Maximilian Butz, Roel Oomen

Internalisation by electronic FX spot dealers

Article (Accepted version) (Refereed)

Original citation:
Butz, M. and Oomen, R. (2018) *Internalisation by electronic FX spot dealers.* Quantitative Finance. ISSN 1469-7688

DOI: 10.1080/14697688.2018.1504167

DEUTSCHE BANK

SALES AND TRADING DEPARTMENT

Internalisation by electronic FX spot dealers

Maximilian Butz* Roel Oomen*

November 2017

Abstract

Dealers in over-the-counter financial markets provide liquidity to customers on a principal basis and manage the risk position that arises out of this activity in one of two ways. They may internalise a customer's trade by warehousing the risk in anticipation of future offsetting flow, or they can externalise the trade by hedging it out in the open market. It is often argued that internalisation underlies much of the liquidity provision in the currency markets, particularly in the electronic spot segment, and that it can deliver significant benefits in terms of depth and consistency of liquidity, reduced spreads, and a diminished market footprint. However, for many market participants the internalisation process can be somewhat opaque, data on it is scarcely available, and even the largest and most sophisticated customers in the market often do not appreciate or measure the impact that internalisation has on their execution costs and liquidity access. This paper formulates a simple model of internalisation and uses queuing theory to provide important insights into its mechanics and properties. We derive closed form expressions for the internalisation horizon and demonstrate – using data from the Bank of International Settlement's triennial FX survey – that a representative tier 1 dealer takes on average several minutes to complete the internalisation of a customer's trade in the most liquid currencies, increasing to tens of minutes for emerging markets. Next, we analyse the costs of internalisation and show that they are lower for dealers that are willing to hold more risk and for those that face more price sensitive traders. The key message is that a customer's transaction costs and liquidity access are determined both by their own trading decisions as well as the dealer's risk management approach. A customer should not only identify the externalisers but also distinguish between passive and aggressive internalisers, and select those that provide liquidity compatible with their execution objectives.

*The authors are employed within the electronic FX spot trading division of Deutsche Bank A.G. Deutsche Bank (DB) is an industry recognised world leader in the foreign exchange business, and offers a full spectrum of foreign exchange products and services, including the trading of foreign exchange products through its Autobahn electronic trading platform. This paper was prepared within the Sales and Trading function of DB, and was not produced, reviewed or edited by the DB Research Department. The views and opinions rendered in this paper reflect the authors' personal views about the subject. No part of the authors' compensation was, is, or will be directly related to the views expressed in this paper. See Disclaimer at end of paper. The authors would like to thank colleagues at DB, Søren Johansen, Vasco Leemans, Richard Louth, Michael Melvin, and seminar participants at the London School of Economics, The London Quantitative Finance Seminar at the CFM-Imperial Institute of Quantitative Finance, INET "Microstructure workshop" at Trinity College Cambridge, the MathFinance Conference 2017 in Frankfurt, TradeTech FX 2017 in Miami, the BIS Working Party on Markets in Buenos Aires, and Jim Gatheral's 60th Birthday Conference at the Courant Institute New York University for helpful comments.

1 Introduction

The foreign exchange market is increasingly polarised with dealers categorised into one of two camps: those that "externalise" and those that "internalise" customer flows. Externalisation refers to the process where a dealer – in their capacity as a liquidity provider – hedges customer trades directly in the open market. Even though the dealer may operate on a principal basis, their hedging activity can make them observationally equivalent to that of an agent or a provider of market-access technology. Internalisation is the polar opposite: here a dealer holds the risk arising from a customer trade until an offsetting trade is received thereby avoiding the need to trade in the external market. The internalising dealer provides immediacy of execution to their customers and effectively sources liquidity by intermediating customer flows across time and bearing the risk over the intervening period.[1]

Despite the fundamental differences between these two approaches, and the widespread adoption of both, in practice customers are often unaware of the underlying mechanics by which the dealer sources liquidity, and are unable to judge the impact this can have on their execution costs and liquidity access. Those that do appreciate the distinction often lack a sufficiently detailed understanding to let them distinguish between dealers in each category. This is less important when choosing between externalisers – they are much alike and typically only differ by the trading speed and routing logic with which they access an ultimately common set of liquidity sources – but more so when differentiating between internalisers: the liquidity pools they maintain are unique and so are the internalisation strategies they adopt. Furthermore, while internalisation is viewed by many as fundamental to the efficient working of the currency market, information on it is scarcely available because the detailed logic is generally considered commercially sensitive intellectual property, and transaction-level trade data is often subject to bi-lateral confidentiality agreements.

The objective of this paper is to contribute towards a better understanding of internalisation. We study the mechanism by which dealers internalise customer flows, compare and contrast different methodologies, and provide a comprehensive analysis of the time scales and costs involved in the operation of this process. The paper builds on the recent work done by the Bank of International Settlements on this topic, in particular the most recent triennial market survey that includes a specific question on the prevalence of internalisation. The results are dis-

[1] An example of externaliser behaviour is found in Virtu Financial, Inc (2014, p. 2) "Our strategies are also designed to lock in returns through precise and nearly instantaneous hedging, as we seek to eliminate the price risk in any positions held". The practice of internalisation is referenced in Bank of England, H.M. Treasury, and Financial Conduct Authority (2014, p. 59) "Market participants have indicated that some dealers with large enough market share can now internalise up to 90% of their client orders in major currency pairs". The notion that dealers are either perfect internalisers or perfect externalisers is of course too constraining and in practice they will use to varying extent a mix of both to manage their risk. This is likely to be true even at a more granular level, e.g. by individual customer, currency pair, time-zone, etc. The distinction between internalisers and externalisers is nevertheless an important one.

cussed in Moore, Schrimpf, and Sushko (2016). They report a use of internalisation that varies greatly by product and access channel as well as by dealers and jurisdictions,[2] a finding that can be explained by the theory presented in this paper. We further dissect the market survey results to make predictions about the time-scales and risks faced by the dealers when internalising flows in the electronic spot market.

The approach we adopt to study internalisation draws on queuing theory: an established field in the operations research branch of mathematics literature and one that has numerous practical applications.[3] We consider a model where the dealer provides liquidity to their customer-base by internalising the risk it absorbs as part of this activity, i.e. they builds up a long position (or reduces a short position) over episodes where their customers are selling and vice versa. A simple reformulation clarifies the applicability of queuing theory: instead of classifying a trade as a buy or a sell and the dealer's position as long or short, we categorise each trade as either increasing or decreasing the dealer's risk position. The time it takes for a risk increasing trade (the "customer") to be internalised by a risk decreasing trade ("serviced") once it has entered the dealer's risk position (the "queue") is what defines the internalisation horizon ("queuing time"). To control the build-up of risk, the dealer can skew their prices to encourage risk-reducing customer flow. This is referred to as position skewing, and can be thought of as the equivalent of deploying additional staff to process the queue.

Within this setup, we make two key contributions. The first is a comprehensive characterisation of a dealer's internalisation horizon. Leveraging a fundamental result in queuing theory – known as Little's Law (Little, 1961) – we show that when a dealer uses position skewing to control risk, the average internalisation horizon is proportional to the ratio of their risk limit over the notional trade flow rate. Internalisation is therefore quicker (and less risky) for larger dealers and over more active trading periods, a prediction that is consistent with the empirical findings of BIS (2016). Smaller dealers, or those with limited risk appetite, can reduce internalisation horizons by more pro-actively soliciting risk reducing flow via more aggressive position skewing. But because this incurs a cost – both direct via reduced spread capture on risk decreasing deals, and indirect via information leakage or signalling risk associated with skewing – those dealers are at a disadvantage to internalise compared to their larger counterparts who benefit from economies of scale and reduced costs. It also highlights that internalisation operates across a continuum from passive to aggressive internalisation: not every internaliser is the same. Next, we take the BIS (2016) foreign exchange market survey results, concentrate on the electronic spot segment, and then apply the developed theory to infer typical internalisation horizons for a representative tier-1 passive internaliser. Our model suggests that even

[2] The highest rates of internalisation are attained in electronic spot, particularly by larger dealers in more active trading centres.

[3] For example, queuing theory is used in the analysis of shopping queues or traffic congestion, design of factory assembly lines, allocation of staffing levels in a hospital's emergency department, and the scheduling and load balancing of large scale calculations or high volumes of search queries across a cluster of servers. See, e.g., Harris (2010); Hopp and Spearman (2000); Wolff (1989).

for the most liquid G10 currency pairs, over the most active period of the day, a tier-1 dealer would still take several minutes to internalise an average sized trade. For less liquid pairs, or over the Asia trading session, internalisation horizons often increase to tens of minutes or even several hours. This counters the common perception that liquidity in the currency markets is virtually unlimited and that the larger dealers are able to find offsetting flows in a matter of seconds if not instantaneously. The quantitative results presented here also provide a practical reference point to the customer who wants to space out their trades in order to give the dealer a fair chance to fully internalise the flow: trading too quickly will stress the dealer's capacity to internalise, whereas trading too slowly will expose the trader to unnecessary market risk.

The second contribution concentrates on the mechanics and costs associated with internalisation. We start with a model similar to that of Oomen (2017a) where two liquidity providers compete for a trader's flow[4] but then allow for two new features (a) the trader can be informed and acts on the perceived dealers' mis-pricing and (b) the liquidity providers' prices explicitly include a position skew used to control their risk exposure. For simplicity we assume that all the trader's requests to deal are accepted by the liquidity provider.[5] We derive explicit expressions for the effective spread and then use this to formulate a measure of the costs associated with position skewing. We show that these costs increase with a reduction in the dealer's risk appetite, but decrease with the informedness of the trader. An extension allowing for asymmetric position skewing between the bid and offer rates, demonstrates that costs of internalisation can be very substantial for dealers with limited risk bearing capacity.

An important message is that the polarisation between internalisers and externalisers is overly simplistic and that instead one needs to view liquidity providers across a continuum with passive internalisers at one end through to aggressive internalisers and eventually externalisers at the other end (in the same way that the market impact produced by liquidity providers is not binary). We show that the execution costs incurred by the customer are directly influenced by the risk management actions of the dealer and that the time-scales over which a trader executes should be aligned with the dealer's internalisation horizon. Any sustained and systematic mismatch between the two will either lead to unnecessary market risk borne by the trader when they executes too slowly, or increased transaction costs due to the liquidity provider needing to risk manage the trader's excessive flow via aggressive position skewing or even externalisation. A careful choice of liquidity provider with a risk management approach that is compatible with the trader's execution objectives is therefore key in the pursuit of efficient execution and optimal

[4]Throughout the paper we use the terms "dealer" and "liquidity provider" interchangeably, and the same for "trader" and "customer".

[5]This assumption is made to simplify exposition. In practice, the dealer reserves the right to reject incoming trade requests if a set of predefined validation checks is not passed. This is commonly referred to as "last-look", see Oomen (2017b) for further discussion. Also, a trader would routinely place more than two dealers in competition for their flow. Oomen (2017a) studies such an execution setup. Generalisations along those lines do not change our basic findings.

liquidity access.

The applicability of the results presented in this paper extend beyond electronic FX spot. Internalisation is a fundamental mechanism for dealers in virtually all over-the-counter markets (see, e.g., Kirk, McAndrews, Sastry, and Weed, 2014, for a discussion of the role internalisation plays in the collateralised debt financing markets). The typical flow rates will vary across these markets and the position skewing policies adopted by the dealers may also differ, but the basic queuing theory results regarding internalisation horizons will still apply. As such, these may be used by regulators to gauge the risks borne by dealers in these markets, and by customers to measure the time-scales that dealers require to internalise their flow.

The remainder of the paper is organised as follows. Section 2 develops the theory needed to establish a dealer's internalisation horizon, followed by an application that uses BIS (2016) survey results to quantify typical internalisation horizons for a tier-1 dealer in the electronic FX spot market. Section 3 analyses the costs associated with internalisation. Section 4 concludes. The appendix contains the proofs.

2 Internalisation horizon

2.1 A queuing theory approach

Let X_t, $t \geq 0$ denote the position of a liquidity provider (LP) at time$-t$, defined as the accumulation of completed buy- and sell-transactions of unit size:

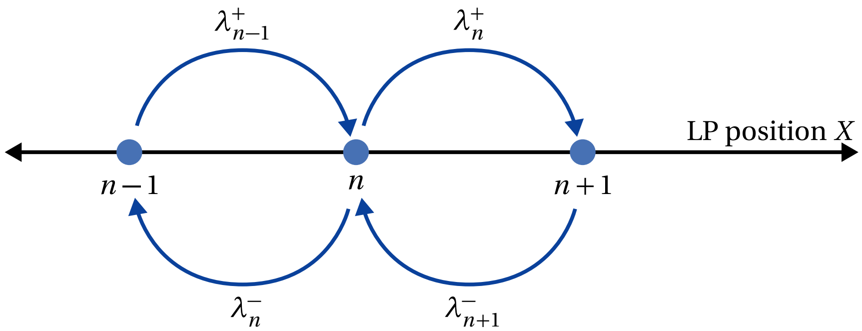
$$X_t = N_t^+ - N_t^-, \tag{1}$$

where N_t^+ and N_t^- count the number of buy and sell transactions up to time$-t$. We assume that X follows a compound Poisson process with position dependent arrival rates of buys and sells, i.e. when $X_t = n$, then $E(dN_t^+) = \lambda_n^+ dt$ and $E(dN_t^-) = \lambda_n^- dt$. See Figure 1 for an illustration. Thus, when the LP's position at time-t is n, they expect a buy transaction after an exponentially distributed waiting time with expectation $1/\lambda_n^+$ and a sell transaction after an exponentially distributed waiting time with expectation $1/\lambda_n^-$, whichever comes first. dN_t^+ and dN_t^- are assumed to be independent given $X_t = n$.

The position dependence of the arrival rates of buys and sells is necessary to guarantee a stationary distribution for X_t: if the rates were equal and constant, the LP's position would follow a random walk. We assume the LP can control the arrival rate of buys and sells by skewing their prices and does so in a manner that encourages risk reducing trades once the position exceeds some specified threshold $n^* \geq 0$, i.e.

$$\lambda_n^+ < \lambda_n^- \quad \text{when} \quad n > n^* \qquad \text{and} \qquad \lambda_n^+ > \lambda_n^- \quad \text{when} \quad n < -n^*. \tag{2}$$

Figure 1: Illustration of the LP's position state transition



The stationary distribution of the LP's position is defined as $\phi_n \equiv \lim_{t\to\infty} \mathbb{P}(X_t = n)$ and – assuming it exists – can be identified via the global balance equation of a Markov chain:

$$\lambda^+_n \phi_n = \lambda^-_{n+1}\phi_{n+1} \tag{3}$$

where $n \in \mathbb{Z}$. This yields:

$$\phi_n = \phi_0 \prod_{k=1}^{n} \frac{\lambda^+_{k-1}}{\lambda^-_k} \quad \text{for} \quad n>0 \qquad \text{and} \qquad \phi_n = \phi_0 \prod_{k=1}^{n} \frac{\lambda^-_{1-k}}{\lambda^+_{-k}} \quad \text{for} \quad n<0, \tag{4}$$

with ϕ_0 determined by $\sum_{n\in\mathbb{Z}} \phi_n = 1$.

Example 1 (binary position skew) *Let* $\lambda^+_n + \lambda^-_n = 2\lambda_0$ *with*

$$\lambda^+_n = \begin{cases} \lambda_0 & |n| \le R \\ \lambda_0(1-\alpha) & n > R \\ \lambda_0(1+\alpha) & n < -R \end{cases} \tag{5}$$

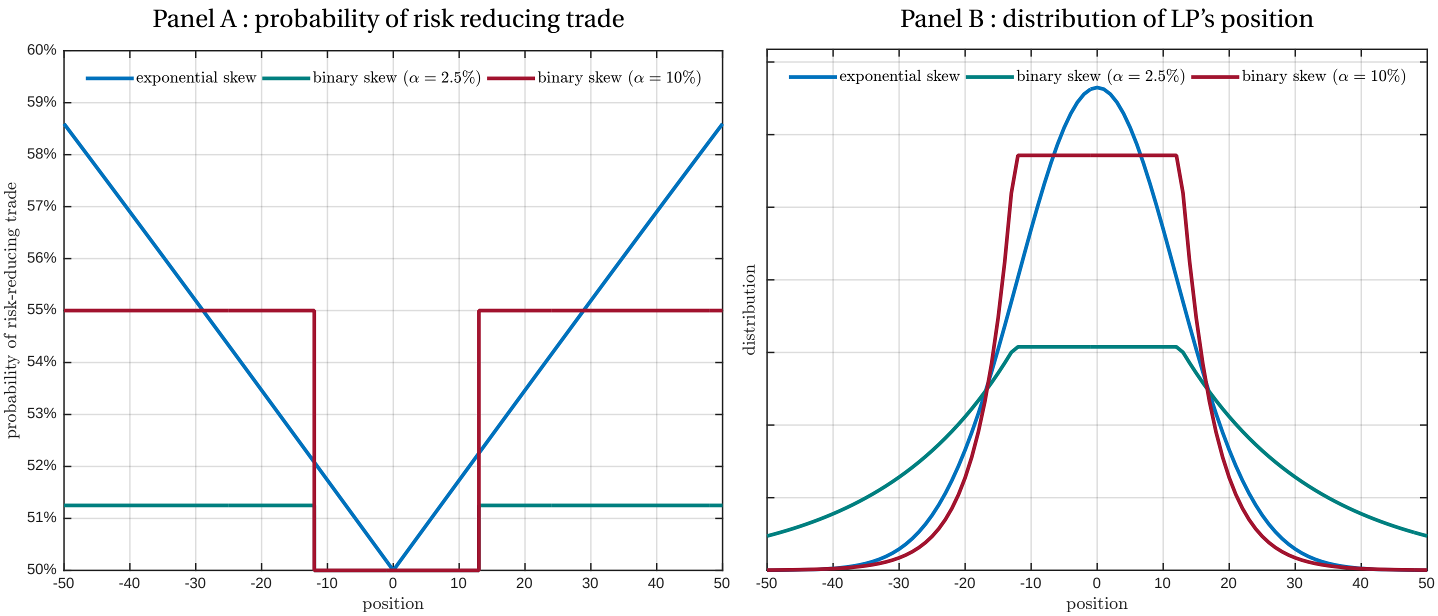
for some fixed threshold $R \ge 0$ *and* $0 < \alpha \le 1$. *Buys and sells arrive at the same constant rate when the LP's absolute position is below the threshold* R*, but beyond that risk reducing trades are stimulated and arrive with probability* $(1+\alpha)/2$. *The distribution of the LP's position is stationary and given (by application of Eq. 4) as:*

$$\phi_n = \phi_0 \quad \textit{for} \quad |n| \le R \qquad \textit{and} \qquad \phi_n = \frac{\phi_0}{1+\alpha}\left(\frac{1-\alpha}{1+\alpha}\right)^{|n|-1-R} \quad \textit{for} \quad |n| > R, \tag{6}$$

where $\phi_0 = (2R+1+\alpha^{-1})^{-1}$.

