The market microstructure approach to foreign exchange: Looking back and looking forward

Michael R. Kinga,*, Carol L. Oslerb, Dagfinn Rimec

University of Western Ontario, Canada
Brandeis University, USA
Norges Bank, Norway

JEL classification:
F31
G12
G15
C42
C82

Keywords:
Exchange rates
Market microstructure
Order flow
Information
Liquidity
Electronic trading

Abstract

Research on foreign exchange market microstructure stresses the importance of order flow, heterogeneity among agents, and private information as crucial determinants of short-run exchange rate dynamics. Microstructure researchers have produced empirically-driven models that fit the data surprisingly well. But FX markets are evolving rapidly in response to new electronic trading technologies. Transparency has risen, trading costs have tumbled, and transaction speed has accelerated as new players have entered the market and existing players have modified their behavior. These changes will have profound effects on exchange rate dynamics. Looking forward, we highlight fundamental yet unanswered questions on the nature of private information, the impact on market liquidity, and the changing process of price discovery. We also outline potential microstructure explanations for long-standing exchange rate puzzles.

© 2013 Elsevier Ltd. All rights reserved.

The authors would like to thank Michael Melvin, Geir Bjønnes, Alain Chaboud, Martin Evans, Ingrid Lo, Lukas Menkhoff, Michael Sager, Elvira Sojli, Mark Taylor, Paolo Vitale, and participants at the Journal of International Money and Finance 30th Anniversary conference in New York for helpful comments and suggestions. We thank Alain Chaboud and Clara Vega for the order flow regressions using EBS data. The views expressed in this paper are those of the authors and do not necessarily represent those of the Norges Bank. Michael King acknowledges financial support from the Bank of Montreal Fellowship.

* Corresponding author. Ivey Business School, UWO, 1151 Richmond Street N, London, Ontario, Canada N6A 3K7. Tel.: +1 519 661 3084.
E-mail address: mking@ivey.ca (M.R. King).

0261-5606/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jimonfin.2013.05.004
The ancient and honorable field of international finance has grown furiously of late in activity, in content, and in scope.

Michael R. Darby

These opening words, written by the editor to introduce the inaugural issue of the Journal of International Money and Finance (JIMF) in 1982, could well have been written about the field of foreign exchange (FX) research today. Over the past thirty years, research on exchange rates has continued to grow in response to the puzzles that naturally arose following the move to floating rates after the breakdown of Bretton Woods.

We survey an important and relatively new line of exchange-rate research known as FX market microstructure. Researchers in this field take a microeconomic approach to understanding the determination of exchange rates, which are after all just prices. They analyze the agents that trade currencies, the incentives and constraints that emerge from the institutional structure of trading, and the nature of equilibrium.

Our survey first looks back at how FX microstructure emerged in the 1990s in response to the disappointing empirical performance of macro-based exchange-rate models. Early microstructure researchers went directly to the market, observing the trading process in action and talking to the FX dealers who actually set this price. These observations prompted research into features of this market that had previously been considered irrelevant, such as trading flows and private information. Progress was slow until the mid-1990s, when FX market activity shifted to electronic platforms that generated large and accurate trading records. Studies confirmed the initial insights from first-hand observation of the market and inspired fruitful new lines of inquiry. In interpreting this evidence, FX research drew on a strong conceptual foundation from existing equity-market microstructure research, always recognizing that research on one market cannot be uncritically be “taken over in to and applied to the FX market because the nature of the markets differ” (Booth, 1984, p. 210).

We survey the extensive body of striking and robust results that has emerged from these efforts and re-visit some early pioneering research that laid the groundwork for recent studies. The new insights from the FX microstructure literature have their own inherent scientific value and are proving valuable in achieving the field’s original goal: understanding macro-level exchange-rate puzzles. Our survey finishes by looking forward to the impact of recent dramatic changes in the FX market structure, and highlight topics that may be a fruitful focus for future research.

This survey follows the FX microstructure literature in focusing primarily on empirical studies, while highlighting the numerous important contributions published by the JIMF. The JIMF, always receptive to the ‘facts first’ approach, has been the leading outlet for this field. The JIMF has published four times as many FX microstructure papers as the next leading journal (see Appendix, Table A). The JIMF has published key microstructure papers even if they adopted methodologies not widely accepted in economics (e.g. surveys), even if they reached conclusions at odds with the rest of international economics (e.g. Evans and Lyons, 2002a), and even if they dealt with microstructural nonlinearities orthogonal to standard exchange-rate models (e.g. Osler, 2005). The JIMF has thus played an important role in establishing this line of inquiry as a respected part of international economics.

Given the breadth of FX microstructure research, some important topics are not covered in this survey. The current study only briefly discusses the changes in electronic trading and, the FX market infrastructure, which are covered in detail by King et al. (2012). We do not discuss the large literature on FX intervention, which is surveyed by Sarno and Taylor (2001), Neely (2005), Melvin et al. (2009) and Menkhoff (2010). We do not review the many studies on the FX reaction to macro news


2 King et al. (2012) provides a comprehensive history of the evolution of FX market structure, with considerable detail on the geography and composition of currency trading, the players in FX markets, and the evolution of electronic trading. The chapter is descriptive and does not consider the microstructure literature or other academic studies of FX.
announcements, which are covered by Andersen et al. (2003), Bauwens et al. (2005), Chen and Gau (2010), and Savaser (2011). Finally, we do not review the literature on FX volatility; interested readers should see Berger et al. (2009) and the references contained therein.

This study has six sections. Section 1 looks back to the origins of FX microstructure. Section 2 reviews the most powerful finding from microstructure, namely the impact of order flow on exchange rate movements. Section 3 discusses liquidity provision and price discovery. Section 4 highlights key changes in the FX infrastructure over the past decade and their potential implications for exchange rate dynamics. Section 5 highlights unresolved questions that warrant more research. Section 6 concludes.

1. The emergence of FX microstructure

When fixed exchange rates were abandoned in the early 1970s, researchers had little evidence to guide the development of exchange-rate models. Few countries had experimented with floating exchange rates and then only briefly. Constrained by the lack of data, economists developed models inductively. The first model, purchasing power parity (PPP), has always been helpful for explaining long-run exchange rate movements but it provided few explanations for short-run movements under floating rates. From this line of inquiry economists gained the important lesson that currencies are a store of value as well as a medium of exchange.

All models are simplifications of reality that are eventually falsified by the data, leading to the development of better models (Popper, 1959). The next generation of FX models included interest rates and focused on rational portfolio selection. Consistent with the efficient markets perspective of the 1980s, these models assumed that all information is public and that uncovered interest rate parity (UIP) and PPP hold continuously. Even covered interest rate parity (CIP) did not hold during turbulent periods (Taylor, 1989). As time went on and the world gained experience with floating rates, these next-generation models inevitably faced their own empirical challenges. Asset supplies could not be connected to exchange rates (e.g., Boughton, 1987) and UIP, like PPP, failed to hold at short horizons (Hodrick, 1987; Engel, 1996). The most high-profile disappointment was the failure of these models to forecast exchange short-run exchange rate movements better than a random walk (Meesee and Rogoff, 1983; Faust et al., 2003).

Scientific progress mandated a third round of exchange-rate models. When scientific frameworks face major empirical challenges some scientists choose to modify the existing model while others develop entirely new models. Researchers attempted to modify these standard models to better fit the data by introducing exogenous elements such as a time-varying risk premium. As noted by Burnside et al. (2007), however, such efforts to patch up standard models are ‘fraught with danger’ because they introduce important sources of model misspecification.

1.1. Deductive, facts-first approach

By the mid-1980s the accumulation of experience with floating rates suggested it would be possible to design new models using a deductive, ‘facts first’ approach. Researchers decided to visit FX dealing rooms, reasoning that “economists cannot just rely on assumption and hypotheses about how speculators and other market agents may operate in theory, but should examine how they work in practice, by first-hand study of such markets” (Goodhart, 1988). Frankel et al. (1996, p. 3) were optimistic that this approach could be fruitful, stating: “It is only natural to ask whether [the] empirical problems of the standard exchange-rate models...might be solved if the structure of foreign exchange markets was to be specified in a more realistic fashion”.

