UNITED KINGDOM · CHINA · MALAYSIA

Discussion Papers in Economics

Discussion Paper
No. 14/03

Decomposing the bid-ask spread in multi-dealer Markets

Michael Bleaney and Zhiyong Li

April 2014

Decomposing the bid-ask spread in multi-dealer markets

Michael Bleaney and Zhiyong Li *
University of Nottingham

Abstract

In this paper, we modify the Huang and Stoll (1997) spread-decomposing model to fit multi-dealer markets. In a multi-dealer market, individual dealers can rebalance their inventories either by trading with other dealers or changing the quote price. Our modified model captures this feature. Using transaction data from the Reuters D2000-1 system, we find that the order-processing and inventory control components of the spread in the foreign exchange market are relatively small and dealers may tolerate the unwanted inventory to keep the spread small to attract informed orders. The asymmetric information component carries the biggest weight. We study the time pattern of the spread and its components. The spread varies significantly with the time of day, but the inventory control and asymmetric information components do not.

Keywords: Bid-ask Spread, Multi-dealer Market, Decomposition
JEL: G10 G15

*Corresponding Author: Michael Bleaney, Professor of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, England; Tel +44 115 95 15464; Email: michael.bleaney@nottingham.ac.uk

1 Introduction

The bid-ask spread provides an important measure of trade costs and therefore market liquidity. It follows that understanding the determination of both the spread and its components is of interest. This paper makes two contributions to the literature. First, it develops a new model for estimating and decomposing the spread, based on the framework of Huang and Stoll (1997) (HS hereafter). In contrast to most existing models which assume there is only one market maker (e.g. New York Stock Exchange, NYSE), the new model is designed specifically for multi-dealer markets. We will refer to the new model as the modified HS model (MHS model thereafter). Second, we introduce time dummies in the MHS model so that we can study the intra-day pattern of the spread and its components. Until now, there has been little research on the intraday spread pattern in the Reuters D2000-1 system, which is an important component of the foreign exchange market, perhaps in part because spreads cannot be observed but need to be estimated.

There are two kinds of inter-dealer market structures: the limited order book market (the order-driven market) and the direct trade or sequential trade market (the quote-driven market).[a] An example of the former is the Electronic Brokerage System (EBS), while an example of the latter is the Reuters D2000-1 system. An order-driven market does not have a market maker as it is organized as a two-sided auction. A quote-driven market has one or more market makers (dealers) who supply the liquidity and offer quotes to other market participants. Models designed to estimate the spread and its decomposition are highly sensitive to the structure of the market. Huang and Stoll (1997) develop a model for estimating the efficient spread and calculating the fraction of each of its components. They show that covariance spread models and trade indicator spread models are special cases of their model.

Although the HS model encompasses all other spread models, it very much reflects the structure of NYSE, which has a single market maker for each stock. For other structures, the HS equations need to be interpreted differently. For example, McGroarty et al. (2007) develop an alternative model for decomposing the spread specifically designed for order-driven intra-dealer markets like the EBS. Their model uses the price difference between the best buy and sell limit orders as the measure of the spread, and then re-interprets one of the HS equations to estimate the decomposition. The interpretation of the HS model in multi-dealer markets such as the Reuters D2000-1 system has not been previously discussed. Multi-dealer markets allow individual dealers to offload their inventory imbalances onto other dealers, and such inter-dealer

[a]See Viswanathan and Wang (2004) for a more detailed description of the two markets.

("hot-potato") trading is quantitatively important (Lyons 1995). This invalidates the HS assumption that the inventory imbalances of dealers are equivalent to the sum of all previous trades.

The unsuitability of the original HS model to the Reuters D2000-1 system is indicated by the fact that one of the key predictions of the HS model fails to hold. In the original HS model, an inventory imbalance is met by price adjustment designed to induce customer trades in the opposite direction to those that have just occurred. Thus the HS model predicts negative serial correlation of trade direction. Hot-potato trading to reduce inventory imbalances is ruled out by assumption in the HS model, but if it is prevalent trade direction will tend to be positively serially correlated, and this is what we find in data from the Reuters D2000-1 system.

Here we derive a model for estimating and decomposing spreads in markets with more than one market maker. The new model shares the same basic structure of the HS model, but has a different interpretation. Using USD/DEM transaction data, we compare the results for decomposing the spread obtained by the MHS and the HS models. Unlike datasets from stock markets such as NYSE and NASDAQ, foreign exchange datasets provide information about inter-dealer transactions.

The intra-day patterns of the bid-ask spread have been widely examined for a variety of markets. For instance, McInish and Wood (1992) find that the spread of NYSE stocks has a U-shaped intra-day pattern, while Chan et al. (1995)) find the inter-dealer spread in NASDAQ exhibits an L-shaped pattern. Among studies of the foreign exchange market, Daníelsson and Payne (2002) find a U-shaped spread pattern in the Reuters D2000-2 trading system and Hua and Li (2011) find that the spread pattern of the JPY/USD pair in Electronic Broking Services (EBS) is U-shaped during Tokyo trading hours and inverse U-shaped during London trading hours.

Because spread information is not always available, the literature studying the intra-day spread pattern on the Reuters D2000-1 system, an important market structure within the foreign exchange market, is very thin. Our modified HS model is specifically designed for the market structure of the Reuters D2000-1 system. We find that spreads of the USD/DEM pair are inverse U-shaped during the Asia trading hours, are stable during the European trading hours and become larger after the closing of European markets. The inventory control part and the asymmetric information part are stable during the day. The components of the spread are stable over the day.

We organize the rest of the paper as follows. In Section 2, we briefly introduce the HS model. In Section 3, we present our new model and compare the results with the HS model. In Section 4, we introduce time dummies in order to study the intra-day pattern of the spread. We use the model to analyse USD/DEM transaction data collected from the Reuters D2000-1 system.

2 Theoretical Background

Bid-ask spreads reflect market-maker costs, which can be broken down into three types. First, the order processing cost reflects dealers' operating costs, such as labour costs and platform commissions. Second, an asymmetric information cost is incurred when a dealer trades with an agent with better information about the fundamental price. Dealers will make losses when trading with better-informed agents unless they set the spread to reflect this risk. Third, inventory control costs arise when a dealer aims to keep inventories within a certain range. Market competition is another factor that can influence the spread. A general model of the spread based on Bollen et al. (2004) can be written as:

$$SP_t = (OP_t, IC_t, AS_t, COM_t) \tag{1}$$

or more specifically,

$$SP_t = OP_t + IC_t + AS_t + COM_t \tag{2}$$

where SP_t is the efficient bid-ask spread, OP_t is the order processing cost, IC_t is the inventory control cost, AS_t is the asymmetric information cost, and COM_t is the degree of competition. A model in which the spread is the dependent variable can be called a spread determination model.

