School of Finance

University of St.Gallen

«CONSTRAINED DEALERS AND MARKET EFFICIENCY»

WENQIAN HUANG
ANGELO RANALDO
ANDREAS SCHRIMPF
FABRICIUS SOMOGYI

WORKING PAPERS ON FINANCE NO. 2021/16

SWISS INSTITUTE OF BANKING AND FINANCE (S/BF - HSG)

NOVEMBER, 2021

Constrained Dealers and Market Efficiency*

Wenqian Huang,† Angelo Ranaldo,‡ Andreas Schrimpf,§ and Fabricius Somogyi¶

This version: November 4, 2021

Working paper and subject to change.

Abstract

We analyse how constraints on dealers' risk bearing capacity affect market efficiency in the foreign exchange (FX) market. Dealers support market efficiency by accommodating their customers' trading demands through elastic liquidity provision in normal times but when they face constraints their elasticity of liquidity provision weakens. Episodes of tight dealer constraints – for instance, due to high leverage, Value-at-Risk, and funding costs – in turn go hand in hand with price inefficiencies due to law of one price deviations and elevated trading costs. We rationalise our novel empirical findings with a tractable model that sheds light on the key mechanisms of how market efficiency can deteriorate when dealers are more constrained.

J.E.L. classification: F31, G12, G15

Keywords: Market Efficiency, Dealer Constraints, Foreign Exchange, Liquidity Provision.

*We welcome comments, including references to related papers we have inadvertently overlooked. We are grateful to Stijn Claessens and Semyon Malamud for helpful comments. We also thank seminar participants at the Bank for International Settlements. Angelo Ranaldo acknowledges financial support from the Swiss National Science Foundation (SNSF grant 182303). All errors are our own.

†Bank for International Settlements, Switzerland. E-mail: wenqian.huang@bis.org.

‡University of St.Gallen, Switzerland. E-mail: angelo.ranaldo@unisg.ch.

§Bank for International Settlements & CEPR, Switzerland. E-mail: andreas.schrimpf@bis.org.

¶University of St.Gallen, Switzerland. E-mail: fabricius.somogyi@unisg.ch.

1. Introduction

A market is efficient when an asset is traded at a unique price reflecting its fundamental value. Financial intermediaries play a crucial role in supporting the efficient functioning of financial markets, as they bring together parties that wish to trade and share risks. This is especially true for over-the-counter (OTC) markets, which are not organised based on centralised exchanges but rely on intermediation of dealers, who provide immediacy to their clients. However, the ability of dealers to provide market liquidity depends on their capacity to bear risks. Therefore, market efficiency may suffer potential setbacks in periods when dealers' risk-bearing capacity is impaired.[1]

Against this backdrop, the fundamental question we address in this paper is whether dealer constraints have adverse implications on market inefficiency by impeding the arbitrage process that leads to efficient price formation. To analyse this question, we study prices and volumes in the foreign exchange (FX) market based on a globally representative FX trade data set from CLS group. The FX market is the largest financial market in the world and often also regarded as one of the most liquid and efficient ones. As such, one may expect our findings about the key mechanisms to also apply to other important, yet less liquid, OTC markets.

The contribution of this paper is threefold. First, we provide a novel analytical method to identify and measure market (in)efficiency in spot FX markets that we label "price inefficiency metric" (PIM). The construction of PIM builds on the well-known triangular relation among exchange rates (e.g., Chaboud, Chiquoine, Hjalmarsson, and Vega, 2014; Foucault, Kozhan, and Tham, 2016) and is composed of two main frictions affecting price efficiency: the fundamental violation of the law of one price and the associated transaction cost that traders would incur to restore no-arbitrage conditions. Our second contribution is to document novel empirical findings connecting market efficiency to dealers' constraints. In particular, we show that when dealers are unconstrained they support market efficiency accommodating their clients' trading demands. Thus, the forces of arbitrage can freely operate through dealers' elastic liquidity supply and eliminate any fundamental mispricing. However, when dealers are constrained – i.e., when they face tightening Value-at-Risk, leverage or funding conditions – their willingness to provide liquidity to their clients decreases. Consequently, the relation between their intermediated volume and our price inefficiency metric breaks down. Lastly, we provide a tractable model that formalises the intuition on how market efficiency deteriorates when markets are more volatile and when dealers are more constrained.

A good understanding of the functioning of OTC markets – especially those that, like the FX market, are crucial to the transmission of monetary policy – is important for a variety of reasons. For central banks, regulators, and financial stability agencies, the question is particularly relevant in light of the post-crisis efforts to strengthen the resilience of OTC markets and the intermediaries that support their functioning.

[1] See "Holistic Review of the March Market Turmoil," Financial Stability Board, November 2020.

Our study sheds light on the interplay between the efficiency of the pricing mechanism and the liquidity in OTC markets on the one hand as well as the financial health of their intermediaries on the other. From an academic perspective, an important tenet in financial economics is that ample market liquidity begets price efficiency. However, prior research has highlighted that frictions faced by financial intermediaries can hamper the arbitrage process.[2] Our study provides new evidence on how constraints faced by dealers impede their liquidity provision and as a consequence adversely affect market efficiency.

Our paper consists of three main parts. The starting point of our analysis is to propose a novel analytical method for measuring market inefficiency: the price inefficiency metric (PIM). The conceptual underpinning of PIM is a no-arbitrage relation that ties together triplets of currency pairs. Every triplet comprises one non-dollar currency pair (e.g., AUDJPY) and two USD legs (e.g., USDAUD and USDJPY) that are required to synthetically replicate the non-dollar pair. Specifically, we decompose PIM into two components: i) the fundamental violation of the law of one price ($VLOOP$), and ii) the transaction costs that traders would incur in triangular arbitrage transactions ($TCOST$). Put differently, VLOOP captures fundamental mispricing in the sense of price dislocations in the triangular no-arbitrage relation across currency crosses, whereas TCOST is the cumulative transaction cost that an arbitrageur would face to eliminate any fundamental mispricing. The main advantages of PIM are threefold: i) it is model- and preference-free, ii) it can be derived from a simple accounting identity, and iii) it is grounded in the principle of no-arbitrage.

Based on our FX trade data from CLS, we document two novel empirical results. First, PIM tends to move up when volatility is higher and FX dealers are more constrained. Second, VLOOP and TCOST exhibit strong commonality but VLOOP is an order of magnitude smaller than TCOST. In addition to confirming the virtues of our methodology, these results suggest that the FX spot market is efficient in normal times, given that price dislocations are relatively small and arbitraging them away is most of the time not profitable for the average trader due to prohibitively high transaction costs. Our results further indicate the important role of dealers in supporting market efficiency, as their intermediation is no longer connected to efficient pricing when they are constrained.

In the second part, we go one step further by linking our price inefficiency metric to dealer-intermediated FX trading volumes. To do so, we mostly rely on so-called logistic smooth transition regression (LSTAR) methods that are particularly well-suited to capture shifts in relations across different regimes. The main goal of our analysis is to investigate how the relation between dealer-intermediated volume and market efficiency changes across two regimes: i) normal times when dealers are largely unconstrained, and ii) stressed periods when dealers face more tightening economic conditions. To capture such regime shifts we construct a composite measure that aggregates a multitude of constraints impacting dealers' liquidity provision. Specifically, we define our time-varying "dealer constraint measure"

[2]See Gromb and Vayanos (2010) for an excellent survey of the literature.

DCM as the first principal component of the top 10 FX dealer banks' i) Value-at-Risk measure (measured at quarterly frequency), ii) He, Kelly, and Manela (2017) leverage ratio (quarterly), iii) credit default swap (CDS) premia (daily), and iv) funding costs (daily) that are particularly relevant for debt-financed positions and market-making functions (see, e.g. Andersen, Duffie, and Song, 2019). To this set of variables we also add the daily JP Morgan Global FX Volatility index, which can be thought of as an analogue of the VIX index in the FX market.

Two novel findings arise from this regime-dependent analysis: First, dealer-intermediated trading volume (VLM) and our price inefficiency metric (PIM) co-move in tandem (correlation of 67% on average between the two measures). In other words, when trading volumes spike we also observe higher volatility and a weakening of market efficiency. This positive correlation holds for each of the two components of PIM: deviations from fundamentals and arbitrage transaction costs. Second, we find that the relation between PIM and VLM significantly weakens when dealers become more constrained. Although both VLM and PIM are on average at least 21–29% higher than normal in the constrained regime, the correlation between the two is insignificant or even negative in periods when dealers are constrained. We interpret this result as the elasticity of dealer banks' liquidity provision, which decreases by up to 70-80% in the constrained regime. While dealer-intermediated volume goes up as dealers accommodate transaction demands by end-users, they also increase bid-ask spreads, rendering it more difficult for arbitrageurs to correct fundamental mispricing. As such, while intermediated FX volumes tend to rise in the constrained regime, this rise seems to be insufficient to compensate for the deterioration in market efficiency. All these findings also hold individually for each price inefficiency component, that is, for deviations from fundamentals (VLOOP) and arbitrage transaction costs (TCOST). In addition, they remain qualitatively unchanged when considering each determinant of the dealer constraint measure as a single regime variable rather than using the composite measure.

In the third part of the paper we rationalise our novel empirical findings with a simple partial equilibrium model. Building on standard microstructure theory (see e.g., Grossman and Miller, 1988; Hendershott and Menkveld, 2014), the model features two periods, three currency pairs, and two types of agent: i) a risk-averse and debt-financed dealer supplying liquidity and ii) liquidity traders with exogenous trading demands. Heterogeneity in private values in the latter group results in demand imbalances across the three currency pairs (Gabaix and Maggiori, 2015), which in turn generate price pressure. Such pressure pushes the exchange rates of the three currency pairs away from their fundamental value, resulting in deviations from the law of one price. However, in equilibrium such deviations are difficult to arbitrage away due to the transaction costs stemming from bid-ask spreads. In other words, the spreads are a manifestation of limits to arbitrage (Shleifer and Vishny, 1997). These two conceptual sources of market inefficiency in our model manifest themselves empirically via the price inefficiency metric PIM. When the dealer is unconstrained, she is able to intermediate among liquidity traders. These unconstrained periods coincide with small spreads and small deviations of exchange rates from the fundamental value, resulting in a small PIM.

When the dealer becomes more constrained, both the spreads and the deviations increase along with a more inelastic liquidity supply by the dealer, leading to a higher PIM.

Our simple model provides an economic reasoning to the novel empirical findings on the positive correlation between PIM and dealer-intermediated volume. When volatility is high, the larger dispersion of private values of the liquidity traders pushes up their demand and the net imbalance. The high volatility, coupled with the large net imbalance gives rise to two effects: First, it leads dealers to increase their quoted spreads, and second it generates deviations of exchange rates from the fundamental value. The combination of these two forces pushes PIM higher. In the case when the dealer is unconstrained, she is capable of providing liquidity and thereby allowing arbitrageurs to channel capital to take advantage of price discrepancies. Dealer-intermediated volume and PIM are hence positively correlated. Thus, such a positive correlation captures the elasticity of liquidity provision. It represents how intermediated market depth (i.e., dealer-intermediated volume) responds to pricing dislocations (i.e., PIM). In an efficient OTC market where liquidity provision is elastic, market depth should increase in pricing dislocations. However, when the dealer is constrained, the spread she charges can be prohibitively high, deterring some market participants from trading. Intermediated volume therefore will still increase but at a slower pace, which in turn mutes the correlation between dealer-intermediated volume and PIM. This attenuated relation between the two variables in turn reflects a drop in dealers' elasticity of liquidity provision.

As suggested by the model, both PIM and dealer-intermediated volume are equilibrium outcomes. Hence, to overcome the classic problem that regressing price inefficiency on trading volume might lead to biased estimates, we use a granular instrumental variable (GIV) approach. Following the work by Gabaix and Koijen (2020), we construct our GIV as the difference between the size- and equal-weighted average of daily dealer-intermediated trading volume. Clearly, when there is not enough cross-sectional granularity in the data, then this approach may not work. However, this does not concern our setup since trading volume exhibits a high degree of heterogeneity in the cross-section of currency pairs. Indeed, the inspection of the main idiosyncratic events identified by the GIV analysis provides a supportive narrative. Overall, we find consistent estimates for the elasticity of PIM with respect to dealer-intermediated volume using GIV as an instrument for the latter.

We contribute to three strands of literature. First, our work is related to the broader literature that emphasises the role of intermediary frictions in affecting asset prices (e.g., Gârleanu and Pedersen, 2011; He and Krishnamurthy, 2011, 2013; Adrian and Boyarchenko, 2012; Adrian, Etula, and Muir, 2014; He et al., 2017) that stresses the importance of constrained financial intermediaries for market outcomes and in particular for asset prices and risk premia. Our main contribution is to show how constrained financial intermediaries reduce their liquidity provision in times of markets stress by raising bid-ask spreads and curtailing volume. This finding is consistent with the idea that market-wide liquidity conditions depend on intermediaries' balance sheet (Adrian and Shin, 2010) and that intermediary leverage and

banks' risk management practices (e.g. following Value-at-Risk methodologies) tend to be pro-cyclical Adrian and Shin (2013). Lastly, our findings pointing to price inefficiency associated with inelastic liquidity provision are consistent with slow-moving intermediated capital causing distortions in pricing relations (Duffie, 2010).

Second, we add to the literature on limits to arbitrage. While prior research has mostly focused on constrained *arbitrageurs* (e.g., Shleifer and Vishny, 1997; Gromb and Vayanos, 2002; Hombert and Thesmar, 2014), we study the role of constrained *dealers* and how their ability to provide immediacy contributes to market inefficiency. Our key contribution is to show that the relation between price inefficiency and trading volume critically depends on the risk-bearing capacity of dealers. In addition, a large body of prior research has studied limits to arbitrage in equity markets (see Gromb and Vayanos, 2010). However, many of the frictions, such as short sale constraints which are considered a major explanation in equities (e.g., Chu, Hirshleifer, and Ma, 2020), do not apply to FX. Related to the stock market literature, recent studies document widespread mispricings in stressful conditions (Pasquariello, 2014), commonality in arbitrage deviations (Rösch, Subrahmanyam, and van Dijk, 2016), and limits to arbitrage impacting market liquidity (Rösch, 2021). We add to this branch of the literature by identifying constrained intermediaries as the main driving force behind such commonalities and by showing how their liquidity provision impacts arbitrage trading activity.

Lastly, we contribute to the literature on FX microstructure on understanding the role of trading volume. Our key angle here is to study the relation of both quantities (i.e., information on trading volume from CLS Group) and prices. In contrast to the FX order flow literature (e.g., Evans, 2002; Evans and Lyons, 2002, 2005), the literature on trading volume is relatively scarce due to the lack of comprehensive data sets. Earlier research has focused largely on the inter-dealer segment, which is dominated by two platforms: Reuters (e.g., Evans, 2002; Payne, 2003; Foucault et al., 2016) and EBS (e.g., Chaboud, Chernenko, and Wright, 2008; Mancini, Ranaldo, and Wrampelmeyer, 2013; Chaboud et al., 2014). Alternative sources of spot FX trading volume are proprietary data sets from specific dealer banks (e.g., Bjønnes and Rime, 2005; Menkhoff, Sarno, Schmeling, and Schrimpf, 2016). The relatively recent public access to CLS data has enabled researchers to study customer-dealer volume at a global scale (Hasbrouck and Levich, 2018, 2021; Ranaldo and Santucci de Magistris, 2018; Cespa, Gargano, Riddiough, and Sarno, 2021; Ranaldo and Somogyi, 2021). We contribute to this strand of literature by investigating the impact of constrained financial intermediaries' liquidity provision on FX market efficiency and its components (i.e., price dislocations and arbitrage transaction costs).

The remainder of the paper is structured as follows. Section 2 describes our data and key variables. Section 3 contains the main empirical analysis and further robustness checks. Section 4 outlines a simple model of how market efficiency depends on dealer constraints. Section 6 concludes. A separate Online Appendix provides all proofs, additional results, and robustness checks not reported in the paper.

2. Measuring market (in)efficiency in FX markets

2.1. Data sources

Note that the data excludes any trades between two market makers or two price takers and hence only includes trading activity that is intermediated by FX dealer banks. This data set is publicly available from CLS or via Quandl.com, a financial and economic data provider. CLS data have been used in prior research, among others, by Fischer and Ranaldo (2011), Hasbrouck and Levich (2018, 2021), Ranaldo and Santucci de Magistris (2018), Cespa et al. (2021), and Ranaldo and Somogyi (2021). The full sample period spans from November 2011 to September 2020 and includes data for 18 major currencies and 33 currency pairs.

Our goal is to construct measures of market efficiency that we derive from triangular no-arbitrage trades involving one non-dollar currency pair (e.g., AUDJPY) and two dollar legs (i.e., USDAUD and USDJPY). Hence, we exclude the USDHKD, USDILS, USDKRW, USDMXP, USDSGD, and USDZAR from our sample because there are no further non-dollar currency pairs involving the respective quote currencies (i.e., HKD, ILS, KRW, MXP, SGD, and ZAR). Furthermore, to maintain a balanced panel, we also remove all currency pairs involving the Hungarian forint (HUF), which enters the data set later, on 7 November 2015.[3]

We pair the hourly FX volume and order flow data with intraday spot bid and ask quotes from Olsen, a well-known provider of high-frequency data. Olsen compiles historical tick-by-tick data from various electronic trading platforms, both from the inter-dealer and dealer-customer segments. The indicative bid and ask quotes are directly available for all 25 currency pairs but do not correspond to actually executable transaction prices. This is not an issue for our purposes for two reasons: First, we are interested in measuring price efficiency rather than identifying actual triangular arbitrage opportunities in the global spot FX market. Second, on average, the correlation of Olsen indicative quotes with tradeable EBS best bid and offer prices is around 99% and the mean absolute error is roughly 3.3%.[4] Eventually, notice that both CLS and Olsen data are sampled at the hourly frequency.

2.2. Key variables

We explore market efficiency in the FX market using a new "price inefficiency metric," or in short PIM. Our measure encompasses two dimensions: i) violations of the law of one price (VLOOP), and ii) arbitrage transaction cost (TCOST). As a first step, we explain how

[3]This filtering leaves us 15 non-dollar currency pairs (i.e., AUDJPY, AUDNZD, CADJPY, EURAUD, EURCAD, EURCHF, EURDKK, EURGBP, EURJPY, EURNOK, EURSEK, GBPAUD, GBPCAD, GBPCHF, and GBPJPY) and 10 dollar pairs (i.e., USDAUD, USDCAD, USDCHF, USDDKK, USDEUR, USDGBP, USDJPY, USDNOK, USDNZD, and USDSEK) that are used to synthetically replicate each of the non-dollar pairs.

[4]To be precise, we estimate correlations and mean absolute errors individually for 25 currency pairs over the full-year of 2016. For brevity, we relegate these results to the Online Appendix.

we derive the two components and then elaborate how they jointly form the comprehensive price inefficiency metric PIM.

VLOOP captures the price dispersion for two assets or trading positions with the same intrinsic value, while TCOST refers to the trading costs to arbitrage such violations. Conceptually, we derive both measures from the well-known triangular arbitrage trade that takes advantage of three exchange rates (e.g., Chaboud et al., 2014; Foucault et al., 2016). The VLOOP component of the triangular arbitrage trade can be computed with midquote prices reflecting the intrinsic values of the direct and indirect positions. The TCOST can be measured with the bid and ask quotes (depending on the base and quote currency).

There are two notable advantages of this decomposition that lead to our price inefficiency measure. First, it is model- as well as preference-free and second, it can be derived from a simple accounting identity.