Given the paucity of data, many microstructure researchers undertook to survey FX market participants directly (e.g. Taylor and Allen, 1992; Cheung and Chinn, 2001; Gehrig and Menkhoff, 2004; Lui and Mole, 1998; Menkhoff, 1998). The surveys revealed three beliefs that are widely shared in the market but that were strikingly inconsistent with standard exchange-rate models. First, FX dealers believe that exchange rates respond to trading flows. To traders, the importance of such flows is self-evident and dealers build their day-to-day trading strategies on this conceptual foundation. Nonetheless, this belief is inconsistent with the focus in standard models on FX holdings, not flows, and the assumption that UIP and PPP hold continuously.
Second, the surveys reveal that FX dealers view private information as an important feature of their market. For example, Cheung and Chinn (2001) report that dealers view larger banks as having an informational advantage due to their larger customer base and network. This view is inconsistent with the standard models’ assumption that all information is public.

Third, market participants view trading flows as the conduit through which private information influences exchange rates. This view is inconsistent with the pure efficient markets property of standard models, where news causes an instantaneous adjustment in the equilibrium exchange rate without any trading required.

1.2. Insights from high-frequency datasets

High-frequency data on FX trading were scarce during the 1970s and 1980s, when deals were agreed via telephone and fax machines. To analyze the market more systematically, pioneers such as Charles Goodhart, Richard Lyons, Richard Olsen and Mark Taylor painstakingly assembled detailed datasets from records generously provided by EBS, Citibank, and Thomson Reuters, among others. Olsen and Associates played a leading role by sponsoring conferences in the early 1990s on high-frequency data in finance.

An early paper by Goodhart and Figliuoli (1991) uses high-frequency exchange-rate data to analyze minute-by-minute quotes in the interdealer market. This path-breaking work documents key stylized facts, such as the tendency for bid-ask spreads to cluster at just a few levels and for exchange-rate returns to be negatively autocorrelated. The authors also found that spreads were not sensitive to market conditions – a finding that was shown by Melvin and Tan (1996) to reflect the general stability of their small sample period. Goodhart and Figliuoli (1991) and Goodhart and Payne (1996) document the absence of negative autocorrelation in traded prices, in contrast to equity and bond markets where it is normally attributed to bid-ask bounce. These studies attribute the zero autocorrelation to a balance between negative autocorrelation associated with bid-ask bounce and long sequences of trades all in one direction that impart positive autocorrelation. Another important pioneer, Lyons (1995) used trade data from a single active FX dealer to document the very brief half-life of an FX dealer’s inventory. Lyons (1995) found that the dealer had daily average profits of $100,000 (or one basis point) on trading volume of $1 billion. Early work tended to rely on indicative quotes rather than trades, raising the possibility of mis-measurement. Danielsson and Payne (2002) show that indicative and firm quotes are quite close to each other except when the market moves quickly.

In the late-1990s trading became automated and high-frequency data became more widely collected on trading floors. The rigorous econometric tests soon undertaken confirmed all three beliefs outlined above and validated Goodhart’s call for ‘first-hand’ study of the markets. The next two sections summarize this evidence.

1.3. Key agents in FX markets

Any microeconomic investigation of a market begins by identifying the key agents. In FX markets, these can be divided into customers and dealers. FX has historically been almost exclusively a wholesale market where the major customers were corporations engaged in international trade, leveraged and unleveraged (‘real money’) asset managers, local and regional banks, and central banks. With the emergence of electronic trading networks in the late 1990s, a retail segment has emerged that may now represent as much as 10 percent of trading volume (King and Rime, 2010). FX trading volume has grown rapidly since the advent of floating rates, but the share of trading between dealers and corporate customers has remained steady at around 20 percent (Fig. 1). Consistent with the contemporaneous explosion of the finance industry, the share of financial trading between dealers and other financial institutions has risen from roughly 20 percent in the 1980s to over 50 percent today. Inter-dealer makes up the remaining 30 percent, a significantly lower share than seen previously.

4 Describing this process, Goodhart recounts “After one of the authors, C. Goodhart, had obtained the original videotapes from Reuters... the date on the tapes was transcribed onto paper by two of the authors’ wives, Mrs. Goodhart and Mrs. Ito, assisted by Yoko Miyao, a painstaking task beyond and above the normal requirements of matrimony” Goodhart et al. (1996).
Standard exchange rate models feature many of the major FX market actors. Hedge funds, for example, resemble the representative rational investor; they use currencies as a store of value, they condition their trades on proprietary exchange-rate forecasts, and they are motivated by profits and risk. Exporters and importers also have identifiable counterparts in some standard models. Such firms rely on foreign currency as a medium of exchange and they purchase more (less) of a currency once it has depreciated (appreciated). Most corporations, however, do not engage in speculative trading and do not condition their trades on exchange rate forecasts (Bodnar et al., 1998).

First-hand study of FX markets uncovered some important agent types that do not exhibit the behavior of agents in the standard models. Most international asset managers, for example, do not condition their portfolio choices on exchange rate forecasts (Taylor and Farstrup, 2006). This choice is arguably rational in light of the inability of exchange-rate forecasts to beat a random walk. Retail traders may condition their trades on exchange rate forecasts or technical trading rules, but on average they appear to lose money (Heimer and Simon, 2011). This recent finding suggests that retail investors do not conform to the standard assumption that all agents are perfectly rational.

FX dealers actually set bid and ask prices but they are not found in standard exchange-rate models. Dealers provide liquidity to customers, manage inventories, and take speculative positions. They are motivated by bonuses based on trading profits. Their risk-taking is constrained by position and loss limits. FX dealers typically close their inventory positions within a few minutes, and generally maintain inventories close to zero at the end of day, as illustrated in Fig. 2 (Lyons, 1998; Bjønnes and Rime, 2005). Many commentators confuse FX dealers with the proprietary traders at the large banks who behave more like hedge funds, and do not make markets for customers.

Interdealer trading accounted for over 60 percent of spot FX trades during the 1980s and early 1990s, though that share is now below 40 percent (BIS, 2010). This decline reflects the enhanced efficiency and transparency that accompanied the emergence of electronic trading networks. Most interdealer trading is now carried out indirectly via limit-order markets run by the electronic brokers, EBS and Thomson Reuters. In such markets, no agent is specifically tasked with providing liquidity. Every agent can either supply (‘make’) liquidity by placing a limit order, or demand (‘take’) liquidity by entering a market order. The midpoint between the brokers’ best posted bid and ask quotes provides a reliable signal of the current equilibrium price and is used by dealers as the basis for their quotes to customers.

The source of a dealer’s profits varies across banks and across individual dealers. Mende and Menkhoff (2006) find that liquidity provision is the dominant source of profits at a relatively small dealing bank. Bjønnes and Rime (2005), by contrast, find that speculation contributed more to profits than liquidity provision at a relatively large dealing bank. This difference can be explained by the fact that small banks focus primarily on serving customers while large banks historically would trade aggressively on the basis of information extracted from observing customer trading flows. These
different business models are neither absolute nor immutable. Lyons (1998) finds that a “jobber” at Citibank earned more from providing liquidity to other dealers than speculative position-taking. Jobbing was unusual even in 1992, and seems to be extinct today, but the widespread adoption of high-frequency trading over recent years has encouraged even the large banks to rely more heavily on customer service.

2. Order flow and exchange-rate returns

To examine the dealers’ belief that trading flows influence returns, researchers first had to measure FX flows, but this was not entirely straightforward. The number or value of total trades would not suffice; what was needed was a measure of demand pressure. But for any given trade one party demands a given currency while the other party sells it, so it was not immediately obvious how to classify them into demand and supply. This ambiguity was resolved by focusing on the trade initiator or “aggressor.” A trade is considered a buy (sell) if the initiator buys (sells) the base currency (with the other currency treated as the medium of exchange). In FX microstructure, demand pressure or “order flow” is thus measured as the number of buyer-initiated trades minus the number of seller-initiated trades. This measure is also sometimes called the “order imbalance”.

Lyons (1995) provides the first estimate of how order flow influences exchange rates. He finds that this dealer would raise his quotes by 0.0001 Deutschemark (DEM) for incoming orders worth $10 million. As he recognizes, however, one cannot necessarily extrapolate from a single dealer to the overall market. Evans and Lyons (2002b) provide more reliable estimates of the market’s response to interdealer order flow using four months of transactions in USD-DEM and USD-JPY during 1996. They regress the base currency’s daily return, \( r_t \), on order flow, \( \Delta x_t \), and fundamentals, \( F_t \):

\[
r_t = \alpha + \beta \Delta x_t + \gamma F_t + \varphi_t,
\]

where the fundamental variables are proxied by interest differentials, either lagged or in first difference, or lagged exchange-rate returns.