Example 2 (exponential position skew) *Let* $\lambda^+_n = \lambda^-_{-n} = \lambda_0 e^{-\frac{1}{2}n/R^2}$ *for* $R > 0$. *Buys and sells are equally likely only when the LP's position is zero and otherwise risk reducing (increasing) trades arrive at a rate that grows (falls) with*

Figure 2: Distribution of the LP's position for different skew methods



Note. Panel A draws the probability of a risk reducing trade (i.e. $\lambda_n^{-\text{sign}(n)}/(\lambda_n^- + \lambda_n^+)$) as a function of the LP's position n for the binary and exponential skew method as defined in Examples 1 & 2 with $R = 12.75$. Panel B draws the corresponding stationary distribution of the LP's position.

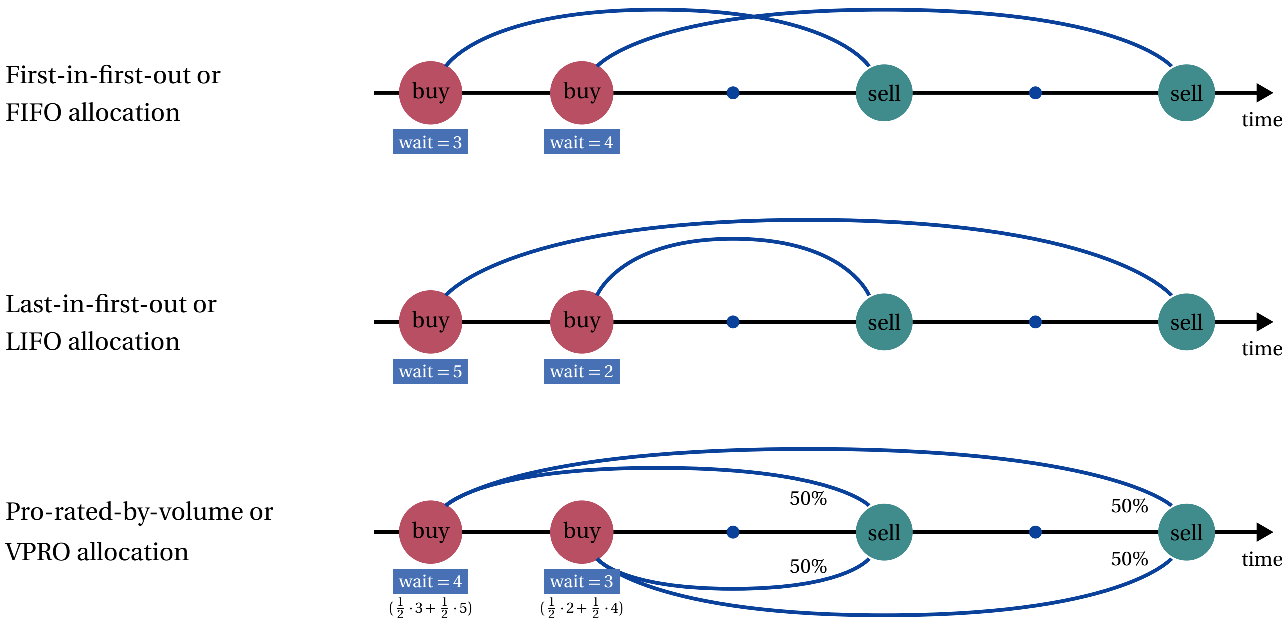
the magnitude of the LP's position. The distribution of the LP's position is stationary and given (by application of Eq. 4) as:

$$\phi_n = \frac{1}{\sqrt{2\pi R^2}} e^{-\frac{1}{2}n^2/R^2}. \tag{7}$$

Figure 2 provides a graphical illustration of these two examples.

The process of internalisation requires the LP to temporarily absorb risk-increasing trades into their inventory in anticipation of future risk-reducing trades, possibly aided by position skewing for risk management purposes. The LP provides immediacy of execution to the trader by agreeing to an instantaneous risk transfer which is subsequently managed as part of the dealer's overall risk position. Within this context, the internalisation horizon is defined as the length of time a given trade forms part of the LP's risk position before it is offset by another trade in the opposite direction. Analysis of the mathematical properties of the internalisation horizon can get quite involved and quickly intractable because it varies – for instance – with the specific methodology used to match risk increasing with risk decreasing trades, e.g. a last-in-first-out (LIFO) allocation methodology will likely produce a materially different distribution of internalisation horizons to that of a first-in-first-out (FIFO) method, pro-rated-by-volume (VPRO) method, or any other time- and/or size-weighted allocation method for that matter (see Figure 3 for an illustration).

Figure 3: Illustration of internalisation times for different allocation methodologies



Note. In this illustration, it is assumed that the liquidity provider's position starts at zero, that trades are for unit size and time increments in one-period steps. The first two trades are risk increasing, whereas the last two are risk decreasing. The time to internalise the risk increasing trades is labelled as "wait".

This is where queuing theory comes in, and in particular a fundamental result known as Little's Law: it states that the average number of customers in a queuing system, denoted L, equals the average arrival rate of customers to the system, λ, multiplied by the average waiting time of a customer in the system, W, or $L = \lambda W$ (Little, 1961, 2011). Crucially, even when the distribution of waiting time is intractable it allows one to establish the average waiting time independent of the arrival process, service distribution, servicing or allocation priorities, etc. Drawing a parallel to the LP's internalisation horizon: we interpret the absolute position $|X_t|$ as the length of the queue, and the arrival rate of risk increasing trades as the arrival of "customers". The expectation of these quantifies can be calculated as:

$$L = \sum_{n\in\mathbb{Z}} \phi_n |n| \qquad \text{and} \qquad \lambda = \sum_{n\geq 0} \phi_n \lambda_n^+ + \sum_{n\leq 0} \phi_n \lambda_n^-. \tag{8}$$

The average internalisation horizon for a risk increasing trade is then given by Little's law as L/λ. Because the internalisation of risk-reducing trades is instantaneous, the average internalisation horizon for any trade – unconditioned on whether it is risk-increasing or risk-decreasing – is:

$$W = \frac{1}{2}\frac{L}{\lambda}. \tag{9}$$

Proposition 1 (Internalisation horizon) *With a binary position skew as specified in Example 1, the LP's average internalisation horizon of a trade is:*

$$W = \frac{R}{\lambda_0}\frac{c_R}{4} \tag{10}$$

where $c_R = \frac{2\alpha^2 R + 2\alpha(1+\alpha)+(1+\alpha)/R}{2\alpha^2 R+\alpha(1+\alpha)} \to_R 1$. *With an exponential position skew as specified in Example 2, the LP's average internalisation horizon of a trade is:*

$$W \approx \frac{R}{\lambda_0}\frac{c_R}{\sqrt{2\pi}} \tag{11}$$

where $c_R = 2(1-\Phi(\frac{1}{2}R^{-1}))e^{1/(8R^2)} + 1/\sqrt{2\pi R^2} \to_R 1$.

Proof See Appendix A. ■

See Figure 4 for an illustration. We see from Panels B – D that the distribution of the time it takes to internalise a trade varies substantially depending on the allocation methodology used. For instance, if a risk decreasing trade is assumed to close out the most recent risk increasing trade (i.e. LIFO), then the distribution of internalisation horizons is more skewed, fatter tailed, and has a lower median compared to when the FIFO method is used to match trades. But provided the LP's position starts and ends at zero, the average internalisation is numerically identical over every sample path realisation. This property of Little's Law is crucial – it lets us apply standard queuing theory to make reliable inference about internalisation horizons without making any assumptions on how the LP conduct their "internal accounting" when matching risk reducing with risk increasing trades.

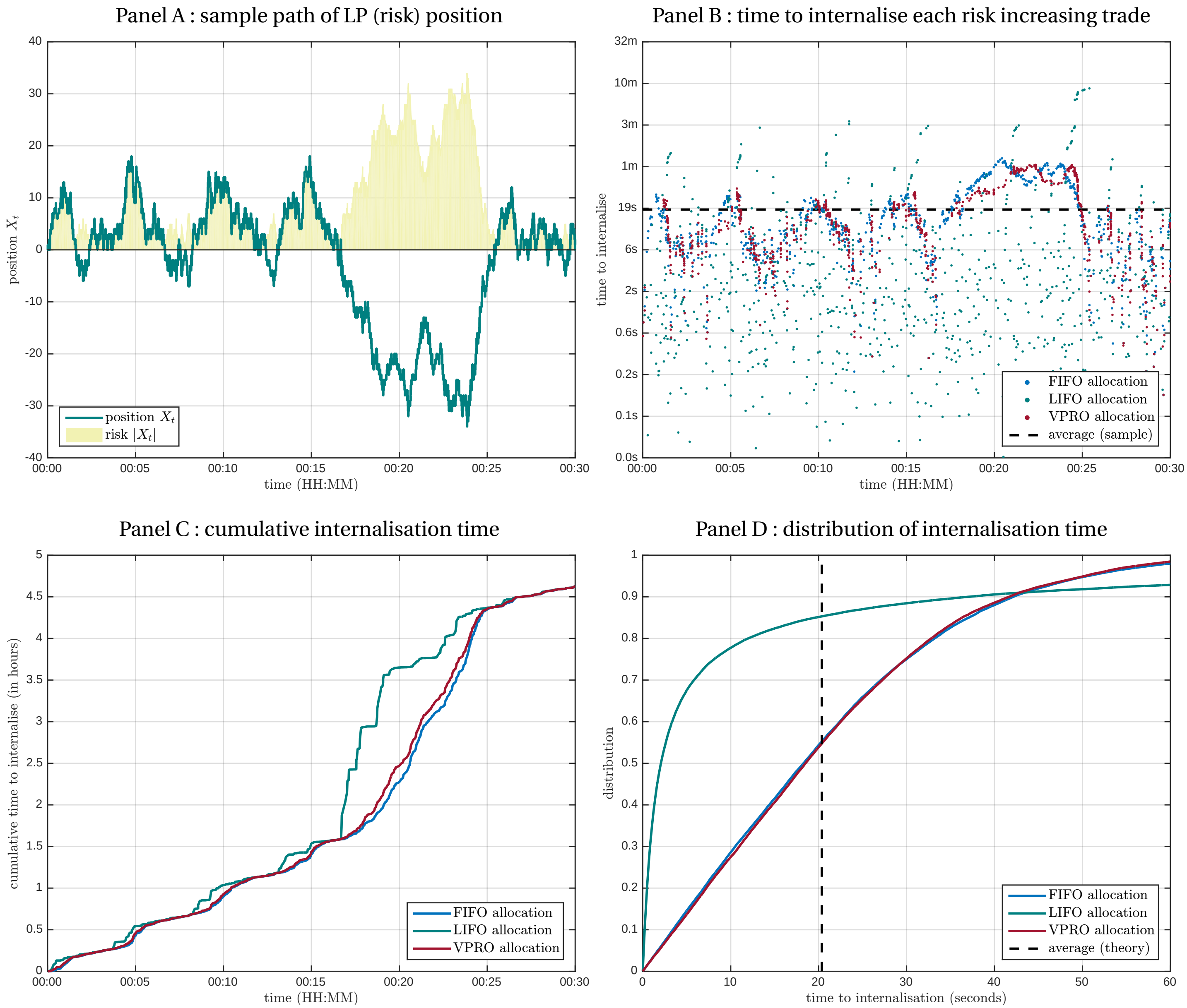
2.2 Internalisation horizons in practice

We now apply the above theory to develop some understanding of what the typical internalisation horizons would be in practice for a large tier-1 liquidity provider in the segment where internalisation is most prevalent, i.e. the electronic FX spot market. For this purpose, we require estimates of the trading flow rates the LP is expected to process. Because customer and dealer trading activity data is difficult to obtain at the required granularity and also to avoid introducing dealer specific biases, we adopt a top-down approach using the statistics published by the BIS (2016) triennial survey. Our starting point is the average daily volume (ADV) figure of \$5,067bn for the FX market as a whole. Our interest is in FX spot only, so we exclude volumes in forwards, swaps, options and other derivatives and this leaves an ADV of \$1,652bn (BIS, 2016, Table 1).[6] Because internalisation only applies to customer dealing,[7] we

[6] Hasbrouck and Levich (2017) propose a correction for the double counting of prime brokerage volumes in the raw BIS figures. If applied here, it would reduce volume numbers by about 17%.

[7] In the BIS survey customers include financial institutions (e.g. non-reporting commercial and investment banks, security houses, leasing companies, financial subsidiaries of corporate firms), real money (e.g. mutual funds, pension funds, asset and wealth managers, currency

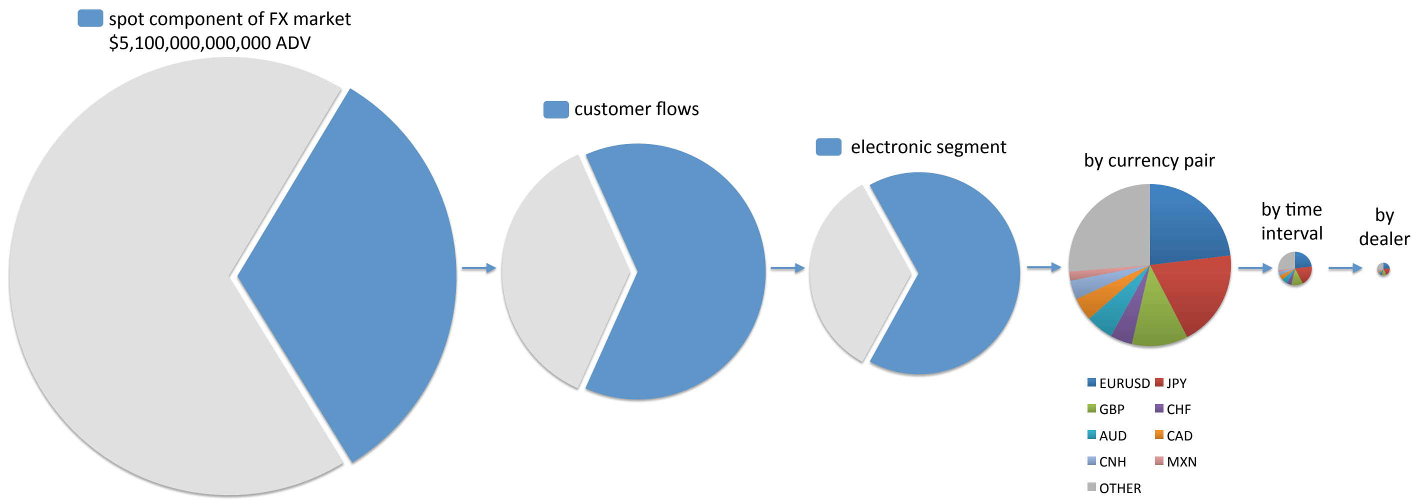
Figure 4: Illustration of internalisation time of risk increasing trades



Note. We simulate from the exponential position skew model as in Example 2 with $R = 12.75$ and $\lambda_0 = \frac{1}{2}$/second (i.e. a trade occurs on average once per second) and establish the internalisation time of the risk increasing trades via the first-in-first-out (FIFO), last-in-first-out (LIFO), and pro-rated by volume (VPRO) allocation methodologies as illustrated in Figure 3. Panels A – C report the results for one 30-minute sample path. Panel D draws the underlying distribution of internalisation time.

funds, money market funds, building societies, insurance and reinsurance companies, endowments), hedge funds and proprietary trading firms (e.g. commodity trade advisers, high frequency trading firms, global macro funds), official sector (e.g. central banks, sovereign wealth funds, development banks and agencies, financial public sector institutions), non-financial (e.g. corporates, non-financial public sector in-

Figure 5: Decomposition of the FX market size



Note. Starting from aggregate FX market-wide figures from BIS (2016), this series of pie-charts progressively isolates the customer-facing electronic FX spot segment, and then further breaks this down by currency pair, hourly time-interval, and tier-1 dealer (assuming 10% market share). The surface area of the pie charts are proportional to $-volume. In the currency pair break down, for each currency listed, the results reflect the combined activity across the EUR and USD crosses, e.g. "JPY" consolidates EURJPY and USDJPY.

