risk-bearing capacities. Second, we introduce strategic competition between dealers for order flow in the D2C market. Third, we assume that dealers are also strategic when trading in the D2D market. This assumption is crucial for our main results: It implies that (heterogeneous) dealers have (heterogeneous) price impact and, hence, the joint distribution of inventories and price impacts (as captured by the liquidity mismatch) matters for equilibrium prices and allocations. Finally, because we apply our model to very short horizons (up to ten seconds), we do not need to impose the

---

<sup>37</sup>See, [Bacchetta and Van Wincoop \(2006\)](#), [Evans et al. \(2011\)](#), [Evans and Lyons \(2005c\)](#); [Cao et al. \(2006\)](#), [Frankel et al. \(2009\)](#), [Lyons et al. \(2001\)](#), [King et al. \(2010\)](#), [Michaelides et al. \(2018\)](#), and [Gargano et al. \(2018\)](#) for FX microstructure models of exchange rates that rely on private, heterogeneous information. Incorporating such informational asymmetries into our model is an important direction for future research.

zero inventory constraint. We believe that all of these new ingredients are important for correctly modeling real-world FX markets and allow us to capture new effects that have not previously been studied in the literature.

Our theoretical model is related to several existing models of market fragmentation. [Dunne et al. \(2015\)](#) develop a dynamic model of dealer intermediation between a monopolistic customer-dealer market with homogeneous dealers and a competitive inter-dealer market. [Vogler \(1997\)](#) develops a fragmented market model with homogeneous dealers. The inter-dealer market in [Vogler \(1997\)](#) operates through the double auction protocol, which is similar to ours, but absent heterogeneity, the notion of liquidity mismatch does not arise. Most importantly, [Vogler \(1997\)](#) assumes Bertrand competition between dealers in the D2C market: Dealers quote a single price, and customers can both buy and sell unlimited amounts at this price.<sup>38</sup> Thus, counterfactually, there is no bid-ask spread in the D2C market. Summarizing, neither [Vogler \(1997\)](#) nor [Dunne et al. \(2015\)](#) can account for the key features of our model: (1) dealer heterogeneity; (2) imperfect competition (price impact) in the inter-dealer market; and (3) illiquid D2C markets. All of these features are crucial for our ability to compute the price-based liquidity mismatch from the observed cross-section of D2C prices and spreads.<sup>39</sup>

[Colliard et al. \(2018\)](#) study the effects of market fragmentation and dealer market power in which dealers differ in their connectivity. However, they deliberately keep the matching process between dealers and their clients simple to focus on the D2D market. While their market mechanisms are well-suited for many OTC markets, such as the bond market, ours could be better suited for modeling forex trading since the market structure is very different.

The closest to ours is the paper by [Babus and Parlato \(2018\)](#), who developed a fragmented market model with identical (homogeneous) dealers and a D2D double auction market protocol as in our paper. In addition, they assume that each dealer runs a local

---

<sup>38</sup>[Dunne et al. \(2015\)](#) also consider an extension of their model with the same trading protocol.

<sup>39</sup>Similarly, the models of [Ho and Stoll \(1983\)](#) and [Biais \(1993\)](#), related to our simple model abstract from dealer heterogeneity.



D2C market and has a customer base that can only trade in this local market with one particular dealer. They show how market fragmentation can arise endogenously when customers endogenously decide upon the local market they want to join. Thus, in [Babus and Parlato \(2018\)](#), conditional on the chosen market participation structure, competition between dealers is nonexistent. Although their model might apply to some real-world fragmented markets, the assumed market structure differs from the structure we observe in real-world FX markets. Yet, introducing dealer heterogeneity into their model would lead to dispersion in dealer-specific prices and spreads (price impacts) in the D2C market. Hence, their model might produce the joint behavior of D2D and D2C prices that we observe in our data. In Appendix G, we extend the model of [Babus and Parlato \(2018\)](#) to allow for dealer heterogeneity and then derive the equilibrium relationship between the price-based liquidity mismatch in the D2C market and the price level in the D2D market. Surprisingly, we find that this model implies the sign of a relationship that is opposite to the one that we observe in the data. By contrast, our model does indeed produce the empirically observed positive relationship between the liquidity mismatch and D2D prices. In Section 3.1, we explain how the nature of competition between dealers drives this difference in model predictions. Namely, in our model (as in the real world), dealers in the D2C market compete in mechanisms (price schedules; see [Biais et al. \(2000, 2013\)](#) and [Back and Baruch \(2013\)](#)). This competition implies that they use D2C markets to aggressively manage their inventories to the extent that their post-D2C inventories are negatively related to pre-D2C inventories, reverting the sign of the link between D2C and D2D markets.

Our notion of liquidity mismatch is indirectly related to the liquidity mismatch defined in [Brunnermeier and Krishnamurthy \(2012\)](#) (see also [Bai et al. \(2018\)](#)) as the mismatch between the market liquidity of assets and the funding liquidity of liabilities. It is natural to expect that the inventory cost in our model is closely related to the dealers' funding liquidity. Hence, the total inventory cost for a given dealer is linked to the liquidity mismatch of

Brunnermeier and Krishnamurthy (2012), defined on a single bank level. However, in stark contrast to Brunnermeier and Krishnamurthy (2012), our liquidity mismatch measure is cross-sectional and captures a mismatch in the distribution of assets across different dealers with different liquidity needs and different price impacts. The fact that one can identify the mismatch directly from prices in the D2C segment is a surprising and novel prediction of our model.

Babus and Parlato (2018) belongs to a larger stream of literature on fragmented markets (see, e.g., Babus and Kondor (2018), Malamud and Rostek (2017), and Babus and Parlato (2018)) that assumes identical, auction-like trading protocols in the two market segments.<sup>40</sup> Under such protocols, effectively, dealers do not play any unique role, resulting in customers and dealers ending up equally providing liquidity to each other. We show that, contrary to our model, this behavior leads to an opposite (negative) sign of the relationship between liquidity mismatch and price level. Empirically, we find strong evidence for the positive relationship, suggesting that our modeling of D2C trading is crucial for matching the data.

Numerous papers provide evidence that order flow contains information about contemporaneous (Evans and Lyons (2002b), Evans and Lyons (2002a), Hau et al. (2002), Fan and Lyons (2003), Froot and Ramadorai (2005), Bjørnnes et al. (2005), Danielsson and Love (2006), Killeen et al. (2006), Berger et al. (2008), Brunnermeier et al. (2008), King et al. (2010), Rime et al. (2010), Breedon and Vitale (2010), and Bjørnnes et al. (2011)) and future (Evans and Lyons (2005b), Danielsson et al. (2012), and Evans and Rime (2016), Collin-Dufresne et al. (2019)) prices. In this paper, our focus is on predictive relationships at horizons that are short enough so that efficient allocation cannot be achieved due to market imperfections. The key insight from our model is that price dispersion in the D2C market can be used to recover information about the distribution of inventories and the liquidity mismatch. This novel, endogenous, purely price-based object is unique to our model.

---

<sup>40</sup>Some papers (see, e.g., Liu et al. (2017)) also assume a competitive D2D market, implying that the allocation of risk among dealers is irrelevant.

Our model assumes that dealer inventories constitute an important driving force behind price dynamics. Recent empirical evidence supports this assumption. For instance, [Friewald et al. \(2019\)](#), [Anderson and Liu \(2019\)](#), and [Randall \(2015\)](#) show that inventory costs explain a significant fraction of yield spread changes in corporate bonds. [Hendershott and Menkveld \(2014\)](#) provide similar evidence for equity markets. Numerous papers have also shown the importance of the D2D market as a key venue for inventory management. See, for example, [Schultz \(2017\)](#) [Schürhoff and Li \(2019\)](#), [Collin-Dufresne et al. \(2019\)](#), [Hollifield et al. \(2017\)](#), and [Anderson and Liu \(2019\)](#). The most closely related to ours is the paper by [Collin-Dufresne et al. \(2019\)](#), who study two-tiered CDS markets and show that bid-ask spreads are wider in the D2C market, and D2C transactions are largely institutional trades that have a large permanent price impact. By contrast, the FX market is relatively less concentrated and has a significant retail segment, which is captured in our data. We find that bid-ask spreads are narrower in the D2C market and have only a small, transitory price impact, consistent with our inventory-driven model.

Our main predictions stem from dealers’ heterogeneity. Data strongly support the presence of such heterogeneity. For example, [Evans \(2002\)](#) found that most short-term volatility in exchange rates is a result of dealers’ heterogeneous trading decisions. Similarly, [Bjønnes and Rime \(2005\)](#) document significant differences in dealers’ trading styles, especially related to how they control their inventories. [Randall \(2015\)](#) provides evidence of heterogeneous and time-varying costs of holding inventory for dealers in the US corporate bond market. Our data on dealer quotes in the D2C market also suggests the presence of significant heterogeneity: Both the bid-ask spreads and the sensitivity of quotes to shocks are highly heterogeneous across dealers. [Han et al. \(2022\)](#) argue that dispersed beliefs represent another important source of dealer heterogeneity. While dealer disagreement is undoubtedly important, in our model, heterogeneous inventory generates a distinctive prediction about the link between prices in the two market segments that heterogeneous beliefs cannot generate.

Investigating the joint role of inventories and beliefs in FX markets is an important direction for future research.

Most of our key predictions depend crucially on the fact that the FX market is not perfectly liquid, and dealers have a price impact. As a result, markets cannot efficiently allocate risk at short horizons, and the distribution of inventories across dealers impacts price dynamics. However, the idea that price impact is linked to order flow is not new (see, for example, [Evans and Lyons \(2005a\)](#)), to the best of our knowledge, our model is the first to micro-found this price impact in a model that accounts for the two-tier structure of the FX market. In particular, our model can be used to recover market liquidity from dealer quotes, providing an explicit, micro-founded measure of liquidity risk and shedding new light on the findings of [Banti et al. \(2012\)](#) and [Mancini et al. \(2013\)](#).

## 8 Conclusions

Foreign exchange markets are highly fragmented and are dominated by a handful of large strategic dealers. Due to balance sheet shocks, different capitalizations, and various frictions, these dealers differ in their willingness and ability to take on risk. This creates a risk-liquidity mismatch: Large risk-averse dealers would like to get rid of their inventory but cannot due to their price impact. This distorts their liquidity provision in the D2C market, ultimately affecting customers' ability to trade efficiently. We develop the first theoretical model that can quantify these frictions. Our model generates several explicit predictions linking prices and spreads in the D2C and D2D market segments. We test these predictions empirically and find strong support for the mechanisms underlying our model.

Our model has two drawbacks. First, it is effectively static (there are only two trading rounds: one D2C and one D2D round). Second, trading happens purely for risk-sharing purposes; hence, we completely ignore frictions due to asymmetric information. Incorporating dynamics and adverse selection into our model is an important direction for future research.

## References

- Anderson, Chris and Weiling Liu**, “Intermediary Trading and Risk Constraints,” Working paper 2019.
- Babus, Ana and Cecilia Parlatore**, “Strategic Fragmented Markets,” Technical Report, NYU 2018.
- and **Peter Kondor**, “Trading and Information Diffusion in Over-the-Counter markets,” *Econometrica*, 2018, *86* (5), 1727–1769.
- Bacchetta, Philippe and Eric Van Wincoop**, “Can information heterogeneity explain the exchange rate determination puzzle?,” *The American Economic Review*, 2006, *96* (3), 552–576.
- Back, Kerry and Shmuel Baruch**, “Strategic liquidity provision in limit order markets,” *Econometrica*, 2013, *81* (1), 363–392.
- Bai, Jenny, Arvind Krishnamurthy, and Charles-Henri Weymuller**, “Measuring Liquidity Mismatch in the Banking Sector,” *Journal of Finance*, 2018, *73* (1), 51–93.
- Banti, Chiara, Kate Phylaktis, and Lucio Sarno**, “Global liquidity risk in the foreign exchange market,” *Journal of International Money and Finance*, 2012, *31* (2), 267–291.
- Berger, David W, Alain P Chaboud, Sergey V Chernenko, Edward Howorka, and Jonathan H Wright**, “Order flow and exchange rate dynamics in electronic brokerage system data,” *Journal of international Economics*, 2008, *75* (1), 93–109.
- Biais, Bruno**, “Price Formation and Equilibrium Liquidity in Fragmented and Centralized Markets,” *The Journal of Finance*, 1993, *48*, 157–185.
- , **David Martimort, and Jean-Charles Rochet**, “Competing mechanisms in a common value environment,” *Econometrica*, 2000, *68* (4), 799–837.
- , —, and —, “Competing mechanisms in a common value environment: A Corrigendum,” *Econometrica*, 2013, *81* (4), 393–406.
- Bjønnes, Geir H, Carol L Osler, Dagfinn Rime, and Norges Bank**, “Sources of information advantage in the foreign exchange market,” *Norges Bank*, 2011, p. 46.
- Bjønnes, Geir Høidal and Dagfinn Rime**, “Dealer behavior and trading systems in foreign exchange markets,” *Journal of Financial Economics*, 2005, *75* (3), 571 – 605.
- , —, and **Haakon O Aa Solheim**, “Liquidity provision in the overnight foreign exchange market,” *Journal of International Money and Finance*, 2005, *24* (2), 175–196.
- Breedon, Francis and Paolo Vitale**, “An empirical study of portfolio-balance and information effects of order flow on exchange rates,” *Journal of International Money and Finance*, 2010, *29* (3), 504–524.
- Brunnermeier, Markus K, Stefan Nagel, and Lasse H Pedersen**, “Carry trades and currency crashes,” *NBER macroeconomics annual*, 2008, *23* (1), 313–348.
- Cao, H Henry, Martin D Evans, Richard K Lyons et al.**, “Inventory Information,” *The Journal of Business*, 2006, *79* (1), 325–364.
- Cenedese, Gino, Pasquale Della Corte, and Tianyu Wang**, “Currency mispricing and dealer balance sheets,” *The Journal of Finance*, 2021, *76* (6), 2763–2803.

- Colliard, Jean-Edouard, Thierry Foucault, and Peter Hoffmann, “Inventory Management, Dealers’ Connections and Prices in OTC Markets,” *Working Paper*, 2018.
- Collin-Dufresne, Pierre, Benjamin Junge, and Trolle anders B, “Market Structure and Transaction Costs of Index CDSs,” *Journal of Finance*, 2019.
- Danielsson, Jon and Ryan Love, “Feedback trading,” *International Journal of Finance & Economics*, 2006, 11 (1), 35–53.
- , Jinhui Luo, and Richard Payne, “Exchange rate determination and inter-market order flow effects,” *The European Journal of Finance*, 2012, 18 (9), 823–840.
- Duffie, Darrell, Nicolae Gârleanu, and Lasse Heje Pedersen, “Over-the-counter markets,” *Econometrica*, 2005, 73 (6), 1815–1847.
- Dunne, Peter G, Harald Hau, and Michael Moore, “A tale of two platforms: dealer intermediation in the European sovereign bond market,” *Journal of the European Economic Association*, 2010.
- Dunne, Peter, Harald Hau, and Michael Moore, “Dealer Intermediation Between Markets,” *Journal of the European Economic Association*, 2015, 13 (5), 770–804.
- Eisfeldt, Andrea L, Bernard Herskovic, and Shuo Liu, “Interdealer Price Dispersion,” *Available at SSRN*, 2022.
- Evans, Martin DD, “FX trading and exchange rate dynamics,” *The Journal of Finance*, 2002, 57 (6), 2405–2447.
- and Dagfinn Rime, “Order flow information and spot rate dynamics,” *Journal of International Money and Finance*, 2016, 69, 45–68.
- and Richard K Lyons, “Informational integration and FX trading,” *Journal of International Money and Finance*, 2002, 21 (6), 807–831.
- and —, “Order flow and exchange rate dynamics,” *Journal of political economy*, 2002, 110 (1), 170–180.
- and —, “Do currency markets absorb news quickly?,” *Journal of International Money and Finance*, 2005, 24 (2), 197–217.
- and —, “Meese-Rogoff Redux: Micro-Based Exchange-Rate Forecasting,” *American Economic Review*, 2005, pp. 405–414.
- Evans, Martin D.D. and Richard K. Lyons, “A New Micro Model of Exchange Rate Dynamics,” Working paper, Georgetown University 2005.
- Evans, Martin DD et al., *Exchange-Rate Dynamics*, Princeton University Press, 2011.
- Fan, Mintao and Richard K Lyons, “Customer trades and extreme events in foreign exchange,” *Monetary History, Exchange Rates and Financial Markets: Essays in Honour of Charles Goodhart*, 2003, pp. 160–179.
- for International Settlements, (BIS) Bank, *Foreign exchange and derivatives market activity in April 2016*, Triennial Central Bank Survey., 2016.
- Frankel, Jeffrey A, Giampaolo Galli, and Alberto Giovannini, *The microstructure of foreign exchange markets*, University of Chicago Press, 2009.
- Friewald, Nils, , and Florian Nagler, “Over-the-Counter Market Frictions and Yield Spread Changes,” *The Journal of Finance*, 2019, 74, 3217–3257.

- Froot, Kenneth A and Tarun Ramadorai**, “Currency Returns, Intrinsic Value and Institutional-Investor Flows,” *The Journal of Finance*, 2005, 60 (3), 1535–1566.
- Gargano, Antonio, Steven J. Riddiough, and Lucio Sarno**, “The Value of Volume in Foreign Exchange,” Working Paper, University of Melbourne 2018.
- Glebkin, Sergei, Semyon Malamud, and Alberto Teguia**, “Illiquidity and Higher Cumulants,” 2022.
- Han, Brandon Yueyang, Sophia Zhengzi Li, and Zhaogang Song**, “Dealer Disagreement and Asset Prices in FX Markets,” *Available at SSRN*, 2022.
- Hau, Harald, William Killeen, and Michael Moore**, “The euro as an international currency: explaining puzzling first evidence from the foreign exchange markets,” *Journal of International Money and Finance*, 2002, 21 (3), 351–383.
- Hendershott, Terrence and Albert J. Menkveld**, “Price pressures,” *Journal of Financial Economics*, 2014, 114, 405–423.
- Ho, Thomas and Hans R Stoll**, “Optimal dealer pricing under transactions and return uncertainty,” *Journal of Financial economics*, 1981, 9 (1), 47–73.
- Ho, Thomas SY and Hans R Stoll**, “The dynamics of dealer markets under competition,” *The Journal of finance*, 1983, 38 (4), 1053–1074.
- Hollifield, Burton, Artem Neklyudov, and Chester Spatt**, “Bid-ask spreads, trading networks and the pricing of securitizations,” *Review of Financial Studies*, 2017, 30, 3048–3085.
- K., Gary Gorton Brunnermeier Markus and Arvind Krishnamurthy**, “Risk topography,” *NBER Macroeconomics Annual*, 2012, 26, 149–176.
- Killeen, William P, Richard K Lyons, and Michael J Moore**, “Fixed versus flexible: Lessons from EMS order flow,” *Journal of International Money and Finance*, 2006, 25 (4), 551–579.
- King, Michael, Lucio Sarno, and Elvira Sojli**, “Timing exchange rates using order flow: The case of the Loonie,” *Journal of Banking & Finance*, 2010, 34 (12), 2917–2928.
- Krishna, Vijay**, *Auction theory*, Academic press, 2009.
- Kyle, Albert S**, “Continuous auctions and insider trading,” *Econometrica: Journal of the Econometric Society*, 1985, pp. 1315–1335.
- , “Informed speculation with imperfect competition,” *The Review of Economic Studies*, 1989, 56 (3), 317–355.
- Li, Dan and Norman Schürhoff**, “Dealer networks,” *The Journal of Finance*, 2019, 74 (1), 91–144.
- Liu, Ying, Sebastian Vogel, and Yuan Zhang**, “Electronic Trading in OTC Markets vs. Centralized Exchange,” Technical Report, SFI 2017.
- Lyons, Richard K**, “Tests of microstructural hypotheses in the foreign exchange market,” *Journal of Financial Economics*, 1995, 39 (2), 321–351.
- et al., *The Microstructure Approach to Exchange Rates*, Vol. 1, The MIT Press, 2001.
- Madhavan, Ananth and Seymour Smidt**, “An analysis of changes in specialist inventories and quotations,” *The Journal of Finance*, 1993, 48 (5), 1595–1628.

- Malamud, Semyon and Marzena Rostek**, “Decentralized Exchange,” *American Economic Review*, 2017, 107 (11), 3320–3362.
- Mancini, Lorian, Angelo Rinaldo, and Jan Wrampelmeyer**, “Liquidity in the foreign exchange market: Measurement, commonality and risk premiums,” *The Journal of Finance*, 2013, 68 (5), 1805–1841.
- Meese, Richard A and Kenneth Rogoff**, “Empirical exchange rate models of the seventies: Do they fit out of sample?,” *Journal of international economics*, 1983, 14 (1-2), 3–24.
- Michaelides, Alexander, Andreas Milidonis, and George Nishiotis**, “Private Information in Currency Markets,” Technical Report, Journal of Financial Economics (forthcoming) 2018.
- Moore, Michael, Andreas Schrimpf, and Vladyslav Sushko**, “Downsized FX markets: causes and implications,” *BIS Quarterly Review*, December 2016.
- Osler, Carol L, Alexander Mende, and Lukas Menkhoff**, “Price discovery in currency markets,” *Journal of International Money and Finance*, 2011, 30 (8), 1696–1718.
- Randall, Oliver**, “How Do Inventory Costs Affect Dealer Behavior in the US Corporate Bond Market?,” *Working Paper*, 2015.
- Reiss, Peter C and Ingrid M Werner**, “Does risk sharing motivate interdealer trading?,” *The Journal of Finance*, 1998, 53 (5), 1657–1703.
- and —, “Anonymity, adverse selection, and the sorting of interdealer trades,” *The Review of Financial Studies*, 2005, 18 (2), 599–636.
- Rime, Dagfinn, Lucio Sarno, and Elvira Sojli**, “Exchange rate forecasting, order flow and macroeconomic information,” *Journal of International Economics*, 2010, 80 (1), 72–88.
- Rostek, Marzena and Marek Weretka**, “Dynamic thin markets,” *The Review of Financial Studies*, 2015, 28 (10), 2946–2992.
- Schultz, Paul**, “Inventory management by corporate bond dealers,” Working paper 2017.
- Schürhoff, Norman and Dan Li**, “Dealer Networks,” *forthcoming in the Journal of Finance*, 2019.
- Vayanos, Dimitri**, “Strategic trading and welfare in a dynamic market,” *The Review of Economic Studies*, 1999, 66 (2), 219–254.
- and **Paul Woolley**, “An institutional theory of momentum and reversal,” *The Review of Financial Studies*, 2013, 26 (5), 1087–1145.
- Vives, Xavier**, “Strategic supply function competition with private information,” *Econometrica*, 2011, 79 (6), 1919–1966.
- Vogler, Karl-Hubert**, “Risk allocation and inter-dealer trading,” *European Economic Review*, 1997, 41, 1615–1634.