The time pattern of the bid-ask spread can be modelled using the following regression:

$$SP_t = \tau_1 + \sum \tau_i timedummy_i + \varepsilon_t \tag{3}$$

This model is particularly useful for studying the intra-day pattern. As in the spread determination model, the spread is the dependent variable in this model, but now, the independent variables are time dummies instead of proxy variables. Equation (3) will be called the spread description model (SD model) in this paper.

The Huang and Stoll model (HS model) aims to estimate the bid-ask spread and to decompose it into its components, using information about transactions prices and trade direction. The structure of the HS model is as follows.

The price of an asset can be decomposed into the bid-ask spread and the mid-price, or the midpoint between the bid price and the ask price. Formally, the price is given by:

$$s_t = M_t + \frac{SP}{2} \cdot BS_t \tag{4}$$

where s is the transaction price and M_t is the mid-price. SP is the bid-ask spread, BS is an indicator that gives the direction of the trade.

$$BS = \begin{cases} 1 & \textit{buy order} \\ -1 & \textit{sell order} \end{cases} \tag{5}$$

The spread will affect the return only when the direction of the trade changes ($BS_t - BS_{t-1} \neq 0$). Then the return is given by:

$$\Delta s_t = \Delta M_t + \frac{SP}{2}(BS_t - BS_{t-1}) \tag{6}$$

where Δ is the first-order difference operator.

The mid-price depends not only on the fundamental value of the asset but also on the degree of divergence from the ideal inventory level. This is because dealers are assumed to adjust the mid-price to correct their inventory imbalances. Crucially, inventory imbalances are equated with the sum of all previous trades. This assumption is reasonable within the context of a mono-dealer market, where all trades are between customers and a single dealer. If the dealer starts with ideal inventory levels before trading begins, then the sum of trades will exactly reflect the deviation of inventories from the ideal level. Formally, the mid-price is given by:

$$M_t = F_t + \beta \cdot \frac{SP}{2} \sum_{i=1}^{t-1} BS_i \tag{7}$$

where F_t is the fundamental, and $\frac{SP}{2}$ is the half spread. Taking the first-order difference of Equation (7) gives:

$$\Delta M_t = \Delta F_t + \beta \cdot \frac{SP}{2} BS_{t-1} \tag{8}$$

Equation (8) shows that the change in the mid-price is a function of the change in the fundamental and the incoming order in the previous period, where $\beta \cdot \frac{SP}{2}$ gives the effect of the inventory level on the mid-price.

If dealers are aware of serial autocorrelation in the order flow, then given the order flow in the previous period, dealers know the conditional expectation of the order flow in the current period. Dealers are aware that customers may be better informed about the fundamental value, and so the direction of the most recent trade (relative to its expectation) affects their beliefs about the fundamental value. Formally, the change in the fundamental value can be written as:

$$\begin{aligned} \Delta F_t &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} [E(BS_{t-1} | BS_{t-2})] \\ &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} (1 - 2\theta) BS_{t-2} \end{aligned} \tag{9}$$

where $\alpha\frac{SP}{2}$is the effect of an incoming order on the dealer's beliefs about the fundamental value, and θ is the probability of an order reversal. The conditional expectation of an incoming order BS_{t-1} given that BS_{t-2} is known can be written as:

$$E(BS_{t-1}|BS_{t-2}) = (1-2\theta)BS_{t-2} \tag{10}$$

Taking this expectation into account, the HS model is given by:

$$\Delta s_t = \frac{SP}{2}BS_t + (\alpha+\beta-1)\frac{SP}{2}BS_{t-1} - \alpha\frac{SP}{2}(1-2\theta)BS_{t-2} + \varepsilon_t \tag{11}$$

Let $\vartheta = (1-2\theta)$. θ can be estimated from

$$BS_{t-1} = \vartheta BS_{t-2} + \varepsilon_t \tag{12}$$

The generalised method of moments is used to estimate the two equations simultaneously.

The weight of inventory control costs on the bid-ask spread is given by β, and the weight of asymmetric information costs is given by α. $1-\alpha-\beta$ is the weight of the other factors influencing the bid-ask spread, which include order processing costs and market competition.

3 Modified HS Model

The question is how to adapt the HS model to a multi-dealer market. The first point is that in such a market there are two types of trade: between dealers and customers, and between one dealer and another. Trades between dealers arise because dealers can adjust their inventories by placing orders with other dealers. According to King et al. (2013, p. 98), inter-dealer trades constitute about 30% of all trades in the foreign exchange market.

Evans and Lyons (2002) develop a model to describe the price formation process in a multi-dealer market such as the Reuters D2000-1 system. In each day there are assumed to be three rounds of dealing. In the first round, dealers trade with the non-dealer public and collect private information from customer order flows. In the second round, dealers trade with each other, and thus the private information spreads through the market and becomes public. In the third round, dealers quote prices based on the aggregate information from the second round and trade with the public again, with the aim of returning their inventories to the desired level. Thus customer order flow is transmitted initially only to individual dealers, and is aggregated at the second stage through inter-dealer trading. At the third stage prices adjust to induce inventory correction, as assumed in the HS model.

Osler et al. (2011) show that, in the foreign exchange market, dealers do not widen spreads to customers to deal with asymmetric information, as the HS model assumes. Instead they even reduce the spread to informed customers in an effort to acquire such information, which they can then exploit profitably in inter-dealer trading. This highlights the possibility that spreads may differ between types of trades: inter-dealer trades, trades with informed customers and trades with uninformed customers.

In a multi-dealer market there is not an equivalence between the sum of all previous trades and the aggregate inventory imbalance of dealers; this would only be true if inter-dealer trades were excluded, for inter-dealer trades are recycling inventory imbalances from one dealer to another, rather than creating new ones. Thus the HS model would only work if the data related exclusively to customer trades. In fact the 1996 data used by Evans and Lyons (2002), and also by us below, relate exclusively to inter-dealer trades. Inter-dealer trades are assumed to result from inventory imbalances of the initiating dealer. These imbalances would initially be a consequence of the initiating dealer's trading with customers, but could subsequently reflect "hot-potato" trading by other dealers.