These regressions reveal a strong positive relationship between order flow and contemporaneous returns. The estimated coefficient on order flow is both statistically and economically significant, with an extra $1 billion in net aggressive interdealer purchases of U.S. dollars associated with a 0.5 percent appreciation of the USD vis-à-vis the DEM. The explanatory power of these regressions is on the order of 40–60 percent, which is extraordinary when compared to the 1 percent R-squared from regressions of returns on fundamentals alone. Evans and Lyons (2002a) show that the explanatory power of order flow has also been found to have a strong influence on equity returns (Chordia et al., 2002) and bond returns (Brandt and Kavajecz, 2004; Pasquariello and Vega, 2007).
flow for contemporaneous returns can be even higher, exceeding 70 percent, when returns are allowed to respond to order flow across additional currency pairs.

FX traders applauded this research as a sign that academics were more attuned to reality, with dealing banks creating teams to analyze their order flow. Within the economics profession some saw directions for new research while others remained skeptical and called for more evidence. In particular, given the extensive evidence of positive- and negative-feedback trading in FX markets, the skeptics directed special concern toward the possibility that exchange-return caused order flow instead of the reverse. Nonetheless, careful investigations provided substantial empirical support for the original hypothesis that order flow causes returns (Evans and Lyons, 2005a,b; Killeen et al., 2006). Notably, Daniélsson and Love’s (2006) study of transaction-level data reveals that the estimated influence from order flow to FX returns is actually stronger when controlling for feedback trading.

The contemporaneous relationship between interdealer order flow and exchange rate returns has been replicated with longer datasets, datasets that cover more currencies, datasets that are more recent, datasets from both large and small dealing banks, and datasets including brokered rather than direct interdealer trades.6 Table 1 presents new evidence based on brokered trades from Thomson Reuters Spot Matching (formerly Reuters D3000-2) and EBS for a broad set of currencies over a long time horizon. Order flow is measured at both intraday and daily frequencies and the sample periods vary by currency but extend as long as 15 years. We regress exchange-rate returns on order flow for different currency pairs at both the daily and the intraday frequencies. The estimated coefficients for order flow are always statistically significant at both frequencies. Explanatory power from this single factor ranges up to 40 percent at the daily frequency and 58 percent intraday.

2.1. Explaining the persistent impact of order flow

Since exchange rates for liquid currency pairs are known to approximate a random walk, the effects documented in Evans and Lyons (2002b) suggest that some part of order flow has a persistent impact on FX returns. Subsequent studies confirm that the impact persists up to a week (Evans and Lyons, 2005a) and potentially longer (King et al., 2010; Killeen et al., 2006). Using interdealer data from 1999 to 2004, Berger et al. (2008) show that the price impact of interdealer order flow declines gradually over longer horizons but remains statistically and economically significant even at one month.

To explain why order flow influences return and why some of this effect is persistent, researchers rely on three mutually consistent theories from the broader microstructure literature. The first focuses on dealers’ inventory management, the second postulates a finite price elasticity of asset demand, and the third focuses on private information. All were originally derived in optimizing models with rational agents.

2.2. Inventory effects

All currencies are quoted with a bid-ask spread, which compensates liquidity providers for operating costs and inventory risk, among other factors. Given a bid-ask spread, buyer-initiated trades necessarily push prices upwards and seller-initiated trades necessarily move them downwards, other things equal. Order flow will therefore automatically have a positive contemporaneous relation with returns. But these “inventory effects” should only persist for a few minutes, given the liquidity of FX markets, so they represent at most a partial explanation for the correlation between order flow and exchange-rate returns.

2.3. Finite elasticity of demand

A lasting effect of order flow on exchange rates emerges when the price elasticity of supply and demand are finite (Shleifer, 1986). Evans and Lyons (2002b) outline an FX trading model that captures many important aspects of FX markets and has become the intellectual workhorse of the microstructure

---

6 For studies linking order flow to exchange rates, see: Berger et al. (2008), Evans (2002), Hau et al. (2002), Killeen et al. (2006), King et al. (2010), Payne (2003) and Rime et al. (2010).
Every trading day the model’s dealers begin with zero inventory and then engage in three rounds of trading. In Round 1, dealers are contacted by random customers to trade. Dealers quote prices, trade with these customers, and accumulate inventory. In Round 2, the dealers trade with each other, effectively redistributing the aggregate inventory among themselves. In Round 3, dealers sell their inventory to a second set of customers and return to their preferred zero overnight inventory position.

The Round-1 customers can be viewed as demanding liquidity from dealers in response to exogenous shocks to their desired currency holdings. Since these customers permanently change their currency holdings, they effectively demand overnight liquidity from the market as a whole. Dealers willingly provide instantaneous liquidity but are reluctant to provide overnight liquidity in the sense that they prefer to end the day with zero inventory. The dealers therefore move bid and ask quotes sufficiently that other customers are induced to buy their remaining inventories by the end of the day. Thus it is the finite elasticity of Round-3 currency demand that accounts for a currency’s appreciation in response to positive order flow from Round-1 customers. The Round-3 customers, while demanding instantaneous liquidity from their dealers, in effect provide overnight liquidity to the market as a whole. This influence of order flow on exchange rates is sometimes referred to as a “liquidity” effect.

2.4. Private information

A persistent impact of order flow on exchange rates would also be observed if order flow is the conduit through which private information becomes embedded in exchange rates. This would be
consistent with the dealers’ beliefs that private information is an important feature of the market and that trading flows aggregate dispersed information that is relevant for pricing exchange rates. It would also be consistent with the classic microstructure theories of Glosten and Milgrom (1985) and Kyle (1985). These papers illuminate the price discovery process in stock markets, however, where the nature of private information is easy to identify. It was not immediately clear that these models would be relevant to FX markets because most exchange rate fundamentals, such as interest rates and general price levels, are publicly announced. On this basis, the possibility of private information in FX markets has often been questioned.

Early evidence consistent with the presence of private information in FX markets came from a number of studies that provided evidence of price leadership of dealers in FX markets. Peiers (1997) finds that Deutsche Bank was a price-leader in the Deutschemark and could anticipate Bundesbank interventions by up to 60 min. Similarly Covrig and Melvin (2002) find that Tokyo-based banks exhibit price leadership in the JPY and their quotes lead the rest of the market when informed players are active. Ito et al. (1998) study how volatility rose over lunch time in response to informed trading by Tokyo-based traders after the local prohibition against dealer trading over lunch time was removed. Since volatility often reflects the arrival of private information, this finding suggests that such information was emerging at that time. Killeen et al. (2006) find that the French franc–DEM exchange rate was cointegrated with cumulative order flow before the rigid parity-rates were announced in May 1998 but not after. This finding also points to a role for private information, since information would only be important when rates are flexible.

Evidence in support of a role for private information does not imply, directly or indirectly, the irrelevance of liquidity effects. Indeed, Payne (2003) and Berger et al. (2008) show that the connection from interdealer order flow to returns is stronger when market liquidity is lowest, which strongly suggests that liquidity effects play a role. Evans (2002) provides further evidence for this conclusion.

2.5. Implications for exchange-rate modeling

The Evans and Lyons (2002b) 3-round framework has important implications for modeling short-run exchange rate dynamics. First, it highlights the importance of finitely elastic currency demand (or, equivalently, currency supply, since the sale of one currency is the purchase of another), which is inconsistent with the infinite price elasticity required for UIP and PPP to hold continuously in standard models.

Second, it highlights the crucial role of heterogeneity among customers. The factors motivating Round-1 customers differ fundamentally from the factors motivating Round-3 customers: one group’s net demand is exogenous to the model, while the second group’s net demand responds endogenously to the exchange rate.

Third, the framework indicates that exchange rate models at daily or longer horizons can be microstructurally rigorous without explicitly including dealers. Though dealers are involved in virtually every FX transaction, they have limited relevance beyond the intraday horizon because they generally return to a zero inventory position when they leave for the day.7

2.6. Order flow and exchange-rate forecasting

If order flow carries information, then it should be possible to forecast exchange rates using order flow aggregates. Evans and Lyons (2005b), the first to examine this question, conclude that daily customer order flow from Citibank can forecast exchange rate returns. Their exchange rate forecasts beat a random walk over forecast horizons from 1 day to 1 month, judged using traditional statistical criteria. In a study of interbank data for four major currency pairs, Daniélsson et al. (2012) also find that order flow beats a random walk at high frequencies for all four currencies and at longer frequencies for the two most liquid currency pairs.