## A Main Proofs

### A.1 Proof of Proposition 2

**Proof of Proposition 2.** The results regarding the dealers' price impact ( $\beta_\ell$ ) and  $\mathcal{B}$ , except (ii), are given in [Malamud and Rostek \(2017\)](#). We now prove (ii) :  $\beta_l$  is monotone increasing  $\Gamma_\ell$  for any  $\ell \neq l$ . Fix  $l$ . It follows from differentiation that

$$\beta_l(\mathcal{B}) = \frac{2\Gamma_l}{\Gamma_l\mathcal{B} - 2 + \sqrt{(\Gamma_l\mathcal{B})^2 + 4}}$$

is increasing in  $\mathcal{B}$ . Let  $\mathbf{\Gamma} = \{\Gamma_\ell\}_{\ell=1}^M$  be a set of dealers' risk-aversion and suppose that  $\tilde{\mathbf{\Gamma}} = \{\tilde{\Gamma}_\ell\}_{\ell=1}^M$  is such that

$$\tilde{\Gamma}_l > \Gamma_l \quad \text{and} \quad \tilde{\Gamma}_\ell = \Gamma_\ell \quad \forall \ell \neq l.$$

Since  $\beta_l(\mathcal{B})$  is increasing, it is enough to show that the corresponding  $\mathcal{B}$  and  $\tilde{\mathcal{B}}$  satisfy

$$\tilde{\mathcal{B}} < \mathcal{B}$$

to prove the statement. By definition,

$$\begin{aligned} \frac{1}{2} &= \sum_{\ell} \frac{1}{\Gamma_\ell\mathcal{B} - 2 + \sqrt{(\Gamma_\ell\mathcal{B})^2 + 4}} + \frac{1}{\Gamma_l\mathcal{B} - 2 + \sqrt{(\Gamma_l\mathcal{B})^2 + 4}} \\ &> \sum_{\ell} \frac{1}{\tilde{\Gamma}_\ell\tilde{\mathcal{B}} - 2 + \sqrt{(\tilde{\Gamma}_\ell\tilde{\mathcal{B}})^2 + 4}}. \end{aligned}$$

It follows that  $\tilde{\mathcal{B}} < \mathcal{B}$  since

$$\sum_{\ell} \frac{1}{\tilde{\Gamma}_\ell\tilde{\mathcal{B}} - 2 + \sqrt{(\tilde{\Gamma}_\ell\tilde{\mathcal{B}})^2 + 4}} = \frac{1}{2}.$$

We finish by proving the remaining parts of the proposition. First, we prove that  $\mathcal{B}$  is the aggregate liquidity, that is

$$\mathcal{B} = \sum_{\ell} \frac{1}{\Gamma_\ell + \beta_\ell}$$

We have

$$\begin{aligned} (\beta_\ell + \Gamma_\ell)^{-1} &= \frac{1}{\Gamma_\ell} \left[ 1 - \frac{2}{\Gamma_\ell \mathcal{B} + \sqrt{(\Gamma_\ell \mathcal{B})^2}} \right] = \frac{1}{2\Gamma_\ell} \left[ 2 + \Gamma_\ell \mathcal{B} - \sqrt{(\Gamma_\ell \mathcal{B})^2} \right] \\ &= 2 \left[ \frac{\mathcal{B}}{2 + \Gamma_\ell \mathcal{B} - \sqrt{(\Gamma_\ell \mathcal{B})^2}} \right]. \end{aligned}$$

Summing over  $\ell$  then yields the result.

Next, we multiply Equation 3 by  $(\Gamma_l + \beta_l)^{-1}$  and rearrange the terms to obtain

$$(\Gamma_l + \beta_l)^{-1}(\mathcal{P}^{D2D} - d) = -(\Gamma_l + \beta_l)^{-1}\Gamma_l \chi_l - Q_l,$$

We can then sum over  $\ell$  and use Equation A.1 (which shows that  $\mathcal{B}$  is equal to aggregate liquidity) to obtain

$$\mathcal{P}^{D2D} = \bar{d} - \mathbf{Q}^* - \sum_{l=1}^M Q_l.$$

Equation 3 then follows from market clearing. We finish the proof by plugging

$$Q_\ell = (\Gamma_\ell + \beta_\ell)^{-1}(d - \mathcal{P}^{D2D} - \Gamma_\ell \chi_\ell) = (\Gamma_\ell + \beta_\ell)^{-1}\mathbf{Q}^* - (\Gamma_\ell + \beta_\ell)^{-1}\Gamma_\ell \chi_\ell$$

into  $\tilde{\chi}_\ell \chi_\ell + Q_\ell$  and Equation (16) to obtain Equations 4 and 5.

Q.E.D.

## A.2 Proof of Proposition 3

**Proof of Proposition 3.** Starting with Equation 6, we have

$$\mathcal{P}^{D2D} = d - \mathcal{B}^{-1} M E[(\Gamma_l + \beta_l)^{-1} \Gamma_l] E[\chi_l] - \mathcal{B}^{-1} M \text{Cov}((\Gamma_l + \beta_l)^{-1} \Gamma_l, \chi_l).$$

The first two terms in the R.H.S. of the equation below remain the same if the three cases (a), (b), and (c). Focusing on the last term, we have

$$-\mathcal{B}^{-1} M \text{Cov}((\Gamma_l + \beta_l)^{-1} \Gamma_l, \chi_l) \stackrel{s}{=} \text{Cov}(\chi_l, \Gamma_l).$$

The result then follows.

Q.E.D.

## A.3 Proof of Proposition 4

We will need the following technical lemmas.

**Lemma 10** *The function  $(xy + 2 + \sqrt{(xy)^2 + 4})^{-1}$  is jointly convex  $\mathbb{R}_+$ .*

**Proof.** Consider the function

$$f(x) = \left(x + 2 + \sqrt{x^2 + 4}\right)^{-1} \quad \implies \quad f''(x) = \frac{(\sqrt{x^2 + 4} - 3)x^2 + 4(\sqrt{x^2 + 4} - 2)}{x^3(x^2 + 4)^{3/2}}.$$

We show that  $f''$  is positive for  $x > 0$ . Clearly, its denominator is positive for  $x > 0$ . Consider its numerator

$$h(x) \equiv \left(\sqrt{x^2 + 4} - 3\right)x^2 + 4\left(\sqrt{x^2 + 4} - 2\right) \quad \implies \quad h'(x) = 3x\left(\sqrt{x^2 + 4} - 2\right).$$

Thus,  $h$  is increasing for  $x \geq 0$ . Moreover,  $h(0) = 0$ . Thus  $f$  is convex for  $x > 0$ . It follows that the function

$$G(x, y) \equiv \left(xy + 2 + \sqrt{(xy)^2 + 4}\right)^{-1}$$

is jointly convex in  $(x, y)$ .

Q.E.D.

**Lemma 11** *The function  $\mathcal{B}(\Gamma)$  is jointly convex in the vector  $\Gamma = (\Gamma_l)_{l=1}^M$ .*

**Proof.** Let

$$f(\Gamma_l, B) = \left(\Gamma_l B + 2 + \sqrt{(\Gamma_l B)^2 + 4}\right)^{-1}$$

Let  $B_1(\Gamma^1)$  and  $B_2(\Gamma^2)$  be defined implicitly by

$$\sum_l f(\Gamma_l^1, B_1) = 1/2,$$

$$\sum_l f(\Gamma_l^2, B_2) = 1/2.$$

We need to show that  $B(\lambda\Gamma_l^1 + (1 - \lambda)\Gamma_l^2) > \lambda B_1 + (1 - \lambda)B_2$ . We have

$$\sum_l f(\lambda\Gamma_l^1 + (1 - \lambda)\Gamma_l^2, \lambda B_1 + (1 - \lambda)B_2) > \lambda \sum_l f(\Gamma_l^1, B_1) + (1 - \lambda) \sum_l f(\Gamma_l^2, B_2) = 1/2.$$

Since  $f(\Gamma_l, B)$  decreases in  $B$ , the  $B$  that solves

$$\sum_l f(\lambda\Gamma_l^1 + (1 - \lambda)\Gamma_l^2, \lambda B_1 + (1 - \lambda)B_2) = 1/2$$

is greater than  $\lambda B_1 + (1 - \lambda)B_2$ .

Q.E.D.

**Proof of Proposition 4.** We have  $BA = 2/\mathcal{B}$ . Whew aggregate liquidity  $\mathcal{B}$  is given by

$$\sum_l \frac{1}{\Gamma_l + \beta_l} = \mathcal{B},$$

where the equality was established in the proof of Proposition 2. By definition, a mean preserving spread is such that we move from  $(\Gamma_l)$  to  $(\tilde{\Gamma}_l = \Gamma_l + \varepsilon_l)$  where  $(\varepsilon_l)$  has  $E[\varepsilon_l] = 0$  and is mean-independent of  $(\Gamma_l)$ . We know from Lemma 11 that  $\mathcal{B}(\Gamma_l)$  is convex in  $(\Gamma_l)$ . Hence,

$$\mathcal{B}(\Gamma_l) < \mathcal{B}(\tilde{\Gamma}_l).$$

Furthermore, the function

$$\beta_l^{-1} = \Gamma_l \mathcal{B} - 2 + \sqrt{(\Gamma_l \mathcal{B})^2 + 4}$$

is clearly increasing and convex in  $\mathcal{B}$ . Thus,  $\beta_l^{-1}$  is convex in  $\Gamma_{-l}$ , and Jensen's inequality implies

$$\beta_l^{-1}(\Gamma_{-l}) \leq E[\beta_l^{-1}(\tilde{\Gamma}_{-l})]$$

that is

$$\beta_l \geq E[\beta_l^{-1}(\tilde{\Gamma}_{-l})]^{-1}.$$

Q.E.D.

## A.4 Proof of Proposition 5

**Proof of Proposition 5.** An increase in dispersion in  $b_l$  corresponds to an increase in dispersion in  $\Gamma_l$ , which by Proposition 4 is associated with smaller  $BA$  in D2D market.

Q.E.D.

## A.5 Proof of Proposition 6

**Proof of Proposition 6.** We number traders according to their inventories, i.e.  $x_l$  increases in  $l$ . Then  $\alpha_l$  decreases in  $l$ . Then  $\chi_l - x_l$  decreases in  $l$ . (Inventories increase for the dealer with highest midprice and decrease for the dealer with lowest midprice. They remain unchanged for other dealers. )

Then we get

$$\begin{aligned} cov(\chi_l - x_l, \Gamma_l) &\stackrel{s}{=} cov(\alpha_l, \Gamma_l) \\ &\stackrel{s}{=} cov(\alpha_l, b_l). \end{aligned}$$

Moreover, for  $q$  large enough we get  $\text{cov}(\chi_l - x_l, \Gamma_l) \approx \text{cov}(\chi_l, \Gamma_l)$ .

Q.E.D.

## A.6 Proof of Proposition 7

**Proof of Proposition 7.** It is a standard result of bidding the reservation value in the second price auction is an equilibrium strategy, see, e.g. Krishna (2009), Chapter 2.

We also state the standard results for the first price auction. Krishna (2009), Chapter 2 is a good reference.

**Lemma 12** *Consider  $M$  sellers with i.i.d. reservation values  $r^a$  with continuous CDF  $F^a(\cdot)$  with support  $[\underline{r}^a, \bar{r}^a]$  participating in the first price auction. The equilibrium bidding strategy is given by*

$$p^a(r) = r + \int_r^{\bar{r}^a} \left( \frac{1 - F^a(t)}{1 - F^a(r)} \right)^{M-1} dt.$$

When  $F^a(\cdot)$  is a CDF of uniform distribution, we have

$$p^a(r) = \left( 1 - \frac{1}{M} \right) r + \frac{1}{M} \bar{r}^a.$$

The equilibrium strategy for buyers with i.i.d. reservation values  $r^b$  with continuous CDF  $F^b(\cdot)$  with support  $[\underline{r}^b, \bar{r}^b]$  is given by

$$p^b(r) = r + \int_{\underline{r}^b}^r \left( \frac{F^b(t)}{F^b(r)} \right)^{M-1} dt,$$

which in the case of uniform distribution becomes

$$p^b(r) = \left( 1 - \frac{1}{M} \right) r + \frac{1}{M} \underline{r}^b.$$

The first two statements follow from the results above.

Consider an equilibrium strategy of a seller when  $\bar{r}^a - \underline{r}^a$  is small. One can write that

$$\left( \frac{F^b(t)}{F^b(r)} \right)^{M-1} = \left( \frac{t - \underline{r}^b}{r - \underline{r}^b} \right)^{M-1} + O\left((\bar{r}^b - \underline{r}^b)^M\right),$$

and so

$$p^b(r) = \left( 1 - \frac{1}{M} \right) r + \frac{1}{M} \underline{r}^b + O\left((\bar{r}^b - \underline{r}^b)^M\right).$$

Thus, (1) holds approximately. Note also that in our case terms  $O\left((\bar{r}^b - \underline{r}^b)^M\right)$  are also of order  $O((\Gamma^*)^2)$  ( $\bar{r}^b - \underline{r}^b$  is  $O(\Gamma^*)$  and  $M > 2$ ), and the statements follow.

Q.E.D.

## A.7 Deriving predictions 4 and 5 in the simple model

In this section, we derive predictions 4 and 5 in our simple model. The formal result is provided in the Proposition below.

**Proposition 13** *Suppose that  $\|\Gamma - \Gamma^*\|^2$  is sufficiently small and that the realisation of  $|\tilde{q}|$  is sufficiently large. Define mid-prices  $\alpha'_l$ ,  $b'_l$  as mid-prices and spreads associated with the pre-D2C inventories equal to post-D2D inventories  $x_l = \tilde{\chi}_l$ , holding  $\{\Gamma_l\}_l$  fixed. Then, the dealer posting the highest (lowest)  $\alpha_l$  will be posting the lowest (highest)  $\alpha'_l$ . Define  $\mathcal{A}'_{mismatch} = \text{Cov}(\alpha'_l, b'_l)$ . Under the stated conditions positive (negative)  $\mathcal{A}_{mismatch}$  will be followed by negative (positive)  $\mathcal{A}'_{mismatch}$ .*

**Proof of Proposition 13.** Suppose that  $\tilde{q} > 0$ , i.e., the customers are selling. Consider a dealer  $l^*$  posting the highest  $\alpha_l$ . This dealer will increase his inventories by  $\tilde{q}$ , and, since  $\tilde{q}$  is assumed to be large, will hold the largest inventory  $\chi_{l^*}$  post-D2C. It follows from (4) that post-D2D inventory can be written as  $\tilde{\chi}_{l^*} = (\Gamma^* + \beta^*)\mathbf{Q}^* + (\Gamma^* + \beta^*)^{-1}\beta^*\chi_{l^*} + O(\|\Gamma - \Gamma^*\|)$ , where  $\beta^* = \Gamma^*/(M - 2)$  is the price impact in the D2D market with homogeneous investors. It follows that the dealer with highest  $\chi_{l^*}$  will also end up with highest  $\tilde{\chi}_{l^*}$ . Then, by (7), this dealer will post lowest price  $\alpha'_{l^*}$ .

We now consider the statement about  $\mathcal{A}_{mismatch}$ . For large  $\tilde{q}$  we have  $\mathcal{A}_{mismatch} \stackrel{s}{=} 1/M(\alpha_{l^*} - E[\alpha_l])(\Gamma_{l^*} - E[\Gamma_l])$  and  $\mathcal{A}'_{mismatch} \stackrel{s}{=} 1/M(\alpha'_{l^*} - E[\alpha'_l])(\Gamma_{l^*} - E[\Gamma_l])$ . Since the signs of  $(\alpha_{l^*} - E[\alpha_l])$  and  $(\alpha'_{l^*} - E[\alpha'_l])$  are the opposites, the statement follows. Q.E.D.

## B Regression tables

### B.1 Main Results

**Table 2: Forecasting FX rates: Ten Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 10s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	0.39215320*** (0.00247951)	0.39272230*** (0.00276747)	0.39241120*** (0.00293718)	0.39278770*** (0.00303912)	0.39268750*** (0.00309929)	0.39265200*** (0.00302175)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-19.34687000*** (2.01114500)	-19.40294000*** (2.04942700)	-19.35697000*** (2.07074800)	-19.40267000*** (2.08287900)	-19.40140000*** (2.08964900)	-19.39819000*** (2.08014200)
ProxyLiqShock <sub>t</sub>		-0.01146765*** (0.00209185)		-0.01149864*** (0.00205862)	-0.01147806*** (0.00205179)	-0.01145615*** (0.00205795)
$D2D.Vol_t$			0.00101904 (0.00075160)	0.00109480 (0.00074938)		
$D2D.Vol.1h_t$					-0.00036481 (0.00109628)	
$D2D.Vol.24h_t$						0.00106639 (0.00074399)
Constant	-5.94916800*** (0.00082863)	-5.94916200*** (0.00074211)	-5.94917400*** (0.00072752)	-5.94916800*** (0.00072555)	-5.94917400*** (0.00072506)	-5.94917700*** (0.00072604)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 3: Forecasting FX Spread: Ten Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+l} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C_t} + controls_t,$$

where  $\ell = 10s$ .

	<i>Spread_D2D<sub>t+1t</sub></i>					
<i>Spread_D2C<sub>t</sub></i>	1.720*** (0.054)	1.722*** (0.055)	1.720*** (0.054)	1.722*** (0.054)	1.720*** (0.054)	1.722*** (0.054)
<i>STD_Spread_D2C<sub>t</sub></i>	-0.612*** (0.020)	-0.613*** (0.021)	-0.612*** (0.020)	-0.613*** (0.021)	-0.612*** (0.020)	-0.613*** (0.021)
<i>D2D_Vol<sub>t</sub></i>		-0.000*** (0.000)		-0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					-0.00000** (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						-0.000*** (0.000)
ProxyLiqShock <sub>t</sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
<i>Note:</i>					*p<0.1; **p<0.05; ***p<0.01	



**Table 4: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 10s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.51084480*** (0.00545883)
Constant	−0.07109310*** (0.00631655)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 5: Reversals in  $\mathcal{A}_{mismatch}$ : 10 seconds**— The table reports estimates for the regression

$$\Delta \mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1 \Delta \mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 10s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta \mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta \mathcal{A}_{mismatch,t,\ell}$	−0.17538010*** (0.00126546)
Constant	−0.00000060*** (0.00000013)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

## C Full Model

### C.1 Model: Replicating the Real World FX Market Structure

This section aims to develop a theoretical model that captures the crucial aspects of the real-world market structure described in the previous section. First, each dealer bank must be able to quote prices to its customers through its own SBP, accounting for the imperfect quote competition with other dealers. Second, each customer must be able to choose a dealer with whom to trade. Third, a customer shall be able to split orders across several dealers. Fourth, in the case of a customer order imbalance, each dealer must be able to offload some of its inventory in the inter-dealer market. Fifth, we need a way of modeling imperfect competition and market power in both segments. Finally, the model should feature dealer heterogeneity to speak to heterogeneity in prices and bid-ask spreads in the data. Below, we outline a model that accounts for all these frictions.

There are five time periods,  $t = -1, 0-, 0, 1, 2$ , and two tradable assets, a risk-free asset with a rate of return normalized to zero and a risky asset with a random payoff  $d$  at time  $t = 2$ . We assume that  $d$  is normally distributed with mean  $\bar{d}$  and variance  $\sigma_d^2$ . The market is populated by  $M$  heterogeneous dealers, indexed by  $l = 1, \dots, M$ , and  $n$  ex-ante identical customers, indexed by  $c = 1, \dots, n$ . All agents start with zero asset inventories. The timeline is as follows:

- At time  $t = -1$ , dealers' inventory shocks  $x_l$ ,  $l = 1, \dots, M$  are realized and are public information.<sup>41</sup> Making such information private is irrelevant for the D2D trading round because of the ex-post nature of the D2D market mechanism. As for the D2C round, the absence of uncertainty about other dealers' holdings is a simplifying assumption. In the real world, dealers have many sources of price-based information, allowing them to make inferences about other dealers' inventory constraints. For example, dealers may observe other dealers' quotes on various platforms. However, modeling price-based inference would drastically complicate the analysis, introducing signaling and belief manipulation aspects into the equilibrium behavior. For this reason, almost all existing models of OTC markets make the same simplifying assumption of publicly observed private types. (See, for example, [Duffie et al. \(2005\)](#), [Schürhoff and Li \(2019\)](#), [Babus and Parlato \(2018\)](#).) We abstract from these effects and leave them for future research. These shocks may originate from previous trading rounds on the dealer-specific SBP.
- At time  $t = 0-$ , customers' *endowment shocks* vector  $\Theta = \{\theta_c\}_{c=1}^n$  is realized; each shock is customers' private information. We assume that these shocks are independent and identically distributed across customers and are drawn from an arbitrary non-degenerate distribution with finite first two moments. Parameters  $\bar{\theta} = E[\theta]$  and  $\sigma_\theta^2 = \text{Var}[\theta]$  are public information. For simplicity, we assume that customers are ex-ante homogeneous and have identical parameters  $\bar{\theta}$ ,  $\sigma_\theta$ . In the real world, customers, might

---

<sup>41</sup>This assumption is made for simplicity.

be heterogeneous across dealers, which may be an important aspect of heterogeneity in dealer behavior. We leave this aspect for future research. As for the case of dealers, we could also interpret  $\theta$  as a private demand/taste shock unrelated to actual asset holdings.

- At the time  $t = 0$ , each dealer trades with customers in the D2C market using the bank's own SBP. Namely, the dealer  $l$  publishes a (dealer-specific) binding price schedule  $p_l(q)$ ,  $l = 1, \dots, M$  to each customer on the SBP, describing the per-unit price at which he is willing to sell  $q$  units of the risky asset. When  $q < 0$ , then  $-p_l(q)$  is the per-unit bid price for many  $-q$  units. Although in real life the price schedules are restricted to step functions (with different prices quoted for orders below USD 1 million and between USD 1 million and USD 5 million), we restrict dealers to use linear schedules instead. That is, we assume  $p_l(q) = \alpha_l + b_l q$ . Such a restriction captures the ability of real-world dealers to quote different prices for different order sizes while preserving analytical tractability.<sup>42</sup>
- We assume that each customer has access to the SBPs of all dealers.<sup>43</sup> Given the quoted price schedules on all the SBPs, each customer  $c$  optimally chooses the vector of quantities  $q_c = (q_l(\theta_c))_{l=1}^M$ , where  $q_l(\theta_c)$  specifies the quantity of the asset acquired from dealer  $l$  by a customer with endowment shock  $\theta_c$ . The total amount paid by customer  $c$  to the dealers is then given by

$$\pi(q_c) = \sum_{l=1}^M q_l(\theta_c) p_l(q_l(\theta_c)).$$

A customer  $c$  ends up holding a total of

$$\bar{q}_c = \theta_c + \sum_{l=1}^M q_l(\theta_c)$$

---

<sup>42</sup>Allowing dealers to optimize in a more general class of functions complicates the analysis significantly. Each dealer selects a mechanism (the price schedule  $p_l(q)$ ), while at the same time competing with other dealers' mechanisms and simultaneously taking into account the fact that all dealers serve as liquidity providers to each other during the second stage of the game (the inter-dealer trading). Such games in competing mechanisms are known to be extremely complex; even with homogeneous traders and a single trading round, optimal price schedules are highly nonlinear, and a symmetric equilibrium often fails to exist. See, for example, [Biais et al. \(2000, 2013\)](#), and [Back and Baruch \(2013\)](#). In our paper, the problem is much more involved because dealers are asymmetric, and there is a second rebalancing stage, introducing another dimension to the strategic interaction. In particular, when posting a price schedule, a dealer  $l$  has to account for the fact that his schedule affects customers' trades with other dealers, which in turn affects other dealers' inventories, thereby affecting the ability of dealer  $l$  to trade with other dealers in the subsequent D2D trading round.

<sup>43</sup>It is a simplifying assumption. If the customer is a regional bank, it may have a subscription to quotes from all major dealers; by contrast, if the customer is a hedge fund, it may decide to subscribe to only a subset of those dealers. Investigating the architecture of the D2C trading network and the endogenous decision of each customer with which dealers connect is an important direction for future research.

units of the asset. Thus, our model features order-splitting by customers. As discussed above, in the real world, order-splitting occurs often for orders above USD 1 million.

- After this D2C trading round on the SBPs, dealer  $l$  receives the vector of orders  $Q_l = (q_l(\theta_c))_{c=1}^n$  and a total cash transfer of

$$\Pi_l(Q_l) \equiv \sum_{c=1}^n p_l(q_l(\theta_c)) q_l(\theta_c) \quad (11)$$

from the customers and ends up holding

$$\chi_l = x_l - \sum_{c=1}^n q_l(\theta_c) \quad (12)$$

units of the asset. The aggregation of the D2C order flow  $Q_l$  into the sum  $\sum_{c=1}^n q_l(\theta_c)$  represents the process of *internalization of order flow* by dealer  $l$ .