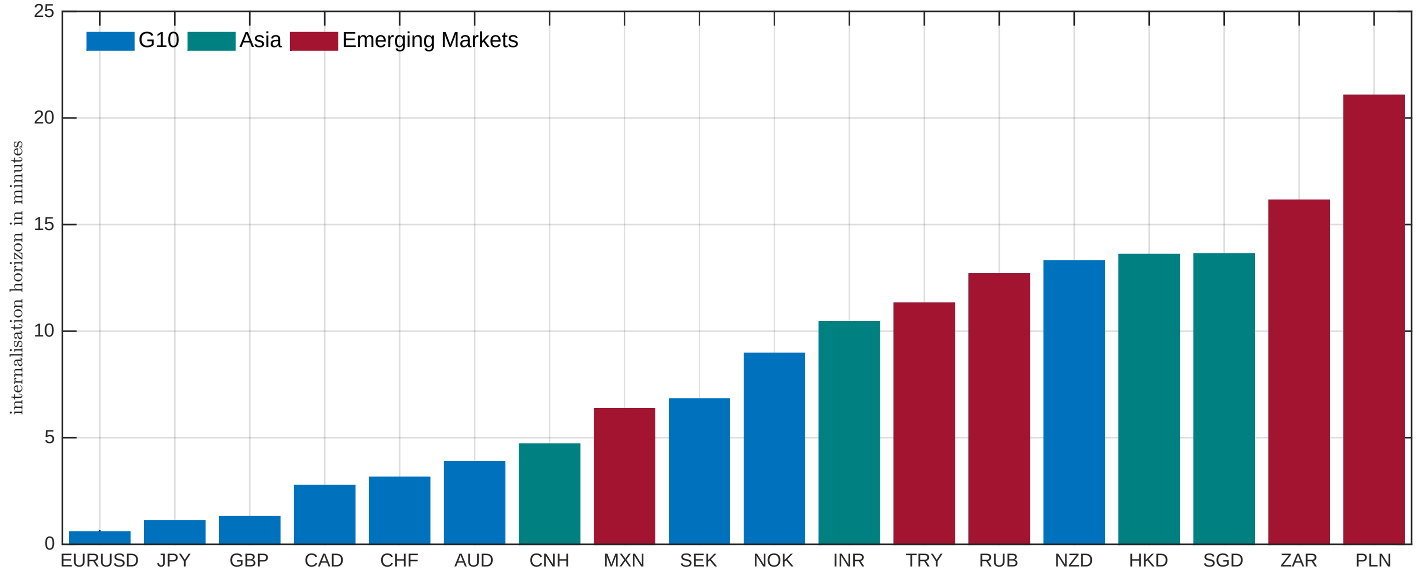
exclude inter-dealer activity from this figure which further lowers the relevant ADV to $1,047bn (BIS, 2016, Table 4 & 5). Next, we separate out electronic flows from voice. The nature of voice flows (large notional and infrequent) and the narrow currency coverage of individual voice traders makes internalisation often impractical or uneconomical. This is different for the electronic flows where individual tickets tend to be of smaller size, come in at a higher rate, and dealers can more easily manage risk at a portfolio level across a broad spectrum of currency pairs. The fraction of electronically executed flow in FX spot is measured at 66% by BIS (2016, see detailed Table 1). This brings us to an ADV figure of $692bn. As a final step, we break this down by currency pair using the market shares reported in Table 3 of BIS (2016). See Figure 5 for a graphical illustration of this decomposition of the FX market size into the segment that is relevant for the analysis of FX spot internalisation horizons.

Let us consider an example. The Euro-Dollar (EURUSD) currency pair accounts for 23.1% of trading volumes and so this translates into a flow rate per minute across the market wide electronic FX spot segment of

$$23.1\% \times \$692\text{bn/day} = \$159{,}852\text{mn}/1440\text{min} = \$111\text{mn/min}.$$

stitutions), and other (e.g. retail aggregators). The BIS defines the internalisation ratio as "the percentage of reported total foreign exchange turnover which was internally matched against offsetting trades by other customers." (BIS, 2015, Table 9). Applying this definition to the aggregate BIS numbers for spot we infer an market wide internalisation ratio of 1,047/1,652 = 63%. This coincides perfectly with the average internalisation ratio across FX spot dealers as reported by Moore, Schrimpf, and Sushko (2016).

Figure 6: Average internalisation horizon by currency pair



Note. The chart draws the internalisation horizon – as implied by the exponential skew model with $R = 12.75$mn and industry volume and market share figures as discussed in the text – by currency pair over the most active trading session. See also column "min" in Table 1. For each currency listed, the results reflect the combined activity across the EUR and USD crosses, e.g. "AUD" consolidates EURAUD and AUDUSD.

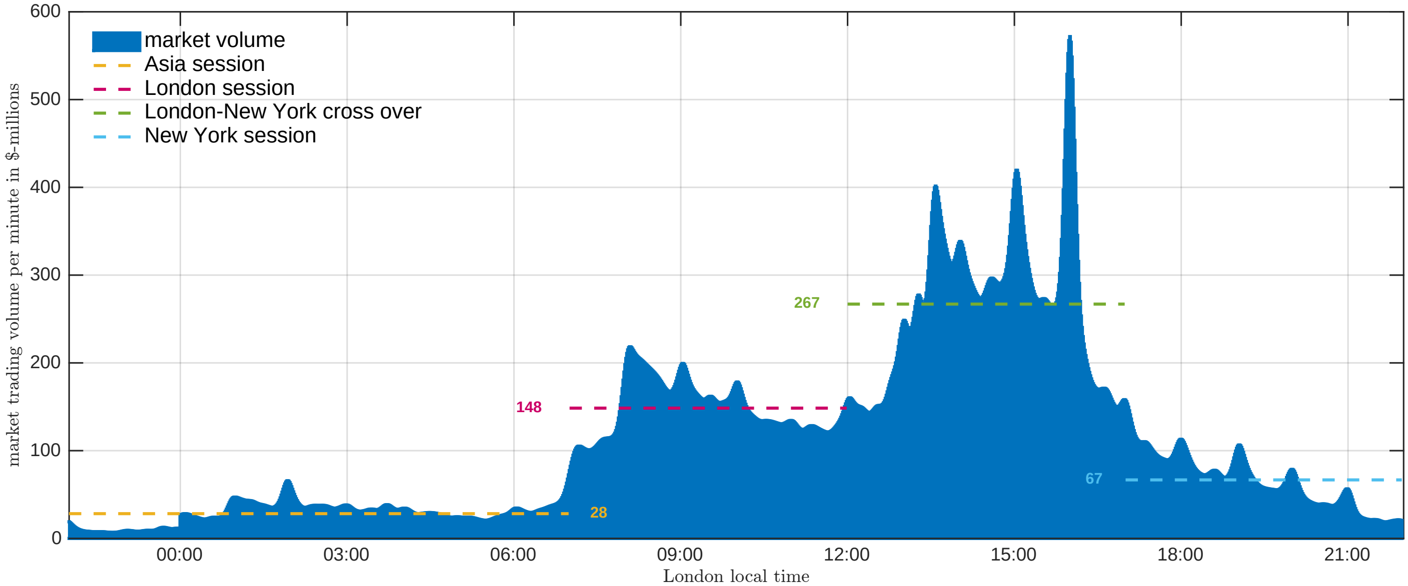
Any individual dealer will only capture a fraction of this figure. Euromoney (2016) estimates that the largest ten FX dealers collectively hold a 66% market share. So, for simplicity, we assume a representative tier-1 dealer sees 6.6% of market wide activity. Half of this will constitute buys and half sells, and similarly half will be risk increasing trades and half risk decreasing trades. So, in this example, the tier-1 dealer will on average trade $\frac{1}{2} \times \$111\text{mn/min} \times 6.6\% = 3.66\text{mn/min}$ of each type. Finally, we assume the dealer adopts exponential price skewing as in Example 2 and sets their risk limit R so that the absolute position is within a \$25mn corridor 95% of the time, i.e. $R = \$25\text{mn}/\Phi^{-1}(97.5\%) = \$12.75\text{mn}$. Collecting the relevant numbers and applying the result in Proposition 1 we obtain a top-down measurement of the average internalisation horizon of a tier-1 dealer in EURUSD as:

$$\frac{\$25\text{mn}/1.96}{\frac{1}{2} \times \$111\text{mn/min} \times 6.6\%} \times \frac{1.00}{\sqrt{2\pi}} = \frac{\$12.75\text{mn}}{3.66\$\text{mn/min}} \times \frac{1.00}{2.51} = 1.39 \text{ minutes.}$$

Table 1 reports the corresponding flow rates and internalisation horizons for a number of other currency pairs in the columns labelled "avg". There is a substantial amount of cross-sectional variation with typical internalisation horizons for the liquid G10 currencies in the single digit minutes, increasing to tens of minutes for most Asian and emerging market currencies as well as some less liquid G10 currencies. See also Figure 6.

Because of the pronounced diurnal activity patterns in the 24-hour FX markets, the internalisation horizons are expected to vary across the day in a predictable manner. To investigate this, we divide up the day into four contigu-

Figure 7: Diurnal volume profile for EURUSD



Note. This chart draws the (smoothed) per minute volume profile for EURUSD, rescaled to match the BIS (2016) implied ADV figures for the customer-facing electronic FX spot segment. The dashed lines and associated numbers report the average trading volume per minute (in $mn) for each of the four trading session. The volume profile is derived from EBS Markets trade data for 2016.

ous trading sessions, namely Asia-Pacific (APAC) from 22:00 – 07:00, London (LON) from 07:00 – 12:00, London – New York cross over (NYLON) from 12:00 – 17:00, and New York (NY) from 17:00 – 22:00, all in local London time. Using tick data from the primary inter-dealer platforms EBS Market and Reuters Matching we calculate the per minute volume profile by currency pair, normalise it, and multiply it by the volume figures implied from the BIS survey, and then compute averages within each trading session. Figure 7 displays the result for EURUSD and highlights the substantial variation of activity within the day: the flow rates during the NYLON session are nearly ten times larger than over the APAC session. From this it follows that average internalisation horizons for a representative tier-1 dealer will also vary substantially across the day. Table 1 reports the internalisation horizons broken down by trading session for a number of different currency pairs.

The results presented here assume the representative LP has a significant risk bearing capacity (i.e. large R) and receives a high volumes of trades (i.e. large λ_0) because of their substantial market share. Because the average internalisation horizon is proportional to R/λ_0, it will be longer for smaller dealers, in less active currencies, and over less active days or periods of the day. Equally, a dealer with reduced risk bearing capacity (i.e. small R) will require more aggressive position skewing to limit their risk exposure, and this naturally comes with shorter internalisation horizons. So the time a dealer takes to internalise a trade is partly determined by their size and partly by a choice of whether they adopt a passive or an aggressive internalisation strategy.

Table 1: Flow rates & internalisation horizons

	market	per minute trade flow rate (in \$mn)					average internalisation horizon (in mins)					
			by trading session						by trading session			
currency	share	avg	APAC	LON	NYLON	NY	min	avg	APAC	LON	NYLON	NY
Panel A: G10 currencies (73.3% market share)												
EURUSD	23.1%	111	28	148	267	67	1	1	5	1	1	2
JPY	19.3%	93	86	103	140	47	1	2	2	1	1	3
GBP	11.2%	54	13	84	119	32	1	3	12	2	1	5
CHF	4.4%	21	4	31	49	14	3	7	40	5	3	11
AUD	5.5%	26	23	32	40	13	4	6	7	5	4	12
CAD	4.6%	22	6	17	56	23	3	7	27	9	3	7
NZD	1.5%	7	6	9	12	4	13	21	26	18	13	40
SEK	2.0%	10	1	19	23	3	7	16	60^+	8	7	48
NOK	1.5%	7	1	14	17	3	9	21	60^+	11	9	60^+
avg		*73*	*36*	*93*	*150*	*42*	*2*	*4*	*19*	*3*	*2*	*8*
Panel B: Asian currencies (8.1% market share)												
CNH	3.8%	18	20	33	14	5	5	8	8	5	11	34
SGD	1.6%	8	7	11	10	3	14	20	21	14	16	59
HKD	1.5%	7	6	11	9	3	14	21	26	14	17	46
INR	1.1%	5	5	15	1	0	10	29	29	10	60^+	60^+
avg		*12*	*13*	*22*	*11*	*3*	*9*	*16*	*17*	*9*	*36*	60^+
Panel C: Emerging markets currencies (5.6% market share)												
MXN	1.8%	9	1	5	24	11	6	18	60^+	34	6	14
TRY	1.3%	6	1	13	14	2	11	25	60^+	12	11	60^+
RUB	1.1%	5	0	8	12	5	13	29	60^+	18	13	34
ZAR	0.8%	4	0	6	10	2	16	40	60^+	24	16	60^+
PLN	0.6%	3	0	6	7	1	21	53	60^+	28	21	60^+
avg		*6*	*1*	*8*	*15*	*5*	*12*	*29*	60^+	*24*	*12*	60^+

Note. This table reports by currency and category, the market share, the per minute trading volume in the customer-facing electronic FX spot segment, and the associated internalisation horizon for a tier-1 LP (i.e. 6.6% market share) derived from the exponential skew model (i.e. Eq. (11) with $R = 12.75$). For each currency listed, the results reflect the combined activity across the EUR and USD crosses, e.g. "NOK" consolidates EURNOK and USDNOK. The results are reported for daily averages ("avg") and by the trading sessions Asia-Pacific (APAC) from 22:00 – 07:00, London (LON) from 07:00 – 12:00, London – New York cross over (NYLON) from 12:00 – 17:00, and New York (NY) from 17:00 – 22:00, all in local London time. Internalisation horizons are truncated at 60 minutes.

2.3 Internalisation horizon by trade size

Throughout the paper we assume that the trader executes exclusively in unit size. The scenario where trade sizes are variable is interesting and relevant in practice but the mathematics quickly become intractable. To nevertheless obtain some basic insights, we conclude this section by taking a little detour and analyse how the internalisation horizon is expected to vary with trade size using a simple simulation. To that end, we modify Eq. (1) to

$$X_t = \sum_{i=1}^{N_t^+} q_i^+ - \sum_{j=1}^{N_t^-} q_j^- \tag{12}$$

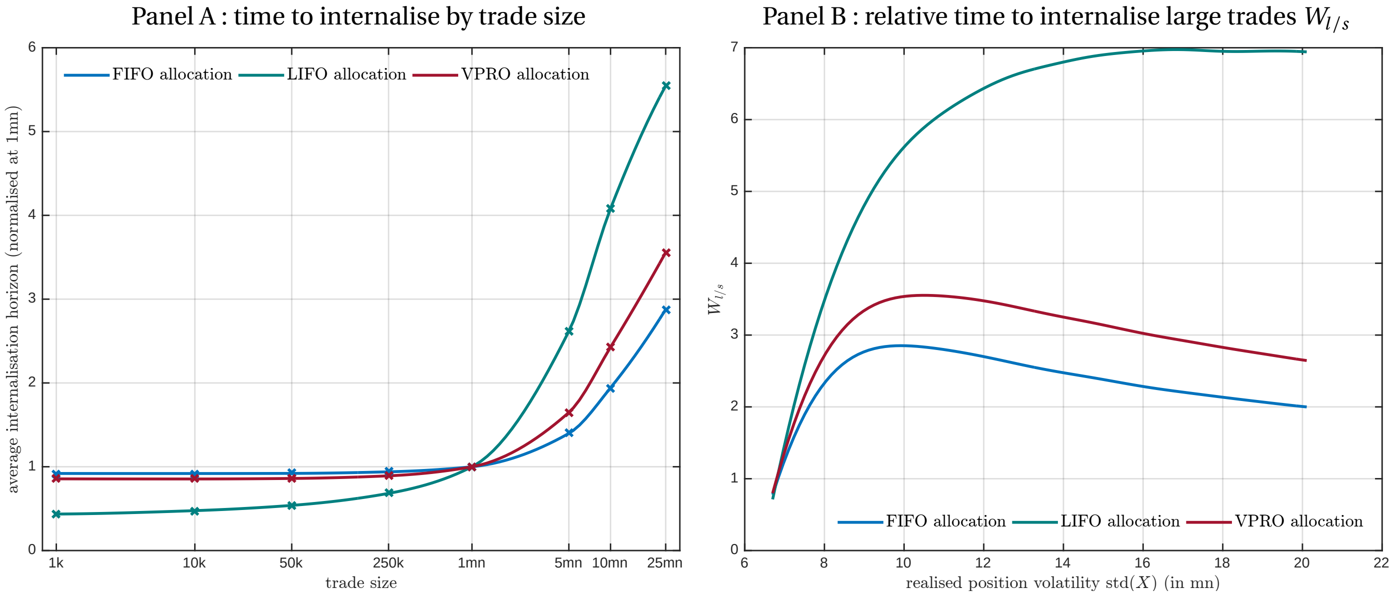
where $q^\pm$ now denotes the random trade size of buys and sells, and is assumed to be independent of all other other state variables (e.g. X, N, λ). Using this model, we simulate a long sequence of (ten million) trades q together with the corresponding LP position X and then calculate for each trade the internalisation horizon using the three allocation methods. We include the LIFO method for comparison purposes only, but in the discussion below we exclusively concentrate on the FIFO and VPRO results as they are of greater practical relevance. The baseline trade arrival rate is normalised to $\lambda_0 = 1$, the trade size distribution is assumed to be discrete with 1mn the most common "standard" size supplemented with relatively frequent smaller trades in the 1k-250k range and infrequent larger trades in the 5mn-25mn range.[8] The position-dependent buy and sell arrival rates $\lambda^\pm$ are determined via the exponential position skewing model as in Example 2 where R is calibrated to produce a realised LP position volatility of 10mn.[9]

Panel A of Figure 8 reports the average internalisation horizon grouped by trade size. We see that for small trades – and concentrating exclusively on FIFO and VPRO here – the internalisation horizon is only marginally shorter than that of a standard sized trade. Intuitively, for small trades, it is the LP's position (the "queue length") that is the primary driver of the time to internalise rather than the size of the trade. For trades that are large (in comparison to the typical magnitude of the LP position) this pattern changes and now the internalisation horizon increases rapidly with trade size. However, if we consider the ratio of the time it takes to internalise a large 25mn trade expressed as a ratio to that of a standard trade – denoted by $W_{l/s}$ for convenience – we see this is still only in the range of 3 – 4 rather than a value of 25 which one might naively expect. Of course, large trades tend to produce large risk positions, which in turn leads to the active solicitation of risk reducing trades via position skewing, and this speeds up the time to internalise.