Note that the order flow from proprietary trading desks at commercial and investment banks is classified as financial order flow, not interdealer order flow.
Studies using economic rather than statistical criteria for assessing predictive power also find that order flow has predictive power for returns beyond the current day. Using one year of high-frequency interdealer trades, Rime et al. (2010) show that conditioning exchange rate forecasts on order flow not only beats a random walk but generates Sharpe ratios above unity. Using 11 years of daily disaggregated customer data, King et al. (2010) find that adding financial order flow to a forecasting model that already includes macroeconomic fundamentals and commodity prices improves the model's ability to predict movements in the Canadian dollar. Finally, Menkhoff et al. (2012b) document that disaggregated daily customer order flow from a leading FX dealer from 2001 to mid-2011 has predictive power for major exchange rates.

Predictive power is also indicated indirectly by other statistical tests. Killeen et al. (2006) find that the French franc–DEM exchange rate is cointegrated with cumulative order flow before the rigid parity-rates were announced but not afterwards, consistent with the hypothesis that order flow gains its lasting impact in part because it carries information. Predictive power is also implied by their finding that cumulative order flow Granger causes exchange rates. Dominguez and Panthaki (2006) use intraday interbank data to test a two-equation VAR and find that order flow leads exchange rate changes for both GBP–USD and EUR–USD.

An important exception is the study by Sager and Taylor (2008), which does not find evidence that order flow has forecasting power for exchange-rate returns. The authors study three commercially available datasets of order flow – one based on interdealer order flow from Reuters and two based on customer order flow from two major dealers, JP Morgan and RBS. The dealer data is daily customer data aggregated and manipulated to create an index, then made available with a lag. The authors find these indices have no forecasting power for FX returns at any horizon. They also find that Granger causality is reversed in their data, with exchange rate changes causing order flow. This finding is important for members of the active trading community, such as hedge funds, who have no direct access to order flow and must purchase such data from banks.

The findings for the JP Morgan and RBS indices may result from the dominance of corporate flows in these series. When constructing the indices, order flow is measured in terms of number of trades rather than the dollar value of trades. Corporate trades tend to be much smaller than financial trades (Osler et al., 2011) so that corporate trades will represent a larger share of trade numbers than trade values. The customer order flow data may be dominated by Round-3 agents, such as risk-averse real-money investors or corporate customers, whose order flow would naturally be Granger-caused by exchange-rate returns (Evans and Lyons, 2002b; Bjønnes et al., 2005). Sager and Taylor (2008, p.621) conclude “that, except for relatively few, particularly well-informed investment bank traders who observe order flow data on a tick-by-tick, real-time, and unfiltered basis, knowledge of customer or interdealer order flow cannot help improve the quality of exchange rate forecasting or the profitability of investment portfolio decision-making.” This caveat is important for researchers to keep in mind.

3. Nature and sources of private information

Given the apparent relevance of private information for FX order flow and returns, researchers have gone on to identify the nature and sources of that information. Microstructure research focuses on three types of private information about currencies: intervention, fundamental variables, and non-fundamental variables (such as dealer inventory imbalances). The sources of private information also potentially vary. It could come from corporate customers, financial customers, retail customers, or the dealers themselves. These agents could acquire the information either actively or passively.

3.1. What constitutes private information?

Information about foreign-exchange market intervention by a central bank could certainly be profitable if one were among the first to learn about it. Central banks are usually extremely secretive about intervention before it begins but once it starts they call individual banks in sequence, so the first banks called gain an informational advantage. Prior to the euro’s adoption, one might have expected that Deutsche Bank, the biggest German bank, to consistently be among the first banks called by the Bundesbank. As this suggests, Peiers (1997) finds that Deutsche Bank was indeed a price-leader in USD-
DEM during the pre-euro period, with its quotes anticipating reports about Bundesbank interventions by up to 60 min. Intervention in most liquid currencies is infrequent, so this type of information can account for only a small fraction of the overall documented influence of order flow on returns. Private exchange-rate information could also concern real-side macroeconomic fundamentals such as economic growth and relative price levels. The necessary delay between a fundamental variable’s realization in the economy and its public announcement creates an opportunity for private information to emerge, as do the revisions associated with GDP and other important macro series (Evans and Lyons, 2005a; Evans, 2010, 2011).

The extent of resources devoted to gathering intelligence about fundamental variables by members of the active trading community provides a first indication that this type of information is highly relevant. More formal evidence is provided in Evans and Lyons (2007) who show that aggregate Citibank customer order flow helps predict future GDP and inflation rates. Rime et al. (2010) also find evidence that interdealer order flow carries information about upcoming macro statistical releases.

The potential relevance of macro fundamentals is supported by evidence showing that order flow is key to the impact of news announcements on exchange rates. Indeed, econometric tests using transactions data show that the impact of macro news operates primarily through order flow (Love and Payne, 2008; Evans and Lyons, 2008; Carlson and Lo, 2006; Rime et al., 2010). These results indicate that heterogeneous interpretations of macro news represent an important source of private information. Evans and Lyons (2005a) present further evidence that order flow aggregates heterogeneous interpretations in response to news for days following a news release.

MacDonald and Marsh (1996) document that professional FX forecasters hold widely differing opinions about the expected path of exchange rates due to the idiosyncratic interpretation of public information. This heterogeneity translates into economically meaningful differences in forecast accuracy with the extent of these disagreements influences trading volume. Dunne et al. (2010) identify another form of informational heterogeneity by providing evidence that FX order flow can explain the cross-section of equity returns. This result suggests that FX order flow captures heterogeneous beliefs about fundamentals that are useful for valuing different asset classes.

The heterogeneity in agents’ views may be influenced by social forces, as indicated by Simon’s (2013) study of a social network of retail traders. He finds that after news announcements agents tend to trade in parallel with their “friends” and that agents with more friends are more profitable, all of which suggests that traders share perspectives with each other. Heterogeneity could also reflect imperfect rationality. MacDonald (2000), who summarizes studies of professional exchange rate forecasts, shows that regardless of sample period or currency pair these forecasts are biased, inefficient, and inconsistent across time horizons. The potential relevance of imperfect rationality is also indicated by Oberlechner and Osler (2012), who find that FX dealers tend to be overconfident and that this tendency does not diminish with trading experience.

3.2. Dealers’ views of fundamental information

We finish this section by examining two observations that might be mistakenly interpreted as indicating that fundamental information is irrelevant. First there is a commonly held view among FX dealers that fundamentals either do not exist or do not matter for explaining exchange-rate returns (Menkhoff, 1998). Though dealers are an important source of insight into the market, their views are not necessarily infallible and this perspective, upon close examination, could be expected from them even if prices are entirely determined by fundamentals. In standard models of price discovery, dealers learn only the direction of their informed customers’ trades, not the private information that motivates those trades. Not only would the dealers be unaware of the content of their customers’ private information, they would be unlikely to trace the long-run impact of each trade so as to identify whether that information had a permanent impact, as required if the information is to be identified as fundamental.

The irrelevance of fundamental information for dealers might also be inferred from the observation that dealers’ speculative positions are typically held only briefly. This could be expected since fundamental information will rapidly influence prices when markets are highly liquid. Holden and Subrahmanyam (1992) demonstrate this familiar principle using a Kyle (1985) model modified to
include multiple informed traders. The principle has been confirmed by empirical studies from many subfields within finance. In FX markets, Carlson and Lo (2006) find that the impact of a 1997 fundamental news shock on USD-DEM was essentially complete within half a minute. With the recent spread of algorithmic trading (Chaboud et al., 2009; King et al., 2012), the market’s reaction to new information may be now substantially faster.

3.3. Non-fundamental information

There are good reasons why non-fundamental information could also influence FX markets. In fact, if demand and supply are provided with finite elasticity, it is not just possible but logical that information about order flow and dealer inventory imbalances will forecast returns at high frequencies. Suppose a dealer learns that a US firm needs 30 billion Norwegian krone to purchase a Norwegian firm. The dealer can be fairly certain that the krone will appreciate in the near term regardless of whether the acquisition represents a lasting shift to fundamentals. Dealers report that it is standard practice to trade on this type of information and that the big banks are better informed in part because they trade more with the customers who make the biggest trades.

