Essays on the Foreign Exchange Market: the Market Microstructure and Evidence of the Behavioural Theory

By Zhiyong Li

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

October 2015

Dedication

This thesis is dedicated to my parents, Fenghua Li and Ming Shang. Thanks for your unconditional love and support.

Acknowledgement

It is great to write down a few words in my thesis without tearing my hair out. Doing research is one of my favourites and writing up is the least. I am sure that my supervisors Professors Michael Bleaney and Spiros Bougheas are quite satisfied with my research and feel like to pull their hair out while reading my English.

My journey of pursuing a PhD was not very smooth. Many years ago, I was an innocent and naïve student who was passionate about research but did not really know what it is. I have overcome all the difficulties and unfairness and finally obtained my degree. I am well-trained researcher now. I am mature and still with passion. I was very lucky to know many brilliant people these years. We have numerous joyful moments together. They also help me with their gracious kindness. I would not be able to complete my PhD without them.

I would like to thank my supervisors Professors Michael Bleaney and Spiros Bougheas for their patient, insightful advice, invaluable guidance and the greatest support during these years. I could not imagine better supervisors.

I would like to thank my dearest friends who give me precious friendship and love, Drs Chou-Hsiung Chen, Hui Jiang, Hao Lan, Hock Ann Lee, Jianan Li, Jinke Li, Yao Liu, Hiroshi Ochiai, Lili Sheng, Fu-min Tseng, Xiaoyu Tian, Hao Wang, Zheng Wang, Fangya Xu, Feng Xu, Veronica Veleanu-Khosla, Jing You, Wei Yun and Ping Zhang. Miss Lu Dong, Mr Lingbo Huang, Miss Nianxin Le, Miss Lefan Liu, Miss Zhaohui Niu, Miss Shubin Yang, Mr Zhixian Yu.

Abstract

This thesis contributes to the literature in two areas: the empirical market microstructure and empirical behavioural finance. Data from the foreign exchange market are used in the thesis. In Chapter 2, we compare three bid-ask spread estimators using simulation experiments. In Chapter 3, based on the Huang and Stoll (1997) model (the HS model thereafter), we develop a new model which can be used to estimate and to decompose the spread. The new model is designed for the multi-dealer market. We then compare our new model and the HS model using the transaction tick-by-tick data of the DEM/USD pair from the Reuters D2000-1 system. In Chapter 4, using mid-price data estimated from transaction data, we investigate whether the professional traders in the foreign exchange market are suffering from gambler's fallacy and/or the hot hand fallacy.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis contributes to the literature in two areas: the empirical market microstructure and empirical behavioural finance. Data from the foreign exchange market are used in the thesis. In Chapter 2, we compare three bid-ask spread estimators using simulation experiments. In Chapter 3, based on the Huang and Stoll (1997) model (the HS model thereafter), we develop a new model which can be used to estimate and to decompose the spread. The new model is designed for the multi-dealer market. We then compare our new model and the HS model using the transaction tick-by-tick data of the DEM/USD pair from the Reuters D2000-1 system. In Chapter 4, using mid-price data estimated from transaction data, we investigate whether the professional traders in the foreign exchange market are suffering from gambler's fallacy and/or the hot hand fallacy. The rest of this chapter reviews the related background and literature.

1.1 The Bid-Ask Spread

"While much of economics abstracts from the mechanics of trading, microstructure theory focuses on how specific trading mechanisms affect the price formation process." (O'Hara 1995)

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).

In a quote-driven market, one or more dealers as liquidity providers (market makers) offer quotes to other participants in the market and promise to accept both buy and sell orders at the quotes. Dealers offer a bid price at which they accept all sell orders and an ask price at which they accept all buy orders. The ask price is usually higher than the bid price, so that dealers can cover their costs of market making by buying at low prices and selling at high prices. The difference between the ask price and the bid price is called the bid-ask spread. NASDAQ and the Reuters D2000-1 system are multi-dealer markets.

An order-driven market does not have a market maker as it is organised as a two-sided auction. Participants send their orders (market orders or limit orders) to the market. Limit orders are the orders executed when the market reaches a certain price. The best bid price is the highest bid price and the best ask price is the lowest ask price. Market orders are the orders that trade at the current best bid or ask prices. The bid-ask spread is the difference between the best ask price and the best bid price. An example of the order-driven market is the Electronic Brokerage System (EBS).

1.1.1 The Spread Keeps Prices Away from the Efficient Market Hypothesis

For scholars who focus on the market microstructure, the spread is one of the objects to be studied. For scholars who focus on other branches of economics, the spread is one of the frictions which keep prices away from the classic hypotheses (e.g. the efficient market hypothesis) and should be eliminated in data pre-treatment procedures.

In a friction-free weakly efficient market, the observed price process can be written as follows:

$$r_t = \Delta M_t = M_t - M_{t-1} = \varepsilon_t$$

M is the price and subscript t represents period t. Δ is the first order difference operator and r_t is the observed return at period t. ε_t is the price innovation which follows a random walk at period t. The return under this condition also follows a random walk.

The spread brings friction into the market. If there is the spread in the market, M is no longer observed. Instead, the observed price is given as follows:

$$s_t = M_t + \frac{1}{2} SP \cdot BS_t$$

where s is the observed price; BS is the trade indicator which is 1 if there is a buy order and is -1 if there is a sell order; SP is the bid-ask spread. The observed return is given as follows:

$$r_t = \Delta s_t = M_t - M_{t-1} + \frac{1}{2} SP \cdot BS_t - \frac{1}{2} SP \cdot BS_{t-1} \quad (1.1)$$

Although, the unobserved M series still follows the efficient market hypothesis, the observed return is now serial correlated because of the spread.

1.1.2 Determination of the Spread: Theories

Inventory Control Costs

Market makers adjust the mid-price to eliminate the unwanted inventory. Thus order flows will influence the mid-price (e.g. Stoll 1978, Amihud and Mendelson 1980, Ho and Stoll 1981, 1983).

In Stoll (1978), market making will introduce a cost to holding stocks at quantities that differ from the optimal portfolio to a dealer. Because of the uncertainty of the return of the stock, extra inventory brings more risk to the dealer. To protect himself, the dealer will offer a bid price which is lower than the true price of the stock. The difference between the bid price and the true price is positively correlated to the dealer's confidence of relative risk aversion, the size of unwanted inventory, and the volatility of the stock return. Symmetrically, the dealer will offer an ask price which is higher than the true price when he faces a buy order.

Asymmetric Information Costs

Order flows include fundamental information. The incoming order will change the market maker's knowledge about the fundamental and thus change the midpoint of the bid-ask prices (e.g. Glosten and Milgrom 1985, Kyle 1985, Easley and O'Hara 1988, Admati and Pfleiderer 1988).

In Glosten and Milgrom (1985), a dealer faces uncertainty about the true value of an asset which could be either v^{high} or v^{low} . There are uninformed and informed traders in the market. When the dealer's quote is higher than the true values, the informed trader will only sell the asset to the dealer and vice versa. Therefore, the dealer suffers a loss when he trades with informed traders. The dealer will use the spread to protect himself.

Furthermore, because a buy order suggests the quote is lower than the true values if it is from the informed trader, the dealer will rise both the bid and the ask prices to reflect the new information after receiving a buy order. The amount of the rising depends on the proportion of the informed traders in the market.

1.1.3 Determination of the Spread: Evidence

The theoretical components of the spread cannot be observed directly, so, researchers have to use proxy variables. Order size (OS) and the number of trades (NT) and the price volatility are proxy variables. In Bollen et al. (2004), OS is the proxy variable for order processing costs, and NT and volatility are the proxy variables for inventory control costs.

Evidence provided by Ding (2009) and by Bjønnes and Rime (2005) suggests that the order size and the spread are independent.

The weight of order processing in the spread represents the various costs of trading in the market, such as computer costs and the rent to use the trading floor, most of which are fixed in short term. Therefore, if OS and NT are large, the average order processing cost will be low. Thus, both OS and NT should have a negative relationship with the spread. McInish and Wood (1992) and Hua and Li (2011) find that OS and NT are negatively related to the spread. Evidence provided by Stoll (1978) suggests that order size and the spread are negatively correlated.

On the other hand, a large trading volume increases the difficulty of inventory management. In other words, considering inventory control costs, both OS and NT should have a positive relationship with the spread. Bollerslev and Domowitz (1993) find that although NT does not have a significant effect on the spread, it does increase the spread volatility. Lyons (1995) finds a positive relationship between the order size and the spread.

Bessembinder (1994) finds that the unforecast trade volume, which implies the existence of asymmetric information, has a positive effect on the spread in the foreign exchange market, although the effect is not very strong. Hartmann (1999) confirms Bessembinder's finding using the daily spread (from Reuters FXFX), the trade volume (from Nikkei Europe), and the USD/YEN rate (from Olsen & Ass. Zurich).

Exchange rate volatility is usually assumed to have a positive relationship with the spread as in McInish and Wood (1992) and Hua and Li (2011). Bollerslev and Melvin

(1994) and Goodhart and Payne (1996), using data from Reuters D2000-2 system, find that as the spread widens the volatility of price returns is higher.

1.1.4 Seasonality of the Spread

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. Unlike datasets from stock markets such as NYSE and NASDAQ, foreign exchange datasets provide information about inter-dealer transactions. Goodhart et al. (1996) and Daníelsson and Payne (2002) compare the pattern of the quote spread data from EFX and the transaction spreads data from the Reuters D2000-2 system. Daníelsson and Payne (2002), Ito and Hashimoto (2006b) and Hua and Li (2011) study the intro-day pattern of the spread on the Reuters D2000-2 system and EBS.

Chapter Three studies the intra-day patterns of the spread of EUR/USD and USD/DEM pairs on the Reuters D2000-1 system.

1.1.5 Estimators of the Spread

Spread data are not always available. Thus the spread estimator is required. The spread estimators can be divided into two groups: the Roll family estimators which base on the serial correlation of transaction returns and others. This model requires data for prices.

Holden (2009) and Goyenko et al. (2009) develop a spread estimator based on price clustering. The estimator assumes that “price clustering is completely determined by the spread size”. The spread is “a probability-weighted average of each possible spread size” divided by the average price. The estimator is called “effective tick”. Holden (2009) develops a spread estimator which is a hybrid of “effective tick” and the Huang and Stoll (1997) model.

Lesmond et al. (1999) (LOT) develop a spread estimator based on Kyle (1985) and Glosten and Milgrom (1985). The LOT estimator assumes that informed trading will move the price, and thus lead to a non-zero return, and that other trading will not move the price, and thus lead to a zero return. This model requires data for price returns.

In Corwin and Schultz (2012), the spread is estimated with daily high-low prices. Corwin and Schultz's model assumes that the price follows a random walk, and the highest price of a day is an ask-price (at which a trader makes a buy order to the market maker) and the lowest price of a day is a bid-price (at which a trader makes a sell order to the market maker). This model requires data for price returns and daily high-low prices.

Roll (1984) establishes a spread estimation model where the midpoint of bid and ask prices is assumed to follow a random walk. In other words, the Roll model considers only the order processing cost. The spread is estimated from the auto-covariance of price returns. Glosten and Harris (1988) introduce a spread estimation model where both the order processing cost and the adverse selection cost are considered. A buy (sell) will raise (decrease) the midpoint of bid and ask prices. For the model to work, price returns and trade indicators which represent the directions of trades are needed. Choi et al. (1988) interoperate the correlation of order flows into the Roll model. Stoll's (1989) model extends the Roll model by incorporating the probability of price reversal. Stoll's model also can be used to infer the components of the spread. George et al. (1991) relax Roll's assumption of random walk of mid-prices. When mid-prices contain positive correlation components, the Roll model will underestimate the true spread. Lyons (1995) develops a model to estimate the three components of the spread in the foreign exchange market. The paper considers the order processing cost, inventory control cost and adverse selection cost . A positive (negative) inventory shock leads to an increase (decrease) of the mid-price. For this model to work, price returns, trade indicators and the inventory level are required. Huang and Stoll (1997) develop a general model which incorporates previous spread estimation and decomposition models such as Roll's, Glosten and Harris's (1988), and Stoll's (1989) model. The Huang and Stoll model requires price returns and trade indicators. Hasbrouck (2004, 2009) extends The Roll model by using the Bayesian Gibbs sampler. Hasbrouck's model performs better than the Roll model.

There are many spread estimators. Which one should be used? The performance of spread estimators are compared in ap Gwilym and Thomas (2002), Anand and Karagozoglu (2006), Goyenko et al. (2009) and Corwin and Schultz (2012). These comparison papers normally use stock markets or future markets data. Under these circumstances, two issues may rise: first, estimators may have different performance when foreign exchange market data are used; second, one cannot isolate the specific influence of the factors such as non-random walk mid-price returns, and autocorrelated order flows on the performances of spread estimators by analysing the results obtained from real market data.

Chapter Two sheds light to these issues. By using simulation experiments, factors which may influence the spread estimation are considered separately. The results of chapter two show that in most cases Huang and Stoll's (1997) model (the HS model) can estimate the spread accurately when order flow data are available.

1.1.6 Decomposing the Spread

Besides estimating the spread, decomposing the spread is another topic. The spread reflects market makers' costs which include the order processing cost which is the labour cost, the cost of the trading floor etc; the adverse selection cost which is the loss when market makers trade with informed traders, and the inventory control cost which is the cost happens when market makers hold unwanted inventory. Though the HS model gives more accurate estimation of the spread than Roll's (1984) and Corwin's model and can be used to decompose the components of the spread, it is not widely used. The main reason is that the HS model relies on the assumption that the order flow should be negatively autocorrelated, and consequently the results of the model may suggest negative fractions of the components of the spread when the order flow is positively autocorrelated. The results may suggest a fraction which is larger than one when the order flow only exhibit weak autocorrelation. Furthermore, the HS model is designed for a single dealer market, which suggests that it may not be appropriate for a multi-dealer market.

Chapter Four presents the spread estimating and decomposing model based on the HS model. The modified HS model is designed for multi-dealer markets and considers placing a direct order as the inventory management method other than moving mid-prices.

1.2 FX Microstructure

Order flow, which is defined as the *net* of the buyer-initiated and seller-initiated orders, plays a key role in the microstructure analysis because it conveys information on determinants of exchange rates. Evans and Lyons (2002) establish the basic framework of the microstructure analysis on the foreign exchange market. In Evans and Lyons (2002) the exchange rate is determined by the dealers who cares about their own profits. Dealers share the information through the order flows in inter-dealer trades. The econometric model supposes the daily exchange rate return is influenced by changes in the interest rate differential and changes in signed order flow

$$\Delta P_t = \Delta r_t + \lambda x_t$$

where Δr_t is the first difference in the interest rate differential $(i_t - i_t^*) - (i_{t-1} - i_{t-1}^*)$. Δx_t is the difference between the number of buyer-initiated trades and seller initiated trades in day t. Δr_t is observed publicly each day before trading, while Δx_t is not observable for customers. A positive value for Δx_t implies that within day t the number of buy orders exceed that of sell ones. This means that a majority of traders has purchased the foreign currency during the day indicating that they consider the foreign currency undervalued.

1.3 Gambler's Fallacy and Hot Hand Fallacy

Both gambler's fallacy (GF) and hot hand fallacy (HHF) are biases when agents face a random series. An agent who suffers from the GF will expect a trend to break in the next period. An agent who suffers from the HHF will expect a trend to be continuous in the next period.

The evidence of the GF and the HHF is normally from laboratory experiments and gambling or lottery market. There is no direct evidence from financial markets. Furthermore, literature normally focuses on the binary outcomes (head or tail; score or miss), but, it is more complicated in financial markets. Traders not only care whether the price goes up or down, but also care about the magnitude of the returns. It seems that both the fallacies are caused by lack of knowledge of probability theory. The experienced agents (e.g. professional traders) may not have the fallacies. Do the professionals in financial markets suffer from the fallacies? Is the effect of a 1% appreciation different from that of a 2% appreciation on traders' behaviour? These are interesting questions to be answered.

Chapter Five extends the binary outcomes literature, where the price trend properties are described by two factors: a) the streak length which is the number of periods that the price continuous to move in the same direction and is similar to the setting in binary-outcome cases; (b) the streak width which is the distance between the current level and the point where the trend starts. We find that the concepts of the GF and the HHF are helpful to explain the pattern of the traders' reactions to the trend of prices.

Tick-by-tick transaction data, which include the data of transaction prices (the uncertain series) and the order flow (traders' behaviour), from Reuters D2000-1 system are used. According to Chapter Two, because the HS model performs best among the spread estimators when both transaction prices and order flows are available, it is used to obtained mid-prices from transaction prices in pre-treatment process.

1.3.1 Heterogeneous Agents Model

The heterogeneous agents model (HAM) explains the asset price as the result of the interaction between various groups of agents, where each group holds different beliefs from others. In financial markets, the HAM can be used to describe the interaction between rational traders and irrational traders who, for example, suffer from the gambler's fallacy or the hot hand fallacy. Frankel and Froot (1986, 1990, 1991) develop a heterogeneous agent model on exchange rate determination. The model assumes that there are three types of agents in the market, chartists, fundamentalists and portfolio managers. Chartists use extrapolation, the information set that they use is the time series data of the exchange rate. In other words, they choose the autoregressive model. If the underlying shocks are random, chartists suffer from the GF or the HHF. Fundamentalists expect the foreign exchange rate to be based on the fundamentals and economic factors. Portfolio managers decide the weight of the expectation of chartists and fundamentalists. Finally, the exchange rate is determined by the decision of portfolio managers. De Grauwe and Grimaldi (2005, 2006) extend earlier works and develop a new heterogeneous agents model of the exchange rate, where only fundamentalists and chartists are in the market. The weights of the two groups change according to previous profits. Their model gives a reasonable explanation of some of the exchange rate puzzles, such as the “excess volatility” puzzle (Baxter and Stockman 1989 and Flood and Rose 1995), and the “disconnect puzzle” (Goodhart and Figliuoli 1991).

1.3.2 The BSV Model

Barberis et al. (1998) develop a similar model based on the conservatism and the representativeness heuristics. Conservatism means individuals update their mind slowly facing new information. Agents believe that the random series is not really “random”, but follows two regimes: trending and reversing. The prediction is a weighted average of the predictions from the two regimes. In contrast to the probability of reversing in the RV model, the weights of regimes are monotonic responses to the feedback from the market. The model suggests individuals underreact to short term shocks and overreact to long term trends. One may notice that the mean-reversing regime can be considered as the GF effect and the trending regime can be explained as the HHF effect.

1.3.3 The RV Model

Rabin and Vayanos (2010) propose “a model to examine the link between the gambler’s fallacy and the hot-hand fallacy”. They argue that these “two fallacies might be related, with the hot-hand fallacy arising as a consequence of the gambler’s fallacy”. In the RV model, agents can observe a signal which is generated by a mean-reversing process and an i.i.d shock. The parameters of the mean-reversing process are not known by the agents. Agents also believe that the shock is negative autocorrelated. Therefore agents use the weighted average of past shocks to predict the random shock and learn the parameters according to the signal. The prediction process represents the agents’ GF effect and the learning process represents agents’ HHF effect. The prediction of the agents is a combination of these two effects. A key implication of the model is that the reversing probability which agents believe is not monotonic in the streak length due to the different converge speed between the GF effect and the HHF effect. The GF effect gets stronger in the first several periods of the streak and then HHF becomes more dominant as the streak is getting longer.

1.3.4 Empirical Evidence

Laboratory Evidence

Laboratory experiments (Ayton and Fischer 2004, Huber et al. 2008) support the explanation that the GF effect happens when individuals face the observable outcomes of a series (for example, heads or tails of the coin), while the HHF effect happens when individuals predict human performance (for example, wins or losses in gambles).

Rao (2009) designs an experiment to study the relation between GF and HHF. In the experiment, the subjects were asked to watch the NBA games and to predict the results of shooting. The experiment suggests that the GF normally happens after short streaks and HHF happens after long streaks. The transition point from GF to HHF is also found.

Field Evidence

According to Gilovich et al. (1985), though the evidence does not support the existence of "the hot hand", basketball players and fans believe a player who continuously scored recently has high probability to score next time.

Clotfelter and Cook (1991) find that lottery players exhibit GF in a numbers game: numbers which are drawn will become less popular among the players remarkably and quickly.

Sundali and Croson (2006), and Croson and Sundali (2005) observe and identify the GF and the HHF "within a given individual" by using data collected from the casino.

Camerer (1989) uses the data from the basketball betting market to study the HHF. Gamblers in the basketball betting market bet on the point spreads which are set by bookmakers. Considering their own benefits, bookmakers should set the point spreads which relate to the "dollar-weighted" position of the bettors. If bettors suffer from the HHF, the point spreads in the betting market should be larger (smaller) than those actually happen when one of the teams in the match is in the winning (losing) streak. Camerer (1989) shows that the difference between the market point spreads and those in the match are positive and statistically significant but not big enough for one to have extra profits on it. Following Camerer (1989), Brown and Sauer (1993) take the team's ability into account. Their results reject the hypotheses that "the hot hand is irrelevant to both the betting market and game outcomes" but do not reject the hypotheses that "people believe in a mythical hot-hand effect" and that "point spreads are rational expectations and real hot-hand effects exist".

The gamblers believe that there is more probability of wining if they buy a ticket from the store which sells a winning ticket recently. This phenomenon can be called the "lucky store" effect. Guryan and Kearney (2008) approve the "lucky store" effect by studying the data on "large lotto games in the Texas Lottery system". The winning ticket does bring the increase of sales to the store which sold it and the stores near it, but the increase in the winning store is significantly larger than others. The effect can last for a long time (up to 40 weeks) and is stronger in the area where economically poorer populations live. Guryan and Kearney (2008) argue that the "lucky store" effect challenges the representativeness explanation to the GF and HHF at two points. In the case of "lucky store", firstly, there is no transformation from GF to HHF; secondly, it does not rely on the streak which is necessary for the representativeness explanation. This challenge can be solved if the streak length and width are distinguished. Since the profits of the winning lottery are very large (from $8,888 to $51200200 in the dataset), the shock is strong enough to make the gambler have a HHF behaviour. If the authors consider the small size of profits, the role of streak may be found.

Chapter 2

Comparing the bid-ask spread estimators

2.1 Introduction

This chapter compares the performance of various spread estimators. The bid-ask spread, which is the difference between the ask and the bid prices, is of interest for several reasons: first, the spread is a useful measure of market participants' trading costs and thus a widely used proxy for market liquidity (e.g. Mancini et al. 2013 and Banti et al. 2012); second, the spread is a type of microstructure friction that keeps the observed price series away from what would be obtained under efficient market hypothesis (e.g. that prices follow a random walk); third, the spread can influence measures of market volatility (e.g. Bandi and Russell 2006).