Deriving PIM from the triangular no-arbitrage relation. To derive VLOOP, consider a trader exchanging one euro (EUR) to some amount of US dollar (USD), exchanging the amount of US dollar to some amount of Canadian dollar (CAD) and exchanging back the amount of Canadian dollar to euro instantaneously at time t. The final amount of such a round-trip transaction measured in euro is given as:

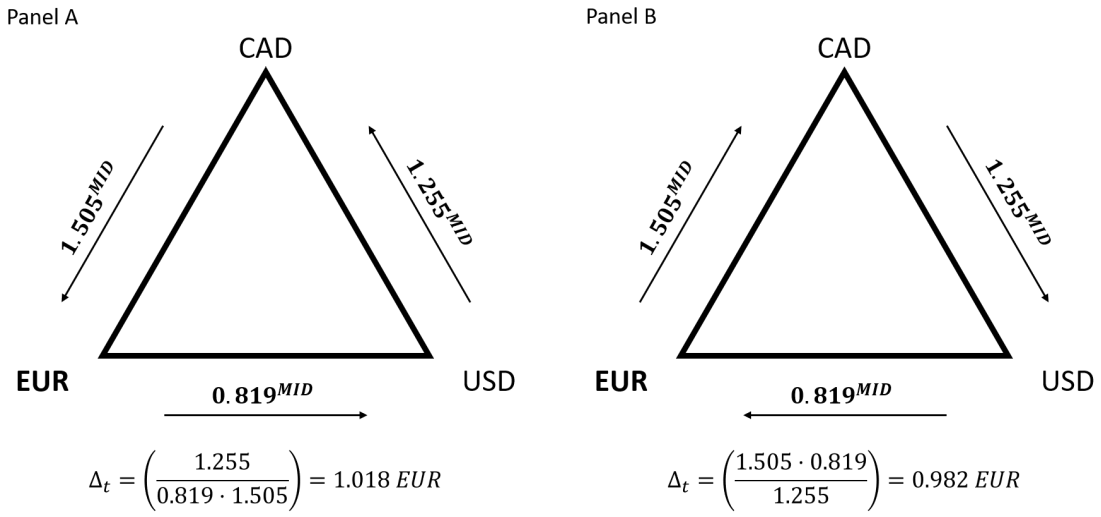
$$\Delta_t \equiv \prod_{i=1}^{3} P_{i,t}, \tag{1}$$

where $P_{1,t} = \frac{1}{USDEUR_t^{mid}}$, $P_{2,t} = USDCAD_t^{mid}$, and $P_{3,t} = \frac{1}{EURCAD_t^{mid}}$ denote midquote exchange rates expressed as the amount of quote currency per unit of base currency, for instance, 1.255 CAD per USD (i.e., indirect quotation).

The trader has identified a seemingly profitable violation of the law of one price if Δ_t is greater than unity. Note that if Δ_t is less than one, then the trader could still try to go into the opposite direction, that is, first exchange one unit of Canadian dollar to euro, and then subsequently euro to US dollar and then US dollar to euro. Note that Δ_t is independent of the currency that an arbitrageur is endowed with at the starting point of the round-trip (i.e., CAD, EUR or USD). Figure 1 provides a schematic overview of how to detect seemingly profitable deviations from no-arbitrage conditions.

To derive TCOST, we consider the same trader as above but now incorporate transaction costs by accounting for bid-ask spreads. Specifically, for every transaction that a trader makes, she pays the midquote price plus the half-spread. To reflect this, we replace the midquote prices in Eq. (1) by bid and ask prices, that is, $P_{1,t} = \frac{1}{USDEUR_t^{ask}}$, $P_{2,t} = USDCAD_t^{bid}$, and $P_{3,t} = \frac{1}{EURCAD_t^{ask}}$, respectively. The superscripts 'bid' and 'ask' refer to the price at which someone sells and buys one currency for another currency. Clearly, a positive Δ_t corresponds to a true arbitrage opportunity as the deviations from the fundamental value are large enough

Figure 1: Identifying a triangular arbitrage opportunity



Note: This figure provides a schematic overview of two triangular arbitrage strategies, where the arrows denote the direction. Panel A shows the hypothetical profit of a trader starting with one euro, first exchanging it to $\frac{1}{0.819} = 1.221$ US dollars, then exchanging 1.221 US dollars to Canadian dollars at the midquote price of 1.255 Canadian dollars per US dollar. This yields 1.532 Canadian dollars that are exchanged back to euros at the CADEUR midquote that is equivalent to $\frac{1}{EURCAD^{MID}} = \frac{1}{1.505}$. Such a round trip yields 1.018 euros or equivalent a positive return of 1.8% in this example. Panel B embraces the same logic but going the opposite direction, that is, first from euro to Canadian dollar, to US dollar and then back to euro yielding a negative return of -1.8%.

relative to the incurred transaction costs.[5] Figure 2 provides an overview of such a triangular arbitrage trade when including bid and ask prices. Note that the bid and ask prices in this example are just for illustrative purposes and do not correspond to actual data.

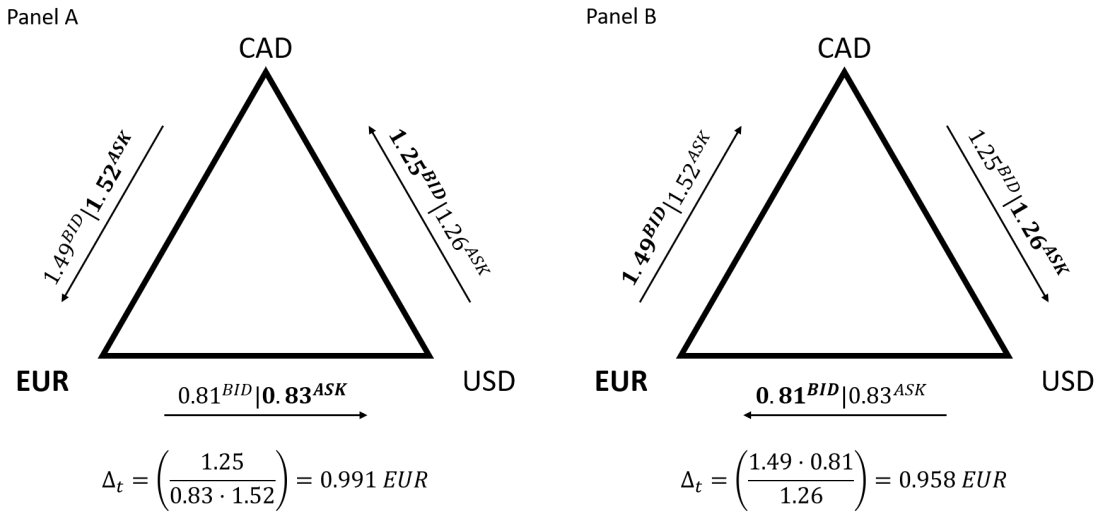
The last step in the derivation of the price inefficiency metric consists of taking the log on both sides of Eq. (1), and exploiting the fact that bid and ask prices are the midquote plus half the bid-ask spread. This yields the following expression:

$$log(\Delta_t) \equiv \underbrace{log\Big(\frac{USDCAD_t^{mid}}{USDEUR_t^{mid} \cdot EURCAD_t^{mid}}\Big)}_{VLOOP_t} + \underbrace{log\Big(\frac{1 - \frac{USDCAD_t^{bas}}{2}}{(1 + \frac{USDEUR_t^{bas}}{2}) \cdot (1 + \frac{EURCAD_t^{bas}}{2})}\Big)}_{TCOST_t}, \quad (2)$$

where the superscripts 'mid' and 'bas' denote the midquote price (i.e., the average of the bid and ask price) and relative bid-ask spread (i.e., the difference between ask and bid price relative to the midquote), respectively. The first part of Eq. (2) (i.e., $VLOOP_t$) captures the violations from the law of one price, whereas the second part (i.e., $TCOST_t$) captures the cumulative arbitrage transaction costs of performing such a triangular arbitrage trade. No-

[5]It is worth to emphasise at this point that the bid and ask prices that we retrieve from Olsen are indicative quotes and hence *not* executable. In our sample, such seemingly profitable opportunities occur on average with a relative frequency of less than 0.2 percent.

Figure 2: Triangular arbitrage trade with transaction costs



Note: This figure provides a schematic overview of two triangular arbitrage strategies, where the arrows denote the direction. Panel A shows the profit of first exchanging one euro to $\frac{1}{0.83} = 1.21$ US dollars at the ask price, then exchanging 1.21 US dollars to Canadian dollars at the bid price of 1.25 Canadian dollars per US dollar. This yields 1.51 Canadian dollars that are exchanged back to euros at the CADEUR bid price that is equivalent to $\frac{1}{EURCAD^{ASK}} = \frac{1}{1.52}$. Such a round trip yields 0.991 euros or equivalent a negative return of -0.9%. Panel B embraces the same logic but going the opposite direction, that is, first from euro to Canadian dollar, to US dollar and then back to euro.

arbitrage violations $VLOOP_t$ may be positive or negative depending on the direction of the trade (see Figure 1) but will be identical in absolute terms (if measured in logs) irrespective of the initial endowment of the trader (i.e., CAD, EUR or USD). On the contrary, arbitrage transaction costs $TCOST_t$ are by nature always negative and correspond to the minimum required return that an arbitrageur would have to earn for the triangular no-arbitrage trade to break even (see Figure 2). Thus, $TCOST_t$ possesses a natural interpretation of being a measure of "limits to arbitrage" (Shleifer and Vishny, 1997). Note that $TCOST_t$ is contingent on the direction of the trade (see Figure 2).

Based on this decomposition, we derive an encompassing price inefficiency measure (PIM) that we define as the sum of VLOOP and the absolute value of TCOST:

$$PIM_t = E_t[VLOOP_t + abs(TCOST_t)|VLOOP_t > 0], \tag{3}$$

where $abs()$ denotes the absolute value and PIM_t is computed conditional on VLOOP being positive. Hence, PIM_t is by construction always positive, with a higher reading of PIM coinciding with greater inefficiency. Following our methodology, we compute $VLOOP_t$, $TCOST_t$, and PIM_t for $k = 1, 2, ..., 15$ triplets of currency pairs. A triplet of currency pairs is defined as one non-dollar currency pair (e.g., EURCAD) plus the two USD legs (e.g., USDEUR and USDCAD). We prune the hourly time-series of $VLOOP_t$, $TCOST_t$, and PIM_t, respectively, for

heavy outliers, which we define as observations in the top and bottom 1.5 percentiles of the data. Since the signs of $VLOOP_t$, $TCOST_t$, and PIM_t are always the same, we hereinafter focus on *absolute values* to support the economic interpretation of these measures. Eventually, we can also compute daily measures of $VLOOP_t$, $TCOST_t$, and PIM_t by summing up hourly observations for each day. Specifically, we interpret $VLOOP_t$ as the daily cumulative deviations from the fundamental value, $TCOST_t$ as a measure of daily liquidity frictions, and PIM_t as the daily composite price inefficiency measure in the FX market.

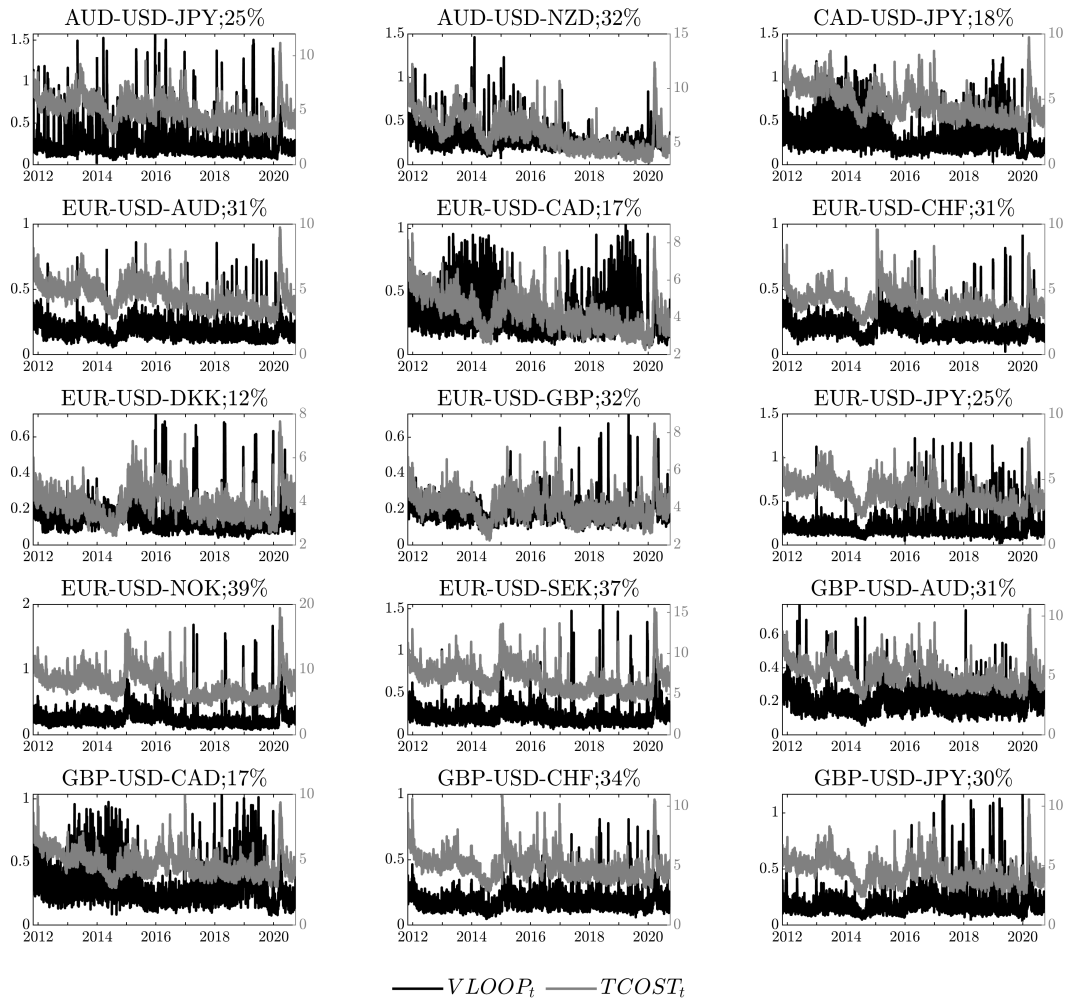
Empirical illustrations. Figure 3 shows the time-series and cross-sectional variation of hourly no arbitrage violations $VLOOP_t$ (left y-axis) and arbitrage transactions costs $TCOST_t$ (right y-axis), respectively. Both measures of market efficiency exhibit intuitive properties in the sense that they surge during periods of market stress and mean-revert during calm periods. The large spike during the Covid-19 market turmoil in March and April 2020 is particularly well pronounced across all 15 triplets of currency pairs and is indicative of the global nature of the stress. The correlation of $VLOOP_t$ and $TCOST_t$ is positive for the entire cross-section and ranges from 12–39%. We interpret this as evidence of commonality in no-arbitrage violations and market liquidity in the broader sense (Rösch, 2021).

Table 1 reports the time-series average of hourly triangular no arbitrage deviations VLOOP and arbitrage transaction costs TCOST. In addition, it also tabulates hourly averages of direct trading volume in non-dollar currency pairs (e.g., AUDJPY) and synthetic trading volume in dollar pairs (e.g., USDAUD and USDJPY). By "synthetic" we refer to the average trading volume in two dollar currency pairs within a currency pair triplet, for instance, AUDJPY, USDAUD, and USDJPY, that we abbreviate as AUD-USD-JPY. Each row corresponds to one triplet of currency pairs.

This simple summary statistics table conveys two key messages: First, deviations from fundamentals VLOOP are an order of magnitude smaller than arbitrage transaction costs TCOST. We interpret this result as a sign of overall market efficiency given that a seemingly profitable violation of triangular no-arbitrage can oftentimes not be exploited by the average trader because transactions costs are prohibitively high. Second, trading volume in non-dollar currency pairs is considerably smaller relative to the synthetic volume in dollar currency pairs.[6] This is essentially the case for all 15 triplets of currency pairs but the effect is less pronounced for the ones involving the NOK and SEK, where the euro crosses play a bigger role. Consequently, the average relative bid-ask spread and realised volatility are somewhat smaller in the more heavily traded dollar currency pairs than in non-dollar pairs.

[6]Somogyi (2021) provides both theoretical and empirical evidence in favour of the idea that strategic complementarity in price impact can explain these cross-sectional differences between dollar and non-dollar pairs.

Figure 3: No-arbitrage violations and arbitrage transaction costs



Note: This figure plots the absolute value of 24-hour moving average of hourly triangular no arbitrage deviations $VLOOP_t$ (left y-axis) and arbitrage trading costs $TCOST_t$ (right y-axis), respectively, for 15 triplets of currency pairs. Both variables are measured in basis points. The numbers in the titles refer to the correlation coefficient of $VLOOP_t$ and $TCOST_t$. The sample covers the period from 1 November 2011 to 30 September 2020.

3. Elasticity of liquidity provision and dealer constraints

This section presents evidence consistent with our hypothesis that dealers support market efficiency during calm periods but that their liquidity provision does not keep up sufficiently with trading demands when they are constrained. The analysis is split into two parts. We first start with some motivating evidence that the relationship between intermediated volumes and price efficiency depends on dealers' financial health. We then draw on logistic smooth transition regressions to study more rigorously the state-dependent nature of this relation.

Table 1: Summary statistics

	PIM in bps		Volume in $bn		Bid-ask spread in bps		Volatility in bps	
	VLOOP	TCOST	Direct	Synthetic	Direct	Synthetic	Direct	Synthetic
AUD-USD-JPY	0.24	4.88	0.18	2.55	4.15	2.93	14.38	11.21
AUD-USD-NZD	0.29	5.85	0.09	1.00	4.44	3.72	9.32	12.70
CAD-USD-JPY	0.30	4.67	0.03	2.66	4.29	2.61	12.66	9.47
EUR-USD-AUD	0.19	4.52	0.14	3.86	3.54	2.82	11.54	10.97
EUR-USD-CAD	0.28	4.25	0.08	3.97	3.55	2.49	10.15	9.12
EUR-USD-CHF	0.21	3.98	0.37	3.38	2.62	2.71	6.38	9.50
EUR-USD-DKK	0.14	3.89	0.09	3.09	2.54	2.65	1.82	9.22
EUR-USD-GBP	0.19	4.07	0.61	4.08	3.19	2.48	9.52	9.62
EUR-USD-JPY	0.21	3.90	0.65	4.83	3.14	2.42	11.43	9.69
EUR-USD-NOK	0.26	7.69	0.24	3.13	6.25	4.70	11.01	11.95
EUR-USD-SEK	0.23	6.86	0.27	3.13	5.41	4.21	9.18	11.28
GBP-USD-AUD	0.20	5.08	0.04	1.80	4.22	2.99	12.53	11.26
GBP-USD-CAD	0.29	4.69	0.03	1.91	4.00	2.67	10.85	9.41
GBP-USD-CHF	0.19	4.94	0.03	1.32	4.09	2.88	10.69	9.89
GBP-USD-JPY	0.19	4.47	0.20	2.77	3.85	2.59	12.78	9.99

Note: This table reports the time-series average of hourly triangular no arbitrage deviations $VLOOP$ in basis points (bps), cumulative trading costs $TCOST$ in bps, direct trading volume in non-dollar currency pairs (e.g., AUDJPY) in $bn, synthetic trading volume in dollar pairs (e.g., the average across USDAUD and USDJPY) in $bn, as well as direct and synthetic relative bid-ask spreads and realised volatility in dollar and non-dollar currency pairs in bps, respectively. Each row corresponds to a triplet of currency pairs, for example, AUDJPY, USDAUD, and USDJPY that we abbreviate as AUD-USD-JPY. The sample covers the period from 1 November 2011 to 30 September 2020.

3.1. Motivating evidence

For motivational purposes, we first present some descriptive evidence of how trading volume and price inefficiency co-move over time. In a second step, we then compute conditional correlations to show how the relation weakens as FX dealer banks become more constrained.