A theoretical reference point for the relevance of order-flow information is provided by Cao et al. (2006), who extend the Evans and Lyons (2002a) model to include two rounds of interdealer trading in the “middle” of the trading day, rather than just one. An individual dealer’s inventory imbalance immediately after it trades in Round-1 is non-fundamental information because it provides a signal of the market-wide inventory imbalance.

Empirical evidence for the importance of non-fundamental information is provided by Cai et al.’s (2001) analysis of high-frequency trade data. They show that customer order flow has an influence on rates distinct from the impact of macroeconomic announcements and central bank intervention. Additional evidence comes from Dominguez and Panthaki’s (2006) analysis of how different types of news events influence exchange rates. They conclude that the definition of news should definitely include non-fundamental factors such as order flow.

3.4. Who is informed?

Dealers consistently stress that, on average, financial customers are informed and corporate customers are not. Indirect support for this comes from the robust finding that financial (corporate) order flow has a positive (negative) relation with contemporaneous returns (Lyons, 2001; Evans and Lyons, 2007; Marsh and O’Rourke, 2005; Bjønnes et al., 2005; King et al., 2010; Osler et al., 2011). Menkhoff et al. (2012b), for example, examine the returns to forecasts based on disaggregated daily customer order flow from a leading FX dealer over an eleven year span and find that corporate order flow produces negative payoffs while financial order flow produces positive payoffs.

The extent of information in corporate order flow has been examined directly using an approach pioneered in the broader microstructure literature. Building on the intuition that after an informed agent buys (sells) a currency its value should rise (fall) on average, this approach measures the information content of a group’s trades by their average price impact, with returns signed positive (negative) when the base currency is bought (sold). Osler and Vandrovych (2009) use this approach to examine the information content of orders placed by ten distinct groups with a large UK dealer, with horizons ranging from five minutes to one week. The authors find that the trades of both large and mid-sized corporations do not predict returns at any horizon.

The dealers’ view that financial customers are generally informed is supported by empirical studies. The aggregate order flow of financial customers is positively correlated with contemporaneous returns and financial trades do predict returns at high frequencies (Carpenter and Wang, 2007; Frömmel et al., 2008). Though financial order flow seems to be more informative than corporate order flow for high-frequency returns, Fan and Lyons (2003) suggest that this ranking may be reversed at longer horizons. This view is consistent with the findings of Evans and Lyons (2007) and Evans (2010), who find that Citibank’s corporate order flow does carry information relevant to horizons from one month to a few years.

Microstructure theories typically assume that dealers are uninformed but gain private information by observing customer order flows. A growing number of studies indicate, however, that FX dealers
bring their own private information to the market. Moore and Payne (2011) find that better-informed dealers trade more frequently, are specialized in a particular exchange rate and are located on larger trading floors and their trades have a greater price impact. Osler and Vandroyvych (2009) show that dealer order flow anticipates returns better than the trades of six distinct customer groups (including leveraged financial investors).

Finally, access to private information seems to be associated with an agent’s location, as well as an agent’s line of business. Menkhoff and Schmeling (2008) find that agents located in centers of political and financial decision-making are better informed than others, consistent with evidence for the importance of location in the broader finance literature (Hau, 2001).

3.5. How is private information acquired?

It is natural to suppose that the information carried by order flow is gained through active search. Hedge funds and other members of the active trading community are known to invest substantial time and resources in gathering market-relevant information. Of course, effort expended in gathering information does not necessarily produce useful information. The trades of retail FX traders, who definitely seek information aggressively, do not anticipate upcoming returns (Nolte and Nolte, 2012) and these investors lose money on average (Heimer and Simon, 2011). This evidence suggests that retail FX traders may improve the overall ability to provide liquidity by fulfilling the role of noise traders in the sense of Black (1986).

It is also possible, however, that information is naturally “dispersed” among agents who do not actively seek it (Lyons, 2001; Evans, 2010). Real-side fundamental information about economic activity or price levels, for example, will automatically influence the trades of corporate importers and exporters. Dealers who observe sufficient corporate trades could potentially identify that underlying economic information by observing broad patterns in corporate order flow. Alternatively, dealers could learn about financial factors such as aggregate risk tolerance by observing patterns in financial order flow (Breeden and Vitale, 2010). Changes in investment flows could reflect changes in risk tolerance, which in turn affects a currency’s “discount rate,” or shifting perceptions of a country’s economic potential, which affect a currency’s anticipated “cash flows.” Both types of shifts influence equilibrium real and thus nominal exchange rates and the influence can be lasting (Osler, 1991).

Actively-acquired information may be more influential than passively-acquired information due to structural differences in incentives and trade timing. Agents who acquire information passively are unlikely to trade on it quickly, either because they are not aware of the information or because they do not choose to engage in speculative trading, as is true at most corporate firms (Goodhart, 1988; Osler, 2006). By contrast, agents who actively acquire information know they must trade quickly to earn a profit (Holden and Subrahmanyam, 1992). In short, passively-acquired information could be learned by leveraged investors and priced into the market by the time it is unintentionally manifested in corporate or other financial order flow.

Evidence for the relevance of actively-acquired information comes from studies showing that leveraged investors flows anticipate returns (Menkhoff et al., 2012a). The relevance of passively-acquired information remains an open question. Corporate customers and most unleveraged asset managers invest little in exchange-rate forecasting (Goodhart, 1988; Bodnar et al., 1998; Taylor and Farstrup, 2006), so resolving this question will depend on the emergence of greater clarity as to whether their order flow has predictive power.

4. Liquidity and price discovery in FX markets

Liquidity provision and price discovery – perhaps the two most important functions of financial markets – have naturally been a focus of FX microstructure research. Research shows that “liquidity is priced” for traditional asset classes such as equities, meaning the most liquid assets have higher prices

---

8 Since both discount-rate and cash-flow information is generally considered fundamental in equity markets, we are comfortable considering financial information about currencies as fundamental.
and lower expected returns (Pastor and Stambaugh, 2003). A number of recent studies document that market liquidity is also priced in the cross-section of FX returns (Banti et al., 2012; Mancini et al., 2013; Menkhoff et al., 2012a). Reflecting its importance, many FX microstructure researchers have studied liquidity and documented key differences relative to equity and bond markets. These differences have important implications for exchange-rate modeling.

4.1. Market liquidity and bid-ask spreads

Liquidity, though notoriously difficult to define, appears to be well-proxied empirically by the bid-ask spread. Classic theories of liquidity provision indicate that bid-ask spreads should rise with market risk or volatility, with the expected time between trades, with adverse selection risk and the rate of information arrival, with trade size, and with dealers’ risk aversion (Ho and Stoll, 1981; Glosten and Milgrom, 1985). Bid-ask spreads on interdealer currency trades generally conform to these predictions. Glassman’s (1987) early examination of daily spreads, published by the JIMF, finds that volatility and trading volume both have a positive effect on interdealer spreads. Bollerslev and Melvin (1994), Hartmann (1998), de Jong et al. (1998) and Melvin and Yin (2000) confirm that interdealer bid-ask spreads rise with trading volumes and volatility. Melvin and Tan (1996) document that bid-ask spreads widen when social unrest and financial-market risk are heightened.

The consistent finding that rising trading volume brings rising interdealer FX spreads might seem surprising, given Demsetz’s (1968) hypothesis that higher trading volume should bring lower spreads because it reduces the waiting times between trades. However, trading volume can rise because new private information comes to the market, intensifying information asymmetries. It appears that, when trading volume rises in FX markets, interdealer bid-ask spreads are more strongly influenced by the intensification of adverse-selection risk than by the decline in waiting time.

A few studies examine the behavior of bid-ask spreads during specific events. Kaul and Sapp (2006) document that FX dealers widened bid-ask spreads from December 1999 to January 2000 as the uncertainty surrounding Y2K brought increased “safe-haven” flows and rising dealer inventories. Mende (2006) confirms that interdealer spreads widened on the day of the 9/11 attacks as uncertainty and volatility both rose dramatically. Notably, interdealer spreads reverted to normal the next day.