[8]The precise distribution for q is specified as follows: trades of sizes 1k and 10k occur with a probability of 15% each, trades of sizes 50k and 250k occur with a probability of 20% each, and then trades of sizes 1mn, 5mn, 10mn, and 25mn occur with a probability 25%, 3%, 1.5%, and 0.5% respectively.

[9]With unit trade sizes, the position volatility is equal to R as per Eq. (7) in Example 2. With variable trade sizes, however, the distribution is not normal anymore and also there is no well defined relationship between the parameter R and the LP position volatility.

Figure 8: Illustration of internalisation horizon by trade size



Note. Based on a simulated sequence of trades with random sizes as described in the text, Panel A draws the average internalisation time grouped by trade size for the first-in-first-out (FIFO), last-in-first-out (LIFO), and pro-rated by volume (VPRO) allocation methodologies as illustrated in Figure 3. Here R is set such that the realised LP position volatility is 10mn. Panel B draws $W_{l/s}$ – i.e. the internalisation horizon of a large 25mn trade expressed as a ratio to that of a standard 1mn trade – across a range of realised LP position volatilities.

Panel B of Figure 8 shows that – for a fixed trade size distribution – the relationship between R and $W_{l/s}$ is non-monotonic and actually peaks in the range discussed above. The intuition for this pattern is as follows. For sufficiently small R, the position skewing that follows a large trade can be so aggressive that the time taken to internalise it approaches, or even drops below, that of a small trade. Equally, when R is big in comparison to the large trade sizes, the main determinant of the internalisation horizon is not the size of the trade but rather the magnitude of the LP's risk position. This latter point is formalised in the proposition below.[10] It demonstrates that as the risk bearing capacity of the LP grows, the internalisation horizon becomes independent of trade size and $W_{l/s} \to 1$.

Proposition 2 (Internalisation horizon with variable trade sizes) *Assume the random trade sizes q are strictly positive and small in comparison to R in the sense that $\tau_2/\tau_1 \ll R$ and $\max q \ll R$ where $\tau_k \equiv E(q^k)$. Further assume the trade sizes are independent of all other state variables in the model, and that the arrival rates of buy and sell orders*

[10]It can also be further illustrated with a queuing theory analogy. Consider a group of, say, 15 friends queuing together at airport security where they will be individually and sequentially served by a single custom's officer on a first-come-first-serve or FIFO basis. If the queue ahead of them is short (e.g. 2 people) then the waiting time for the group of friends to all pass security is primarily driven by the size of the group itself. However, if the queue ahead is large in comparison to the size of the group (e.g. 200 people) then the waiting time for them to all pass security is primarily driven by the length of the queue.

conditional on a position of $X_t = x$ is given by

$$\lambda_x^+ = \lambda_{-x}^- = \lambda_0 e^{-\frac{1}{2}\frac{\tau_2}{\tau_1} x/R^2}. \tag{13}$$

Then, the distribution of the LP's position is stationary and approximately Gaussian with mean zero and standard deviation R. The average internalisation horizon with FIFO or VPRO allocation is:

$$W \approx \frac{R/\tau_1}{\lambda_0}\frac{1}{\sqrt{2\pi}}, \tag{14}$$

and (approximately) independent of the trade size q.

Proof See Appendix A. ∎

3 Cost of internalisation

The market risk borne by the internalising dealer on their inventory, and the time required to internalise a trade, is increasing with R,[11] which in turn is determined by the position skewing strategy adopted. Naturally, with more aggressive skewing, the market risk and internalisation horizon diminishes. But what are the costs associated with this? Before we present the full model and rigorous mathematical results, we first build some intuition via an informal and approximate approach. It proceeds in three simple steps. First, we assume that LP–1 uses exponential position skewing as in Example 2, i.e. when they hold a position of $X_t = n$, they skew their prices to ensure that a risk reducing trade is received with probability

$$P_n \equiv \frac{\lambda_n^{-\mathrm{sign}(n)}}{\lambda_n^- + \lambda_n^+} \approx \frac{1}{2} + \frac{|n|}{4R^2}. \tag{15}$$

Second, we assume that LP–1 competes on best price with LP–2 for a noise trader's flow (i.e. the trader's buy and sell orders are driven by exogenous factors and appear random to the LPs). The "unskewed" price of LP–i is normally distributed around the true (unobserved) price, i.e. $p_t^{(i)} \sim \mathcal{N}(p_t^*, \kappa^2)$ with $\mathrm{corr}(p_t^{(1)}, p_t^{(2)}) = \rho$. When LP–1 holds a short position, i.e. $X_t = n < 0$, in order to attain the required risk reduction, they need to skew their price by an amount θ_t that satisfies $\Pr(p_t^{(1)} + \theta_t > p_t^{(2)}) = \Phi\big(\theta_t \kappa_-^{-1}\big) = P_n$ where $\kappa_-^2 = 2\kappa^2(1-\rho)$. By inversion we get the following approximation:

$$\theta_t = \kappa_- \Phi^{-1}\Big(\frac{1}{2} + \frac{|n|}{4R^2}\Big) \approx -\kappa_- \frac{\sqrt{2\pi}}{4R^2} n. \tag{16}$$

This is intuitive: for a long position $n > 0$ a negative price skew is required to make the LP more (less) likely to win the trader's buy (sell) orders. The desire for risk decreasing trades, and the magnitude of the price skew, grows with

[11] For instance, with exponential skewing, the dealer's position distribution is Gaussian. The average risk position is therefore $E(|X|) \propto R$ which after multiplying it with the asset's volatility gives a measure of market risk.

a decrease in the dealer's risk bearing capacity R. As a final step, we note that the LP sacrifices spread capture on the side where deals are risk reducing and that the rate at which those deals come in is elevated. The (gross) skew costs can thus be approximated as:

$$\mathbb{C}_S^- \propto \sum_n |\theta_n| e^{\frac{|n|}{2R^2}} \phi_n \approx \frac{\kappa_-}{2R} + \frac{\sqrt{2\pi}\kappa_-}{8R^2} e^{\frac{1}{8R^2}}. \tag{17}$$

At the same time, however, the LP will demand additional spread capture on risk increasing deals, i.e.

$$\mathbb{C}_S^+ \propto \sum_n |\theta_n| e^{-\frac{|n|}{2R^2}} \phi_n \approx \frac{\kappa_-}{2R} - \frac{\sqrt{2\pi}\kappa_-}{8R^2} e^{\frac{1}{8R^2}}. \tag{18}$$

The net costs of skewing is then:

$$\mathbb{C}_S = \mathbb{C}_S^- - \mathbb{C}_S^+ \approx \frac{\sqrt{2\pi}\kappa_-}{4R^2} e^{\frac{1}{8R^2}}. \tag{19}$$

This illustrates a few important points: (a) skew costs increase with a reduction in risk appetite R, or similarly, aggressive internalisation is more costly than passive internalisation, (b) the individual skew cost components $\mathbb{C}_S^-$ and $\mathbb{C}_S^+$ are proportional to R^{-1} and can be of substantial magnitude but because they are largely offsetting the net skew costs $\mathbb{C}_S$ tend to be much smaller and of order R^{-2}, (c) the higher the correlation between competing LPs' prices, the less skew is needed to win a deal on the desired side and the lower the skew costs. See Panel A of Figure 9 for an illustration of the skew cost components.

The informal approach taken here ignores a number of effects that further influence the cost of skewing. For instance, when the dealer skews their prices, we should not only consider how this affects the spread captured but also take into account the impact this has on the adverse selection the dealer is exposed to when competing for the trader's flow. Also, what if the trader is price sensitive and the flow the LP receives is a function of their pricing and therefore their position skew? What if the LP skews prices asymmetrically between the bid and the offer rates? The remainder of this section addresses these points via a more rigorous approach.

3.1 The model setup

Our starting point is the model introduced by Oomen (2017a) and we consider the simplest case where two LPs compete for a trader's order flow. There is an unobserved true (logarithmic) price process, p_t^*, which follows a random walk. The two LPs each make an independent assessment of where they set their reference (mid) price, i.e. LP–i sets $p_t^{(i)} = p_t^* + d_t^{(i)}$ where $d_t^{(i)} \sim$ i.i.d. $\mathcal{N}(0, \omega^2)$ for $i \in \{1, 2\}$ and $\text{corr}(d_t^{(1)}, d_t^{(2)}) = \rho_d$. Based on this, LP–$i$ then determines their bid- and offer-prices as $b_t^{(i)} = p_t^{(i)} - \frac{1}{2}s$ and $a_t^{(i)} = p_t^{(i)} + \frac{1}{2}s$ where s is a nominal spread which, for now, is assumed to be the same for both LPs. For simplicity we assume that there is no last look and so every trade request submitted by the trader is accepted by the LP at the bid or offer price depending on the direction of the trade request (see Oomen, 2017b, for a detailed study of trading with last look).

Position skewing is introduced via the process d, which tracks the deviation of the LP's reference price away from the true price. We decompose it into two distinct components (a) a measurement error m incurred when making inference on p^* and (b) a position skew θ. The measurement error is a random variable with $m_t^{(i)} \sim$ i.i.d. $\mathcal{N}(0, \kappa^2)$ and $\text{corr}(m_t^{(1)}, m_t^{(2)}) = \rho_m$. The position skew is a known function of the LP's position X. Note that the information set for each participant is different, i.e. LP–i observes $p_t^{(i)}$ and $\theta_t^{(i)}$ but does not observe $p_t^{(\neq i)}$ whereas the trader observes both $p_t^{(i)}$ and $p_t^{(\neq i)}$ but cannot isolate the skew component.

Different from Oomen (2017a) – where the trader is uninformed – here we consider one that is informed with trade and routing decision dependent on their own assessment of the true price process. Specifically, the trader sets their reference mid-price as $p_t^{(0)} = p_t^* + d_t^{(0)}$ where $d_t^{(0)} \sim$ i.i.d $\mathcal{N}(0, \omega_T^2)$ and $d^{(0)} \perp d_t^{(i)}$ for $i \in \{1, 2\}$. So when $\omega_T = 0$ they are fully informed and have exact knowledge of the true price, while they are a pure noise trader when $\omega_T = \infty$. The trader is assumed to demand liquidity at exogenously given time points t_j and determines trade direction and selects the LP by executing against the bid or offer price that is most favourable in comparison to the reference price, i.e. the lowest cost or highest revenue trade when marked-to-market against $p_{t_j}^{(0)}$. In particular, the trader sells to LP–i when:

$$
\begin{aligned}
b_{t_j}^{(i)} &> b_{t_j}^{(\neq i)} && \text{(i.e. LP–}i\text{ shows best bid),} && (20)\\
p_{t_j}^{(0)} - b_{t_j}^{(i)} &< \min(a_{t_j}^{(i)} - p_{t_j}^{(0)}, a_{t_j}^{(\neq i)} - p_{t_j}^{(0)}) && \text{(i.e. selling is more attractive than buying).} && (21)
\end{aligned}
$$

Analogously, the trader buys from LP–i when:

$$
\begin{aligned}
a_{t_j}^{(i)} &< a_{t_j}^{(\neq i)} && \text{(i.e. LP–}i\text{ shows best offer),} && (22)\\
a_{t_j}^{(i)} - p_{t_j}^{(0)} &< \min(p_{t_j}^{(0)} - b_{t_j}^{(i)}, p_{t_j}^{(0)} - b_{t_j}^{(\neq i)}) && \text{(i.e. buying is more attractive than selling).} && (23)
\end{aligned}
$$

It is instructive to distinguish between the nominal spread s that the individual LP charges, the observed spread that the trader sees after consolidating the competing LPs prices, the effective spread as a measure of actual transaction costs that the trader incurs, and the equilibrium spread that the LP would charge if their aim were to maximise overall revenues.

Proposition 3 (spread metrics) *Assume two identical LPs and absence of position skewing (i.e. $\theta = 0$ and $\kappa = \omega$). The expected observed spread is defined as $S = E(\min_i a_t^{(i)} - \max_i b_t^{(i)})$ and equal to:*

$$S = s - \kappa_- \sqrt{2/\pi}, \tag{24}$$

where $\kappa_\pm^2 = 2\kappa^2(1 \pm \rho_m)$. Let $\{x_{t_j}\}$ denote the set of prices at which the trader executes by crossing the LPs' spread and following the order placement rules given by Eqs. (20–23). The expected effective spread is defined as $\mathbb{S} = 2E(|x_{t_j} - p_{t_j}^*|)$

and equal to:

$$\mathbb{S} = S - \xi^{-1}\kappa_+^2\sqrt{2/\pi}, \tag{25}$$

where $\xi^2 = \kappa_+^2 + 4\omega_T^2$. Finally, the equilibrium spread at which neither LP can profit by unilaterally making a change to their own spread is equal to:

$$s^* = \frac{4\kappa^2 + \pi\xi\kappa_-}{\xi + \kappa_-}\sqrt{2/\pi}. \tag{26}$$

Proof See Appendix A. ■

Starting with a noise trader ($\omega_T = \infty$), we note that the observed spread is equal to the effective spread,[12] i.e. $\mathbb{S} = S$. While the trader is uninformed, they do benefit from the spread compression introduced by LP competition (measured as the difference between the nominal and observed spread) and so their actual transaction costs are given by the observed spread rather than the wider nominal spread. If in turn, the LPs set the nominal spread to the revenue maximising equilibrium spread, the trader's effective spread is equal to $\mathbb{S} = s^*(1-\pi^{-1})$ where $s^* = \sqrt{2\pi}\kappa_-$. The informed trader achieves lower execution costs by leveraging the information they holds regarding the true price to more effectively exploit any "mis-pricing" by the LPs. The effective spread now deviates from the observed spread and we have $\mathbb{S} < S < s$ provided $\rho < 1$. Additionally, it is easy to see that $\mathbb{S}_{s=s^*} > 0$, i.e. the effective spread the trader pays when the LPs charge a nominal spread s^* is always positive regardless of how informed they are and regardless of the magnitude of measurement error incurred by the LPs. Of course, for nominal spread values $s \neq s^*$ the effective spread paid by the informed trader can be negative. Panel B of Figure 9 provides an illustration of the different spread measures for varying degrees of trader informedness.