There are several types of spreads, namely, the quoted spread, the effective spread and the realised spread. In this chapter, we focus on the effective spread. The quoted spread is the difference between the bid and the ask prices quoted by the market maker at any instant. The effective spread is equal to twice the difference between the actual executed price and the mid-price. Normally, the effective spread is less than the quoted spread because traders can sometimes obtain tighter spreads that are not recorded by quoted data. The effective spread and the quoted spread are equal only if all the transactions take place at the exact ask or bid price.

Because spread data are not always available, many indirect methods have been suggested for estimating the spread. Some estimators are designed for specific market structures (e.g. Huang and Stoll 1997 propose an estimator designed for single dealer markets; McGroarty et al. 2006 propose an estimator designed for order-driven markets). Some estimators are designed to consider data availability (e.g. Huang and Stoll 1997 can be used when both transaction prices and order flow are available; Roll 1984 requires transaction prices only; Corwin and Schultz 2012 requires daily high and low prices). Some estimators reflect various components of the spread (Roll 1984 can only be used to estimate the order processing cost ; Huang and Stoll 1997 reflects three components of the spread). Some estimators are designed according to the sampling frequency of data (Corwin and Schultz 2012 is designed for low frequency data; Roll 1984 is a high frequency estimator).The existence of many spread estimators raises the question of which one is most appropriate. This chapter assesses the performances of several popular spread estimators over different time intervals and under different conditions.

In contrast to previous literature such as ap Gwilym and Thomas (2002), Anand and Karagozoglu (2006) and Holden (2009), which estimates the spread using real market data, we discuss factors that might bias estimations using simulated data.

Though tick-by-tick data record more market information than lower frequency data, they are not always available. Therefore researchers are searching for methods to estimate the spread from low frequency data (daily or less often). Goyenko et al. (2009) Holden (2009) compare dozens of estimators by using low-frequency data (daily data) from the stock market. It is interesting to assess the performance of estimators using high-frequency data, because these estimators are supposed to be unaffected by the frequency of the data. In the appendix of Corwin and Schultz (2012), the high-low estimator is tested with high-frequency data (15-minute). This chapter evaluates the estimators over a wide range of frequencies (from tick-by-tick to daily data).

The comparison is based on the following criteria: differences between the estimate and the benchmark and the standard deviation of estimates. The root mean square error (RMSE), i.e., the standard deviation of the spread estimates about the true value (rather than the estimated mean) can also be a useful measure. The RMSE captures both the bias in the estimate and its variability. Better performing estimators are those with less differences and standard deviations, and thus small RMSEs.

The rest of the chapter is organized as follows. In Section (2.2), the relevant literature is briefly reviewed. In Section (2.3), the motivation behind the estimator selection is discussed. In Section (2.4), we discuss theoretical issues related to the estimators and their errors. In Section (2.5), the design of the simulation experiments is introduced. In Section (2.6), the estimators and their errors are compared based on the results of the simulation experiments according to the procedure described in Section (2.5).In Section (2.7), we use estimate the spread of transaction data for Deutsche-Mark against US dollar transactions using the estimators.

2.2 Related Literature

The literature on the spread normally focuses on four aspects: first, the factors that determine the spread (e.g. Bessembinder 1994 and Lyons 1995 study the effects of trade volume and inventory control cost on the spread. Bollerslev and Melvin 1994 and Goodhart and Payne 1996, using data from the Reuters D2000-2 system, find that as the spread widens, the volatility of price returns increases); second, the methods that decompose the spread (e.g. McGroarty et al. 2007 derive a spread decomposing model specifically to the inter-dealer foreign exchange market, EBS); third, patterns of the spread over time; (e.g. Goodhart et al. 1996 and Daníelsson and Payne 2002 compare the pattern of quote spread data from EFX and transaction spreads data from the Reuters D2000-2 system. Daníelsson and Payne 2002, Ito and Hashimoto 2006b and Hua and Li 2011 examine intro-day pattern of the spread on the Reuters D2000-2 system or EBS); and fourth, the methods that estimate the spread. When the spread is not available, the fourth aspect is the foundation of the other three.

Stoll (1978) mentions three components of the spread that correspond to the ways in which the market maker uses the spread to cover three types of costs consisting of (a) the order processing cost such as the commission fee and the cost of the trading floor, (b) the inventory control cost such as the opportunity cost of holding inventory and the cost of inventory imbalance, which Bessembinder (1994) and Lyons (1995) suggest influences the spread along with its proxies in the foreign exchange market, and (c) the asymmetric information cost such as the loss when the market maker trades with an informed trader. Order flow includes fundamental information. Market makers adjust the mid-price to eliminate any unwanted inventory. Thus, the cumulated order flow influences the mid-price (e.g. Stoll 1978, Amihud and Mendelson 1980, Ho and Stoll 1981). An incoming order will change the market maker's knowledge about the fundamental and thus alter the midpoint of the bid-ask prices (e.g. Glosten and Milgrom 1985, Kyle 1985). Bessembinder (1994) finds that the unforecasted trade volume, which implies the existence of asymmetric information, has a positive, but weak, effect on the spread in the foreign exchange market. Hartmann (1999) confirm Bessembinder's finding using the daily spread (from Reuters FXFX), trade volume (from Nikkei Europe), and the USD/YEN exchange rate (from Olsen & Ass. Zurich). When no inventory control costs or asymmetric information costs exist in the market, incoming orders will not change the mid-price of the bid and the ask prices.

Roll (1984) presents a spread estimation model in which the midpoint of the bid and the ask prices is assumed to follow a random walk. That is, the Roll model considers only the order processing cost. The spread is estimated from the auto-covariance of price returns. Glosten and Harris (1988) introduce a spread estimation model in which both the order processing cost and the adverse selection cost are considered. A buy (sell) will raise (decrease) the midpoint of bid and ask prices. For the model to work, price returns and trade indicators representing the directions of trades are necessary. Choi et al. (1988) interoperate the correlation of order flow into the Roll model. Stoll's (1989) model extends the Roll model by incorporating the probability of price reversal. Stoll's model also can be used to infer the components of the spread.George et al. (1991) relax Roll's assumption that the mid-prices follow a random walk. When mid-prices contain positive correlation components, the Roll model underestimates the true spread.Lyons (1995) develops a model to estimate the three components of the spread in the foreign exchange market. The Lyons paper considers order processing costs, inventory control costs and adverse selection costs. A positive (negative) inventory shock leads to an increase (decrease) in the mid-price. For this model to work, price returns, trade indicators and the inventory levels are required. Huang and Stoll (1997) develop a general model that incorporates previous spread estimate and decomposition models such as Roll's, Glosten and Harris's (1988), and Stoll's (1989) models. Huang and Stoll's model requires price returns and trade indicators. Hasbrouck (2004, 2009) extends Roll's model by using the Bayesian Gibbs sampler. Hasbrouck's model outperforms the Roll model. In Corwin and Schultz (2012), the spread is estimated with daily high-low prices. Corwin and Schultz's model assumes that the price follows a random walk, and the highest price of a day is an ask-price (at which a trader places a buy order to the market maker) and the lowest price of a day is a bid-price (at which a trader places a sell order to the market maker). This model requires data for price returns and daily high-low prices.

Generally, two types of methods can be used to estimate the bid-ask spread in the market. The first type estimates the spread using the series properties of the prices as in Roll (1984), Stoll (1989), Hasbrouck (2004, 2009) and Corwin and Schultz (2012). Methods based on series properties possess several advantages including easily obtainable data of which only price data are required; compatibility with data of various frequencies, and acceptable errors of the estimators.

The other type of methods uses both the price and trading direction to estimate the spreads. Trading direction data allow researchers to estimate the spread and its theoretical components simultaneously (e.g. Glosten and Harris 1988, Lyons 1995 and Huang and Stoll 1997). Because these models require more information, they are expected to produce more accurate estimates than models that employ only price returns.

2.3 The Choice of Estimators

In this chapter, we compare three spread estimators: the Roll estimator which is the most widely used, the Huang and Stoll (HS) estimator, which is the most general model, and the Corwin and Schultz (CS) estimator which is one of more recent developments not requiring intensive computation. The family of Roll estimators (e.g. Roll 1984, Choi et al. 1988, Stoll 1989, George et al. 1991 and Hasbrouck 2004, 2009) is the most important set of spread estimators. Because all these estimators are based on Roll (1984), factors that influence the Roll estimator may also affect the estimators as we discuss below. Thus, this chapter focuses on the performance of the original Roll estimator and discusses some extensions of this estimator. The HS estimator introduces a general framework of interactions among order flow, spreads, mid-prices and transaction prices, and other estimators (e.g. Roll 1984, Choi et al. 1988, Stoll 1989) can be considered special cases. If the estimator performs well, only the HS estimator is needed; therefore, it is helpful to evaluate this estimator's performance. Although the HS estimator is comprehensive, the data on order flow required by the estimator are not always available. Like the Roll estimator, the

CS estimator requires a relatively low level of data; only the high-low ratio for which even historical data are available is needed. The calculation of the CS estimator is also very simple and can be applied to address a considerable volume of data easily. Because the CS estimator is very new, it has been applied rarely. This chapter provides a first attempt to examine the performance of this estimator in detail.

These estimators are developed to estimate the true spread accurately; therefore, it is important to evaluate their performance. A good estimator should be able to identify the true spread. Theoretically, the estimate should converge to the true spread; if the estimate does not converge, the error should be predictable (i.e., the direction, magnitude, and other characteristics of the errors should be known). Additionally, the difference between the estimate and the true spread should be small. The estimator that produces the smallest errors is the best one. A good estimator should also produce stable estimates. In other words, the variance of the error should be small to increase the confidence of each estimate. Furthermore, given that the properties of real data often diverge from ideal conditions, a good estimator should perform acceptably when assumptions are not satisfied. Additionally, a good estimator should not be computing-intensive.

Because spread estimators are designed for the stock market, they are normally tested with stock market data. ap Gwilym and Thomas (2002) and Anand and Karagozoglu (2006) compare estimators by using data from the futures market. The performance of estimators can be affected by several factors, such as the variation of the spreads over time, autocorrelation of the price or order flow and the relationship between order flow and price returns. By using simulated data, the effects of the different factors can be controlled and discussed separately.

Table 2.1 gives a summary about major bid-ask spread estimators and gives further reasons why an estimator is (not) selected by this chapter.

Table 2.1: Bid-Ask Spread Estimators

Estimator	Data requirement	Market designed for	reasons (not) selected by the paper	comments
Roll (1984)	Transaction price data	Stock markets	Earliest spread estimator; wildly used	Only consider the transaction cost
Choi et al. (1988)	Transaction price data; trade direction	Stock markets	Directly related to the Roll estimator; Discussed in sections 2.4 and 2.6.7, pages 23 and 69	Based on the Roll estimator and takes autocorrelation of the trade direction into account
Stoll (1989)	Transaction price data; trade direction	Stock markets	Directly related to the Roll estimator	
George et al. (1991)	Transaction price data; mid-prices	Stock markets	Directly related to the Roll estimator; Mid-prices are unusually not known; Discussed in sections 2.4 and 2.6.4, pages 23 and 56	Roll estimator taking mid-price autocorrelation into account
Huang and Stoll (1997)	Transaction price data and trade direction	Stock markets	Most general estimator, which incorporates most previous estimators	Can be used to decompose the components of the spread
Lesmond, Ogden and Trzcinka, (1999)	Transaction price data	Stock markets	Not suitable for the simulation experiment, and according to CS(2012), the estimator is worse than the CS estimator	Using zero returns to estimate the spread
Hasbrouck (2004, 2009)	Transaction price data	Stock markets	Directly related to the Roll estimator; Performance is similar to Roll	Using Bayesian method to estimate the Roll estimator
Holden (2009)	Transaction price data	Stock markets	Not suitable for the simulation experiment and the FX market because there is not 1/8 dollar minimum tick in the FX market	Using minimum tick to estimate the spread
Corwin and Schultz (2012)	High low transaction data	Stock markets	Latest spread estimator; good performance according to their paper	Not an unbiased estimator

2.4 Estimators and Errors

The following equations describe the quoted and effective spreads:

$$\textit{quoted spread} = \textit{quoted ask} - \textit{quoted bid}$$

$$\textit{effective spread} = 2 \cdot |\textit{transaction price} - \textit{mid price}|$$

where $\textit{mid price} = 0.5 \cdot (\textit{ask price} + \textit{bid price})$. Let s_t be the transaction price. It is equal to the ask/bid price if a buy/sell order is executed:

$$s_t = \begin{cases} ask_t & \textit{Buy order} \\ bid_t & \textit{Sell order} \end{cases} \tag{2.1}$$

The effective spread and the quoted spread are equal only if all transactions take place at the exact ask or bid prices.

Observed prices can be divided into two parts: the bid-ask spread and the unobserved mid-price. Formally, the price is given as follows:

$$s_t = M_t + \frac{SP}{2} \cdot BS_t \tag{2.2}$$

where s is the log exchange rate, and M_t is the mid-price. SP is the efficient bid-ask spread, and BS is the trade indicator noting the direction of trade.

$$BS = \begin{cases} 1 & \textit{buy order} \\ -1 & \textit{sell order} \end{cases} \tag{2.3}$$

By taking the first-order difference of the above equation, we obtain an expression for the price return.

$$\Delta s_t = \Delta M_t + \frac{SP}{2}(BS_t - BS_{t-1}) \tag{2.4}$$

where Δ is the first order difference operator. This equation suggests that the spread will increase (decrease) the observed return when the change in the trade direction has the same (opposite) sign as the mid-price change. If the trade direction does not change

($BS_t - BS_{t-1} = 0$), the observed return is equal to the mid-price change.

Roll (1984) obtains the covariance of price returns from the equation above. The covariance is as follows:

$$\begin{aligned}
&Cov(\Delta s_t, \Delta s_{t-1}) = E(\Delta s_t \Delta s_{t-1}) \\
&= E[[\Delta M_t + \tfrac{SP}{2}(BS_t - BS_{t-1})] \cdot [\Delta M_{t-1} + \tfrac{SP}{2}(BS_{t-1} - BS_{t-2})]] \\
&= E(\Delta M_t \cdot \Delta M_{t-1}) + E[\Delta M_{t-1} \cdot \tfrac{SP}{2}(BS_t - BS_{t-1})] \\
&+ E[\Delta M_t \cdot \tfrac{SP}{2}(BS_{t-1} - BS_{t-2})] + (\tfrac{SP}{2})^2 E[(BS_t - BS_{t-1}) \cdot (BS_{t-1} - BS_{t-2})] \\
&= E(\Delta M_t \cdot \Delta M_{t-1}) + E[\Delta M_{t-1} \cdot \tfrac{SP}{2}(BS_t - BS_{t-1})] \\
&+ E[\Delta M_t \cdot \tfrac{SP}{2}(BS_{t-1} - BS_{t-2})] + (\tfrac{SP}{2})^2 E[(BS_t \cdot BS_{t-1}) \\
&- (BS_{t-1})^2 - (BS_t \cdot BS_{t-2}) + (BS_{t-1} \cdot BS_{t-2})]
\end{aligned} \tag{2.5}$$

where E is the expectation operator. Roll assumes that the mid-price and order flow follow a random walk and that spreads are fixed. Thus, several terms in the above equation would be zeros and the variance of the order flow equals one.

$$\begin{aligned}
&E(\Delta M_t \cdot \Delta M_{t-1}) \\
&= E[\Delta M_{t-1} \cdot \tfrac{SP}{2}(BS_t - BS_{t-1})] \\
&= E[\Delta M_t \cdot \tfrac{SP}{2}(BS_{t-1} - BS_{t-2})] \\
&= 0 \\
&E(BS_t \cdot BS_{t-1}) = E(BS_t \cdot BS_{t-2}) = E(BS_{t-1} \cdot BS_{t-2}) = 0 \\
&E(BS_{t-1} \cdot BS_{t-1}) = 1
\end{aligned} \tag{2.6}$$

By re-arranging the above equation, we obtain an equation as follows,

$$Cov(\Delta s_t, \Delta s_{t-1}) = E\left(-\frac{SP^2}{4}\right) \tag{2.7}$$

Finally, the Roll spread estimator can be written as follows:

$$SP = 2\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} \tag{2.8}$$

When some elements of (2.6) are invalid or spreads are not fixed, the Roll estimator is biased. Therefore, it is of interest to evaluate the influence of these factors on the Roll estimator.

Error one: the Roll estimator assumes that the spread is fixed. When the spread is time-varying, Equation (2.7) becomes as follows:

$$Cov(\Delta s_t, \Delta s_{t-1}) = E\left(-\frac{{SP_t}^2}{4}\right) \tag{2.9}$$

The spread in period t equals the mean of spreads plus a shock. Formally,

$$SP_t = \overline{SP} + u_t \tag{2.10}$$

where $\overline{SP}$ is the mean of spreads, and u_t is the shock at period t, which has an expected value of zero. The second moment of the spread is given as follows,

$$E\left({SP_t}^2\right) = \overline{SP}^2 + E(u_t^2) \tag{2.11}$$

Therefore the true spread can be represented as follows:

$$SP = 2 \cdot \sqrt{-Cov(\Delta s_t, \Delta s_{t-1}) - \frac{E(u_t^2)}{4}} \tag{2.12}$$

The error is represented by the following equation:

$$Error = 2 \cdot \left[\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - \sqrt{-Cov(\Delta s_t, \Delta s_{t-1}) - \frac{E(u_t^2)}{4}}\right] \tag{2.13}$$

The equation above suggests that the Roll estimator would overestimate the mean of spreads when spreads are time-varying and the error is positively correlated with the variance of spreads.

Error two: the Roll estimator assumes that ΔM_t is independently and identically distributed (iid). When this assumption is not valid, that is, when the mid-price does not follow a random walk, the covariance between the transaction price returns and the spread is as follows:

$$\begin{aligned} Cov(\Delta s_t, \Delta s_{t-1}) &= Cov(\Delta M_t, \Delta M_{t-1}) - \frac{SP^2}{4} \\ SP &= 2\sqrt{Cov(\Delta M_t, \Delta M_{t-1}) - Cov(\Delta s_t, \Delta s_{t-1})} \end{aligned} \tag{2.14}$$

Then, the Roll estimator would be biased if the error term is negatively serially correlated.

Formally, the error of the estimate is given as follows

$$\begin{aligned} &Error = 2 \cdot \left[\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - \sqrt{Cov(\Delta M_t, \Delta M_{t-1}) - Cov(\Delta s_t, \Delta s_{t-1})}\right] \\ &\begin{cases} Error > 0 & \textit{if } Cov(\Delta M_t, \Delta M_{t-1}) < 0 \\ Error < 0 & \textit{if } Cov(\Delta M_t, \Delta M_{t-1}) > 0 \end{cases} \end{aligned} \tag{2.15}$$

The estimation error is positive if the mid-price returns are negatively autocorrelated and vice versa. George et al. (1991) analyse a modified Roll model which is unbiased when the mid-price is autocorrelated. Their method requires either a portfolio of assets or data on quote spreads. Because this chapter focuses on the estimators' performance for a single time series, George et al.'s (1991) work is not evaluated.

Error three: when order flow is autocorrelated, the Roll model is biased. Choi et al. (1988) modify the Roll model to overcome this problem. Let δ be the probability that order flow does not change directions between two periods, while δ is assumed to be 0.5 in the original Roll model. Formally, when $Pr(BS_t = BS_{t-1}) = \delta$, Equation (2.5) becomes:

$$\begin{aligned} &Cov(\Delta s_t, \Delta s_{t-1}) = (\tfrac{SP}{2})^2 \, E[(BS_t - BS_{t-1}) \cdot (BS_{t-1} - BS_{t-2})] \\ &= SP^2 \, (1-\delta)^2 \end{aligned} \tag{2.16}$$

Choi et al.'s (1988) version of the Roll model is given as follows:

$$SP = \frac{\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})}}{1-\delta} \tag{2.17}$$

Roll's error is given as follows:

$$Error = 2\left[\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - \frac{2\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})}}{2(1-\delta)}\right] \tag{2.18}$$

When $\delta = 0.5$, the Choi et al. (1988) model reduces to the original Roll model. The Roll model underestimates the true spread when $\delta > 0.5$ and vice versa. The error is positively correlated with the autocorrelation of order flow (i.e., the value of δ).

Error four: when inventory control and asymmetric information components of the spread exist, as assumed in the HS model, mid-price returns are influenced by past order flow. Formally,

$$\Delta M_t = \varrho \frac{SP}{2} \cdot BS_{t-1} + \epsilon_t \tag{2.19}$$

where ϱ is the proportion of the inventory control and asymmetric information component of the spread and ϵ_t is a random shock with mean zero , and $0 \leq \varrho \leq 1$. Under these circumstances, the Roll model underestimates the true spread, because it only considers the order processing cost. Formally, substituting Equation (2.19) into Equation (2.5), one can obtain:

$$Cov(\Delta s_t, \Delta s_{t-1}) = (\frac{SP}{2})^2 (1 - \varrho) \tag{2.20}$$

Thus, the true spread should be as follows:

$$SP = 2 \cdot \sqrt{\frac{-Cov(\Delta s_t, \Delta s_{t-1})}{(1 - \varrho)}} \tag{2.21}$$

Roll's error is given as follows:

$$Error = 2 \cdot \left[\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - \sqrt{\frac{-Cov(\Delta s_t, \Delta s_{t-1})}{(1 - \varrho)}} \right] \tag{2.22}$$

Stoll (1989) analyses the components of the spread using Roll's framework and obtains an equivalent result. This model cannot be estimated with transaction data alone. The equation above suggests that when $\varrho = 0$ (i.e., when there are no inventory control or asymmetric information components), the Stoll (1989) model reduces to the original Roll model. The error of the model is positively correlated with the proportion of inventory control and asymmetric information components (i.e. the value of ϱ).

Error five: Harris (1990) derives the expression of the Roll estimator for a finite sample.

$$Cov(\Delta s_t, \Delta s_{t-1}) \approx -\frac{-SP^2}{4} - \frac{\sigma^2}{n} \tag{2.23}$$

and thus,

$$SP \approx 2\sqrt{-\left[Cov(\Delta s_t, \Delta s_{t-1}) + \frac{\sigma^2}{n} \right]} \tag{2.24}$$

where σ^2 is the variance of transaction price returns and n is the sample size. The bias is a decreasing function of n, and when n is infinite, the Roll estimator is unbiased. The error is given as follows:

$$Error = 2 \cdot \left[\sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - \sqrt{-\left[Cov(\Delta s_t, \Delta s_{t-1}) + \frac{\sigma^2}{n} \right]} \right] \tag{2.25}$$

The equation suggests that when the volatility of transaction prices is high, the Roll estimator underestimates the spread, and, overcoming the error requires additional observations.