To measure dealer banks' risk-bearing capacity we derive a composite dealer constraint measure (DCM) that we compute in two steps. First, we create four time-series based on cross-sectional averages of the top 10 FX dealer banks' (see Euromoney FX surveys) i) value-at-risk of the trading book (measured at quarterly frequency), ii) He et al. (2017) leverage ratio (quarterly), iii) credit default swap (CDS) premia (daily) and iv) debt funding costs (daily). To this set of variables we also add the daily JP Morgan Global FX Volatility index as a measure of FX market stress. See the Online Appendix for a detailed description of how we retrieve and compute each of these variables.

All of these factors can typically generate important constraints on dealer behaviour. For example, self-imposed or regulatory-driven VaR-limits force dealers to scale back their market making or proprietary trading. Similarly, their need for to reduce their market making activity

and liquidity provision is linked to their risk profile reflected in high leverage and CDS premia. High risk exposure can lead to an immediate increase in funding costs and valuation adjustments (XVA), including debt and funding value adjustments (Andersen et al., 2019). The latter two are both a measure of dealers' funding liquidity that can in turn induce dealers to curtail their intermediation activities.

Since FX dealers typically operate on a global scale and provide liquidity in many currencies (and other asset classes) at once, they are even more exposed to these issues. In this setting, intermediaries might be forced to reduce liquidity provision in *all* currency pairs when they endure trading losses (Kyle and Xiong, 2001) and/or experience funding constraints in a specific position (Cespa and Foucault, 2014).

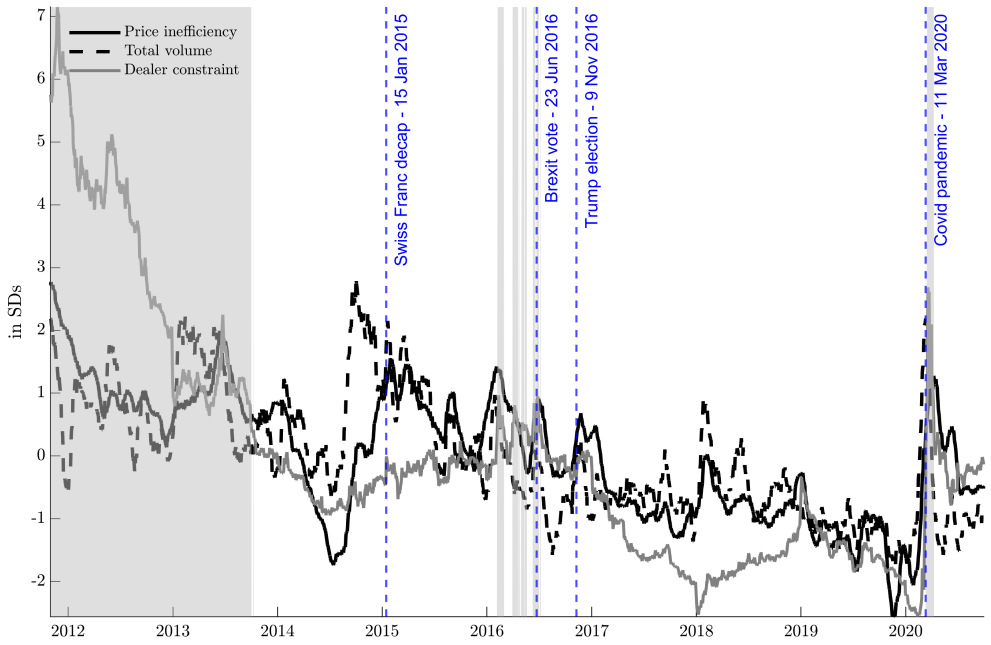
Second, we extract the first principal component that explains around 72% of the total variance and serves as our composite measure of dealer constraints. Hence, the key advantage of our dealer constraint measure DCM is that it encompasses a range of different factors that can all impact dealers' risk appetite and willingness to warehouse risk. By doing so, we are able to extract common information of all these factors and obtain a measure of the financial health of global dealers available at the daily frequency.

Figure 4 plots our price inefficiency measure PIM and total trading volume VLM against our dealer constraint measure DCM. For ease of illustration, we show the cross-sectional average of PIM and VLM across triplets of currency pairs. There are two key takeaways from this figure: First, price inefficiency and trading volume covary positively and exhibit a correlation of 67%. In Section 4 we provide a simple theory that can explain why the contemporaneous relationship between trading volume and price inefficiency is positive in a well-functioning market under normal conditions. This result directly stems from the fact that both volume and price inefficiency positively hinge on the volatility of exchange rates, which (as we will show later based on our model) in turn reflects investors' disagreement about fundamentals. Second, the correlation between trading volume VLM and price inefficiency PIM tends to weaken whenever there is a large spike in the dealer constraint measure DCM (grey shaded areas).

Table 2 takes the descriptive analysis one step further by computing conditional correlation coefficients of (log) changes in price inefficiency PIM, trading volume VLM, and realised volatility VOL across the quantiles of the dealer constraint measure DCM. Consistent with our intuition, we find that the correlation weakens as our dealer constraint measure DCM and/or realised volatility increases. For instance, the conditional correlation based on the highest DCM decile (i.e., when dealers are most constrained) is a mere 17%, and hence economically and statistically significantly lower than the full-sample correlation of 31%. We also observe an almost monotonic increase in the price inefficiency measure across the DCM quantiles.

These initial results suggest that market efficiency deteriorates when dealers are more constrained. Taken together, we interpret this as circumstantial evidence that in constrained

Figure 4: Market efficiency, dealer constraints, and intermediated volume



Note: This figure plots the cross-sectional average of our price inefficiency measure (PIM, black solid line) against total trading volume (VLM, black dashed line) in units of standard deviations. Both time-series correspond to 22-day moving averages. The grey line plots our dealer constraint measure (DCM) that we define as the first principal component of the top 10 FX dealers' (based on the Euromoney FX survey) quarterly Value-at-Risk measure (VaR), quarterly He et al. (2017) leverage ratio (HKM), daily credit default spread (CDS), daily debt funding cost (DFC), as well as the daily JP Morgan Global FX Volatility index (VXY). The grey shaded areas correspond to times when DCM exceeds its 75% quantile. The sample covers the period from 1 November 2011 to 30 September 2020.

periods there is less trading volume than what would be price efficient.

3.2. Smooth transition regression

We next briefly describe our econometric approach that is based on a simple smooth transition regression (LSTAR) model (e.g., van Dijk, Teräsvirta, and Franses, 2002; Christiansen, Ranaldo, and Söderlind, 2011). Let $G(z_{t-1})$ be a logistic function depending on the 1-day lagged regime variable z_{t-1}

$$G(z_{t-1}) = (1 + exp(-\gamma'(z_{t-1} - c)))^{-1}, \tag{4}$$

Table 2: Price inefficiency characteristics across top DCM quantiles

Top DCM quantile		$cor(PIM, VLM)$	t-stat	Mean(PIM) in %	Mean(VLM) in $bn	Mean(VOL) in %	#Obs
least constrained	1.00	0.31		1.18	148.66	0.56	2,185
	0.90	0.31	0.12	1.20	150.73	0.57	1,957
	0.80	**0.29	2.03	1.24	154.06	0.59	1,728
	0.70	***0.29	3.13	1.28	158.29	0.61	1,501
	0.60	***0.26	5.95	1.32	160.40	0.63	1,272
	0.50	***0.25	6.45	1.34	159.70	0.63	1,043
	0.40	***0.24	7.49	1.36	163.30	0.63	815
	0.30	***0.17	12.19	1.39	169.44	0.63	586
	0.20	***0.19	9.04	1.42	177.17	0.60	358
most constrained	0.10	***0.17	6.34	1.52	180.63	0.65	129

Note: This table shows the conditional correlation coefficient of our price inefficiency measure PIM and total trading volume VLM across the top quantiles of the dealer constraint measure (DCM, column 1). The underlying data are based on a panel of 15 currency pair triplets. The asterisks *, **, and *** indicate that the correlation is significantly different from the full sample (in the first line) estimate at the 90%, 95%, and 99% levels. Column 3 reports the corresponding test statistic for the conditional correlation cor^{τ} being equal to the full sample correlation $cor^{\tau=1.00}$, where $\tau \in 0.1, 0.2, ..., 0.9$ corresponds to top DCM_t quantiles. The test statistic for equal correlations is based on the Fisher z-transformation. Columns 4 to 7 report the average $PIM_{k,t}$ (Mean(PIM)) in %, the average $VLM_{k,t}$ (Mean(VLM)) in $bn, the average realised volatility (Mean(VOL)) in % in dollar pairs, and the average number of observations (#Obs). The full sample covers the period from 1 November 2011 to 30 September 2020.

where the parameter c is the central location and the vector γ determines the steepness of $G(z_{t-1})$. Hence, the LSTAR model is of the form

$$y_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}, \tag{5}$$

where the dependent variable is a measure of market (in)efficiency (i.e., VLOOP, TCOST or PIM) and $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors. We include both cross-sectional α_k and time-series λ_t fixed effects to control for any unobservable heterogeneity that is constant across triplets of currency pairs k and time t, respectively. For estimation, we use the generalised method of moments (GMM) and determine the optimal parameters γ and c by nonlinear least squares minimising the concentrated sum of squared errors.[7] Note that the slope coefficients in Eq. (5) vary smoothly with the regime variable z_{t-1} from β_1 at low values of $\gamma' z_{t-1}$ to β_2 at high values of $\gamma' z_{t-1}$. There are two interesting boundary cases: First, if $\beta_1 = \beta_2$ we effectively have a linear regression. Second, the limit case where $\gamma \to \infty$ is equivalent to a linear regression with a dummy variable.

The state-dependent explanatory variable $f_{k,t}$ is the total trading volume (i.e., $VLM_{k,t}$) that is defined as the sum of all trading volume in one non-dollar as well as two dollar currency pairs within within a particular currency pair triplet k. The state-independent variable is the realised variance (i.e., $RV_{k,t}$) in the direct non-dollar currency pair (e.g., AUDJPY) that we

[7] Our inference is based on a Driscoll and Kraay (1998) covariance matrix that allows for random clustering and serial correlation up to 8 lags. We choose the optimal bandwidth using the plug-in procedure for automatic lag selection by Andrews and Monahan (1992) and Newey and West (1994), respectively.

estimate following Barndorff-Nielsen and Shephard (2002) as the sum of squared intraday midquote returns. Note that across all regression specifications both LHS and RHS variables are taken in logs and first differences. The obvious advantage of this is twofold: First, regression coefficients are straightforward to be interpreted as percentage point changes. Second, FX volume in levels is non-stationary and persistent (see Ranaldo and Santucci de Magistris, 2018), hence taking first-differences is an effective remedy to render the series stationary.

Table 3 shows the passage from a linear model with a dummy or an interaction term to a non-linear smooth transition regression (LSTAR). To be specific, the first three columns in this table report the results from estimating a linear model (OLS) of the form

$$PIM_{k,t} = \lambda_t + \alpha_k + \beta_1' f_{k,t} + \rho' f_{k,t} \cdot D_{t-1} + \delta' w_{k,t} + \epsilon_{k,t}, \tag{6}$$

where $f_{k,t}$ and $w_{k,t}$ collect all regressors and D_{t-1} is a 1-day lagged interaction variable capturing periods of heightened dealer constraints. Note that the estimate of ρ corresponds to the difference between the constrained and unconstrained regime coefficient (i.e., $\beta_2 - \beta_1$) of the LSTAR model in Eq. (5). In column 1, we simply interact our dealer constraint measure DCM_t with total trading volume. In column 2, D_t is equal to one if DCM_t is above its 75% quantile in period t. Note that the specifications in columns 1 and 2 are a simple, yet intuitive, approximation of non-linear regression models. In column 3, D_t is a logistic transformation of DCM_t based on $1/[1 + exp(-\gamma DCM_t)]$, where γ determines the steepness of the function. For simplicity, we set $\gamma = 1$ but the results in column 3 are robust for values of γ ranging from 1 to 12. Lastly, in column 4 the table shows results from a smooth transition regression (LSTAR) as in Eq. (5), which constitutes our preferred econometric approach. Again, the logistic function $G(z_{t-1})$ depends on 1-day lagged values of our dealer constraint measure DCM_t. Note that across each of the four specifications we control for the realised variance in the direct non-dollar currency pair (e.g., AUDJPY).[8] Moreover, in the Online Appendix we show that our findings are robust to including the cross-currency basis (Du, Tepper, and Verdelhan, 2018) as a control for funding liquidity (Andersen et al., 2019; Rime, Schrimpf, and Syrstad, 2021).

There is a consistent picture that arises across all four specifications in Table 3: the difference between the slope coefficient on total trading volume in constrained and unconstrained periods is negative and highly statistically significant. Moreover, the estimated slope coefficients are almost identical for the linear model with dummy (column 2) and the LSTAR model (column 4). In both cases, the difference in the slope coefficients (i.e., $\beta_2 - \beta_1$) is at least 80% lower when dealer banks' are constrained and hence less willing or able to cater their customers' trading demands.

Taken together, these results highlight the state-dependent nature of the relation between

[8] In principle, one might also want to control for liquidity, for instance, by including relative bid-ask spreads. However, this is problematic since the price inefficiency measure PIM already includes different half-spreads.

our price inefficiency metric and dealer-intermediated trading volume.[9] We can interpret the estimated coefficient linking PIM and DCM as an elasticity. From this perspective, our results suggest that dealers' liquidity provision is generally elastic in normal times, supporting market efficiency. But, when dealers face constraints their liquidity supply becomes less elastic, which adversely affects both the efficiency of prices and the level of transaction costs.

Table 3: From linear model with dummies to smooth transition regression

	PIM			
	Linear	Dummy	Logistic	LSTAR
γ			1.00	***4.90
c				**−0.04
Unconstrained volume	***0.07	***0.09	***0.14	***0.11
	[10.00]	[11.40]	[9.41]	[10.64]
Constrained volume	***0.06	0.01	0.00	*0.02
	[6.90]	[0.38]	[0.06]	[1.72]
Realised variance	***0.03	***0.03	***0.03	***0.03
	[8.13]	[8.17]	[8.15]	[8.17]
Constrained-Unconstrained	***−0.02	***−0.09	***−0.13	***−0.09
	[4.81]	[4.92]	[4.97]	[4.64]
R^2 in %	3.41	3.45	3.46	3.46
Avg. #Time periods	2,185	2,185	2,185	2,185
#Currency triplets	15	15	15	15
Currency triplet FE	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes

Note: In columns 1 to 3 this table reports results from estimating a linear model (OLS) of the form $PIM_{k,t} = \lambda_t + \alpha_k + \beta_1' f_{k,t} + \rho' f_{k,t} \cdot D_{t-1} + \delta' w_{k,t} + \epsilon_{k,t}$, where $f_{k,t}$ and $w_{k,t}$ collect all regressors and D_{t-1} is a 1-day lagged interaction variable capturing distressed market periods. Note that the estimate of ρ corresponds to the difference between the constrained and unconstrained regime coefficient (i.e., $\beta_2 - \beta_1$) in column 4. In column 1, D_t is the dealer constraint measure DCM_t. In column 2, D_t is equal to one if DCM_t is above its 75% quantile in period t. In column 3, D_t is a logistic transformation of DCM_t based on $1/[1 + exp(-\gamma DCM_t)]$, where γ determines the steepness of the function. In column 4 the table shows results from a smooth transition regression (LSTAR) of the form $PIM_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \delta' w_{k,t} + \varepsilon_{k,t}$, where $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors and $G(z_{t-1})$ is a logistic function depending on the state variable z_{t-1}. The regime variable is the 1-day lagged value of DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

To hone some further intuition for the LSTAR model, we plot the resulting time path of

[9]Note that our estimates for the difference in the slope coefficients across constrained and unconstrained regimes are very similar when using PIM as a regressor and VLM as the dependent variable.

the fitted regime function $G(DCM_t)$ in Graph A of Figure 5. Except for the first two years of the sample, the fitted $G(DCM_t)$ is mostly close to 0 (it is less than 0.25 on 45% of the days in the sample) and occasionally increases above 0.75 (29% of the days). These upward spikes are particularly pronounced during the European sovereign debt crisis through 2014/15, uncertainty around Brexit and US elections in 2016, and the Covid-19 period in 2020. Hence, the unconstrained regime (when β_1 is the effective slope coefficient) corresponds to a normal market situation, while the constrained regime (when the effective slope coefficient is close to β_2) represents periods when dealers face constraints in their risk-bearing capacity.

To assess the economic importance of dealers' risk-bearing capacity for changes in the price inefficiency measure PIM, we consider the fitted value of the LSTAR model in column 4 of Table 3. Panels B and C in Figure 5 show our fitted log changes in PIM split into two parts: the first part (Panel B) is driven by the unconstrained regime $[1 - G(DCM_t)]\hat{\beta}_1' f_{k,t}$, whereas the second part (Panel C) is determined by the constrained regime $G(DCM_t)\hat{\beta}_2' f_{k,t}$. By construction, the total fitted changes in PIM add up to the sum of these two subcomponents. Hence, it is precisely during normal periods, when dealers' risk-bearing capacity is most important for explaining changes in PIM. This result is driven by the observation that trading volume and price inefficiency co-move significantly during times when dealers are largely unconstrained. On the contrary, during constrained regimes the channel is almost completely muted as illustrated by the small fitted values in Panel C of Figure 5.

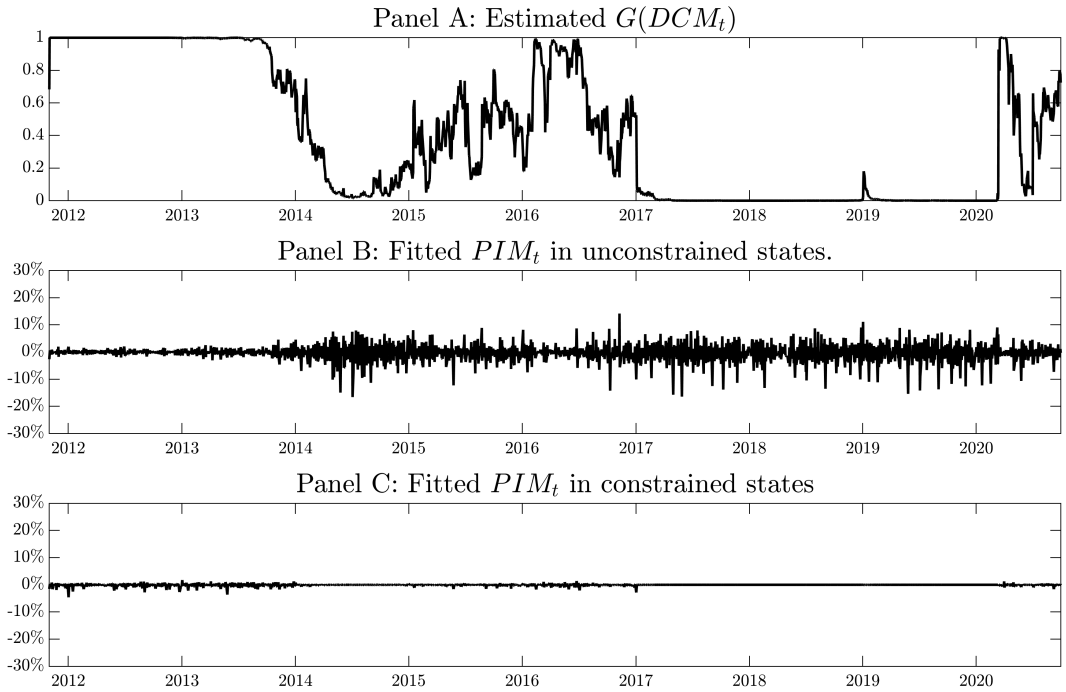
Thus far, we have mainly focused on the time-series dimension of the relation between trading volume and price inefficiency but have not delved deeper into the cross-section of currency pair triplets. To explore the cross-sectional heterogeneity, we estimate the LSTAR model individually for 15 triplets of currency pairs. We further contrast the result with a simple linear model that does not distinguish between constrained and unconstrained regimes. We report these analyses in Table 4. In particular, Panel A shows the results from estimating a linear model (OLS) of the form

$$PIM_{k,t} = \alpha_k + \beta_k \, VLM_{k,t} + \delta_k \, RV_{k,t} + \epsilon_{k,t}, \tag{7}$$

where $VLM_{k,t}$ is the total trading volume within each currency pair triplet k and $RV_{k,t}$ the realised variance in the non-dollar currency pair. In Panel B the same table shows results based on the LSTAR model in Eq. (5), where the regime variable is again the 1-day lagged value of the dealer constraint measure DCM_t. As before, both dependent and independent variables are taken in logs and changes to support the interpretation of the regression coefficients as percentage point changes (or, equivalently, elasticities).