3.2 Position skewing

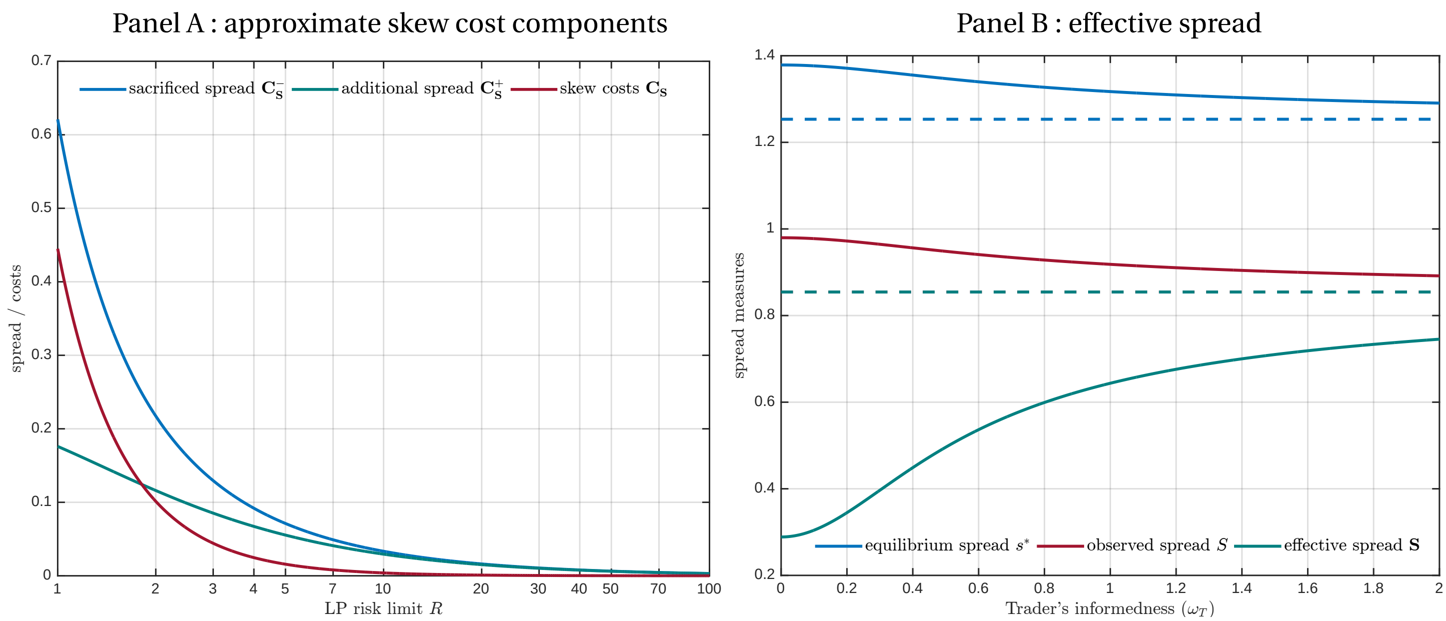
Let us consider the simplest possible position skewing rule, i.e. $\theta_t^{(i)} = -\gamma X_t^{(i)}$. It can be seen from Panel A of Figure 2 and also Eq. (16) that this strategy is a linear approximation to the exponential skewing model discussed in Example 2. When the LP's position is long, they skew their prices down to stimulate risk reducing trades by making it more attractive for the trader to buy at a lower offer and less attractive to sell at a lower bid. Also, the magnitude of the skew grows linearly in X and so the LP is more likely to hold a smaller risk position than a larger one. This skewing rule is symmetric in that the bid and offer prices are moved by the same amount in the same direction. We will also consider asymmetric skewing where the LP sets the bid and offer prices as follows:

$$b_t^{(i)} = m_t^{(i)} - \frac{1}{2}s - \gamma_{\text{in}}\min(X_t^{(i)},0) - \gamma_{\text{out}}\max(X_t^{(i)},0), \tag{27}$$

$$a_t^{(i)} = m_t^{(i)} + \frac{1}{2}s - \gamma_{\text{in}}\max(X_t^{(i)},0) - \gamma_{\text{out}}\min(X_t^{(i)},0), \tag{28}$$

[12]This relies on the assumed absence of trade rejects via the last look mechanism, see Oomen (2017b) for further discussion.

Figure 9: Skew cost components and effective spread



Note. Panel A draws the net skew costs $\mathbb{C}_S$ and its components $\mathbb{C}_S^-$ and $\mathbb{C}_S^+$ as a function of the LP's risk limit R and when faced with a noise trader, i.e. see Eqs. (17 – 19). Panel B draws the effective spread given in Proposition 3 as a function the trader's informedness ω_T. The dashed lines indicate the limiting spread values for $\omega_T \to \infty$. In both illustrations, $\omega = 0.5, \rho = 0.5$ and $s = s^*$.

with $\gamma_{\text{in}}, \gamma_{\text{out}} \geq 0$. The factor γ_{in} (γ_{out}) controls skewing on the side of the quote where – if traded on – the LP's risk position would reduce (increase further).

Proposition 4 (position distribution) *For the model with two LPs and an informed trader as in Section 3.1, with only LP–1 skewing according to Eqs. (27 – 28), and $m_t^{(1)} \perp \theta_t$, the stationary distribution of the positon of LP–1 is:*

$$\phi_n \xrightarrow[R]{} \frac{1}{\sqrt{2\pi R^2}} e^{-\frac{1}{2}n^2/R^2}, \tag{29}$$

where

$$R^2 = \frac{\sqrt{2\pi}}{\gamma_{\text{in}} + \gamma_{\text{out}}} \left(1 - \frac{1}{\pi} \arccos \frac{\kappa_1^2 - \kappa_2^2}{\xi \kappa_-}\right) \vartheta^{-1}, \tag{30}$$

and $\kappa_\pm^2 = \kappa_1^2 + \kappa_2^2 \pm 2\rho\kappa_1\kappa_2$, $\vartheta = \kappa_-^{-1} + \xi^{-1}$, and ξ as defined in Proposition 3. When $\kappa_1 = \kappa_2$, Eq. (30) simplifies to:

$$R^2 = \frac{\sqrt{\pi/2}}{\gamma_{\text{in}} + \gamma_{\text{out}}} \vartheta^{-1}. \tag{31}$$

Proof See Appendix A. ■

This result shows that a linear skewing rule produces a position that is asymptotically Gaussian. It also makes explicit how the skewing parameter γ relates to R: to attain a position volatility of R, the LP should set the symmetric skewing parameter $\gamma = \sqrt{\pi/8}\vartheta^{-1}R^{-2}$, or double that value if they only skew in or out. From Eqs. (30 – 31) it can be seen that $\partial R/\partial\omega_T > 0$ which suggests that a better informed trader is more responsive to the LP's position skewing.

To study the costs of position skewing, we compare the revenues of LP–1 with and without skewing while assuming that LP–2 does not control their position via skewing (i.e. $\theta_t^{(2)} = 0$), i.e.

$$\mathbb{C}_\gamma = \mathbb{R}_\gamma - \mathbb{R}_0, \tag{32}$$

where $\mathbb{R}_c$ denotes the revenue for LP–1 given a level of skewing $\gamma = c$ defined as:

$$\mathbb{R}_c = E(p^*_{t_j} - b^{(1)}_{t_j} | \gamma = c, 20 \,\&\, 21) \cdot \Pr(20 \,\&\, 21 | \gamma = c) + E(a^{(1)}_{t_j} - p^*_{t_j} | \gamma = c, 22 \,\&\, 23) \cdot \Pr(22 \,\&\, 23 | \gamma = c). \tag{33}$$

Note that $\mathbb{R}_0 = \mathbb{S}$.

Proposition 5 (internalisation metrics) *For the model with two LPs and an informed trader as in Section 3.1, with only LP–1 skewing according to Eqs. (27 – 28), and $m_t^{(1)} \perp \theta_t$, the costs of position skewing for LP–1 is:*

$$\mathbb{C}_\gamma = \vartheta(s^* - s)\frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{4\pi}R + \mathcal{O}_C\left(R^{-2}\right). \tag{34}$$

The market share of LP–1, defined as the probability of winning a trader's buy or sell order as per Eqs. (20 – 23), is:

$$\mathbb{M}_\gamma = \frac{1}{2} + \vartheta\frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{2\pi}R + \mathcal{O}_M\left(R^{-2}\right). \tag{35}$$

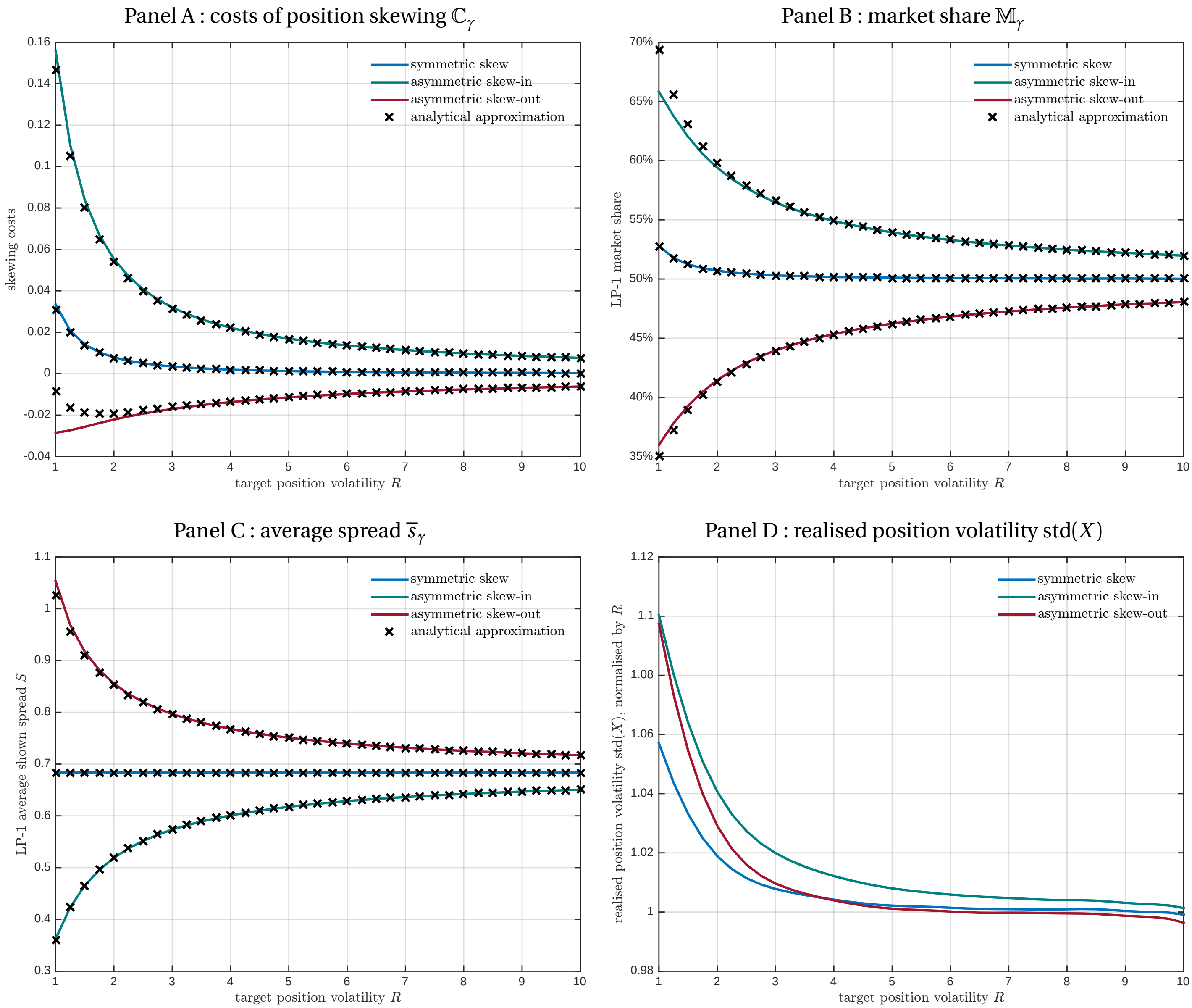
The average spread shown by LP–1, i.e. $E(a_t^{(1)} - b_t^{(1)})$, is:

$$\overline{s}_\gamma = s - \frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{\sqrt{\pi/2}}R + \mathcal{O}_{\overline{s}}\left(R^{-2}\right). \tag{36}$$

Proof See Appendix A. ■

From Eq. (31) we know that $\gamma \propto R^{-2}$ and so all the above quantities – skewing costs, market share away from 50%, and average shown spread away from nominal spread – are of order $1/R$ for asymmetric skewing (i.e. $\gamma_{\text{in}} \neq \gamma_{\text{out}}$) and of order $1/R^2$ for symmetric skewing (i.e. $\gamma_{\text{in}} = \gamma_{\text{out}}$). When $\gamma_{\text{in}} > \gamma_{\text{out}}$, the LP move their quote on the risk reducing side in more than they moves the quote on the risk increasing side out. This amounts to an effective tightening of the spread, hence $\overline{s}_\gamma < s$. In that scenario it is also intuitive that the market share of LP–1 exceeds that of LP–2, i.e. $\mathbb{M}_\gamma > \frac{1}{2}$. The cost of position skewing, however, varies not only by skewing policy but also by how the nominal spread compares to the equilibrium spread. In particular, position skewing is costly when the LP charges a nominal spread below (above) the equilibrium value and skews asymmetrically with $\gamma_{\text{in}} > \gamma_{\text{out}}$ ($\gamma_{\text{in}} < \gamma_{\text{out}}$) leading to an effective spread tightening (widening). In the reverse scenario, however, position skewing is actually revenue generating:

Figure 10: Illustration of internalisation metrics



Note. This chart draws the internalisation metrics listed in Proposition 5 plus the realised position volatility as a function of the target position volatility R. Three different linear position skewing methodologies are considered: symmetric with $\gamma_{\text{in}} = \gamma_{\text{out}}$, asymmetric skewing in only with $\gamma_{\text{out}} = 0$, and asymmetric skewing out only with $\gamma_{\text{in}} = 0$. The solid lines draw the simulated values. The crosses draw the second order analytical approximations given in Propositions 5 and 6. The model parameters are set as $\kappa = 0.5$, $\rho = 0.5$, $\omega_T = \frac{1}{2}\kappa$, and $s = \frac{1}{2}s^*$, i.e. a scenario with a highly informed trader and LPs that charge a tight nominal spread.

the effective spread widening or tightening associated with asymmetric skewing (partially) counteracts the revenue loss associated with charging a spread away from its equilibrium value.

To obtain explicit expressions for the metrics above in terms of the LP's risk appetite R, we can use Eq. (31) to obtain expressions for the skewing parameters in terms of R and then substitute those into Eqs. (34 – 36). For example, when the LP is only skewing in, then $\gamma_{\text{in}} = \sqrt{\pi/2}\vartheta^{-1}R^{-2}$ and $\gamma_{\text{out}} = 0$ from which it follows that the cost of position skewing is $(s^* - s)(4\sqrt{2\pi}R)^{-1}$the market share is $\frac{1}{2} + (2\sqrt{2\pi}R)^{-1}$, and the average spread shown is $s - \vartheta^{-1}R^{-1}$.

Before studying the internalisation metrics in more detail, we first present the second order terms in Proposition 5 which provide more accurate approximations across a wide range of values for R including smaller ones.

Proposition 6 (internalisation metrics) *The second order terms in Proposition 5 are:*

$$\mathscr{O}_C\left(R^{-2}\right) = \vartheta(s^* - s)\frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{4\pi}\Lambda + s\frac{\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{4\pi\xi\kappa_-}R^2 + \left(\vartheta - \frac{1}{4\kappa_-} - \frac{\kappa_+^2}{4\xi^3} - \frac{s}{\xi\kappa_-\sqrt{2\pi}}\right)\frac{\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2}{2\sqrt{2\pi}}R^2 + \mathscr{O}\left(R^{-3}\right), \quad (37)$$

$$\mathscr{O}_M\left(R^{-2}\right) = \vartheta\frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{2\pi}\Lambda + \frac{\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{2\pi\xi\kappa_-}R^2 + \mathscr{O}\left(R^{-3}\right), \quad (38)$$

$$\mathscr{O}_{\overline{s}}\left(R^{-2}\right) = -\frac{\gamma_{\text{in}} - \gamma_{\text{out}}}{\sqrt{\pi/2}}\Lambda + \mathscr{O}_{\overline{s}}\left(R^{-3}\right), \quad (39)$$

where

$$\Lambda = c_1\vartheta\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)R^2 + c_2\left(\gamma_{\text{out}}^2 - \gamma_{\text{in}}^2\right)\left(\frac{2}{\xi\kappa_-} - \vartheta^2\right)R^4,$$

and $c_1 = \frac{2-\pi}{2\pi}$, *and* $c_2 = \frac{3\pi-4}{3\pi^2}\sqrt{\pi/2}$.

Proof See Appendix A. ■

Panels A – C of Figure 10 illustrate the accuracy of the analytical approximations for the three internalisation metrics by comparing them to their true values obtained by simulation. Note that for small values of R we get (in this example) positive costs associated with symmetric skewing whereas the first order approximation gives a zero value. Similarly, with symmetric skewing the market share deviates from 50% for small values of R – a pattern only captured via the second order terms. Despite the symmetric skewing of the dealer, the trader is informed and more likely to trade on the side where the dealer is skewing in, effectively trading on a tighter spread thereby increasing the market share of LP–1. Panel D of Figure 10 shows that a linear skewing strategy that targets a position volatility of R on the basis of the "asymptotic large-R" Gaussian result in Proposition 4 is reasonably accurate even for small values of R, in that the targeted volatility closely aligns with the realised position volatility.