Error six: when there is feedback trading, the Roll estimator is biased. The existence of feedback trading suggests that order flow is influenced by past mid-price returns. The simplest type of feedback trading occurs when a trader decides whether to buy or sell according to the most recent mid-price return. Formally, feedback trading can be given as follows,

$$OF_t = M(\Delta M_t) \tag{2.26}$$

where $M(\cdot)$ is a function of past mid-price return. We assume that the trader first observes the mid-price return and then places the order, thus ΔM_t is the trader's past return at period t.

When there is feedback trading ($E[\Delta M_{t-1} \cdot \frac{SP}{2}(BS_t - BS_{t-1})] \neq 0$, Equation (2.5) and the true spread become

$$\begin{aligned} Cov(\Delta s_t, \Delta s_{t-1}) &= \frac{SP}{2} \cdot [-Cov(\Delta M_{t-1}, BS_{t-1})] - \frac{SP^2}{4} \\ SP &= -Cov(\Delta M_{t-1}, BS_{t-1}) + \sqrt{[Cov(\Delta M_{t-1}, BS_{t-1})]^2 - 4Cov(\Delta s_t, \Delta s_{t-1})} \end{aligned} \tag{2.27}$$

Thus the error is given as follows:

$$\begin{aligned} Error = 2 \cdot \sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} + Cov(\Delta M_{t-1}, BS_{t-1}) \\ - \sqrt{[Cov(\Delta M_{t-1}, BS_{t-1})]^2 - 4Cov(\Delta s_t, \Delta s_{t-1})} \end{aligned} \tag{2.28}$$

The error is positive when there is positive feedback trading and vice versa.

Along with the theoretical analysis of the biases of the Roll model, later in this section, cases of time-varying spreads, auto-correlated mid-price returns, auto-correlated order flow, and mid-price returns affected by order flow, feedback trading are also considered for empirical evaluations.

Huang and Stoll (1997) relax the assumption that the mid-price follows a random walk and assume instead that the mid-price is affected by two factors. One factor is the fundamental value and the other is the inventory level. It is also assumed that these two factors are equally weighted. Then, the sum of the trade indicators is employed as a proxy for the inventory level. Formally, the mid-price is given as follows:

$$M_t = F_t + \beta \frac{SP}{2} \sum_{i=1}^{t-1} BS_i + \varepsilon_{1,t} \tag{2.29}$$

where F_t is the fundamental value, β is equal to the fraction of inventory control cost of the spread. $\varepsilon_{1,t}$ is a random shock. Taking the first order difference of the equation above we obtain:

$$\Delta M_t = \Delta F_t + \beta \cdot \frac{SP}{2} BS_{t-1} + \varepsilon'_{1,t} \tag{2.30}$$

where $\varepsilon'_{1,t} = \Delta\varepsilon_{1,t}$. The order flow conveys fundamental information. A market maker uses the spread to protect himself when trading with an informed trader. If the autocorrelation of order flow series is known by market makers, then only unexpected incoming orders convey fundamental information. In other words, market makers have rational expectations about the incoming order based on past order flow, and thus the difference between the incoming order and its rational expectation contains the information about changes in the fundamental value. Formally, the change in the fundamental value can be written as follows:

$$\begin{aligned} \Delta F_t &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} [E(BS_{t-1} | BS_{t-2})] + \varepsilon_{2,t} \\ &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} (1 - 2\theta) BS_{t-2} + \varepsilon_{2,t} \end{aligned} \tag{2.31}$$

where $\alpha \frac{SP}{2}$ is the effect of incoming orders on the dealer's beliefs about fundamental value, and the conditional expectation on the right hand side is the market makers' rational expectation. θ is the probability of order reversal. $\varepsilon_{2,t}$ is a random shock. Because market makers may adjust prices to avoid inventory disequilibrium,which would increase the likelihood that the next order will be in the opposite direction, it is assumed that $\theta \geq 0.5$. By substituting Equations (2.31) and (2.30) into Equation (2.2), the model becomes,

$$\begin{aligned} &\Delta s_t = \tfrac{SP}{2} BS_t + (\alpha + \beta - 1) \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} (1 - 2\theta) BS_{t-2} + \epsilon_t \\ &\epsilon_t = \varepsilon'_{1,t} + \varepsilon_{2,t} \end{aligned} \tag{2.32}$$

The equation corresponds to the Huang and Stoll model. θ can be estimated as follows,

$$BS_{t-1} = (1 - 2\theta) BS_{t-2} + \varepsilon_t \tag{2.33}$$

The generalised moment method is applied to estimate the two equations simultaneously.

When order flow is correlated with mid-price returns (feedback trading), endogeneity biases the HS model. If the order flow is actually determined by the return, the bid price is only obtained when return is negative and the ask price when the return is positive. Formally, if feedback trading exists:

$$\begin{aligned} BS_t &= \psi \Delta M_t + (1-2\theta) BS_{t-1} + \mu_t \\ &= \psi \left[\beta \cdot \tfrac{SP}{2} BS_{t-1} + \varepsilon'_{1,t} + \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} (1-2\theta) BS_{t-2} + \varepsilon_{2,t} \right] \\ &\quad + (1-2\theta) BS_{t-1} + \mu_t \end{aligned} \tag{2.34}$$

where μ_t is a random shock. We assume that order flow is not autocorrelated ($\theta = 0.5$) and that the only component of the spread is transaction costs ($\alpha = \beta = 0$) to focus on feedback trading. Then the equation above becomes,

$$BS_t = \psi \left(\varepsilon'_{1,t} + \varepsilon_{2,t} \right) + \mu_t = \psi \epsilon_t + \mu_t \tag{2.35}$$

The HS model should be adjusted to the following simultaneous equation model,

$$\begin{aligned} \Delta s_t &= \tfrac{SP}{2} BS_t + (\alpha + \beta - 1) \tfrac{SP}{2} BS_{t-1} + \epsilon_t \\ BS_t &= \psi \epsilon_t + \mu_t \end{aligned} \tag{2.36}$$

The second equation describes feedback trading. When ψ is positive, there is positive feedback trading and thus the covariance of mid-price returns and order flow are positive. It is easy to show that when $cov(BS_t, \epsilon_t) \neq 0$ in Equation (2.36), the estimated value $\widehat{\frac{SP}{2}}$ is given as follows:

$$\frac{\widehat{SP}}{2} = \frac{SP}{2} + cov(BS_t, \epsilon_t) \tag{2.37}$$

Then, the error should be

$$Error = 2 \cdot cov(BS_t, \epsilon_t) \tag{2.38}$$

Equation (2.38) suggests that when there is positive feedback trading, the HS estimator overestimates the true spread and vice versa.

The feedback effect depends greatly on market organisation and other factors and cannot be evaluated using only the transaction and order flow data, therefore ψ cannot be known. The HS model is biased when feedback trading exists.

Later in this section, the case of feedback trading is considered to empirically evaluate the influence of feedback trading on the HS model.

Corwin and Schultz's (2012) spread estimator uses the daily high-low prices to estimate the spread. The estimator assumes that the mid-price follows a random walk, the spread is fixed and the highest (lowest) mid-price corresponds to the highest (lowest) transaction price and the buy (sell) order.

From Equation (2.2), the observed high (low) price is assumed to be the highest (lowest) mid-price plus (minus) half of the spread.

$$H_t^O = TH_t^M + \frac{SP}{2} \tag{2.39}$$

$$L_t^O = TL_t^M - \frac{SP}{2} \tag{2.40}$$

where H_t^O is the logarithm of the observed daily high price which is a transaction price and L_t^O is the logarithm of the observed daily low price which is also a transaction price, TH_t^M is logarithm of the daily high mid-price and TL_t^M is the logarithm of the daily low mid-price. From equations above, one can obtain the following equation:

$$H_t^O - L_t^O = TH_t^M - TL_t^M + SP \tag{2.41}$$

This equation suggests that the high-low ratio of transaction prices is the summation of the high-low ratio of mid-prices and the spread. Squaring both sides of the equation, we have,

$$\left(H_t^O - L_t^O\right)^2 = \left(TH_t^M - TL_t^M\right)^2 + 2\left(TH_t^M - TL_t^M\right) \cdot SP + SP^2 \tag{2.42}$$

Equations (2.42) and (2.41) describe the basic relationship between the high-low ratio of transaction price and mid-price and the spread.

Assume the movement of mid-prices is a Wiener process and the daily high price is at the ask price and the daily low price is at the bid price. Using the results from Parkinson (1980) and Garman and Klass (1980), the first and second order moments of the ratio of the high price to the low price can be written as follows,

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left(TH_t^M - TL_t^M\right)\right\} = k_2 \cdot \sigma \tag{2.43}$$

$$E\left\{\frac{1}{T}\sum_{t=1}^{T}\left(TH_t^M - TL_t^M\right)^2\right\} = k_1 \cdot \sigma^2 \tag{2.44}$$

where σ^2 is the variance of the daily price, $k_1 = 4\ln(2)$ and $k_2 = \sqrt{\frac{8}{\pi}}$. By considering one day high and low prices and two-day high and low prices, one can solve the equations and obtain the spread. Formally, according to Equations (2.42), (2.43) and (2.44), the following equations can be obtained,

$$2 \cdot k_1 \sigma^2 + 4 \cdot k_2 \cdot \sigma SP + 2SP^2 - \beta = 0 \tag{2.45}$$

where

$$\beta = E\left\{\sum_{J=0}^{1}\left(H_{t+J}^O - L_{t+J}^O\right)^2\right\} \tag{2.46}$$

β is the square of the summation of two adjacent daily high-low ratios. Equation (2.45) suggests that the high-low ratio of the transaction prices can be represented by the volatility of mid-price returns and the spread. According to Equation (2.42), the high-low ratio over two-period can be written as follows:

$$\left(H_{t,t+1}^O - L_{t,t+1}^O\right)^2 = \left(TH_{t,t+1}^M - TL_{t,t+1}^M\right)^2 + 2\left(TH_{t,t+1}^M - TL_{t,t+1}^M\right) \cdot SP + SP^2 \tag{2.47}$$

where subscript $(t, t+1)$ represents the value over a two-day interval. Considering that the variance of the summation of two variables drawn from some distributions equals the summation of the variances, one can use the summation of the high-low ratios of two adjacent days to represent a high-low ratio over two days. Formally, the relationship between high-low ratios in one period and in two periods is based on the following expressions.

$$\begin{aligned} E\left(TH_{t,t+1}^M - TL_{t,t+1}^M\right) &= \sqrt{2} \cdot E\left[\sum_{J=0}^{1}\left(TH_{t+J}^M - TL_{t+J}^M\right)\right] \\ E\left(TH_{t,t+1}^M - TL_{t,t+1}^M\right)^2 &= E\left[\sum_{J=0}^{1}\left(TH_{t+J}^M - TL_{t+J}^M\right)^2\right] \end{aligned} \tag{2.48}$$

The right hand sides of the equations represent expectations of the high-low ratio and its square over two periods. On the left hand side, there are the summations of the high-low ratios for two periods. Substituting Equations (2.43) and (2.44) into Equation (2.48), we have,

$$2 \cdot k_1 \sigma^2 + 2\sqrt{2} \cdot k_2 \cdot \sigma SP + SP^2 - \gamma = 0 \tag{2.49}$$

where

$$\gamma = \left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2 \tag{2.50}$$

γ is the square of the high-low ratio for a two-day interval.

Equations (2.45) and (2.49) are the core of the CS estimator. In general, the system of equations does not have analytical solutions. Assuming that $\sqrt{k_1} = k_2$, analytical solutions can be obtained. The solutions to the equations system (2.45) and (2.49) for this special case are given as follows,

$$\sigma = \frac{\sqrt{\beta/2} - \sqrt{\beta}}{k_2(3 - 2\sqrt{2})} - \sqrt{\frac{\gamma}{k_2^2(3 - 2\sqrt{2})}} \tag{2.51}$$

$$SP = \frac{\sqrt{2\beta} - \sqrt{\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma}{3 - 2\sqrt{2}}} \tag{2.52}$$

In practice, the estimation of the spread is obtained by the averaging the solutions.

$$\overline{SP} = \frac{1}{n}\sum_{t=1}^{n} SP_t \tag{2.53}$$

The CS estimator is biased for two reasons. First, some assumptions about the measure of the spread are not valid, and the means of the errors are not zeros. These errors can be called the spread errors. Second, the average of spreads is a biased estimator of the expectation of spreads due to the non-linearity of the equations system (Corwin and Schultz 2012). Due to non-linearity, errors, with means equal to zeros, cannot be completely eliminated by increasing the sample size. The second type of error can be called equation errors. Formally, suppose there are spread and equation errors; then the system becomes

$$\begin{aligned} &2 \cdot k_1\sigma^2 + 4 \cdot k_2 \cdot \sigma(\widehat{SP} + e_{sp}) + 2(\widehat{SP} + e_{sp})^2 - \beta = e_\beta \\ &2 \cdot k_1\sigma^2 + 2\sqrt{2} \cdot k_2 \cdot \sigma(\widehat{SP} + e_{sp}) + (\widehat{SP} + e_{sp})^2 - \gamma = e_\gamma \end{aligned} \tag{2.54}$$

where e_{sp} is a spread error and e_γ and e_β are equation errors. The solutions are given as follows:,

$$\sigma = \frac{\sqrt{(\beta + e_\beta)/2} - \sqrt{\beta + e_\beta}}{k_2(3 - 2\sqrt{2})} - \sqrt{\frac{\gamma + e_\gamma}{k_2^2(3 - 2\sqrt{2})}} \tag{2.55}$$

$$SP = \frac{\sqrt{2(\beta + e_\beta)} - \sqrt{\beta + e_\beta}}{3 - 2\sqrt{2}} - \sqrt{\frac{\gamma + e_\gamma}{3 - 2\sqrt{2}}} - e_{sp} \tag{2.56}$$

The solutions will reflect the true spread when $E(e_{sp}) = 0$, but will not converge to the true spread even when $E(e_{\beta,t}) = E(e_{\gamma,t}) = 0$.

$$\begin{aligned} \overline{SP} &= \lim_{n\to+\infty} \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\sqrt{2(\beta_t + e_{\beta,t})} - \sqrt{\beta_t + e_{\beta,t}}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma_t + e_{\gamma,t}}{3-2\sqrt{2}}} - e_{sp} \right] \\ &\neq \lim_{n\to+\infty} \frac{1}{n} \sum_{t=1}^{n} \left[\frac{\sqrt{2\beta_t} - \sqrt{\beta_t}}{3-2\sqrt{2}} - \sqrt{\frac{\gamma_t}{3-2\sqrt{2}}} \right] \end{aligned} \tag{2.57}$$

Equation (2.57) is the general form of the estimation of σ or the spread. The following paragraphs will discuss the specific errors in sequence.

Error one (a spread error): the CS estimator assumes that the mid-price of the highest (lowest) transaction price corresponds to the highest (lowest) mid-price. The assumption is not true when the standard deviation of mid-price is much larger than the spread. Under this circumstance, the true highest (lowest) mid-prices may be not used but the less (greater) prices can be; as a result, the true spread is underestimated.

Let us consider the above formally. According to the assumption of the CS estimator, the relationship between the highest (lowest) observed prices and the highest (lowest) mid-prices is given as follows:

$$\begin{aligned} H_t^O &= TH_t^M + \frac{\widehat{SP}}{2} \\ L_t^O &= TL_t^M - \frac{\widehat{SP}}{2} \end{aligned} \tag{2.58}$$

where TH^M (TL^M) is the true highest (lowest) mid-price and $\widehat{SP}$ is the estimated spread. Then the $\widehat{SP}$ is given as follows:

$$\begin{aligned} \frac{\widehat{SP}}{2} &= H_t^O - TH_t^M \\ \frac{\widehat{SP}}{2} &= TL_t^M - L_t^O \\ (H_t^O - L_t^O) - (TH_t^M - TL_t^M) &= \widehat{SP} \end{aligned} \tag{2.59}$$

In fact, the real values are as follows:

$$\begin{aligned} H_t^O &= H_t^M + \frac{SP}{2} \\ L_t^O &= L_t^M - \frac{SP}{2} \end{aligned} \tag{2.60}$$

where

$$\begin{aligned} TH_t^M &\geq H_t^M \\ TL_t^M &\leq L_t^M \end{aligned} \tag{2.61}$$

The inequalities suggest that in reality, the assumption is not always true, that is, not all the highest (lowest) prices are used. [a] Combine the equations above, we can obtain that,

$$\begin{aligned} TH_t^M + \frac{\widehat{SP}}{2} &= H_t^M + \frac{SP}{2} \\ TL_t^M - \frac{\widehat{SP}}{2} &= L_t^M - \frac{SP}{2} \end{aligned} \tag{2.62}$$

If the assumption were correct, which suggests that $TH_t^M = H_t^M$ and $TL_t^M = L_t^M$, then the spread is correctly estimated. If the assumption were not true, which suggests that $TH_t^M > H_t^M$ or $TL_t^M < L_t^M$, then the spread is underestimated. Formally,

$$\begin{aligned} TH_t^M - H_t^M &= \frac{SP}{2} - \frac{\widehat{SP}}{2} > 0 \\ TL_t^M - L_t^M &= -\frac{SP}{2} + \frac{\widehat{SP}}{2} < 0 \end{aligned} \tag{2.63}$$

thus,

$$SP > \widehat{SP} \tag{2.64}$$

Equation (2.64) suggests that when the assumption is not valid, the spread is underestimated, and the error is given as follows:

$$e_m = SP - \widehat{SP} = D_H + D_L \tag{2.65}$$

where Let D_H and D_L are given as follows:,

$$\begin{aligned} D_H &= TH_t^M - H_t^M \\ D_L &= L_t^M - TL_t^M \end{aligned} \tag{2.66}$$

The probability of the events that $TH_t^M > H_t^M$ or $TL_t^M < L_t^M$ occur is positively correlated with the mid-price volatility and negatively related to the number of observations in an interval. An infinite number of observations is not problematic because the distance ap-

[a] The highest (lowest) mid-price may be with a sell (buy) order (50% probability). In this circumstance, if a mid-price which is close to highest (lowest) mid-price is with a buy (sell) order, it will surpass the highest (lowest) mid-price to be the one corresponding to the highest (lowest) transaction price. The fact can be written as follows,

$$\begin{aligned} TH_t^M - \frac{SP}{2} &< H_t^M + \frac{SP}{2} \\ TL_t^M + \frac{SP}{2} &< L_t^M - \frac{SP}{2} \end{aligned}$$

In this case, TH_t^M, the true highest mid-price, is with a sell order, and thus it's corresponding transaction price is sometimes less than the transaction price which is corresponding to a mid-price H_t^M which is with a buy order.

proaches zero. Given that the CS estimator usually functions with daily data, the number of observations in a single interval is limited; the distance is positively correlated with the volatility of mid-price returns. If the number of observations is fixed over an interval, more volatility yields greater diffusion and larger average distances between two observations. Higher volatility and less samples introduce larger errors.

Error two (a spread error): Equation (2.45) also assumes that the highest (lowest) transaction price normally happens on a buy (sell) order. The assumption is not true when the number of trades in the interval is small. Formally, when the assumption is not true, the transaction prices are given as follows:,

$$\begin{aligned} H_t^O &= TH_t^M - \tfrac{SP}{2} \quad (a) \\ L_t^O &= TL_t^M + \tfrac{SP}{2} \quad (b) \end{aligned} \tag{2.67}$$

When both (a) and (b) hold, Equation (2.41) becomes the following:

$$H_t^O - L_t^O = TH_t^M - TL_t^M - SP \tag{2.68}$$

When either (a) or (b) hold, Equation (2.41) becomes the following:

$$H_t^O - L_t^O = TH_t^M - TL_t^M \tag{2.69}$$

The equations above suggest that the spread will be underestimated. The error is one or two times as large as the spread. The probability that events (a) and (b) occur is negatively correlated with the number of observations in an interval. Thus the error is also negatively correlated with the number of observations in an interval. One can easily show that if the probability of events (a) or (b) happening is $1 - \eta$ where $0 \leq \eta \leq 1$, the error is given the as following:

$$e_o = 2\eta \cdot SP \tag{2.70}$$

Error three (an equation error): Equation (2.49) is not always valid. Equation (2.49) links the one-period volatility and the two-period volatility based on Parkinson's (1980) volatility estimator (equations 2.43 and 2.44). While Parkinson's (1980) volatility estimator performs when the sample size is large, Equation (2.49) applies the estimator using one observation (a high-low ratio over two periods) on the left hand side and applies it using two observations (summation of two high-low ratios for two periods) on the right hand side. It is reasonable that the two-period estimation is more accurate than the one-period estimation. The imbalance in the accuracy produces errors. The CS estimator assumes that the relationship of the high-low ratios is valid for each pair of two adjacent periods:

$$\begin{aligned} \left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) &= \sqrt{2}\cdot\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)\right] \\ \left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2 &= \sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)^2 \end{aligned} \tag{2.71}$$

In fact, the true relationship should be as descibed by Equation (2.48). Comparing these two equations, the high-low ratios should be the following,

$$\begin{aligned} &E\left[\left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2\right] = \left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2 + \epsilon_1 \\ &E\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) = \left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) + \epsilon_2 \\ &E\left[\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2\right] = \left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2 + \epsilon_3 \\ &E\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) = \sqrt{2}E\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)\right] = \sqrt{2}\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)\right] + \epsilon_4 \\ &E\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2 = E\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)^2\right] = \sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)^2 + \epsilon_5 \end{aligned} \tag{2.72}$$

where ϵs are random shocks with means of zero.

The correct version of Equation (2.47) should be as following:

$$E\left[\left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2\right] = E\left[\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2\right] + 2SP\cdot E\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) + SP^2 \tag{2.73}$$

By substituting Equations (2.72) into the equation above, we obtain:

$$\left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2 + \epsilon_1 = \left(TH^M_{t,t+1} - TL^M_{t,t+1}\right)^2 + \epsilon_3 + 2SP\cdot\left[\left(TH^M_{t,t+1} - TL^M_{t,t+1}\right) + \epsilon_2\right] + SP^2 \tag{2.74}$$

Because ϵ_1, ϵ_2 and ϵ_3 are shocks corresponding to the same high-low ratio of the one-period estimation, and thus are highly correlated, they can be partially offset out in Equation (2.47). Formally, this is described as follows:

$$e_{v1} = \epsilon_1 - 2SP_t \cdot \epsilon_2 - \epsilon_3 \approx 0 \tag{2.75}$$

Equations (2.74) are two periods versions of Equation (2.45), therefore, e_{v1} shares the same intuition as the error in the first equation of the estimator system (i.e. e_β).