- At the time  $t = 1$ , dealers trade in the centralized inter-dealer market to rebalance their inventories. In the real world, D2D platforms function closely to a centralized limit-order market. We capture this fact by assuming that the D2D market operates as the standard uniform-price double auction (see, e.g., Kyle (1989), Vives (2011), Rostek and Weretka (2015), and Malamud and Rostek (2017)). Dealer  $l$  submits a (net) demand schedule  $Q_l(\mathcal{P}^{D2D}) : \mathbb{R} \rightarrow \mathbb{R}$ , which specifies demanded quantity of the asset given its price  $\mathcal{P}^{D2D}$  in the inter-dealer market. All dealers are strategic; in particular, there are no noise traders. As is standard in strategic centralized market models for divisible goods or assets, we study the Nash equilibrium in linear bid schedules (hereafter, *equilibrium*). With divisible goods, equilibrium is invariant to the distribution of independent private uncertainty.<sup>44</sup> We denote by  $Q_l(\chi_l, \mathcal{P}^{D2D})$  the D2D net trade of dealer  $l$  with inventory  $\chi_l$  (see (12)). The latter is given by the total initial inventory plus the total non-internalized customer order flow in the D2C market. Post-D2D trade, the dealer ends up with an inventory of

$$\tilde{\chi}_l = \chi_l + Q_l(\chi_l, \mathcal{P}^{D2D}). \quad (13)$$

---

<sup>44</sup> That is, the linear Bayesian Nash Equilibrium with independent private endowment values has an *ex post* property and coincides with the linear equilibrium that is robust to adding noise in trade (robust Nash Equilibrium; e.g., Vayanos (1999) and Rostek and Weretka (2015)). Equilibrium is *linear* if schedules have the functional form of  $q_l(\cdot) = \alpha_0 + \alpha_{l,q} q_l^0 + \alpha_{l,p} p$ . Strategies are not restricted to linear schedules; rather, it is optimal for a trader to submit a linear demand, given that others do. The approach of analyzing the symmetric linear equilibrium is common in centralized market models (e.g., Kyle (1989), Vayanos (1999), and Vives (2011)). Our analysis does not assume equilibrium symmetry. As equilibrium schedules are optimal even if traders learn the independent value endowments  $q_l^0$  (or equivalently, stochastic marginal utility intercepts,  $\tilde{d} = d - \alpha \Sigma \tilde{q}_l^0$ ) of all other agents, equilibrium is *ex post* Bayesian Nash. The key to the *ex post* property is that permitting pointwise optimization – for each price – equilibrium demand schedules are optimal for any distribution of independent private information and are independent of agents' expectations about others' endowments.

- At time  $t = 2$ , the asset pays off.

We assume that all agents (dealers and customers) incur quadratic costs for holding inventories, equivalent to linearly decreasing marginal values. Importantly, these inventory holding costs are heterogeneous across dealers: Although customers are assumed to be homogeneous, all having the same cost  $\gamma$ , dealers are heterogeneous, with dealer  $l$  incurring cost  $\Gamma_l$ .<sup>45</sup> Thus, customers' total expected utility is given by

$$u(q_c) = -\pi(q_c) + E [d \cdot \bar{q}_c - 0.5\gamma \bar{q}_c^2] . \quad (14)$$

Dealers' total expected utility has two components:

$$U_l = E [\Pi_l(Q_l) + \mathcal{U}_l(\tilde{\chi}_l, \mathcal{P}^{D2D})] , \quad (15)$$

where  $\Pi_l(Q_l)$  is the total transfer (11) received from customers, while

$$\mathcal{U}_l(\tilde{\chi}_l, \mathcal{P}^{D2D}) = d \cdot \tilde{\chi}_l - 0.5\Gamma_l \tilde{\chi}_l^2 - \mathcal{P}^{D2D} Q_l(\chi_l, \mathcal{P}^{D2D}) \quad (16)$$

is dealers' quadratic utility of their post-D2D trade inventory (13) net of the total price  $\mathcal{P}^{D2D} Q_l(\chi_l, \mathcal{P}^{D2D})$  paid for the  $Q_l(\chi_l, \mathcal{P}^{D2D})$  units of the asset in the D2D market.

We follow the standard route used in most of the market microstructure literature and confine our attention to linear equilibria, characterized in the following definition.

**Definition 14** *A linear Nash equilibrium is a collection of the following policies:*

- price schedules  $p_l(q) = \alpha_l + b_l q$  in the D2C market segment,  $l = 1, \dots, M$ ;
- customer demand

$$q(\theta_c) = (q_l(\theta_c))_{l=1}^M, \quad q_l(\theta_c) = \delta_l + \eta_l \theta_c; \text{ and}$$

- dealer demand schedules  $Q_l(\mathcal{P}^{D2D}) = Q_l^{(0)} + Q_l^{(1)} \mathcal{P}^{D2D}$  in the inter-dealer market

such that

- dealer demand schedules form a robust Nash equilibrium in the D2D market;<sup>46</sup>
- customers' demand maximizes customers' utility (14); and

---

<sup>45</sup>The assumption of homogeneous customer compositions across dealers is made for two reasons: First, it simplifies the analysis and, second, we cannot observe customer composition empirically. In the real world, heterogeneity in customer composition might be responsible for a large fraction of variation in the effective inventory cost. We view the inventory cost  $\Gamma$  as a theoretical shortcut for the unobservable sources of heterogeneity. Allowing for dealers' heterogeneity is crucial for matching their highly heterogeneous empirically observed behavior.

<sup>46</sup>As in Rostek and Weretka (2015, p. 2955), the robust Nash equilibrium is in one in which large traders' demands are optimal even after adding full-support uncertain additive noise to their residual demand.

- dealers choose  $\alpha_l$  and  $b_l$  to maximize expected utility (15) given customers' demand functions  $q(\theta_c)$  and provided the equilibrium allocation from the second stage of the game.

Given the definition of equilibrium, we follow the standard backward induction procedure: First, we solve for the unique, robust linear Nash equilibrium of the second stage of the game (the D2D market). Second, we use this equilibrium to calculate dealers' utilities (15). Third, we solve for customers' optimal demand given the linear dealer price schedules. Fourth, we use customers' demand schedules as well as the dealers' utilities from the second trading round to solve for the equilibrium in the liquidity provision game in the D2C market.

## C.2 Linking Dealer Characteristics to D2C Prices and Spreads in the Full Model

This section shows how unobservable dealer characteristics can be linked to prices and spreads in the D2C market. The equilibrium relationship between prices and liquidity in the D2C market and dealers' inventories and risk aversions is generally complex. However, it is possible to derive analytical approximations when dealer heterogeneity is small. Such an approximation allows us to capture the first-order effects of heterogeneity on equilibrium quantities while preserving analytical tractability. Under such approximation, we show that: (i) more risk-averse dealers quote wider spreads, that is,  $b_l^{-1}$  is negatively related to  $\Gamma_l$  (see Proposition 22); and (ii) dealers with higher prices end up holding higher post-D2C inventories, that is,  $\chi_l$  is positively related to  $\alpha_l$  (see Lemma 24). Then,  $Y_{mismatch} \stackrel{s}{=} \mathcal{A}_{mismatch}$ .

Both (i) and (ii) are intuitive. More risk-averse dealers are less efficient at holding inventory and require higher compensation, resulting in wider spreads. Dealers posting the highest prices would attract a disproportionate share of sell volume from customers (since customers would choose to sell to dealers offering the highest price) and would end up with the highest inventories post-D2C.

We thus formulate the analog to Proposition 3, where unobservable dealer characteristics are substituted by observable prices and spreads in the D2C market.

**Proposition 15** *Suppose that  $\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2$  is sufficiently small and that  $n$  is sufficiently large. Then  $Y_{mismatch} \stackrel{s}{=} \mathcal{A}_{mismatch}$ . Suppose further that  $\{\Gamma_l\}_l$  are fixed. Consider three possibilities for the joint distribution of mid-prices and bid-ask spreads in the D2C market: (a)  $\mathcal{A}_{mismatch} > 0$ , (b)  $\mathcal{A}_{mismatch} > 0 = 0$ , and (c)  $\mathcal{A}_{mismatch} < 0$ . Suppose that in all three cases,  $E[\alpha_l]$  is the same. Then,  $\mathcal{P}_{(a)}^{D2D} < \mathcal{P}_{(b)}^{D2D} < \mathcal{P}_{(c)}^{D2D}$ .*

As in Proposition 6, Proposition 15 states that  $Y_{mismatch} \stackrel{s}{=} \mathcal{A}_{mismatch}$  when heterogeneity among dealers is small and when customer order flow is significant. In Proposition 15, the last requirement is captured by the condition that  $n$  must be large enough.

We also derive the following proposition from the fact that bid-ask spreads in the D2C market are positively related to dealers' risk aversion.

**Proposition 16** *An increase in the dispersion of D2C bid-ask spreads  $b_l$  (defined as a mean-preserving spread) is associated with an increase in the liquidity of the D2D market.*

## D Equilibrium in the Full Model

### D.1 Customers' Problem

In this section, we study customers' optimization problems. Given the  $M$  different price schedules offered by the dealers, a customer optimally decides how much to trade with each dealer. Substituting dealers' price schedules into customers' utility (14), we get that customers' optimization problem takes the form of

$$\max_{\{q_{c,l}\}_{l=1}^M} \left\{ - \sum_{l=1}^M q_{c,l} p_l(q_{c,l}) + \left( \sum_{l=1}^M q_{c,l} + \theta_c \right) d - 0.5\gamma \left( \sum_{l=1}^M q_{c,l} + \theta_c \right)^2 \right\},$$

taking dealers' schedules

$$p_l(q_{c,l}) = \alpha_l + \lambda_l^{-1} q_{c,l}$$

as given. (Thus, we denote  $\lambda_l = b_l^{-1}$ ). Define  $a_l = \lambda_l(d - \alpha_l)$  to be the bid-ask spread-normalized deviation from the fundamental value  $d$  of the mid-price quoted by dealer  $l$ . It plays the role of the risk premium in our model: the higher the premium, the lower the mid-price relative to the fundamental value  $d$ . In the sequel, we will therefore refer to this quantity as the "risk premium". Let also  $a_{-l} = \sum_{\ell \neq l} a_\ell$  be the total risk premium of dealers  $\ell \neq l$ , and let  $\lambda_{-l} = \sum_{\ell \neq l} \lambda_\ell$  be the measure of total liquidity provided by these dealers.

Writing down the first-order conditions, we arrive at the following result.

**Lemma 17** *The optimal demand of customer  $j$  is given by  $q_l = \{q_{c,l}\}_{l=1}^M$  with*

$$q_{c,l} = \delta_l(a_l, \lambda_l, \lambda_{-l}, a_{-l}) + \eta_l(a_l, \lambda_l, \lambda_{-l}, a_{-l}) \theta_c \quad (17)$$

with

$$\begin{aligned} \delta_l &= \frac{0.5a_l + 0.25\gamma(a_l\lambda_{-l} - a_{-l}\lambda_l)}{1 + 0.5\gamma(\lambda_l + \lambda_{-l})} \\ \eta_l &= -\frac{0.5\gamma\lambda_l}{1 + 0.5\gamma(\lambda_l + \lambda_{-l})}, \quad l = 1, \dots, M. \end{aligned}$$

The intuition behind the optimal order-splitting strategy of Lemma 17 is as follows: Ideally, the customer would like to buy from the dealer with the lowest mid-quote (equivalently, the highest risk premium  $a_l$ ) and the highest offered liquidity. Thus, his average demand addressed to dealer  $l$  is increasing in  $a_l$  and is decreasing in  $a_{-l}$  (the attractiveness of trading with other dealers). The customer's demand curve is naturally downward sloping in his inventory (that is,  $\eta_l < 0$  for all  $l$ ), and the size of the slope is proportional to the liquidity

$\lambda_l$  offered by dealer  $i$ , as well as to the customer's cost of holding inventory,  $\gamma$ . The stylized linear-quadratic setting of the customer problem makes the analysis tractable and explicit while capturing the key realistic features of real-world demand functions in the D2C segment of FX markets: Quoting higher (lower) mid-prices relative to other dealers attracts sell (buy) volume, whereas dealers quoting wider spreads get less customer volume.

## D.2 Dealers' Optimal Price Schedules and Equilibrium in the D2C Market

At the time  $t = 1$ , dealer  $l$  selects the optimal price schedule  $p_l(q)$  that he quotes to all customers, taking as given other dealers' price schedules  $(a_{-l}, \lambda_{-l})$ , as well as customers' optimal response (Lemma 17). Substituting (5) into (15), we get that dealers' objective is to maximize

$$\begin{aligned} U_l(a_l, \lambda_l; a_{-l}, \lambda_{-l}) = & E \left[ \Pi_l(Q_l) + \chi_l d - 0.5 \Gamma_l \chi_l^2 \right. \\ & \left. + (0.5 \Gamma_l + \beta_l)(\Gamma_l + \beta_l)^{-2} \left( \mathcal{B}^{-1} \sum_{\ell=1}^M (\Gamma_\ell + \beta_\ell)^{-1} \Gamma_\ell \chi_\ell - \Gamma_l \chi_l \right)^2 \right] \end{aligned} \quad (18)$$

over  $a_l, \lambda_l$  subject to (12) and (17). Thus, the dealer's objective function has three components:

- total revenues from trading with customers, as given by

$$\Pi_l(Q_l) = \sum_{c=1}^n p_l(q_{c,l}(a_l, \lambda_l; a_{-l}, \lambda_{-l})) q_{c,l}(a_l, \lambda_l; a_{-l}, \lambda_{-l})$$

- the expected utility  $\chi_l d - 0.5 \Gamma_l \chi_l^2$  from holding the *post-D2C trading round* inventory

$$\chi_l(a_l, \lambda_l; a_{-l}, \lambda_{-l}, x_l) = x_l - \sum_{c=1}^n q_{c,l}(a_l, \lambda_l; a_{-l}, \lambda_{-l}),$$

where  $x_l$  is the dealer's initial inventory.

- the utility surplus from trade in the inter-dealer market

$$(0.5 \Gamma_l + \beta_l)(\Gamma_l + \beta_l)^{-2} \left( \mathcal{B}^{-1} \sum_{\ell=1}^M (\Gamma_\ell + \beta_\ell)^{-1} \Gamma_\ell \chi_\ell - \Gamma_l \chi_l \right)^2. \quad (19)$$

The latter is determined by the deviation of dealer's post-D2C inventory  $\chi_\ell$  from the



liquidity-weighted<sup>47</sup> average of post-D2C inventories of other dealers, as given by

$$\mathcal{B}^{-1} \sum_{\ell=1}^M (\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell} \chi_{\ell}. \quad (20)$$

In particular, dealer  $l$  has incentives to select a quote policy in the D2C market that pushes his inventory  $\chi_l$  and the other dealers' inventory (20) apart.

Importantly, dealers' choice of the price schedule characteristics  $a_l, \lambda_l$  influence all three components, including *the gains from trades made in the D2D market*. The latter effect is particularly subtle and itself consists of two sub-components: the dealer's impact on equilibrium price and the impact of the dealer's liquidity provision in the D2C market on the inventories of all other dealers. Indeed, the more liquidity the dealer provides in the D2C market, the fewer clients will trade with other dealers, directly influencing other dealers' inventories  $\chi_{\ell}$ ,  $\ell \neq l$ , and, hence, also influencing the surplus from trade in (19) through (20).

Let

$$\Psi = \left( \frac{\Gamma_1}{\mathcal{B}(\Gamma_1 + \beta_1)}, \dots, \frac{\Gamma_M}{\mathcal{B}(\Gamma_M + \beta_M)} \right)^T$$

be the vector of weights that define the “aggregate risk” in the D2D market, and let also  $\Psi_l \equiv \Psi - \Gamma_l \mathbf{1}_{\ell=l}$ . Denote also  $\boldsymbol{\delta} \equiv (\delta_l)_{l=1}^M$ ,  $\boldsymbol{\eta} \equiv (\eta_l)_{l=1}^M$  to be the vectors of coefficients of customers' demand (see (17)), and let  $x = (x_{\ell})_{\ell=1}^M$  be the vector of dealer initial inventories. Recall that we use  $\bar{\theta}$  and  $\sigma_{\theta}^2$  to denote expected customer endowment and the variance of customer endowments, respectively. Evaluating the expectation in (18), we get the following expression for dealers' indirect utility.

**Lemma 18** *We have*

$$\begin{aligned} U_l(\alpha_l, \lambda_l; \alpha_{-l}, \lambda_{-l}) &= x_l d - n \lambda_l^{-1} a_l (\delta_l + \eta_l \bar{\theta}) + n \lambda_l^{-1} \left[ (\delta_l + \eta_l \bar{\theta})^2 + \eta_l^2 \sigma_{\theta}^2 \right] \\ &\quad - 0.5 \Gamma_l \left[ (x_l - n \delta_l - n \bar{\theta} \eta_l)^2 + n \sigma_{\theta}^2 \eta_l^2 \right] \\ &\quad + (0.5 \Gamma_l + \beta_l) (\Gamma_l + \beta_l)^{-2} \left[ \left( \Psi_l \cdot (x - n \boldsymbol{\delta} - n \bar{\theta} \boldsymbol{\eta}) \right)^2 + n \sigma_{\theta}^2 (\Psi_l \cdot \boldsymbol{\eta})^2 \right] \end{aligned}$$

and a Nash equilibrium in linear price schedules is a collection of  $(\alpha_l, \lambda_l)_{l=1}^M$  such that, for all  $\ell$ ,

$$(\alpha_l, \lambda_l) = \arg \max_{\alpha_l, \lambda_l} U_l(\alpha_l, \lambda_l; \alpha_{-l}, \lambda_{-l}). \quad (21)$$

The first term,  $x_l d$ , is just the expected payoff from the initial inventory of the dealer. The second term,  $n \lambda_l^{-1} a_l (\delta_l + \eta_l \bar{\theta})$ , is the amount the dealer is losing (relative to the asset payoff

---

<sup>47</sup>The weights are inversely related to price impacts  $\beta_{\ell}$  in the D2D market.

d) from the trades in the D2C market: By quoting the price at a premium of  $a_l$ , he is selling the asset “too cheap” and his losses are proportional to the number  $n$  of customers. The third term,  $n\lambda_l^{-1}[(\delta_l + \eta_l\bar{\theta})^2 + \eta_l^2\sigma_\theta^2]$  represents the rents the dealer extracts from the customer using the increasing price schedule. These rents are proportional to the slope  $\lambda_l^{-1}$ : The higher the slope, the more the dealer charges for liquidity provision. The fourth term is just the expected quadratic inventory cost,

$$0.5\Gamma_l[(x_l - n\delta_l - n\bar{\theta}\eta_l)^2 + n\sigma_\theta^2\eta_l^2] = 0.5\Gamma_l\chi_l^2.$$

Finally, the last term is the expected utility gain from trading in the D2D market. Importantly, contrary to the quadratic inventory cost term, the other quadratic terms are convex. This is a novel aspect of our model: Dealer market power in both segments introduces incentives for risk-taking whereby amplifying volatility of demand shocks also amplifies the rents the dealer is able to extract in the D2C and the D2D market.

In the Appendix, we write down the system of first conditions for (21). In general, this system cannot be solved explicitly. However, it is possible to derive analytical approximations to its solution when dealer heterogeneity is small. Such an approximation allows us to capture the first-order effects of heterogeneity on equilibrium quantities while preserving analytical tractability.

### D.3 The Joint D2C-D2D Equilibrium

Everywhere in the sequel, we assume that dealers’ heterogeneity is small. This technical assumption will allow us to derive approximate closed-form expressions for equilibrium objects. We start with the case when dealers do not hold any inventory and risk aversions are homogeneous,  $\Gamma_l = \Gamma$ . Then, the following is true.

**Proposition 19** *If  $x_l = 0$  (zero initial dealer inventories) and  $\Gamma_l = \Gamma$  for all  $l = 1, \dots, M$ , and  $\bar{\theta} = 0$  (zero expected customer endowment), then there exists a unique symmetric equilibrium, and the inverse of the price function slope is given by*

$$\lambda^*(\Gamma) = \frac{(M-2)\gamma - 2\Gamma + \sqrt{((M-2)\gamma - 2\Gamma)^2 + 8\gamma\Gamma(M-1)}}{2\gamma\Gamma(M-1)},$$

while  $a_l = 0$  for all  $l$ .  $\lambda^*(\Gamma)$  is decreasing in both  $\Gamma$  and  $\gamma$  and is increasing in  $M$ .

The result of Proposition 19 is very intuitive: Liquidity in the D2C market, as captured by the inverse slope  $\lambda^*$ , is determined by two forces: the willingness (and the ability) of dealers to take on risk (that is, their cost of holding inventory,  $\Gamma$ ) and the dealers’ market power, determined by their number,  $M$ . When  $\Gamma$  increases, or when clients are more aggressively trying to get rid of their inventory (that is,  $\gamma$  is large), dealers optimally widen the spread. At the same time, an increase in  $M$  creates competitive pressure on equilibrium spreads, driving D2C market liquidity up. In particular, in the competitive limit, as  $M \rightarrow \infty$ , we

have  $\lambda^* \rightarrow 1/\Gamma$ , consistent with the standard competitive CAPM, whereby price sensitivity to inventory shocks equals the reciprocal of the risk aversion.

Having solved for the equilibrium in the limiting case  $x_l = 0$  and  $\Gamma_l = \Gamma$ , we can now use Taylor approximation to compute the equilibrium for the case when initial (prior to D2C trading) dealer inventories ( $x_l$ ), mean customer endowments  $\bar{\theta}$ , and heterogeneity in  $\Gamma_l$  are all sufficiently small. Let  $\bar{\alpha} = M^{-1} \sum_l \alpha_l$  be the average mid quote, and recall the definition of the observable price-based liquidity mismatch,  $\mathcal{A}_{mismatch}$ , in the D2C market. The following Proposition characterizes the equilibrium link between prices in the two market segments

**Proposition 20** *The price  $P = P^{D2D}$  in the D2D market is linked to prices in the D2C market via*

$$\begin{aligned} \mathcal{P}^{D2D} = & \left[ 1 + \frac{2\gamma\lambda^*}{(2 + M\gamma\lambda^*)[2 + (M-1)\gamma\lambda^*]} \right] (\bar{\alpha} - \mathcal{A}_{mismatch}) \\ & - (\lambda^*)^2 \left[ \Gamma^*(M\varphi_{0,\lambda}^A + \varphi_1^a) + \frac{M\varphi^\lambda(\varphi_0^a - 0.5n)}{M^2 - 2M + 2} \right] \mathcal{A}_{mismatch} \\ & - \frac{2\gamma\lambda^*}{(2 + M\gamma\lambda^*)[2 + (M-1)\gamma\lambda^*]} d - \Gamma^* \varphi_0^\theta \bar{\theta} - \gamma \frac{1}{2 + \gamma\Lambda} \lambda^* \Gamma^* \Theta \\ & + O(\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2) \end{aligned}$$

Proposition 20 provides an explicit expression linking the price in the D2D market to:

- mean and *distribution of quotes* in the D2C market, as captured by  $\bar{\alpha}$  and  $\alpha_{mismatch}$ ;
- expected client endowment,  $\bar{\theta}$ , as well as the aggregate client endowment shock,  $\Theta$ ;<sup>48</sup>
- the fundamental,  $d$ .<sup>49</sup>

The most important consequence of Proposition 20 is that, in the presence of heterogeneity, the joint cross-sectional distribution of quotes and bid-ask spreads is directly linked to the price level in the D2D market. Namely, in addition to the average mid-quote  $\bar{\alpha}$ , the price also depends on  $\mathcal{A}_{mismatch}$  which is effectively a spread between low bid-ask spread (large  $\lambda_l$ ) mid-quotes, and high bid-ask spread (low  $\lambda_l$ ) quotes. As we explain in Section 3.1, the term originates from (i) strategic interactions between dealers in the D2D market; (ii) dealers' bid shading in the D2C market; and (iii) non-exclusive competition and customers' optimal order splitting across dealers in the D2C market.

---

<sup>48</sup>Thus,  $E[\Theta] = n\bar{\theta}$ .