Table 4 strongly supports the idea that constraints on dealers non-trivially impact the relation between trading volume and price inefficiency. In particular, the difference between the parameter estimates of constrained and unconstrained regimes (i.e., $\beta_2 - \beta_1$) is significantly negative for 11 out of 15 triplets of currency pairs. In line with this finding, the R^2s of these regressions are rather close to the linear model. This is entirely expected, given that the co-

Figure 5: Time-series of fitted $G(\text{CDS})$ and price inefficiency measure



Note: Panel A of this figure shows the fitted regime function $G(DCM_t)$, using the point estimates in column 4 of Table 5. Panel B shows the cross-sectional average of the part of the fitted log changes in $PIM_{k,t}$ that is driven by unconstrained state coefficients ($[1 - G(z_{t-1})]\beta_1' f_{k,t}$). Panel C shows the cross-sectional average of the part driven by the constrained state coefficients ($G(z_{t-1})\beta_2' f_{k,t}$). By construction, the fitted values for log changes in $PIM_{k,t}$ are the sum of Panels B and C. The sample covers the period from 1 November 2011 to 30 September 2020.

efficient with respect to trading volume in constrained regimes is close to zero. In sum, both results are consistent with the idea that in calm periods dealers channel liquidity to the FX market supporting market efficiency but are less willing to do so when they are constrained. In other words, the elasticity of liquidity provision by dealers is significantly lower during periods when our measure of dealer constraints is high.

We dig deeper into the relation between trading volume VLM and the two determinants of our price inefficiency measure, namely, violations of triangular no-arbitrage VLOOP and the transaction costs TCOST incurred by arbitrage trades. Table 5 reports results from daily fixed effects LSTAR panel regressions based on Eq. (5), where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., VLOOP, TCOST or PIM) and the regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. Moreover, in columns 2, 4, and 6 we include currency pair specific realised variance in non-dollar pairs as a control variable.

Two findings stand out: First, trading volume covaries significantly less with VLOOP and TCOST during times when dealers are more constrained. Second, it is above all the relation between VLOOP and trading volume that strongly diverges and even exhibits negative

Table 4: Linear model and smooth transition regression with DCM as state variable

Panel A: OLS	AUDJPY	AUDNZD	CADJPY	EURAUD	EURCAD	EURCHF	EURDKK	EURGBP	EURJPY	EURNOK	EURSEK	GBPAUD	GBPCAD	GBPCHF	GBPJPY
Intercept (α)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.21]	[0.20]	[0.35]	[0.19]	[0.20]	[0.24]	[0.07]	[0.09]	[0.11]	[0.14]	[0.12]	[0.18]	[0.20]	[0.19]	[0.27]
Volume	***0.12	***0.10	***0.06	***0.09	***0.06	***0.08	***0.09	***0.07	***0.13	*0.02	***0.04	***0.06	0.02	***0.04	***0.09
	[4.94]	[9.08]	[3.36]	[6.46]	[3.35]	[7.58]	[8.19]	[4.79]	[6.11]	[1.90]	[3.46]	[3.86]	[1.54]	[2.81]	[3.01]
Realised variance	***0.07	***0.05	***0.07	***0.08	***0.05	***0.04	***0.05	***0.05	***0.06	***0.09	***0.06	***0.07	*0.03	***0.06	**0.07
	[3.54]	[6.54]	[5.14]	[6.58]	[3.50]	[2.64]	[6.04]	[3.27]	[3.07]	[12.03]	[7.76]	[3.83]	[1.84]	[3.27]	[2.44]
Panel B: LSTAR	AUDJPY	AUDNZD	CADJPY	EURAUD	EURCAD	EURCHF	EURDKK	EURGBP	EURJPY	EURNOK	EURSEK	GBPAUD	GBPCAD	GBPCHF	GBPJPY
γ	1.05	12.00	1.31	1.00	12.00	12.00	12.00	12.00	1.29	1.99	12.00	12.00	12.00	12.00	7.82
c	−0.34	1.14	0.47	−1.08	0.17	−0.08	0.03	0.02	−0.14	0.47	1.08	−0.21	0.12	−0.08	−0.09
Unconstrained volume	***0.26	***0.13	**0.14	0.21	***0.07	***0.11	***0.11	***0.10	***0.25	**0.05	***0.05	***0.12	0.04	***0.07	***0.18
	[3.01]	[9.63]	[2.32]	[0.34]	[3.44]	[6.68]	[5.79]	[3.96]	[2.64]	[2.11]	[4.15]	[5.34]	[1.59]	[3.21]	[6.53]
Constrained volume	0.03	*0.04	−0.01	0.03	0.02	**0.04	0.03	0.02	0.01	−0.02	−0.03	0.03	−0.03	0.00	0.03
	[0.74]	[1.72]	[0.80]	[0.19]	[0.55]	[2.02]	[0.71]	[0.71]	[0.21]	[0.77]	[1.02]	[1.33]	[1.01]	[0.10]	[1.07]
Intercept (α)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	[0.66]	[1.57]	[0.36]	[0.28]	[0.27]	[0.05]	[0.99]	[0.36]	[0.10]	[1.02]	[0.28]	[0.60]	[0.30]	[0.47]	[0.46]
Realised variance	***0.05	***0.05	***0.06	***0.07	***0.05	**0.04	***0.05	***0.05	***0.05	***0.09	***0.06	***0.06	0.03	***0.06	**0.06
	[3.11]	[5.82]	[4.44]	[6.19]	[3.71]	[2.47]	[6.12]	[3.00]	[2.92]	[11.91]	[7.42]	[3.36]	[1.50]	[3.01]	[1.97]
Constrained-Unconstrained	**−0.23	***−0.09	**−0.15	−0.17	*−0.05	***−0.07	**−0.08	**−0.08	*−0.23	**−0.07	**−0.08	**−0.09	−0.07	*−0.06	***−0.15
	[2.23]	[3.28]	[2.46]	[0.22]	[1.65]	[2.61]	[2.44]	[2.03]	[1.66]	[2.41]	[2.43]	[2.50]	[1.54]	[1.92]	[3.92]
Adj. R^2 in % - OLS	18.40	13.33	11.22	17.44	6.55	10.19	12.70	11.16	18.01	11.81	7.00	12.15	1.80	9.78	14.33
Adj. R^2 in % - LSTAR	19.97	14.67	11.89	17.66	6.06	10.20	11.32	11.05	18.66	12.44	7.17	12.65	1.82	9.49	16.34
#Obs	2,285	2,285	2,278	2,285	2,285	2,285	2,285	2,285	2,285	2,283	2,285	2,285	2,285	2,285	2,277

Note: In Panel A this table reports results from estimating a linear model (OLS) of the form $PIM_{k,t} = \alpha_k + \beta' v_{k,t} + \epsilon_{k,t}$, where $v_{k,t}$ collects all regressors. In Panel B the table shows results from a smooth transition regression (LSTAR) of the form $PIM_{k,t} = \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \delta' w_{k,t} + \varepsilon_{k,t}$, where $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors and $G(z_{t-1})$ is a logistic function depending on the state variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on robust standard errors (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

coefficients in the constrained regime.

Our preliminary explanation, which we formalise below based on a simple model (see Section 4), is that there are two main forces at play when dealers are constrained. On the one hand, dealers will post more conservative bid and ask quotes thus increasing transaction costs (higher TCOST). On the other hand, a higher spread will discourage trading including arbitrage activity, which in turn makes dealers' incoming customer flows less imbalanced and hence implies a lower VLOOP. In our model (see Section 4 below), volatility is one of the underlying forces of market dynamics and price inefficiency. The primitive source of volatility (and hence trading volume) in the model are diverging private values among market participants (i.e., disagreement).

To account for the effect of disagreement, we have explored regression specifications including alternative high-frequency measures of disagreement. For instance, we consider the dispersion of order flows of corporates, funds, non-bank financials, and banks (Cespa et al., 2021) as well as the volume-volatility ratio (Liu and Tsyvinski, 2020) as control variables and have found that our key empirical results remain qualitatively unchanged.[10] Furthermore, the estimates for the composite price inefficiency measure PIM strongly resemble the ones based on TCOST. This result is not surprising given that TCOST is both theoretically and empirically an order of magnitude larger than VLOOP (see Table 1 and Figure 6, respectively).

4. A simple model of constrained market efficiency

This section presents a static partial equilibrium model that rationalises the two main empirical findings from the previous section:

1. Price inefficiency is higher when volatility is higher and FX dealer banks are more constrained (Figure 4 and Table 2).
2. Price inefficiency and dealer-intermediated volume co-move when the dealers are unconstrained but the positive correlation weakens significantly when dealers become more constrained (Tables 3 to 5).

The model features two periods ($t = 0, 1$), three currency pairs (e.g., EURCAD, USDEUR, USDCAD) and two types of agents: liquidity traders and one representative dealer. At $t = 0$, liquidity traders arrive and trade with the dealer. At $t = 1$, the uncertainty is resolved.

4.1. Trading environment

FX spot contracts. Let us denote the three currency pairs as x, y, and z. Let $\mathbf{p}$ denote the exchange rates of the three currency pairs $[p^x, p^y, p^z]^T$ at $t = 0$ whereby the base currency for

[10] These additional findings are available upon request.

Table 5: Smooth transition panel regression with DCM as state variable

	VLOOP		TCOST		PIM	
	(1)	(2)	(3)	(4)	(5)	(6)
γ	*12.01	*12.01	***12.05	***12.04	***4.95	***4.90
c	***0.80	***0.80	***0.76	***0.74	***−0.11	***−0.04
Unconstrained volume	***0.10	***0.08	***0.12	***0.09	***0.14	***0.11
	[3.69]	[2.96]	[15.46]	[11.04]	[13.79]	[10.64]
Constrained volume	−0.06	−0.07	**0.03	0.01	***0.05	*0.02
	[1.15]	[1.43]	[2.41]	[0.72]	[3.96]	[1.72]
Realised variance		**0.02		***0.03		***0.03
		[2.05]		[7.98]		[8.17]
Constrained-Unconstrained	***−0.15	***−0.15	***−0.09	***−0.08	***−0.09	***−0.09
	[2.70]	[2.66]	[5.24]	[4.89]	[4.89]	[4.64]
R^2 in %	0.10	0.13	2.47	3.79	2.26	3.46
Avg. #Time periods	2,182	2,182	2,186	2,185	2,185	2,185
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $y_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., *VLOOP*, *TCOST*, or *PIM*), $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors, and $G(z_{t-1})$ is a logistic function depending on the state variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

the first currency pair is the EUR and that for the other two it is the USD. For instance, for currency pair x, 1 EUR = p^x CAD. More specifically, let a^k and b^k denote the ask and bid price of currency k, and m^k and s^k denote the mid-quote and the spread of currency k.

The agents trading the currency pairs at $t = 1$ get the fundamental value, in which there is no difference between the direct FX rate (EURCAD) and the correspondent synthetic rate computed with two indirect rates (USDEUR and USDCAD). We denote the fundamental value of the three currency pairs as $\tilde{\mathbf{e}} = [\tilde{e}^x, \tilde{e}^y, \tilde{e}^z]^T$. The mean and variance of the fundamental value are $\mathbf{e} = [e^x, e^y, e^z]^T$ and $\boldsymbol{\sigma} = [\sigma, \sigma, \sigma]^T$, respectively. Note that the three fundamental values are intimately linked via $e^x = e^y e^z$.

Liquidity traders. We model liquidity demand in a reduced form, following the classic market microstructure literature (see e.g., Grossman and Miller, 1988; Hendershott and

Menkveld, 2014). At $t = 0$, there is L unit mass of liquidity traders in each currency pair where L is increasing in σ. In addition, L is a decreasing function of the transaction cost, that is, the (absolute) spread s quoted by the dealer. Let λ denote the slope coefficient, and $L = \lambda\sigma(1-s)$. A π fraction of the liquidity traders in currency pair x are buyers and the rest are sellers. For currency pair y, a π fraction of liquidity traders are sellers and the rest are buyers. For currency pair z, half of them are buyers, whereas the other half are sellers.

The demand is imbalanced across the three currency pairs due to different private values, following the spirit of Gabaix and Maggiori (2015). For simplicity, suppose π is larger than $1/2$, the liquidity traders impose net buying pressure $(2\pi - 1)$ in currency pair x and net selling pressure $(1 - 2\pi)$ in currency pair y. As a result, the liquidity traders' demand imbalance (i.e., the net buying pressure) in each of the three currencies is given as

$$\mathbf{d} = \lambda\sigma(1-s) \times [2\pi - 1, 1 - 2\pi, 0]^T. \tag{8}$$

Dealer. There is a representative and competitive dealer, à la Foucault, Pagano, and Roell (2013, Sec. 3.5). The dealer is risk-averse and is endowed with equity capital that is worth $\mathbf{E}$ and makes the market at $t = 0$. Being competitive and starting with zero inventory, she decides on her positions in the three currency pairs $\mathbf{q}$, where $q^k > 0$ means the dealer sells the currency pair k, taking the market-clearing prices $\mathbf{p}$ as given.

The dealer is assumed to be debt-financed and incurs a funding cost (e.g., Scott, 1976; Binsbergen, Graham, and Yang, 2010). For simplicity, suppose the debt financing costs across the three currency pairs are the same and are denoted as δ. Moreover, the dealer's leverage (i.e., asset-equity ratio) across the three currency pairs is the same and denoted as γ. The cost of debt financing is proportional to the net positions in the three currency pairs: $\delta\gamma\mathbf{1}^\mathbf{T}|\mathbf{q}|$ where $\mathbf{1} = [1,1,1]^T$. Let ρ denote the risk-aversion coefficient. Hence, the utility of the leveraged dealer is given as

$$U^D = E\Big(\mathbf{p}^\mathbf{T}\mathbf{q} - \mathbf{e}^\mathbf{T}\mathbf{q} - \underbrace{\delta\gamma\mathbf{1}^\mathbf{T}|\mathbf{q}|}_{\text{Debt cost}}\Big) - \frac{\rho}{2}Var\Big(\mathbf{p}^\mathbf{T}\mathbf{q} - \mathbf{e}^\mathbf{T}\mathbf{q} - \underbrace{\delta\gamma\mathbf{1}^\mathbf{T}|\mathbf{q}|}_{\text{Debt cost}}\Big). \tag{9}$$

Thus, the constraints examined in Section 3 are captured via i) the mean-variance type of risk aversion (i.e., Value-at-Risk and FX volatility) and ii) the debt funding cost (i.e., leverage ratio in He et al. (2017), CDS spreads, and dealer debt financing cost).[11] Note that the mean-variance utility function is commonly used in the literature. The only deviation of our setup is to explicitly introduce the debt financing cost.

[11] "Constrained" in our context refers to impaired risk-bearing capacity, which does not necessarily imply binding restrictions. Hence, in the model, the dealer is not subject to any additional (regulatory) binding constraints.

Market clearing. The market clearing condition is the following: At $t = 0$, the liquidity traders' demand is equal to the dealer's position $\mathbf{q}$. In other words,

$$\mathbf{d} = \mathbf{q} \tag{10}$$

4.2. Price inefficiency measure (PIM)

At $t = 0$, the supply function of the dealer is pinned down by her first order condition:

$$\frac{\partial U^D}{\partial q^k} = \begin{cases} \underbrace{a^k - \delta\gamma - e^k}_{\text{marginal value of selling}} - \underbrace{\rho\sigma^2 q^k}_{\text{price impact}} & \text{if } q^k > 0, \\ \underbrace{b^k + \delta\gamma - e^k}_{\text{marginal value of buying}} - \underbrace{\rho\sigma^2 q^k}_{\text{price impact}} & \text{if } q^k < 0. \end{cases} \tag{11}$$

The first order condition suggests that there are two components in the dealer's supply function. One is about the marginal valuation of buying and selling due to the debt financing cost. The other component is the price impact, which depends on the dealer's risk aversion.

Let η denote $\delta\gamma$. When the debt funding cost δ is higher, or when the dealer is more leveraged (a higher γ), η is higher. Thus, η is a measure of how constrained the dealer is. The higher η, the more constrained the dealer is. Thus, the supply function is

$$q^k = \begin{cases} \frac{a^k - \eta - e^k}{\rho\sigma^2} & \text{if } q^k > 0, \\ \frac{b^k + \eta - e^k}{\rho\sigma^2} & \text{if } q^k < 0. \end{cases} \tag{12}$$

Facing the liquidity traders' demand, the bid and ask prices for the three currency pairs are pinned down by the following market clearing conditions (see Eqs (13) to (15)). There are two equations for each currency pair: one is when the dealer is buying (i.e., bid price) and the other is when the dealer is selling (i.e., ask price). Taking the first condition as an example, the left-hand side is the selling amount from the liquidity traders and the right-hand side is the buying amount from the dealer for currency pair x. Eventually, in equilibrium, the bid price of currency pair x is determined by market clearing:

$$-\lambda\sigma(1 - s^x)(1 - \pi) = \frac{b^x + \eta - e^x}{\rho\sigma^2}, \qquad \lambda\sigma(1 - s^x)\pi = \frac{a^x - \eta - e^x}{\rho\sigma^2}; \tag{13}$$

$$-\lambda\sigma(1 - s^y)\pi = \frac{b^y + \eta - e^y}{\rho\sigma^2}, \qquad \lambda\sigma(1 - s^y)(1 - \pi) = \frac{a^y - \eta - e^y}{\rho\sigma^2}; \tag{14}$$

$$-\frac{1}{2}\lambda\sigma(1 - s^z) = \frac{b^z + \eta - e^z}{\rho\sigma^2}, \qquad \frac{1}{2}\lambda\sigma(1 - s^z) = \frac{a^z - \eta - e^z}{\rho\sigma^2}. \tag{15}$$

Solving the system of equations, the bid-ask spreads for the three currency pairs turn out to be the same. The intuition for this hinges on the simplified assumption that the dealers'

debt financing cost and leverage as well as the volatility of fundamental values are the same across the three currency pairs.[12] Therefore, in this setup, the bid-ask spread is

$$s = \frac{2\eta + \lambda\rho\sigma^3}{1 + \lambda\rho\sigma^3}. \tag{16}$$

Thus, the bid-ask spread increases in η. In other words, when the dealer is more constrained, either because leverage is high, or due to high debt financing cost, the bid-ask spread is also higher. Moreover, as the dealer-intermediated volume is proportional to $(1 - s)$, the model implicitly assumes that $s < 1$ since volume cannot be negative. That means $\eta < 1/2$. In this case, one finds that the bid-ask spread increases in σ. Substituting s into the market clearing conditions, the mid-quotes are[13]

$$m^x = e^x + \frac{\lambda\rho\sigma^3(1-2\eta)}{1+\lambda\rho\sigma^3}\left(\pi - \frac{1}{2}\right), \quad m^y = e^y + \frac{\lambda\rho\sigma^3(1-2\eta)}{1+\lambda\rho\sigma^3}\left(\frac{1}{2} - \pi\right), \quad m^z = e^z. \tag{17}$$

The mid-quotes of currency pair x and y deviate from their fundamental values. The effects of volatility and the dealer's constraint on the mid-quotes are directional and depend on the (net) demand imbalance of the three currency pairs. Take currency pair x as an example. The buying pressure from the liquidity traders dominates their selling pressure, which pushes up its mid-quote above its fundamental value. In other words, due to the buying pressure, the dealer charges a mark-up on currency pair x. This mark-up increases in the imbalance, which is $\lambda\sigma(1 - s)(2\pi - 1)$ and the dealer's risk aversion ($\rho\sigma^2/2$). When the dealer is more constrained, the higher spread she charges will deter part of the liquidity demand. The actions of the dealer will thus ultimately reduce trading volume and hence curb the imbalance. As a result, the mark-up decreases in the dealer's constraint η.