By substituting Equations (2.72) into Equations (2.74), the CS estimator links the high-low ratio for two periods with the two high-low ratios for the single periods.

$$\begin{aligned} E\left[\left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2\right] &= E\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)^2\right] + 2SP \\ &\quad \cdot\sqrt{2}E\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)\right] + SP^2 \\ \left(H^O_{t,t+1} - L^O_{t,t+1}\right)^2 + \epsilon_1 &= \left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right)^2 + \epsilon_5\right] + 2SP \\ &\quad \cdot\sqrt{2}\left[\sum_{J=0}^{1}\left(TH^M_{t+J} - TL^M_{t+J}\right) + \epsilon_4\right] + SP^2 \end{aligned} \tag{2.76}$$

In contrast to Equations (2.74), ϵ_4 and ϵ_5 correspond to two-period estimation, and the correlation between ϵ_1, ϵ_4 and ϵ_5 is not as high as the one with ϵ_2 and ϵ_3. Thus,

$$e_{v2} = \epsilon_1 - 2SP_t \cdot \epsilon_4 - \epsilon_5 \neq 0 \tag{2.77}$$

e_{v2} is the error of Equation (2.49) (e_γ). The expectations of both e_{v1} and e_{v2} are zeros. The variances could be more important in a non-linear system. The variance of e_{v2} is much greater than that of e_{v1} and is less than $3k_1\sigma^2$ but is still positively correlated with the volatility of mid-price returns.

Furthermore, the irregular timed spaces between trades can also increase error three. The CS estimator assumes that the numbers of trades within time intervals are the same, so that Equations (2.48) are valid. When there are irregular timed spaces, Equations (2.48) are not valid and thus error three increases further.

Error four (an equation error): the approximation that $\sqrt{k_1} = k_2$ can also introduce an error to the system. Formally, the error is given as follows:

$$e_a = 2k_1\sigma^2 - 2k_2^2\sigma^2 = 0.452\sigma^2 \tag{2.78}$$

e_a influences both e_β and e_γ. It is straightforward that the error is positively correlated with the volatility of mid-price returns.

Error five: the CS estimator is derived under the assumption that the mid-price follows a Brownian motion. When the mid-price is autocorrelated, Parkinson's (1980) estimator on which the CS estimator is based is no longer unbiased, and the CS estimator could underestimate the variance of the price returns if the mid-price is negatively correlated and thus overestimate the spread.

Error six: similar to the Roll estimator, the time-varying spread can also influence the accuracy of the estimator due to non-linearity.

Error seven: similar to the Roll estimator, the CS estimator does not consider the existence of the IC & AS components of the spread. Therefore, the CS estimator will underestimate the true spread when these components are not zeros.

Furthermore, due to non-linearity, the joint effect of errors could be several times as large as the summation of single errors.

Most errors of the CS estimator are related to the mid-price returns volatility. Thus, various sampling frequencies that affect the volatility and various ratios of returns to spreads are designed to empirically evaluate the impact of errors. Cases of time-varying spread, autocorrelated mid-price returns, irregular timed spaces are also used to examine the influence of the errors.

2.5 Comparison Procedure

In this section, we introduce the procedure that we will employ to comparing the estimators empirically and analyse the errors.

The estimators are compared step-wise using simulation experiments.

First, we consider the ideal case where random mid-price returns and order flow, low mid-price returns volatility and fixed bid-ask spreads are simulated and all the assumptions of the estimators are satisfied.

Second, we study the case of higher mid-price returns volatility and keep other factors the same as the ideal case. Thus, we can attribute the difference in performance of the estimators in this case and case one to the change in the volatility of mid-price returns. The mid-price returns volatility is close to the real volatility in the foreign exchange market.

Because we focus on the performance of estimators in the foreign exchange market, this case will be the benchmark.

Third, we allow the bid-ask spread to be time-varying and keep other conditions as in the previous case. This way we can ensure that all the differences between these cases are caused by the time-variability of the spread.

Fourth, settings in case two are kept except that we allow the mid-prices to be auto-correlated. Goodhart et al. (1996) and Daníelsson and Payne (2002) show that quote mid-point returns of both EFX and Reuters D2000-2 are negatively autocorrelated.Thus, we can ensure that all the differences between the case and case two are caused by the auto-correlation of mid-prices. Goodhart and Payne (1996) show that negative autocorrelation is caused by the limit order book. When the market is "thin", the difference between the best bid (ask) price and the second best bid (ask) price is relatively large. After the best bid (ask) price is exhausted, the mid-price moves down (up) sharply. The market maker, whose inventory level was lower because of the buy order raises (lowers) the mid-price to reverse the inventory flow.

Fifth, settings in case two are maintained, and larger spreads relative to the standard deviation of the mid-price returns are used. Thus the differences between these cases are due to the larger spreads.

Sixth, settings in case two are maintained, except that we allow mid-price returns to be influenced by order flow, which suggests that there are inventory control and adverse selection components of the spread. Then, we can identify the influence of these factors. Bessembinder (1994) confirms the positive relationship between the spread and inventory control cost in the foreign exchange market. Lyons (1995) finds both the inventory control and adverse selection costs in the FX market. Because this chapter does not focus on decomposing the components of the spread, we do not distinguish the difference between inventory control and adverse selection costs.

Seventh, settings in case two are maintained except that the auto-correlated order flow is considered. Therefore, the differences of results between this case and case two are caused by the autocorrelation of order flow. The autocorrelation of the order flow could be caused by herding behaviour, hot potato trading, clustering or other reasons.

Eighth, the case of feedback trading is considered. Feedback trading suggests that order flow might be influenced by price returns.De Long et al. 1990 introduce a positive-feedback trading model. Hasbrouck 1991, Nofsinger and Sias 1999 and Daníelsson and Love (2006) show the existence of feedback trading in stock market and the FX market. Settings in case two are maintained except that order flow is assumed to be correlated with the quoted price returns. In this case, the divergence of results from case two are due to feedback trading.

Ninth, time stamps of the real market data (EUR/USD) are used. The settings are similar to case two except for the irregular timed spaces and the number of trades in a time interval. The influence of the irregularity on the estimators can be evaluated.

Tenth, settings in case two are maintained, and various ratios of the mean of the spread to the standard deviation of mid-price returns are considered. Therefore, the influence of the ratio can be identified by comparing the results of cases two and ten.

To sum up, we change one condition and hold the others for each case, so that the unchanged conditions can serve as baselines. By comparing the latest version with the baseline case, we can isolate the influence of a condition on the estimators.

2.6 Simulation Experiments

In this section, simulated data are used. The aim of this section is to assess effects of the following factors on the performance of the estimators: the variation of spreads, the autocorrelation of mid-price returns, the magnitude of the spread relative to the returns volatility, the number of trades in a time interval, the mid-price changes caused by order flow and feedback trading.

There are 1000 replications simulated for each case. There are 432000 periods in a replication. Let one period be one minute, and there is one trade per minute. Thus there are 300 trading days (1440 minutes and 1440 trades per day). For each replication, data are considered for various sampling periods: tick-by-tick, five-minute, fifteen-minute, one-hour, four-hour, 12-hour and 24-hour. To obtain five-minute and longer sampling period data, only close observations in a sampling period are included from the tick-by-tick data. Thus, there are eight subgroups for each replication.

Each replication includes data on order flow, bid-ask spreads, mid-prices, and transaction prices. The data are generated according to the following system. An order has two possible values 1 and −1. The order flow series is given by the following:

$$\begin{aligned} &BS_t = \varphi F(BS_{t-1}) + \psi M(\Delta M_t) + \phi \varpi_t \\ &\varphi = 0\ or\ 1; \psi = 0\ or\ 1; \phi = 0\ or\ 1 \\ &\varphi + \psi + \phi = 1 \end{aligned} \tag{2.79}$$

where BS_t is the order flow, $F(BS_{t-1})$ is a function of the past order flow, which suggests that the order flow is autocorrelated, $M(\Delta M_t)$ is a function of the past mid-price return (assume the trader observes the mid-price return first, then places the order, and ΔM_t is the past return for the trader at period t), which suggests the existence of feedback trading. The function $M(\cdot)$ reflects the following relationship between order flow and past mid-price returns.

$$BS_t \sim \begin{cases} B(1,\kappa) & if \quad \Delta M_t > 0 \\ B(1, 1-\kappa) & if \quad \Delta M_t < 0 \end{cases} \tag{2.80}$$

where $B(1,\kappa)$ is a binomial distribution with one trial and κ probability. When $\kappa = 0.5$, there is no feedback trading, and when $\kappa > 0.5$, there is positive feedback trading and vice versa. ϖ_t is a binomial random variable, which follows a binomial distribution with one trial and 50% probability i.e., $B(1, 0.5)$, which suggests that order flow is randomly drawn from a binomial distribution and both the buy and sell orders possess the same weight. φ, ψ and ϕ are the weighting coefficients, which may take two possible values, 0 or 1. Let the sum of these coefficients equal one, which suggests that only one of the coefficients could be one. This setting ensures that only one factor is considered at a time, thus we can identify the influence of each factor separately.

Mid-price returns are generated using the following equation:

$$\begin{aligned} &\Delta M_t = \tau\, \xi \Delta M_{t-1} + \omega \chi\, BS_{t-1} \cdot \tfrac{SP_t}{2} + \varepsilon_t \\ &\tau = 0\ or\ 1; \omega = 0\ or\ 1; \\ &\tau + \omega = 1 \end{aligned} \tag{2.81}$$

where ξ describes the autocorrelation of mid-price returns. ε_t follows a normal distribution with zero mean and standard deviation σ; SP_t is the bid-ask spread that follows a normal distribution $N(\mu, \varsigma^2)$, where μ is the mean and ς is the standard deviation. When $\varsigma = 0$, spreads are fixed. χ is the fraction of inventory control and asymmetric information cost of the spread, and $(1-\chi)$ suggests the order processing cost of the spread. When $\chi = 0$, the order processing cost is the only component of the spread. When $\xi = 0$ and $\chi = 0$, the mid-price follows a random walk process. Let the summation of τ and ω equal one, which suggests that only one of the coefficients could be one. The setting makes sure only one factor is considered at a time, thus we can identify the influence of each factor separately.

Transaction prices are generated by the following:

$$s_t = M_t + \frac{SP_t}{2} \cdot BS_t \tag{2.82}$$

where s_t is the transaction price.

2.6.1 Fixed Spread

In this section, the ideal case for the estimators is considered. The order flow is random; mid-prices follow a random walk, volatility of mid-price returns is small; and the spread is fixed. Under these conditions, both the HS and the Roll estimators are unbiased, and the error of the CS estimator should be small for the σ is small. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process. Let $\mu = 0.0003$, $\varsigma^2 = 0$ which suggest that the spread is a constant. Let the standard deviation of mid-price returns be small, $\sigma = 0.000003$. The system is given as follows:

$$\begin{aligned}
&BS_t = \varpi_t \\
&\varpi_t \sim B(1, 0.5) \\
&\Delta M_t = \varepsilon_t \\
&\varepsilon_t \sim N(0, 9 \times 10^{-12}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned} \tag{2.83}$$

One thousand replications, each of which has 432000 periods, are generated according to the system above. As we mentioned at the beginning of Section (2.6), each replication consists of eight subgroups from various sampling periods.

Transaction returns are used to calculate the Roll estimator. High-low ratios are used to calculate the CS estimator. Transaction returns and order flow are used to calculate the HS estimator. The standard deviation of mid-price returns is also calculated. For each subgroup, 1000 estimated spreads and 1000 standard deviations of mid-price returns are produced for each estimator.

The results are presented in Table (2.2). The four panels report the summary statistics and the results of the estimators. *Midstd* reports the average of the standard deviations of mid-price returns. *Estimates* represents the average of estimated spreads. *Est-Std* reports the standard deviations of the estimated spreads. *T-test* reports the results of a T-test of which the null-hypothesis is that the difference between estimated spreads and the means of true spreads equals zero. The first row of *T-test* reports the means of the differences, and the second row of *T-test* reports the t-statistics. $\widehat{SP}$, η and *Bridge* are related to the first three errors of the CS estimator. Error four concerns the difference between the analytical and numerical solutions, and the numerical solution greatly depends on the initial values and algorithm. Thus we do not report the numerical solution to avoid introducing unrelated issues. We argue that the first three errors are the main sources and the biases of the CS estimator. Error one of the CS estimator suggests that because the highest (lowest) price may not be corresponding to the highest (lowest) mid-price, the estimated spread could be less than the true one. $\widehat{SP}$ is the estimated spread if there are no errors except for error one, and is calculated from Equation (2.59). Error two suggests that because the highest (lowest) observed price may not be with the buy (sell) order, the estimator may underestimate the true spread. The variable η is the proportion that the highest (lowest) observed price is with the buy (sell) order. Error three suggests that when the CS estimator tries to link the high-low ratios of one period and two periods, the error may happen because the differences between the high-low ratios and the expectations of them cannot be cancelled out. *Bridge* reports error three, which is calculated from Equation (2.77). Because in all the cases, both the mean and the standard deviation of e_{v1} are very small, only e_{v2} is reported. To eliminate the influences of errors one and two, the transaction prices are re-calculated based on the mid-prices using Equations (2.39) and (2.40). *BridgeStd* reports the standard deviation of *Bridge*, because the non-linearity, not only *Bridge* but also its standard deviation, which could play an even more important role, will influence the solutions

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are a positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 3×10^{-6} which is the same as the setting of the system. In the 24-hour case, the standard deviation is 1.14×10^{-4}. Thus the ratio of the spread to the standard deviation ranges from 100 to 2.63.

The Roll estimator estimates the spread correctly, according to the first row of the panel *Roll*. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. In the 24-hour case, the standard deviation is less than 10% of the spread, which suggests that the results are stable across the replications. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true spreads except for the tick-by-tick case. The error in the tick-by-tick case is 4.06×10^{-8} and is much less than the true spread. Therefore, one can conclude that the Roll estimator performs well under ideal conditions.

The HS estimator does estimate the spread correctly, according to the first row of the panel *Huang and Stoll*. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. In the 24-hour case, the standard deviation is about 0.3% of the spread. The standard deviations are much less than those of the Roll estimator, which suggests that the results are stable across the replications and are more reliable than the Roll estimator. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one when the time interval is longer than 30 minutes. Although they are significantly different from zero, the scale of the errors is 10^{-9} which is much less than the true spread and the ones of the Roll estimator. Therefore, one can conclude that the HS estimator performs well under the ideal conditions and performs better than the Roll estimator. The reason why the HS estimator performs better than the Roll estimator is that the former uses more information (order flow) than the latter.

The CS estimator provides approximate estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. We start with the five-minute case. The total error is 4.6×10^{-5}, which is about 15% of the true spread, and is significantly different from zero. The standard deviation of the estimates is 5.51×10^{-7} which is much less than the true spread, which suggests that the estimator is stable in the case. $\widehat{SP}$ is 2.78×10^{-4} which implies that not all the highest (lowest) observed prices are corresponding to the highest (lowest) mid-prices. Error one makes the CS estimation be 2.2×10^{-5} less than the true spread. η suggests that in most cases (about 96.9%), the highest (lowest) prices are with buy (sell) orders. As discussed in the previous section, error two makes the estimate be 6.2%[a] (1.86×10^{-5}) less than the true spread. Both the errors are from to sources: the number of observations in the interval and the volatility of mid-price returns. In this particular case, because the volatility of mid-price returns is low and there are only five observations in one interval, the limited number of observations is the dominant factor which causes errors. *Bridge* suggests that error three in this case is of the order of magnitude of 10^{-9}, which is much lower than the true spread, thus error three can be neglected. When the time interval is longer than five minutes, the CS estimator produces estimates that are more accurate. The total errors are less than 2.34%, although they are significantly different from zero according to the results of t-tests. The main reason of this could be that the number of observations in an interval is large enough. The second row of the panel suggests standard deviation of the estimates is positively correlated with the time interval. In the 24-hour case, the standard deviation is 5.98×10^{-6} which is the maximum and is still much less than the true spread, thus the CS estimator is stable. Furthermore, the CS estimator has the smallest standard deviation among the three estimators. Both errors one and two are much lower than those in the five-minute case. As a result, the total errors of the estimations are also less, since errors one and two contribute more than 85% of total error in the five-minute case. Similar to the five-minute case, the number of observations in an interval is the main source of the error when the volatility of mid-price returns is low. Here, both the errors are much lower than the true spread when there are more than 15 observations in an interval, although error one will never be zero. Nevertheless, it needs to be noted that when the time interval is very long (12 hours or longer), the errors become larger again. This is caused by the second source of the errors: the volatility of mid-price returns. The longer sampling period should be with the higher volatility. Error three (*Bridge* and *BridgeStd*) which is influenced by the volatility increases as the time interval lengthens, and can be neglected for all time intervals.

To sum up, in this section, both the Roll and HS estimators yield unbiased estimates. The estimation of the CS estimator is acceptable (the error is less than 2.34%) except for the case of five-minute intervals. In terms of the standard deviations of the estimations,

[a] $2 \times (1 - 96.9\%) = 6.2\%$

the CS estimator has the best performance and the Roll estimator has the largest standard deviation. According to the rows of RMSE, for long time intervals (longer than 4 hours), the CS estimator is preferred and the HS estimator is preferred in other time intervals. All the estimators have reliable performances under the ideal conditions which are fixed spreads, low volatility of mid-price returns, random mid-price returns, random order flow.

Table 2.2: Fixed Spread and Low Returns Volatility

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-4}$	0.0300	0.0671	0.116	0.164	0.232	0.465	0.805	1.14
Roll 1984								
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Relative Estimate	1	1	1	1	1	1	1	1
Est-Std$\times 10^{-4}$	0.00520	0.0113	0.0195	0.0284	0.0412	0.0822	0.156	0.224
T-test $\times 10^{-8}$	4.06*	3.51	1.97	6.71	-9.07	-9.64	2.92	-1.74
	(2.47)	(0.98)	(0.32)	(0.75)	(-0.70)	(-0.37)	(0.06)	(-0.02)
RMSE$\times 10^{-4}$	0.00520	0.0113	0.0195	0.0284	0.0412	0.0822	0.156	0.224
Huang and Stoll 1997								
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Relative Estimate	1	1	1	1	1	1	1	1
Est-Std$\times 10^{-5}$	9.68×10^{-4}	4.54×10^{-3}	0.0135	0.0275	0.0549	0.204	0.639	1.31
T-test $\times 10^{-8}$	0.401***	0.670***	1.49***	1.65	-0.338	2.11	-6.39	-15.5
	(13.10)	(4.66)	(3.49)	(1.90)	(-0.19)	(0.33)	(-0.32)	(-0.38)
RMSE$\times 10^{-5}$	9.68×10^{-4}	4.54×10^{-3}	0.0135	0.0275	0.0549	0.204	0.639	1.31
Corwin and Schultz 2010								
Estimates $\times 10^{-3}$		0.254	0.293	0.294	0.294	0.295	0.297	0.298
Relative Estimate		0.847	0.977	0.98	0.98	0.983	0.99	0.993
Est-Std$\times 10^{-5}$		0.0515	0.00782	0.0139	0.0265	0.108	0.313	0.598
T-test$\times 10^{-5}$		-4.59***	-0.710***	-0.644***	-0.597***	-0.490***	-0.347***	-0.201***
		(-2819.15)	(-2869.95)	(-1463.39)	(-711.24)	(-143.83)	(-35.08)	(-10.61)
RMSE$\times 10^{-4}$		0.460	0.0700	0.0600	0.0601	0.0512	0.0434	0.0631
$\widehat{SP}$ $\times 10^{-3}$		0.278	0.297	0.298	0.298	0.298	0.297	0.297
η		0.969	1.000	1.000	1.000	1.000	0.999	0.998
Bridge $\times 10^{-8}$		-0.234	-0.518	-0.791	-1.17	-2.48	-4.39	-6.27
BridgeStd$\times 10^{-8}$		0.186	0.312	0.448	0.659	1.52	3.16	5.22

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 9×10^{-12}, i.e. $\Delta M_t \sim N(0, 9 \times 10^{-12})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.

Midstd is the average of the standard deviations of mid-price returns.

Estimates is the average of the estimated spreads.

Relative Estimate represents the average of estimated spreads divided by the true spread. It is one if the estimate equals the true spread.

Est-Std is the standard deviation of the estimated spreads.

T-test is results of T-test of which the null-hypothesis is that the difference between estimated spreads and the means of true spreads equals to zero. The first row reports the means of the differences, and the second row reports the t-statistics.

RMSE is the Root Mean Square Error.

$\widehat{SP}$ is the estimated spread of the CS estimator if there are no errors except for error one, and it is calculated from Equation (2.59). Error one of the CS estimator suggests that because the highest (lowest) price may not be corresponding to the highest (lowest) mid-price, the estimated spread could be less than the true one.

η is the proportion that the highest (lowest) observed price does come with the buy (sell) order. Error two suggests that because the highest (lowest) observed price may not come with the buy (sell) order, the estimator may underestimate the true spread.

Bridge reports error three of the CS estimator (e_{v1}), and it is calculated from Equation (2.77). Because in all the cases, both the mean and the standard deviation of e_{v1} are very small, only e_{v2} is reported. Error three suggests that when the CS estimator tries to link the high-low ratios of one period and two periods, the error may happen because the differences between the high-low ratios and the expectations of which cannot be cancelled out. To eliminate the influences of errors one and two, the transaction prices are re-calculated based on the mid-prices using Equations (2.39) and (2.40)

BridgeStd reports the standard deviation of *Bridge*, because the non-linearity, not only the mean will influence the solutions but the standard deviation which could play an even more important role.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.6.2 Fixed Spread and High Returns Volatility

In this section, most settings are the same as in the previous section except that here a greater standard deviation of mid-price returns is used. Thus all the differences of the performance of the estimators can be imputed to the change of the standard deviation. Although, none of the estimators imposes any restrictions on the mid-price volatility, as mentioned earlier, the volatility does affect the errors of the estimators, and intuitively, when the data are more volatile, the more difficult it is for the estimators to identify the spread from the data. The standard deviations of mid-price returns is $\sigma = 0.0002$, which is similar to the case for real foreign exchange markets and is much greater than the one in the previous section. In this section, order flow is random; mid-prices follow a random walk; and the spread is fixed. Under these circumstances, both the HS and the Roll estimator are unbiased, and the error of the CS estimator should be large for the σ is big. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process and $\mu = 0.0003$, $\varsigma^2 = 0$ which suggests that the spread is a constant. The system is given as follows:

$$\begin{aligned}
&BS_t = \varpi_t \\
&\varpi_t \sim B(1, 0.5) \\
&\Delta M_t = \varepsilon_t \\
&\varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \frac{SP_t}{2} \cdot BS_t
\end{aligned} \tag{2.84}$$

The results are presented in Table (2.3), which is similar to Table (2.2).

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns have a positive relationship. In the tick-by-tick case, the average standard deviation of mid-price returns is 2×10^{-4} which is the same as the setting of the system. In the 24-hour case, the standard deviation is 7.58×10^{-3}. Thus the ratio of the spread to the standard deviation ranges from 1.5 to 0.0396.

The Roll estimator does estimate the spread correctly for short time intervals, and underestimates the spread in the 30-minute case and the one-hour case, and then overestimates the spread in even longer time intervals, according to the first row of the panel *Roll*.