<sup>49</sup>Note that, although the coefficient on the fundamental value,  $d$ , is negative, the price  $\mathcal{P}^{D2D}$  depends positively on  $d$  (at it should) due to a positive link between  $\alpha_l$  and  $d$ .

# Internet Appendix for “Dealer Heterogeneity and Exchange Rates”

## E Additional Regression Tables

### E.1 Additional Lags: Prices

**Table 6: Forecasting FX rates: One Second Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 1s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	0.03873706*** (0.00150778)	0.03873713*** (0.00149743)	0.03873832*** (0.00149056)	0.03873685*** (0.00147956)	0.03873748*** (0.00147841)	0.03873737*** (0.00148198)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-0.56428000*** (0.09301252)	-0.56422310*** (0.09263100)	-0.56407690*** (0.09239439)	-0.56404540*** (0.09204492)	-0.56409150*** (0.09201776)	-0.56406020*** (0.09212495)
ProxyLiqShock <sub>t</sub>		-0.00009274*** (0.00001961)		-0.00009318*** (0.00001959)	-0.00009184*** (0.00001955)	-0.00009253*** (0.00001959)
$D2D.Vol_t$			0.00001781** (0.00000786)	0.00001841** (0.00000784)		
$D2D.Vol.1h_t$					0.00002316** (0.00001073)	
$D2D.Vol.24h_t$						0.00001853** (0.00000828)
Constant	-1.71384700*** (0.00000773)	-1.71384700*** (0.00000772)	-1.71384700*** (0.00000772)	-1.71384700*** (0.00000772)	-1.71384600*** (0.00000771)	-1.71384700*** (0.00000772)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 7: Forecasting FX rates: Three Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 3s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

	<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	0.15443070*** (0.00275667)	0.15448450*** (0.00269613)	0.15442730*** (0.00266416)	0.15448870*** (0.00260009)	0.15448250*** (0.00258719)	0.15449370*** (0.00260993)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-6.78784800*** (0.35444290)	-6.79415300*** (0.35270430)	-6.78695500*** (0.35152090)	-6.79361600*** (0.34968290)	-6.79418600*** (0.34926670)	-6.79402600*** (0.35003480)
ProxyLiqShock <sub>t</sub>		-0.00151414*** (0.00019632)		-0.00152120*** (0.00019598)	-0.00151069*** (0.00019555)	-0.00151283*** (0.00019598)
$D2D.Vol_t$			0.00023801*** (0.00006697)	0.00024877*** (0.00006786)		
$D2D.Vol_{1h_t}$					0.00008968 (0.00009398)	
$D2D.Vol_{24h_t}$						0.00019939*** (0.00006881)
Constant	-3.00424400*** (0.00006610)	-3.00424300*** (0.00006681)	-3.00424400*** (0.00006609)	-3.00424400*** (0.00006678)	-3.00424000*** (0.00006677)	-3.00424500*** (0.00006680)
<i>Note:</i>					*p<0.1; **p<0.05; ***p<0.01	

**Table 8: Forecasting FX rates: Five Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 5s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

	<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	0.07769800*** (0.00083556)	0.07774450*** (0.00078643)	0.07770872*** (0.00075440)	0.07775902*** (0.00070019)	0.07774352*** (0.00068278)	0.07775173*** (0.00071012)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-2.70442700*** (0.23356760)	-2.70748000*** (0.23231550)	-2.70419400*** (0.23101990)	-2.70734300*** (0.22951550)	-2.70741500*** (0.22896130)	-2.70739900*** (0.22978560)
ProxyLiqShock <sub>t</sub>		-0.00120330*** (0.00018013)		-0.00120895*** (0.00017970)	-0.00120421*** (0.00017920)	-0.00120235*** (0.00017971)
<i>D2D.Vol<sub>t</sub></i>			0.00019518*** (0.00006373)	0.00020319*** (0.00006463)		
<i>D2D.Vol.1h<sub>t</sub></i>					-0.00002541 (0.00009007)	
<i>D2D.Vol.24h<sub>t</sub></i>						0.00013867** (0.00006451)
Constant	-4.00277800*** (0.00006123)	-4.00277700*** (0.00006205)	-4.00277800*** (0.00006166)	-4.00277800*** (0.00006243)	-4.00277800*** (0.00006262)	-4.00277900*** (0.00006239)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 9: Forecasting FX rates: Fifteen Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 15s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

	<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1.16306400*** (0.01064314)	1.16934900*** (0.01107594)	1.16308500*** (0.01125033)	1.16871400*** (0.01144837)	1.16901000*** (0.01153718)	1.16898700*** (0.01143152)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-61.66987000*** (8.27103100)	-62.19161000*** (8.39204400)	-61.66666000*** (8.38484100)	-62.15246000*** (8.45649900)	-62.17151000*** (8.47251200)	-62.17090000*** (8.45500600)
ProxyLiqShock <sub>t</sub>		-0.04617202*** (0.00915775)		-0.04621309*** (0.00906875)	-0.04612746*** (0.00905460)	-0.04611698*** (0.00907106)
$D2D\_Vol_t$			0.00220351 (0.00316746)	0.00250521 (0.00318910)		
$D2D\_Vol\_1h_t$					0.00108077 (0.00490193)	
$D2D\_Vol\_24h_t$						0.00467389 (0.00316280)
Constant	-7.24852800*** (0.00317889)	-7.24852900*** (0.00314778)	-7.24853700*** (0.00308585)	-7.24853600*** (0.00310774)	-7.24848800*** (0.00311596)	-7.24859000*** (0.00311092)
<i>Note:</i>						*p<0.1; **p<0.05; ***p<0.01

**Table 10: Forecasting FX rates: Twenty Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 20s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	55.27935000*** (0.64341760)	55.95169000*** (0.66535810)	55.35582000*** (0.67244350)	55.95703000*** (0.68518880)	55.96041000*** (0.68683830)	56.07694000*** (0.68614430)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-2,568.69100000*** (511.42220000)	-2,616.21800000*** (523.08240000)	-2,571.99300000*** (521.02910000)	-2,616.19000000*** (528.29340000)	-2,616.50000000*** (528.65020000)	-2,622.50100000*** (529.35510000)
ProxyLiqShock <sub>t</sub>		-2.83902100*** (0.58742300)		-2.84338600*** (0.58306350)	-2.83783900*** (0.58271080)	-2.84296900*** (0.58427560)
$D2D\_Vol_t$			0.13538540 (0.19575270)	0.15440600 (0.19862270)		
$D2D\_Vol\_1h_t$					0.06463395 (0.32186150)	
$D2D\_Vol\_24h_t$						0.28245960 (0.19635790)
Constant	-8.49130500*** (0.19717880)	-8.49206700*** (0.19740960)	-8.49198900*** (0.19191970)	-8.49244100*** (0.19469150)	-8.48947800*** (0.19596210)	-8.49601700*** (0.19523400)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01



**Table 11: Forecasting FX rates: Thirty Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 30s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

	<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>					
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1,566.5700000*** (17.69701000)	1,565.37900000*** (17.78775000)	1,566.51100000*** (17.91811000)	1,565.33200000*** (17.95429000)	1,565.35200000*** (17.96782000)	1,565.34700000*** (17.95716000)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-64,260.21000000*** (20,827.76000000)	-65,350.25000000*** (20,947.31000000)	-64,261.95000000*** (21,091.92000000)	-65,354.01000000*** (21,104.00000000)	-65,349.77000000*** (21,112.65000000)	-65,341.64000000*** (21,106.79000000)
ProxyLiqShock <sub>t</sub>		-118.77080000*** (22.55573000)		-118.93810000*** (22.56407000)	-118.73120000*** (22.55518000)	-118.68820000*** (22.55972000)
$D2D\_Vol_t$			6.09856300 (8.30987900)	6.84660500 (8.38996000)		
$D2D\_Vol\_1h_t$					0.97244140 (12.74730000)	
$D2D\_Vol\_24h_t$						7.60809900 (8.26973200)
Constant	-13.33470000 (8.60622500)	-13.09005000 (8.53245700)	-13.37672000 (8.30059400)	-13.13668000 (8.36590300)	-13.06131000 (8.39622200)	-13.20670000 (8.36528100)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 12: Forecasting FX rates: Forty Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 40s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1,589.64100000*** (24.23399000)	1,589.24200000*** (24.11451000)	1,589.62400000*** (24.10898000)	1,589.20100000*** (24.09673000)	1,589.36500000*** (24.09822000)	1,589.20200000*** (24.09093000)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-46,788.79000000 (28,980.55000000)	-48,816.19000000* (28,953.86000000)	-46,821.91000000 (28,955.54000000)	-48,858.07000000* (28,890.47000000)	-48,847.99000000* (28,911.21000000)	-48,808.76000000* (28,913.72000000)
ProxyLiqShock <sub>t</sub>		-170.78310000*** (30.38525000)		-171.02140000*** (30.99576000)	-170.96800000*** (30.87123000)	-170.72770000*** (30.85773000)
$D2D.Vol_t$			9.01325500 (12.39544000)	10.08355000 (12.70833000)		
$D2D.Vol\_1h_t$					-4.35846800 (17.88818000)	
$D2D.Vol\_24h_t$						4.52796600 (12.62489000)
Constant	-9.39330200 (11.86934000)	-9.07833600 (12.18556000)	-9.45775800 (12.31642000)	-9.15001100 (12.60861000)	-9.19266600 (12.55940000)	-9.14831800 (12.52307000)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 13: Forecasting FX rates: Forty Five Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 45s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1,564.86300000*** (25.94795000)	1,564.43900000*** (25.97072000)	1,564.85300000*** (25.97618000)	1,564.40800000*** (26.29842000)	1,564.51400000*** (26.14544000)	1,564.42000000*** (26.12031000)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-29,096.55000000 (33,040.74000000)	-31,747.72000000 (33,008.52000000)	-29,115.59000000 (32,937.88000000)	-31,772.62000000 (32,349.44000000)	-31,854.48000000 (32,594.45000000)	-31,737.83000000 (32,621.40000000)
ProxyLiqShock <sub>t</sub>		-190.63380000*** (34.74468000)		-190.92880000*** (37.19062000)	-191.12930000*** (36.34754000)	-190.59840000*** (36.25084000)
$D2D.Vol_t$			10.49881000 (14.84299000)	11.79139000 (16.06530000)		
$D2D.Vol\_1h_t$					-7.95174300 (21.38842000)	
$D2D.Vol\_24h_t$						3.38251000 (15.55294000)
Constant	-6.84904300 (13.83944000)	-6.56124700 (14.42244000)	-6.92537100 (14.78349000)	-6.64890000 (15.99845000)	-6.75173800 (15.50183000)	-6.61365700 (15.41197000)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 14: Forecasting FX rates: Fifty Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 50s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1,492.57600000*** (27.65986000)	1,492.79700000*** (27.92664000)	1,492.56500000*** (28.19467000)	1,492.77800000*** (28.77858000)	1,493.28100000*** (29.85946000)	1,492.79000000*** (29.63336000)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	-19,421.15000000 (36,216.16000000)	-22,605.30000000 (35,710.05000000)	-19,417.72000000 (35,344.00000000)	-22,607.35000000 (34,692.74000000)	-22,934.81000000 (33,526.69000000)	-22,599.22000000 (33,732.68000000)
ProxyLiqShock <sub>t</sub>		-200.95030000*** (39.17355000)		-201.31230000*** (42.66249000)	-201.90000000*** (46.73086000)	-200.93340000*** (46.00511000)
$D2D.Vol_t$			11.58196000 (17.99090000)	13.09564000 (19.39545000)		
$D2D.Vol_{1h_t}$					-12.62028000 (29.69561000)	
$D2D.Vol_{24h_t}$						1.43107800 (21.94732000)
Constant	-4.75412200 (15.73834000)	-4.49671600 (17.04564000)	-4.83464300 (17.94762000)	-4.58935800 (19.38305000)	-4.77335500 (22.36620000)	-4.51637500 (21.79624000)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 15: Forecasting FX rates: Sixty Seconds Ahead.**— The table reports estimates for the regression

$$\Delta p_{t+\ell, \ell}^{D2D} = a_0 + a_1 \Delta \bar{\alpha}_{t, \ell} + a_2 \Delta \mathcal{A}_{mismatch, t, \ell} + controls_t,$$

where  $\Delta X_{t, \ell} = X_t - X_{t-\ell}$  and  $\ell = 60s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>10^4 \times \Delta \mathcal{P}_{t+\ell, \ell}^{D2D}</math></i>						
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta \bar{\alpha}_{t, \ell}$	1,312.67100000*** (32.41165000)	1,314.95000000*** (31.22029000)	1,312.68500000*** (31.05948000)	1,314.96800000*** (30.95963000)	1,315.48800000*** (31.05480000)	1,314.95100000*** (30.97912000)
$\Delta \mathcal{A}_{mismatch, t, \ell}$	15,553.21000000 (38,385.91000000)	11,127.53000000 (39,984.52000000)	15,586.30000000 (40,174.35000000)	11,153.72000000 (40,446.33000000)	10,948.05000000 (40,290.41000000)	11,125.51000000 (40,407.15000000)
ProxyLiqShock <sub>t</sub>		−220.35690000*** (43.49203000)		−220.78200000*** (42.04641000)	−221.07010000*** (42.47202000)	−220.35710000*** (42.14161000)
<i>D2D.Vol<sub>t</sub></i>			14.28575000 (19.13201000)	15.88489000 (18.48285000)		
<i>D2D.Vol.1h<sub>t</sub></i>					−15.49562000 (25.57480000)	
<i>D2D.Vol.24h<sub>t</sub></i>						−0.16187510 (18.74520000)
Constant	−4.10795700 (23.08267000)	−3.80036300 (19.47854000)	−4.20619900 (19.10517000)	−3.91228900 (18.48088000)	−4.14984500 (18.84784000)	−3.79718700 (18.57503000)

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

## E.2 Additional Lags: Spread

**Table 16: Forecasting FX Spread: One Second Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 1s$ .

<i>Spread D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.791*** (0.049)	1.793*** (0.049)	1.791*** (0.049)	1.793*** (0.049)	1.791*** (0.049)	1.793*** (0.049)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.637*** (0.018)	−0.638*** (0.018)	−0.637*** (0.018)	−0.638*** (0.018)	−0.637*** (0.018)	−0.638*** (0.018)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					−0.00000* (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
ProxyLiqShock <sub>t</sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

**Table 17: Forecasting FX Spread: Three Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 3s$ .

<i>Spread_D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.769*** (0.052)	1.770*** (0.052)	1.769*** (0.052)	1.770*** (0.052)	1.769*** (0.052)	1.770*** (0.052)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.629*** (0.019)	−0.630*** (0.020)	−0.629*** (0.019)	−0.630*** (0.020)	−0.629*** (0.019)	−0.630*** (0.020)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					−0.00000** (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
<i>ProxyLiqShock<sub>t</sub></i>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 18: Forecasting FX Spread: Five Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 5s$ .

<i>Spread_D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.753*** (0.051)	1.754*** (0.054)	1.753*** (0.054)	1.754*** (0.054)	1.753*** (0.054)	1.755*** (0.054)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.624*** (0.019)	−0.624*** (0.020)	−0.624*** (0.020)	−0.624*** (0.020)	−0.624*** (0.020)	−0.625*** (0.020)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					−0.00000* (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
ProxyLiqShock <sub>t</sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
<i>Note:</i>					*p<0.1; **p<0.05; ***p<0.01	



**Table 19: Forecasting FX Spread: Fifteen Seconds Ahead.**— The table reports estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C_t} + controls_t,$$

where  $\ell = 15s$ .

	<i>Spread</i> <i>D2D</i> <sub><i>t</i>+<i>ℓ</i></sub>					
<i>Spread.D2C</i> <sub><i>t</i></sub>	1.699*** (0.054)	1.701*** (0.055)	1.699*** (0.054)	1.701*** (0.054)	1.699*** (0.054)	1.701*** (0.054)
<i>STD.Spread.D2C</i> <sub><i>t</i></sub>	−0.605*** (0.019)	−0.605*** (0.020)	−0.605*** (0.019)	−0.605*** (0.020)	−0.605*** (0.019)	−0.605*** (0.020)
<i>D2D.Vol</i> <sub><i>t</i></sub>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D.Vol.1h</i> <sub><i>t</i></sub>					−0.00000** (0.00000)	
<i>D2D.Vol.24h</i> <sub><i>t</i></sub>						−0.000*** (0.000)
ProxyLiqShock <sub><i>t</i></sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

**Table 20: Forecasting FX Spread: Twenty Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 20s$ .

<i>Spread_D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.683*** (0.056)	1.685*** (0.056)	1.683*** (0.055)	1.685*** (0.056)	1.683*** (0.055)	1.686*** (0.055)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.601*** (0.027)	−0.602*** (0.027)	−0.601*** (0.026)	−0.602*** (0.027)	−0.601*** (0.026)	−0.602*** (0.027)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					−0.00000** (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
<i>ProxyLiqShock<sub>t</sub></i>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)	0.00004*** (0.00000)

*Note:* \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 21: Forecasting FX Spread: Fifteen Seconds Ahead.**— The table reports estimates for the regression

$$Spread\_D2D_{t+l} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C_t} + controls_t,$$

where  $\ell = 30s$ .

	<i>Spread</i> <i>D2D</i> <sub><i>t</i> + <i>ℓ</i></sub>					
<i>Spread.D2C<sub>t</sub></i>	1.644*** (0.053)	1.646*** (0.053)	1.644*** (0.053)	1.646*** (0.053)	1.644*** (0.053)	1.646*** (0.052)
<i>STD.Spread.D2C<sub>t</sub></i>	−0.585*** (0.020)	−0.586*** (0.020)	−0.585*** (0.020)	−0.586*** (0.020)	−0.585*** (0.020)	−0.586*** (0.020)
<i>D2D.Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D.Vol_1h<sub>t</sub></i>					−0.000 (0.00000)	
<i>D2D.Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
ProxyLiqShock <sub><i>t</i></sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00005*** (0.00000)	0.00004*** (0.00000)	0.00005*** (0.00000)	0.00004*** (0.00000)	0.00005*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

**Table 22: Forecasting FX Spread: Forty Seconds Ahead.**— The table reports estimates for the regression

$$Spread\_D2D_{t+l} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C_t} + controls_t,$$

where  $\ell = 40s$ .

	<i>Spread.D2D<sub>t+t</sub></i>					
<i>Spread.D2C<sub>t</sub></i>	1.632*** (0.052)	1.634*** (0.053)	1.632*** (0.052)	1.634*** (0.053)	1.632*** (0.052)	1.634*** (0.052)
<i>STD.Spread.D2C<sub>t</sub></i>	-0.581*** (0.020)	-0.582*** (0.020)	-0.581*** (0.020)	-0.582*** (0.020)	-0.581*** (0.020)	-0.582*** (0.020)
<i>D2D.Vol<sub>t</sub></i>		-0.000*** (0.000)		-0.000*** (0.000)		
<i>D2D.Vol_1h<sub>t</sub></i>					-0.00000 (0.00000)	
<i>D2D.Vol_24h<sub>t</sub></i>						-0.000*** (0.000)
ProxyLiqShock <sub>t</sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00004*** (0.00000)	0.00005*** (0.00000)	0.00004*** (0.00000)	0.00005*** (0.00000)	0.00004*** (0.00000)	0.00005*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

**Table 23: Forecasting FX Spread: Forty Five Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 45s$ .

<i>Spread_D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.618*** (0.054)	1.621*** (0.057)	1.618*** (0.054)	1.621*** (0.056)	1.618*** (0.054)	1.621*** (0.357)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.576*** (0.021)	−0.578*** (0.034)	−0.576*** (0.021)	−0.578*** (0.033)	−0.576*** (0.021)	−0.578 (0.613)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					−0.00000 (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
<i>ProxyLiqShock<sub>t</sub></i>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)

*Note:*

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

**Table 24: Forecasting FX Spread: Fifty Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+l} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C_t} + controls_t,$$

where  $\ell = 50s$ .

	<i>Spread D2D<sub>t+t</sub></i>					
<i>Spread.D2C<sub>t</sub></i>	1.621*** (0.052)	1.623*** (0.054)	1.621*** (0.053)	1.623*** (0.053)	1.621*** (0.053)	1.623*** (0.052)
<i>STD.Spread.D2C<sub>t</sub></i>	-0.578*** (0.019)	-0.578*** (0.020)	-0.578*** (0.019)	-0.578*** (0.019)	-0.578*** (0.019)	-0.579*** (0.019)
<i>D2D.Vol<sub>t</sub></i>		-0.000*** (0.000)		-0.000*** (0.000)		
<i>D2D.Vol_1h<sub>t</sub></i>					0.000 (0.00000)	
<i>D2D.Vol_24h<sub>t</sub></i>						-0.000*** (0.000)
ProxyLiqShock <sub>t</sub>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

**Table 25: Forecasting FX Spread: Sixty Seconds Ahead.**— The table report estimates for the regression

$$Spread\_D2D_{t+\ell} = a_0 + a_1 STD\_Spread\_D2C_t + a_2 \overline{Spread\_D2C}_t + controls_t,$$

where  $\ell = 60s$ .

<i>Spread_D2D<sub>t+ℓ</sub></i>						
<i>Spread_D2C<sub>t</sub></i>	1.605*** (0.053)	1.607*** (0.054)	1.605*** (0.053)	1.607*** (0.054)	1.605*** (0.053)	1.607*** (0.053)
<i>STD_Spread_D2C<sub>t</sub></i>	−0.571*** (0.019)	−0.572*** (0.019)	−0.571*** (0.019)	−0.572*** (0.019)	−0.571*** (0.019)	−0.572*** (0.019)
<i>D2D_Vol<sub>t</sub></i>		−0.000*** (0.000)		−0.000*** (0.000)		
<i>D2D_Vol_1h<sub>t</sub></i>					0.00000 (0.00000)	
<i>D2D_Vol_24h<sub>t</sub></i>						−0.000*** (0.000)
<i>ProxyLiqShock<sub>t</sub></i>			0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Constant	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)	0.00005*** (0.00000)
<i>Note:</i>				*p<0.1; **p<0.05; ***p<0.01		

### E.3 Additional lags: $\mathcal{A}_{mismatch}$

**Table 26: Reversals in  $\mathcal{A}_{mismatch}$ : 1 second**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 1s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.11960410*** (0.00145343)
Constant	−0.00000010** (0.00000004)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	



**Table 27: Reversals in  $\mathcal{A}_{mismatch}$ : 3 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 3s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.16057820*** (0.00138495)
Constant	−0.00000044*** (0.00000007)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 28: Reversals in  $\mathcal{A}_{mismatch}$ : 5 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 5s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.17032590*** (0.00130624)
Constant	−0.00000051*** (0.00000009)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 29: Reversals in  $\mathcal{A}_{mismatch}$ : 15 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 15s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.18127190*** (0.00122993)
Constant	−0.00000075*** (0.00000023)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 30: Reversals in  $\mathcal{A}_{mismatch}$ : 20 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 20s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.19437510*** (0.00161862)
Constant	−0.00000080*** (0.00000018)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 31: Reversals in  $\mathcal{A}_{mismatch}$ : 30 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 30s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.22155150*** (0.00185767)
Constant	−0.00000086*** (0.00000030)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 32: Reversals in  $\mathcal{A}_{mismatch}$ : 40 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 40s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.23728540*** (0.00228447)
Constant	−0.00000099*** (0.00000026)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 33: Reversals in  $\mathcal{A}_{mismatch}$ : 45 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 45s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.24642560*** (0.00246791)
Constant	−0.00000112*** (0.00000027)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 34: Reversals in  $\mathcal{A}_{mismatch}$ : 50 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 50s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.25461630*** (0.00262684)
Constant	−0.00000119*** (0.00000028)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 35: Reversals in  $\mathcal{A}_{mismatch}$ : 60 seconds**— The table reports estimates for the regression

$$\Delta\mathcal{A}_{mismatch,t+\ell,\ell} = a_0 + a_1\Delta\mathcal{A}_{mismatch,t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 60s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta\mathcal{A}_{mismatch,t+\ell,\ell}</math></i>	
$\Delta\mathcal{A}_{mismatch,t,\ell}$	−0.27091950*** (0.00293203)
Constant	−0.00000135*** (0.00000030)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

#### E.4 Additional lags: price competitiveness

**Table 36: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 1s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.54754750*** (0.00512132)
Constant	−0.04765977*** (0.00626592)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 37: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 3s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.50713180*** (0.00549015)
Constant	−0.07107528*** (0.00631872)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 38: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 5s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.50531160*** (0.00517648)
Constant	−0.06504597*** (0.00636469)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 39: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 15s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.55226020*** (0.00506787)
Constant	−0.04755946*** (0.00626638)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 40: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 20s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.50487600*** (0.00493113)
Constant	−0.06732368*** (0.00627096)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 41: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 30s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.45087080*** (0.00493493)
Constant	−0.05297284*** (0.00630028)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	



**Table 42: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 40s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.50649200*** (0.00503544)
Constant	−0.06503918*** (0.00636303)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 43: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 45s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.51483110*** (0.00534165)
Constant	−0.07110893*** (0.00631328)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 44: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 50s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.55169720*** (0.00498781)
Constant	−0.04757562*** (0.00627074)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

**Table 45: Reversals in price competitiveness  $pc$** — The table reports estimates for the regression

$$\Delta pc_{t+\ell,\ell} = a_0 + a_1 \Delta pc_{t,\ell}$$

where  $\Delta X_{t,\ell} = X_t - X_{t-\ell}$  and  $\ell = 60s$ . Heteroscedasticity and autocorrelation robust standard errors are shown in parenthesis.