The effect of volatility is a result from three channels: First, higher volatility leads to a higher spread, dampening the trading volume and henceforth the mark-up. Second, more volatility itself is a reflection of greater disagreement (e.g., Hong and Sraer, 2016), hence a higher demand for liquidity provision and a higher mark-up. Third, more volatility increases the dealer's risk aversion and hence raises the mark-up. It turns out that the overall effect of volatility is to increase the mark-up implying that the second and the third channels dominate the first one. Note that for currency pair y, the mark-down is a symmetric case of the mark-up in x. Contrarily, for currency pair z, there is no mark-up or mark-down as there is no (net) order flow imbalance by assumption.

[12] Relaxing the assumption of homogeneous volatility across currency pairs (i.e., having currency pair specific volatility) does not change the results qualitatively because the market clearing conditions Eqs (13) to (15) are also currency pair specific.

[13] Although $m^z = e^z$ in our model, it is an arbitrary choice. With a slightly different setup on the demand imbalance, the same could hold for the other two currency pairs. More generally, as long as the demand imbalance across the three currency pairs is not exactly the same, the size of the deviations of the mid-quotes from the fundamental values is different, generating violations of the law of one price. For brevity, our modeling focuses on one simple example.

The deviation of the mid-quotes charged by the dealer from the fundamental values represents a violation of the law of one price, leading to a positive VLOOP. However, such positive deviations are not necessarily profitable arbitrage opportunities because of transaction costs (TCOST), that is, bid-ask spreads, which define the arbitrage bounds (Shleifer and Vishny, 1997). Thus, the combination of VLOOP and TCOST reflects market (in)efficiency. Following our definition of PIM in Eq. (3), we have that

$$PIM = VLOOP + |TCOST| = log\left(\frac{m^x}{m^y m^z}\right) + log\left(\frac{(1-\frac{s}{2m^x})}{(1+\frac{s}{2m^y})(1+\frac{s}{2m^z})}\right). \tag{18}$$

4.3. The links between the model and the empirical results

As shown in Eq. (18), the effect of volatility and the dealer's constraint on PIM is transmitted both via the (relative) bid-ask spreads and the mid-quotes. Figure 6 visualizes these effects. Intuitively, when volatility is high, the deviations of the mid-quotes from the fundamental value are larger (due to the larger mark-up in x and larger mark-down in y), pushing up the VLOOP. Meanwhile, a higher volatility also means a higher bid-ask spread and thus a higher TCOST. As a result, PIM increases in volatility.

Regarding the effect of the dealer's constraint, there are two forces. On the one hand, when the dealer is more constrained, the spread is higher, that is, a higher TCOST. On the other hand, a higher spread also discourages trading and makes the net order flow less imbalanced. As a consequence, the deviation from the law of one price is smaller and hence VLOOP is smaller. Overall, the first channel dominates the second one, and hence PIM increases in the dealer's leverage and her debt financing cost.

Proposition 1 summarizes these results. It formalizes the intuition behind the empirical findings in Table 2 showing that more severe dealer constraints (higher DCM) (reflecting higher η) are associated with higher volatility and price inefficiency (PIM). In addition, across different parameter specifications, it turns out that the VLOOP is much smaller than TCOST, which is also one of the stylized facts from the data (see Table 1). Proposition 1 summarises these results, which are consistent with the empirical findings in Tables 3 to 5.

Proposition 1: *The price inefficiency metric (PIM) is higher when*

i) the volatility of the currency pairs is higher;

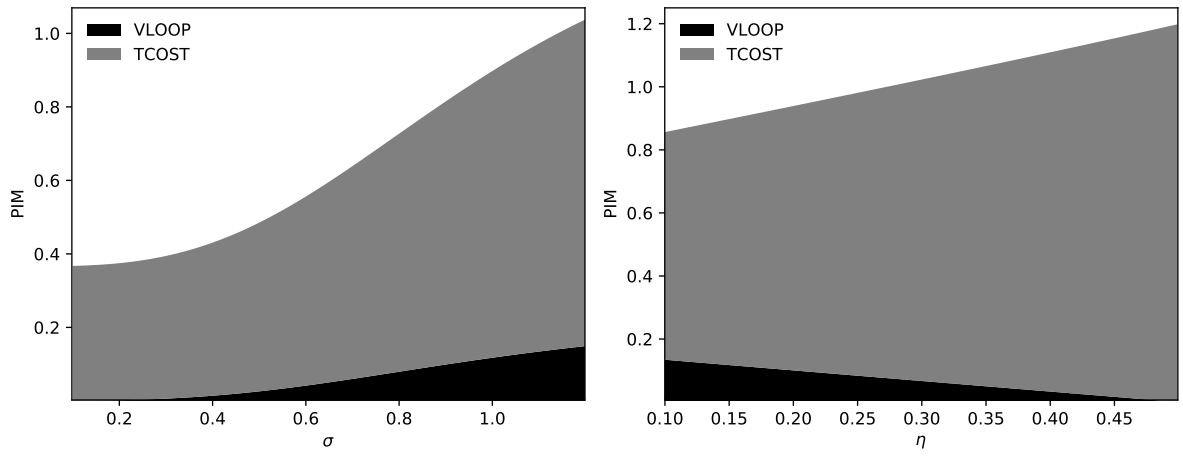
ii) the (representative) dealer is more levered;

iii) the debt funding cost for the dealer is higher.

Proof. See Online Appendix. □

Next, we investigate the dealer's elasticity of liquidity provision, which corresponds to the regression coefficients on dealer-intermediated volume in the previous section. Note that

Figure 6: The effects of volatility and the dealer's constraint on PIM



Note: This figure shows how volatility and the dealer's constraint affect PIM. The baseline parameters are $\pi = 0.7, \lambda = 1, \sigma = 1, \gamma = 1.5, \delta = 0.1, \rho = 1, e^x = 1.32, e^y = 1.1, e^z = 1.2$.

both volume and PIM are equilibrium outcomes. The dealer-intermediated volume is[14]

$$VLM = 3\lambda\sigma\frac{1-2\eta}{1+\lambda\rho\sigma^3}. \tag{19}$$

Taking the first order derivative with respect to σ, we have

$$\frac{\partial VLM}{\partial\sigma} = 3\lambda(1-2\eta)\frac{(1-2\lambda\rho\sigma^3)}{(1+\lambda\rho\sigma^3)^2}. \tag{20}$$

As $\eta < 1/2$, $(1-2\eta)$ is positive. Thus, when $\rho < 1/(2\lambda\sigma^3)$, the dealer-intermediated volume increases in volatility. Intuitively, as discussed above, a higher volatility affects volume via two channels: First, it reflects a higher trading demand and second, it deters some trading volume due to the concurrent rise in the bid-ask spread. When the dealer's risk aversion coefficient ρ is small, the former dominates the latter. As shown in Table 2, the dealer-intermediated volume increases in volatility, indicating that the parameter space of interest is $\rho < 1/(2\lambda\sigma^3)$. Thus, for the rest of this section, we focus on this case.

When the dealer is more constrained, that is, η is high, the higher bid-ask spread results into a slower increase of the dealer-intermediated volume. Lemma 1 summarises this result.

Lemma 1: *When the dealer is more constrained, volume increases in volatility at a slower pace.*

Proof. See the discussions preceding this lemma. □

[14]Note that the scalar 3 comes from assuming three currency pairs.

Recall that PIM increases in volatility for all η (see Proposition 1). Hence, as volatility increases, the dealer-intermediated volume positively correlates with PIM. Such a positive correlation captures the elasticity of liquidity provision and corresponds to the coefficient on trading volume in the regression analysis. In essence, such a positive correlation between dealer-intermediated volume and PIM represents how intermediated market depth responds to price dislocations. If both market depth and pricing distortions are large, then liquidity provision is apparently more elastic and resilient to demand and supply shocks.

Figure 7 visualises the elasticity of liquidity provision. When the dealer is unconstrained, that is, η is small, the dealer-intermediated volume is high and PIM is low (see the grey line). In general, liquidity provision seems to be rather elastic as both variables are positively correlated conditional on different realisations of volatility. However, when the dealer is constrained, the line shifts towards the lower right (see black line). To see this, compare the slope of the two tangent lines, capturing the elasticity of liquidity provision across the two market states: i) unconstrained dealers and low volatility and ii) constrained dealers and high volatility. Beyond doubt, the slope of the latter is much flatter and hence indicates that the elasticity of liquidity provision is smaller when dealers are more constrained. This is because the dealer charges a higher spread that leads to a slower increase in the amount of equilibrium trading volume. Proposition 2 summarises these results.

Proposition 2: *The dealer-intermediated volume positively correlates with PIM. This positive correlation, capturing the elasticity of liquidity provision, weakens when the dealer is more constrained.*

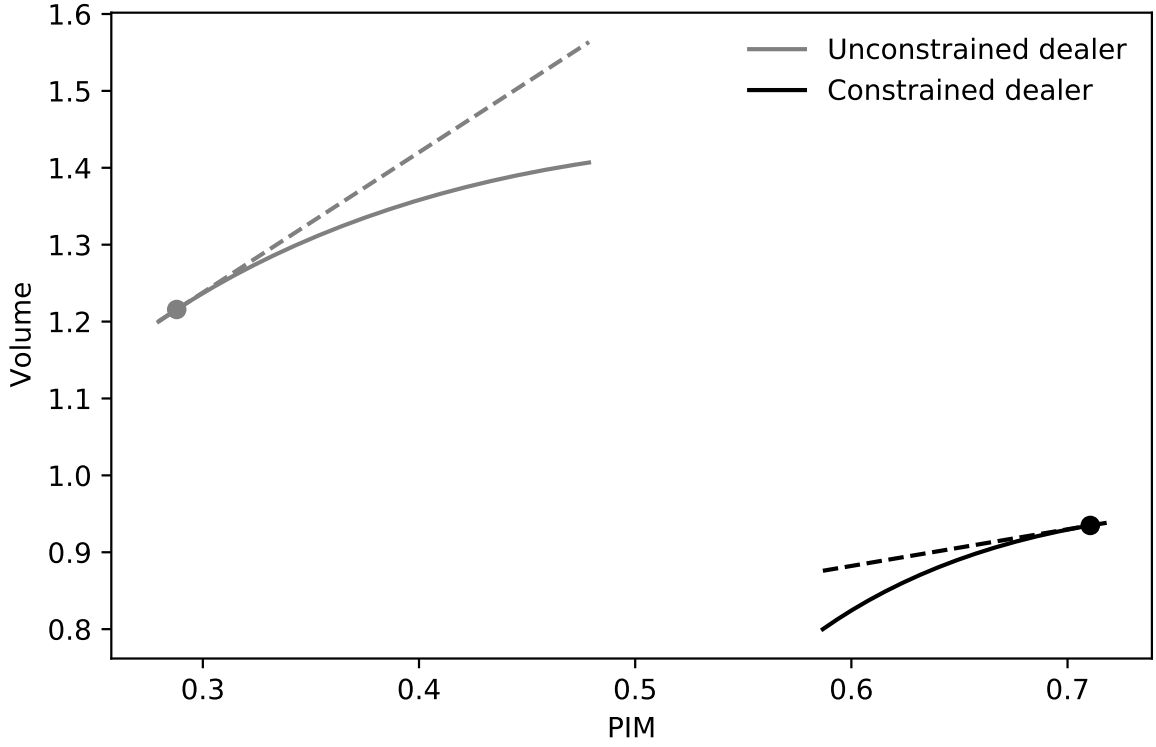
Proof. See the Online Appendix. □

Proposition 2 finds compelling support in our empirical analyses. Specifically, Table 2 shows that dealer-intermediated trading volume increases even when dealers are more constrained. However, the state-dependent correlation between intermediated volume and PIM weakens when dealer constraints intensify. This key result holds true no matter which econometric model is applied (Table 3), for almost all currency pair triplets (Table 4), for both price inefficiency components (VLOOP and TCOST), and after controlling for volatility (Table 5).

5. Additional tests and robustness

Our model suggests that both volume and PIM are equilibrium outcomes. In this case, regressing PIM on volume could lead to biased estimates when the shocks (or disturbances) on liquidity demand and supply are correlated due to potentially omitted variables. To address this issue, we follow Gabaix and Koijen (2020) and use a granular instrumental variable (GIV) approach. On top of that, we provide two additional robustness tests.

Figure 7: The elasticity of liquidity provision



Note: This figure plots the dealer-intermediated volume against PIM. The baseline parameters are $\pi = 0.7, \lambda = 1, \rho = 1, e^x = 1.32, e^y = 1.1, e^z = 1.2$. σ varies between 0.5 and 0.7. The unconstrained dealer is with $\eta = 0.05$ while the constrained one is with $\eta = 0.2$. The solid lines indicate the equilibrium outcomes when varying volatility. The grey dashed line indicates the derivative of volume with respect to PIM when the dealer is unconstrained and the volatility is low, while the black dashed line is the derivative of volume with respect to PIM when the dealer is constrained and the volatility is high.

Granular instrumental variables. To mitigate the omitted variable bias (OVB) in the smooth transition regressions setup, we construct our granular instrumental variable (GIV) following the work by Gabaix and Koijen (2020). Their approach is particularly well-suited for our analysis because it allow us to systematically identify idiosyncratic spikes in trading volume by leveraging the cross-sectional heterogeneity of the data. Given the interconnected nature of the FX market, these idiosyncratic shocks do not only affect the aggregate level of trading volume in the FX market but also the extent to which trading volume is concentrated in individual triplets of currency pairs. With these assumptions in place, we define our GIV as the difference between the size and equal weighted average of hourly dealer-intermediated trading volume:

$$GIV_t = \sum_{i=1}^{15} S_{i,t-1} VLM_{i,t} - \frac{1}{15} \sum_{i=1}^{15} VLM_{i,t} \tag{21}$$

where $S_{i,t} = \frac{VLM_{i,t}}{\sum_{i=1}^{15} VLM_{i,t}}$ is the relative volume share of currency pair triplet i at time t. We sum up hourly GIV estimates over the course of the trading day to obtain a daily time-series of idiosyncratic shocks. The intuition is that by taking the difference between size and equal weighted averages the common component is washed out across currency pair triplets and the residual corresponds to the idiosyncratic shock part. Note that these idiosyncratic spikes are driven by triplets of currency pairs that are "large" in the sense that they amass a lot of trading volume compared to the average currency triplet. Clearly, when there is not enough granularity in the data, then the GIV approach may not work. This concern does certainly not apply to our volume data set, which exhibits a high degree of heterogeneity in the cross-section of currency pair triplets (see Table 1).

Table 6 shows the results of applying the GIV approach to our LSTAR model Eq. (5). Note that we cannot include time-series fixed effects because the GIV is the same for all 15 triplets of currency pairs. The (unreported) first stage coefficient is highly significant (with an F-statistic well above 10) suggesting that we have identified an economically relevant instrument. The second stage results in Table 6 are fully in line with the OLS estimates in Table 6 suggesting that the OVB is negligible. Specifically, the difference between the trading volume coefficient across constrained and unconstrained regimes remains negative and statistically significant. This result does not only hold for our price inefficiency measure PIM but also its components based on violations of triangular no-arbitrage conditions VLOOP and arbitrage transaction costs TCOST. To be specific, the elasticity of PIM with respect to dealer-intermediated volume is 0.11 percentage points in normal times but just around 0.05 percentage points when dealers are constrained. Put differently, dealers' elasticity of liquidity provision drops by more than 50% in times of market stress. This results is consistent with the idea that dealer-intermediated trading volume promotes market efficiency in normal times but fails to do so when dealer constraints become (economically) binding.

To hone some economic intuition for our GIV we perform a narrative check in the spirit of Gabaix and Koijen (2020). The top 20 idiosyncratic shocks can be classified into three broad categories: monetary policy announcements, large asset price moves, and political uncertainty in the US, EU, and Japan, respectively. This is consistent with the idea that the GIV approach defines idiosyncratic shocks as the difference between the size and equal weighted averages of the endogenous variable. Table 7 gives an overview of the economic narratives for the top 20 idiosyncratic events. For instance, notice that the Swiss franc de-pegging on 15 January 2015, which was arguably exogenous and largely unanticipated, shows up very prominently in this list. As an additional robustness test, we include the first 3 principal components of the total trading volume within each currency pair triplet as a control variable (see Table B.3 in the Online Appendix).

Different components of dealer constraints. Next, we consider the same LSTAR specification as in Eq. (5) but instead of our dealer constraint measure DCM we use its five con-

Table 6: LSTAR panel regression with DCM as state variable and granular instrument

	VLOOP		TCOST		PIM	
γ	12.08	12.10	4.99	12.03	12.01	12.01
c	***1.11	***1.11	**1.11	*0.84	−0.12	−0.11
Unconstrained volume	***0.16	***0.12	***0.12	***0.08	***0.15	***0.11
	[6.78]	[5.18]	[8.02]	[6.61]	[9.76]	[7.25]
Constrained volume	0.07	0.04	***0.06	**0.04	***0.08	***0.05
	[1.61]	[1.01]	[2.61]	[2.42]	[3.78]	[2.85]
Realised variance		***0.08		***0.06		***0.07
		[10.05]		[6.37]		[6.92]
Constrained-Unconstrained	*−0.09	−0.08	**−0.06	**−0.05	***−0.07	***−0.06
	[1.67]	[1.63]	[2.28]	[2.31]	[2.60]	[2.59]
R^2 in %	0.95	1.77	7.30	11.16	7.44	11.29
Avg. #Time periods	2,280	2,280	2,285	2,284	2,284	2,284
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	no	no	no	no	no	no

Note: This table reports results from the second stage of a granular instrumental variable approach that we apply to the daily fixed effects LSTAR panel regressions of the form $y_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., $VLOOP$, $TCOST$, or PIM), $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors, and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

stituents. In particular, we use the 1-day lagged value of primary FX dealer banks' quarterly Value-at-Risk measure (VaR), quarterly He et al. (2017) leverage ratio (HKM), daily credit default spread (CDS), daily funding cost yield (DFC), and the daily JP Morgan Global FX Volatility index (VXY) as regime variables. Figure 8 shows the time-series variation of these five measures that exhibit reasonably high correlations ranging from 35% to 90% percent.

Table 8 reports the estimates of using each of the five aforementioned measures as a state variable and jointly considers all five of them in the last column. The difference between the constrained and unconstrained coefficients is negative and significant across all six specifications. Moreover, the estimates are fully in line with our baseline specification based on our dealer constraint measure DCM in terms of economic magnitudes. Again, the robustness of our results is not surprising given the strong co-movement across these five different regime variables.