Table 2 provides a further illustration of the skew costs for varying levels of trader informedness, dealer's risk appetite R, and different nominal spread settings. It emphasises a number of important patterns. Consistent with Eq. (34), when the dealer charges the equilibrium spread, any skewing strategy is costly. When they charge less than the equilibrium spread, then only symmetric and asymmetric skewing-in is costly but skewing-out is revenue

Table 2: Illustration of position skewing costs

	Nominal spread $s = \frac{1}{2}$ equilibrium spread s^*				Nominal spread $s =$ equilibrium spread s^*			
Trader informedness	perfect ($\omega_T = 0$)	high ($\omega_T = \kappa$)	medium ($\omega_T = 2\kappa$)	none ($\omega_T = \infty$)	perfect ($\omega_T = 0$)	high ($\omega_T = \kappa$)	medium ($\omega_T = 2\kappa$)	none ($\omega_T = \infty$)
Panel A: spread metrics in absence of position skewing ($\times 100$)								
Nominal half-spread ($\frac{1}{2}s$)	34.5	33.7	32.9	31.3	68.9	67.4	65.9	62.7
Observed half-spread ($\frac{1}{2}S$)	14.5	13.7	13.0	11.4	49.0	47.4	45.9	42.7
Effective half-spread ($\frac{1}{2}\mathbb{S}$)	−20.0	−8.9	−0.7	11.4	14.4	24.8	32.2	42.7
Panel B: cost of symmetric position skewing ($\times 100$)								
Low risk ($R = 1$)	3.0	3.9	4.7	6.0	2.0	3.1	4.1	6.1
Medium risk ($R = 5$)	0.1	0.1	0.2	0.2	0.1	0.1	0.2	0.3
High risk ($R = 10$)	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1
Panel C: cost of asymmetric-in position skewing ($\times 100$)								
Low risk ($R = 1$)	14.6	17.4	20.4	34.8	9.1	12.1	15.2	30.0
Medium risk ($R = 5$)	1.7	1.7	1.8	1.8	0.3	0.4	0.5	0.6
High risk ($R = 10$)	0.8	0.8	0.7	0.8	0.1	0.1	0.1	0.1
Panel D: cost of asymmetric-out position skewing ($\times 100$)								
Low risk ($R = 1$)	−3.3	−2.2	−1.2	0.4	1.5	2.6	3.5	5.0
Medium risk ($R = 5$)	−1.2	−1.1	−1.0	−0.8	0.1	0.2	0.3	0.4
High risk ($R = 10$)	−0.6	−0.6	−0.6	−0.5	0.0	0.1	0.1	0.1

Note. Panel A reports the spread metrics from Proposition 3 for different levels of trader informedness (ω_T, in columns) and nominal spread settings ($s = \frac{1}{2}s^*$ in left panel and $s = s^*$ in right panel). Panels B – D report the costs of symmetric ($\gamma_{\text{in}} = \gamma_{\text{out}}$), asymmetric-in ($\gamma_{\text{in}} > 0, \gamma_{\text{out}} = 0$), and asymmetric-out ($\gamma_{\text{in}} = 0, \gamma_{\text{out}} > 0$) position skewing targeting a low, medium, and high level of position volatility of R. The model parameters are set as $\kappa = 0.5$, $\rho = 0.5$. For ease of readability, all numbers in the table are multiplied by one hundred. Note that if the spread is expressed in basis points, then the numbers in the table are expressed as "per million" figures.

positive. As trader informedness grows, the skewing costs diminish, i.e. $\partial\mathbb{C}/\partial\omega_T > 0$. In our model, a better informed trader is more price sensitive and the dealer therefore requires a smaller skew to facilitate a risk reducing trade. An alternative interpretation is one where the informed trader becomes an opportunistic liquidity provider to the dealer that is willing to pay for risk reducing flow. An informed trader is insensitive to this because their trade direction is unaffected by the position skew, whereas the informed trader is more likely to trade on the side that is risk reducing for the dealer. Finally, the costs of position skewing can be material when the LP is required to keep the position in a tight range. As a proportion of the revenues that could be obtained in the absence of position skewing – the effective (half) spread $\mathbb{S}$ – the costs of skewing when targeting a small R range from double digit percentages to multiples of the effective spread depending on the skewing strategy adopted and the informedness of the trader. For instance, when the trader is moderately informed ($\omega_T = 2\kappa$), and the dealer charges an equilibrium (half) spread of 65.9 (in appropriate units, e.g. per million of base currency notional), the inside spread observed by the trader after consolidating the competing LPs prices is 47.4. The effective spread that can be earned by the dealer is still less at 32.2 due to the informed trading. Now, if the dealer also has limited risk bearing capacity and targets a position volatility of $R = 1$, then the associated skewing costs will further reduce their revenues by 3.5 – 4.1 (or 11% – 13% of the effective half spread) for symmetric or asymmetric-out skewing and by 15.2 (or 47% of the effective half spread) for asymmetric-in skewing. Even stronger effects are observed when the dealer charges a nominal spread that is less than the equilibrium spread, e.g. when the trader is uninformed the effective half-spread available to the dealer is 11.4 but skewing costs in the examples considered can rise to as high as 34.8 thereby rendering the trading relationship economically unviable.

3.3 Further discussion

The simplicity of the model analysed here allows us to retain analytic tractability and build a basic understanding of the mechanics and trade-offs involved. However, it necessarily omits a number aspects one may want to consider. For instance, we assume that the measurement error incurred by the trader is independent of the dealers' (i.e. $d^{(0)} \perp d^{(i)}$) but in practice one would expect these to be correlated because the trader is likely to form a view of the efficient price process – at least in part – on the basis of the prices provided by the dealers. Similarly, we assume the dealer's position (skew) is independent of the error they make in the mid-price construction (i.e. $\theta \perp m^{(1)}$). The points at which the trader demands liquidity are exogenously given but the flow rate – particularly of informed and opportunistic traders – naturally depends on the prices shown by the dealers: with tighter spreads, more signals become economically viable for the trader to act upon thereby increasing the flow rate. Another important and restrictive assumption is that there are only two LPs, that charge the same nominal spread, and that LP–2 is risk neutral. In practice, traders routinely place more than two LPs in competition for their flow, and the LPs' pricing and hedging strategies are not determined in isolation but influenced by the actions of their competitor LPs (e.g. a spread tightening by one LP may lead another LP to also tighten in order to remain competitive or retain market share). A general equilibrium approach would be required to study such interactions. From a queuing theory perspective, further analysis of internalisation horizon of non-unit trade sizes would be of interest, as well as the internalisation horizon conditional on the dealer's risk position.

Perhaps the most important aspect that has not been modelled is the signalling risk or market impact associated with position skewing. Price discovery in the electronic FX market operates on milli-second timescales whereas the risk held by dealers often spans several minutes, if not longer. Consequently, the (random) oscillations in the measurement error m are likely to be orders of magnitude quicker than that of the position skew component θ. This in turn, may enable a trader to statistically identify the LP's position skew. They may even attempt to infer the LP's precise risk position or imminent hedging demand. This is one channel through which position skewing can lead to information leakage, and a trader's actions can cause market impact that is adverse to the economic interests of the LP. Oomen (2017a) emphasises the importance of distinguishing between LPs that externalise and those that internalise. Externalisers tend to cause (more) market impact as part of their liquidity provision and this can adversely affect the execution costs of the trader. What we argue here, is that additionally it is important to distinguish between passive and aggressive internalisers because the information leakage associated with position skewing can also impact on a trader's execution costs. Instead of polarising LPs between externalisers and internalisers or those that create market impact and those that do not, one should view the LPs across a continuum ranging from passive internalisers to aggressive internalisers to externalisers and their varying degrees of associated market impact. A formal analysis of this – and more – is left for future research.

4 Concluding remarks

In the world's largest financial market of foreign exchange, where the majority of spot transactions are conducted electronically at a pace approaching the speed of light, the popular perception is that dealers tend to hold risk positions for only a matter of seconds. Using standard queuing theory in conjunction with the authoritative BIS (2016) triennial market survey, this paper shows that internalising dealers hold risk for minutes, if not hours. This is an important observation, not least because it has implications for how a customer should best interact with the dealers and it provides a more granular perspective on the access to liquidity in this over-the-counter market. We study the mechanics of internalisation, and show that position skewing for dealers with limited risk bearing capacity can be very costly.

The key message of the paper is that a trader's speed of execution should be aligned with the internalisation horizon of the dealer. Any sustained and systematic mismatch between the two either leads to unnecessary market risk borne by the trader when they execute too slowly, or increased transaction costs due to the LP needing to risk manage the excessive trader's flow via aggressive position skewing or even externalisation. This highlights a basic but fundamental trade-off between the speed at which a trader can access liquidity and the transaction costs they incur. Patient traders that aim to minimise transaction costs and market impact should therefore focus on passive internaliser LPs whereas those that require quicker liquidity access are more naturally served by aggressive internaliser or externaliser LPs.

A Proofs

Proof of Proposition 1. The expression in Eq. (10) follows directly from the specification of the binary position skewing in Example 1 and Eqs. (8) and (9). To obtain the expression in Eq. (11) we note that L in Eq. (8) can be viewed as a Riemann sum and can therefore be approximated by the corresponding integral $2\int_0^\infty n\phi_n dn = R\sqrt{2/\pi}$. Similarly, λ in Eq. (8) can be approximated by $\phi_0 + 2\int_0^\infty \phi_n e^{-n/(2R^2)} dn$. ■

Proof of Proposition 2. For finite R, let X_R be distributed according to the stationary distribution of the LP. For the first statement, we have to show that X_R/R converges in distribution to a standard normal random variable as $R \to \infty$. The existence of such a limit is obtained by a compactness argument and we focus on verifying that the limit is actually standard normal. To this end, let f_R and φ_R be the characteristic functions of X_R and X_R/R respectively,

$$\begin{aligned} f_R(k) &= \mathbb{E}\left[e^{-ikX_R}\right] \\ \varphi_R(k) &= \mathbb{E}\left[e^{-ikX_R/R}\right] = f_R(k/R). \end{aligned}$$

As the distribution of X_R can be bounded from above by a Gaussian, f_R and φ_R are actually defined on the entire complex plane and there is a limit

$$\varphi_R \to \varphi, \tag{40}$$

uniformly on compact subsets (and similar for all derivatives of φ_R). All that is left to do is to identify φ. Let ψ be the characteristic function of the trade size q, so, in particular,

$$\psi(0) = 1, \qquad \psi'(0) = -i\tau_1, \qquad \psi''(0) = -\tau_2. \tag{41}$$

If μ_R is the distribution of X_R, its stationarity can be written as

$$\left(A^+\lambda^+ + A^-\lambda^- - \lambda^+ - \lambda^-\right)\mu_R = 0, \tag{42}$$

where the operators $A^\pm$ describe the right/left jump with a jump size distributed like q. Expressing (42) in characteristic functions,

$$f_R\left(p - i\frac{\tau_2}{2\tau_1 R^2}\right)\left(\psi(p) - 1\right) + f_R\left(p + i\frac{\tau_2}{2\tau_1 R^2}\right)\left(\psi(-p) - 1\right) = 0, \tag{43}$$

for any $p \in \mathbb{R}$. Now fix $k \in \mathbb{R}$, multiply (43) by R^2 and plug in $p = k/R$,

$$R^2\varphi_R\left(k - i\frac{\tau_2}{2\tau_1 R}\right)\left(\psi(k/R) - 1\right) + R^2\varphi_R\left(k + i\frac{\tau_2}{2\tau_1 R}\right)\left(\psi(-k/R) - 1\right) = 0. \tag{44}$$

Taking the $R \to \infty$ limit using (40) and (41),

$$\begin{aligned} 0 &= -i\varphi'(k)\psi'(0)\frac{\tau_2}{\tau_1}k + \varphi(k)\psi''(0)k^2 \qquad (45) \\ &= -\tau_2\left(k\varphi'(k) + k^2\varphi(k)\right). \qquad (46) \end{aligned}$$

So

$$\varphi'(k) = -k\varphi(k), \tag{47}$$

$$\varphi(k) = e^{-k^2/2}, \tag{48}$$

and the limit distribution of X_R/R is indeed standard normal.

We now have to show that in both FIFO and VPRO case, the conditional waiting time T for a risk-increasing trade (say, sell to the LP) of size h to be risk-reduced is approximately

$$E[T_{\text{FIFO}}|q=h] = E[T_{\text{VPRO}}|q=h] = \frac{R/\tau_1}{\lambda_0}\frac{1}{\sqrt{2\pi}}. \tag{49}$$

From Little's law we have the unconditional expectation

$$E[T_{\text{FIFO}}] = E[T_{\text{VPRO}}] = \frac{R/\tau_1}{\lambda_0}\frac{1}{\sqrt{2\pi}}, \tag{50}$$

and so all we have to check is that the first two terms of Eq. (49) are independent of h in the large R limit. Assume that the trade arrives when the LP holds a position $X_0 > -h$. To verify independence of h for FIFO, it is enough to see that the conditional waiting time T is defined as the time when we have received a volume of $X_0 + h$ in off-setting trades. Note that X_0 is typically $O(R)$, so $h \leq \max(q) \ll X_0$ and thus $X_0 + h$ and X_0 are asymptotically equal as R goes to infinity, and we can drop the h–dependence.

In case of VPRO, each risk-reducing trade of size q_t^- is split over all trades not completely off-set yet, with the portion reducing the trade in consideration being hq_t/X_t. The conditional waiting time is given as the smallest time T such that

$$T_{\text{VPRO}} = \arg\min_T \sum_{t\leq T} h\frac{q_t^-}{X_t} \geq h. \tag{51}$$

Cancelling h and noting that the paths of X_t, q_t are approximately independent of h (by same argument as for FIFO), again removes the h–dependence. ■

Proof of Proposition 3. Here and in all proofs below, we drop the time subscript t. The expression for the observed spread in Eq. (24) is obtained by setting $N=2$ in Proposition 1 of Oomen (2017a). The expression for the effective spread paid by the trader in Eq. (25) is obtained as:

$$\begin{aligned}
S &= s + E\left(d^{(1)}|d^{(1)} < d^{(2)}, d^{(1)} + d^{(2)} < 2d^{(0)}\right) - E\left(d^{(1)}|d^{(1)} > d^{(2)}, d^{(1)} + d^{(2)} > 2d^{(0)}\right),\\
&= s + 2E\left(d^{(1)}|d^{(1)} < d^{(2)}, d^{(1)} + d^{(2)} < 2d^{(0)}\right),\\
&= s + E\left(Y + Z \mid Y < 2d^{(0)}, Z < 0\right),\\
&= s + E(Y|Y < 2d^{(0)}) - E(|Z|),\\
&= s - \sqrt{2/\pi}\left(\overline{\omega}_- + \frac{\overline{\omega}_+^2}{\sqrt{\overline{\omega}_+^2 + 4\omega_T^2}}\right).
\end{aligned}$$

We use the transformation of variables $Y = d^{(1)} + d^{(2)}$ and $Z = d^{(1)} - d^{(2)}$ and note that because the variance of the true price deviations $d^{(i)}$ are the same for both LPs, Y and Z are independent and normally distributed with variances $\overline{\omega}_+^2 \equiv V(Y) = 2\omega^2(1+\rho)$ and $\overline{\omega}_-^2 \equiv V(Z) = 2\omega^2(1-\rho)$.

As for the proof of Eq. (26), while we eventually want to identify the equilibrium spread s^* that both LPs will use, for now we need to accommodate for unilateral spread changes and define $s^{(1)}$ and $s^{(2)}$ to be the spreads of the individual LPs. Taking into account the decision making of the trader, the expected revenues of LP–1 equal

$$\begin{aligned}
\mathscr{R}^{(1)} &= \mathbb{E}\Bigg[\left(m^{(1)}+\frac{s^{(1)}}{2}\right)\mathbb{1}\left(m^{(1)}-m^{(2)}<\frac{s^{(2)}-s^{(1)}}{2}, m^{(1)}-d^{(0)}<0, m^{(1)}+m^{(2)}-2d^{(0)}<\frac{s^{(2)}-s^{(1)}}{2}\right)\\
&\qquad -\left(m^{(1)}-\frac{s^{(1)}}{2}\right)\mathbb{1}\left(m^{(1)}-m^{(2)}>\frac{s^{(1)}-s^{(2)}}{2}, m^{(1)}-d^{(0)}>0, m^{(1)}+m^{(2)}-2d^{(0)}>\frac{s^{(1)}-s^{(2)}}{2}\right)\Bigg]\\
&= \mathbb{E}\Bigg[\left(m^{(1)}+\frac{s^{(1)}}{2}\right)\mathbb{1}\left(m^{(1)}-m^{(2)}<\frac{s^{(2)}-s^{(1)}}{2}, m^{(1)}+m^{(2)}-2d^{(0)}<\frac{s^{(2)}-s^{(1)}}{2}\right)\\
&\qquad -\left(m^{(1)}-\frac{s^{(1)}}{2}\right)\mathbb{1}\left(m^{(1)}-m^{(2)}>\frac{s^{(1)}-s^{(2)}}{2}, m^{(1)}+m^{(2)}-2d^{(0)}>\frac{s^{(1)}-s^{(2)}}{2}\right)\Bigg]+\mathscr{O}\left(\left(s^{(1)}-s^{(2)}\right)^2\right)\\
&= -\frac{\kappa_-}{\sqrt{2\pi}}\Phi\left(\frac{s^{(2)}-s^{(1)}}{2\sqrt{\kappa_+^2+4\omega_T^2}}\right)-\frac{\kappa_+^2}{\sqrt{2\pi}\sqrt{\kappa_+^2+4\omega_T^2}}\Phi\left(\frac{s^{(2)}-s^{(1)}}{2\kappa_-}\right)+s^{(1)}\Phi\left(\frac{s^{(2)}-s^{(1)}}{2\sqrt{\kappa_+^2+4\omega_T^2}}\right)\Phi\left(\frac{s^{(2)}-s^{(1)}}{2\kappa_-}\right)+\mathscr{O}\left(\left(s^{(1)}-s^{(2)}\right)^2\right).
\end{aligned}$$