<i>Dependent variable: <math>\Delta pc_{t+\ell,\ell}</math></i>	
$\Delta pc_{t,\ell}$	−0.52398920*** (0.00506093)
Constant	−0.06499731*** (0.00654184)
<i>Note:</i> *p<0.1; **p<0.05; ***p<0.01	

## F Additional Proofs

### F.1 Proof of Proposition 15

**Proof of Proposition 15.** Suppose that  $\mathcal{A}_{mismatch} \stackrel{s}{=} Y_{mismatch}$ .

Starting with Equation 6, we have

$$\mathcal{P}^{D2D} = d - \mathcal{B}^{-1} M E[(\Gamma_l + \beta_l)^{-1} \Gamma_l] E[\chi_l] - \mathcal{B}^{-1} M \text{Cov}((\Gamma_l + \beta_l)^{-1} \Gamma_l, \chi_l).$$

The first two terms in the R.H.S. of the equation below remain the same if the three cases (a), (b), and (c). Focusing on the last term, we have

$$-\mathcal{B}^{-1} M \text{Cov}((\Gamma_l + \beta_l)^{-1} \Gamma_l, \chi_l) \stackrel{s}{=} -\text{Cov}(\chi_l, \Gamma_l) = -Y_{mismatch}.$$

Thus,

$$-\mathcal{B}^{-1} M \text{Cov}((\Gamma_l + \beta_l)^{-1} \Gamma_l, \chi_l) = -\mathcal{A}_{mismatch}.$$

Thus, the  $\mathcal{P}^{D2D}$  price has the reverse ordering of the  $\mathcal{A}_{mismatch}$  under the conditions of the proposition. To complete the proof, we need to show that

$$\mathcal{A}_{mismatch} \stackrel{s}{=} Y_{mismatch}.$$

We prove this result in the next three lemmas.

Q.E.D.

**Lemma 21** *Suppose that*

$$\lambda_l = a_\lambda + b_\lambda \Gamma_l + O(\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2)$$

and

$$\chi_l = a_\chi + b_\chi \alpha_l + O(\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2)$$

with  $b_\lambda < 0 < b_\chi$ . Then,

$$\mathcal{A}_{mismatch} \stackrel{s}{=} Y_{mismatch}.$$

**Proof.**

$$\begin{aligned} Y_{mismatch} &= E[(\Gamma_l - \Gamma^*)(\chi_l - \chi^*)] \\ &\stackrel{s}{=} -E[(\lambda_l - \lambda^*)(\alpha_l - \alpha^*)] \stackrel{s}{=} E[(\lambda_l^{-1} - (\lambda^*)^{-1})(\alpha_l - \alpha^*)] = E[(b_l - b^*)(\alpha_l - \alpha^*)] \\ &= \mathcal{A}_{mismatch}. \end{aligned}$$

Q.E.D.

**Lemma 22** Suppose that  $x_l$  are sufficiently small and  $\Gamma_l$  are sufficiently close to  $\Gamma^*$  and that  $\bar{\theta}$  is sufficiently small. Then, there exists a unique equilibrium with<sup>50</sup>

$$\lambda_l = \lambda^*(\Gamma^*) + \Phi^\Gamma(\Gamma^*)(\Gamma_l - \Gamma^*) + O(\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2),$$

where  $\Phi^\Gamma(x)$  is negative and increasing in  $M$  for  $x > 0$  and  $M > 1$ . Furthermore,

$$\begin{aligned} a_l = & \Phi_0^x x_l + \Phi_0^X \bar{X} + \Phi_0^\Gamma X_{mismatch} + \Phi_0^\theta \bar{\theta} + (\Gamma_l - \Gamma^*)[\Phi_1^x x_l + \Phi_1^X \bar{X} + \Phi_1^\theta \bar{\theta}] \\ & + O((\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2)^{3/2}), \end{aligned}$$

where for some coefficients  $\Phi_0^x, \Phi_0^X, \Phi_0^\Gamma, \Phi_0^\theta, \Phi_1^x, \Phi_1^X, \Phi_1^\Gamma, \Phi_1^\theta$  with  $\Phi_0^x > 0$ .

**Proof of Lemma 22.** Recall that

$$\begin{aligned} \boldsymbol{\delta} &= \frac{[2 + \gamma\Lambda]\mathbf{a} - \gamma A\boldsymbol{\lambda}}{4 + 2\gamma\Lambda}; & \boldsymbol{\eta} &= -\frac{\gamma\boldsymbol{\lambda}}{2 + \gamma\Lambda}; \\ \partial_{a_l}\delta_j &= \frac{\mathbb{1}_l(j)[2 + \gamma\Lambda] - \gamma\lambda_l}{4 + 2\gamma\Lambda}; & \partial_{a_l}\boldsymbol{\delta} &= \frac{1}{4 + 2\gamma\Lambda} \left[ (2 + \gamma\Lambda)\mathbf{e}_l - \gamma\boldsymbol{\lambda} \right], \end{aligned}$$

where  $\mathbf{e}_l$  is the  $l$ th coordinate vector. Consider the first FOC:

$$\begin{aligned} 0 = & n(\partial_{a_l}\delta_l) \left[ \Gamma_l x_l - \lambda_l^{-1} a_l \right] + n\lambda_l^{-1} \left[ (2 - n\Gamma_l \lambda_l)(\partial_{a_l}\delta_l) - 1 \right] (\delta_l + \eta_l \bar{\theta}) \\ & - n(\Gamma_l + 2\beta_l)(\Gamma_l + \beta_l)^{-2} \left( \Psi_l \cdot \partial_{a_l}\boldsymbol{\delta} \right) \left( \Psi_l \cdot (x - n\boldsymbol{\delta} - n\bar{\theta}\boldsymbol{\eta}) \right). \end{aligned}$$

---

<sup>50</sup>As usual, we use  $\|x\| = (\sum_l x_l^2)^{1/2}$  to denote the Euclidean norm.

We cancel  $n$  and multiply by  $\lambda_l$  and get

$$\begin{aligned}
0 &= (\partial_{a_l} \delta_l) \left[ \lambda_l \Gamma_l x_l - a_l \right] + \left[ (2 - n \Gamma_l \lambda_l) (\partial_{a_l} \delta_l) - 1 \right] (\delta_l + \eta_l \bar{\theta}) \\
&\quad - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \Psi_l \cdot (x - n \boldsymbol{\delta} - n \bar{\theta} \boldsymbol{\eta}) \right) \\
&= (\partial_{a_l} \delta_l) \left[ \lambda_l \Gamma_l x_l - a_l \right] + \left[ (2 - n \Gamma_l \lambda_l) (\partial_{a_l} \delta_l) - 1 \right] \frac{[2 + \gamma \Lambda] a_l - \gamma (A + 2\bar{\theta}) \lambda_l}{4 + 2\gamma \Lambda} \\
&\quad - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \Psi_l \cdot x - n \Psi_l \cdot \frac{[2 + \gamma \Lambda] \mathbf{a} - \gamma (A + 2\bar{\theta}) \boldsymbol{\lambda}}{4 + 2\gamma \Lambda} \right) \\
&= (\partial_{a_l} \delta_l) \left[ \lambda_l \Gamma_l x_l - a_l \right] + \left[ (2 - n \Gamma_l \lambda_l) (\partial_{a_l} \delta_l) - 1 \right] \frac{[2 + \gamma \Lambda] a_l - \gamma (A + 2\bar{\theta}) \lambda_l}{4 + 2\gamma \Lambda} \\
&\quad - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \tilde{\chi} - \frac{n}{2} \tilde{A} + n \frac{\gamma (A + 2\bar{\theta})}{4 + 2\gamma \Lambda} \tilde{\Lambda} - \Gamma_l x_l + \frac{n \Gamma_l}{2} a_l - \frac{\gamma (A + 2\bar{\theta}) n \Gamma_l}{4 + 2\gamma \Lambda} \lambda_l \right) \\
&= -\frac{1}{2} \left[ n \Gamma_l \lambda_l (\partial_{a_l} \delta_l) + 1 + n \lambda_l \Gamma_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \right] a_l \\
&\quad + \lambda_l \Gamma_l \left[ (\partial_{a_l} \delta_l) + \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \right] x_l - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \tilde{\chi} \\
&\quad + \frac{n}{2} \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \tilde{A} \\
&\quad - \lambda_l \left[ n \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \tilde{\Lambda} - \Gamma_l \lambda_l \right) + (2 - n \Gamma_l \lambda_l) (\partial_{a_l} \delta_l) - 1 \right] \frac{\gamma (A + 2\bar{\theta})}{4 + 2\gamma \Lambda} \\
&= -\frac{1}{2} [1 + n c_{0x_l}] a_l + c_{0x_l} x_l - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) \\
&\quad - \frac{\gamma \lambda_l}{4 + 2\gamma \Lambda} \left[ n \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \tilde{\Lambda} - n c_{0x_l} + 2 (\partial_{a_l} \delta_l) - 1 \right] (A + 2\bar{\theta}),
\end{aligned}$$

where

$$\begin{aligned}
c_{0x_l} &= \lambda_l \Gamma_l \left[ (\partial_{a_l} \delta_l) + \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \right] \\
&= \lambda_l \Gamma_l \left[ \frac{[2 + \gamma \Lambda] - \gamma \lambda_l}{4 + 2\gamma \Lambda} + \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \frac{\Gamma_l}{2} \left( \frac{1}{\mathcal{B}(\Gamma_l + \beta_l)} - 1 \right) - \frac{\gamma (\tilde{\Lambda} - \Gamma_l \lambda_l)}{4 + 2\gamma \Lambda} \right) \right].
\end{aligned}$$

It follows that

$$\begin{aligned}
a_l &= \frac{2c_{0x_l}}{1 + n c_{0x_l}} x_l - \frac{2\lambda_l}{1 + n c_{0x_l}} \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) \\
&\quad - \frac{2}{1 + n c_{0x_l}} \frac{\gamma \lambda_l}{4 + 2\gamma \Lambda} \left[ n \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \tilde{\Lambda} - n c_{0x_l} + 2 (\partial_{a_l} \delta_l) - 1 \right] (A + 2\bar{\theta}),
\end{aligned}$$

We now write the Taylor series expansion of this FOC. We shall use the following lemma:

**Lemma 23** *When the dispersion of  $\Gamma_{i,t}$  is sufficiently small, we have*

$$\begin{aligned}\frac{\Gamma_{i,t}}{\mathcal{B}_t(\Gamma_{i,t} + \beta_{i,t})} &\approx \frac{\Gamma_t^*}{M} + \frac{1}{M^2 - 2M + 2}(\Gamma_{i,t} - \Gamma_t^*) \\ \frac{1}{\Gamma_{i,t} + \beta_{i,t}} &\approx \frac{M - 2}{\Gamma_t^*(M - 1)} - \frac{(M - 2)^2}{(M^2 - 2M + 2)(\Gamma_t^*)^2}(\Gamma_{i,t} - \Gamma_t^*) \\ \frac{\beta_{i,t}}{\Gamma_{i,t} + \beta_{i,t}} &\approx \frac{1}{M - 1} - \frac{M(M - 2)}{(M^2 - 2M + 2)(M - 1)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*).\end{aligned}$$

It thus follows that

$$\begin{aligned}\frac{\Gamma_{i,t}}{\Gamma_{i,t} + \beta_{i,t}} &\approx \frac{M - 2}{M - 1} + \frac{M(M - 2)}{(M^2 - 2M + 2)(M - 1)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \\ \frac{\Gamma_{i,t} + 2\beta_{i,t}}{(\Gamma_{i,t} + \beta_{i,t})^2} &= \frac{\Gamma_{i,t}}{(\Gamma_{i,t} + \beta_{i,t})^2} + \frac{2\beta_{i,t}}{(\Gamma_{i,t} + \beta_{i,t})^2} \\ &\approx \frac{M - 2}{M - 1} \left[ 1 + \frac{M}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \frac{M - 2}{\Gamma_t^*(M - 1)} \left[ 1 - \frac{(M - 1)(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \\ &\quad + \frac{2}{M - 1} \left[ 1 - \frac{M(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \frac{M - 2}{\Gamma_t^*(M - 1)} \left[ 1 - \frac{(M - 1)(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \\ &= \frac{(M - 2)^2}{\Gamma_t^*(M - 1)^2} \left[ 1 + \frac{M}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) - \frac{(M - 1)(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \\ &\quad + \frac{2(M - 2)}{\Gamma_t^*(M - 1)^2} \left[ 1 - \frac{M(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) - \frac{(M - 1)(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right] \\ &= \frac{M(M - 2)}{\Gamma_t^*(M - 1)^2} \left[ 1 - \frac{M(M - 2)}{(M^2 - 2M + 2)\Gamma_t^*}(\Gamma_{i,t} - \Gamma_t^*) \right].\end{aligned}$$

Moreover,

$$\lambda_l \Gamma_l \approx \lambda^* \Gamma^* + [\Phi^\Gamma(\Gamma^*)\Gamma^* + \lambda^*](\Gamma_l - \Gamma^*) \quad \text{and} \quad \tilde{\Lambda} \approx \lambda^* \Gamma^*$$

and

$$\begin{aligned}\tilde{A} &= \sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} a_j \approx \frac{\Gamma^*}{M} A + \frac{1}{M^2 - 2M + 2} A^\Gamma \\ \tilde{\chi} &= \sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} x_l \approx \frac{\Gamma^*}{M} X + \frac{1}{M^2 - 2M + 2} X_{mismatch},\end{aligned}$$

where

$$A^\Gamma = \sum_{l=1}^M (\Gamma_l - \Gamma^*) a_j \quad \text{and} \quad X_{mismatch} = \sum_{l=1}^M (\Gamma_l - \Gamma^*) x_l$$

Using the lemma above, to a first-order approximation, we have

$$\begin{aligned}
a_l &= [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)]x_l - [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)] \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) - [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)](A + 2\bar{\theta}) \\
&= [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)]x_l - [k_{0\theta} + k_{1\theta}(\Gamma_l - \Gamma^*)](A + 2\bar{\theta}) - [k_{0X} + k_{1X}(\Gamma_l - \Gamma^*)] \frac{\Gamma^*}{M} \left( X - \frac{n}{2} A \right) \\
&\quad - [k_{0X} + k_{1X}(\Gamma_l - \Gamma^*)] \frac{1}{M^2 - 2M + 2} \left( X_{mismatch} - \frac{n}{2} A^\Gamma \right) \\
&= [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)]x_l - [k_{0\theta} + k_{1\theta}(\Gamma_l - \Gamma^*)](A + 2\bar{\theta}) - [k_{0X} + k_{1X}(\Gamma_l - \Gamma^*)] \frac{\Gamma^*}{M} \left( X - \frac{n}{2} A \right) \\
&\quad - \frac{k_{0X}}{M^2 - 2M + 2} \left( X_{mismatch} - \frac{n}{2} A^\Gamma \right) \\
&= [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)]x_l - [k_{0X} + k_{1X}(\Gamma_l - \Gamma^*)] \frac{\Gamma^*}{M} X - \frac{k_{0X}}{M^2 - 2M + 2} X_{mismatch} \\
&\quad - \left[ \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{2M} + \frac{2Mk_{1\theta} - nk_{1X}\Gamma^*}{2M} (\Gamma_l - \Gamma^*) \right] A + \frac{n}{2} \frac{k_{0X}}{M^2 - 2M + 2} A^\Gamma \\
&\quad - [k_{0\theta} + k_{1\theta}(\Gamma_l - \Gamma^*)]2\bar{\theta},
\end{aligned}$$

Summing over  $i$  we get

$$\begin{aligned}
A^\Gamma &= k_{0x} X_{mismatch} \\
A &= k_{0x} X + k_{1x} X_{mismatch} - k_{0X} \Gamma^* X - \frac{Mk_{0X}}{M^2 - 2M + 2} X_{mismatch} \\
&\quad - \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{2} A + \frac{n}{2} \frac{Mk_{0X}}{M^2 - 2M + 2} A^\Gamma - 2Mk_{0\theta}\bar{\theta}.
\end{aligned}$$

Thus,

$$A = \left[ 1 + \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{2} \right]^{-1} \left( [k_{0x} - k_{0X}\Gamma^*] X + \left[ k_{1x} - \frac{(2 - nk_{0x})Mk_{0X}}{2(M^2 - 2M + 2)} \right] X_{mismatch} - 2Mk_{0\theta}\bar{\theta} \right)$$

Substituting into the equation for  $a_l$  we obtain

$$\begin{aligned}
a_l = & [k_{0x} + k_{1x}(\Gamma_l - \Gamma^*)]x_l - [k_{0X} + k_{1X}(\Gamma_l - \Gamma^*)]\frac{\Gamma^*}{M}X \\
& - \left[ \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{M} + \frac{2Mk_{1\theta} - nk_{1X}\Gamma^*}{M}(\Gamma_l - \Gamma^*) \right] \frac{k_{0x} - k_{0X}\Gamma^*}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*}X \\
& - \left( 1 - \frac{n}{2}k_{0x} \right) \frac{k_{0X}}{M^2 - 2M + 2} X_{mismatch} \\
& - \left[ \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{M} + \frac{2Mk_{1\theta} - nk_{1X}\Gamma^*}{M}(\Gamma_l - \Gamma^*) \right] \frac{k_{1x} - \frac{(2-nk_{0x})Mk_{0X}}{2(M^2-2M+2)}}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} X_{mismatch} \\
& + \left[ \frac{2Mk_{0\theta} - nk_{0X}\Gamma^*}{M} + \frac{2Mk_{1\theta} - nk_{1X}\Gamma^*}{M}(\Gamma_l - \Gamma^*) \right] \frac{2Mk_{0\theta}}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} \bar{\theta} \\
& - 2[k_{0\theta} + k_{1\theta}(\Gamma_l - \Gamma^*)]\bar{\theta}.
\end{aligned}$$

It follows that

$$\begin{aligned}
a_l = & \Phi_0^x x_l + \Phi_0^X X + \Phi_0^\Gamma X_{mismatch} + \Phi_0^\theta \bar{\theta} + (\Gamma_l - \Gamma^*)[\Phi_1^x x_l + \Phi_1^X X + \Phi_1^\theta \bar{\theta}] \\
& + O((\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2)^{3/2}),
\end{aligned}$$

with

$$\begin{aligned}
\Phi_0^x &= k_{0x}; & \Phi_1^x &= k_{1x} \\
\Phi_0^X &= - \left[ \frac{2k_{0X}\Gamma^* + k_{0x}(2Mk_{0\theta} - nk_{0X}\Gamma^*)}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} \right] \frac{1}{M}; \\
\Phi_1^X &= - \left[ \frac{k_{1X}\Gamma^*(2 + 2Mk_{0\theta}) + 2Mk_{1\theta}(k_{0x} - k_{0X}\Gamma^*) - nk_{1X}k_{0x}\Gamma^*}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} \right] \frac{1}{M} \\
\Phi_0^\Gamma &= - \left[ \frac{k_{0X}(2 - nk_{0x})}{M^2 - 2M + 2} + \frac{k_{1x}(2Mk_{0\theta} - nk_{0X}\Gamma^*)}{M} \right] \frac{1}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} \\
\Phi_0^\theta &= \frac{-4k_{0\theta}}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*} \\
\Phi_1^\theta &= - \frac{2[2k_{1\theta} - n(k_{0X}k_{1\theta} - k_{0\theta}k_{1X})\Gamma^*]}{2 + 2Mk_{0\theta} - nk_{0X}\Gamma^*}.
\end{aligned}$$

Q.E.D.