Table 7: Narrative analysis

Date	Event
03/12/2015	ECB decides to decrease the interest rate on the deposit facility by 10 BPS to -0.30%
31/10/2014	Bank of Japan expands its monetary stimulus and injects JPY 80 tn into the economy
04/09/2014	ECB cuts rates and buys governments bonds as part of its quantitative easing programme
10/11/2011	Italy auctions EUR 5 tn worth of governments bonds as investors lose confidence in the country
06/11/2014	Key ECB interest rate targets remain unchanged
04/04/2013	Bank of Japan has surprised markets with the size of its latest stimulus package
10/03/2016	ECB sets the main refinancing rate to 0% and accelerates its monthly purchase of assets to EUR 80 bn
29/01/2016	Bank of Japan adopts negative interest rates
03/11/2011	ECB cuts benchmark interest rates by 0.25%
05/06/2014	ECB announces a cut in benchmark interest rates
13/01/2012	Standard & Poor's downgrades France and eight other eurozone countries due to the ongoing debt crisis
15/01/2015	The Swiss National Bank (SNB) abandons the peg against the EUR
09/11/2011	Italy heads deeper into the financial crisis as the 10-year government bond yield spikes above 7%
25/05/2012	Spain's fourth largest bank, Bankia, asks the government for a bailout worth EUR 19 bn
04/12/2014	Markets fall as the ECB does not announce quantitative easing and keeps interest rates unchanged
24/05/2013	Japan's central bank governor has vowed to stabilise the country's bond market
18/01/2012	IMF adjusts its growth forecast downward as the European sovereign debt crisis continues
25/02/2013	Japanese markets are boosted by new central bank nominee, Haruhiko Kuroda
09/11/2016	US stock market rallies as Donald Trump wins the US presidential election
24/08/2015	Sharp fall in the US and European stock market as concerns over China's slowing economic growth.

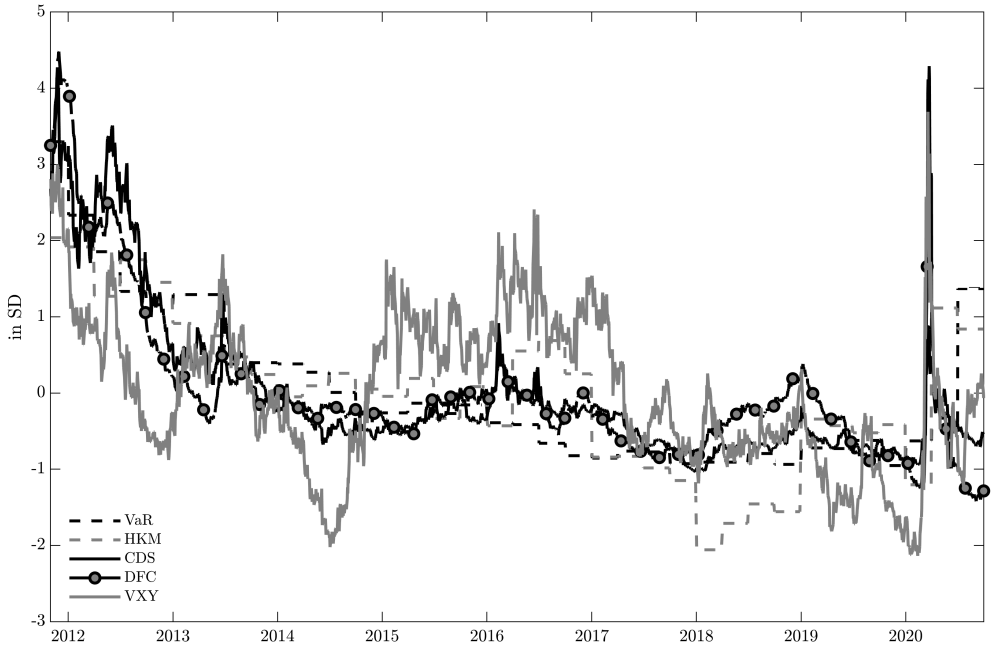
Note: This table summarises the economic narratives for the top 20 idiosyncratic events encompassing our granular instrumental variables approach. The top 20 idiosyncratic shocks can be classified into three broad categories: monetary policy announcements, large stock market moves, and political uncertainty in the US, EU, and Japan, respectively.

Inter-bank vs customer volumes Lastly, we decompose trading volume into to inter-bank and customer-bank volume to better understand which market segments suffer the most from reduced liquidity provision when dealer constraints tighten. Specifically, the CLS customer-bank order flow data comprises three groups of customers, that is, corporates, funds, and non-bank financials.[15] Note that bilateral trades between two such customer groups are quasi non-existent given the two-tier structure of the FX market (Rime and Schrimpf, 2013) and hence also do not form part of the data that CLS provides. As a result, the customer-bank data only contains trades that pass through an FX dealer bank (e.g., Citi Bank or UBS). Moreover, the inter-bank data includes trades between two banks that are members of the CLS system. Some of these banks are GSIBs, whereas others include lower-tier banks outside of the main dealer community (e.g., Danske Bank or Commerzbank).

Table 9 reports the results of estimating the LSTAR model in Eq. (5) based on inter-bank and customer-bank volume rather than total trading volume. To be precise, we define total trading volume in each customer group as the sum of buy and sell volume in a given currency

[15]See Cespa et al. (2021) and Ranaldo and Somogyi (2021) for a detailed description of the CLS flow data set.

Figure 8: State variables



Note: This figure plots different state variables that we observe at the daily and quarterly frequency, respectively. Observations have been standardised by subtracting the full sample mean and dividing by the full sample standard deviation of every state variable. The five state variables are the 1-day lagged value of primary FX dealer banks' i) quarterly Value-at-Risk measure (VaR, dashed black line), ii) quarterly He et al. (2017) leverage ratio (HKM, dashed grey line), iii) daily credit default spread (CDS, solid black line), iv) daily funding cost yield (DFC, solid black line with grey circles), and v) the daily JP Morgan Global FX Volatility index (VXY, solid grey line). The sample covers the period from 1 November 2011 to 30 September 2020.

pair. There is an interesting picture that arises: On the one hand, the coefficients related to unconstrained volume of the inter-bank segment are higher than those of the customer-bank segment suggesting a more elastic of liquidity provision in the former. On the other hand, the elasticity of liquidity provision significantly weakens with dealer constraints for both customer-bank and inter-bank trading activity. However, the economic magnitudes of the constrained minus unconstrained coefficients suggest that large dealer banks mainly curtail their liquidity provision in transactions with other banks. Of course, this does not rule out the possibility that dealers charge higher spreads to their customers as well when they are more constrained.

To summarise, the first robustness check has delved into the determinants of our composite dealer constraint measure, whereas the second robustness test has sought to better understand how dealer constraints affect customer-dealer and inter-dealer relations. Overall, these tests corroborate our previous results and support the main mechanisms of our model.

Table 8: Smooth transition regression with different state variables using market shares

	PIM					
	VaR	HKM	CDS	DFC	VXY	All
γVaR	***12.04					*12.10
γHKM		***4.98				**12.10
γCDS			***12.03			**12.10
γDFC				***4.33		12.10
γVXY					***12.02	*12.09
c	***0.36	***0.60	***−0.21	***−0.27	***0.26	***0.50
Unconstrained volume	***0.09	***0.09	***0.11	***0.12	***0.09	***0.09
	[11.46]	[10.96]	[11.64]	[7.27]	[10.43]	[11.24]
Constrained volume	0.02	0.02	*0.02	***0.04	***0.04	0.00
	[1.04]	[1.55]	[1.68]	[3.28]	[3.14]	[0.22]
Realised variance	***0.03	***0.03	***0.03	***0.03	***0.03	***0.03
	[8.41]	[8.44]	[8.43]	[8.21]	[8.58]	[8.16]
Constrained-Unconstrained	***−0.08	***−0.07	***−0.09	***−0.08	***−0.06	***−0.10
	[4.46]	[3.73]	[5.39]	[3.15]	[3.59]	[4.87]
R^2 in %	3.33	3.26	3.40	3.34	3.28	3.45
BIC	51.57	51.58	51.56	50.82	51.58	50.80
Avg. #Time periods	2,284	2,284	2,284	2,185	2,284	2,185
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $PIM_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variables are the 1-day lagged value of primary FX dealer banks': quarterly Value-at-Risk measure (VaR, column 1), quarterly He et al. (2017) leverage ratio (HKM, column 2), daily credit default spread (CDS, column 3), daily funding cost yield (DFC, column 4), and the daily JP Morgan Global FX Volatility index (VXY, column 5). Note that we assign an equal weight to each top 10 FX dealer bank (based on the Euromoney FX survey) when computing a cross-sectional average. Column 6 includes all 5 regime variables jointly. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

Dealers promote market efficiency in normal times through elastic liquidity provision. As such, dealer intermediation contributes to greater market efficiency: tight spreads and informative prices. However, during periods of market stress the FX dealer banks are constrained and as a result their intermediation activities cannot keep up with the deterioration of market efficiency.

Table 9: Smooth transition regression with different customer groups

	VLOOP		TCOST		PIM	
	Non-bank	Bank	Non-bank	Bank	Non-bank	Bank
γ	14.90	**20.10	*20.08	**20.00	*20.03	**20.08
c	***1.10	***0.84	***0.18	***0.76	***0.15	***0.76
Unconstrained volume	*0.03	***0.08	***0.02	***0.09	***0.03	***0.09
	[1.86]	[2.66]	[5.56]	[9.60]	[5.25]	[9.90]
Constrained volume	−0.04	−0.12	0.00	−0.01	0.00	−0.02
	[1.30]	[1.51]	[0.01]	[0.40]	[0.16]	[0.76]
Realised variance	**0.02	*0.02	***0.04	***0.03	***0.04	***0.03
	[2.28]	[1.71]	[9.02]	[7.15]	[9.13]	[7.35]
Constrained-Unconstrained	**−0.07	**−0.19	***−0.02	***−0.10	***−0.03	***−0.11
	[2.04]	[2.36]	[2.99]	[3.82]	[3.16]	[3.90]
R^2 in %	0.10	0.14	3.00	3.92	2.72	3.57
Avg. #Time periods	1979	1979	1982	1982	1982	1982
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $y_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., $VLOOP$, $TCOST$, or PIM), $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors, and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 September 2012 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

6. Conclusion

In this paper, we have studied whether dealer constraints have adverse implications on market efficiency. Using a unique data set of prices and volumes in the FX market, we provide a novel analytical method to identify and measure market (in)efficiency, and its main components: the fundamental violation of the law of one price and the associated transaction cost that traders would incur to restore no-arbitrage conditions. Second, we uncover the following novel empirical findings: On the one hand, when dealers are unconstrained they support market efficiency by providing liquidity to their customers. On the other hand, when dealers are constrained, for instance, when they face Value-at-Risk, leverage, or funding constraints, their liquidity provision is impaired and consequently the relation between intermediated volume and price efficiency breaks down. We rationalise our findings with a theoretical model outlining how market efficiency deteriorates when markets are more volatile and when

dealers are more constrained.

We obtain our results for the FX spot market, which is commonly regarded as one of the world's most liquid markets in the world. However, we believe that our findings also have implications for other over-the-counter (OTC) markets. For instance, broadly similar mechanisms could be at play when pricing distortions emerge between similar government bonds (Hu, Pan, and Wang, 2013) with pronounced deviations from a smooth yield curve (as observed during the Covid-19 crisis). We leave the study of the role of dealer constraints on the efficiency of other important OTC markets (e.g., government and corporate bonds, OTC derivatives) to future research.

References

Adrian, T. and Boyarchenko, N., 2012. Intermediary leverage cycles and financial stability. *SSRN Electronic Journal*.

Adrian, T. and Shin, H. S., 2010. Liquidity and leverage. *Journal of Financial Intermediation*, 19(3): 418–437.

Adrian, T. and Shin, H. S., 2013. Procyclical leverage and value-at-risk. *Review of Financial Studies*, 27 (2):373–403.

Adrian, T., Etula, E., and Muir, T., 2014. Financial intermediaries and the cross-section of asset returns. *The Journal of Finance*, 69(6):2557–2596.

Andersen, L., Duffie, D., and Song, Y., 2019. Funding value adjustments. *The Journal of Finance*, 74(1): 145–192.

Andrews, D. W. K. and Monahan, J. C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966.

Barndorff-Nielsen, O. E. and Shephard, N., 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(2):253–280.

Binsbergen, J. H. V., Graham, J. R., and Yang, J., 2010. The cost of debt. *The Journal of Finance*, 65(6): 2089–2136.

Bjønnes, G. H. and Rime, D., 2005. Dealer behavior and trading systems in foreign exchange markets. *Journal of Financial Economics*, 75(3):571–605.

Cespa, G. and Foucault, T., 2014. Illiquidity contagion and liquidity crashes. *Review of Financial Studies*, 27(6):1615–1660.

Cespa, G., Gargano, A., Riddiough, S. J., and Sarno, L., 2021. Foreign exchange volume. *Review of Financial Studies*.

Chaboud, A. P., Chernenko, S. V., and Wright, J. H., 2008. Trading activity and macroeconomic announcements in high-frequency exchange rate data. *Journal of the European Economic Association*, 6 (2-3):589–596.

Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., and Vega, C., 2014. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5):2045–2084.

Christiansen, C., Ranaldo, A., and Söderlind, P., 2011. The time-varying systematic risk of carry trade strategies. *Journal of Financial and Quantitative Analysis*, 46(04):1107–1125.

Chu, Y., Hirshleifer, D., and Ma, L., 2020. The causal effect of limits to arbitrage on asset pricing anomalies. *The Journal of Finance*, 75(5):2631–2672.

Driscoll, J. C. and Kraay, A. C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560.

Du, W., Tepper, A., and Verdelhan, A., 2018. Deviations from covered interest rate parity. *The Journal of Finance*, 73(3):915–957.

Duffie, D., 2010. Presidential address: Asset price dynamics with slow-moving capital. *The Journal of Finance*, 65(4):1237–1267.

Evans, M. D., 2002. FX trading and exchange rate dynamics. *The Journal of Finance*, 57(6):2405–2447.

Evans, M. D. and Lyons, R. K., 2002. Order flow and exchange rate dynamics. *Journal of Political Economy*, 110(1):247–290.

Evans, M. D. and Lyons, R. K., 2005. Do currency markets absorb news quickly? *Journal of International Money and Finance*, 24(2):197–217.

Fischer, A. M. and Ranaldo, A., 2011. Does FOMC news increase global FX trading? *Journal of Banking and Finance*, 35(11):2965–2973.

Foucault, T., Pagano, M., and Roell, A., 2013. *Market Liquidity*. Oxford University Press.

Foucault, T., Kozhan, R., and Tham, W. W., 2016. Toxic arbitrage. *The Review of Financial Studies*, 30(4): 1053–1094.

Gabaix, X. and Koijen, R. S. J., 2020. Granular instrumental variables. *SSRN Electronic Journal*.

Gabaix, X. and Maggiori, M., 2015. International liquidity and exchange rate dynamics. *The Quarterly Journal of Economics*, 130(3):1369–1420.

Gârleanu, N. and Pedersen, L. H., 2011. Margin-based asset pricing and deviations from the law of one price. *Review of Financial Studies*, 24(6):1980–2022.

Gromb, D. and Vayanos, D., 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics*, 66(2-3):361–407.

Gromb, D. and Vayanos, D., 2010. Limits of arbitrage. *Annual Review of Financial Economics*, 2(1): 251–275.

Grossman, S. J. and Miller, M. H., 1988. Liquidity and market structure. *The Journal of Finance*, 43(3): 617–633.

Hasbrouck, J. and Levich, R. M., 2018. FX market metrics: New findings based on CLS bank settlement data. *SSRN Electronic Journal*.

Hasbrouck, J. and Levich, R. M., 2021. Network structure and pricing in the FX market. *Journal of Financial Economics*, 141(2):705–729.

He, Z. and Krishnamurthy, A., 2011. A model of capital and crises. *The Review of Economic Studies*, 79 (2):735–777.

He, Z. and Krishnamurthy, A., 2013. Intermediary asset pricing. *American Economic Review*, 103(2): 732–770.

He, Z., Kelly, B., and Manela, A., 2017. Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

Hendershott, T. and Menkveld, A. J., 2014. Price pressures. *Journal of Financial Economics*, 114(3): 405–423.

Hombert, J. and Thesmar, D., 2014. Overcoming limits of arbitrage: Theory and evidence. *Journal of Financial Economics*, 111(1):26–44.

Hong, H. and Sraer, D. A., 2016. Speculative betas. *The Journal of Finance*, 71(5):2095–2144.

Hu, G. X., Pan, J., and Wang, J., 2013. Noise as information for illiquidity. *The Journal of Finance*, 68(6): 2341–2382.

Kyle, A. S. and Xiong, W., 2001. Contagion as a wealth effect. *The Journal of Finance*, 56(4):1401–1440.

Liu, Y. and Tsyvinski, A., 2020. Risks and returns of cryptocurrency. *The Review of Financial Studies*, 34 (6):2689–2727.

Mancini, L., Ranaldo, A., and Wrampelmeyer, J., 2013. Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance*, 68(5):1805–1841.

Menkhoff, L., Sarno, L., Schmeling, M., and Schrimpf, A., 2016. Information flows in foreign exchange markets: Dissecting customer currency trades. *The Journal of Finance*, 71(2):601–634.

Newey, W. K. and West, K. D., 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.

Pasquariello, P., 2014. Financial market dislocations. *Review of Financial Studies*, 27(6):1868–1914.

Payne, R., 2003. Informed trade in spot foreign exchange markets: an empirical investigation. *Journal of International Economics*, 61(2):307–329.

Ranaldo, A. and Santucci de Magistris, P., 2018. Trading volume, illiquidity, and commonalities in FX markets. *SSRN Electronic Journal*.

Ranaldo, A. and Somogyi, F., 2021. Asymmetric information risk in FX markets. *Journal of Financial Economics*, 140(2):391–411.

Rime, D. and Schrimpf, A., 2013. The anatomy of the global FX market through the lens of the 2013 Triennial Survey. *BIS Quarterly Review*.

Rime, D., Schrimpf, A., and Syrstad, O., 2021. Covered interest parity arbitrage. *Review of Financial Studies*.

Rösch, D., 2021. The impact of arbitrage on market liquidity. *Journal of Financial Economics*, 142(1): 195–213.

Rösch, D. M., Subrahmanyam, A., and van Dijk, M. A., 2016. The dynamics of market efficiency. *The Review of Financial Studies*, 30(4):1151–1187.

Scott, J. H., 1976. A theory of optimal capital structure. *The Bell Journal of Economics*, 7(1):33.

Shleifer, A. and Vishny, R. W., 1997. The limits of arbitrage. *The Journal of Finance*, 52(1):35–55.

Somogyi, F., 2021. Dollar dominance in FX trading. *SSRN Electronic Journal*.

van Dijk, D., Teräsvirta, T., and Franses, P. H., 2002. Smooth transition autoregressive models — a survey of recent developments. *Econometric Reviews*, 21(1):1–47.

Appendix A. Other data sources

Major FX dealer banks are at the heart of our composite dealer constraint measure. For each year from 2011 to 2020, we retrieve the ranking of the top 10 FX dealer banks from the Euromoney FX surveys, which are publicly available. See Table 1 for an overview of the top 10 FX dealer banks over the sample period from 2011 to 2020. Note that this implies that we do not include any non-bank financial liquidity providers (i.e., XTX, HC Tech or Jump Trading), which are privately held companies. What follows lists the data source for each of the five subcomponents of our composite dealer constraint measure (DCM).

- **Value-at-Risk (VaR)** is retrieved directly from the financial statements for each of the top 10 dealer banks. The frequency is quarterly.
- **Leverage ratio (HKM)** is computed following the work by He et al. (2017) as book debt (i.e., short plus long term debt) relative to the sum of market equity (i.e., shares outstanding times share price) and book debt that are retrieved from Bloomberg for each dealer bank. The frequency is quarterly.
- **Credit default spread (CDS)** with 5 year maturity is retrieved from Bloomberg for each dealer bank. The CDS premia are denominated in dollars for US banks and in euros for all Europeans banks, including the UK domiciled ones. The frequency is daily.
- **Debt Funding cost (DFC)** is retrieved from iBoxx for each dealer bank corresponding to the average bond issuance cost across different maturities and major currencies (i.e., USD, EUR, and GBP). The frequency is daily.
- **JP Morgan Global Volatility index (VXY)** is not dealer bank specific but a global stress factor that we retrieve directly from Bloomberg. The frequency is daily.