4.2. Liquidity and order types

The behavior of order choice in interdealer markets also generally conforms to the predictions from the broader literature on limit-order markets. Consistent with the models of Parlour (1998), an FX dealers’ choice between supplying liquidity (by submitting a limit order) and demanding liquidity (by submitting a market order) depends on the previous type of order submitted as well as volatility (Lo and Sapp, 2008). Menkhoff, (2010) show that interdealer bid-ask spreads respond to changes in market conditions much as they do in other markets. In the interdealer market for Russian roubles, limit orders are submitted relatively frequently when volatility and bid-ask spreads are high, when depth on the same side of the order book is low, and at the beginning of the trading day. The authors also find that higher trade waiting times increase the frequency of limit orders, thereby increasing liquidity. This is consistent with Glassman’s (1987) finding that spreads and liquidity rise with trading volume. The inference common to both is that FX trading volume is high when private information arrives more frequently.

Menkhoff et al. (2012a) show that the response of liquidity to market conditions is dominated by informed agents, an empirical finding that is new to the microstructure literature. They explain this in terms of picking-off risk: uninformed agents will respond less aggressively to changes in market conditions because it puts them at risk of losses to informed agents. Menkhoff and Schmeling (2010) also show that informed traders rely most heavily on market orders early in the trading day, when information flows into the market; by contrast, uninformed traders rely most on market orders late in the day when they are constrained to achieve inventory goals. This contrast between the order choices of informed and uninformed agents supports experimental results in Bloomfield et al. (2005).
4.3. Demand and supply of overnight liquidity

To model exchange rates it is important to identify which customers demand and supply liquidity during the opening (Round-1) and closing (Round-3) rounds of trading. Empirical evidence shows consistently that financial customers demand liquidity while corporate customers provide overnight liquidity. The finding that corporate order flow has a negative relation with contemporaneous exchange-rate returns while the opposite is true for financial order flow has been replicated with many different datasets (Lyons, 2001; Marsh and O'Rourke, 2005; Bjønnes et al., 2005; Evans and Lyons, 2007; King et al., 2010; Osler et al., 2011).

Intraday order-flow data allow researchers to use timing as an additional identification device. Bjønnes et al. (2005) find that financial order flow Granger-causes corporate order flow while the reverse is not true, consistent with the hypothesis that financial customers generally demand overnight liquidity while corporate customers provide it. Marsh and O'Rourke (2005) show that financial order flow does not respond to lagged returns, consistent with the behavior of Round-1 agents, while corporate order flow responds negatively to lagged returns, consistent with the behavior of Round-3 agents.

4.4. Interdealer vs. customer spreads

There are a few notable ways in which interdealer spreads in FX markets do not conform to standard microstructure theory. First, increased transparency and trading volume may lead to wider bid-ask spreads, not narrower. Hau et al. (2002) find that the percentage bid-ask spreads on the newly created EUR were wider than spreads on the DEM prior to the common currency. This result was considered surprising, given the expansion in trading volume and apparently stable information flows. Hau et al. (2002) argue that spreads widened due to the higher transparency of order flow in the interdealer market: with only one currency to trade vis-à-vis the USD, dealers had fewer options for hiding their inventory-management trades from other dealers.

The behavior of bid-ask spreads has also deviated from standard theory insofar as historically FX dealers do not “shade” their interdealer bid-ask quotes in response to inventory shifts (Bjønnes and Rime, 2005; Osler et al., 2011). That is, they did not shift prices down (up) when their inventory exceeded (fell below) the desired level (Ho and Stoll, 1981; Madhavan and Smidt, 1993). FX dealers explained that quote shading in interdealer markets would reveal information about their inventory position that could leave them vulnerable to other dealers. Instead of shading quotes, dealers preferred to unload inventory quickly in the liquid interdealer market. In today’s market, the dominant FX dealers can typically find customers with whom to trade very quickly so inventory warehousing and price shading have become standard practice. We return to this matter below.

While interdealer spreads largely conform to the predictions of standard models, the spreads quoted by dealers to customers do not. The orthodox view is that dealers protect themselves from adverse selection by widening the spreads charged to informed customers (Glosten and Milgrom, 1985; Madhavan and Smidt, 1993). But Osler et al. (2011) show that FX spreads are narrower, not wider, for the most informed FX customers – specifically customers making bigger trades and financial customers. Anecdotal evidence suggests that dealers advertise large inventory positions (known as “axes”) to informed customers at attractive bid-ask spreads in order to clear an inventory position without relying on interdealer markets.

Why might a dealer provide narrower spreads to informed customers? The discrimination in favor of customers making bigger trades could reflect their lower per-unit operating costs or their greater relative bargaining power. The narrower spreads for financial customers also reflects their greater bargaining power as financial customers tend to be informed about upcoming exchange-rate moves. Dealers have a strategic incentive to learn whether financial customers are buying or selling by trading with them (Naik et al., 1999; Osler et al., 2011). This strategic incentive is directly linked to the potential for dealers to profit from trades with informed customers, which is possible due to the two-tier structure of FX markets.
4.5. Two-tier market and spread behavior

These findings highlight the importance of market structure for the behavior of bid-ask spreads. The classic microstructure theory assumes a one-tier market in which adverse selection dominates customer bid-ask spreads (Glosten and Milgrom, 1985; Holden and Subrahmanyam, 1992). Due to adverse selection, market makers incur losses when trading against better informed customers in any market. If market makers have no way to make indirect gains from such trades, then informed traders will be charged a higher bid-ask spread when they can be identified. In one-tier markets, like the NYSE, market makers (or specialists) have no source of indirect gains and NYSE spreads thus widen as trade size rises (Peterson and Sirri, 2003).

This theory cannot be directly applied to FX, which is a two-tier market. In two-tier markets, indirect gains can be achieved when dealers use the information learned from customers in the first tier to profit on interdealer trades in the second tier. These indirect gains create an incentive for dealers to quote narrower spreads to informed customers. Adverse selection therefore appears to have little to no influence on customer bid-ask spreads.

Osler et al. (2011) model the price discovery process in the two-tier FX markets. They assume that private information originates with a subset of customers and hypothesize that price discovery in FX markets involves three stages. In Stage 1, an informed customer trades with a dealer who gains an indication of the customer’s private information. In contrast to the standard theory, this signal does not immediately affect the traded price because the informed customer pays a narrower spread than uninformed customers. In Stage 2, a dealer profits from his new information by making an aggressive trade in the interdealer market, which moves the bid-ask spread. Thereafter, price discovery within the interdealer market is hypothesized to follow the standard paradigm. In Stage 3, the prices quoted to other customers reflect the new information because they are based on the interdealer quotes. This completes the price discovery process.

There is substantial evidence consistent with this three-stage price discovery process. Osler et al. (2011) confirm that dealers are most likely to trade aggressively after informed customer trades. Goodhart and Payne (1996) and Menkhoff (2010) show that other dealers adjust their quotes in the direction of the most recent observed trade, thereby contributing to the impact of new information. This response is smaller for informed dealers, presumably because they have less to learn from the trades of others. Less informed dealers, by contrast, will even reverse the direction of their trades so that it matches the direction of aggressive dealers who are viewed as better informed.

5. Topics for future research

Having looked back at key findings from FX microstructure over the past 30 years, we next examine issues that are likely to prove important in future research. Many of these stem from the proliferation of new electronic trading platforms in recent decades.

5.1. Impact of electronic trading

Electronic trading first transformed the interdealer market when electronic messages replaced the telephone in the late 1980s. It transformed that market again a few years later when electronic limit-order markets largely replaced voice brokers in the liquid currencies. Both of these innovations brought speedier and more efficient trading. A variety of electronic trading platforms reached the customer market in the late 1990s. Proprietary single-bank platforms such as Barclays’ BARX or Deutsche Bank’s Autobahn allow customers to interact electronically with a single dealer. Anonymous multibank platforms such as Currenex and Hotspot allow customers to supply liquidity to the market, if they choose, in a limit-order market. Request-for-quote (RFQ) systems such as FXall allow customers to compare quotes from several dealing banks simultaneously. The straight-through-processing of these electronic platforms reduces operational errors and lowers trading costs. The enhanced transparency associated with RFQ systems improves customers’ negotiating power vis-à-vis dealers. Bid-ask spreads for customers that previously had low bargaining power, most notably corporate customers, tumbled the most.
Electronic trading brought substantial economies of scale to FX dealing because electronic trading platforms are expensive to design, develop, and maintain. This heavy investment has contributed to increased market concentration, as seen in Fig. 3. *Euromoney* reports that the top three banks’ share of wholesale FX trading reached 40 percent in 2010, up from only 19 percent in 1998. To exploit their expanded market share, the large dealers have developed new automated approaches to extract information from customer flows. As the large banks consolidate their information advantage relative to smaller banks, the latter have responded by focusing on trading local currencies and providing credit to customers, both activities where they still have a comparative advantage.