**Lemma 24** *After the D2C trading round, dealer inventories become*

$$\begin{aligned}
x_l \approx & -(\varphi_0^a - 0.5n)\lambda^* \alpha_l - (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^* \bar{\alpha} - M\lambda^*[(\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
& + \lambda^*[\varphi_0^a + M\varphi_0^A + n\eta_0^*]d + \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Theta \\
& + (\lambda_l - \lambda^*)[-(\varphi_0^a + \varphi_1^a \lambda^* - 0.5n)\alpha_l + M\lambda^*(0.5n\gamma\eta_0^* - \varphi_1^A)\bar{\alpha}] \\
& + (\lambda_l - \lambda^*)[(\varphi_0^a + n\eta_0^* + \lambda^* \varphi_1^a + M\lambda^* \varphi_1^A)d - \gamma\eta_0^* \Theta + \varphi_1^\theta \bar{\theta}],
\end{aligned}$$



where

$$\bar{\alpha} = \frac{1}{M} \sum_l \alpha_i; \quad \alpha_{mismatch} = \frac{1}{\Lambda} \sum_l (\lambda_i - \lambda^*) \alpha_i; \quad \text{and} \quad \eta_0^* = -\frac{1}{2 + \gamma \Lambda}.$$

**Proof of Lemma 24.**

$$\begin{aligned} \chi_l &= x_l - \sum_j (\delta_l + \eta_l \theta_{j,t-}) = x_l - 0.5na_l - \eta_l(\Theta_t + 0.5nA) \\ &= \varphi_0^a a_l + \varphi_0^A A + \varphi_{0,\lambda}^A A^\lambda + \varphi_0^\theta \bar{\theta} + (\lambda_l - \lambda^*)[\varphi_1^a a_l + \varphi_1^A A + \varphi_{1,\lambda}^A A^\lambda + \varphi_1^\theta \bar{\theta}] \\ &\quad - 0.5na_l - \eta_l(\Theta + 0.5nA) \\ &= (\varphi_0^a - 0.5n)a_l + (\varphi_0^A - 0.5n\eta_l)A + \varphi_{0,\lambda}^A A^\lambda + \varphi_0^\theta \bar{\theta} - \eta_l \Theta + (\lambda_l - \lambda^*)[\varphi_1^a a_l + \varphi_1^A A + \varphi_1^\theta \bar{\theta}]. \end{aligned}$$

For simplicity, we assume that the total customer inventory shocks  $\Theta_t = \sum_l \theta_{j,t}$  are i.i.d. over time. We rewrite both  $a_l$  and  $A$  in terms of mid-prices:

$$\begin{aligned} a_l &= \lambda_l(d - \alpha_l) \\ A &= \sum_l a_j = d \sum_l \lambda_l - \sum_l \lambda_l \alpha_j = d\Lambda - \sum_l \lambda_l \alpha_j \\ &= M\lambda^* [d - (\alpha_{mismatch} + \bar{\alpha})] \\ A^\lambda &= \sum_l (\lambda_l - \lambda^*) a_j = d \sum_l (\lambda_l - \lambda^*) \lambda_l - \sum_l (\lambda_l - \lambda^*) \lambda_l \alpha_j \approx -\lambda^* \sum_l (\lambda_l - \lambda^*) \alpha_j \\ &= -M(\lambda^*)^2 \hat{\alpha}. \end{aligned}$$

Thus,

$$\begin{aligned}
\tilde{\chi}_l &= (\varphi_0^a - 0.5n)\lambda_l(d - \alpha_l) + (\varphi_0^A - 0.5n\eta_i)M\lambda^*[d - (\alpha_{mismatch} + \bar{\alpha})] - \varphi_{0,\lambda}^A M(\lambda^*)^2 \hat{\alpha} \\
&\quad + \varphi_0^\theta \bar{\theta} - \eta_i \Theta + (\lambda_l - \lambda^*)[\varphi_1^a \lambda_l(d - \alpha_l) + \varphi_1^A M\lambda^*[d - (\alpha_{mismatch} + \bar{\alpha})] + \varphi_1^\theta \bar{\theta}] \\
&= -(\varphi_0^a - 0.5n)\lambda_l \alpha_l - (\varphi_0^A - 0.5n\gamma\lambda_l \eta_0^*)M\lambda^* \bar{\alpha} - M\lambda^*[(\varphi_0^A - 0.5n\gamma\lambda_l \eta_0^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
&\quad + [(\varphi_0^a - 0.5n)\lambda_l + (\varphi_0^A - 0.5n\gamma\lambda_l \eta_0^*)M\lambda^*]d + \varphi_0^\theta \bar{\theta} - \gamma\lambda_l \eta_0^* \Theta \\
&\quad + (\lambda_l - \lambda^*)[-\varphi_1^a \lambda_l \alpha_l - \varphi_1^A M\lambda^*(\alpha_{mismatch} + \bar{\alpha}) + (\varphi_1^a \lambda_l + \varphi_1^A M\lambda^*)d + \varphi_1^\theta \bar{\theta}] \\
&\approx -(\varphi_0^a - 0.5n)\lambda^* \alpha_l - (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^* \bar{\alpha} - M\lambda^*[(\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
&\quad + [(\varphi_0^a - 0.5n)\lambda^* + (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^*]d + \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Theta \\
&\quad - (\varphi_0^a - 0.5n)(\lambda_l - \lambda^*)\alpha_l + 0.5n\gamma\eta_0^*(\lambda_l - \lambda^*)M\lambda^* \bar{\alpha} \\
&\quad + [(\varphi_0^a - 0.5n)(\lambda_l - \lambda^*) - 0.5n\gamma\eta_0^*(\lambda_l - \lambda^*)M\lambda^*]d - \gamma\eta_0^*(\lambda_l - \lambda^*)\Theta \\
&\quad + (\lambda_l - \lambda^*)[-\varphi_1^a \lambda^* \alpha_l - \varphi_1^A M\lambda^* \bar{\alpha} + \lambda^*(\varphi_1^a + M\varphi_1^A)d + \varphi_1^\theta \bar{\theta}] \\
&\approx -(\varphi_0^a - 0.5n)\lambda^* \alpha_l - (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^* \bar{\alpha} - M\lambda^*[(\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
&\quad + [(\varphi_0^a - 0.5n)\lambda^* + (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^*]d + \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Theta \\
&\quad + (\lambda_l - \lambda^*)[-(\varphi_0^a - 0.5n)\alpha_l + 0.5nM\gamma\eta_0^* \lambda^* \bar{\alpha}] \\
&\quad + (\lambda_l - \lambda^*)[(\varphi_0^a - 0.5n) - 0.5nM\gamma\eta_0^* \lambda^*]d - (\lambda_l - \lambda^*)\gamma\eta_0^* \Theta \\
&\quad + (\lambda_l - \lambda^*)[-\varphi_1^a \lambda^* \alpha_l - \varphi_1^A M\lambda^* \bar{\alpha} + \lambda^*(\varphi_1^a + M\varphi_1^A)d + \varphi_1^\theta \bar{\theta}],
\end{aligned}$$

where we used the definition

$$\eta_i = -\frac{\gamma\lambda_l}{2 + \gamma\Lambda} = \gamma\eta_0^* \lambda_l \implies 1 + M\gamma\eta_0^* \lambda^* = -2\eta_0^*.$$

Thus,

$$\begin{aligned}
\tilde{\chi}_l &\approx -(\varphi_0^a - 0.5n)\lambda^* \alpha_l - (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*)M\lambda^* \bar{\alpha} - M\lambda^*[(\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
&\quad + \lambda^*[\varphi_0^a + M\varphi_0^A + n\eta_0^*]d + \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Theta \\
&\quad + (\lambda_l - \lambda^*)[-(\varphi_0^a - 0.5n)\alpha_l + 0.5nM\gamma\eta_0^* \lambda^* \bar{\alpha}] \\
&\quad + (\lambda_l - \lambda^*)[(\varphi_0^a + n\eta_0^*)d - \gamma\eta_0^* \Theta] \\
&\quad + (\lambda_l - \lambda^*)[-\varphi_1^a \lambda^* \alpha_l - \varphi_1^A M\lambda^* \bar{\alpha} + \lambda^*(\varphi_1^a + M\varphi_1^A)d + \varphi_1^\theta \bar{\theta}],
\end{aligned}$$

Q.E.D.

## F.2 Proof of Proposition 16

**Proof of Proposition 16.** The proof follows along the same lines as that of Proposition 4 so we skip some details.

By Proposition 2, our goal is to show that aggregate liquidity  $\mathcal{B}$  is monotone increasing in spread dispersion. Proposition 25 shows that, for small heterogeneity in risk aversion, there is a one-to-one mapping between  $\lambda_l$  and  $\Gamma_l$ . Therefore, for small heterogeneity in risk aversion, proving Proposition 16 is equivalent to proving that  $\mathcal{B}$  is monotone increasing in

the dispersion in risk aversions. This follows directly because  $\mathcal{B}$  is a convex function of risk aversions (see Lemma 11).

Q.E.D.

### F.3 Proof of Lemma 17

**Proof of Lemma 17.** The first order condition is

$$-\alpha_l - 2\lambda_l^{-1}q_l + d - \gamma\left(\sum_l q_l + \theta\right) = 0$$

where  $q = (q_l)$  and this gives

$$q = (2B + \gamma\mathbf{1})^{-1}((d - \gamma\theta)\mathbf{1} - \alpha)$$

where  $B = \text{diag}(\lambda_l^{-1})$ ,  $\alpha = (\alpha_l)$ .

Q.E.D.

### F.4 Proof of Lemma 18

**Proof of Lemma 18.** We have

$$E[\Pi_l(Q_l)] = E\left[\sum_l ((d - \lambda_l^{-1}a_l) + \lambda_l^{-1}(\delta_l + \eta_l\bar{\theta}_c))(\delta_l + \eta_l\bar{\theta}_c)\right] = (d - \lambda_l^{-1}a_l)n(\delta_l + \eta_l\bar{\theta}) + \lambda_l^{-1}(n(\delta_l + \eta_l\bar{\theta})^2 + n\eta_l^2\sigma_\theta^2).$$

Thus, the utility becomes

$$\begin{aligned} E[U_l(\chi_l; (\chi_l)_{j \neq i})] &= (d - \lambda_l^{-1}a_l)n(\delta_l + \eta_l\bar{\theta}) + \lambda_l^{-1}(n(\delta_l + \eta_l\bar{\theta})^2 + n\eta_l^2\sigma_\theta^2) \\ &\quad + (x_l - n\delta_l - \eta_l n\bar{\theta})d - 0.5\Gamma_l[(x_l - n\delta_l - \eta_l n\bar{\theta})^2 + \eta_l^2 n\sigma_\theta^2] \\ &\quad + (0.5\Gamma_l + \beta_l)(\Gamma_l + \beta_l)^{-2} \left[ \left( \Psi_l \cdot (x - n\delta - n\bar{\theta}\eta) \right)^2 + n\sigma_\theta^2(\Psi_l \cdot \eta)^2 \right] \\ &= x_l d - n\lambda_l^{-1}a_l(\delta_l + \eta_l\bar{\theta}) + n\lambda_l^{-1} \left[ (\delta_l + \eta_l\bar{\theta})^2 + \eta_l^2\sigma_\theta^2 \right] \\ &\quad - 0.5\Gamma_l \left[ (x_l - n\delta_l - n\bar{\theta}\eta_l)^2 + n\sigma_\theta^2\eta_l^2 \right] \\ &\quad + (0.5\Gamma_l + \beta_l)(\Gamma_l + \beta_l)^{-2} \left[ \left( \Psi_l \cdot (x - n\delta - n\bar{\theta}\eta) \right)^2 + n\sigma_\theta^2(\Psi_l \cdot \eta)^2 \right]. \end{aligned}$$

where we have defined

$$\Psi = \left( \frac{\Gamma_1}{\mathcal{B}(\Gamma_1 + \beta_1)}, \dots, \frac{\Gamma_M}{\mathcal{B}(\Gamma_M + \beta_M)} \right)^T \quad \text{and} \quad \Psi_l \equiv \Psi - \Gamma_l \mathbf{1}_{l=i}.$$

Now, we have

$$\begin{aligned} \delta &= \frac{[2 + \gamma\Lambda]\mathbf{a} - \gamma A\boldsymbol{\lambda}}{4 + 2\gamma\Lambda} \\ \eta &= -\frac{\gamma\boldsymbol{\lambda}}{2 + \gamma\Lambda}. \end{aligned}$$

Q.E.D.

## F.5 Proof of Proposition 19

**Proof of Proposition 19.** It follows from Lemma 17 that

$$\begin{aligned}\partial_{a_l}\delta_j &= \frac{\mathbb{1}_l(j)[2 + \gamma\Lambda] - \gamma\lambda_l}{4 + 2\gamma\Lambda} \\ \partial_{a_l}\eta_j &= 0 \\ \partial_{\lambda_l}\delta_j &= -\frac{2\gamma\left(2a_j + \gamma[a_j(\Lambda - \lambda_l) - (A - a_j)\lambda_l]\right)}{(4 + 2\gamma\Lambda)^2} + \gamma\frac{a_j - A\mathbb{1}_l(j)}{4 + 2\gamma\Lambda} \\ \partial_{\lambda_l}\eta_j &= \frac{\gamma^2\lambda_l}{(2 + \gamma\Lambda)^2} - \mathbb{1}_l(j)\frac{\gamma}{2 + \gamma\Lambda},\end{aligned}$$

where  $\mathbb{1}_l(\cdot)$  is the indicator function. We make the following definition:

$$\begin{aligned}\partial_{a_l}\boldsymbol{\delta} &\equiv (\partial_{a_l}\delta_1, \dots, \partial_{a_l}\delta_n)^T = \frac{1}{4 + 2\gamma\Lambda} \left[ (2 + \gamma\Lambda)\mathbf{e}_l - \gamma\boldsymbol{\lambda} \right] \\ \partial_{a_l}\boldsymbol{\eta} &= 0 \\ \partial_{\lambda_l}\boldsymbol{\delta} &\equiv (\partial_{\lambda_l}\delta_1, \dots, \partial_{\lambda_l}\delta_n)^T = -\frac{2\gamma}{(4 + 2\gamma\Lambda)^2} \left( (2 + \gamma\Lambda)\mathbf{a} - \gamma A\boldsymbol{\lambda} \right) + \frac{\gamma}{4 + 2\gamma\Lambda} (\mathbf{a} - A\mathbf{e}_l) \\ \partial_{\lambda_l}\boldsymbol{\eta} &\equiv (\partial_{\lambda_l}\eta_1, \dots, \partial_{\lambda_l}\eta_n)^T = \frac{\gamma^2}{(2 + \gamma\Lambda)^2} \boldsymbol{\lambda} - \frac{\gamma}{2 + \gamma\Lambda} \mathbf{e}_l,\end{aligned}$$

where  $\mathbf{e}_l$  is the  $l$ th coordinate vector. Then, the FOCs are

$$\begin{aligned}0 &= n(\partial_{a_l}\delta_l) \left[ \Gamma_l x_l - \lambda_l^{-1} a_l \right] + n\lambda_l^{-1} \left[ (2 - n\Gamma_l \lambda_l)(\partial_{a_l}\delta_l) - 1 \right] (\delta_l + \eta_l \bar{\theta}) \\ &\quad - n(\Gamma_l + 2\beta_l)(\Gamma_l + \beta_l)^{-2} \left( \Psi_l \cdot \partial_{a_l}\boldsymbol{\delta} \right) \left( \Psi_l \cdot (x - n\boldsymbol{\delta} - n\bar{\theta}\boldsymbol{\eta}) \right) \\ 0 &= n\lambda_l^{-2} a_l (\delta_l + \eta_l \bar{\theta}) - n\lambda_l^{-2} \left[ (\delta_l + \eta_l \bar{\theta})^2 + \eta_l^2 \sigma_\theta^2 \right] + n\lambda_l^{-1} \sigma_\theta^2 \left[ 2 - \Gamma_l \lambda_l \right] (\partial_{\lambda_l}\eta_l) \eta_l \\ &\quad + \lambda_l^{-1} n \left[ 2(\delta_l + \eta_l \bar{\theta}) + \lambda_l \Gamma_l (x_l - n\delta_l - n\bar{\theta}\eta_l) - a_l \right] (\partial_{\lambda_l}\delta_l + \bar{\theta}\partial_{\lambda_l}\eta_l) \\ &\quad - n(\Gamma_l + 2\beta_l)(\Gamma_l + \beta_l)^{-2} \left[ \left( \Psi_l \cdot (\partial_{\lambda_l}\boldsymbol{\delta} + \bar{\theta}\partial_{\lambda_l}\boldsymbol{\eta}) \right) \left( \Psi_l \cdot (x - n\boldsymbol{\delta} - n\bar{\theta}\boldsymbol{\eta}) \right) - \sigma_\theta^2 (\Psi_l \cdot \partial_{\lambda_l}\boldsymbol{\eta})(\Psi_l \cdot \boldsymbol{\eta}) \right].\end{aligned}$$

Suppose that

$$x_l = 0, \quad \Gamma_l = \Gamma \quad \forall l; \quad \text{and} \quad \bar{\theta} = 0.$$

It follows that

$$\begin{aligned}
\beta_i &= \frac{\Gamma}{M-2} \quad \forall i; \quad \mathcal{B}(\Gamma_i + \beta_i) = M \quad \forall i; & \mathcal{B} &= \frac{M(M-2)}{\Gamma(M-1)}; & \Psi &= \frac{\Gamma}{M} \mathbf{1}; \\
\tilde{A} &= \frac{\Gamma}{M} A; & \tilde{\Lambda} &= \frac{\Gamma}{M} \Lambda; & \frac{\Gamma_i + 2\beta_i}{(\Gamma_i + \beta_i)^2} &= \frac{M(M-2)}{\Gamma(M-1)^2}; \\
\delta_l &= \frac{2a}{4 + 2\gamma\Lambda}; & \eta_l &= -\frac{\gamma\lambda}{2 + \gamma\Lambda}; & \Psi_l \cdot \boldsymbol{\delta} &= 0; & \Psi_l \cdot \boldsymbol{\eta} &= 0;
\end{aligned}$$

Then, the FOCs become

$$\begin{aligned}
0 &= n(\partial_{a_l} \delta_l) \left[ \Gamma_l x_l - \lambda_l^{-1} a_l \right] + n\lambda_l^{-1} \left[ (2 - n\Gamma_l \lambda_l)(\partial_{a_l} \delta_l) - 1 \right] (\delta_l + \eta_l \bar{\theta}) \\
0 &= n\lambda_l^{-2} a_l (\delta_l + \eta_l \bar{\theta}) - n\lambda_l^{-2} \left[ (\delta_l + \eta_l \bar{\theta})^2 + \eta_l^2 \sigma_\theta^2 \right] + n\lambda_l^{-1} \sigma_\theta^2 \left[ 2 - \Gamma_l \lambda_l \right] (\partial_{\lambda_l} \eta_l) \eta_l \\
&\quad + \lambda_l^{-1} n \left[ 2(\delta_l + \eta_l \bar{\theta}) + \lambda_l \Gamma_l (x_l - n\delta_l - n\bar{\theta} \eta_l) - a_l \right] (\partial_{\lambda_l} \delta_l + \bar{\theta} \partial_{\lambda_l} \eta_l).
\end{aligned}$$

We simplify the first of the FOCs and obtain

$$0 = 2 \left[ \gamma(M-1)(\gamma M + \Gamma n)\lambda^2 + 2(2\gamma M + \Gamma n)\lambda + 4 \right] \frac{a}{(4 + 2\gamma\Lambda)^2}.$$

The only solution is

$$a = 0,$$

since the quadratic equation in  $\lambda$  admits only negative solutions (if any). Substituting  $a = 0$  into the second of the FOCs and simplifying yields

$$0 = 1 + (2 - \Gamma\lambda) \left[ \frac{\gamma\lambda}{2 + M\gamma\lambda} - 1 \right].$$

This equation has a unique positive solution, which is

$$\lambda = \frac{\gamma(M-2) - 2\Gamma + \sqrt{\gamma^2(M-2)^2 + 4\Gamma(\Gamma + 4\gamma)M}}{2\gamma(M-1)\Gamma}.$$

We now present the proof of the monotonicity results.

$$\begin{aligned}
\frac{\partial \lambda^*}{\partial \gamma} &= \frac{\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} - (\gamma M + 2\Gamma)}{\gamma^2(M-1)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}} \\
&= \frac{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M - (\gamma M + 2\Gamma)^2}{\gamma^2(M-1)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}[\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} + (\gamma M + 2\Gamma)]} \\
&= \frac{-4\gamma^2(M-1)}{\gamma^2(M-1)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}[\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} + (\gamma M + 2\Gamma)]} \\
&< 0
\end{aligned}$$

since  $M \geq 2$ .

$$\frac{\partial \lambda^*}{\partial \Gamma} = \frac{-(M-2)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} - \gamma(M-2)^2 - 2M\Gamma}{2\Gamma^2(M-1)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}} < 0.$$

$$\frac{\partial \lambda^*}{\partial M} = \frac{\gamma^2(M-2) + (\gamma + 2\Gamma)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} - 2\Gamma(\gamma + 2\Gamma + \gamma M)}{2\gamma\Gamma(M-1)^2\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}}.$$

To show that this derivative is always positive for  $M \geq 2$  it is enough to show that its numerator is positive for  $M \geq 2$ . We do so by showing that the numerator is increasing and positive for  $M \geq 2$ . Consider the numerator. Its derivative with respect to  $M$  is

$$\gamma \left( \gamma - 2\Gamma + (\gamma + 2\Gamma) \frac{2\Gamma + \gamma(M-2)}{\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M}} \right).$$

We now show that the term in brackets is positive. First,

$$1 \geq \frac{\gamma(M-2) + 2\Gamma}{\sqrt{\gamma^2(M-2)^2 + 4\gamma\Gamma M + 4\Gamma^2}} \geq 1 - \frac{2\gamma}{\gamma M + 2\Gamma}.$$

Then,

$$\begin{aligned}
\gamma + 2\Gamma &\geq \frac{(\gamma + 2\Gamma)[\gamma(M-2) + 2\Gamma]}{\sqrt{\gamma^2(M-2)^2 + 4\gamma\Gamma M + 4\Gamma^2}} \geq (\gamma + 2\Gamma) - \frac{2\gamma(\gamma + 2\Gamma)}{\gamma M + 2\Gamma} \\
2\gamma &\geq \gamma - 2\Gamma + \frac{(\gamma + 2\Gamma)[\gamma(M-2) + 2\Gamma]}{\sqrt{\gamma^2(M-2)^2 + 4\gamma\Gamma M + 4\Gamma^2}} \geq 2\gamma \left[ 1 - \frac{\gamma + 2\Gamma}{\gamma M + 2\Gamma} \right] > 0
\end{aligned}$$

for  $M > 1$ . Thus, to show that  $\lambda^*$  is increasing, it remains to

$$\gamma^2(M-2) + (\gamma + 2\Gamma)\sqrt{4\Gamma^2 + \gamma^2(M-2)^2 + 4\gamma\Gamma M} - 2\Gamma(\gamma + 2\Gamma + \gamma M)$$

evaluated at  $M = 2$  is non-negative. When  $M = 2$ , the expression above is

$$2 \left[ (\gamma + 2\Gamma) \sqrt{\Gamma^2 + 2\gamma\Gamma} - \Gamma(2\Gamma + 3\gamma) \right] > 0$$

for  $\gamma, \Gamma > 0$ .

Q.E.D.