Consider the case where customers are net buyers of dollars. In general, a minority of individual dealers will find themselves accumulating dollars in the first round, because their customers are net sellers. Thus the aggregate of absolute inventory imbalances of individual dealers will be greater than or equal to the aggregate market imbalance. It is easiest to focus on the case where this is an equality, i.e. all dealers have customers who are net buyers of dollars. Assume that at the first stage of round two, each dealer tries to buy a proportion v of his inventory imbalance from another dealer, where $0 \leq v \leq 1$. However, this only recycles the aggregate imbalance between dealers. At the second stage dealers try to recycle a proportion v, and so on. The aggregate imbalance is evenly distributed among traders eventually. Then, if the aggregate inventory imbalance is X, total interdealer trades Z will be given by:

$$Z = k_1 \cdot X \tag{13}$$

where $k_1 \geq 1$. The appendix shows that when there are N dealers in the market, if each dealer trades $\frac{1}{N}$ of his total imbalance with every other dealer, i.e. $v = \frac{N-1}{N}$, the equilibrium is achieved at the first stage. Then the sum of total inter-dealer trades equals the average inventory imbalance. In any other cases, the sum of total inter-dealer trades should exceed the aggregate inventory imbalance. Thus, k_1 is a measure of the efficiency of imbalance re-distribution. When $k_1 = 1$, the re-distribution is at its most efficient. Generally, aggregate inventory imbalance is given as follows,

$$k_1 \sum_{i=1}^{t-1} BS_i \tag{14}$$

Another difference between the HS model and our modified HS model is how the dealer's inventory level is determined. This difference is driven by different market structures. Dealers in a multi-dealer market control inventory differently from dealers in a single dealer market. A dealer can control his inventory using either the passive or the active method. The passive method is to adjust the mid-quote to attract an order flow in the opposite direction of the previous flow. The active method is to either initiate a trade directly (in quote-driven markets such as Reuters D2000-1) or make a market or a limit order (in order-driven markets such as EBS). The direction of this order is the same as that of the previous order. The first method is considered and explained in spread-estimating models of the stock market (e.g., the HS model). The second method is first modelled in a theoretical paper specifically describing the foreign exchange market (Lyons 1997). Dealers in the HS model can only use the first method, whereas dealers who are in a multi-dealer market can use both methods.

Using the first method, as in the HS model, a dealer has to wait for an incoming order and faces uncertainty about this incoming order. The second method is more efficient because a dealer can return the inventory to the ideal level immediately and surely. The literature suggests that dealers rebalance inventory very quickly (e.g. Lyons 1997, Bjønnes and Rime 2005, Osler et al. 2011). King et al. (2013) suggest that a dealer does not change the price when his inventory is different from the desired level. Lyons (1997) and Evans and Lyons (2002) emphasise that in the foreign exchange inter-dealer market, a quote-driven market, dealers exchange price information and control the inventory through hot potato trading, in which dealers make an order right after receiving an order in the same direction. Thus, they use the second method to control the inventory.

In the HS model, for a dealer in a single dealer market, an inventory imbalance is the accumulated past incoming order ($\sum_{i=1}^{t-1} BS_i$ in equation 7). Having an efficient method to control the inventory, dealers in a multi-dealer market can get rid of unwanted inventory fast and can tolerate some amount of unwanted inventory without moving the price. We define the inventory imbalance which would influence mid-prices as the *intolerable inventory*. The intolerable inventory is less than or equal to the unwanted inventory because of dealers' tolerance. Formally, the amount of intolerable inventory before the most recent order is given by:

$$k_2 \sum_{i=1}^{t-2} BS_{i,d} \tag{15}$$

where $0 < k_2 < 1$. k_2 is a measure of the dealer's ability to keep the intolerable inventory close to zero by using his/her own order (the active method). $k_2 = 0$ suggests that the dealer can eliminate all the unwanted inventory or tolerate them. $k_2 = 1$ suggests that the dealer cannot eliminate the unwanted inventory at all, which is the case of the HS model. The most recent order is yet to be taken into account. Therefore, if dealer d received an incoming order at period $t-1$, dealer d's inventory level at period t is the most recent order and the intolerable inventory. Formally, the inventory level is given by:

$$BS_{t-1,d} + k_2 \sum_{i=1}^{t-2} BS_{i,d} \tag{16}$$

We assume dealers in the market are identical. Then an initiating trader chooses a quote dealer randomly. Under these circumstances, dealers' cumulated incoming orders are identical in the long run. If there are N dealers in the market, the cumulated incoming orders are evenly distributed among dealers. Taking Equation (14) into account, dealer d's cumulated incoming order at period t-2 is given as follows:

$$k_1 BS_{t-1,d} + k_2 \cdot \sum_{i=1}^{t-2} BS_{i,d} = k_1 BS_{t-1} + \frac{1}{N} \cdot k_1 \cdot k_2 \cdot \sum_{i=1}^{t-2} BS_i$$

If the trade is observed by other participants in the market, then, dealer d's inventory information is known by the participants i.e. it has become new public information. Then, all participants will update their quotes as a response to the trade that has just happened. If the no-arbitrage condition, which suggests all dealers' quotes at any time are the same, is valid, it makes sure that dealer d's quote increment be the quote increment in the whole market. Therefore, we re-write Equation (7) as follows:

$$M_t = F_t + \beta \cdot \frac{SP}{2} \left[k_1 BS_{t-1} + \frac{1}{N} \cdot k_1 \cdot k_2 \cdot \sum_{i=1}^{t-2} BS_i \right] \tag{17}$$

Taking the first-order difference of Equation (17), we obtain that:

$$\Delta M_t = \Delta F_t + \beta \cdot \frac{SP}{2} \cdot k_1 BS_{t-1} + \beta \cdot \frac{SP}{2} \cdot \left(\frac{k_1 k_2}{N} - 1 \right) \cdot BS_{t-2} \tag{18}$$

Equation (18) suggests that inventory control costs influence the mid-price through the two most recent orders, whereas in the original HS model, inventory control costs affected the mid-price only through the most recent order. When there is only one dealer in the market and thus the deal cannot use the active method to manage his/her inventory ($k_1 = k_2 = N = 1$), the equation above reduces to the HS model.

The setting of asymmetric information costs is the same as the HS model. Substituting Equation (9) into Equation (18) gives the change in mid-price,

$$\begin{aligned}\Delta M_t &= \alpha \cdot \tfrac{SP}{2} \cdot BS_{t-1} - \alpha(1-2\theta) \cdot \tfrac{SP}{2} \cdot BS_{t-1} + \beta \cdot \tfrac{SP}{2} \cdot k_1 BS_{t-1} + \beta \cdot \tfrac{SP}{2} \cdot \left(\tfrac{k_1 k_2}{N} - 1\right) BS_{t-2} \\ &= (\alpha + k_1\beta) \cdot \tfrac{SP}{2} \cdot BS_{t-1} - \left[\alpha(1-2\theta) + \beta\left(1 - \tfrac{k_1 k_2}{N}\right)\right] \cdot \tfrac{SP}{2} \cdot BS_{t-2}\end{aligned} \tag{19}$$

Taking the order possessing cost into account, we can finally obtain the modified HS model. Substituting Equation (19) into Equation (6), we have,

$$\Delta s_t = \frac{SP}{2} BS_t + (\alpha + k_1\beta - 1)\frac{SP}{2} BS_{t-1} - \left[\alpha(1-2\theta) + \beta\left(1 - \frac{k_1 k_2}{N}\right)\right] \frac{SP}{2} BS_{t-2} + \varepsilon_t \tag{20}$$

The regression of our new model is the same as that of the HS model, but the third term has a different meaning. It now represents both α and β rather than α only. The new model will be called the modified HS model (MHS model for short) in this paper. The HS model is a special case of the MHS model when there is only one market maker in the market.