Table 1: Top 10 FX dealer banks

Rank	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
1	Deutsche Bank	Deutsche Bank	Deutsche Bank	Citi Bank	Citi Bank	Citi Bank	Citi Bank	JP Morgan	JP Morgan	JP Morgan
2	Barclays	Citi Bank	Citi Bank	Deutsche Bank	Deutsche Bank	JP Morgan	JP Morgan	UBS	Deutsche Bank	UBS
3	UBS	Barclays	Barclays	Barclays	Barclays	UBS	UBS	Bank of America	Citi Bank	Deutsche Bank
4	Citi Bank	UBS	UBS	UBS	JP Morgan	Deutsche Bank	Bank of America	Citi Bank	UBS	Citi Bank
5	JP Morgan	HSBC	HSBC	HSBC	UBS	Bank of America	Deutsche Bank	HSBC	State Street	HSBC
6	HSBC	JP Morgan	JP Morgan	JP Morgan	Bank of America	Barclays	HSBC	Goldman Sachs	HSBC	Goldman Sachs
7	RBS	RBS	RBS	Bank of America	HSBC	Goldman Sachs	Barclays	Deutsche Bank	Bank of America	State Street
8	Credit Suisse	Credit Suisse	Credit Suisse	RBS	BNP Paribas	HSBC	Goldman Sachs	Standard Chartered	Goldman Sachs	Bank of America
9	Goldman Sachs	Morgan Stanley	Morgan Stanley	BNP Paribas	Goldman Sachs	Morgan Stanley	Standard Chartered	State Street	Barclays	BNP Paribas
10	Morgan Stanley	Goldman Sachs	Bank of America	Goldman Sachs	RBS	BNP Paribas	BNP Paribas	Barclays	BNP Paribas	Barclays

Note: This table reports the ranking of the top 10 FX dealer banks for the years 2011 to 2020 from the Euromoney FX surveys. Note that this ranking only includes banks and excludes any non-bank financial liquidity providers (i.e., XTX, HC Tech or Jump Trading), which are privately held companies.

Appendix B. Additional empirical results

Table 2: Summary statistics

	PIM in bps		Volume in $bn		Bid-ask spread in bps		Volatility in bps		VLOOP>TCOST in %
	VLOOP	TCOST	Direct	Synthetic	Direct	Synthetic	Direct	Synthetic	
AUD-USD-JPY	0.24	4.88	0.18	2.55	4.15	2.93	14.38	11.21	0.18
AUD-USD-NZD	0.29	5.85	0.09	1.00	4.44	3.72	9.32	12.70	0.02
CAD-USD-JPY	0.30	4.67	0.03	2.66	4.29	2.61	12.66	9.47	0.43
EUR-USD-AUD	0.19	4.52	0.14	3.86	3.54	2.82	11.54	10.97	0.04
EUR-USD-CAD	0.28	4.25	0.08	3.97	3.55	2.49	10.15	9.12	0.07
EUR-USD-CHF	0.21	3.98	0.37	3.38	2.62	2.71	6.38	9.50	0.10
EUR-USD-DKK	0.14	3.89	0.09	3.09	2.54	2.65	1.82	9.22	0.05
EUR-USD-GBP	0.19	4.07	0.61	4.08	3.19	2.48	9.52	9.62	0.03
EUR-USD-JPY	0.21	3.90	0.65	4.83	3.14	2.42	11.43	9.69	0.66
EUR-USD-NOK	0.26	7.69	0.24	3.13	6.25	4.70	11.01	11.95	0.05
EUR-USD-SEK	0.23	6.86	0.27	3.13	5.41	4.21	9.18	11.28	0.05
GBP-USD-AUD	0.20	5.08	0.04	1.80	4.22	2.99	12.53	11.26	0.02
GBP-USD-CAD	0.29	4.69	0.03	1.91	4.00	2.67	10.85	9.41	0.05
GBP-USD-CHF	0.19	4.94	0.03	1.32	4.09	2.88	10.69	9.89	0.03
GBP-USD-JPY	0.19	4.47	0.20	2.77	3.85	2.59	12.78	9.99	0.62

Note: This table reports the time-series average of hourly triangular no arbitrage deviations $VLOOP$ in basis points (bps), cumulative trading costs $TCOST$ in bps, direct trading volume in non-dollar currency pairs (e.g., AUDJPY) in $bn, synthetic trading volume in dollar pairs (e.g., the average across USDAUD and USDJPY) in $bn, as well as direct and synthetic relative bid-ask spreads and realised volatility in dollar and non-dollar currency pairs in bps, respectively. The last column shows the relative share of VLOOP>TCOST in %. Each row corresponds to a triplet of currency pairs, for example, AUDJPY, USDAUD, and USDJPY that we abbreviate as AUD-USD-JPY. The sample covers the period from 1 November 2011 to 30 September 2020.

Table B.1: Comparison EBS vs Olsen bid and ask quotes

	RMSE		MAE		CORR	
	BID	ASK	BID	ASK	BID	ASK
AUDJPY	0.286	0.232	0.131	0.126	0.996	0.997
AUDNZD	0.001	0.001	0.001	0.001	0.999	0.999
AUDUSD	0.001	0.002	0.001	0.001	0.998	0.997
CADJPY	0.335	0.315	0.148	0.147	0.995	0.996
EURAUD	0.003	0.002	0.002	0.002	0.998	0.998
EURCAD	0.002	0.002	0.001	0.001	0.998	0.998
EURCHF	0.001	0.001	0.001	0.001	0.993	0.992
EURDKK	0.001	0.001	0.001	0.001	0.991	0.988
EURGBP	0.001	0.001	0.001	0.001	1.000	1.000
EURJPY	0.211	0.201	0.124	0.124	0.999	0.999
EURNOK	0.016	0.012	0.008	0.008	0.996	0.998
EURSEK	0.009	0.009	0.006	0.006	0.999	0.999
EURUSD	0.002	0.002	0.001	0.001	0.997	0.998
GBPAUD	0.004	0.005	0.003	0.003	1.000	0.999
GBPCAD	0.004	0.004	0.003	0.003	0.999	0.999
GBPCHF	0.003	0.003	0.002	0.002	0.999	0.999
GBPJPY	0.523	0.590	0.251	0.257	0.999	0.999
GBPUSD	0.002	0.003	0.001	0.001	1.000	0.999
NZDUSD	0.001	0.001	0.001	0.001	0.999	0.999
USDCAD	0.002	0.002	0.001	0.001	0.999	0.999
USDCHF	0.001	0.001	0.001	0.001	0.998	0.998
USDDKK	0.007	0.007	0.005	0.005	0.999	0.999
USDJPY	0.178	0.194	0.112	0.111	1.000	0.999
USDNOK	0.014	0.016	0.010	0.010	0.998	0.998
USDSEK	0.018	0.013	0.009	0.008	0.999	0.999

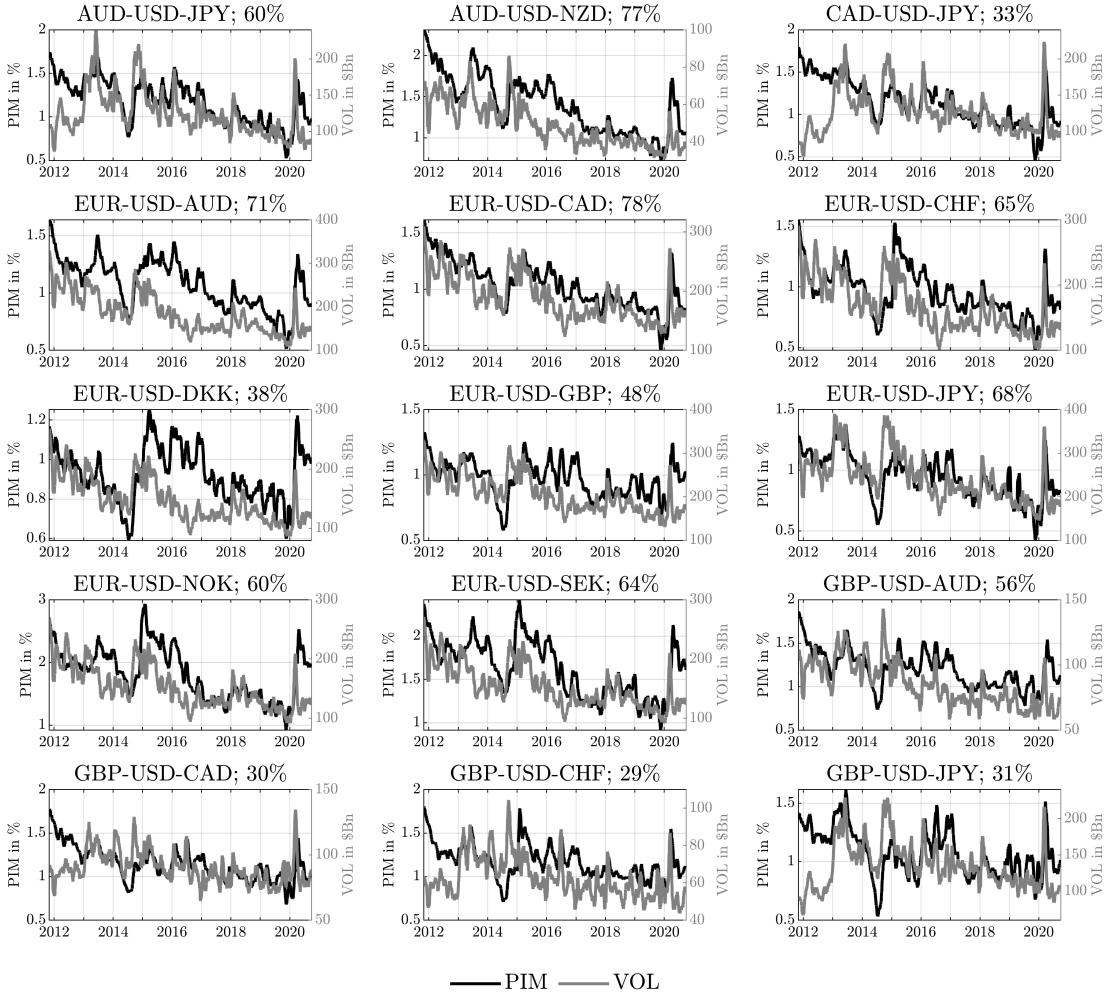
Note: This table reports the root mean squared error (*RMSE*, columns 1 and 2), the mean absolute error (*MAE*, columns 3 and 4), and the pairwise correlation coefficient (*CORR*, columns 5 and 6) for bid and ask quotes based on EBS and Olsen data, respectively. The sample covers the period from 4 January 2016 to 30 December 2016.

Table B.2: Correlations in percent

	VLOOP	TCOST	VOD	VOS	BAD	BAS	RVD
TCOST	***28.10						
VOD	***−0.48	***6.19					
VOS	***4.93	***16.53	***61.87				
BAD	***23.24	***74.34	***1.85	***9.59			
BAS	***20.68	***75.85	***17.01	***34.91	***83.36		
RVD	***15.52	***37.78	***27.43	***37.57	***54.60	***49.05	
RVS	***13.22	***44.43	***31.47	***55.92	***45.59	***69.84	***74.85

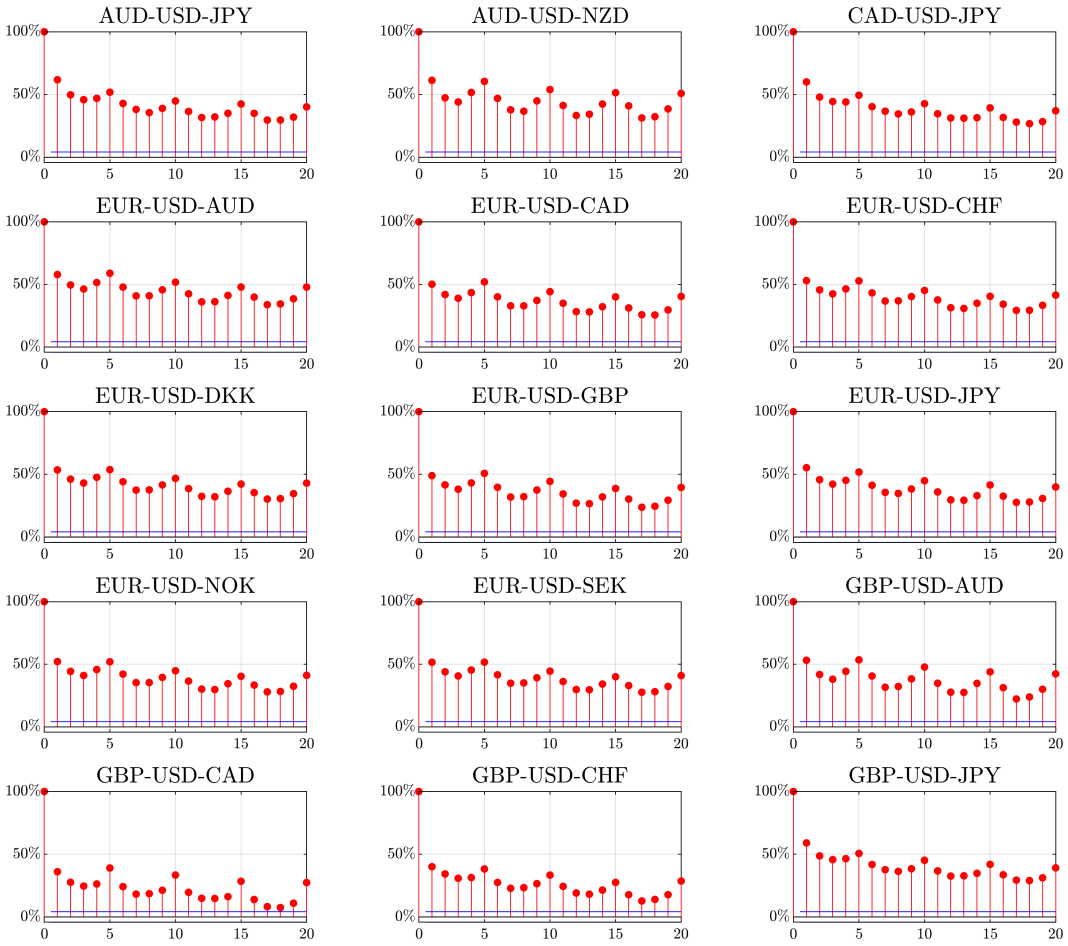
Note: This table reports the pairwise correlation coefficient of hourly triangular no arbitrage deviations $VLOOP$, cumulative trading costs $TCOST$, direct trading volume VOD in non-dollar currency pairs (e.g., AUDJPY), synthetic trading volume VOS in dollar pairs (e.g., the average across AUDUSD and USDJPY), relative bid-ask spread BAD and realised volatility RVD in non-dollar currency pairs, as well as relative bid-ask spreads BAS and realised volatility RVS in dollar currency pairs in percent (%). Significant correlations at the 90%, 95%, and 99% levels are represented by asterisks *, **, and ***, respectively. The sample covers the period from 1 November 2011 to 30 September 2020.

Figure B.1: Trading volume and market efficiency



Note: This figure plots total trading volume $VOL_{k,t}$ against our price inefficiency measure $PIM_{k,t}$ for 15 triplets of currency pairs. Currency pair triplets are denoted as XXX-USD-YYY, consisting of two dollar currency pairs (i.e., USDXXX, and USDYYY) as well as one non-dollar currency pair (i.e., XXXYYY). The percentages in the titles report the Pearson correlation coefficient between $VOL_{k,t}$ and $PIM_{k,t}$. Both time-series correspond to 22-day moving averages. The sample covers the period from 1 November 2011 to 30 September 2020.

Figure B.2: Autocorrelated trading volume



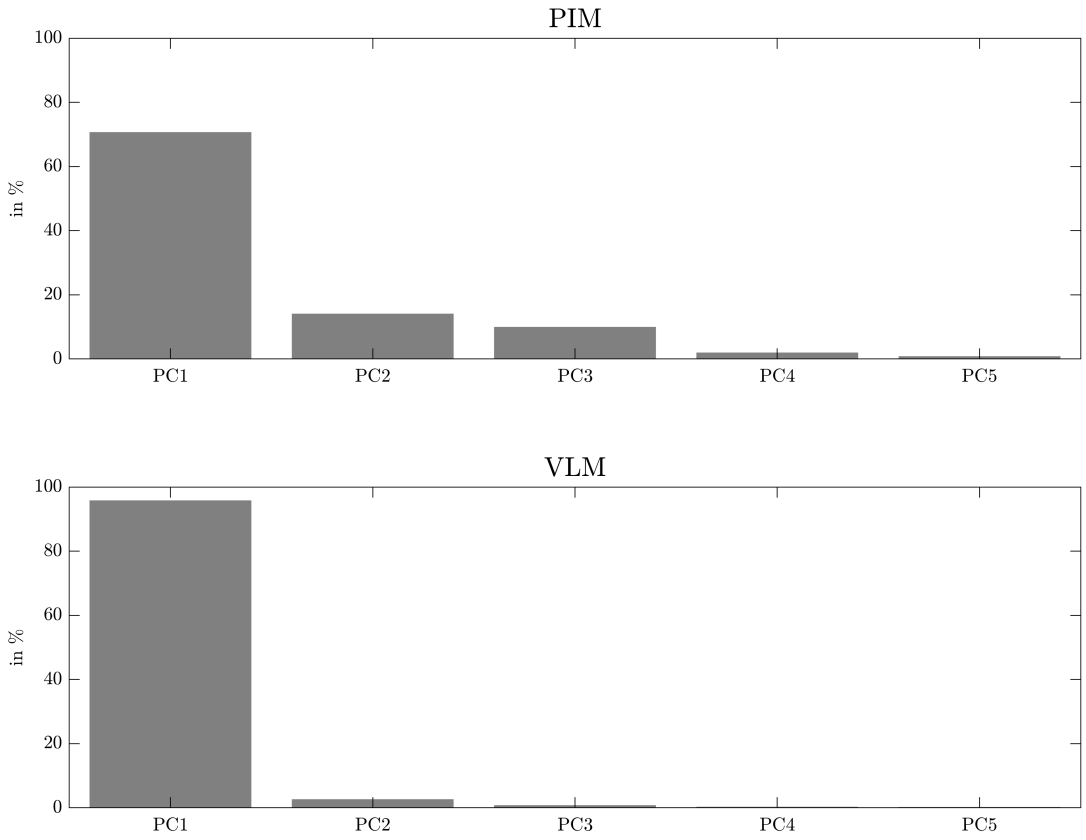
Note: This figure plots the autocorrelation coefficient of total trading volume $VOL_{k,t}$ for 15 triplets of currency pairs. Currency pair triplets are denoted as XXX-USD-YYY, consisting of two dollar currency pairs (i.e., USDXXX, and USDYYY) as well as one non-dollar currency pair (i.e., XXXYYY). The solid lines are approximate 95% confidence bounds. Both time-series correspond to 22-day moving averages. The sample covers the period from 1 November 2011 to 30 September 2020.