All the differences between this section and the previous section are caused by the higher mid-prices volatility and thus related to error five of the Roll estimator. In the 30-minute case, the error is approximately of the scale of 10^{-5} which is the same as the prediction of Equation (2.25), where the error is of the scale of $\sqrt{\sigma^2/n}$ and is 10^{-5}[b] in this case. The situation is similar in the one-hour case. The Roll estimator requires more observations to obtain an acceptable result. Obtaining a sufficient number of observations for low frequency data is not easy. In the one-hour case, there are 3600 observations per replication, the error is greater than 10% of the spread. According to Equation (2.25), to obtain a result with error less than 3%, 36000 observations are needed, which is 10 years data. The mission could be impossible for longer time intervals. The estimator significantly overestimates the spread in longer time intervals, which could be imputed to the appearance of the positive co-variance. The estimator will be undefined when the co-variance is positive, and the results are set to be zeros in the circumstance. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. In the one-hour case, the standard deviation is greater than 50% of the spread, which suggests that the results are not stable across the replications. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one for short time intervals. When time intervals is longer than one hour, the error becomes too large to be ignored (more than 15% of the spread). One can conclude that the Roll estimator performs well when the volatility of prices is low. When the volatility of prices is high, the estimator is unstable and the error cannot be ignored. Because of the higher volatility, the Roll estimator performs worse than it did in the previous section.

The HS estimator does estimate the spread correctly, according to the first row of the panel *Huang and Stoll*. The error becomes greater as the time interval become longer. In the 24-hour case, the error, which is the largest, is -1.73×10^{-5} which is less than 6% of the true spread. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. When the time interval is longer than 1 hour, the standard deviation is greater than 33% of the spread, which suggests that the estimator is not quite stable in this case. Compared to the previous section, the HS estimator, becomes

[b]The variance is 0.0011^2 and the number of observations is 14400, thus $\sqrt{0.0011^2/14400} = 9.17 \times 10^{-6}$

less stable, which could be imputed to the higher volatility in these cases. If there were more observations, the HS estimator would perform better. The standard deviations are much lower than those of the Roll estimator, which suggests that the results are stable across the replications and are more reliable than the estimates of the Roll estimator. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one. Therefore, one can conclude that the HS estimator performs well in this section and performs better than the Roll estimator. The reason why the HS estimator performs better than the Roll estimator does is that the former uses more information (order flow) than the latter. But because of the higher volatility of prices, the HS estimator performs worse than it does in the previous section.

The CS estimator does not yield good estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. It can be seen that the estimate is an increasing function of time interval (or the volatility of prices). In the five and fifteen-minute cases, the CS estimator has negative results, which might be caused by the joint effect of errors. In the 24-hour case, the estimator significantly overestimates the true spread: the estimate is about 250% of the true spread. The standard deviations of the results are the smallest among the estimators, according to the second row in the panel. The total errors are significantly different from zero, and cannot be ignored. $\widehat{SP}$ ranges from 1.28×10^{-4} to 1.74×10^{-4}, which is much less than the true spread and implies that at least 42% (in the 24-hour case) of the observations do not satisfy the assumption that the highest (lowest) observed prices are corresponding to the highest (lowest) mid-prices, and thus error one causes the estimator to underestimate the true spread by at least 42%. Although $\widehat{SP}$ is an increasing function of the time interval (or the volatility of prices), it is still much less than the true spread, and the function exhibits a concave shape. As it was mentioned earlier, error one decreases in the number of observations in a time interval and increases in the volatility of the prices. For short time intervals, the error one decreases since the number of observations is the dominant factor. In longer time intervals, the volatility becomes the dominant factor. The results of error one suggest that when the volatility of prices is high, the CS estimator will never obtain an accurate estimation. η ranges from 86.8% to 96.0%, which implies that the percentages of that the highest (lowest) prices are with buy (sell) orders are lower than those in previous sections because of the higher volatility.

Error two makes the estimation be at least 8%[a] lower than the true spread. Error two is not as sensitive as error one is to the volatility. The main source of error two is the number of observations in a time interval. Error two could be eliminated with a sufficiently large number of the observations in an interval. *Bridge* suggests that e_{v2} ranges from the scale of 10^{-7} to 10^{-6}, which is much lower than the true spread but much greater than the ones in the previous case. Furthermore, *BridgeStd*, which is the the standard deviation of *Bridge*, ranges from the scale of 10^{-7} to 10^{-4} and is much greater than that in the previous section. When the time interval is longer than one hour, *BridgeStd* becomes at least more than 16.9% of the true spread, and according to Equation (2.57), the influence of could be greater than 40.8%[c]. Therefore although both errors one and two make the estimator underestimate the true spread, in the 12-hour and the 24-hour cases, the CS estimator finally overestimates the true spread.

To sum up, the only change in the conditions of the system in this section is in the volatility of mid-prices. Because of the higher volatility, all the estimators perform worse. The Roll estimator can be used for short time intervals. The HS estimator is still stable and has acceptable results. The CS estimator is very sensitive to the change of the volatility and the results are far away from the true spread. Similar to those shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator at low frequencies.

[a] $2 \times (1 - 96.0\%) = 8\%$

[c] Because of the non-linearity of Equation (2.57), one can only has an approximation of the influence of e_γ. When $e_\gamma = e_{v2} = 16.9\%$, the influence of the error is close to $\frac{16.9\%}{\sqrt{3-2\sqrt{2}}} \approx 40.8\%$

Table 2.3: FIXED Spread and High Returns Volatility

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-2}$	0.0200	0.0447	0.0775	0.110	0.155	0.310	0.536	0.758
				Roll 1984				
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.292	0.265	0.422	0.921	1.58
Relative Estimate	1.000	1.000	1.000	0.973	0.883	1.407	3.070	5.267
Est-Std$\times 10^{-3}$	0.000886	0.00587	0.0246	0.0729	0.184	0.471	1.05	1.77
T-test $\times 10^{-3}$	1.26e-05 (0.45)	1.38e-05 (0.07)	-3.56e-04 (-0.46)	-0.00772*** (-3.35)	-0.0349*** (-5.99)	0.122*** (8.21)	0.621*** (18.76)	1.28*** (22.96)
RMSE$\times 10^{-3}$	0.000886	0.00587	0.0246	0.0733	0.187	0.487	1.22	2.18
				Huang and Stoll 1997				
Estimates $\times 10^{-3}$	0.300	0.300	0.301	0.301	0.298	0.298	0.295	0.283
Relative Estimate	1.000	1.000	1.003	1.003	0.993	0.993	0.983	0.943
Est-Std$\times 10^{-4}$	0.00622	0.0296	0.0898	0.176	0.364	1.42	4.21	8.80
T-test $\times 10^{-5}$	3.53e-05 (0.02)	4.17e-03 (0.45)	0.0738** (2.60)	0.0847 (1.53)	-0.154 (-1.34)	-0.242 (-0.54)	-0.527 (-0.40)	-1.73 (-0.62)
RMSE$\times 10^{-3}$	0.000622	0.00296	0.00904	0.0176	0.0365	0.142	0.421	0.880
				Corwin and Schultz 2010				
Estimates $\times 10^{-3}$		-0.143	-0.0478	0.00979	0.0740	0.258	0.492	0.732
Relative Estimate		-0.477	-0.159	0.033	0.247	0.860	1.640	2.440
Est-Std$\times 10^{-3}$		0.00167	0.00447	0.00806	0.0170	0.0665	0.207	0.397
T-test$\times 10^{-3}$		-0.443*** (-8377.83)	-0.348*** (-2460.10)	-0.290*** (-1139.12)	-0.226*** (-421.31)	-0.0422*** (-20.05)	0.192*** (29.27)	0.432*** (34.36)
RMSE$\times 10^{-3}$		0.443	0.348	0.290	0.227	0.0787	0.282	0.587
$\widehat{SP}$ $\times 10^{-3}$		0.128	0.152	0.160	0.165	0.171	0.174	0.174
η		0.868	0.919	0.933	0.942	0.954	0.959	0.960
Bridge $\times 10^{-5}$		-0.0156	-0.0345	-0.0527	-0.0783	-0.165	-0.292	-0.416
BridgeStd$\times 10^{-4}$		0.00466	0.0131	0.0257	0.0509	0.202	0.603	1.20

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}, i.e. $\Delta M_t \sim N(0, 4 \times 10^{-8})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected. The other settings are the same as table (2.2)

2.6.3 Time-varying Spreads

In this section, the environment is the same as the one in the previous section except that we let the spread follow a normal distribution with a mean of 0.0003 and a standard deviation of 0.0001, i.e. $N\left(0.0003, 10^{-8}\right)$. Thus all the differences in the results of the estimators can be imputed to the time-varying spread. In this section, order flow is random and mid-prices follow a random walk. Under these circumstances, the HS estimator is unbiased, both the Roll and the CS estimators should have another error. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process and the standard deviations of mid-price returns is $\sigma = 0.0002$. $\mu = 0.0003$, $\varsigma^2 = 0.0001$ which suggests that the spread is time-varying. The system is given as follows:

$$\begin{aligned} & BS_t = \varpi_t \\ & \varpi_t \sim B(1, 0.5) \\ & \Delta M_t = \varepsilon_t \\ & \varepsilon_t \sim N(0, 4 \times 10^{-8}) \\ & SP_t \sim N\left(0.0003, 10^{-8}\right) \\ & s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t \end{aligned} \tag{2.85}$$

The results are presented in Table (2.4), of which the contents are similar to Table (2.2). Because most settings are the same and thus some results are not very different from those in Table (2.3), the analysis here emphasises the differences between the two sections rather than the full description of the table.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns have a positive relationship. In the tick-by-tick case, the average standard deviation of mid-price returns is 2×10^{-4} which is the same as the baseline setting of the system. In the 24-hour case, the standard deviation is 7.58×10^{-3}. Thus the ratio of the spread to the standard deviation ranges from 1.5 to 0.0396.

According to the first row of the panel *Roll*, the estimates of the Roll estimator are slightly greater (1.6×10^{-5}) than the ones in the previous section, which is caused by the time-varying spread (error one). Error one suggests that the Roll estimator will over-estimate the spread when the spread is time-varying. In the tick-by-tick case, the estimate is 3.16×10^{-4}, and the covariance of the transaction price returns is $Cov(\Delta s_t, \Delta s_{t-1}) = -2.5 \times 10^{-8}$. The variance of the spread is $E(u^2) = 10^{-8}$, which is the same as in the baseline setting of the system. According to Equation (2.12), one can find the result is 0.0003[d] which is the same as the one in the previous section and equals the true spread. Furthermore the error is 0.00016 which coincides with the prediction of Equation (2.13). The prediction is also valid for other time intervals. Thus, all the differences between this and previous sections are caused by the time-varying spread and can be explained by Equation (2.13). When the volatility of the spread is low, the Roll estimator will have acceptable results. The other contents of the table are similar to those of Table (2.2), which has already been discussed above. The Roll estimator has larger estimates compared to those in the previous section. The differences are caused by the time-variability of the spread. When the volatility of the spread is low, the differences can be ignored.

The results of the HS estimator are not very different from the ones in the previous section. Because of the linearity of the HS estimator, it can obtain the correct value from the time-varying spread. The HS estimator does estimate the spread correctly, according to the first row of the panel *Huang and Stoll*. The error becomes greater as the time interval become longer. In the 24-hour case, the error, which is the largest, is equal to -1.22×10^{-5} which is less than 5% of the true spread. The second row suggests that the standard deviations of the estimated spreads are slightly greater than those in the previous section, but the differences are not quite large (less than 1% of the standard deviation). The standard deviations are less than those of the Roll estimator, which suggests that the results are stable across the replications and are more reliable than those of the Roll estimator. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one. To sum up, the time-varying spread does not bring extra errors to the HS estimator. The HS estimator performs well in this section and better than the Roll estimator. The reason why the HS estimator performs better than the Roll estimator does is that the former uses more information (order flow) than the latter, and thus avoids the non-linearity of the Roll estimator.

Similar to the Roll estimator, the CS estimator has slightly larger estimates compared

---

[d] $SP = 2 \cdot \sqrt{-(-2.5 \times 10^{-8}) - \frac{10^{-8}}{4}} = 0.0003$

to the previous section because of the time-varying spread and the non-linearity. The error is positively correlated to the volatility of the spread. All the differences between this section and the previous one are caused by the time-varying spread. The CS estimator does not yield good estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. In the five-minute and the fifteen-minute cases, the CS estimator has negative results, which might be caused by the joint effect of errors. In the 24-hour case, the estimator significantly overestimates the true spread: the estimate is about 250% of the true spread. The standard deviations of the results are the smallest among all the estimators, according to the second row in the panel. The total errors are significantly different from zero, and cannot be ignored. $\widehat{SP}$ ranges from 1.40×10^{-4} to 1.94×10^{-4}, which is slightly greater than those in previous section, which suggests the estimator is better in terms of error one. In other words, the time-varying spread increase the proportion (about 4%) of observations that satisfy the assumption that the highest (lowest) transaction price corresponds the highest (lowest) mid-price. η, *Bridge* and *BridgeStd* are not very different from the previous section, which suggests that the time-varying spread does not have big influence on errors two and three. For brevity, the analysis in previous section is not repeated. One can conclude that the time-varying spread makes the CS estimator have slightly larger estimates than the fixed spread through error one. The results are far away from the true spread, thus the estimator should not be used in this type of environments.

To sum up, the only change of the conditions of the system in this section is the time-variability of the spread. Because of the time-varying spread, both the Roll and the CS estimators have slightly larger estimates can be used for short time intervals. The HS estimator is not influenced by the time-varying spread. Similar to those shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator at low frequencies.

Table 2.4: Random Spreads

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-2}$	0.0200	0.0447	0.0775	0.110	0.155	0.310	0.536	0.758
Roll 1984								
Estimates $\times 10^{-3}$	0.316	0.316	0.315	0.310	0.286	0.454	0.958	1.70
Relative Estimate	1.053	1.053	1.05	1.033	0.953	1.513	3.193	5.667
Est-Std$\times 10^{-3}$	0.000923	0.00540	0.0241	0.0732	0.178	0.476	1.03	1.79
T-test $\times 10^{-3}$	0.0162***	0.0164***	0.0148***	0.00960***	-0.0138*	0.154***	0.658***	0.00140***
	(559.12)	(96.36)	(19.33)	(4.15)	(-2.44)	(10.23)	(20.27)	(24.71)
RMSE$\times 10^{-3}$	0.0160	0.0169	0.0284	0.0739	0.179	0.500	1.222	2.272
Huang and Stoll 1997								
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.301	0.300	0.301	0.302	0.288
Relative Estimate	1	1	1	1.003	1	1.003	1.007	0.96
Est-Std$\times 10^{-4}$	0.00627	0.0306	0.0901	0.178	0.385	1.49	4.44	8.86
T-test $\times 10^{-4}$	5.68e-05	-8.29e-05	-0.00186	0.00432	0.00387	0.0114	0.0151	-0.121
	(0.30)	(-0.09)	(-0.65)	(0.77)	(0.32)	(0.24)	(0.11)	(-0.43)
RMSE$\times 10^{-3}$	0.000627	0.00306	0.00901	0.0178	0.0385	0.149	0.444	0.886
Corwin and Schultz 2010								
Estimates $\times 10^{-3}$		-0.136	-0.0358	0.0236	0.0898	0.277	0.527	0.781
Relative Estimate		-0.453	-0.119	0.0787	0.299	0.923	1.7567	2.603
Est-Std $\times 10^{-4}$		0.0166	0.0443	0.855	0.172	0.664	2.01	3.91
T-test $\times 10^{-3}$		-0.436***	-0.336***	-0.276***	-0.210***	-0.0226***	0.227***	0.481***
		(-8275.56)	(-2388.65)	(-1017.18)	(-386.90)	(-10.74)	(35.75)	(38.83)
RMSE$\times 10^{-3}$		0.436	0.336	0.277	0.210	0.0230	0.227	0.482
$\widehat{SP}$ $\times 10^{-3}$		0.140	0.168	0.177	0.183	0.191	0.193	0.194
η		0.861	0.913	0.928	0.937	0.947	0.950	0.949
Bridge $\times 10^{-5}$		-0.0156	-0.0345	-0.0527	-0.0783	-0.165	-0.292	-0.417
BridgeStd $\times 10^{-4}$		0.00466	0.0131	0.0257	0.0509	0.202	0.601	1.20

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}, i.e. $\Delta M_t \sim N(0, 4 \times 10^{-8})$. The spread is time-varying and follows a normal distribution which mean is 0.0003 and standard deviation is 10^{-4}, i.e. $SP_t \sim N(0.0003, 10^{-8})$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four-hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.

The other settings are the same as Table (2.2)

2.6.4 Auto Mid-Price Returns

In this section, most settings are the same as in Section (2.6.2) except that mid-price returns are autocorrelated. Thus all the differences in the performance of the estimators can be imputed to autocorrelation. Let $\tau = 1$ in Equation (2.81), which suggests that the mid-price is autocorrelated. The coefficient is given by $\zeta = -0.3303$. Thus, the mid-price returns are negatively autocorrelated, which coincide with the results in Goodhart et al. (1996) and Daníelsson and Payne (2002), where mid-point returns of both the EFX and the Reuters D2000-2 system are shown to be negatively autocorrelated, although the autocorrelation is weaker in the Reuters D2000-2 system. In this section, order flow is random; mid-prices are autocorrelated; and the spread is fixed. Under these conditions, both the CS and the Roll estimator are biased, and the error of the HS estimator is unbiased. Because the autocorrelation reduces the volatility of mid-prices, the CS estimator may have less errors compared to Section (2.6.2). Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\mu = 0.0003$, $\varsigma^2 = 0$ which suggests that the spread is a constant. The system is given as follows:

$$
\begin{aligned}
&BS_t = \varpi_t \\
&\varpi_t \sim B(1, 0.5) \\
&\Delta M_t = -0.3303\,\Delta M_{t-1} + \varepsilon_t \\
&\varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned}
\tag{2.86}
$$

The results are presented in Table (2.5), whose contents are similar to Table (2.2). *CovarianceMid* reports the average covariance of mid-price returns. Because most settings are the same and thus many results are not very different from those in Table (2.3), the analysis here emphasises the differences between the sections rather than the full description of the table.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2.12×10^{-4} which is similar to the standard deviation of the white noise ε. In the 24-hour case, the standard deviation is 5.69×10^{-3} which is slightly less than that in Section (2.6.2) because of the autocorrelation. Thus the ratio of the spread to the standard deviation ranges from 1.41 to 0.0527.

The Roll estimator overestimates the spread in all time intervals and the errors are significantly different from zero, according to the first and third rows of panel *Roll*. All the differences between this section and Section (2.6.2) are caused by the autocorrelated mid-price returns and thus related to error two of the Roll estimator. Error two suggests that when the mid-price returns are negatively (positively) autocorrelated, the Roll estimator will overestimate (underestimate) the spread. In this section, the mid-price returns are negatively autocorrelated and the estimator overestimates the spread, which is consistent with the prediction of the theoretical analysis. In the tick-by-tick case, the estimate is 3.86×10^{-4}; covariance of the transaction price returns is 3.725×10^{-8}; and the covariance of mid-price returns is -1.48×10^{-8}. If Equation (2.14), which holds when there is autocorrelation, is used, the estimated spread is 0.0003 [c]. Thus, the error is 8.6×10^{-5}, which can be fully explained by Equation (2.15). The findings for other time intervals are similar. As the time interval lengthens, the autocorrelation becomes weaker, as results of which, the error is less in cases of longer time intervals. The estimations are all greater than the ones in Section (2.6.2). The standard deviations of estimates are slightly lower, according to the second row of the penal, the reason of which may be that the autocorrelation makes the volatility of the mid-price returns less than the one when mid-prices returns are random. When the time interval is longer than one hour, the standard deviation of the estimate is close to the estimate itself, which suggests the estimations are not stable for long time intervals. To sum up, the Roll estimator performs worse than it does in Section (2.6.2) because of autocorrelation of the mid-price returns. It overestimates the true spread by at least 20% and is not stable for long time intervals.

The HS estimator does not produce much different results from Section (2.6.2), which suggests that the estimator is not influenced by the autocorrelation of mid-price returns. The HS estimator does estimate the spread correctly in most cases, according to the first row of the panel *Huang and Stoll*. The error becomes greater as the time interval become longer. All the errors are not significant different from zeros. The estimator becomes unstable when the time interval us longer than one hour, because the standard deviation

[c] $2 \cdot \sqrt{-1.48 \times 10^{-8} - (-3.725 \times 10^{-8})} = 0.0003$

of the estimates are greater than the true spread. In the 24-hour case, the error, which is the largest, is equal to -1.73×10^{-5} which is less than 6% of the true spread. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. In the 4-hour case, the standard deviation is greater than 33% of the spread, which suggests that the estimator is not quite stable in this case. Compared to Section (2.6.2), the HS estimator, become more stable, which could be imputed to the autocorrelation which lowers the volatility of the mid-price returns. If there were more observations, the HS estimator will have a better performance. The standard deviations are less than those of the Roll estimator, which suggests that the results are stable across the replications and are more reliable. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one. Therefore, one can conclude that the HS estimator performs well in this section and performs better than the Roll estimator. The reason why the HS estimator performs better than the Roll estimator does is that the former uses more information (order flow) than the latter. The autocorrelation of mid-price returns does not influence the HS estimator greatly. The performance of the estimator is roughly the same as that in Section (2.6.2) and slight more stable than that in Section (2.6.2).

Compared to Section (2.6.2), the CS estimator has roughly the same results, and because of the autocorrelation of mid-price returns which lower the volatility of mid-price returns, the estimator performs slightly better than in Section (2.6.2): less errors (including the particular errors), more stable. Although the autocorrelation breaks Parkinson's assumption, which introduces biases to the estimation of the volatility of mid-price returns, it is good for spread estimation. The following findings are roughly the same as the ones in Section (2.6.2). The CS estimator has slightly larger estimates for short time intervals and slightly less results in longer time intervals, and does not yield good estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. It can be seen that, the estimate is an increasing function of time interval (or the volatility of prices). In five-minute case, the CS estimator has negative results, which might be caused by the joint effect of errors. In the 24-hour case, the estimator significantly overestimates the true spread: the estimate is about 230% of the true spread. The standard deviations of the results are the smallest among the estimators and are less than those in Section (2.6.2), according to the second row in the panel. The total errors, though less than those in Section (2.6.2), are significantly different from zero, and cannot be ignored, according to the third row. $\widehat{SP}$ ranges from 1.32×10^{-4} to 1.79×10^{-4}, which is much less than the true spread and implies that at least 40% (in the 24-hour case) of the observations are not satisfied the assumption that the highest (lowest) observed prices are corresponding to the highest (lowest) mid-prices, and thus error one causes the estimator to underestimate the true spread at least 40%. Compared to Section (2.6.2), the volatility is lower because of the autocorrelation, and thus $\widehat{SP}$ is slightly greater. The results of error one suggest that when the volatility of prices is high, the CS estimator will never obtain an accurate estimation. η ranges from 88.1% to 96.8%, which implies that the percentages of the highest (lowest) prices are with buy (sell) orders are higher than those in Section (2.6.2) because of the lower volatility caused by the autocorrelation. Error two makes the estimation be at least 6.4%[a] less than the true spread. Both *Bridge* (e_{v2}), which ranges from the scale of 10^{-7} to 10^{-6}, and *BridgeStd*, which is the the standard deviation of *Bridge* and ranges from the scale of 10^{-7} to 10^{-5}, are less than those in Section (2.6.2). When the time interval is longer than one-hour, *BridgeStd* becomes at least more than 10.1% of the true spread, and according to Equation (2.57), the influence of could be greater than 24.4%[f]. Therefore although both errors one and two make the estimator underestimate the true spread, in 12-hour and 24-hour cases, the CS estimator overall overestimates the true spread. One can conclude that because of the autocorrelation of mid-price returns the CS estimator performs better here than in Section (2.6.2). However, the results are far away from the true spread, thus it should not be usedi this type of environments.