Equation (20) suggests that when k_1 is very big, i.e. when the inter-dealer trades significantly exaggerate the aggregate inventory imbalance, the inventory components should be zero ($\beta = 0$). The inventory component is negatively correlated with the number of dealers in the market and dealers' ability of tolerating/eliminating unwanted inventory. When there are many dealers in the market (N is big) or the dealer can keep intolerable inventory small (k_2 is small), the inventory component is small.

We cannot know the values of k_1 and k_2. Therefore, the MHS model can generate only a range for the components of the spread, rather than a precise number.

Assuming that the imbalance re-distribution process is efficient ($k_1 = 1$), when the intolerable inventory is zero ($k_2 = 0$) or there are infinite dealers in the market ($N = \infty$), the MHS model becomes

$$\Delta s_t = \frac{SP}{2} BS_t + (\alpha + \beta - 1)\frac{SP}{2} BS_{t-1} - [\alpha(1-2\theta) + \beta] \frac{SP}{2} BS_{t-2} + \varepsilon_t \tag{21}$$

The decomposition results show the lower bound of the inventory component and the upper bound of the asymmetric information component. Equation (21) can be called low-inventory-MHS model (LIMHS model).

When there is only one dealer in the market ($N = 1$), and thus the dealer does not tolerate the unwanted inventory ($k_2 = 1$), the MHS model is the same as the HS model, and the results show the upper bound of the inventory component and the lower bound of the asymmetric information components.

4 Decomposing the Spread in the Reuters D2000-1 System

In this section, we decompose the bid-ask spread of the USD/DEM pair in the Reuters D2000-1 system. We first discuss how the MHS model fits the microstructure of the Reuters D2000-1 system. Then we discuss the range of the components of the spread.

4.1 The Reuters D2000-1 system and the MHS model

In the previous section, we developed the MHS model for a multi-dealer market. The Reuters D2000-1 system is a multi-dealer market. However, the MHS model requires trade information to be known by all participants in the market, which is not the case for the Reuters D2000-1 system. We argue here that as long as the no-arbitrage condition is valid, the MHS model fits the microstructure of the Reuters D2000-1 system.

We first introduce the basic organisation of the Reuters D2000-1 system. Trades on D2000-1 happen between two anonymous dealers: a calling dealer who requires quotes and a quoting dealer. The quoting dealer offers bid and ask prices to the calling dealer. The calling dealer has to make a quick decision to buy dollars (make a positive order flow) or sell dollars (make a negative order flow) or reject the quote. If a transaction is made, the time and the direction will be recorded by the system. Two things need to be mentioned. First, traders can only observe their own trading records. Second, though both bid and ask prices (two series of exchange rates) were quoted by the calling dealers, only the price that reflects the direction of actual trade is in the dataset (and the price may be slightly more favourable to the trader than the quote). Both prices and volumes of trades are not observed by traders other than the two participants.

Dealers on the Reuters D2000-1 system can keep requesting quotes from other dealers, so that price information is de facto known by all dealers in the market all the time, and thus quotes from different dealers should be the same at every point of time, otherwise there will be arbitrage. When dealer d receives an incoming order, he/she will update his/her quote based on the private information in the order and the inventory imbalance caused by the order. At the time, dealer d is the only one who receives an order; he/she therefore is the only one in the market who update his/her quote, because other dealers do not have new information. By requesting quotes, all dealers know dealer d's price increment and update their own quotes instantly. Then, though the incoming order flow is unobserved, information about it is incorporated into the price instantly. The assumption of the MHS model is actually satisfied. Therefore, one can conclude that the MHS model fits the microstructure of the Reuters D2000-1 system.

4.2 Empirical Results

Our empirical results are based on inter-dealer transaction data for the USD/DEM pair on the Reuters D2000-1 system from 1 May 1996 to 2 September 1996, the same dataset as used by Evans and Lyons (2002).

Our data have several features. First, the quote data have irregular time spaces. Second, the trade densities vary with the time of the day. For example, the number of trades in GMT 10:00-11:00 is much greater than the number in GMT 1:00-2:00.

We begin by estimating the spread Equation (20) on the assumption that the spread remains constant throughout. As mentioned earlier, the estimating equation is the same for the HS and LIMHS models; it is merely the interpretation that is different. Table (1) shows the results of the regressions of the HS and the LIMHS models. Table (2) gives the interpretations of the coefficients of these regressions, where we can find the ranges of the components of the spread. There are 382 dealers who trade USD/DEM pair in the Reuter D2000-1 system (Evans 1998), so the LIMHS model is likely to provide a more accurate estimate.

The estimate of the half-spread is γ_1 in Table (1), and so the average bid-ask spread of USD/DEM is 0.0000794 in percentage terms or 1.2 pips. Goodhart et al. (2002) find a 2.84 pips spread on average on the Reuters D2000-2 system. Lyons (1995) suggests a 3 pips spread on average from a big USD/DEM dealer. Evans (1998) finds 6 pips of quoted spread from the FXFX dataset. Our data are values of the tradable spread in the market, so this spread should be narrower than the others. γ_2 reflects the effect of order-processing costs and market competition, and according to both models these factors explain 14.41% of the spread according to Table (2). The difference between the HS and the LIMHS models is the interpretation of γ_3, which influences the decomposition of the spread into inventory control and asymmetric information components. According to Table (2), the HS model suggests that the contributions of the inventory control cost and the asymmetric information cost to the spread are 20.4% and 65.2% respectively. The LIMHS model suggests that the shares of the inventory control cost and the asymmetric information cost in the spread are -1.52% and 87.11% respectively. Both the HS and the LIMHS models find that the asymmetric information cost is a dominant component of the spread. According to the Evans and Lyons (2002) model, the main purpose of the inter-dealer trading is to exchange the private information in the customer order flow. It is risky to trade with traders who have private information. Furthermore, having an efficient inventory control method implies that the inventory control cost is relatively low. Therefore, it is not a surprise that the asymmetric information cost is greater than the inventory control cost. Because the HS model does not consider the active method of inventory management, it might overestimate the share of the inventory control cost. In contrast, the negative share of the inventory control cost in the spread given by the LIMHS model is very interesting. The most important finding is that the inventory control cost is very close to zero, which coincides with the main point of the LIMHS model that dealers use an efficient method to manage their inventory level. This finding is consistent with Bjønnes and Rime (2005), who use dealer inventory data from Reuters D2000-1 and find that inventory control does not have a big price effect. Furthermore, there is a debatable explanation of the negative share of inventory control cost, which needs further research in the future. We can interpret the negative share as a compensation of the asymmetric information cost. Dealers in the market do not worry very much about the inventory because they can get rid of unwanted inventory quickly, while the risk arising because of asymmetric information is high. To protect themselves from the loss of trading with an informed trader, dealers set big spreads. However, the big spread reduces liquidity and thus the information exchange in the market. To keep liquidity high so that they can collect private information and cover the asymmetric information cost, dealers may sacrifice the inventory control cost. In other words, they could offer a negative spread in terms of the inventory control cost to encourage trading.