Table B.3: LSTAR panel regression with DCM as state variable and granular instrument

	VLOOP		TCOST		PIM	
	(1)	(2)	(3)	(4)	(5)	(6)
γ	12.01	12.03	12.03	12.03	4.98	12.07
c	***1.10	***1.10	*0.93	−0.14	**1.10	−0.11
Unconstrained volume	***0.17	***0.12	***0.12	***0.11	***0.13	***0.11
	[7.05]	[5.46]	[8.31]	[7.30]	[9.05]	[7.42]
Constrained volume	*0.08	0.04	***0.07	***0.04	***0.07	***0.05
	[1.70]	[1.11]	[3.33]	[2.60]	[2.95]	[3.07]
Realised variance		***0.08		***0.06		***0.06
		[9.99]		[6.48]		[6.92]
PC1	*0.00	*0.00	*0.00	**0.00	*0.00	**0.00
	[1.76]	[1.83]	[1.92]	[2.00]	[1.95]	[2.01]
PC2	***−0.01	***−0.01	***0.00	***0.00	***0.00	***0.00
	[3.99]	[4.01]	[3.81]	[3.65]	[3.91]	[3.72]
PC3	0.00	0.00	0.00	0.00	0.00	0.00
	[0.72]	[0.40]	[0.96]	[0.53]	[0.71]	[0.23]
Constrained-Unconstrained	*−0.09	−0.08	**−0.06	***−0.06	**−0.07	**−0.06
	[1.65]	[1.62]	[2.26]	[2.69]	[2.24]	[2.54]
R^2 in %	1.11	1.91	7.91	11.99	7.79	11.77
Avg. #Time periods	2,280	2,280	2,285	2,284	2,284	2,284
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	no	no	no	no	no	no

Note: This table reports results from the second stage of a granular instrumental variable approach that we apply to the daily fixed effects LSTAR panel regressions of the form $y_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., $VLOOP$, $TCOST$, or PIM), $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors, and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. $PC1$, $PC2$, and $PC3$ are the first three principal components of endogenous trading volume. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

Figure B.3: Principal component analysis



Note: This figure plots the share of variation (in %) across currency pair triplets explained by the first 5 principal components (PCs). The top figure is based on our price inefficiency metric PIM, whereas the bottom figure is based on total trading volume VLM. The sample covers the period from 1 November 2011 to 30 September 2020.

Table B.4: Smooth transition regression with different state variables

	PIM					
γVaR	***12.07					11.02
γHKM		***12.03				7.55
γCDS			***8.78			10.17
γDFC				***5.31		10.12
γVXY					***12.07	9.16
c	***0.36	***0.68	***−0.23	***−0.09	***0.26	***0.36
Unconstrained volume	***0.09	***0.09	***0.11	***0.11	***0.09	***0.09
	[11.46]	[11.45]	[11.15]	[9.03]	[10.43]	[11.44]
Constrained volume	0.02	0.01	*0.02	*0.03	***0.04	0.00
	[1.04]	[0.34]	[1.87]	[1.81]	[3.14]	[0.26]
Realised variance	***0.03	***0.03	***0.03	***0.03	***0.03	***0.03
	[8.41]	[8.40]	[8.43]	[8.17]	[8.58]	[8.17]
Constrained-Unconstrained	***−0.08	***−0.08	***−0.09	***−0.08	***−0.06	***−0.09
	[4.46]	[4.75]	[5.30]	[3.78]	[3.59]	[5.02]
R^2 in %	3.33	3.33	3.39	3.37	3.28	3.46
BIC	51.57	51.57	51.56	50.81	51.58	50.80
Avg. #Time periods	2,284	2,284	2,284	2,185	2,284	2,185
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

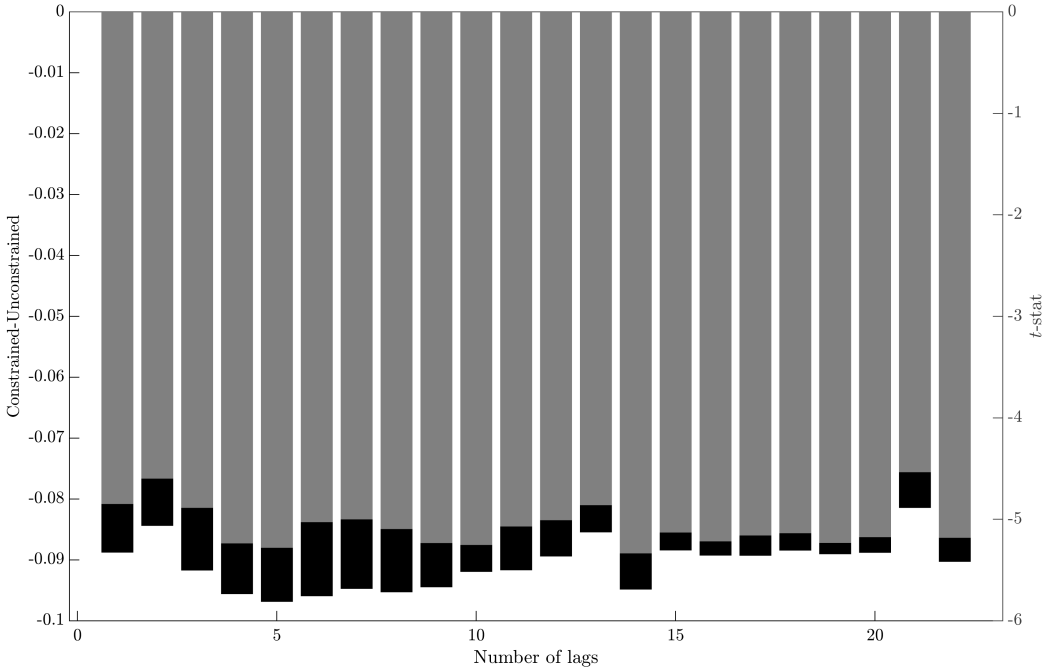
Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $PIM_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variables are the 1-day lagged value of primary FX dealer banks': quarterly Value-at-Risk measure (VaR, column 1), quarterly He et al. (2017) leverage ratio (HKM, column 2), daily credit default spread (CDS, column 3), daily funding cost yield (DFC, column 4), and the daily JP Morgan Global FX Volatility index (VXY, column 5). Note that we weight each top 10 FX dealer bank (based on the Euromoney FX survey) by its relative market share when computing a cross-sectional average. Column 6 includes all 5 regime variables jointly. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

Table B.5: LSTAR panel regression with DCM as state variable and CIP as a control

	VLOOP		TCOST		PIM	
	(1)	(2)	(3)	(4)	(5)	(6)
γ	*12.02	**12.08	***12.04	**12.06	***4.93	**12.02
c	***0.80	***−0.41	***0.76	***0.74	***−0.11	***0.74
Unconstrained volume	***0.10	***0.12	***0.12	***0.12	***0.14	***0.13
	[3.69]	[2.81]	[15.46]	[12.60]	[13.79]	[12.65]
Constrained volume	−0.06	0.00	**0.03	*0.03	***0.05	0.03
	[1.15]	[0.02]	[2.41]	[1.85]	[3.96]	[1.39]
CIP basis		0.00		0.00		*0.00
		[1.05]		[0.65]		[1.73]
Constrained-Unconstrained	***−0.15	**−0.12	***−0.09	***−0.09	***−0.09	***−0.10
	[2.70]	[2.10]	[5.24]	[4.81]	[4.89]	[4.69]
R^2 in %	0.10	0.08	2.47	2.32	2.26	2.17
AIC	90.48	91.04	48.16	48.95	50.44	51.05
BIC	89.06	88.91	46.74	46.82	49.02	48.93
Avg. #Time periods	2,182	1,978	2,186	1,981	2,185	1,981
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $PIM_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. The sample covers the period from 1 September 2012 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

Figure B.4: Constrained-unconstrained coefficient and *t*-stat



Note: This figure plots the difference between the constrained and unconstrained regime coefficient (i.e., $\beta_2 - \beta_1$) of the LSTAR model in Eq. (5) conditional on varying the number of lags in the regime variable DCM_{t-n} for $n = 1, 2, ..., 22$. The sample covers the period from 1 November 2011 to 30 September 2020.

Table B.6: LSTAR panel regression 'London hours' with DCM as state variable

	VLOOP		TCOST		PIM	
γ	**12.03	**12.04	**12.07	***4.98	***4.91	***4.98
c	***1.03	***1.02	***0.70	***0.52	***−0.09	**−0.03
Unconstrained volume	***0.10	**0.06	***0.13	***0.10	***0.16	***0.12
	[3.94]	[2.17]	[11.91]	[7.90]	[12.45]	[9.11]
Constrained volume	*−0.10	**−0.13	0.03	0.00	**0.04	0.01
	[1.80]	[2.40]	[1.53]	[0.13]	[2.40]	[0.42]
Realised variance		***0.05		***0.03		***0.04
		[4.13]		[3.81]		[5.44]
Constrained-Unconstrained	***−0.20	***−0.19	***−0.10	***−0.09	***−0.12	***−0.11
	[3.32]	[3.19]	[4.61]	[4.24]	[5.68]	[5.16]
R^2 in %	0.11	0.25	1.80	2.73	1.91	2.91
Avg. #Time periods	2,173	2,173	2,186	2,185	2,185	2,185
#Currency triplets	15	15	15	15	15	15
Currency triplet FE	yes	yes	yes	yes	yes	yes
Time-series FE	yes	yes	yes	yes	yes	yes

Note: This table reports results from daily fixed effects LSTAR panel regressions of the form $PIM_{k,t} = \lambda_t + \alpha_k + [1 - G(z_{t-1})]\beta_1' f_{k,t} + G(z_{t-1})\beta_2' f_{k,t} + \beta_3' w_{k,t} + \varepsilon_{k,t}$, where the dependent variable $y_{k,t}$ is a measure of market efficiency (i.e., $VLOOP$, $TCOST$, or PIM), $f_{k,t}$ ($w_{k,t}$) are state-dependent (*state-independent*) regressors, and $G(z_{t-1})$ is a logistic function depending on the regime variable z_{t-1}. The regime variable is the 1-day lagged value of the dealer constraint measure DCM_t. The optimal parameters γ and c are determined by nonlinear least squares minimising the concentrated sum of squared errors. Both dependent and independent variables are taken in logs and changes. When aggregating hourly to daily data we omit any observations outside of the main London stock market trading hours (i.e. from 7 am to 5 pm GMT). The sample covers the period from 1 November 2011 to 30 September 2020. The test statistics based on Driscoll and Kraay (1998) robust standard errors allowing for random clustering and serial correlation (using the plug-in procedure for automatic lag selection by Andrews and Monahan, 1992; Newey and West, 1994) are reported in brackets. Asterisks *, **, and *** denote significance at the 90%, 95%, and 99% levels.

Appendix C. Proofs for the model

Proposition 1.

Proof. Let Σ denote $\lambda\rho\sigma^3$, we can rewrite Eq. (18) as

$$e^{PIM} = \frac{1}{2}\left(\frac{m^x}{m^y m^z}\right)^2 \frac{(2m^y+s)(2m^z+s)}{2m^x-s}. \tag{C.1}$$

As the logarithmic transformation is linear, $\frac{\partial e^{PIM}}{\partial \Sigma}$ shares the same sign with $\frac{\partial PIM}{\partial \sigma}$. Note that $\frac{\partial m^x}{\partial \Sigma} = (\pi - \frac{1}{2})(1 - s - \Sigma\frac{\partial s}{\partial \Sigma})$, $\frac{\partial m^y}{\partial \Sigma} = (\frac{1}{2} - \pi)(1 - s - \Sigma\frac{\partial s}{\partial \Sigma})$, and $\frac{\partial m^x}{\partial \Sigma} = 0$. Based on the chain rule, we have

$$\begin{aligned}\frac{\partial e^{PIM}}{\partial \Sigma} &= \frac{\partial e^{PIM}}{\partial m^x}\frac{\partial m^x}{\partial \Sigma} + \frac{\partial e^{PIM}}{\partial m^y}\frac{\partial m^y}{\partial \Sigma} + \frac{\partial e^{PIM}}{\partial s}\frac{\partial s}{\partial \Sigma},\\ &= \frac{1}{2m^{z2}}\left(\frac{1-s-\Sigma\frac{\partial s}{\partial \Sigma}}{m^{y2}}\frac{(2m^y+s)(2m^z+s)}{2m^x-s}\left(\pi-\frac{1}{2}\right)(m^x+m^y)\right)\\ &+ \frac{1}{2m^{z2}}\left(\frac{m^x}{m^y}\right)^2\left(\frac{2m^y+s}{2m^x-s}\frac{\partial s}{\partial \Sigma} + (2m^z+s)\frac{2(m^x+m^y)\left((1-s)(1-2\pi)+(\Sigma(2\pi-1)+1)\frac{\partial s}{\partial \Sigma}\right)}{(2m^x-s)^2}\right).\end{aligned} \tag{C.2}$$

Note that $1 - s = \frac{1-2\eta}{1+\Sigma}$, the model is only meaningful if $\eta < \frac{1}{2}$ as otherwise the volume will turn negative. Furthermore, as $\frac{\partial s}{\partial \Sigma} = \frac{1-2\eta}{(1+\Sigma)^2}$, $\frac{\partial e^{PIM}}{\partial \Sigma}$ can be rewritten as

$$\begin{aligned}\frac{\partial e^{PIM}}{\partial \Sigma} &= \frac{1}{2m^{z2}}\left(\frac{1-2\eta}{(1+\Sigma)^2}\frac{(2m^y+s)(2m^z+s)}{m^{y2}(2m^x-s)}\left(\pi-\frac{1}{2}\right)(m^x+m^y)\right)\\ &+ \frac{1}{2m^{z2}}\left(\frac{m^x}{m^y}\right)^2\left(\frac{2m^y+s}{2m^x-s}\frac{1-2\eta}{(1+\Sigma)^2} + (2m^z+s)\frac{4(m^x+m^y)\left((1-\pi)\frac{1-2\eta}{(1+\Sigma)^2}\right)}{(2m^x-s)^2}\right).\end{aligned} \tag{C.3}$$

As $\frac{1}{2} < \pi < 1$, $\frac{\partial e^{PIM}}{\partial \Sigma} > 0$. In other words, PIM increases in σ. Similarly, $\frac{\partial e^{PIM}}{\partial \eta}$ shares the same sign with $\frac{\partial PIM}{\partial \eta}$. Note that $\frac{\partial m^x}{\partial \eta} = -(2\pi-1)\frac{\Sigma}{1+\Sigma}$, $\frac{\partial m^y}{\partial \eta} = -(1-2\pi)\frac{\Sigma}{1+\Sigma}$, $\frac{\partial m^z}{\partial \eta} = 0$ and $\frac{\partial s}{\partial \eta} = \frac{2}{1+\Sigma}$. Based on the chain rule, we have

$$\begin{aligned}\frac{\partial e^{PIM}}{\partial \eta} &= \frac{\partial e^{PIM}}{\partial m^x}\frac{\partial m^x}{\partial \eta} + \frac{\partial e^{PIM}}{\partial m^y}\frac{\partial m^y}{\partial \eta} + \frac{\partial e^{PIM}}{\partial s}\frac{\partial s}{\partial \eta},\\ &= \frac{1}{2m^{z2}}\left(\frac{(2\pi-1)}{m^{y2}}\frac{(2m^y+s)(2m^z+s)}{2m^x-s}\left(\frac{\Sigma}{1+\Sigma}\right)(m^x+m^y)\right)\\ &+ \frac{1}{2m^{z2}}\left(\frac{m^x}{m^y}\right)^2\left(\frac{2m^y+s}{2m^x-s}\frac{2}{1+\Sigma} + (2m^z+s)\frac{4(m^x+m^y)\left((2\pi-1)\Sigma+1\right)}{(1+\Sigma)(2m^x-s)^2}\right).\end{aligned} \tag{C.4}$$

As $\frac{1}{2} < \pi < 1$, $\frac{\partial e^{PIM}}{\partial \eta} > 0$. In other words, PIM increases in the dealer's leverage (γ) and the debt financing cost (δ). □

Proposition 2.

Proof. The positive correlation between volume and PIM comes directly from Proposition 1. As both PIM and volume increases in volatility, the two are also positively correlated. Such a positive correlation captures the elasticity of liquidity provision (ELP) and can be written as

$$ELP = \frac{\frac{\partial PIM}{\partial \sigma}}{\frac{\partial VLM}{\partial \sigma}} = \frac{\rho\sigma^2}{\frac{(1-2\lambda\rho\sigma^3)}{(1+\lambda\rho\sigma^3)^2}} \frac{e^{-PIM}\frac{\partial e^{PIM}}{\partial \Sigma}}{(1-2\eta)}. \tag{C.5}$$

Taking the derivative of ELP with respect to η, we have

$$\frac{\partial ELP}{\partial \eta} = \frac{\rho\sigma^2}{\frac{(1-2\lambda\rho\sigma^3)}{(1+\lambda\rho\sigma^3)^2}} \frac{\partial e^{-PIM}}{\partial \eta} \frac{\partial \frac{\frac{\partial e^{PIM}}{\partial \Sigma}}{(1-2\eta)}}{\partial \eta}. \tag{C.6}$$

Obviously, the first part of Eq. (C.6) is independent of η. The second part is negative as

$$\frac{\partial e^{-PIM}}{\partial \eta} = -\frac{1}{e^{PIM2}} \underbrace{\frac{\partial e^{PIM}}{\partial \eta}}_{>0, \text{ from Proposition 1}}. \tag{C.7}$$

From Eq. (C.3), the third part can be simplified as

$$\begin{aligned}
\frac{\partial \frac{\partial e^{PIM}}{\partial \Sigma}\frac{1}{1-2\eta}}{\partial \eta} &= \frac{\partial}{\partial \eta}\frac{1}{2m^{z2}}\left(\frac{1}{(1+\Sigma)^2}\frac{(2m^y+s)(2m^z+s)}{m^{y2}(2m^x-s)}\left(\pi-\frac{1}{2}\right)(m^x+m^y)\right) \\
&+ \frac{\partial}{\partial \eta}\frac{1}{2m^{z2}}\left(\frac{m^x}{m^y}\right)^2\left(\frac{2m^y+s}{2m^x-s}\frac{1}{(1+\Sigma)^2} + (2m^z+s)\frac{4(m^x+m^y)\left((1-\pi)\frac{1}{(1+\Sigma)^2}\right)}{(2m^x-s)^2}\right), \\
&= \frac{\partial}{\partial \eta}\frac{e^{PIM}(m^x+m^y)}{(1+\Sigma)^2}\left(\frac{1}{m^{x2}}\left(\pi-\frac{1}{2}\right) + \frac{1}{(2m^z+s)(m^x+m^y)} + \frac{4(1-\pi)}{(2m^x-s)(2m^y+s)}\right)
\end{aligned} \tag{C.8}$$

Note that $\frac{\partial m^x}{\partial \eta} = -(2\pi-1)\frac{\Sigma}{1+\Sigma}$, $\frac{\partial m^y}{\partial \eta} = -(1-2\pi)\frac{\Sigma}{1+\Sigma}$, $\frac{\partial m^z}{\partial \eta} = 0$ and $\frac{\partial s}{\partial \eta} = \frac{2}{1+\Sigma}$, thus m^z and $m^x + m^y$ do not vary with η. In addition, PIM increases in η. Thus, the first part of Eq. (C.8) increases in η. For the second part, we can see that it increases in η because

$$\frac{\partial}{\partial \eta}\left(\frac{1}{m^{x2}}\left(\pi-\frac{1}{2}\right) + \frac{1}{(2m^z+s)(m^x+m^y)} + \frac{4(1-\pi)}{(2m^x-s)(2m^y+s)}\right)$$

$$= \frac{2(2\pi-1)}{m^{x3}}\frac{\Sigma}{1+\Sigma} + \frac{2}{(m^x+m^y)(2m^z+s)^2}\frac{1}{1+\Sigma} + \\ + \frac{16(1-\pi)(m^x+m^y)}{((2m^x-s)(2m^y+s))^2}\left((2\pi-1)\frac{\Sigma}{1+\Sigma}+\frac{1}{1+\Sigma}\right) > 0. \tag{C.9}$$

Hence, the third part of Eq. (C.6) is positive. As a result, ELP decreases in η. □

References: Online Appendix

Andrews, D. W. K. and Monahan, J. C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60(4):953–966.

Driscoll, J. C. and Kraay, A. C., 1998. Consistent covariance matrix estimation with spatially dependent panel data. *Review of Economics and Statistics*, 80(4):549–560.

He, Z., Kelly, B., and Manela, A., 2017. Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1–35.

Newey, W. K. and West, K. D., 1994. Automatic lag selection in covariance matrix estimation. *The Review of Economic Studies*, 61(4):631–653.