To sum up, the only change of the conditions of the system in this section is related to the autocorrelation of mid-price returns. Because of the lower volatility caused by the autocorrelation, the Roll estimator significantly overestimates the true spread compared to Section (2.6.2); and both the HS and CS estimators perform better than those in Section (2.6.2). Similar to those shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS

[a] $2 \times (1 - 96.8\%) = 6.4\%$

[f] Because of the non-linearity of Equation (2.57), one can only has an approximation of the influence of e_γ. When $e_\gamma = e_{v2} = 10.1\%$, the influence of the error is close to $\frac{10.1\%}{\sqrt{3-2\sqrt{2}}} \approx 24.4\%$

estimator at low frequencies.

Table 2.5: Autocorrelated Mid-price Returns

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-3}$	0.212	0.360	0.596	0.833	1.17	2.33	4.03	5.69
CovarianceMid$\times 10^{-8}$	-1.48	-0.846	-0.834	-0.837	-0.820	-1.28	-1.57	-5.80
				Roll 1984				
Estimates $\times 10^{-3}$	0.386	0.352	0.351	0.350	0.331	0.382	0.694	1.17
Relative Estimate	1.287	1.173	1.17	1.167	1.103	1.273	2.313	3.9
Est-Std$\times 10^{-3}$	0.000856	0.00352	0.0136	0.0367	0.122	0.368	0.793	1.34
T-test $\times 10^{-3}$	0.0864***	0.0518***	0.0508***	0.0497***	0.0313***	0.0819***	0.394***	0.873***
	(3193.12)	(466.15)	(118.35)	(42.82)	(8.08)	(7.04)	(15.72)	(20.62)
RMSE $\times 10^{-3}$	0.0860	0.0521	0.0528	0.0620	0.126	0.377	0.885	1.598
				Huang and Stoll 1997				
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.300	0.301	0.301	0.315	0.329
Relative Estimate	1	1	1	1	1.003	1.003	1.05	1.0967
Est-Std$\times 10^{-4}$	0.00632	0.0255	0.0711	0.143	0.282	1.08	3.28	6.73
T-test $\times 10^{-5}$	-0.00143	-0.00507	0.00144	0.0477	0.0982	0.0520	1.48	2.89
	(-0.72)	(-0.63)	(0.06)	(1.06)	(1.10)	(0.15)	(1.42)	(1.36)
RMSE $\times 10^{-3}$	0.000632	0.00255	0.00711	0.0143	0.0282	0.108	0.328	0.674
				Corwin and Schultz 2010				
Estimates $\times 10^{-3}$		-0.0325	0.0734	0.128	0.185	0.333	0.528	0.704
Relative Estimate		-0.108	0.245	0.427	0.617	1.11	1.76	2.347
Est-Std$\times 10^{-3}$		0.00135	0.00346	0.00645	0.0133	0.0508	0.149	0.300
T-test$\times 10^{-3}$		-0.332***	-0.227***	-0.172***	-0.115***	0.0326***	0.228***	0.404***
		(-7761.46)	(-2072.74)	(-842.27)	(-273.42)	(20.28)	(48.15)	(42.60)
RMSE$\times 10^{-3}$		0.333	0.227	0.172	0.116	0.0606	0.272	0.503
$\widehat{SP}$ $\times 10^{-3}$		0.132	0.158	0.165	0.171	0.177	0.179	0.179
η		0.881	0.935	0.948	0.956	0.963	0.966	0.968
Bridge$\times 10^{-5}$		-0.0143	-0.0291	-0.0430	-0.0624	-0.128	-0.224	-0.317
BridgeStd $\times 10^{-5}$		0.0326	0.0834	0.158	0.304	1.17	3.45	6.85

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is autocorrelated and is obtained from $\Delta M_t = -0.3303\,\Delta M_{t-1} + \varepsilon_t$, where ε is a random noise of which the mean is zero and the variance is 4×10^{-8}, i.e. $\varepsilon_t \sim N(0, 4 \times 10^{-8})$. The spread is fixed and equals to 0.0003. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.

CovarianceMid reports the average covariance of mid-price returns.

The other settings are the same as Table (2.2)

2.6.5 Big Spreads

In this section, most settings are the same as in Section (2.6.2) except that here we introduce a higher value for the spread. Thus all the differences in the performance of the estimators can be imputed to the bigger spread. According to the differences between Sections (2.6.1) and (2.6.2), that is the greater volatility of mid-price returns, one finds that the volatility makes the performances of estimators worse. In this section, the results derived under the restriction that the spread is large will show that the ratio of the spread to the volatility of the mid-price returns influences the performance of the estimators rather than the volatility itself. The spread is fixed and equals to 0.03 which is much greater than that in Section (2.6.2). In this section, order flow is random; mid-prices follow a random walk; and the spread is fixed and big. Under these circumstances, both the HS and the Roll estimators are unbiased. The error of the CS estimator should be small, because the SP is large and thus the error caused by the volatility of mid-price returns can be ignored. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process. The standard deviations of mid-price returns is $\sigma = 0.0002$. $\mu = 0.03$, $\varsigma^2 = 0$ which suggests that the spread is a constant. The system is given as follows:

$$
\begin{aligned}
& BS_t = \varpi_t \\
& \varpi_t \sim B(1, 0.5) \\
& \Delta M_t = \varepsilon_t \\
& \varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
& SP_t = 0.03 \\
& s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned}
\tag{2.87}
$$

The results are presented in Table (2.6), whose contents are similar to those of Table (2.2). Because most settings are the same and thus the results are not very different from those in Table (2.3), the analysis here emphasises the differences between the sections rather than the full description of the table for brevity.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2.00×10^{-4} which is similar to the standard deviation of the white noise ε. In the 24-hour case, the standard deviation is 7.58×10^{-3} which is the same as that in Section (2.6.2). The ratio of the spread to the standard deviation ranges from 150 to 3.96.

The Roll estimator performs much better than it does in Section (2.6.2). Intuitively, because the spread is 100 times as big as that in Section (2.6.2), it is much easier for the estimator to obtain an accurate estimation. The Roll estimator does estimate the spread correctly, according to the first row of the panel *Roll*. The second row suggests that the standard deviation of the estimated spreads is at most less than 6% of the spread (in the 24-hour case), which suggests that the results are stable across the replications. The third row suggests that the results of the t-test suggest that the estimated spreads are not significantly different from the true spreads except for the one of tick-by-tick case. The error in the tick-by-tick case is 3.65×10^{-6} and is much less than the true spread. Compared to Section (2.6.1) where both the spread and mid-price volatility are much less but similar ratio of of the spread to the standard deviation of mid-price returns, the results are quite similar: the Roll estimator performs accurately; the standard deviations of the estimates are much small relative to the spread, and thus the estimator is stable; the error is much less than the spread. Comparing the performance of the estimator between the three sections, one finds that ratio of the spread to the standard deviation of mid-price returns is one of the important factors that influences the performance of the estimator. One can conclude that the Roll estimator performs well when the ratio is large enough.

The HS estimator performs better than it does in Section (2.6.2). The HS estimator does estimate the spread correctly even in the longer time intervals, according to the first row of the panel *Huang and Stoll*. The second row suggests that the standard deviation of the estimated spreads increases as the time interval lengthens. In the 24-hour case, the standard deviation is about 0.3% of the spread. The standard deviations are much lower than those of the Roll estimator, which suggests that the results are stable across the replications and are more reliable than the Roll estimator. The third row suggests that the estimated spreads are not significantly different from the true one except for the fifteen-minute case. The error is 10^{-7} which is much less than the true spread. Similar to the Roll estimator, the HS estimator performs better when the ratio of the spread to the standard deviation of mid-price returns is big. The HS estimator performs better than the Roll estimator. The reason why the HS estimator performs better than the Roll estimator is that the former uses more information (order flow) than the latter.

The CS estimator performs better than it does in Section (2.6.2). The CS estimator offers estimation of the true spread, according to the first row of the panel *Corwin and Schultz*. The total error ranges from -4.32×10^{-3} to -2.47×10^{-4}, which is at most 14.4% of the true spread, and is significantly different from zero. When the time interval is longer than five minutes, the estimator is very close to the true spread. The standard deviation of the estimates is 5.51×10^{-4} which is much less than the true spread, which suggests that the estimator is stable in this case. Both errors one and two are very similar to those in Table (2.6.1). In Section (2.6.1), errors one and two are small compared to those in Section (2.6.2) because the volatility of mid-price returns is small. In this section, the reason is that the true spread is big. Thus the ratio of the spread to the standard deviation of mid-price returns is negatively correlated to errors one and two. Although *Bridge* and *BridgeStd* are as large as those in Section (2.6.2), they are much lower than the true spread in this case, and thus can be neglected in all time intervals in this section. Therefore, the ratio does not influences error three, but when the volatility is fixed, the big spread, i.e. big ratio, can undermine the influence of error three. One can conclude that the CS estimator performs well when the ratio is big enough and the time interval is long.

To sum up, the ratio of the spread and the standard deviation of mid-price returns is an important factor that influences the performance of the estimators. The ratio is negatively correlated to the errors. In this section, both the Roll and HS estimators have unbiased estimates. The estimation of the CS estimator is acceptable (the error is less than 1.67%) except for the case of five-minute intervals. In terms of the standard deviations of the estimations, the CS estimator has the best performance and the Roll estimator has the highest standard deviation. Similar to those shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of 12 hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator at low frequencies. All the estimators perform reliably when the ratio is big, fixed spreads, random mid-price returns, and random order flow.

Table 2.6: Fixed Big Spread

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-3}$	0.200	0.447	0.775	1.10	1.55	3.10	5.36	7.58
				Roll 1984				
Estimates $\times 10^{-1}$	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Relative Estimate	1	1	1	1	1	1	1	1
Est-Std $\times 10^{-3}$	0.0520	0.113	0.195	0.283	0.410	0.807	1.47	2.02
T-test $\times 10^{-5}$	0.365*	0.313	0.150	0.587	-0.902	-1.05	0.856	1.22
	(2.22)	(0.88)	(0.24)	(0.66)	(-0.70)	(-0.41)	(0.18)	(0.19)
RMSE$\times 10^{-3}$	0.0520	0.113	0.195	0.283	0.410	0.807	1.47	2.02
				Huang and Stoll 1997				
Estimates $\times 10^{-1}$	0.300	0.300	0.300	0.300	0.300	0.300	0.300	0.300
Relative Estimate	1	1	1	1	1	1	1	1
Est-Std$\times 10^{-3}$	0.000621	0.00303	0.00898	0.0183	0.0366	0.136	0.426	0.871
T-test $\times 10^{-5}$	-0.000486	0.0175	0.0720*	0.0829	-0.0498	0.113	-0.453	-1.06
	(-0.25)	(1.82)	(2.53)	(1.43)	(-0.43)	(0.26)	(-0.34)	(-0.39)
RMSE$\times 10^{-3}$	0.000621	0.00303	0.00898	0.0183	0.0366	0.136	0.426	0.871
				Corwin and Schultz 2010				
Estimates $\times 10^{-1}$		0.257	0.295	0.296	0.296	0.296	0.297	0.298
Relative Estimate		0.857	0.983	0.987	0.987	0.987	0.99	0.993
Est-Std$\times 10^{-3}$		0.0515	0.00558	0.00928	0.0177	0.0722	0.210	0.401
T-test$\times 10^{-3}$		-4.32***	-0.479***	-0.436***	-0.408***	-0.356***	-0.299***	-0.247***
		(-2652.43)	(-2714.35)	(-1485.23)	(-727.77)	(-155.83)	(-45.05)	(-19.44)
RMSE$\times 10^{-3}$		4.300	0.500	0.400	0.400	0.406	0.366	0.448
$\widehat{SP} \times 10^{-1}$		0.279	0.298	0.298	0.298	0.298	0.298	0.298
η		0.969	1.000	1.000	1.000	1.000	0.999	0.998
Bridge $\times 10^{-5}$		-0. 0156	-0. 0345	-0. 0527	-0. 0783	-0. 165	-0. 292	-0. 417
BridgeStd $\times 10^{-4}$		0. 00466	0. 0131	0. 0257	0. 0510	0. 202	0. 602	1.20

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 9×10^{-12}, i.e. $\Delta M_t \sim N(0, 9 \times 10^{-12})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.03$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.

The other settings are the same as Table (2.2)

2.6.6 Inventory Control Costs and Asymmetric information Costs

In this section, most settings are the same as in Section (2.6.2) except that the mid-price return is now influenced by the past order, thus capturing the inventory control and the asymmetric information (IC & AS) components of the spread. Because we focus on the spread estimation rather than the spread decomposition, we do not distinguish the IC & AS costs. Thus all the differences in the performance of the estimators can be imputed to the existence of these components. Let $\omega = 1$ in Equation (2.81), which suggests that the mid-price is influenced by the past order flow. The coefficient is given by $\chi = \frac{1}{3}$ which suggests that the IC & AS costs contribute one third of the total spread. In this section, order flow is random; mid-prices are influenced by the past order flow; and the spread is fixed. Under these conditions, the Roll and the CS estimators are biased, and the error of the HS estimator is unbiased. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\mu = 0.0003$, $\varsigma^2 = 0$ which suggests that the spread is a constant. The system is given as follows:

$$
\begin{aligned}
&BS_t = \varpi_t \\
&\varpi_t \sim B(1, 0.5) \\
&\Delta M_t = \tfrac{1}{3} BS_{t-1} \cdot \tfrac{SP_t}{2} + \varepsilon_t \\
&\varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned}
\tag{2.88}
$$

The results are presented in Table (2.7), whose the contents are similar to those of Table (2.2). ϱ reports the coefficient of Equation (2.19) and represents the proportion of the IC & AS components of the spread. Because most settings are the same and thus the results are not very different from those in Table (2.3), the analysis here emphasises the differences between the sections rather than the full description of the table.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2.06×10^{-4} which is similar to the standard deviation of the white noise ε. In the 24-hour case, the standard deviation is 7.81×10^{-3}. Thus the ratio of the spread to the standard deviation ranges from 1.46 to 0.0384. ϱ is about $\frac{1}{3}$, which is the same as in the baseline setting, when the time interval is short, which suggests that the IC & AS components carry one third weight of the spread. When the time interval is longer than one hour, ϱ becomes unstable, because in relatively long runs the microstructure effects are weaker.

The Roll estimator underestimates the spread for short time intervals and overestimates the spread for long time intervals. The errors are significantly different from zero, according to the first and third rows of panel *Roll*. All the differences between this section and Section (2.6.2) are caused by the existence of the IC & AS components and thus related to error four of the Roll estimator. Error four suggests that the IC & AS components of the spread will cause the Roll estimator to underestimate the spread. The results are the same as the theoretical prediction. In the tick-by-tick case, the estimate is 2.45×10^{-4}; covariance of the transaction price returns is -1.50×10^{-8}; and the fraction of IC& AS components is $\frac{1}{3}$. If Equation (2.21), is used, which holds when the IC & AS components exist, the estimated spread is 0.0003 [g]. Thus, the error is -5.5×10^{-5}, which can be fully explained by Equation (2.21). The findings are the same for other time intervals. The standard deviations of estimates are slightly greater than those in Section (2.6.2), according to the second row of the penal. When the time interval is longer than one hour, the standard deviation of the estimate is close to the estimate itself, which suggests the estimations are not stable for long time intervals. To sum up, the Roll estimator underestimates the spread because it does not consider its IC & AS components. It underestimates the true spread at least 20% and is not stable for long time intervals.

The HS estimator does not show much different results from Section (2.6.2), which suggests that the estimator is not influenced by the existence of the IC & AS components. The reason is that the HS estimator is designed for decomposing the components of the spread, and incorporates the transaction costs and IC & AS components into the model. Therefore, for the HS estimator, Section (2.6.2) is a special case where the IC & AS components are zero. The results in Table (2.7) are not very different from those in Table (2.3), thus we do not repeat the analysis. Generally speaking, the HS estimator is not influenced by the IC & AS components of the spread. It performs better than the Roll estimator does.

---

[g] $2 \cdot \sqrt{\frac{-(-1.5 \times 10^{-8})}{\left(1-\frac{1}{3}\right)}} = 0.0003$

Compared to Section (2.6.2), the CS estimator produces less estimates, the reason of which is similar to the one offered for the Roll estimator: the IC & AS components are ignored. The CS estimator does not yield good estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. In five-minute fifteen- minute and thirty-minute cases, the CS estimator has negative results. In the 24-hour case, the estimator significantly overestimates the true spread: the estimate is about 230% of the true spread. The standard deviations of the results are the smallest among the estimators and are slightly greater than those in Section (2.6.2), according to the second row in the panel. The total errors are greater than those in Section (2.6.2), and cannot be ignored, according to the third row. $\widehat{SP}$ ranges from 8.81×10^{-5} to 1.23×10^{-4}, which is much less than the true spread and is about 70%, which is greater than but close to the proportion of the other components of the spread in this section. The fact suggests that the missing the IC & AS components influence the estimator mainly through error one. η ranges from 83.1% to 93.0%, which about 3% less than that in Section (2.6.2). The differences of errors one and two between this section and Section (2.6.2) can explain most of the differences in the performance of the estimator between these two sections. Both *Bridge* (e_{v2}), which ranges from the scale of 10^{-7} to 10^{-6}, and *BridgeStd*, which is the standard deviation of *Bridge* and ranges from the scale of 10^{-7} to 10^{-4}, are not very different from those in Section (2.6.2). Thus the existence of IC & AS does not influence error three of the estimator. But error three still cannot be ignored, the analysis of which is presented in Section (2.6.2). One can conclude that the IC & AS components make the CS estimator obtain less estimates than those in Section (2.6.2). The results are far away from the true spread, thus it should not be used in this type of environments.

To sum up, the only change in the conditions of the system in this section has been the introduction of the IC & AS components of the spread. Because the components are not considered, the Roll estimator significantly underestimates the true spread and yields less estimates compared to Section (2.6.2); and the HS model's performances is similar to that in Section (2.6.2). The CS estimator, as the Roll estimator, produces less estimates than those in Section (2.6.2) and cannot estimate the spread correctly. Similar to the findings of Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator has at low frequencies.

Table 2.7: ASY INVEN

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-2}$	0.0206	0.0461	0.0798	0.113	0.160	0.319	0.553	0.781
ϱ	0.333	0.333	0.333	0.333	0.328	0.312	0.247	0.200
Roll 1984								
Estimates $\times 10^{-3}$	0.245	0.245	0.242	0.225	0.223	0.445	0.981	1.65
Relative Estimate	0.817	0.817	0.807	0.75	0.743	1.483	3.27	5.5
Est-Std $\times 10^{-3}$	0.000919	0.00681	0.0338	0.102	0.184	0.494	1.07	1.84
T-test $\times 10^{-3}$	-0.0551***	-0.0546***	-0.0576***	-0.0747***	-0.0767***	0.145***	0.681***	1.35***
	(-1895.19)	(-253.45)	(-53.95)	(-23.06)	(-13.16)	(9.29)	(20.12)	(23.18)
RMSE $\times 10^{-3}$	0.0550	0.0554	0.0671	0.126	0.199	0.515	1.268	2.282
Huang and Stoll 1997								
Estimates $\times 10^{-3}$	0.300	0.300	0.299	0.299	0.299	0.298	0.305	0.343
Relative Estimate	1	1	0.997	0.997	0.997	0.993	1.0167	1.143
Est-Std$\times 10^{-4}$	0.00622	0.0303	0.0906	0.186	0.396	1.52	4.46	9.05
T-test $\times 10^{-5}$	0.000743	0.00518	-0.0608*	-0.0759	-0.0712	-0.154	0.498	4.33
	(0.38)	(0.54)	(-2.12)	(-1.29)	(-0.57)	(-0.32)	(0.35)	(1.51)
RMSE$\times 10^{-3}$	0.000622	0.00303	0.00912	0.0186	0.0396	0.152	0.446	0.906
Corwin and Schultz 2010								
Estimates $\times 10^{-3}$		-0.178	-0.0915	-0.0361	0.0286	0.216	0.472	0.710
Relative Estimate		-0.593	-0.305	-0.120	0.0953	0.72	1.573	2.367
Est-Std $\times 10^{-3}$		0.00170	0.00446	0.00888	0.0168	0.0711	0.211	0.425
T-test $\times 10^{-3}$		-0.478***	-0.391***	-0.336***	-0.271***	-0.0837***	0.172***	0.410***
		(-8875.04)	(-2774.45)	(-1196.78)	(-511.84)	(-37.22)	(25.68)	(30.51)
RMSE$\times 10^{-3}$		0.478	0.392	0.336	0.272	0.110	0.272	0.591
$\widehat{SP}$ $\times 10^{-3}$		0.0881	0.106	0.112	0.116	0.121	0.123	0.123
η		0.831	0.884	0.901	0.913	0.924	0.929	0.930
Bridge $\times 10^{-5}$		-0.0161	-0.0356	-0.0544	-0.0807	-0.170	-0.301	-0.429
BridgeStd $\times 10^{-4}$		0.00491	0.0139	0.0273	0.0540	0.214	0.638	1.27

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is influenced by the past order flow and a random shock drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}. Thus there are the inventory control and the asymmetric information components of the spread. Formally, the mid-price returns are given as follows:, $\Delta M_t = \frac{1}{3} BS_{t-1} \cdot \frac{SP_t}{2} + \varepsilon_t$ where $\varepsilon_t \sim N(0, 4 \times 10^{-8})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.
ϱ reports the coefficient of the Equation (2.19) and represents the proportion of the inventory control and asymmetric information costs of the spread. The other settings are the same as Table (2.2)

2.6.7 Autocorrelated Order Flow

In this section, most settings are the same as those in the Section (2.6.2) with only exception the introduction positively autocorrelated order flow. Thus all the differences of the performance of the estimators can be imputed to the autocorrelation of order flow. In this section, order flow is positively autocorrelated; mid-prices follow a random walk; and the spread is fixed. Under these circumstances, the Roll estimator will underestimates the true spread as was mentioned in theoretical analysis. The HS estimator is unbiased; the CS estimator should have similar results to those in Section (2.6.2). Formally, let $\varphi = 1$, $\psi = 0$, $\phi = 0$ in Equation (2.79), which suggests that order flow is autocorrelated. Let the function $F(BS_{t-1})$ make sure $Pr(BS_t = BS_{t-1}) = 0.53$. In other words, let the probability of the order flow to be the same direction as the past one is 53%, which is slight light higher than 50%, and thus order flow is positively autocorrelated and the autocorrelation is not very strong. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process and $\mu = 0.0003$, $\varsigma^2 = 0$ which suggests that the spread is constant. The standard deviations of mid-price returns is $\sigma = 0.0002$. The system is given as follows:

$$\begin{aligned}
&Pr(BS_t = BS_{t-1}) = 0.53 \\
&\Delta M_t = \varepsilon_t \\
&\varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned} \tag{2.89}$$

The results are presented in Table (2.8), which is similar to Table (2.2). δ represents the probability of order flow continuing in the same direction. Because most settings are the same and thus the results are not very different from those in Table (2.3), the analysis here emphasises the differences between the sections rather than the full description of the table.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2×10^{-4} which is the same as in the baseline setting of the system. In the 24-hour case, the standard deviation is 7.58×10^{-3}. Thus the ratio of the spread to the standard deviation ranges from 1.5 to 0.0396. δ is 0.532 in tick-by-tick case, which is the same as the setting in this section. It equals to 0.500 in other time intervals,which suggests that the autocorrelation is too weak to influence the longer time intervals, and thus, there is no autocorrelations in longer time intervals.