Compared to the HS model, the asymmetric information cost has an even greater share in the spread in the LIMHS model. As mentioned earlier, the HS model does not match the Reuters D2000-1 system, while the LIMHS model captures many features of the system. Therefore, the results of the LIMHS model are more likely to be true or yield values closer to the true values than the HS model.

Table 1: Regressions

$\Delta s_t = \gamma_1 BS_t + \gamma_2 BS_{t-1} - \gamma_3 BS_{t-2} + constant + \varepsilon_t$					
$BS_t \times 10^{-3}$	$BS_{t-1} \times 10^{-3}$	$BS_{t-2} \times 10^{-3}$	$constant \times 10^{-3}$	R^2	N
0.0397*** (52.17)	-0.00572*** (-7.52)	-0.00184* (-2.42)	-1.55×10^{-4} (-0.20)	0.0105	257387
$BS_{t-1} = constant + \mu BS_{t-2} + \varepsilon_t$					
BS_{t-2}			$constant$	R^2	N
0.0707*** (35.95)			0.465*** (334.34)	0.0050	257391

This table present the results of the following regressions

$\Delta s_t = \gamma_1 BS_t + \gamma_2 BS_{t-1} - \gamma_3 BS_{t-2} + constant + \varepsilon_t$

$BS_{t-1} = constant + \mu BS_{t-2} + \varepsilon_t$.

Both the HS and the MHS models use these regressions.

Tick-by-tick USD/DEM transaction data from 1996.5.1 to 1996.9.2 on the Reuters D2000-1 system are used.

BS_t is the trade direction indicator in period t which is 1 if there is a buy order and is -1 if there is a sell order.

N is the number of the observations.

T-statistics is in the parenthesis

*Significant at 5% level ***Significant at 0.1% level

Table 2: Explanations of the Regressions

USD/DEM					
	spread	OP	IC	AS	θ
HS	0.0000794	14.41%	20.4%	65.2%	0.4647
LIMHS	0.0000794	14.41%	-1.52%	87.11%	0.4647

This table present the results of the HS model (regressions 11 and 12) and the LIMHS model (regressions 21 and 12).
The row of HS is about the HS model.
The row of LIMHS is about the LIMHS model.
spread is the estimated spread
OP is the weight of the order processing cost on the spread ($1 - \alpha - \beta$ in the HS and the LIMHS model).
IC is the weight of the inventory control cost on the spread (β in the HS and the LIMHS model).
AS is the weight of the asymmetric information cost on the spread.(α in the HS and the LIMHS model).
θ is probability of order direction reversal, and is calcualted from the results of regression (12)

5 Time Patterns of the Spread and Its Components

In this section, we study how the estimated bid-ask spread varies over time. Most previous papers that have investigated this issue have used quoted spread data, which can be regressed directly on time dummies. Here the spread has to be estimated. By introducing time dummies and interaction terms between time dummies and the trade indicator variables in equation (20), we can capture the time pattern of the spread as well as the time patterns of the components of the spread in one single regression.

To incorporate the time dummy variables into the MHS model, we can substitute Equation (3) into Equation (20) and control the time dummy variables in the intercept term, so that the new model is now given by:

$$\begin{aligned}\Delta s_t = & \tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \cdot BS_t + \tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \\ & \cdot(\alpha + k_1\beta - 1) \cdot BS_{t-1} - \tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \\ & \cdot \left[\alpha(1-2\theta) + \beta\left(1 - \tfrac{k_1 k_2}{N}\right)\right] \cdot BS_{t-2} \\ & + \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t\end{aligned} \tag{22}$$

Re-arranging the equation, we have,

$$\begin{aligned}\Delta s_t = & \tfrac{1}{2}[\tau_1 \cdot BS_t + \tau_1 \cdot (\alpha + k_1\beta - 1) \cdot BS_{t-1} - \tau_1 \cdot \left[\alpha(1-2\theta) + \beta\left(1 - \tfrac{k_1 k_2}{N}\right)\right] \cdot BS_{t-2}] \\ & + \tfrac{1}{2} \cdot \Sigma[\tau_i \cdot timedummy_i \cdot BS_t + \tau_i \cdot (\alpha + k_1\beta - 1) \cdot timedummy_i \cdot BS_{t-1} - \tau_i \\ & \cdot \left[\alpha(1-2\theta) + \beta\left(1 - \tfrac{k_1 k_2}{N}\right)\right] \cdot timedummy_i \cdot BS_{t-2}] + \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t \\ = & \tfrac{1}{2}[\tau_1 \cdot BS_t + \Phi_1 \cdot BS_{t-1} - \Lambda_1 \cdot BS_{t-2}] + \tfrac{1}{2} \cdot \Sigma[\tau_i \cdot timedummy_i \\ & \cdot BS_t + \Phi_i \cdot timedummy_i \cdot BS_{t-1} - \Lambda_i \cdot timedummy_i \cdot BS_{t-2}] \\ & + \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t\end{aligned} \tag{23}$$

where $\Phi_i = \tau_i \cdot (\alpha + k_1\beta - 1)$ and $\Lambda_i = \tau_i \cdot \left[\alpha(1-2\theta) + \beta\left(1 - \frac{k_1 k_2}{N}\right)\right]$, then we can obtain all the parameters.

There are three groups of time dummies: (1) for studying the intra-hour pattern, we use 11 minute-dummies which represent 12 five-minute intervals in each hour (FIVE1 = 1 if the quote is in the first five minutes in the hour, in other words within the interval :00-:05); (2) for studying the intra-day pattern, we use 23 hour-dummies which represent 24 hours in each day (H1 = 1 if the quote is in the interval 1:00-2:00); and (3) for studying the week pattern, we use four day-name-dummies which represent five working days in a week (D1 = 1 if the quote occurs on Monday). There are three regressions with different groups of time dummies.

For brevity, we do not show the results of the regressions directly. Instead, Tables (3), (4) and (5) show coefficients of trade indicators (BS) and decomposition results at each time interval, which are calculated using the regression results. Because the components of the spread are closer to the LIMHS model, we only report the decomposition results of the LIMHS model.