The concentration of market share among a handful of large FX dealers may reverse in future, as policymakers and regulators address the moral hazard problem associated with too-big-to-fail banks. While there is little to no research on this topic, the market disruption following Lehman Brothers’ bankruptcy raised questions about issues such as counterparty credit risk, the concentration of prime brokers, and the stability of the financial system (Duffie, 2013). We return to these issues below.

Electronic trading has also transformed the economics of dealer inventory management. Dealers historically tended to lay inventory off in the interdealer market, even though this usually meant they paid the bid-ask spread. But this is no longer the preferred approach, now that major banks have access to large pools of liquidity through their proprietary single-bank platforms. When dealers accumulate inventory through providing liquidity to customers they now typically hold or “warehouse” that inventory until they can lay it off on other customers. The larger dealers report that by 2010 they crossed up to 80 percent of trades internally, up from around 25 percent in 2007 (King and Rime, 2010). The largest banks’ primary source of revenues has, in consequence, shifted from speculative positioning in the interdealer market to liquidity provision for customers.

### 5.2. Retail trading

New electronic trading platforms known as retail aggregators allow individuals of modest wealth to trade FX. By bundling many small retail trades into trades that meet the minimum $1 million size for interdealer trades, retail aggregators can profit despite charging these small customers very narrow spreads. Retail FX trading, which was essentially non-existent in 2000, is estimated to have reached 8 percent to 10 percent of the market in 2010 (King and Rime, 2010). Retail traders strive to be rational, informed investors, but they lose money, on average. This raises an important question: Why does this business continue to grow? Researchers could also investigate the role that retail traders play with respect to overnight liquidity. The role these traders strive to play is that of Round-1 liquidity-demanding agents, but if they mistake noise for information (Black, 1986), they might instead serve as Round-3 agents and effectively supply overnight liquidity. The fierce competition among large banks for the business of retail aggregators, suggests that retail customers might primarily serve as Round-3 agents, since the banks essentially use their order flow to offset the liquidity demand from informed customers.

### 5.3. Algorithmic trading

Electronic trading has made it possible for order-submission strategies to be programmed and executed entirely by computers (Chaboud et al., 2009). With algorithmic (or algo) trading, humans design the program but thereafter monitor its activity and adjust trading parameters as necessary. Some form of algo trading is now used by most FX market participants, though their objectives differ widely. Institutional investors use trading algorithms to manage their trade flows more intelligently. Execution algorithms split larger trades into smaller transactions, thereby reducing price impact and transaction costs (Bertsimas and Lo, 1998). Other algorithms monitor market liquidity and depth on different electronic trading platforms across different currencies to help investors find opportunities to earn the bid-ask spread rather than paying it while achieving speculative trading goals. FX dealers use

---

9 This rising concentration also reflects the merger of large FX dealing banks, such as Swiss Bank Corporation and Union Bank of Switzerland in 1998 and JP Morgan and Chase Manhattan in 2000.
algorithms to match warehoused customer trades or to efficiently clear inventory positions. Hedge funds use algorithms to engage in macro bets, statistical arbitrage, and technical trading.

High-frequency trading algorithms use superior execution speeds to exploit tiny discrepancies in the prices or quote revisions (“latency”) across different electronic platforms. In FX, high-frequency traders concentrate on the spot FX markets for the most liquid currency pairs where they can trade with little price impact. The thousands of limit orders submitted daily by high-frequency trading firms now provide a substantial share of market liquidity. Traditional banks, finding their profitability severely squeezed, have responded in part by adopting aggressive tactics to exclude such “predatory” flows from their single-bank platforms. They have also begun providing less liquidity on multibank platforms.

5.4. Role of voice brokers

Despite the many benefits of electronic trading, dealers seem to be shifting back to using voice brokers, reversing a trend over the past decade. Voice brokers’ share of spot trade execution seems to have bottomed out and actually rose slightly (from 8 percent to 9 percent) between 2007 and 2010 (BIS, 2010). This share rose not just in emerging-markets, where voice brokers always remained important, but also in some of relatively sophisticated markets including Germany, the United States and Canada. The reason for this shift is not understood. Some suspect that voice brokers are used to manage liquidity and risk around London 4 p. m. fix (Melvin and Prins, 2010), when prices tend to be volatile and market manipulation is a concern. Others wonder whether voice brokers are used because of their relative opacity.

5.5. Electronic trading and market liquidity

The effect of electronic trading on liquidity is an important and under-researched topic. Liquidity may be viewed as a public good that benefits end-customers by reducing the rents of financial intermediaries and permitting more efficient risk-sharing. Viewed from this perspective, electronic trading has unequivocally lowered transaction costs and led to greater integration of FX markets globally. Chaboud et al. (2009) show that that algorithmic trading is associated with lower volatility in high-frequency data, which suggests that algorithmic trading is also associated with greater liquidity. At lower frequencies, liquidity could also have been affected by the fragmentation of trading across proliferating electronic platforms. In equity markets the effect of fragmentation on liquidity has been offset by legislation, specifically Regulation NMS, which requires that all equity trades take place at the national best bid or offer price (O’Hara and Ye, 2011). In the unregulated FX markets the effect of fragmentation on liquidity has been offset by the development of “liquidity aggregators,” electronic tools that collect streaming price quotes from many competing platforms and allow customers to trade...
at the best prices. If liquidity is measured by the bid-ask spread, FX liquidity likewise seems to have been sustained or improved (Mancini et al., 2013). But the bid-ask spread is only an appropriate measure of liquidity for small trades. In equity and bond markets, trade sizes in have declined along with spreads, increasing the challenges associated with executing big trades. Trade sizes also seem to have declined in FX markets. Future research could investigate whether high-frequency traders influence liquidity not just by placing limit orders but also by linking liquid pools across different trading platforms and reducing market fragmentation.

The stability of FX market liquidity is an important topic for research. Several major tail events, including the sharp depreciation of JPY in 2007, cannot be explained in terms of news. The possibility that FX markets are subject to sudden shortages of liquidity or “liquidity black holes” (Morris and Shin, 2004) follows logically from the market’s common reliance on stop-loss orders (Osler and Savaser, 2011). The stability of liquidity provided by high-frequency traders is also a source of concern.

5.6. FX market integration

The 2007–2009 financial crisis revealed the extent to which FX markets are integrated, both across borders and across asset classes globally. Though the crisis originated in the US sub-prime mortgage markets, it caused significant disruption to FX markets everywhere. FX volatility spiked to unseen levels, liquidity disappeared entirely in some currencies and instruments, and the cost of trading increased dramatically (Baba and Packer, 2009; Melvin and Taylor, 2009). Agents minimized counterparty credit risk by trading in smaller size and using shorter-dated contracts. Even at the worst, however, spot FX trading continued uninterrupted.

The integration of markets has important effects on volatility. Engle et al. (1990) show that, in daily data, high volatility in one region tends to be associated with high volatility in the next as the trading day moves around the globe, the so called “meteor shower” effect. Melvin and Melvin (2003) and Cai et al. (2008) use high-frequency data to reveal a substantial “heat-wave” effect, whereby high volatility in a given region is associated with high volatility in that same region the next day. Melvin and Melvin (2003) outline various explanations for these effects, but empirical evidence has yet to trace them to a source. Berger et al. (2009), who examine lower-frequency movements in volatility, show that the dominant influence is movement in the price impact of order flow. Little is known about the low-frequency determinants of price impact in FX markets.

5.7. Counterparty credit risk

The importance of counterparty credit risk (i.e. default risk) in FX markets was starkly highlighted by Lehman’s failure in September 2008. Customers and dealers alike pulled back from FX products and maturities that would leave them with a credit exposure to their trading counterparties. The Chicago Mercantile Exchange saw a sharp increase in activity in exchange-traded FX futures and options, which are better protected from credit risk because of centralized clearing and margin requirements. Most hedge funds, who rely on prime brokerage arrangements with dealing banks to gain access to the interbank market, saw their credit lines cut back. Other hedge funds lost assets posted as collateral in prime brokerage accounts when their prime broker, Lehman Brothers, filed for bankruptcy.