The Roll estimator underestimates the true spread in the tick-by-tick case. The performance of the estimator is similar to that in Section (2.6.2): it does estimate the spread correctly for short time intervals, and underestimates the spread in the 30-minute case and in the one-hour case, then overestimates the spread in even longer time intervals. All the differences between this section and the Section (2.6.2) are caused by the autocorrelation of order flow and thus related to error three of the Roll estimator. To emphasise the influence of the autocorrelation, we focus on the tick-by-tick case and according to the theoretical analysis of error three of the Roll estimator, when the order flow is positively autocorrelated, the estimator will underestimates the true spread, which coincides with the empirical results of tick-by-tick case in this section. If δ were known, one can apply the Choi et al. (1988) version Roll model to obtain the true spread. In the tick-by-tick case, the covariance of the transaction price returns is 1.97×10^{-8}, $\delta = 0.532$, the estimated spread is 3.00×10^{-4}, according to Equation (2.17). Therefore, error three completely explains the error caused by the autocorrelation. In longer time intervals, the autocorrelation and its influence disappear, thus the Roll estimator performance is similar to that in Section (2.6.2). Given that most of the analysis is similar analysis to that in Section (2.6.2), it is not repeated here. To sum up, the Roll estimator can be influenced by the autocorrelation of order flow. However, in relatively long time intervals the autocorrelation and the influence of it becomes weak and the estimator can still obtain accurate results.

The HS estimator performance is similar to that in Section (2.6.2). The estimator estimates the spread correctly. It does not rely on the assumption that the order flow is random. Thus, it is not influenced by the autocorrelation of order flow. The results in Table (2.8) are not very different from those in Table (2.3), and thus we do not repeat the analysis. Generally speaking, the HS estimator can be used when order flow is autocorrelated. It performs better than the Roll estimator.

Like the HS estimator, the CS estimator does not assume order flow to be random. The results in this section are not very different from those in Section (2.6.2), and thus we do not repeat the analysis. Generally speaking, the CS estimator does not yield good estimates of the true spread, although it is not influenced by the autocorrelation of order flow, it should not be used under similar conditions.

To sum up, the only change of the conditions of the system in this section has been in the autocorrelation of order flow. The Roll estimator underestimates the true spread because of the autocorrelation, but when the autocorrelation is weak, it still can be used for short time intervals. The HS estimator is still stable and has acceptable results. The results of the CS estimator are far away from the true spread.Similar to the findings in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then either the HS or the Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator applied at low frequencies.

Table 2.8: Autocorrelated Order Flow

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-3}$	0.200	0.447	0.774	1.09	1.55	3.09	5.36	7.58
δ	0.532	0.500	0.500	0.500	0.500	0.500	0.500	0.500
				Roll 1984				
Estimates $\times 10^{-3}$	0.281	0.300	0.300	0.291	0.274	0.432	0.916	1.63
Relative Estimate	0.937	1.000	1.000	0.970	0.913	1.440	3.053	5.433
Est-Std$\times 10^{-3}$	0.000910	0.00556	0.0255	0.0791	0.176	0.472	1.04	1.75
T-test $\times 10^{-3}$	-0.0187*** (-650.73)	0.000350* (1.99)	-0.000310 (-0.38)	-0.00919*** (-3.67)	-0.0262*** (-4.70)	0.132*** (8.84)	0.616*** (18.68)	1.33*** (23.93)
RMSE$\times 10^{-3}$	0.0190	0.00556	0.0255	0.0796	0.1779	0.490	1.209	2.198
				Huang and Stoll 1997				
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.300	0.300	0.297	0.292	0.286
Relative Estimate	1.000	1.000	1.000	1.000	1.000	0.990	0.973	0.953
Est-Std $\times 10^{-4}$	0.00606	0.0317	0.0926	0.179	0.362	1.43	4.48	8.88
T-test $\times 10^{-5}$	0.00163 (0.85)	-0.00404 (-0.40)	0.0115 (0.39)	-0.0191 (-0.34)	0.0318 (0.28)	-0.339 (-0.75)	-0.790 (-0.56)	-1.36 (-0.48)
RMSE $\times 10^{-3}$	0.000606	0.00317	0.00926	0.0179	0.0362	0.1430	0.448	0.888
				Corwin and Schultz 2010				
Estimates $\times 10^{-3}$		-0.147	-0.0515	0.00649	0.0718	0.256	0.494	0.741
Relative Estimate		-0.49	-0.172	0.022	0.239	0.853	1.647	2.470
Est-Std $\times 10^{-3}$		0.0171	0.0446	0.0868	0.174	0.683	2.01	3.89
T-test$\times 10^{-3}$		-0.447*** (-8294.95)	-0.352*** (-2494.71)	-0.294*** (-1069.82)	-0.228*** (-414.90)	-0.0436*** (-20.21)	0.194*** (30.53)	0.441*** (35.82)
RMSE $\times 10^{-3}$		0.447	0.352	0.294	0.229	0.0812	0.279	0.588
$\widehat{SP}$ $\times 10^{-3}$		0. 123	0. 148	0. 156	0. 162	0. 168	0. 171	0. 172
η		0.857	0.912	0.928	0.937	0.950	0.954	0.960
Bridge $\times 10^{-5}$		-0. 0156	-0. 0345	-0. 0527	-0. 0783	-0. 165	-0. 292	-0. 416
BridgeStd $\times 10^{-4}$		0. 00466	0. 0131	0. 0257	0. 0509	0. 201	0. 601	1.20

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is positively autocorrelated. The probability of order flow continuance is set to be 53%. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}, i.e. $\Delta M_t \sim N(0, 4 \times 10^{-8})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected. δ is the probability of the order flow keeping the same direction as the past one.
The other settings are the same as Table (2.2)

2.6.8 Feedback Trading

In this section, we consider the influence of an alternative order flow generating process. most settings are the same as in Section (2.6.2) except that now feedback trading is introduced. Thus all the differences in the performance of the estimators can be imputed to feedback trading. In this section, order flow is related to past mid-price returns; mid-prices follow a random walk process; the volatility of mid-price returns is small; and the spread is fixed. Under these circumstances, both the HS and the Roll estimator are biased, and the error of the CS estimator should also be influenced by feedback trading. Formally, let $\varphi = 0$, $\psi = 1$, $\phi = 0$ in Equation (2.79), which suggests that order flow is a function of past mid-price returns. Let $\kappa = 0.65$, which suggest there is positive feedback trading. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process. Let $\mu = 0.0003$, $\varsigma^2 = 0$ which suggest that the spread is a constant. Let standard deviations of mid-price returns is small, $\sigma = 0.000003$. The system is given as follows:

$$
\begin{aligned}
& BS_t \sim \begin{cases} B(1, 0.65) & \text{if} \quad \Delta M_t > 0 \\ B(1, 0.35) & \text{if} \quad \Delta M_t < 0 \end{cases} \\
& \Delta M_t = \varepsilon_t \\
& \varepsilon_t \sim N(0, 9 \times 10^{-12}) \\
& SP_t = 0.0003 \\
& s_t = M_t + \frac{SP_t}{2} \cdot BS_t
\end{aligned}
\tag{2.90}
$$

The results are presented in Table (2.9), which is similar to Table (2.2). $Cov(\Delta M_t, BS_t)$ represents the covariance of mid-price returns and order flow, which suggests the existence of feedback trading.

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2×10^{-4} which is the same as the setting of the system. In the 24-hour case, the standard deviation is 7.58×10^{-3}. Thus the ratio of the spread to the standard deviation ranges from 1.5 to 0.0396. $Cov(\Delta M_t, BS_t)$ is about 4.79×10^{-5} and is stable across all time intervals except for the cases of 12-hour and 24-hour. Unlike the autocorrelations of mid-price returns or order flow, feedback trading is persistent in low sampling frequencies, thus it will influence the performance of the estimators in all time intervals.

The Roll estimator performs worse than it does in Section (2.6.2), because of feedback trading. All the errors are significantly different from zeros. For short time intervals, the estimator overestimates the spread, which coincides with the theoretical prediction of error six of the estimator. In the tick-by-tick case, the estimate is 3.45×10^{-4}, and the covariance of the transaction price returns is $Cov(\Delta s_t, \Delta s_{t-1}) = -2.98 \times 10^{-8}$. The covariance of mid-price returns and order flow is $Cov(\Delta M_t, BS_t) = -4.79 \times 10^{-5}$. According to Equation (2.27), the result is 0.0003[h] which is the roughly the same as the one in the previous section and equals the true spread. Furthermore the error is 0.00045 which coincides with the prediction of Equation (2.28). The prediction is also valid for other time intervals. Thus, all the differences between this and previous sections are caused by feedback trading and can be explained by the analysis of error six. For brevity, the similar analysis in Section (2.6.2) is not repeated. To sum up, the Roll estimator overestimates the true spread because of positive feedback trading. Since the errors are positively correlated to the strength of feedback trading, when feedback trading is weak, the estimator can be used for short time intervals.

The HS estimator overestimates the true spread in all time intervals. The estimator overestimates the spread, which coincides with the theoretical prediction of the error of the estimator discussed early in this chapter. According to Equation (2.38), the error should be twice as big as the covariance of mid-price returns and order flow. In tick-by-tick case, the theoretical error is 9.58×10^{-5} which is very similar to the real one which is 9.57×10^{-5}. The situation is similar in other time intervals. Therefore all the differences between this and previous sections are caused by feedback trading and can be explained by the analysis of the error of the estimator. For brevity, the similar analysis in Section (2.6.2) is not repeated. To sum up, the HS estimator overestimates the true spread because of positive feedback trading. The errors are greater than those of the Roll estimator when the time interval is shorter than four hours. The HS estimator is more stable than the Roll estimator. When the time interval is short, one should not use the HS estimator but the Roll estimator.

The CS estimator produces larger estimates in all time intervals than those in Section (2.6.2). $\widehat{SP}$, which is related to error one of the estimator, is about 13% closer to the true spread than that in Section (2.6.2). η, which is related to error two of the estimator is at least 13% (24-hour case) and at most 25% (5-minute case) higher than that in Section

[h] $SP = -4.79 \times 10^{-5} + \sqrt{(4.79 \times 10^{-5})^2 - 4 * (-2.98 \times 10^{-8})} = 3.01 \times 10^{-4}$

(2.6.2). These results suggest that positive feedback trading can reduce errors one and two. If there is positive feedback trading, the probability of a positive (negative) shock being with a buy (sell) order is higher than 50%, which is lower the probability of the highest (lowest) being with a sell (buy) order. As was discussed, both errors one and two unlikely happen when the probability of the highest (lowest) being with a sell (buy) order is low. The results of *Bridge* and *BridgeStd* are exactly the same as those in Section (2.6.2), which suggests that error three is independent to feedback trading. Because error three is the dominant error when the ratio of the spread to the volatility of mid-price returns is low, although positive feedback trading improves the estimation in terms of errors one and two, the total error of the estimate is still too big to be ignored. In the four-hour case, the estimator has an accurate estimation. Because the estimates of the CS estimator is an increasing function of the time interval and the range of which ranges from a negative value to a value larger than the true spread, there are always several accurate estimates for various time intervals, but one cannot know in which time interval the estimate is accurate. Therefore, it is an occasional event that the accurate estimation is in four-hour case rather than that the estimator can always obtain accurate estimates in four-hour case. Since other information in the panel is similar to those in Section (2.6.2), we do not repeat the analysis. Generally speaking, the CS estimator does not yield good estimates of the true spread, even if positive feedback trading reduces errors one and two, it should not be used under similar conditions.

To sum up, positive feedback trading makes both the Roll and the HS estimators overestimate the true spread. The estimates of the CS estimator do not improve because of positive feedback trading and still should not be used. For short time intervals, the Roll estimator has low errors than the HS estimator does. For long time intervals, both the Roll and the HS estimators are not very stable, but the latter is relatively more stable and accurate. Similar to those shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator at low frequencies.

Table 2.9: Feedback Trading

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-2}$	0.02	0.0447	0.0775	0.110	0.155	0.310	0.536	0.758
$Cov(\Delta M_t, BS_t)$ $\times 10^{-4}$	0.479	0.479	0.479	0.480	0.477	0.447	0.584	0.544
Roll 1984								
Estimates $\times 10^{-3}$	0.345	0.345	0.344	0.336	0.307	0.450	0.923	1.57
Relative Estimate	1.15	1.15	1.147	1.12	1.0233	1.5	3.0767	5.233
Est-Std$\times 10^{-3}$	0.925	0.00523	0.0219	0.0661	0.174	0.489	1.07	1.78
T-test $\times 10^{-3}$	0.0446***	0.0446***	0.0441***	0.0362***	0.00657	0.150***	0.623***	1.27***
	(1523.78)	(269.71)	(63.66)	(17.32)	(1.19)	(9.69)	(18.38)	(22.61)
RMSE$\times 10^{-3}$	0.926	0.0453	0.0491	0.0753	0.174	0.511	1.238	2.187
Huang and Stoll 1997								
Estimates $\times 10^{-3}$	0.396	0.396	0.396	0.396	0.398	0.406	0.412	0.431
Relative Estimate	1.32	1.32	1.32	1.32	1.327	1.353	1.373	1.437
Est-Std$\times 10^{-4}$	0.00587	0.0302	0.0919	0.179	0.373	1.43	4.37	8.62
T-test $\times 10^{-4}$	0.957***	0.957***	0.956***	0.958***	0.975***	1.06***	1.12***	1.31***
	(5156.49)	(1001.98)	(329.06)	(169.50)	(82.74)	(23.49)	(8.10)	(4.82)
RMSE $\times 10^{-3}$	0.0960	0.0960	0.0964	0.0977	0.105	0.178	0.451	0.872
Corwin and Schultz 2010								
Estimates $\times 10^{-3}$		-0.116	-0.0135	0.0460	0.111	0.298	0.545	0.770
Relative Estimate		-0.387	-0.045	0.153	0.37	0.993	1.817	2.567
Est-Std$\times 10^{-3}$		0.00169	0.00451	0.00869	0.0169	0.0678	0.212	0.407
T-test$\times 10^{-3}$		-0.416***	-0.313***	-0.254***	-0.189***	-0.00227	0.245***	0.470***
		(-7814.87)	(-2197.73)	(-923.95)	(-353.69)	(-1.06)	(36.63)	(36.49)
RMSE$\times 10^{-3}$		0.416	0.314	0.254	0.190	0.0678	0.324	0.622
$\widehat{SP}$ $\times 10^{-3}$		0.164	0.192	0.200	0.206	0.213	0.216	0.216
η		0.893	0.940	0.952	0.960	0.970	0.974	0.973
Bridge $\times 10^{-5}$		-0.0156	-0.0345	-0.0527	-0.0783	-0.165	-0.292	-0.417
BridgeStd$\times 10^{-4}$		0.00466	0.0131	0.0257	0.0509	0.202	0.603	1.20

There are 1000 replications. There are 432000 periods, each of which represents one minute, in each replication. Data of each replication are generated according to the following system. The order flow is positively autocorrelated. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}, i.e. $\Delta M_t \sim N(0, 4 \times 10^{-8})$. Order flow is positively correlated to mid-price returns. The probability of a buy (sell) order being after a positive (negative) return is 65%. i.e. The spread is fixed and equals to 0.0003, i.e. $BS_t \sim B(1, 0.65)$ if $\Delta M_t > 0$ and $BS_t \sim B(1, 0.35)$ if $\Delta M_t < 0$. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.

$Cov(\Delta M_t, BS_t)$ is the covariance of mid-price returns and order flow, which reflects the existence of feedback trading.

The other settings are the same as Table (2.2)

2.6.9 Irregular Timed Spaces

In this section, most settings are the same as in Section (2.6.2) except that now irregular timed spaces are introduced. Thus all the differences in the performance of the estimators can be imputed to irregular timed spaces. The time stamps of EUR/USD indicative data are used. Daníelsson and Payne 2002 show that the quote/transaction frequencies of indicative and transaction data are "fairly strongly positively correlated". So the time stamps of EUR/USD indicative data reflect real market conditions. In this section, order flow is random; mid-prices follow a random walk; the spread is fixed; the time spaces between trades are irregular. Under these circumstances, both the HS and the Roll estimator are unbiased, and error three of the CS estimator should be greater than Section (2.6.2). Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process and the standard deviation of mid-price returns is $\sigma = 0.0002$ and $\mu = 0.0003$, $\varsigma^2 = 0$ which suggests that the spread is a constant. The system is given as follows:

$$\begin{aligned}
&BS_t = \varpi_t \\
&\varpi_t \sim B(1, 0.5) \\
&\Delta M_t = \varepsilon_t \\
&\varepsilon_t \sim N(0, 4 \times 10^{-8}) \\
&SP_t = 0.0003 \\
&s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t
\end{aligned} \tag{2.91}$$

The time stamps of quote EUR/USD data are used. The data are recorded using Greenwich Mean Time (GMT). There are 200 days and 1408725 observations in total. Figure 2.1 shows the average daily time patten of the number of trades in five-minute intervals. The horizontal axis shows the time intervals, where 1 represents the interval from 00:00 to 00:05 and 288 represents the interval from 23:56 to 24:00. The number of trades in each interval is the 200 days average. In the afternoon (GMT), the number of trades in a five-minute interval is more than 30, because both the US and European markets are open. During the night, the trading frequency is lower because only Asia markets are open at that time. There are about 25 trades in a five-minute interval on average.

One thousand replications are generated according to the system above. As we men-

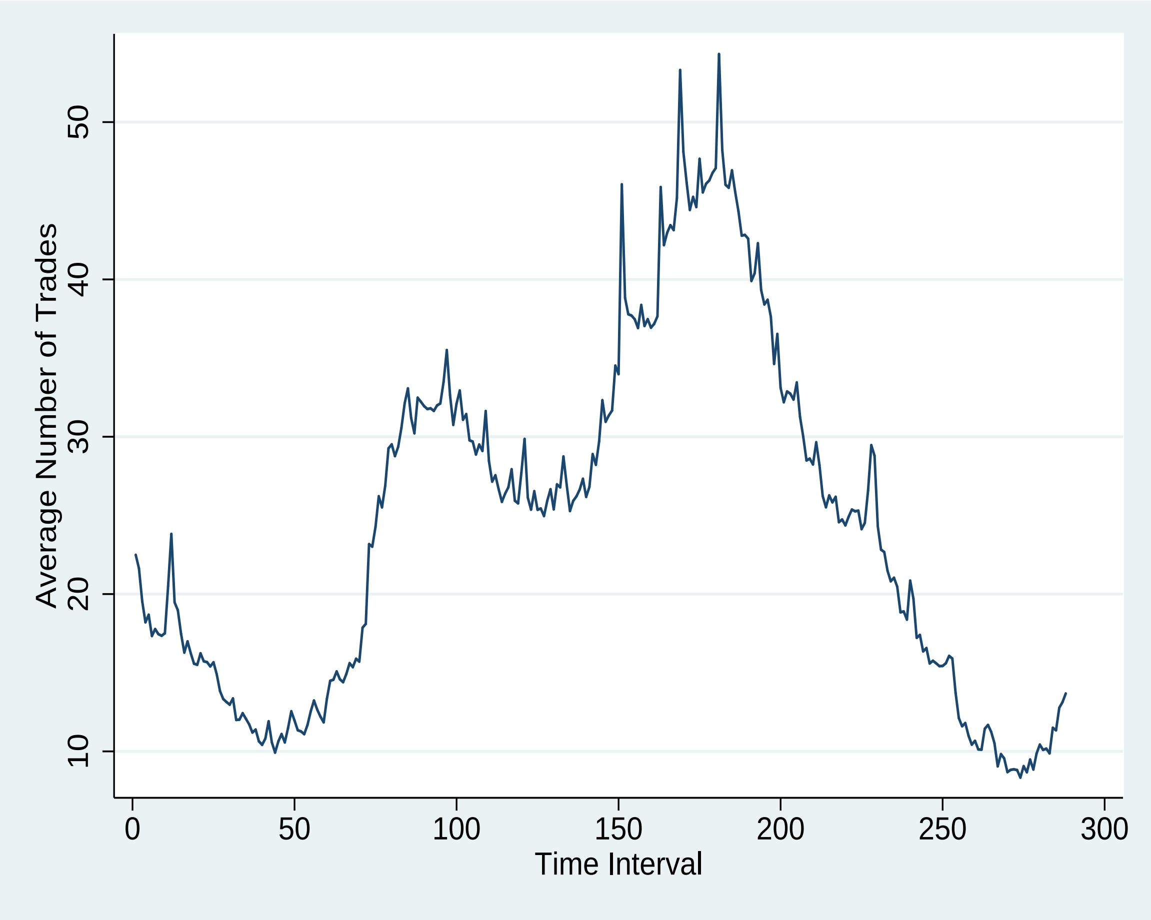

Figure 2.1: Average Number of Trades

The figure shows the average numbers of trades in five-minute intervals in a day. The data are recorded using Greenwich Mean Time (GMT). There are 200 days and 1408725 observations in total. The horizontal axis shows the time intervals, where 1 represents the interval from 00:00 to 00:05 and 288 represents the interval from 23:56 to 24:00. The number of trades in each interval is the 2oo days average.

tioned at the beginning of Section (2.6), each replication has eight subgroups according to various sampling periods. The results are presented in Table (2.10), which is similar to Table (2.2).