Table (3) shows the intra-hour pattern of the spread. We use 11 minute-dummies in regression (23). According to the regression, overall, spreads are not significantly different among the five-minute intervals in an hour, however, spreads in the first five minutes are significantly lower than at other times. The decomposition results are shown in the last three columns of the tables. Asymmetric information costs are the dominant source of spreads on the D2000-1 trading system, and the weights of the components of the spread do not change significantly over time. In most cases, the shares of the asymmetric information costs are between 70% and 90%, which coincides with the results in section (3). In several time intervals (:10-:14, :25-:33, :55-:59), the share of asymmetric information costs is more than 90% and the share of inventory control costs is negative, which suggests that the inventory control costs are used to compensate for the asymmetric information costs. The official economic data and policy are usually released at :00, :15, :30, and, therefore, it is reasonable to expect that the share of asymmetric information costs is higher at those times.

Table (4) shows the intra-day pattern of the spread. We use 23 hour-dummies in regression (23). These intra-day spreads are also highly volatile and do not follow the smooth reverse J-curve found in the NYSE (McInish and Wood 1992). Similar to the finding in Bollerslev and Domowitz (1993), who use USD/DEM data collected from "Reuters' network screens", the spreads at 5:00, which is lunchtime in the Japanese market, are much higher than those at other hours. The trading hours of the Tokyo market are 1:00 to 9:00, so the spread has an inverse U-shape in the Tokyo market, which operates during the off-peak trading hours of the USD/DEM pair. Hua and Li (2011) also find that spreads of JPY/USD have an inverse U-shape in their off-peak trading hours, which are during the London market. During heavy trading hours, spreads are stable. After the closing of the London market at 17:00, spreads become larger. Contrary to previous findings, spreads during Asian trading hours are much lower than during the hours of other markets except for the peak around 5:00. The decomposition results are shown in the last three columns in the tables. Similar to the regression which uses five-minute dummies, asymmetric information costs are the dominant source of spreads, and the weights of the components of the spread do not significantly change over time. In peak-trading hours, the share of the asymmetric information costs is between 70% and 90%, which coincides with the results in section (3). In some time interval (1:00-3:00, 17:00-18:00, 21:00-22:00, 23:00-24:00) the share of order processing costs is negative (the coefficient of BS_{t-1} is positive). This is because the coefficient is not significantly different from zero, and the coefficient is not stable. In the off-peak trading hours, the share of asymmetric information costs are much greater than peak trading hours. In off-peak trading hours, end-users may have a greater share of fundamental information, because at this time end-users' of curries is for trading with foreigners or hedging their foreign exchange risk rather than speculation. At this time, order flows in the interdealer market may include more fundamental information from end-users, and thus, dealers may face a greater asymmetric information cost.

Table (5) shows the week pattern of the spread. We use four day-name-dummies in regression (23). Though the coefficients on the day name dummies suggest that spreads are slightly higher on Friday than on other days, only the coefficient on the Thursday dummy is statistically significant at the 10% level. Unlike the EBS data in Ito and Hashimoto (2006), these data do not exhibit a U-shaped intra-day spread pattern. The shares of the components of the spread are not different on different days. The share of asymmetric information cost on Thursday is slightly higher than other days.

Figure (1) shows the pattern of spreads of the USD/DEM pair for the whole week. There are no significant differences across the trading days in a week. The spreads have an inverse U-shape pattern during the off-peak trading hours of the USD/DEM pair. Spreads during European trading hours are lower than those after the closing of the European markets.

Table 3: Coefficient the Spread Against the Five-minute Dummies USD/DEM

Minutes	$BS_t \times 10^{-5}$	$BS_{t-1} \times 10^{-5}$	$BS_{t-2} \times 10^{-5}$	θ	other	IC	AS	Fraction
0 - 4	3.34	-0.274	-0.471	0.471	8.2 %	9.23 %	82.57 %	8.38 %
5 - 9	3.619	-0.907	-0.22	0.47	25.06 %	1.64 %	73.3 %	8.62 %
10 - 14	4.051	-0.307	-0.115	0.462	7.58 %	-4.47 %	96.89 %	8.20 %
15 - 19	4.511	-0.496	-0.392	0.461	10.99 %	1.96 %	87.05 %	8.23 %
20 - 24	4.161	-0.622	-0.249	0.466	14.95 %	0.2 %	84.85 %	8.13 %
25 - 29	4.113	-0.718	0.044	0.465	17.46 %	-7.39 %	89.93 %	8.14 %
30 - 34	3.761	-0.885	0.349	0.465	23.53 %	-15.67 %	92.14 %	8.21 %
35 - 39	4.571	-0.63	-0.351	0.466	13.78 %	2.02 %	84.19 %	8.81 %
40 - 44	3.449	-0.418	-0.297	0.463	12.12 %	2.2 %	85.68 %	8.71 %
45 - 49	4.191	-0.189	-0.562	0.464	4.51 %	6.95 %	88.54 %	8.37 %
50 - 54	3.96	-1.016	-0.025	0.463	25.66 %	-5.31 %	79.65 %	8.09 %
55 - 59	3.88	-0.422	0.0923	0.46	10.88 %	-10.34 %	99.46 %	8.09 %
F-value	2.1*	1.05	0.95					

Tick-by-tick USD/DEM transaction data from 1996.5.1 to 1996.9.2 on the Reuters D2000-1 system are used.

The results are obtained by the MHSD model (Equation 23) with 11 five-minute-dummies.

BS_t is the trade direction indicator in period t which is 1 if there is a buy order and is -1 if there is a sell order.

OP is the weight of the order processing cost on the spread ($1 - \alpha - \beta$ in the LIMHS model).

IC is the weight of the inventory control cost on the spread (β in the LIMHS model).

AS is the weight of the asymmetric information cost on the spread.(α in the LIMHS model).

θ is probability of order direction reversal, and is calcualted from the results of regression (12)

Fraction is the fraction of total number of trades in relevant intervals

F-value is the results of the F-test which the null-hypothesis is all the dummies corresponding to the variable are zeros.