Thus, if LP–1 changes their spread away from an original equal setting of $s^{(1)}=s^{(2)}=s$, the revenues change by

$$\left.\frac{\partial}{\partial s^{(1)}}\mathscr{R}^{(1)}\right|_{s^{(1)}=s^{(2)}=s}=\frac{\kappa_-}{4\pi\sqrt{\kappa_+^2+4\omega_T^2}}+\frac{\kappa_+^2}{4\pi\sqrt{\kappa_+^2+4\omega_T^2}\kappa_-}+\frac{1}{4}-\frac{s}{4\sqrt{2\pi}}\left(\frac{1}{\kappa_-}+\frac{1}{\sqrt{\kappa_+^2+4\omega_T^2}}\right).$$

By setting

$$\left.\frac{\partial}{\partial s^{(1)}}\mathscr{R}^{(1)}\right|_{s^{(1)}=s^{(2)}=s}=0,$$

one obtains

$$s^*=\sqrt{\frac{2}{\pi}}\left(\frac{4\kappa^2+\pi\kappa_-\sqrt{\kappa_+^2+4\omega_T^2}}{\sqrt{\kappa_+^2+4\omega_T^2}+\kappa_-}\right).$$

■

Proof of Proposition 4. Our first goal is to approximate the probability $p_-(\theta_a,\theta_b)$ of the trader buying from LP–1 by a linear function of the bid and ask skew θ_b, θ_a in the vicinity of $\theta_a=\theta_b=0$. We have

$$\begin{aligned}
p_-(\theta_a,\theta_b) &= \mathbb{P}\left(m^{(1)}+\theta_a-m^{(2)}<0, m^{(1)}+\theta_a+m^{(2)}<2d^{(0)}, 2m^{(1)}+\theta_a+\theta_b<2d^{(0)}\right)\\
&= \mathbb{P}\left(m^{(1)}+\theta_a-m^{(2)}<0, m^{(1)}+\theta_a+m^{(2)}<2d^{(0)}\right)+\mathscr{O}\left((\theta_a-\theta_b)^2\right)\\
&=: p_-(\theta_a)+\mathscr{O}\left((\theta_a-\theta_b)^2\right),
\end{aligned}$$

and can ignore the quadratic term for our linear approximation. Observe that for $\kappa_1\neq\kappa_2$ the random variables

$$\begin{aligned}
m^{(1)}+m^{(2)} &= Y,\\
m^{(1)}-m^{(2)} &= Z
\end{aligned}$$

are not independent. By a simple geometric argument, for $\theta_a=0$,

$$\begin{aligned}
p_-(0) &= \mathbb{P}\left(Y<2d^{(0)}, Z<0\right)=\frac{\pi-\phi}{2\pi}, \text{ with}\\
\phi &= \arccos\left(\operatorname{corr}\left(Y-2d^{(0)}, Z\right)\right),\\
\operatorname{corr}\left(Y-2d^{(0)}, Z\right) &= \frac{\kappa_1^2-\kappa_2^2}{\sqrt{(\kappa_1^2+\kappa_2^2+2\rho\kappa_1\kappa_2+4\omega_T^2)(\kappa_1^2+\kappa_2^2-2\rho\kappa_1\kappa_2)}},
\end{aligned}$$

while

$$\left.\frac{\partial}{\partial\theta}p_-(\theta)\right|_{\theta=0} = -\int_{-\infty}^{\infty} \mathrm{d}y\, f(y,0)\Phi\left(\frac{-y}{2\omega_T}\right) - \int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{0} \mathrm{d}z\, f(y,z)\frac{1}{\sqrt{8\pi\omega_T^2}}\, e^{-y^2/(8\omega_T^2)}, \tag{52}$$

with f the joint density of Y and Z,

$$\begin{aligned}
f(y,z) &= \frac{1}{2\pi\sqrt{\det(\Sigma)}}\exp\left(-\frac{1}{2}(y,z)\Sigma^{-1}(y,z)^T\right),\\
\Sigma &= \begin{pmatrix} \kappa_1^2+\kappa_2^2+2\rho\kappa_1\kappa_2 & \kappa_1^2-\kappa_2^2 \\ \kappa_1^2-\kappa_2^2 & \kappa_1^2+\kappa_2^2-2\rho\kappa_1\kappa_2 \end{pmatrix},\\
\det\Sigma &= 4(1-\rho^2)\kappa_1^2\kappa_2^2.
\end{aligned}$$

For the first line in (52)

$$-\int_{-\infty}^{\infty} \mathrm{d}y\, f(y,0)\Phi\left(\frac{-y}{2\omega_T}\right) = -\frac{1}{\sqrt{8\pi\left(\kappa_1^2+\kappa_2^2-2\rho\kappa_1\kappa_2\right)}}, \tag{53}$$

while

$$\begin{aligned}
&-\int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{0} \mathrm{d}z\, f(y,z)\frac{1}{\sqrt{8\pi\omega_T^2}}\, e^{-y^2/(8\omega_T^2)}\\
&= -\frac{1}{\sqrt{32\pi\omega_T^2}}\int_{-\infty}^{\infty} \mathrm{d}y \int_{-\infty}^{\infty} \mathrm{d}z\, f(y,z)e^{-y^2/(8\omega_T^2)}\\
&= -\left(8\pi\left(\kappa_1^2+\kappa_2^2+2\rho\kappa_1\kappa_2+4\omega_T^2\right)\right)^{-1/2},
\end{aligned}$$

so

$$\left.\frac{\partial}{\partial\theta}p_-(\theta)\right|_{\theta=0} = -\left(8\pi\left(\kappa_1^2+\kappa_2^2-2\rho\kappa_1\kappa_2\right)\right)^{-1/2} - \left(8\pi\left(\kappa_1^2+\kappa_2^2+2\rho\kappa_1\kappa_2+4\omega_T^2\right)\right)^{-1/2}. \tag{54}$$

By symmetry, the probability of the trader selling to LP–1 fulfills

$$p_+(\theta_a,\theta_b) = p_+(\theta_b) + \mathcal{O}\left((\theta_a-\theta_b)^2\right)$$

with

$$\begin{aligned}
p_+(0) &= p_-(0),\\
\left.\frac{\partial}{\partial\theta}p_+(\theta)\right|_{\theta=0} &= -\left.\frac{\partial}{\partial\theta}p_-(\theta)\right|_{\theta=0}.
\end{aligned}$$

For sufficiently small γ_{out} and γ_{in}, one has

$$\sqrt{\gamma_{\text{out,in}}} \ll \sqrt{p_+(0)\left(\left.\frac{\partial}{\partial\theta}p_+(\theta)\right|_{\theta=0}\right)^{-1}},$$

and also

$$\sqrt{\gamma_{\text{out,in}}} \ll \sqrt{\left(\left.\frac{\partial}{\partial\theta}p_+(\theta)\right|_{\theta=0}\right)\cdot\max_\theta\left(\frac{\partial^2}{\partial\theta^2}p_+(\theta)\right)^{-1}}.$$

Thus, the approximate stationary distribution of the position of LP–1 fulfills for, say, $n > 0$

$$\phi_n \left(p_+(0) + \gamma_{\text{in}} p_+(0) n\right) \approx \phi_{n-1} \left(p_+(0) - \gamma_{\text{out}} p_+(0)(n-1)\right)$$

or

$$\phi_n \approx \phi_0 e^{-n^2/(2R^2)}, \text{ with}$$
$$R^2 = \frac{p_+(0)}{\gamma_{\text{in}} + \gamma_{\text{out}}} \left(\frac{\partial}{\partial \theta} p_+(\theta) \bigg|_{\theta=0} \right)^{-1},$$

which proves (30). ■

Proof of Propositions 5 and 6. In the absence of skewing from LP–2, the trader buys from LP–1 if

$$m^{(1)} - m^{(2)} < -\theta_a^{(1)}, \tag{55}$$
$$2m^{(1)} - 2d^{(0)} < -\theta_a^{(1)} - \theta_b^{(1)}, \tag{56}$$
$$m^{(1)} + m^{(2)} - 2d^{(0)} < -\theta_a^{(1)}, \tag{57}$$

and sells to LP–1 if

$$m^{(1)} - m^{(2)} > -\theta_b^{(1)}, \tag{58}$$
$$2m^{(1)} - 2d^{(0)} > -\theta_a^{(1)} - \theta_b^{(1)}, \tag{59}$$
$$m^{(1)} + m^{(2)} - 2d^{(0)} > -\theta_b^{(1)}. \tag{60}$$

For each trade the trader makes, LP–1 realizes an average profit of

$$\mathbb{E}\left[\left(m^{(1)} + s/2 + \theta_a^{(1)}\right) \mathbb{1}\,((55)\&(56)\&(57)) - \left(m^{(1)} - s/2 + \theta_b^{(1)}\right) \mathbb{1}\,((58)\&(59)\&(60))\right] \tag{61}$$
$$= \frac{s}{4} - \frac{\kappa_-}{2\sqrt{2\pi}} - \frac{\kappa_+^2}{2\sqrt{2\pi}\sqrt{\kappa_+^2 + 4\omega_T^2}} \tag{62}$$
$$+ \left(\frac{s}{4\sqrt{2\pi}} \left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}} \right) - \frac{\kappa^2}{\pi\kappa_- \sqrt{\kappa_+^2 + 4\omega_T^2}} - \frac{1}{4} \right) \cdot \mathbb{E}\left[\theta_b^{(1)} - \theta_a^{(1)}\right] \tag{63}$$
$$+ \left[\frac{1}{8\sqrt{2\pi}} \left(\frac{1}{\kappa_-} + \frac{\kappa_+^2}{(\kappa_+^2 + 4\omega_T^2)^{3/2}} \right) + \frac{s}{4\pi\kappa_- \sqrt{\kappa_+^2 + 4\omega_T^2}} - \frac{1}{2\sqrt{2\pi}} \left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}} \right) \right] \cdot \mathbb{E}\left[\left(\theta_a^{(1)}\right)^2 + \left(\theta_b^{(1)}\right)^2\right] \tag{64}$$
$$- \frac{s}{4\pi\kappa_- \sqrt{\kappa_+^2 + 4\omega_T^2}} \mathbb{E}\left[\max\left(\theta_b^{(1)} - \theta_a^{(1)}, 0\right)^2\right] \tag{65}$$
$$+ \mathcal{O}\left(\frac{1}{\min(\kappa_\pm^2)} \mathbb{E}\left[\left(\theta_b^{(1)}\right)^3\right] \right). \tag{66}$$

For a fixed position $X^{(1)} = x \geq 0$ of LP–1, the probability that the trader sells to LP–1 equals

$$\lambda_+(x) = \frac{1}{4} - \frac{1}{2\sqrt{2\pi}} \left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}} \right) \gamma_{\text{out}}\, x + \frac{2\gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{4\pi\kappa_- \sqrt{\kappa_+^2 + 4\omega_T^2}} x^2 + \mathcal{O}\left(\frac{\gamma_{\text{in/out}}^3 x^3}{\min(\kappa_\pm^3)} \right), \tag{67}$$

and the probability that the trader buys from LP–1 is

$$\lambda_-(x) = \frac{1}{4} + \frac{1}{2\sqrt{2\pi}}\left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}}\right)\gamma_{\text{in}} x + \frac{2\gamma_{\text{in}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{4\pi\kappa_-\sqrt{\kappa_+^2 + 4\omega_T^2}} x^2 + \mathcal{O}\left(\frac{\gamma_{\text{in/out}}^3 x^3}{\min(\kappa_\pm^3)}\right). \tag{68}$$

For $x < 0$, the probabilities can then be found by symmetry, $\lambda_\pm(x) = \lambda_\mp(-x)$. Renormalizing to have jump rates 1 at $x = 0$, one has for $x \geq 0$

$$\tilde{\lambda}_+(x) = 1 - \beta_{\text{out}} x + \zeta_{\text{out}} x^2 + \mathcal{O}\left(|x|^3 R^{-6}\right),$$
$$\tilde{\lambda}_-(x) = 1 + \beta_{\text{in}} x + \zeta_{\text{in}} x^2 + \mathcal{O}\left(|x|^3 R^{-6}\right),$$

with

$$R^2 = \left(\beta_{\text{in}} + \beta_{\text{out}}\right)^{-1}, \tag{69}$$

$$\beta_{\text{out}} = \sqrt{\frac{2}{\pi}}\left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}}\right)\gamma_{\text{out}} = \sqrt{\frac{2}{\pi}}\vartheta\gamma_{\text{out}} = \mathcal{O}\left(R^{-2}\right), \tag{70}$$

$$\beta_{\text{in}} = \sqrt{\frac{2}{\pi}}\left(\frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}}\right)\gamma_{\text{in}} = \sqrt{\frac{2}{\pi}}\vartheta\gamma_{\text{in}} = \mathcal{O}\left(R^{-2}\right), \tag{71}$$

$$\zeta_{\text{out}} = \frac{2\gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{\pi\kappa_-\sqrt{\kappa_+^2 + 4\omega_T^2}} = \frac{2\gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{\pi\kappa_-\xi} = \mathcal{O}\left(R^{-4}\right), \tag{72}$$

$$\zeta_{\text{in}} = \frac{2\gamma_{\text{in}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{\pi\kappa_-\sqrt{\kappa_+^2 + 4\omega_T^2}} = \frac{2\gamma_{\text{in}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{\pi\kappa_-\xi} = \mathcal{O}\left(R^{-4}\right). \tag{73}$$

Here we use the shorthands

$$\vartheta = \frac{1}{\kappa_-} + \frac{1}{\sqrt{\kappa_+^2 + 4\omega_T^2}},$$
$$\xi = \sqrt{\kappa_+^2 + 4\omega_T^2}.$$

The stationary distribution of the position $X = X^{(1)}$ of LP–1 then is

$$\phi(x) = \phi(0)\exp\left(-\frac{x^2}{2R^2} + m_1|x| + m_3|x|^3\right)\exp(c(x)) \quad (x \in \mathbb{Z}, |x| \leq R^{3/2}),$$
$$m_1 = \frac{\beta_{\text{out}} - \beta_{\text{in}}}{2} = \frac{\vartheta}{\sqrt{2\pi}}\left(\gamma_{\text{out}} - \gamma_{\text{in}}\right) = \mathcal{O}\left(R^{-2}\right)$$
$$m_3 = \frac{\beta_{\text{in}}^2 - \beta_{\text{out}}^2}{6} + \frac{\zeta_{\text{out}} - \zeta_{\text{in}}}{3} = \left(\gamma_{\text{in}}^2 - \gamma_{\text{out}}^2\right)\left(\frac{\vartheta^2}{3\pi} - \frac{2}{3\pi\kappa_-\xi}\right) = \mathcal{O}\left(R^{-4}\right)$$
$$\phi(0) = \frac{1}{\sqrt{2\pi}R\alpha_R}$$
$$\alpha_R = 1 + \sqrt{\frac{2}{\pi}}m_1 R + 2\sqrt{\frac{2}{\pi}}m_3 R^3 + \mathcal{O}\left(R^{-2}\right) = 1 + \mathcal{O}\left(R^{-1}\right)$$
$$|c(x)| \leq \mathcal{O}\left(\frac{x^2}{R^4} + \frac{x^4}{R^6}\right)$$

and $\phi(x)$ at least exponentially decaying for $|x| > R^{3/2}$. Observe that under this distribution ϕ,

$$\mathbb{E}_\phi\left[|X|\right] = \left(1+\mathcal{O}(R^{-2})\right)\alpha_R^{-1}\left(\sqrt{\frac{2}{\pi}}R + m_1R^2 + 3m_3R^4\right) = \sqrt{\frac{2}{\pi}}R + \underbrace{\left(1-\frac{2}{\pi}\right)m_1R^2 + \left(3-\frac{4}{\pi}\right)m_3R^4}_{\mathcal{O}(1)} + \mathcal{O}\left(R^{-1}\right),$$

$$\mathbb{E}_\phi\left[X^2\right] = \left(1+\mathcal{O}(R^{-2})\right)\alpha_R^{-1}\left(R^2 + 2\sqrt{\frac{2}{\pi}}m_1R^3 + 8\sqrt{\frac{2}{\pi}}m_3R^5\right) = R^2 + \underbrace{\sqrt{\frac{2}{\pi}}\left(m_1R^3 + 6m_3R^5\right)}_{\mathcal{O}(R)} + \mathcal{O}(1).$$

With this, we can explicitly express the skew dependency of (63-65),

$$\begin{aligned}
\mathbb{E}\left[\theta_b^{(1)} - \theta_a^{(1)}\right] &= \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)\mathbb{E}_\phi\left[|X|\right],\\
\mathbb{E}\left[\left(\theta_a^{(1)}\right)^2 + \left(\theta_b^{(1)}\right)^2\right] &= \left(\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2\right)\mathbb{E}_\phi\left[X^2\right],\\
\mathbb{E}\left[\max\left(\theta_b^{(1)} - \theta_a^{(1)}, 0\right)^2\right] &= \max\left(\gamma_{\text{in}} - \gamma_{\text{out}}, 0\right)^2\mathbb{E}_\phi\left[X^2\right]
\end{aligned}$$

and finally, the cost of internalisation equals

$$\frac{1}{4\pi}\vartheta(s^*-s)\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)R \tag{74}$$

$$+\left[\frac{1}{2\sqrt{2\pi}}\vartheta - \frac{1}{8\sqrt{2\pi}}\left(\frac{1}{\kappa_-} + \frac{\kappa_+^2}{\xi^3}\right) - \frac{s}{4\pi\kappa_-\xi}\right]\left(\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2\right)R^2 \tag{75}$$

$$+\frac{s}{4\pi\kappa_-\xi}\max\left(\gamma_{\text{in}} - \gamma_{\text{out}}, 0\right)^2R^2 \tag{76}$$

$$+\frac{1}{4\sqrt{2\pi}}\vartheta(s^*-s)\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)\left(\left(1-\frac{2}{\pi}\right)m_1R^2 + \left(3-\frac{4}{\pi}\right)m_3R^4\right) \tag{77}$$

$$+\mathcal{O}\left(R^{-3}\right), \tag{78}$$

with s^* from Proposition 3, (74) typically of order R^{-1}, and (75-77) of order R^{-2}. This proves Eqs. (34) and (37).