The crisis taught a number of lessons (Duffie, 2013; Melvin and Taylor, 2009). Hedge funds learned to have multiple prime brokers to avoid exposure to one sole credit provider. Regulators learned the value of central clearing, and new laws including the 2010 Dodd-Frank Act now require central clearing for many over-the-counter securities in the United States. Though FX markets have largely been exempted from this regulation, central counterparties have sprung up that offer voluntary clearing for FX products. The magnitude of counterparty credit risk in FX, and the extent to which central counterparties might mitigate that risk, are important topics for future research.

---

10 This risk is typically managed in FX using counterparty risk limits set bilaterally and master netting agreements that specify the conditions and procedures associated with default (NY Foreign Exchange Committee, 2010).
5.8. Order flow and exchange-rate modeling

FX microstructure evidence is beginning to inform the design of exchange-rate models and helpful insights have emerged for addressing long-standing macro-level puzzles. We illustrate the contribution of FX microstructure by highlighting recent research on the failure of UIP. Under UIP, equilibrium expected exchange-rate returns compensate investors for the interest rate differential and risk:

$$E[S_{t+1} - S_t] = (i_t^* - i_t) + \rho_s$$  \hspace{1cm} (2)

where $S_t$ is the (log) price of home currency in terms of the foreign currency, $i_t^*$ and $i_t$ are domestic and foreign interest rates, and $\rho_s$ is the time-varying risk premium. Economists generally infer from Equation (1) that high-interest currencies will depreciate, on average, but numerous studies show that high-interest currencies generally appreciate (Engel, 1996). Hedge funds and other financial investors exploit this regularity by borrowing low-interest currencies and investing the proceeds in high-interest currencies, a strategy known as the “carry trade”. To explain the puzzling profitability of the carry trade, economists have long focused on the risk premium in Equation (2), with risk interpreted in terms of the variance of returns (or its covariance with some market basket).

Recent papers incorporate microstructure variables to explain the initial success of the carry trade and the rapid carry-trade unwinds. Plantin and Shin (2011) note that the order flow associated with carry trades will itself tend to perpetuate the appreciation of the high-interest currency relative to the low-interest currency. Plantin and Shin (2011) develop a model that predicts the persistent positive returns to carry trades and views it as a financial bubble. This view of the carry trade is consistent with the view among FX traders that carry-trade returns are negatively skewed. In common parlance, carry-trade profits “go up by the stairs and down by the lift” (Breedon, 2001). Indeed, returns to investment (funding) currencies are negatively (positively) skewed and crashes are so common among carry-trade currencies that they are referred to as “carry-trade unwinds” (Brunnermeier et al., 2009).

While carry-trade unwinds are exogenous in the model of Plantin and Shin (2011), they emerge endogenously in the model of Osler and Savaser (2011). This model includes feedback trading consistent with market practice in addition to the influence of order flow on returns. Though not included in standard exchange rate models, positive- and negative-feedback trading are ubiquitous in currency markets and are associated with price-contingent trading strategies such as stop-loss orders (Osler, 2003, 2005). A stop-loss buy order instructs a dealer to buy a specific quantity of a currency if and when the currency’s value rises to a pre-specified level. Stop-loss orders are used by many investors prior to the release of macro statistics, and many technical traders automatically place protective “stops” whenever they open a speculative position. Customers can place these orders free of charge but their position is transparent to the FX dealer who monitors a book of such orders. Markets with stop-loss orders will be subject to rapid self-reinforcing price movements known as price cascades (Osler, 2005). Stop-loss induced price cascades are familiar to FX traders, who describe the associated currency moves as extremely rapid and “gappy,” meaning the rate will jump over levels without trading (something that happens infrequently otherwise). The overall likelihood and intensity of price cascades is increased by the empirically observed tendency of stop-loss orders to cluster in certain ways near round numbers (Osler, 2003). Price cascades are a feature of rapid carry-trade unwinds, and contribute to the negative skewness associated with carry-trade returns.

6. Concluding remarks

Our brief tour of FX microstructure research highlights how it emerged as a natural response to the empirical failure of early models of floating exchange rates. Due to the absence of historical experience, early macro models were designed inductively. As time went on, however, most of the elegant assumptions and implications of these models were falsified by a growing body of evidence, paving the way for the development of more micro-founded models.

Microstructure researchers adopted a deductive approach to understanding exchange rates. They surveyed FX market participants and studied large, hand-collected data sets of high-frequency data. This effort led to the insight is that FX order flow is the most powerful single driver of exchange rates.
This finding, in turn, has led to a number of research agendas. It is now recognized that FX traders hold heterogeneous beliefs and that private information is a key source of the influence from order flow to exchange rates. Financial customers appear to be the best informed and tend to demand overnight liquidity. Corporate customers, whose trades do not typically carry information, serve a crucial function nonetheless because they provide overnight liquidity. In short, the interaction between informed and uninformed agents is now recognized as an essential mechanism driving short-run exchange-rate dynamics.

FX market microstructure research has also shown us that the market's structure differs in important ways from equity and bond markets. This distinction underscores the need for researchers to be selective in their reliance on the broader microstructure literature when developing exchange rate models.

Given the dramatic changes associated with the rise of electronic trading and entry of new participants over the past decade, many established relationships may need to be revisited while the number of open questions has multiplied. No doubt FX researchers will continue to look to the JIMF for leadership in publishing the innovative and controversial articles that will move the field forward over coming decades.

Appendix A

Table A
Peer reviewed articles on FX microstructure in top journals, 1982–2012.

<table>
<thead>
<tr>
<th>Journal</th>
<th>(1) Number of scholarly articles</th>
<th>(2) (Exchange W1 rate) OR (foreign W1 exchange) in abstract</th>
<th>(3) Microstructure OR (order W1 flow) in abstract</th>
<th>(4) Column 3 as % of total</th>
</tr>
</thead>
<tbody>
<tr>
<td>JIMF</td>
<td>1510</td>
<td>776</td>
<td>30</td>
<td>2.0%</td>
</tr>
<tr>
<td>JBF</td>
<td>3807</td>
<td>175</td>
<td>7</td>
<td>0.2%</td>
</tr>
<tr>
<td>JIE</td>
<td>2033</td>
<td>374</td>
<td>7</td>
<td>0.3%</td>
</tr>
<tr>
<td>JFM (1998–2012)</td>
<td>282</td>
<td>11</td>
<td>4</td>
<td>1.4%</td>
</tr>
<tr>
<td>JFE</td>
<td>2092</td>
<td>23</td>
<td>3</td>
<td>0.1%</td>
</tr>
<tr>
<td>JF</td>
<td>3399</td>
<td>75</td>
<td>2</td>
<td>0.1%</td>
</tr>
<tr>
<td>JFQA</td>
<td>1226</td>
<td>30</td>
<td>1</td>
<td>0.1%</td>
</tr>
<tr>
<td>JEF</td>
<td>6174</td>
<td>172</td>
<td>1</td>
<td>0.0%</td>
</tr>
<tr>
<td>JPE</td>
<td>1736</td>
<td>37</td>
<td>1</td>
<td>0.1%</td>
</tr>
<tr>
<td>RFS (1988–2012)</td>
<td>1401</td>
<td>23</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>QE</td>
<td>1421</td>
<td>35</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Note: Column 1 shows the number of peer-reviewed (scholarly) articles published between 1982 and 2012 in eleven top finance and economics journals based on a search using Business Source Complete. Column 2 shows the number of FX articles based on a search of the abstract for “exchange rate” or “foreign exchange”. Column 3 shows the number of FX microstructure articles from column 2 that include the words microstructure or order flow in the abstract. Column 4 shows the FX microstructure articles as a percentage of all articles published. The journals, shown in descending order based on column 3, are: Journal of International Money and Finance (JIMF), Journal of Banking and Finance (JBF), Journal of International Economics (JIE), Journal of Financial Markets (JFM), Journal of Financial Economics (JFE), Journal of Finance (JF), Journal of Financial and Quantitative Analysis (JFQA), American Economic Review (AER), Journal of Political Economy (JPE), Review of Financial Studies (RFS), and Quarterly Journal of Economics (QE).

References