## F.6 Proof of Proposition 20

**Proposition 25** *Suppose that  $a_l$ , the dispersion in bid-ask spreads  $\lambda_l^{-1}$ , and  $\bar{\theta}$  are sufficiently small, then,*

$$\Gamma_l = \Gamma^*(\lambda^*) + \varphi^\lambda(\lambda^*)(\lambda_l - \lambda^*) + O(\|\lambda - \lambda^*\|^2 + \|a\|^2 + \bar{\theta}^2)$$

where

$$\lambda^* = \frac{1}{M} \sum_{l=1}^M \lambda_l \quad \text{and} \quad \varphi^\lambda(\cdot) = \frac{1}{\Phi^\Gamma(\cdot)} < 0.$$

Furthermore,

$$\begin{aligned} x_l = & \varphi_0^a a_l + \varphi_0^A A + \varphi_{0,\lambda}^A A_{mismatch} + \varphi_0^\theta \bar{\theta} + (\lambda_l - \lambda^*) [\varphi_1^a a_l + \varphi_1^A A + \varphi_{1,\lambda}^A A^\lambda + \varphi_1^\theta \bar{\theta}] \\ & + O((\|\Gamma - \Gamma^*\|^2 + \|x\|^2 + \bar{\theta}^2)^{3/2}), \end{aligned}$$

where

$$A = \sum_{l=1}^M a_l; \quad A_{mismatch} = \sum_{l=1}^M (\lambda_l - \lambda^*) a_l,$$

and  $\varphi_0^a, \varphi_0^A, \varphi_{0,\lambda}^A, \varphi_0^\theta, \varphi_1^a, \varphi_1^A, \varphi_{1,\lambda}^A$ , and  $\varphi_1^\theta$  are rational functions of  $n, M, \gamma$  and  $\lambda^*$ . Moreover,

$$\varphi_0^a(\cdot) = 1/\Phi_0^x(\cdot) > 0.$$

**Proof of Proposition 25.** The proof is similar to that of Proposition 22. The FOC is

$$\begin{aligned} 0 = & -\frac{1}{2} [1 + nc_{0x_l}] a_l + c_{0x_l} x_l - \lambda_l \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} (\Psi_l \cdot \partial_{a_l} \boldsymbol{\delta}) \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) \\ & - \frac{\gamma \lambda_l}{4 + 2\gamma\Lambda} \left[ n \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} (\Psi_l \cdot \partial_{a_l} \boldsymbol{\delta}) \tilde{\Lambda} - nc_{0x_l} + 2(\partial_{a_l} \delta_l) - 1 \right] (A + 2\bar{\theta}). \end{aligned}$$



It follows that

$$\begin{aligned}
x_l &= \frac{1 + nc_{0x_l}}{2c_{0x_l}} a_l + \frac{\lambda_l}{c_{0x_l}} \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) \\
&\quad + \frac{1}{c_{0x_l}} \frac{\gamma \lambda_l}{4 + 2\gamma\Lambda} \left[ n \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \Psi_l \cdot \partial_{a_l} \boldsymbol{\delta} \right) \tilde{\Lambda} - nc_{0x_l} + 2(\partial_{a_l} \delta_l) - 1 \right] (A + 2\bar{\theta}) \\
&\approx [h_{0a} + h_{1a}(\lambda_l - \lambda^*)] a_l + [h_{0X} + h_{1X}(\lambda_l - \lambda^*)] \left( \tilde{\chi} - \frac{n}{2} \tilde{A} \right) + [h_{0\theta} + h_{1\theta}(\lambda_l - \lambda^*)] (A + 2\bar{\theta}) \\
&= [h_{0a} + h_{1a}(\lambda_l - \lambda^*)] a_l + \frac{\Gamma^*[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M} \left( X - \frac{n}{2} A \right) + [h_{0\theta} + h_{1\theta}(\lambda_l - \lambda^*)] (A + 2\bar{\theta}) \\
&\quad + \frac{\varphi^\lambda [h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M^2 - 2M + 2} \left( X^\lambda - \frac{n}{2} A^\lambda \right),
\end{aligned}$$

where

$$A^\lambda = \sum_{l=1}^M (\lambda_l - \lambda^*) a_j \quad \text{and} \quad X^\lambda = \sum_{l=1}^M (\lambda_l - \lambda^*) x_l,$$

and we used

$$\begin{aligned}
\tilde{A} &= \sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} a_j \approx \frac{\Gamma^*}{M} A + \frac{\varphi^\lambda}{M^2 - 2M + 2} A^\Gamma \\
\tilde{\chi} &= \sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} x_l \approx \frac{\Gamma^*}{M} X + \frac{\varphi^\lambda}{M^2 - 2M + 2} X_{mismatch}.
\end{aligned}$$

Summing over  $i$  yield

$$\begin{aligned}
X^\lambda &= h_{0a} A^\lambda \\
X &= h_{0a} A + h_{1a} A^\lambda + \Gamma^* h_{0X} \left( X - \frac{n}{2} A \right) + M h_{0\theta} (A + 2\bar{\theta}) + \frac{M h_{0X} \varphi^\lambda}{M^2 - 2M + 2} \left( X^\lambda - \frac{n}{2} A^\lambda \right),
\end{aligned}$$

implying that

$$X = \frac{h_{0a} - \frac{n}{2} \Gamma^* h_{0X} + M h_{0\theta}}{1 - \Gamma^* h_{0X}} A + \frac{2M h_{0\theta}}{1 - \Gamma^* h_{0X}} \bar{\theta} + \frac{h_{1a} + \frac{M h_{0X} \varphi^\lambda}{M^2 - 2M + 2} (h_{0a} - \frac{n}{2})}{1 - \Gamma^* h_{0X}} A^\lambda.$$

Substituting into the equation for  $x_l$  we obtain

$$\begin{aligned}
x_l = & [h_{0a} + h_{1a}(\lambda_l - \lambda^*)]a_l + \frac{\Gamma^*[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M} \frac{h_{0a} - \frac{n}{2}\Gamma^*h_{0X} + Mh_{0\theta}}{1 - \Gamma^*h_{0X}} A \\
& - \frac{n}{2} \frac{\Gamma^*[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M} A + [h_{0\theta} + h_{1\theta}(\lambda_l - \lambda^*)]A + \left(h_{0a} - \frac{n}{2}\right) \frac{\varphi^\lambda[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M^2 - 2M + 2} A^\lambda \\
& + \frac{\Gamma^*[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{M} \frac{h_{1a} + \frac{Mh_{0X}\varphi^\lambda}{M^2 - 2M + 2} \left(h_{0a} - \frac{n}{2}\right)}{1 - \Gamma^*h_{0X}} A^\lambda \\
& + 2[h_{0\theta} + h_{1\theta}(\lambda_l - \lambda^*)]\bar{\theta} + \frac{2h_{0\theta}\Gamma^*[h_{0X} + h_{1X}(\lambda_l - \lambda^*)]}{1 - \Gamma^*h_{0X}} \bar{\theta}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\varphi_0^a &= h_{0a} \\
\varphi_0^A &= h_{0\theta} + \frac{\Gamma^*h_{0X}}{M} \frac{h_{0a} - \frac{n}{2}\Gamma^*h_{0X} + Mh_{0\theta}}{1 - \Gamma^*h_{0X}} - \frac{n}{2} \frac{\Gamma^*h_{0X}}{M} \\
&= h_{0\theta} + \frac{\Gamma^*h_{0X}}{M} \frac{1}{1 - \Gamma^*h_{0X}} \left[ h_{0a} + Mh_{0\theta} - \frac{n}{2} \right] \\
\varphi_{0,\lambda}^A &= \left(h_{0a} - \frac{n}{2}\right) \frac{\varphi^\lambda h_{0X}}{M^2 - 2M + 2} + \frac{\Gamma^*h_{0X}}{M} \frac{h_{1a} + \frac{Mh_{0X}\varphi^\lambda}{M^2 - 2M + 2} \left(h_{0a} - \frac{n}{2}\right)}{1 - \Gamma^*h_{0X}} \\
\varphi_0^\theta &= 2h_{0\theta} \left[ 1 + \frac{\Gamma^*h_{0X}}{1 - \Gamma^*h_{0X}} \right] = \frac{2h_{0\theta}}{1 - \Gamma^*h_{0X}} \\
\varphi_1^a &= h_{1a} \\
\varphi_1^A &= h_{1\theta} + \frac{\Gamma^*h_{1X}}{M} \frac{h_{0a} - \frac{n}{2}\Gamma^*h_{0X} + Mh_{0\theta}}{1 - \Gamma^*h_{0X}} - \frac{n}{2} \frac{\Gamma^*h_{1X}}{M} \\
&= h_{1\theta} + \frac{\Gamma^*h_{1X}}{M} \frac{1}{1 - \Gamma^*h_{0X}} \left[ h_{0a} + Mh_{0\theta} - \frac{n}{2} \right] \\
\varphi_{1,\lambda}^A &= \left(h_{0a} - \frac{n}{2}\right) \frac{\varphi^\lambda h_{1X}}{M^2 - 2M + 2} + \frac{\Gamma^*h_{1X}}{M} \frac{h_{1a} + \frac{Mh_{0X}\varphi^\lambda}{M^2 - 2M + 2} \left(h_{0a} - \frac{n}{2}\right)}{1 - \Gamma^*h_{0X}} \\
&= \frac{1}{1 - \Gamma^*h_{0X}} \left[ \left(h_{0a} - \frac{n}{2}\right) \frac{\varphi^\lambda h_{1X}}{M^2 - 2M + 2} + \frac{\Gamma^*h_{1X}h_{1a}}{M} \right] \\
\varphi_1^\theta &= 2h_{1\theta} + \frac{2h_{0\theta}\Gamma^*h_{1X}}{1 - \Gamma^*h_{0X}}.
\end{aligned}$$

Q.E.D.

**Proof of Proposition 20.** We have

$$P^{D2D} = d - \hat{X} \quad \text{and} \quad \hat{X} = \Psi \cdot \tilde{\chi}.$$

To first order,

$$\begin{aligned}
\frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} &\approx \frac{\Gamma^*}{M} + \frac{\varphi^\lambda}{M^2 - 2M + 2}(\lambda_l - \lambda^*) \\
\sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} &\approx \Gamma^* \\
\sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \alpha_l &\approx \Gamma^* \bar{\alpha} + \frac{M\varphi^\lambda \lambda^*}{M^2 - 2M + 2} \alpha_{mismatch} \\
\sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} (\lambda_l - \lambda^*) &\approx 0 \\
\sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \alpha_l (\lambda_l - \lambda^*) &\approx \lambda^* \Gamma^* \alpha_{mismatch}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\Psi \cdot \tilde{\chi} &= -(\varphi_0^a - 0.5n)\lambda^* \left[ \Gamma^* \bar{\alpha} + \frac{M\varphi^\lambda \lambda^*}{M^2 - 2M + 2} \alpha_{mismatch} \right] \\
&\quad - (\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) M \lambda^* \Gamma^* \bar{\alpha} - M \lambda^* \Gamma^* [(\varphi_0^A - 0.5n\gamma\eta_0^* \lambda^*) + \varphi_{0,\lambda}^A \lambda^*] \hat{\alpha} \\
&\quad + \lambda^* \Gamma^* [\varphi_0^a + M\varphi_0^A + n\eta_0^*] d + \Gamma^* \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Gamma^* \Theta \\
&\quad - (\varphi_0^a + \varphi_1^a \lambda^* - 0.5n) \lambda^* \Gamma^* \alpha_{mismatch} \\
&= -\lambda^* \Gamma^* [\varphi_0^a + M\varphi_0^A + n\eta_0^*] (\bar{\alpha} + \alpha_{mismatch}) - (\lambda^*)^2 \left[ \Gamma^* (M\varphi_{0,\lambda}^A + \varphi_1^a) + \frac{M\varphi^\lambda (\varphi_0^a - 0.5n)}{M^2 - 2M + 2} \right] \alpha_{mismatch} \\
&\quad + \lambda^* \Gamma^* [\varphi_0^a + M\varphi_0^A + n\eta_0^*] d + \Gamma^* \varphi_0^\theta \bar{\theta} - \gamma\eta_0^* \lambda^* \Gamma^* \Theta.
\end{aligned}$$

Q.E.D.

## G Fragmented Double Auction Model/Double Auction Benchmark

There are a few existing theoretical papers on two-tiered OTC markets, with different assumptions than our model. Our model tracks real-world FX markets and its predictions are supported by empirical evidence. However, it remains a question whether or not existing models would make the same predictions as ours. We examine this question in this Appendix and show that our model is essential in understanding the dynamics of exchange rates. In particular, we show that a leading two-tiered OTC model, developed by [Babus and Parlatore \(2018\)](#), yields predictions that are not supported by the data from the FX market. This result

points to the importance of specific the market structure used in each market, as the model in [Babus and Parlatores \(2018\)](#) was not developed for the specific markets we consider here.

We will not reproduce the model of [Babus and Parlatores \(2018\)](#) here. The key difference between their model and ours is the request for quote mechanism in our model (and the heterogeneity in the Dealer's holding cost/risk aversion). In their model, customers exogenously trade with only one dealer, thus there is no competition in liquidity provision between dealers in the D2C market.

We now re-derive the predictions about price dynamics in a two-tiered OTC market where, instead of the request-for-quote mechanism, customers can only trade with one dealer.

In this appendix, we modify the model of our paper along the lines of [Babus and Parlatores \(2018\)](#) and assume that customers are split into  $K$  groups of equal size of  $L$  so that  $n = KL$ . Each group only trades with a single dealer. This drastically simplifies the problem because it eliminates strategic competition in liquidity provision between dealers in the D2C market.

The derivation of the equilibrium follows that of the main model and we will provide fewer details.

**Lemma 26** *Dealer number  $l$  trades with  $n$  customers with risk aversions  $\gamma$ . Hence, his post-D2C trade inventory is given by*

$$\chi_l = (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} \mathbf{Q}_l^{D2C} + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} \beta_l^{D2C} \tilde{x}_l.$$

The inter-dealer market does not change, and we can write down the dealer utility as

$$U_l = -(\chi_l - x_l) \mathcal{P}_l^{D2C} + \chi_l d - 0.5 \Gamma_l \chi_l^2 + \frac{1}{2} E \left[ \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \mathcal{B}^{-1} \sum_{\ell=1}^M (\Gamma_\ell + \beta_\ell)^{-1} \Gamma_\ell \chi_\ell - \Gamma_l \chi_l \right)^2 \right]$$

and we can rewrite this utility as

$$U_l = -(\chi_l - x_l) \mathcal{P}_l^{D2C} + (d + Z_{-l}) \chi_l - 0.5 \Gamma_l^{D2C} \chi_l^2 + \text{const},$$

where

$$\Gamma_l^{D2C} = \Gamma_l - \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left[ \left( \frac{1}{\mathcal{B}(\Gamma_l + \beta_l)} - 1 \right) \Gamma_l \right]^2 \approx \frac{2\Gamma^*}{M} + \frac{2}{M^2 - 2M + 2} (\Gamma_l - \Gamma^*),$$

and

$$Z_{-l} = \Delta_l \sum_{\ell \neq l} \frac{\Gamma_\ell}{\mathcal{B}(\Gamma_\ell + \beta_\ell)} E[\chi_\ell]$$

where we have defined

$$\Delta_l = \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \frac{1}{\mathcal{B}(\Gamma_l + \beta_l)} - 1 \right) \Gamma_l \approx -\frac{M-2}{M-1} - \frac{M(M-2)}{\Gamma(M-1)(M^2-2M+2)} (\Gamma_l - \Gamma^*),$$

Given this quadratic objective, the optimal demand schedule for the dealer in the D2C market is given by

$$Q_l(\mathcal{P}_l^{D2C}) = (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}(d + Z_{-l} - \mathcal{P}_l^{D2C} - \Gamma_l^{D2C} x_l)$$

**Lemma 27** *We have*

$$\begin{aligned} \mathcal{B}_l^{D2C} &= \frac{4nM\gamma\Gamma_l^{D2C}(n^2 - 3n + 1) + 4n^3(n - 2)(\Gamma_l^{D2C})^2 - (2n - 1)M^2\gamma^2}{4n(n - 1)(\gamma M + 2n\Gamma_l^{D2C})\gamma\Gamma_l^{D2C}} \\ &\quad + \frac{(M\gamma + 2n^2\Gamma_l^{D2C})\sqrt{\gamma^2 M^2(1 - 2n)^2 + 4Mn(2n^2 - 3n + 2)\gamma\Gamma_l^{D2C} + 4(n - 2)^2 n^2 (\Gamma_l^{D2C})^2}}{4n(n - 1)(\gamma M + 2n\Gamma_l^{D2C})\gamma\Gamma_l^{D2C}} \\ \beta_l^{D2C} &= \frac{2\Gamma_l^{D2C}}{\Gamma_l^{D2C}\mathcal{B}_l^{D2C} - 2 + \sqrt{(\Gamma_l^{D2C}\mathcal{B}_l^{D2C})^2 + 4}} \\ \beta_{c,l}^{D2C} &= \frac{2\gamma}{\gamma\mathcal{B}_l^{D2C} - 2 + \sqrt{(\gamma\mathcal{B}_l^{D2C})^2 + 4}}. \end{aligned}$$

Now, we obtain the post-D2C round inventory

$$\chi_l = x_l + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}(d + Z_{-l} - \mathcal{P}_l^{D2C} - \Gamma_l^{D2C} x_l)$$

and the price

$$\mathcal{P}_l^{D2C} = d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}Z_{-l} - Q_l^{D2C}$$

where

$$Q_l^{D2C} = (\mathcal{B}_l^{D2C})^{-1} \left( (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} \Gamma_l^{D2C} x_l + (\gamma + \beta_{c,l}^{D2C})^{-1} \gamma \sum_{i=1}^n \theta_i \right).$$

Note that

$$Z_{-l} = \Delta_l(\bar{Z} - \mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1}\Gamma_l E[\chi_l]) \quad (22)$$

where we have defined

$$\bar{Z} = \sum_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell} E[\chi_{\ell}] \quad (23)$$

Taking expectations and substituting, we get

$$E[\chi_l] = x_l + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}(d + Z_{-l} - E[\mathcal{P}_l^{D2C}] - \Gamma_l^{D2C} x_l) \quad (24)$$

First, we compute the expected price:

$$\begin{aligned}
E[\mathcal{P}_l^{D2C}] &= d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}E[Z_{-l}] - E[Q_l^{D2C}] \\
&= d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}E[Z_{-l}] \\
&\quad - E\left[(\mathcal{B}_l^{D2C})^{-1}\left((\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}\Gamma_l^{D2C}x_l + (\gamma + \beta_{c,l}^{D2C})^{-1}\gamma\sum_{i=1}^n\theta_i\right)\right] \\
&= d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}Z_{-l} \\
&\quad - (\mathcal{B}_l^{D2C})^{-1}\left((\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}\Gamma_l^{D2C}x_l + (\gamma + \beta_{c,l}^{D2C})^{-1}\gamma nE[\theta]\right)
\end{aligned} \tag{25}$$

Substituting (25) into (24), we get

$$\begin{aligned}
E[\chi_l] &= x_l + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}\left(d + Z_{-l} - \left(d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}Z_{-l}\right.\right. \\
&\quad \left.\left. - (\mathcal{B}_l^{D2C})^{-1}\left((\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}\Gamma_l^{D2C}x_l + (\gamma + \beta_{c,l}^{D2C})^{-1}\gamma nE[\theta]\right)\right) - \Gamma_l^{D2C}x_l\right) \\
&= C_l^Z Z_{-l} + C_l^x x_l + C_l^\theta E[\theta]
\end{aligned}$$

where

$$\begin{aligned}
C_l^Z &= \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \left[1 - \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})}\right] \\
C_l^x &= 1 + \frac{\Gamma_l^{D2C}}{\Gamma_l^{D2C} + \beta_l^{D2C}} \left[-1 + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})}\right] \\
C_l^\theta &= \frac{n\gamma}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})(\gamma + \beta_{c,l}^{D2C})}.
\end{aligned}$$

Now, we are ready to solve for  $E[\chi_l]$ . We have from (22) that

$$E[\chi_l] = C_l^Z \Delta_l(\bar{Z} - \mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1}\Gamma_l E[\chi_l]) + C_l^x x_l + C_l^\theta E[\theta]$$

which gives

$$E[\chi_l] = (1 + C_l^Z \Delta_l \mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1}\Gamma_l)^{-1} (C_l^Z \Delta_l \bar{Z} + C_l^x x_l + C_l^\theta E[\theta])$$

Using (23), we get

$$\bar{Z} = \sum_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1}\Gamma_{\ell}(1 + C_{\ell}^Z \Delta_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1}\Gamma_{\ell})^{-1} (C_{\ell}^Z \Delta_{\ell} \bar{Z} + C_{\ell}^x x_{\ell} + C_{\ell}^\theta E[\theta]) ,$$

which is equivalent to

$$\begin{aligned}
\bar{Z} &= \left( 1 - \sum_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell} (1 + C_{\ell}^Z \Delta_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell})^{-1} C_{\ell}^Z \Delta_{\ell} \right)^{-1} \\
&\quad \left( \sum_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell} (1 + C_{\ell}^Z \Delta_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell})^{-1} (C_{\ell}^x x_{\ell} + C_{\ell}^{\theta} E[\theta]) \right) \\
&= \left( 1 - \sum_{\ell} [1 + (C_{\ell}^Z \Delta_{\ell} \mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell})^{-1}]^{-1} \right)^{-1} \\
&\quad \left( \sum_{\ell} ((\mathcal{B}^{-1}(\Gamma_{\ell} + \beta_{\ell})^{-1} \Gamma_{\ell})^{-1} + C_{\ell}^Z \Delta_{\ell})^{-1} (C_{\ell}^x x_{\ell} + C_{\ell}^{\theta} E[\theta]) \right) \\
&= \bar{C}_{\bar{Z}}^x \cdot x + \bar{C}_{\bar{\theta}}^{\bar{Z}} E[\theta]
\end{aligned}$$

where we have defined

$$\begin{aligned}
Dem_{\bar{Z}} &= 1 - \sum_{\ell} \frac{1}{1 + \frac{1}{C_{\ell}^Z \Delta_{\ell} \frac{\Gamma_{\ell}}{\mathcal{B}(\Gamma_{\ell} + \beta_{\ell})}}} \\
\bar{C}_{\bar{Z},l}^x &= \frac{1}{Dem_{\bar{Z}}} \left( \left( \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \right)^{-1} + C_l^Z \Delta_l \right)^{-1} C_l^x \\
\bar{C}_{\bar{Z},l}^{\bar{\theta}} &= \frac{1}{Dem_{\bar{Z}}} \left( \left( \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \right)^{-1} + C_l^Z \Delta_l \right)^{-1} C_l^{\theta} \\
\bar{C}_{\bar{Z}}^x &= (\bar{C}_{\bar{Z},1}^x, \bar{C}_{\bar{Z},2}^x, \dots, \bar{C}_{\bar{Z},M}^x)^T \\
\bar{C}_{\bar{Z}}^{\bar{\theta}} &= \sum_{\ell} \bar{C}_{\bar{Z},\ell}^{\bar{\theta}}.
\end{aligned}$$

We proceed with linking the prices in D2D and D2C markets. We have

$$\begin{aligned}
\chi_l &= x_l + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} (d + Z_{-l} - \mathcal{P}_l^{D2C} - \Gamma_l^{D2C} x_l) \\
&= (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} (d + Z_{-l} - \mathcal{P}_l^{D2C}) + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} \Gamma_l^{D2C} x_l
\end{aligned}$$

We need to invert  $(x_l)$  from the vector of  $\mathcal{P}_l^{D2C}$ . We have

$$\begin{aligned}
Z_{-l} &= \Delta_l \left( \bar{Z} - \mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1} \Gamma_l (1 + C_l^Z \Delta_l \mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1} \Gamma_l)^{-1} (C_l^Z \Delta_l \bar{Z} + C_l^x x_l + C_l^\theta E[\theta]) \right) \\
&= \Delta_l \left( \bar{Z} - \left( \frac{1}{\mathcal{B}^{-1}(\Gamma_l + \beta_l)^{-1} \Gamma_l} + C_l^Z \Delta_l \right)^{-1} (C_l^Z \Delta_l \bar{Z} + C_l^x x_l + C_l^\theta E[\theta]) \right) \\
&= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{Z} - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^x x_l - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^\theta \bar{\theta} \\
&= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l (\bar{C}_Z^x \cdot x + \bar{C}_\theta^{\bar{Z}} \bar{\theta}) \\
&\quad - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^x x_l - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^\theta \bar{\theta} \\
&= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l (\bar{C}_Z^x \cdot x) - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^x x_l \\
&\quad + \left[ \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{C}_\theta^{\bar{Z}} - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^\theta \right] \bar{\theta} \\
&= C_x^{Z_{-l}} \cdot x + C_{x_l}^{Z_{-l}, l} x_l + C_{\bar{\theta}}^{Z_{-l}, l} \bar{\theta},
\end{aligned}$$

where we have defined

$$\begin{aligned}
C_x^{Z_{-l}, l} &= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{C}_{Z, l}^x; & C_x^{Z_{-l}} &= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{C}_Z^x; \\
C_{x_l}^{Z_{-l}, l} &= - \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^x; & C_{\bar{\theta}}^{Z_{-l}, l} &= \frac{\Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \left[ \mathcal{B}(\Gamma_l + \beta_l) \bar{C}_\theta^{\bar{Z}} - \Gamma_l C_l^\theta \right].
\end{aligned}$$

Returning to the D2C price for Dealer  $l$ , we have



$$\begin{aligned}
\mathcal{P}_l^{D2C} &= d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}Z_{-l} - Q_l^{D2C} \\
&= d + (\mathcal{B}_l^{D2C})^{-1}(\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}Z_{-l} - (\mathcal{B}_l^{D2C})^{-1} \left( (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1}\Gamma_l^{D2C}x_l + (\gamma + \beta_{c,l}^{D2C})^{-1}\gamma \sum_{i=1}^n \theta_i \right) \\
&= d + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_x^{Z_{-l}} \cdot x + C_{x_l}^{Z_{-l},l}x_l + C_{\bar{\theta}}^{Z_{-l}}\bar{\theta}) \\
&\quad - \left( \frac{\Gamma_l^{D2C}}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})}x_l + \frac{\gamma}{\mathcal{B}_l^{D2C}(\gamma + \beta_{c,l}^{D2C})} \sum_{i=1}^n \theta_i \right) \\
&= d + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_x^{Z_{-l}} \cdot x) + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_{x_l}^{Z_{-l},l}x_l) - \frac{\Gamma_l^{D2C}}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})}x_l \\
&\quad + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_{\bar{\theta}}^{Z_{-l}}\bar{\theta}) - \frac{\gamma}{\mathcal{B}_l^{D2C}(\gamma + \beta_{c,l}^{D2C})} \sum_{i=1}^n \theta_i \\
&= d + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_x^{Z_{-l}} \cdot x) + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} [C_{x_l}^{Z_{-l},l} - \Gamma_l^{D2C}]x_l \\
&\quad + \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} (C_{\bar{\theta}}^{Z_{-l}}\bar{\theta}) - \frac{\gamma}{\mathcal{B}_l^{D2C}(\gamma + \beta_{c,l}^{D2C})} \sum_{i=1}^n \theta_i \\
&= \text{where } d + C_x^{\mathcal{P},l} \cdot x + C_{x_l}^{\mathcal{P},l}x_l + C_{\Theta}^{\mathcal{P},l}\Theta + C_{\bar{\theta}}^{\mathcal{P},l}\bar{\theta}, \\
&\quad C_{x,l}^{\mathcal{P}} = \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} C_x^{Z_{-l},l}; \quad C_x^{\mathcal{P},l} = \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} C_x^{Z_{-l}}; \\
&\quad C_{x_l}^{\mathcal{P},l} = \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} [C_{x_l}^{Z_{-l},l} - \Gamma_l^{D2C}]; \\
&\quad C_{\Theta}^{\mathcal{P},l} = -\frac{\gamma}{\mathcal{B}_l^{D2C}(\gamma + \beta_{c,l}^{D2C})}; \quad C_{\bar{\theta}}^{\mathcal{P},l} = \frac{1}{\mathcal{B}_l^{D2C}(\Gamma_l^{D2C} + \beta_l^{D2C})} C_{\bar{\theta}}^{Z_{-l},l}.
\end{aligned}$$