* Significant at 5% level

Table 4: Coefficient the Spread Against the Hour Dummies USD/DEM

Hours	$BS_t \times 10^{-5}$	$BS_{t-1} \times 10^{-5}$	$BS_{t-2} \times 10^{-5}$	θ	other	IC	AS	Fraction
0 - 1	2.94	-0.167	-0.0823	0.463	5.68 %	-4.5 %	98.82 %	0.50%
1 - 2	1.181	0.147	0.0224	0.417	-12.45 %	-24.72 %	137.17 %	0.72%
2 - 3	0.905	0.552	-0.541	0.431	-60.99 %	43.64 %	117.35 %	1.89%
3 - 4	4.44	-1.757	-0.0662	0.449	39.57 %	-5.27 %	65.7 %	2.04%
4 - 5	7.98	-1.317	1.496	0.448	16.5 %	-30.63 %	114.12 %	1.60%
5 - 6	1.751	-0.867	0.514	0.417	49.51 %	-45.22 %	95.71 %	0.62%
6 - 7	3.92	-1.067	-0.119	0.45	27.22 %	-4.7 %	77.48 %	1.06%
7 - 8	3.85	-0.877	0.115	0.463	22.78 %	-9.34 %	86.56 %	3.04%
8 - 9	3.86	-0.557	-0.0668	0.472	14.43 %	-3.17 %	88.74 %	6.49%
9 - 10	3.83	-0.567	0.0147	0.468	14.8 %	-6.18 %	91.37 %	9.52%
10 - 11	3.86	-0.353	-0.423	0.476	9.15 %	7 %	83.85 %	9.45%
11 - 12	4.16	-0.301	-0.39	0.47	7.24 %	4.06 %	88.7 %	7.44%
12 - 13	4.2	-0.617	0.0698	0.467	14.69 %	-7.89 %	93.2 %	6.36%
13 - 14	4.16	-0.657	-0.236	0.471	15.79 %	0.77 %	83.44 %	7.20%
14 - 15	3.92	-0.567	-0.108	0.468	14.46 %	-2.91 %	88.45 %	9.79%
15 - 16	3.62	-0.627	-0.12	0.468	17.32 %	-2.18 %	84.85 %	9.84%
16 - 17	5.36	-0.907	-0.0649	0.463	16.92 %	-5.32 %	88.4 %	8.80%
17 - 18	5.02	0.253	-1.524	0.46	-5.04 %	23.94 %	81.1 %	5.87%
18 - 19	5.21	-0.937	-1.264	0.446	17.98 %	17.31 %	64.7 %	3.06%
19 - 20	4.48	-0.365	-1.194	0.446	8.15 %	18.81 %	73.05 %	1.61%
20 - 21	4.04	-0.547	1.186	0.473	13.54 %	-35.94 %	122.4 %	1.19%
21 - 22	4.06	0.174	-0.0711	0.474	-4.29 %	-3.93 %	108.22 %	0.75%
22 - 23	4.77	-0.947	-1.524	0.469	19.85 %	28.8 %	51.35 %	0.61%
23 - 24	1.59	0.553	-0.0642	0.458	-34.78 %	-7.84 %	142.62 %	0.55%
F-value	5.48***	1.04	1.22					

Tick-by-tick USD/DEM transaction data from 1996.5.1 to 1996.9.2 on the Reuters D2000-1 system are used.
The results are obtained by the MHSD model (Equation 23) with 23 hour-dummies.
Settings in this table are the same as in Table (3)
*** Significant at 0.1% level

Table 5: Coefficient the Spread Against the Day Dummies USD/DEM

Days	$BS_t \times 10^{-5}$	$BS_{t-1} \times 10^{-5}$	$BS_{t-2} \times 10^{-5}$	θ	other	IC	AS	Fraction
Monday	3.92	-0.437	-0.23	0.467	11.15 %	-0.02 %	88.87 %	15.90 %
Tuesday	4	-0.475	-0.22	0.469	11.88 %	0.09 %	88.04 %	21.47 %
Wedsday	3.89	-0.644	-0.14	0.465	16.56 %	-2.44 %	85.88 %	21.34 %
Thursday	3.82	-0.669	0	0.458	17.51 %	-7.53 %	90.02 %	21.11 %
Friday	4.22	-0.606	-0.35	0.465	14.36 %	2.44 %	83.2 %	20.18 %
F-value	0.85	0.35	0.61					

Tick-by-tick USD/DEM transaction data from 1996.5.1 to 1996.9.2 on the Reuters D2000-1 system are used.
The results are obtained by the MHSD model (Equation 23) with 4 day-dummies.
Settings in this table are the same as Table (3)

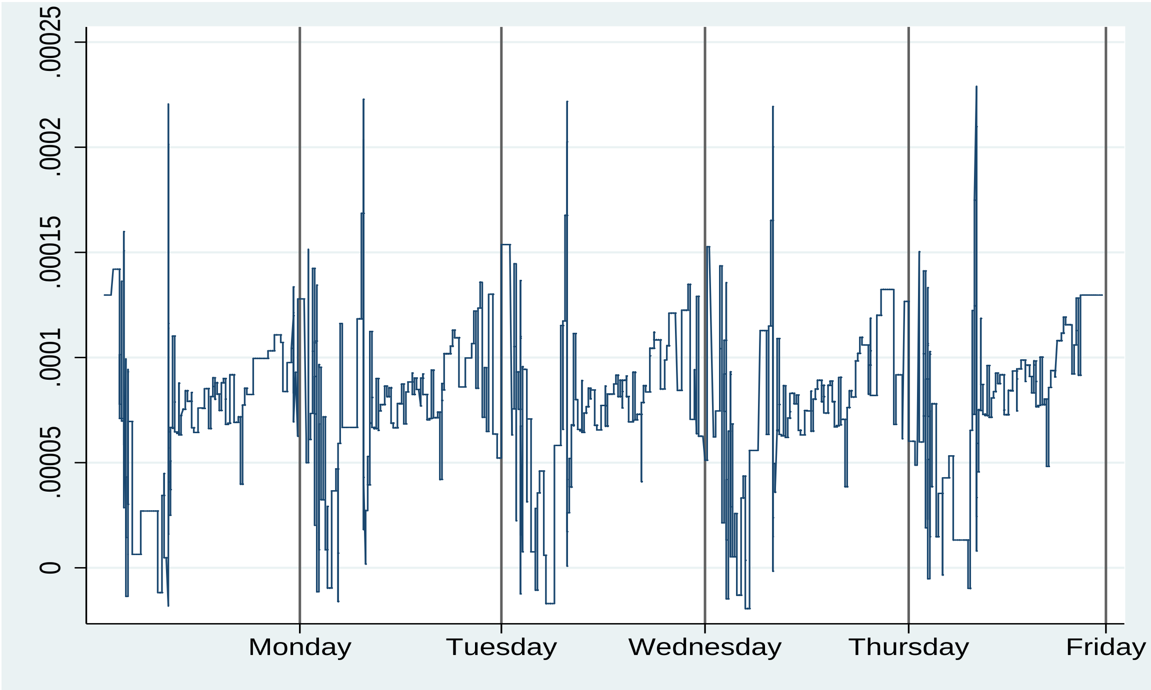

Figure 1: Week Spread USDDEM

This figure shows the time pattern of the spread in a whole week.
The spread is estimated by the MHSD model (Equation 23) with 11 five-minute-dummies, 23 hour-dummies and 4 day-dummies.