As for the market share of LP–1, one can deduce Eqs. (35) and (38) from Eqs. (67 – 68)

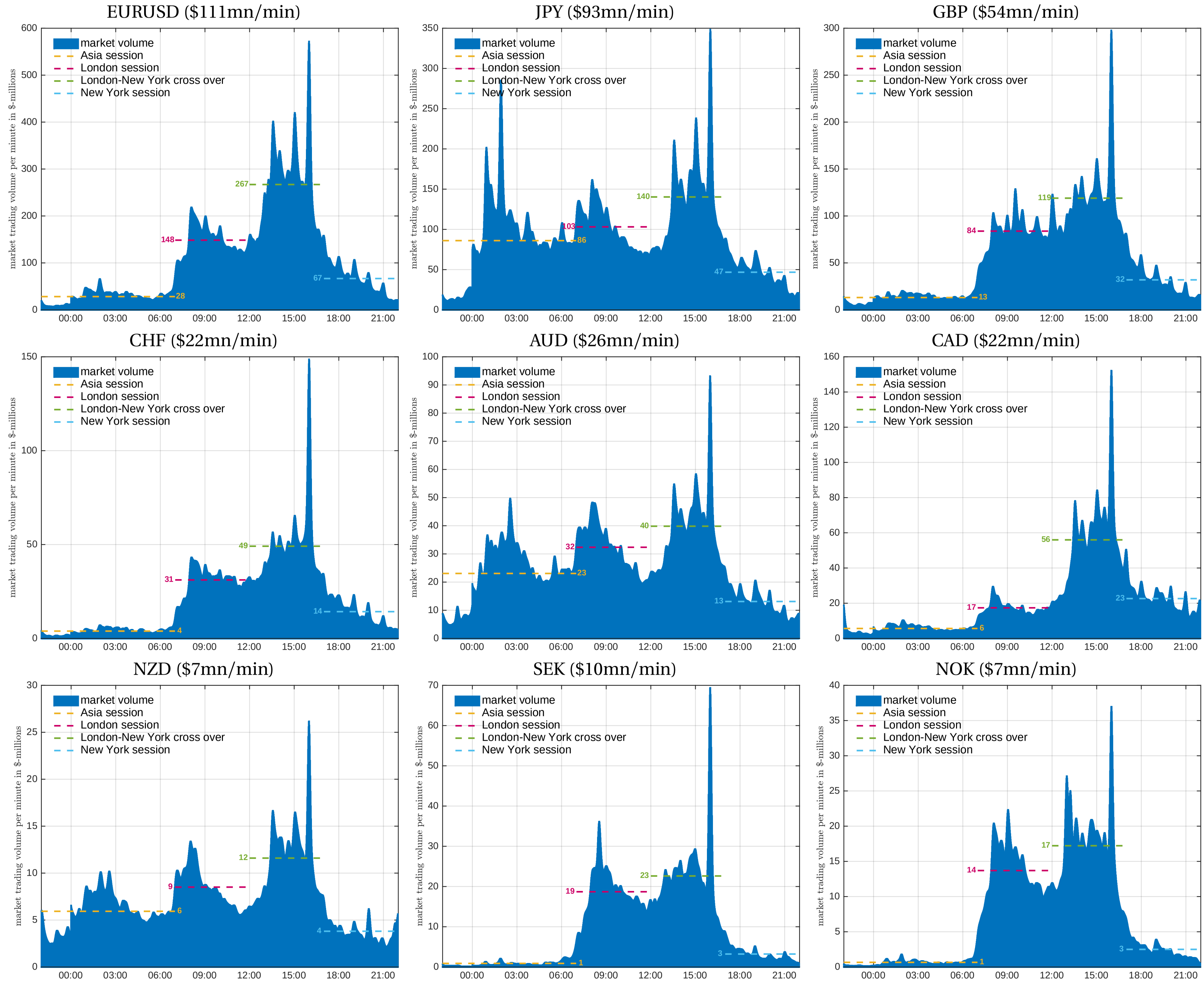
$$\begin{aligned}
&\mathbb{E}\left[\lambda_+\left(X^{(1)}\right) + \lambda_-\left(X^{(1)}\right)\right]\\
&= \frac{1}{2} + \frac{1}{2\sqrt{2\pi}}\vartheta\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)\mathbb{E}_\phi\left[\left|X^{(1)}\right|\right] + \frac{\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{2\pi\kappa_-\xi}\mathbb{E}_\phi\left[\left(X^{(1)}\right)^2\right] + \mathcal{O}\left(R^{-3}\right)\\
&= \frac{1}{2} + \underbrace{\frac{1}{2\pi}\vartheta\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)R}_{\mathcal{O}(R^{-1})}\\
&\quad + \underbrace{\frac{\gamma_{\text{in}}^2 + \gamma_{\text{out}}^2 - \left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)_+^2}{2\pi\kappa_-\xi}R^2 + \frac{1}{2\sqrt{2\pi}}\vartheta\left(\gamma_{\text{in}} - \gamma_{\text{out}}\right)\left(\left(1-\frac{2}{\pi}\right)m_1R^2 + \left(3-\frac{4}{\pi}\right)m_3R^4\right)}_{\mathcal{O}(R^{-2})}\\
&\quad + \mathcal{O}\left(R^{-3}\right).
\end{aligned}$$

Finally, as asymmetric skewing affects the bid and ask price in a different manner, the average spread shown by LP–1 now equals

$$\begin{aligned}\mathbb{E}\left[a^{(1)}-b^{(1)}\right] &= s+\mathbb{E}\left[\theta_a^{(1)}-\theta_b^{(1)}\right] = s+(\gamma_{\text{out}}-\gamma_{\text{in}})\mathbb{E}_\phi\left[\left|X^{(1)}\right|\right] \\ &= s+(\gamma_{\text{out}}-\gamma_{\text{in}})\left(\sqrt{\frac{2}{\pi}}R+\left(1-\frac{2}{\pi}\right)m_1R^2+\left(3-\frac{4}{\pi}\right)m_3R^4\right)+\mathcal{O}\left(R^{-3}\right),\end{aligned}$$

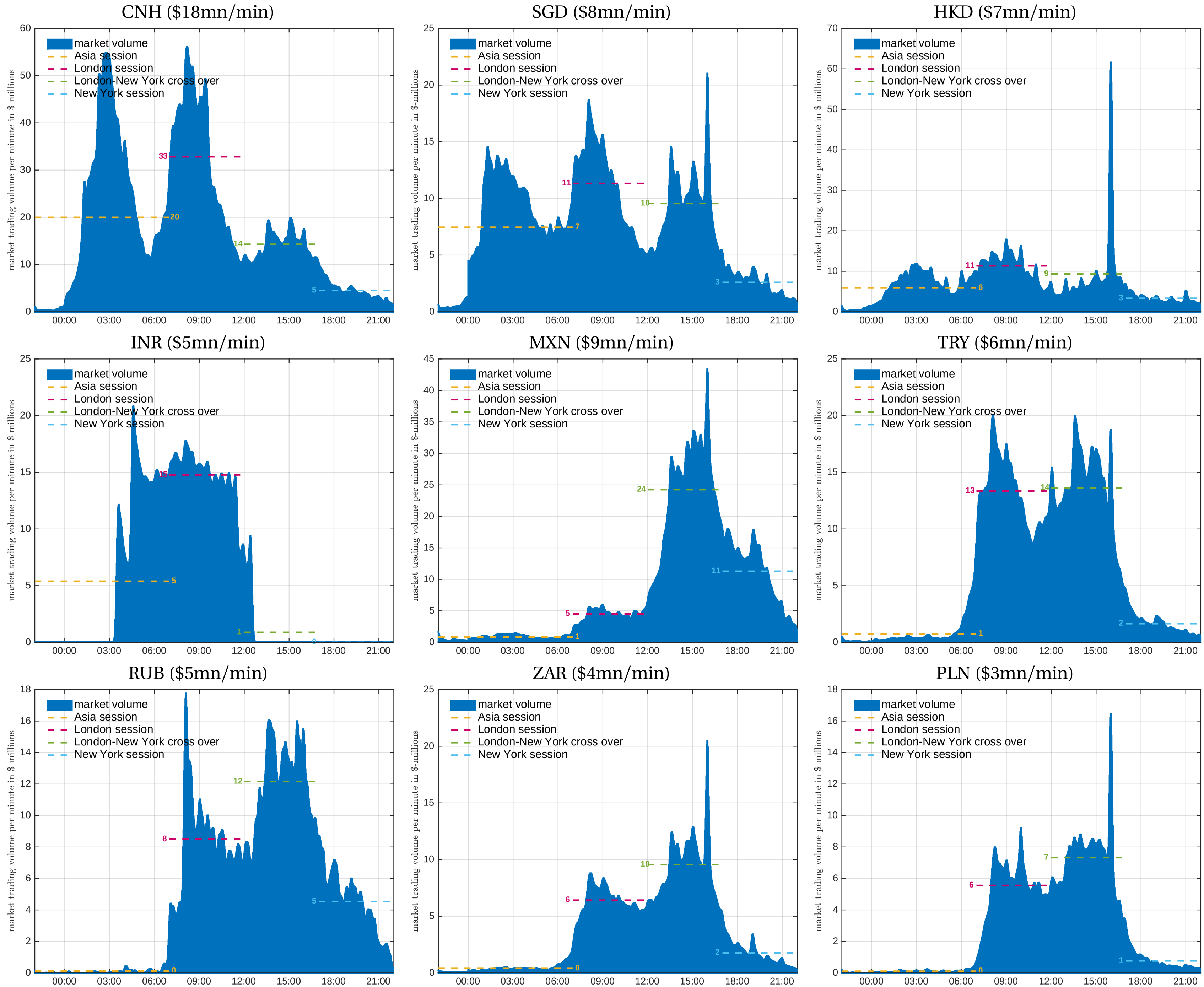
proving (36) and (39). ■

Figure 11: Diurnal volume profile for the customer-facing electronic FX spot segment (G10 currencies)



Note. These charts draws the (smoothed) per minute volume profile for selected currencies, rescaled to match the BIS (2016) implied ADV figures for the customer-facing electronic FX spot segment. X-axis in London local time. The dashed lines indicate the average trading volume per minute (in $mn) for each of the four trading session with the daily figure reported in the chart title. For each currency listed, the results reflect the combined activity across the EUR and USD crosses, e.g. "GBP" consolidates EURGBP and GBPUSD. The volume profiles are derived from EBS Markets trade volume data and Reuters matching trade count data for 2016.

Figure 12: Diurnal volume profile for the customer-facing electronic FX spot segment (Asian and emerging markets currencies)



Note. These charts draws the (smoothed) per minute volume profile for selected currencies, rescaled to match the BIS (2016) implied ADV figures for the customer-facing electronic FX spot segment. X-axis in London local time. The dashed lines indicate the average trading volume per minute (in $mn) for each of the four trading session with the daily figure reported in the chart title. For each currency listed, the results reflect the combined activity across the EUR and USD crosses, e.g. "TRY" consolidates EURTRY and USDTRY. The volume profiles are derived from EBS Markets trade volume data and Reuters matching trade count data for 2016.

SALES AND TRADING DEPARTMENT – DISCLAIMER

This document is intended for discussion purposes only and does not create any legally binding obligations on the part of Deutsche Bank AG and/or its affiliates ("DB"). Without limitation, this document does not constitute an offer, an invitation to offer or a recommendation to enter into any transaction. DB is not acting as your legal, financial, tax or accounting adviser or in any other fiduciary capacity with respect to any proposed transaction mentioned herein. This document does not constitute the provision of investment advice and is not intended to do so, but is intended to be general information. Any product(s) or proposed transaction(s) mentioned herein may not be appropriate for all investors and before entering into any transaction you should take steps to ensure that you fully understand the transaction and have made an independent assessment of the appropriateness of the transaction in the light of your own objectives, needs and circumstances, including the possible risks and benefits of entering into such transaction. You should also consider seeking advice from your own advisers in making any assessment on the basis of this document. If you decide to enter into a transaction with DB, you do so in reliance on your own judgment. The information contained in this document is based on material we believe to be reliable; however, we do not represent that it is accurate, current, complete, or error free. Assumptions, estimates and opinions contained in this document constitute our judgment as of the date of the document and are subject to change without notice.

This material was prepared by a Sales or Trading function within DB, and was not produced, reviewed or edited by the Research Department (which is independent from the Sales or Trading function). Any opinions expressed herein may differ from the opinions expressed by other DB departments including the Research Department. Sales and Trading functions are subject to additional potential conflicts of interest which the Research Department does not face. DB may engage in transactions in a manner inconsistent with the views discussed herein. In general, Sales and Trading personnel are compensated in part based on the volume of transactions effected by them. DB seeks to transact business on an arm's length basis with sophisticated investors capable of independently evaluating the merits and risks of each transaction, with investors who make their own decision regarding those transactions.

The distribution of this document and availability of these products and services in certain jurisdictions may be restricted by law. You may not distribute this document, in whole or in part, without our express written permission. DB SPECIFICALLY DISCLAIMS ALL LIABILITY FOR ANY DIRECT, INDIRECT, CONSEQUENTIAL OR OTHER LOSSES OR DAMAGES INCLUDING LOSS OF PROFITS INCURRED BY YOU OR ANY THIRD PARTY THAT MAY ARISE FROM ANY RELIANCE ON THIS DOCUMENT OR FOR THE RELIABILITY, ACCURACY, COMPLETENESS OR TIMELINESS THEREOF. DB is authorized under German Banking Law (competent authority: BaFin - Federal Financial Supervising Authority) and DB in the UK is authorised by the Prudential Regulation Authority and is subject to limited regulation by the Prudential Regulation Authority and Financial Conduct Authority. Details about the extent of our authorisation and regulation are available on request.

References

Bank of England, H.M. Treasury, and Financial Conduct Authority, 2014, "Fair and effective markets review," available at http://www.bankofengland.co.uk/markets/Documents/femr/consultation271014.pdf.

BIS, 2015, "Triennial Central Bank Survey of Foreign Exchange and OTC Derivatives Market Activity. Reporting guidelines for turnover in April 2016," Monetary and Economic Department, Bank for International Settlements, available at http://www.bis.org/statistics/triennialrep/guidelines_cbanks.htm.

———, 2016, "Triennial central bank survey. Foreign exchange turnover in April 2016," Monetary and Economic Department, Bank for International Settlements, available at http://www.bis.org/publ/rpfx16.htm.

Euromoney, 2016, "All change in the 2016 Euromoney FX rankings," available at http://www.euromoney.com/Article/3556871/All-change-in-the-2016-Euromoney-FX-rankings.html.

Harris, M., 2010, "Little's Law: The science behind proper staffing," *Emergency Physicians Monthly*, Feb.

Hasbrouck, J., and R. M. Levich, 2017, "FX Market Metrics: New Findings Based on CLS Bank Settlement Data," NBER Working Paper No. w23206, available at http://www.nber.org/papers/w23206.

Hopp, W. J., and M. Spearman, 2000, *Factory Physics: Foundations of Manufacturing Management*. Irwin/McGraw-Hill, New York, 2nd edn.

Kirk, A., J. McAndrews, P. Sastry, and P. Weed, 2014, "Matching Collateral Supply and Financing Demands in Dealer Banks," *FRBNY Economic Policy Review*, December, 127 – 151.

Little, J. D., 1961, "A proof for the queuing formula: $L = \lambda W$," *Operations Research*, 9 (3), 383 – 387.

———, 2011, "Little's Law as Viewed on Its 50th Anniversary," *Operations Research*, 59 (3), 536 – 549.

Moore, M., A. Schrimpf, and V. Sushko, 2016, "Downsized FX markets: causes and implications," *BIS Quarterly Review*, December, 35 – 51.

Oomen, R. C., 2017a, "Execution in an aggregator," *Quantitative Finance*, 17 (3), 383 – 404.

———, 2017b, "Last look," *Quantitative Finance*, 17 (7), 1057 – 1070.

Virtu Financial, Inc, 2014, "FORM S-1, REGISTRATION STATEMENT UNDER THE SECURITIES ACT OF 1933," Registration No. 333, available at https://www.sec.gov/Archives/edgar/data/1592386/000104746914002070/a2218589zs-1.htm.

Wolff, R. W., 1989, *Stochastic Modeling and the Theory of Queues*. Prentice-Hall, Inc., Englewood Cliffs, NJ.