It follows that

$$\mathcal{P}^{D2C} = d\mathbf{1} + \mathcal{A}^x x + \mathcal{V}^{\Theta}\Theta + \mathcal{V}^{\bar{\theta}}\bar{\theta},$$

with

$$\mathcal{A}^x = \mathcal{A}_0^x + \text{diag}(C_{x_l}^{\mathcal{P},l}); \quad \mathcal{A}_0^x = \sum_{\ell=1}^M e_{\ell} \otimes C_x^{\mathcal{P},\ell}; \quad \mathcal{V}^{\Theta} = \left( C_{\Theta}^{\mathcal{P},\ell} \right)_{\ell=1}^M; \quad \mathcal{V}^{\bar{\theta}} = \left( C_{\bar{\theta}}^{\mathcal{P},\ell} \right)_{\ell=1}^M.$$

Inverting, we get

**Lemma 28** *We have*

$$x = (\mathcal{A}^x)^{-1} \left[ \mathcal{P}^{D2C} - d\mathbf{1} - \mathcal{V}^{\Theta}\Theta - \mathcal{V}^{\bar{\theta}}\bar{\theta} \right].$$

Now we can write the link between D2D and D2C prices:

$$\begin{aligned}
\mathcal{P}^{D2D} &= d - \mathcal{B}^{-1} \sum_{l=1}^M (\Gamma_l + \beta_l)^{-1} \Gamma_l \chi_l \\
&= d - \mathcal{B}^{-1} \sum_{l=1}^M (\Gamma_l + \beta_l)^{-1} \Gamma_l \left( (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} (d + Z_{-l} - \mathcal{P}_l^{D2C}) + (\Gamma_l^{D2C} + \beta_l^{D2C})^{-1} \Gamma_l^{D2C} x_l \right) \\
&= \left[ 1 - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \right] d + \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \mathcal{P}_l^{D2C} \\
&\quad - \mathcal{B}^{-1} \sum_{l=1}^M (\Gamma_l + \beta_l)^{-1} \Gamma_l \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} [Z_{-l} + \Gamma_l^{D2C} x_l] \\
&= \left[ 1 - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \right] d + \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \mathcal{P}_l^{D2C} \\
&\quad - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \left[ C_x^{Z_{-l}} \cdot x + C_{x_l}^{Z_{-l}, l} x_l + C_{\bar{\theta}}^{Z_{-l}, l} \bar{\theta} + \Gamma_l^{D2C} x_l \right] \\
&= \left[ 1 - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \right] d + \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \mathcal{P}_l^{D2C} \\
&\quad - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} C_x^{Z_{-l}} \cdot x - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} (C_{x_l}^{Z_{-l}, l} + \Gamma_l^{D2C}) x_l \\
&\quad - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} C_{\bar{\theta}}^{Z_{-l}, l} \bar{\theta} \\
&= \hat{C}_d^{\mathcal{P}, D2D} d + \hat{C}_{\mathcal{P}}^{\mathcal{P}, D2D} \mathcal{P}^{D2C} + \hat{C}_x^{\mathcal{P}, D2D} x + \hat{C}_{\bar{\theta}}^{\mathcal{P}, D2D} \bar{\theta},
\end{aligned}$$

where we defined

$$\begin{aligned}
\tilde{C}_x^{\mathcal{P},D2D} &= - \left( \frac{\Gamma_1}{\mathcal{B}(\Gamma_1 + \beta_1)} \frac{C_{x_1}^{Z_{-1},1} + \Gamma_1^{D2C}}{\Gamma_1^{D2C} + \beta_1^{D2C}}, \dots, \frac{\Gamma_M}{\mathcal{B}(\Gamma_M + \beta_M)} \frac{C_{x_M}^{Z_{-M},M} + \Gamma_M^{D2C}}{\Gamma_M^{D2C} + \beta_M^{D2C}} \right)^T \\
\hat{C}_d^{\mathcal{P},D2D} &= 1 - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \\
\hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} &= \left( \frac{\Gamma_1}{\mathcal{B}(\Gamma_1 + \beta_1)} \frac{1}{\Gamma_1^{D2C} + \beta_1^{D2C}}, \dots, \frac{\Gamma_M}{\mathcal{B}(\Gamma_M + \beta_M)} \frac{1}{\Gamma_M^{D2C} + \beta_M^{D2C}} \right)^T \\
\hat{C}_x^{\mathcal{P},D2D} &= \tilde{C}_x^{\mathcal{P},D2D} - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} C_x^{Z_{-l}} \\
\hat{C}_{\bar{\theta}}^{\mathcal{P},D2D} &= -\mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} C_{\bar{\theta}}^{Z_{-l},l}.
\end{aligned}$$

Substituting

$$x = (\mathcal{A}^x)^{-1} \left[ \mathcal{P}^{D2C} - d\mathbf{1} - \nu^\Theta \Theta - \nu^{\bar{\theta}} \bar{\theta} \right],$$

we will get a linear relationship between  $\mathcal{P}^{D2D}$  and the vector of  $(\mathcal{P}_l^{D2C})_l$ :

$$\begin{aligned}
\mathcal{P}^{D2D} &= \hat{C}_d^{\mathcal{P},D2D} d + \hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} \mathcal{P}^{D2C} + \hat{C}_x^{\mathcal{P},D2D} x + \hat{C}_{\bar{\theta}}^{\mathcal{P},D2D} \bar{\theta} \\
&= \hat{C}_d^{\mathcal{P},D2D} d + \hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} \mathcal{P}^{D2C} + \hat{C}_x^{\mathcal{P},D2D} (\mathcal{A}^x)^{-1} \left[ \mathcal{P}^{D2C} - d\mathbf{1} - \nu^\Theta \Theta - \nu^{\bar{\theta}} \bar{\theta} \right] + \hat{C}_{\bar{\theta}}^{\mathcal{P},D2D} \bar{\theta} \\
&= \left[ \hat{C}_d^{\mathcal{P},D2D} - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \mathbf{1} \right] d + \left[ \hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} + ((\mathcal{A}^x)^{-1})^T \hat{C}_x^{\mathcal{P},D2D} \right] \mathcal{P}^{D2C} \\
&\quad + \left[ \hat{C}_{\bar{\theta}}^{\mathcal{P},D2D} - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \nu^{\bar{\theta}} \right] \bar{\theta} - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \nu^\Theta \Theta \\
&= C_d^{\mathcal{P},D2D} d + C_{\mathcal{P}}^{\mathcal{P},D2D} \cdot \mathcal{P}^{D2C} + C_{\Theta}^{\mathcal{P},D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P},D2D} \bar{\theta},
\end{aligned}$$

where we have defined

$$\begin{aligned}
C_d^{\mathcal{P},D2D} &= \hat{C}_d^{\mathcal{P},D2D} - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \mathbf{1} \\
C_{\mathcal{P}}^{\mathcal{P},D2D} &= ((\mathcal{A}^x)^{-1})^\top \hat{C}_x^{\mathcal{P},D2D} + \hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} \\
C_{\bar{\theta}}^{\mathcal{P},D2D} &= \hat{C}_{\bar{\theta}}^{\mathcal{P},D2D} - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \nu^{\bar{\theta}} \\
C_{\Theta}^{\mathcal{P},D2D} &= - (\hat{C}_x^{\mathcal{P},D2D})^T (\mathcal{A}^x)^{-1} \nu^\Theta.
\end{aligned}$$

**Lemma 29** *For small heterogeneity in risk aversion, we have*

$$C_{\mathcal{P}}^{\mathcal{P},D2D} = \bar{\pi}^* \mathbf{1} + \pi^{(1)} (\Gamma - \Gamma^*) + O(\|\Gamma - \Gamma^*\|^2),$$

where  $\mathbf{1} = (1, 1, \dots, 1)^T$ ,  $\mathbf{\Gamma} = (\Gamma_\ell)_{\ell=1}^M$ , and both  $\bar{\pi}^*$  and  $\pi^{(1)}$  are constant.

**Proof.** It is straightforward to show that, for small heterogeneity in risk aversion, we have

$$C_{\mathcal{P}}^{\mathcal{P}, D2D} = \bar{\Pi}^* + \Pi^{(1)}(\mathbf{\Gamma} - \mathbf{\Gamma}^*) + O(\|\mathbf{\Gamma} - \mathbf{\Gamma}^*\|^2),$$

where  $\bar{\Pi}^*$  is a vector and  $\Pi^{(1)}$  is a matrix, and both are constant. Equilibrium considerations imply that

$$\begin{aligned}\bar{\Pi}^* &= \bar{\pi}^* \mathbf{1} \\ \Pi^{(1)} &= a \mathbf{1} \otimes \mathbf{1} + \pi^{(1)} \text{Id}.\end{aligned}$$

for scalars  $a, \bar{\pi}^*$ , and  $\pi^{(1)}$ . Moreover,

$$\mathbf{1} \otimes \mathbf{1}(\mathbf{\Gamma} - \mathbf{\Gamma}^*) = \mathbf{0}.$$

The result then follows. Q.E.D.

The natural proxy for the (half) bid-ask spread in the D2C market is the price impact that customers have when trading with a particular dealer, given by  $\beta_{c,l}^{D2C}$ . By direct calculation,

$$\beta_{c,l}^{D2C} = \beta_{c,*}^{D2C} + \beta_c^{(1)}(\Gamma_l - \Gamma^*) + O(\|\mathbf{\Gamma} - \mathbf{\Gamma}^*\|^2).$$

The sign of  $\beta_c^{(1)}$  will play an important role later on. The proposition below follows directly the results in [Malamud and Rostek \(2017\)](#):

**Proposition 30** *We always have  $\beta_c^{(1)} > 0$ .*

Thus, we can write

$$\Gamma_l - \Gamma^* \approx (\beta_{c,l}^{D2C} - \beta_{c,*}^{D2C}) / \beta_c^{(1)}.$$

This formula allows us to express the latent risk aversion through the observable bid-ask spreads. Thus, we rewrite

$$\begin{aligned}\mathcal{P}^{D2D} &= C_d^{\mathcal{P}, D2D} d + C_{\mathcal{P}}^{\mathcal{P}, D2D} \cdot \mathcal{P}^{D2C} + C_{\Theta}^{\mathcal{P}, D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P}, D2D} \bar{\theta} \\ &\approx C_d^{\mathcal{P}, D2D} d + (\bar{\pi}^* \mathbf{1} + \pi^{(1)}(\mathbf{\Gamma} - \mathbf{\Gamma}^*)) \cdot \mathcal{P}^{D2C} + C_{\Theta}^{\mathcal{P}, D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P}, D2D} \bar{\theta} \\ &\approx C_d^{\mathcal{P}, D2D} d + \left( \bar{\pi}^* \mathbf{1} + \frac{1}{\beta_c^{(1)}} \pi^{(1)} ((\beta_{c,l}^{D2C})_{l=1}^M - \beta_{c,*}^{D2C}) \right) \cdot \mathcal{P}^{D2C} + C_{\Theta}^{\mathcal{P}, D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P}, D2D} \bar{\theta}.\end{aligned}$$

We have

$$\beta_{c,l}^{D2C} = \lambda_l^{-1}$$

and hence, up to second-order terms in the size of heterogeneity,

$$\begin{aligned} ((\beta_{c,l}^{D2C})_{l=1}^M - \beta_{c,*}^{D2C}) \cdot \mathcal{P}^{D2C} &= ((\beta_{c,l}^{D2C})_{l=1}^M - \beta_{c,*}^{D2C}) \cdot \alpha \, M \operatorname{Cov}(\lambda_l^{-1}, \alpha) \\ &\approx -M\lambda_*^{-2} \operatorname{Cov}(\lambda, \alpha) = -M\lambda_*^{-1} \alpha_{mismatch}. \end{aligned}$$

Hence,

$$\mathcal{P}^{D2D} \approx C_d^{\mathcal{P},D2D} d + \bar{\pi}^* \bar{\alpha} - M\lambda_*^{-1} (\pi^{(1)} / \beta_c^{(1)}) \alpha_{mismatch} + C_\Theta^{\mathcal{P},D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P},D2D} \bar{\theta}.$$

We know that  $\beta_c^{(1)}$  is positive, and hence we just need to figure out the sign of  $\pi^{(1)}$ . While we cannot solve for the sign of  $\pi^{(1)}$  analytically, we use extensive numerical analysis to show that:

**Numerical Claim:**

$$\pi^{(1)} > 0$$

when  $n$  is sufficiently large.

We now determine  $\pi^{(1)}$ . Recall from Lemma 29 that  $\pi^{(1)}$  is defined via the Taylor series expansion of  $C_{\mathcal{P}}^{\mathcal{P},D2D}$  and that

$$C_{\mathcal{P}}^{\mathcal{P},D2D} = ((\mathcal{A}^x)^{-1})^\top \widehat{C}_x^{\mathcal{P},D2D} + \widehat{C}_{\mathcal{P}}^{\mathcal{P},D2D},$$

with

$$\mathcal{A}^x = \mathcal{A}_0^x + \operatorname{diag}(C_{x_l}^{\mathcal{P},l}); \quad \mathcal{A}_0^x = \sum_{\ell=1}^M e_\ell \otimes C_x^{\mathcal{P},\ell}.$$

We examine the components of matrix  $\mathcal{A}^x$ , starting with the terms  $C_x^{\mathcal{P},l}$ .

$$C_x^{\mathcal{P},l} = \frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} C_x^{Z-l} = \frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{C}_Z^x.$$

It follows that

$$\mathcal{A}_0^x = \sum_{\ell=1}^M e_\ell \otimes C_x^{\mathcal{P},\ell} = v \otimes \bar{C}_Z^x$$

where

$$v = \left( \frac{1}{\mathcal{B}_\ell^{D2C} (\Gamma_\ell^{D2C} + \beta_\ell^{D2C})} \frac{\mathcal{B}(\Gamma_\ell + \beta_\ell)}{\mathcal{B}(\Gamma_\ell + \beta_\ell) + \Gamma_\ell \Delta_\ell C_\ell^Z} \Delta_\ell \right)_{\ell=1}^M$$

and recall that

$$\begin{aligned}
Dem_{\bar{Z}} &= 1 - \sum_{\ell} \frac{1}{1 + \frac{1}{C_{\ell}^Z \Delta_{\ell} \frac{1}{\mathcal{B}(\Gamma_{\ell} + \beta_{\ell})}}} \\
\bar{C}_{\bar{Z},l}^x &= \frac{1}{Dem_{\bar{Z}}} \left( \left( \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \right)^{-1} + C_l^Z \Delta_l \right)^{-1} C_l^x \\
\bar{C}_{\bar{Z}}^x &= (\bar{C}_{\bar{Z},1}^x, \bar{C}_{\bar{Z},2}^x, \dots, \bar{C}_{\bar{Z},M}^x)^T \\
C_l^x &= 1 + \frac{\Gamma_l^{D2C}}{\Gamma_l^{D2C} + \beta_l^{D2C}} \left[ -1 + \frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} \right]
\end{aligned}$$

In addition,

$$\begin{aligned}
\Delta_l &= \frac{\Gamma_l + 2\beta_l}{(\Gamma_l + \beta_l)^2} \left( \frac{1}{\mathcal{B}(\Gamma_l + \beta_l)} - 1 \right) \Gamma_l \\
C_l^Z &= \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \left[ 1 - \frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} \right].
\end{aligned}$$

We use all these expressions in the numerical work regarding  $\mathcal{A}_0^x$ . Now, we turn our attention to the term  $diag(C_{x_l}^{\mathcal{P},l})$ :

$$\begin{aligned}
C_{x_l}^{\mathcal{P},l} &= \frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} [C_{x_l}^{Z-l,l} - \Gamma_l^{D2C}] \\
&= -\frac{1}{\mathcal{B}_l^{D2C} (\Gamma_l^{D2C} + \beta_l^{D2C})} \left[ \frac{\Gamma_l \Delta_l}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} C_l^x + \Gamma_l^{D2C} \right].
\end{aligned}$$

We use these expressions to evaluate  $\mathcal{A}^x$  and thus

$$C_{\mathcal{P}}^{\mathcal{P},D2D} = ((\mathcal{A}^x)^{-1})^{\top} \hat{C}_x^{\mathcal{P},D2D} + \hat{C}_{\mathcal{P}}^{\mathcal{P},D2D},$$

where,

$$\begin{aligned}
\hat{C}_{\mathcal{P}}^{\mathcal{P},D2D} &= \left( \frac{\Gamma_\ell}{\mathcal{B}(\Gamma_\ell + \beta_\ell)} \frac{1}{\Gamma_\ell^{D2C} + \beta_\ell^{D2C}} \right)_{\ell=1}^M \\
\hat{C}_x^{\mathcal{P},D2D} &= \tilde{C}_x^{\mathcal{P},D2D} - \mathcal{B}^{-1} \sum_{l=1}^M \frac{\Gamma_l}{\Gamma_l + \beta_l} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} C_x^{Z-l} \\
\tilde{C}_x^{\mathcal{P},D2D} &= - \left( \frac{\Gamma_\ell}{\mathcal{B}(\Gamma_\ell + \beta_\ell)} \frac{C_{x_\ell}^{Z-\ell,\ell} + \Gamma_\ell^{D2C}}{\Gamma_\ell^{D2C} + \beta_\ell^{D2C}} \right)_{\ell=1}^M \\
C_x^{Z-l} &= \frac{\mathcal{B}(\Gamma_l + \beta_l)}{\mathcal{B}(\Gamma_l + \beta_l) + \Gamma_l \Delta_l C_l^Z} \Delta_l \bar{C}_Z^x \\
\hat{C}_x^{\mathcal{P},D2D} &= \tilde{C}_x^{\mathcal{P},D2D} - \left[ \sum_{l=1}^M \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \frac{1}{\Gamma_l^{D2C} + \beta_l^{D2C}} \frac{1}{1 + \frac{\Gamma_l}{\mathcal{B}(\Gamma_l + \beta_l)} \Delta_l C_l^Z} \Delta_l \right] \bar{C}_Z^x.
\end{aligned}$$

We are now in the position to evaluate  $\pi^{(1)}$  numerically. The exogenous parameters are (i) the number of customers  $n$ , (ii) the number of dealers  $M$ , (iii) the customers' coefficient of risk-aversion  $\gamma$ , and (iv) the dealers' average coefficient of risk-aversion  $\Gamma^*$ . Following [Vayanos and Woolley \(2013\)](#), we choose

$$\Gamma^* = 30.$$

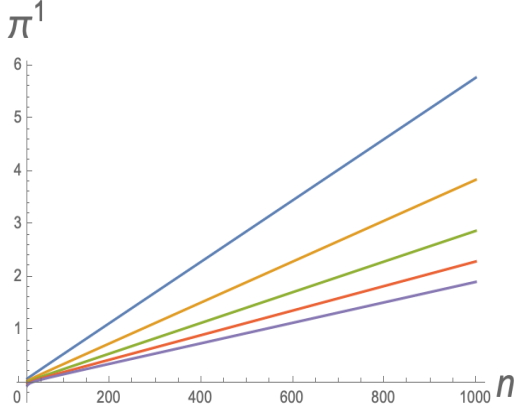
We define

$$\kappa = \frac{\gamma}{\Gamma^*},$$

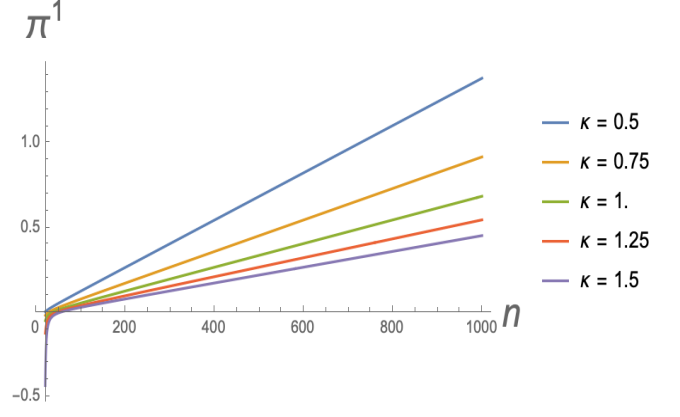
and use the values

$$\kappa \in \{0.5, 0.75, 1, 1.25, 1.5\}.$$

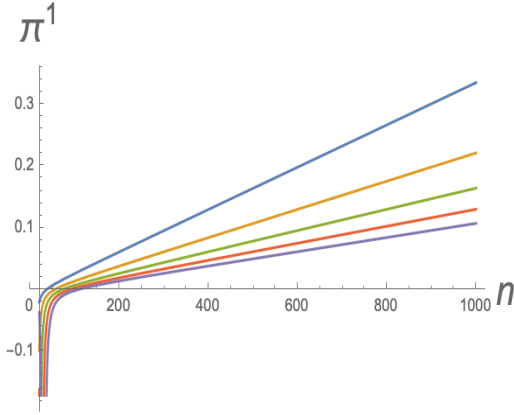
Figure 1 plots  $\pi^{(1)}$  as a function of  $n$ , for  $M \in \{5, 10, 20, 40\}$ . The figure shows that  $\pi^{(1)}$  is negative for  $n$  sufficiently large. In our data, the number of dealers  $M$  fluctuates between 8 and 12.



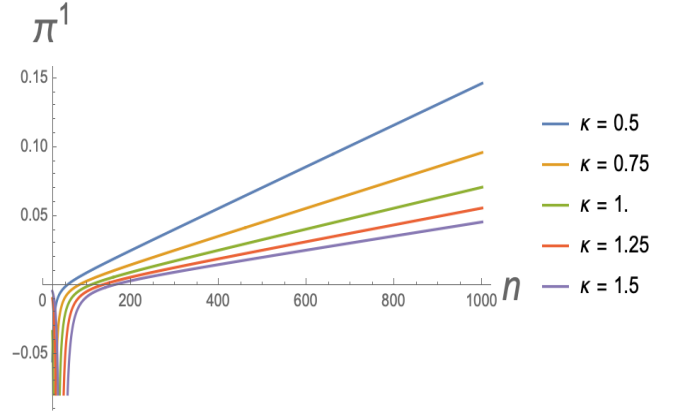
(a)  $M = 5$ .



(b)  $M = 10$ .



(c)  $M = 20$ .



(d)  $M = 30$ .

**Figure 1: Liquidity Mismatch in a Fragmented Double Auction Model** The relation between the D2D price and  $\alpha_{mismatch}$  in a market such as the one in [Babus and Parlato \(2018\)](#) is given by

$$\mathcal{P}^{D2D} \approx C_d^{\mathcal{P}, D2D} d + \bar{\pi}^* \bar{\alpha} - \frac{M \lambda_*^{-1}}{\beta_c^{(1)}} \pi^{(1)} \alpha_{mismatch} + C_{\Theta}^{\mathcal{P}, D2D} \Theta + C_{\bar{\theta}}^{\mathcal{P}, D2D} \bar{\theta},$$

where  $\lambda_*$  and  $\beta_c^{(1)}$  are positive. We plot  $\pi^{(1)}$  as a function of  $n$ . The other parameters are

$$\Gamma^* = 30 \quad \text{and} \quad \kappa = \frac{\gamma}{\Gamma^*}.$$