6 Conclusion

The spread-estimating and decomposition model of Huang and Stoll (1997) was designed with a mono-dealer market such as the New York Stock Exchange in mind. The model assumes that dealers' inventory imbalances can only be resolved through price adjustment. In a multi-dealer market this assumption is true of aggregate inventories, but individual dealers can rebalance their inventories quickly by trading with other dealers. Two assumptions of the HS model therefore break down in multi-dealer markets: trade direction tends to be positively rather than negatively serially correlated, because of this "hot-potato" trading, and aggregate inventory imbalances are not in general equal to the signed sum of all previous trades.

We have modified the HS model to take account of these features of multi-dealer markets. The original HS model is a special case of this modified model. The estimating equation is the same as in the HS model, but its interpretation is different. Our model tends to assign a higher proportion of the spread to the asymmetric information part of the spread, at the expense of the inventory control component. Applying the new model to inter-dealer transaction data of the USD/DEM pair on the Reuters D2000-1 system, we have found that the asymmetric information part is the dominant component of the spread and the weight of the inventory control part is very low.

Incorporating time dummies, the model can be used to analyse the intra-day pattern of the spread as well as the components of the spread. In the foreign exchange market spreads have a hill-shaped pattern during off-peak trading hours. Spreads are stable during peak trading hours and are significantly lower after the closing of the European markets. The asymmetric information and the inventory control components of spreads do not change significantly over time.

References

Bjønnes, G. H. and D. Rime (2005). Dealer behavior and trading systems in foreign exchange markets. *Journal of Financial Economics 75*(3), 571–605. 7, 12

Bollen, N. P. B., T. Smith, and R. E. Whaley (2004). Modeling the bid/ask spread: measuring the inventory-holding premium. *Journal of Financial Economics 72*(1), 97–141. 3

Bollerslev, T. and I. Domowitz (1993). Trading patterns and prices in the interbank foreign exchange market. *The Journal of Finance 48*(4), 1421–1443. 16

Chan, K. C., W. G. Christie, and P. H. Schultz (1995). Market structure and the intraday pattern of bid-ask spreads for nasdaq securities. *The Journal of Business 68*(1), 35–60. 2

Daníelsson, J. and R. Payne (2002). Real trading patterns and prices in spot foreign exchange markets. *Journal of International Money and Finance 21*(2), 203–222. 2

Evans, M. D. D. (1998). The microstructure of foreign exchange dynamics. 11

Evans, M. D. D. and R. K. Lyons (2002). Order flow and exchange rate dynamics. *The Journal of Political Economy 110*(1), 170–180. 5, 6, 7, 11

Goodhart, C., R. Love, R. Payne, and D. Rime (2002). Analysis of spreads in the dollar/euro and deutsche mark/dollar foreign exchange markets. *Economic Policy*. 11

Hua, M. and C.-Y. Li (2011, 2012/02/27). The intraday bid ask spread behaviour of the jpy/usd exchange rate in the ebs electronic brokerage system. *Applied Economics 43*(16), 2003–2013. 2, 16

Huang, R. D. and H. R. Stoll (1997). The components of the bid-ask spread: A general approach. *The Review of Financial Studies 10*(4), 995–1034. 0, 1, 21

Ito, T. and Y. Hashimoto (2006). Intra-day seasonality in activities of the foreign exchange markets evidence from the electronic broking system. 17

King, M. R., C. L. Osler, and D. Rime (2013). The market microstructure approach to foreign exchange: Looking back and looking forward. *Journal of International Money and Finance 38*(0), 95 – 119. 30th Anniversary of the Journal of International Money and Finance. 5, 7

Lyons, R. K. (1995). Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics 39*(2-3), 321–351. 2, 11

Lyons, R. K. (1997). A simultaneous trade model of the foreign exchange hot potato. *Journal of International Economics 42*(3-4), 275–298. 7

McGroarty, F., O. a. Gwilym, and S. Thomas (2007). The components of electronic inter-dealer spot fx bid-ask spreads. *Journal of Business Finance& Accounting 34*, 1635–1650. 1

McInish, T. H. and R. A. Wood (1992). An analysis of intraday patterns in bid/ask spreads for nyse stocks. *The Journal of Finance 47*(2), 753–764. 2, 16

Osler, C. L., A. Mende, and L. Menkhoff (2011). Price discovery in currency markets. *Journal of International Money and Finance 30*(8), 1696 – 1718. 6, 7

Viswanathan, S. and J. J. D. Wang (2004). Inter- dealer trading in financial markets. *The Journal of Business 77*(4), 987–1040. 1

Appendix

Assume there are two dealers, a and b, in the market. After receiving the customers' orders, dealer a's inventory imbalance is given by I_a and b's is given by I_b. The total inventory imbalance is $I_a + I_b$ and $0.5(I_a + I_b)$ on average. Now dealer a and dealer b trade with each other to get rid of their imbalance. Suppose, at the first stage, both of them choose a proportion v of their inventory imbalance to trade. Then dealer a's inventory imbalance is given by $(1-v)I_a + vI_b$ Dealer b's imbalance is given by $(1-v)I_b + vI_a$ The aggregate trade volume is $v(I_a + I_b)$ Suppose, at the second stage, they choose a proportion v to trade. dealer a's inventory imbalance now is given by

$$(1-v)[(1-v)I_a + vI_b] + v[(1-v)I_b + vI_a] \tag{24}$$

dealer b's imbalance is given by

$$v[(1-v)I_a + vI_b] + (1-v)[(1-v)I_b + vI_a] \tag{25}$$

The aggregate trading volume is $2v(I_a + I_b)$ from stage one. The trading will be continuous until the imbalance is the same for each dealer, i.e. $0.5(Ia + Ib)$. Suppose there are n stages. Then the aggregate trading volume is $n \cdot v(I_a + I_b)$. Dealers' inventory imbalance at stage n is given by the following equations.

$$\begin{aligned} I_{a,n} &= \tfrac{[1-(1-2v)^{n-1}]I_b + [1+(1-2v)^{n-1}]I_a}{2} \\ I_{b,n} &= \tfrac{[1+(1-2v)^{n-1}]I_b + [1-(1-2v)^{n-1}]I_a}{2} \end{aligned} \tag{26}$$

When $v = 0.5$, the equilibrium is reached after one trade, and the aggregate trade volume is the average inventory imbalance. When v is different from 0.5, more trades are needed and the aggregate trade volume is greater than average inventory imbalance. If there are N traders in the market, the optimal strategy is that the v should be $\frac{N-1}{N}$, and a dealer should trade with all others with $\frac{1}{N}$ of his inventory imbalance.

The intuition is that the aggregate trade volume is determined by the choice of v and the trading dealers' strategies. The any strategy different from the optimal strategy will make the aggregate trade volume to be greater than the average inventory imbalance.