The row of *Midstd* suggests that the time interval and the standard deviation of mid-price returns are positively related. In the tick-by-tick case, the average standard deviation of mid-price returns is 2×10^{-4} which is the same as the setting of the system. In the 24-hour case, the standard deviation is 1.68×10^{-2}. Thus the ratio of the spread to the standard deviation ranges from 1.5 to 0.0179. Because the trading frequency is higher than previous sections, the standard deviations of mid-price returns are greater than previous sections respectively.

The Roll estimator does estimate the spread correctly in the tick-by-tick case, and underestimates the spread from the 5-minute case to the 30-minute case, and then overestimates the spread in even longer time intervals, according to the first row of the panel *Roll*. In 5-minute case, the error is approximately of the scale of 10^{-6} which is the same as the prediction of Equation (2.25), where the error is of the scale of $\sqrt{\sigma^2/n}$ and is 10^{-6}[i] in this case. The estimator significantly overestimates the spread in longer time intervals, which could be imputed to the appearance of the positive co-variance. The estimator will be undefined when the co-variance is positive, and the results are set equal to zeros. The standard deviation of the estimated spreads in the tick-by-tick case is less than that in Table 2.3) because here there are much more observations. The standard deviations are greater than those in Table 2.3) because the observations are the same and the mid-price volatility is higher. In the 15-minute case, the standard deviation is greater than 50% of the spread, which suggests that the results are not stable across the replications. The third row suggests that the error increases when the time interval lengthens. The results of the t-test suggest that the estimated spreads are not significantly different from the true one in the tick-by-tick case. In longer time intervals, the error becomes too big to be ignored (more than 15% of the spread). The differences between this section and Section (2.6.2) can be explained by the higher mid-prices volatility. One can conclude that irregular timed spaces do not influence the estimates of the estimator.

The HS estimator performance is similar to in Section (2.6.2). The estimator does estimate the spread accurately. It does not rely on the assumption that the time spaces between trades are regular. Thus, it is not influenced by the irregular timed spaces. The results in Table (2.10) are not very different from those in Table (2.3), except that because of the higher mid-price volatility the estimator has greater standard deviations of the estimates, thus we do not repeat the whole analysis. Generally speaking, the HS estimator can be used when time spaces between trades are irregular. It performs better than the Roll estimator does.

The CS estimator does not yield good estimates of the true spread, according to the first row of the panel *Corwin and Schultz*. The estimates are greater than those in Table (2.3). The error is positively correlated to the volatility of the spread. In the five-minute case,

[i]The variance is 0.00101^2 and the number of observations is 86400, thus $\sqrt{0.00101^2/86400} = 3.44 \times 10^{-6}$

the CS estimator has negative results, which might be caused by the joint effect of errors. In the 24-hour case, the estimator significantly overestimates the true spread: the estimate is about 550% of the true spread. The standard deviations of the estimates are greater than those of the HS estimator and lower than those of the Roll estimator, according to the second row in the panel, which suggests that irregular timed spaces lower the stability of the estimates. The total errors are significantly different from zero, and cannot be ignored. $\widehat{SP}$ ranges from 1.51×10^{-4} to 1.76×10^{-4} and is greater than that in Table (2.3), since the number of trades in each time interval is much greater. In other words, compared to Section (2.6.2), the magnitude of error one is less. η ranges from 91.4% to 96.6% and is also greater than that in Section (2.6.2), which can be imputed to the greater number of trades in each time interval as well. *Bridge* suggests that e_{v2} ranges from the scale of 10^{-7} to 10^{-6}, of which the absolute value is greater than that in Section (2.6.2). In the five-minute case, the absolute value of *Bridge* is three times as big as the one in Table (2.3). In the 24-hour case, the absolute value of *Bridge* is 19% greater than the one in Table (2.3). As mentioned before, the irregular timed spaces can bring in error three. The numbers of trades will be closer to each other in longer time intervals, because when sample size increases, the value will be closer to it's mean. Therefore, the phenomenon that the scale of the relative difference between *Bridge* in Tables (2.10) and (2.3) decreases when the time interval becomes longer can be explained by the fact that error three which is partially caused by irregular timed spaces is less in longer time intervals. The higher volatility of mid-price returns is another reason caused larger error three, which has been discussed above. *BridgeStd*, the the standard deviation of *Bridge*, is much greater than that in Section (2.6.2), which is caused by the higher volatility of mid-price returns. We do not repeat the whole analysis. One can conclude that irregular timed space makes the CS estimator have greater errors than those in Section (2.6.2). The results are far away from the true spread, thus it should not be used in similar environments.

To sum up, irregular timed spaces do not influence the Roll and the HS estimators but only the CS estimator. Similar to those results shown in Table (2.2), the CS estimator has the lowest RMSE at time intervals of four hours or more, but if high-frequency data are available, then the HS or Roll estimator applied at high frequencies has a much lower RMSE than the CS estimator at low frequencies.

Table 2.10: Irregular Timed Spaces

	Tick	5-Min	15-Min	30-Min	1-Hour	4-Hour	12-Hour	24-Hour
Midstd $\times 10^{-2}$	0.0200	0.101	0.173	0.245	0.346	0.688	1.19	1.68
				Roll 1984				
Estimates $\times 10^{-3}$	0.300	0.297	0.261	0.298	0.408	1.01	2.27	4.00
Relative Estimate	1	0.99	0.87	0.993	1.36	3.367	7.567	13.333
Est-Std$\times 10^{-3}$	0.000499	0.0348	0.174	0.279	0.441	1.16	2.58	4.49
T-test $\times 10^{-3}$	2.82e-05	-0.00315**	-0.0394***	-0.00198	0.108***	0.711***	1.97***	3.70***
	(1.79)	(-2.86)	(-7.18)	(-0.22)	(7.77)	(19.46)	(24.18)	(26.03)
RMSE$\times 10^{-3}$	0.000499	0.0349	0.178	0.279	0.454	1.360	3.246	5.818
				Huang and Stoll 1997				
Estimates $\times 10^{-3}$	0.300	0.300	0.300	0.302	0.305	0.282	0.233	0.242
Relative Estimate	1	1	1	1.007	1.0167	0.94	0.777	0.807
Est-Std$\times 10^{-3}$	0.000335	0.00856	0.0250	0.0517	0.100	0.390	1.18	2.38
T-test $\times 10^{-4}$	-3.15e-05	0.00461	0.00490	0.0183	0.0538	-0.180	-0.671	-0.579
	(-0.30)	(1.70)	(0.62)	(1.12)	(1.70)	(-1.46)	(-1.80)	(-0.77)
RMSE$\times 10^{-3}$	0.000335	0.00856	0.025	0.0517	0.100	0.390	1.182	2.381
				Corwin and Schultz 2010				
Estimates $\times 10^{-3}$		-0.0275	0.0783	0.163	0.271	0.604	1.17	1.73
Relative Estimate		-0.0917	0.261	0.543	0.903	2.013	3.9	5.767
Est-Std$\times 10^{-3}$		0.00383	0.0114	0.0232	0.0459	0.174	0.541	1.15
T-test$\times 10^{-3}$		-0.328***	-0.222***	-0.137***	-0.0293***	0.304***	0.868***	1.43***
		(-2705.15)	(-614.40)	(-185.91)	(-20.18)	(55.43)	(50.75)	(39.32)
RMSE $\times 10^{-3}$		0.328	0.222	0.139	0.0543	0.350	1.0245	1.835
$\widehat{SP}$ $\times 10^{-3}$		0.151	0.164	0.168	0.171	0.174	0.175	0.176
η		0.914	0.941	0.948	0.953	0.959	0.962	0.966
Bridge $\times 10^{-5}$		-0.0442	-0.0838	-0.123	-0.177	-0.356	-0.528	-0.479
BridgeStd$\times 10^{-4}$		0.0248	0.0722	0.142	0.277	1.02	2.94	5.87

There are 1000 replications. There are 200 days and 1408725 observations in each replication. The time stamps of EURUSD quote data are used. Data of each replication are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The mid-price return is drawn from a normal distribution of which the mean is zero and the variance is 4×10^{-8}, i.e. $\Delta M_t \sim N(0, 4 \times 10^{-8})$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. Thus, there are eight subgroups for each replication. For each subgroup, the standard deviation of mid-price returns, and the estimated spread are collected.
The other settings are the same as Table (2.2)

2.6.10 The Ratio of the Spread to the Standard Deviation of Mid-Price Returns

In this section, the relationship between the errors of the estimators and the ratio of the spread to the standard deviation of mid-price returns (the ratio hereafter) is examined. In previous sections, it is confirmed that the ratio is one of the most important factors which influence the results of the estimators. Unlike previous sections which focus on analysing the sources of errors, this section is going to identify ranges in which the estimators will generate acceptable estimates. The conclusion in this section may be used to infer the accuracy of the estimate.

The system in Section (2.6.1) is used. Order flow is random; mid-prices follow a random walk process; and the spread is fixed. Thirty volatilities of mid-price returns are used. Formally, let $\varphi = 0$, $\psi = 0$, $\phi = 1$ in Equation (2.79), which suggests that order flow is random. Let $\tau + \omega = 0$ in Equation (2.81), which suggests that the mid-price follows a random walk process. Let $\mu = 0.0003$, $\varsigma^2 = 0$ which suggest that the spread is constant. There are thirty values for the standard deviation of mid-price returns which ranges from 0.000018 to 0.000453. The increment is 0.000015. The system is given as follows:

$$\begin{aligned} & BS_t = \varpi_t \\ & \varpi_t \sim B(1, 0.5) \\ & \Delta M_t = \varepsilon_t \\ & \varepsilon_t \sim N(0, \sigma^2) \\ & 0.000018 \leq \sigma \leq 0.000453 \\ & SP_t = 0.0003 \\ & s_t = M_t + \tfrac{SP_t}{2} \cdot BS_t \end{aligned} \tag{2.92}$$

Since there are thirty values of the standard deviations, there are thirty systems that are generated according to the equations above. In each system there are five hundreds replications, each of which has 432000 periods. As we mentioned at the beginning of Section (2.6), each replication has eight subgroups according to various sampling periods. For each time intervals, there are thirty pairs of errors and the corresponding ratios. The pairs are used to draw the figures. To keep things simple, only the results of the four-hour case are reported. The reason why the four-hour case is selected is that firstly, there are

240 observations in a four hours interval, where the number of observations in a time interval will not influence the estimators, and thus one can focus on the influence of the ratio; secondly, in the four-hour case, the results fully reflect the pattern of the relationship of the errors of the estimators and the ratio. In four-hour case, the ratio varies from 0.043 to 1.08. In the figures, the solid curve reports the relative error of the estimator, which is the ratio of the error of the estimator to the true spread, Formally, it is calculated from $\frac{\widehat{SP}-SP}{SP}$, where $\widehat{SP}$ is the estimated spread and SP is the true spread. The dot line reports the relative standard deviation of the estimates, which reflects the degree of stability of the estimation. The relative standard deviation is the ratio of the standard deviation to the true spread.

Figure (2.2) shows the pattern of the ratio against the error of the Roll estimator in the four-hour case. Both the curves are like a reverse J shape. The error curve decreases, then increasing and finally approaches the zero line when the ratio takes higher values. When the ratio is less than 0.18, the estimator tends to overestimate the true spread. The standard deviation decreases as the ratio increases. When the ratio is small, both the error and the standard deviation are very high relative to the true spread, which suggests that the estimator has big errors and is very unstable in this range of the ratio. When the ratio is larger than 0.15, the error is less than 10%, which is acceptable. But the standard deviation is still around 100% of the true spread, which is still unstable. When the ratio is in the interval from 0.2 to 0.3, the error becomes greater again but the absolute value of the error is roughly less than 10%. Thus, in this interval, the estimate is still accurate. When the ratio is greater than 0.3, the estimate is coverage to the true spread, which also has been confirmed in previous sections. When the ratio is greater than 0.6, the standard deviation declines to less than 20% of the true spread, which is stable. To sum, up, the Roll estimator can produce accurate estimates when the ratio is greater than 0.15, which suggests that the Roll estimator can be used if there is a sufficient number of replications, although the estimate is not stable until the ratio is greater than 0.6, when even a single replication would generate an acceptable estimate.

Figure (2.3) shows the pattern of the ratio against the error of the HS estimator in the four-hour case. The error is always less than 10%, which suggests that the estimator can produce accurate estimates. As the ratio becomes greater, the estimation coverages to the true spread. The standard deviation, which decreases in the ratio, is not less than 20% until the ratio is greater than 0.25. Thus the HS estimator can produce accurate estimates when the ratio is greater than 0.25, which suggests that the HS estimator performs better than the Roll estimator.

Figure (2.4) shows the pattern of the ratio against the error of the CS estimator in the four-hour case. The error curve decreases, then increasing and finally approaches the zero line when the ratio takes greater values. The standard deviation is a decreasing function of the ratio. When the ratio is greater than 0.48 the estimator always underestimates the true spread and vice versa. Similarly to previous sections, the standard deviation of the estimates of the CS estimator is the smallest one among the estimators. When the ratio ranges from 0.043 and 0.08, the error is less than 10%. Although the standard deviation in this range is higher than 20%, it is still less than that of the other estimators. When the ratio is greater than 0.25, the error is less than 10% again. Furthermore, the standard devastation is less than 20%, when the ratio is greater than 0.1. Thus when the ratio is greater than 0.25, the estimator can yield accurate and stable estimations. However, as was discussed above, because of errors one and two, the error of the CS estimator is normally higher than those of the other estimators. When the ratio is 1.08, at which errors of both the Roll and the HS estimators are less than 1%, the error of the CS estimator is greater than 5%. The CS estimator performs well in the simulation experiments in Corwin and Schultz (2012), for the reasons which are similar to those that can explain the performance in this section. First, the ratios are 0.16, 0.33, 1.67 and 2.67 and most of which lie at the interval where the estimated errors are acceptable. Second, there are 390 observations in a time interval, which reduces errors one and two.

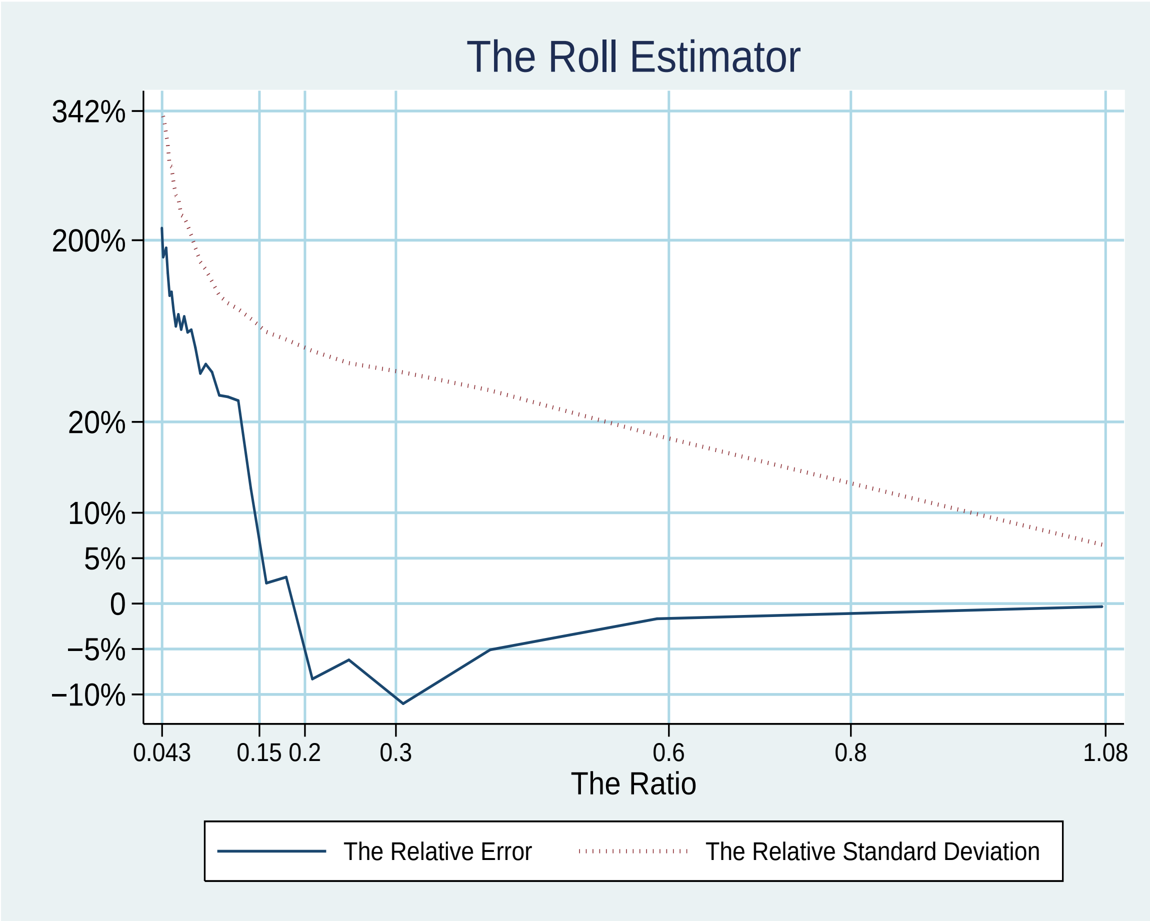

Figure 2.2: The Error of the Roll Estimator and the Ratio

Data are generated according to the following system. The order flow is drawn from a binomial distribution, i.e. $BS_t \sim B(1, 0.5)$. The spread is fixed and equals to 0.0003, i.e. $SP_t = 0.0003$. The transaction price is the mid-price plus or minus a half spread, i.e. $s_t = M_t + \frac{SP_t}{2} \cdot BS_t$. The mid-price return is drawn from a normal distribution of which the mean is zero and the standard deviation of mid-price returns σ, i.e. $\Delta M_t \sim N(0, \sigma^2)$. σ ranges from 0.000018 to 0.000453, the increment of which is 0.000015. Thus, there are thirty different values of σ. Hence, there are thirty systems are generated. In each system there are 500 replications, each of which has 432000 periods. Each replication is also sampled into longer time intervals: five-minute, fifteen-minute, thirty-minute, one-hour, four-hour, twelve-hour and twenty-four- hour, and only the close observations are kept. For each time interval, there are thirty pairs of the errors and the corresponding ratios. The pairs are used to draw the figures. For brevity, only the results of four-hour case are reported in this figure. The reason why the four-hour case is selected is that firstly, there are 240 observations in four hours interval, the number of observations in a time interval will not influence the estimators, thus one can focus on the influence of the ratio; secondly, in four-hour case, the results fully reflect the pattern of the relationship of the errors of the estimators and the ratio. The horizontal axis is the ratio which, in four-hour case, ranges from 0.043 to 1.08. The solid curve reports the relative error of the estimator, which is the ratio of the error of the estimator to the true spread, Formally, it is calculated from $\frac{\widehat{SP} - SP}{SP}$, where $\widehat{SP}$ is the estimated spread and SP is the true spread. The dot line reports the relative standard deviation of the estimates, which reflects the degree of stability of the estimation. The relative standard deviation is the ratio of the standard deviation to the true spread.

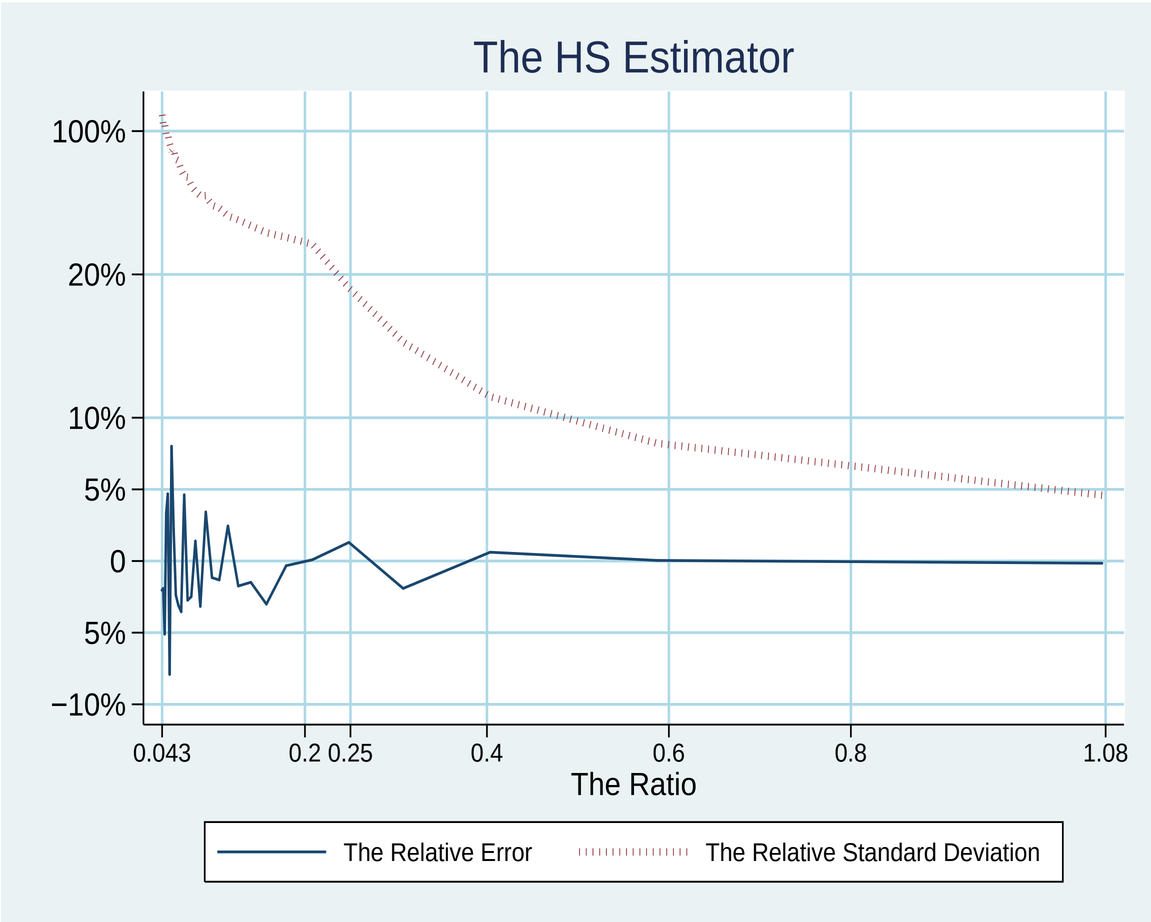

Figure 2.3: The Error of the HS Estimator and the Ratio

The figure shares the same settings as Figure (2.2)

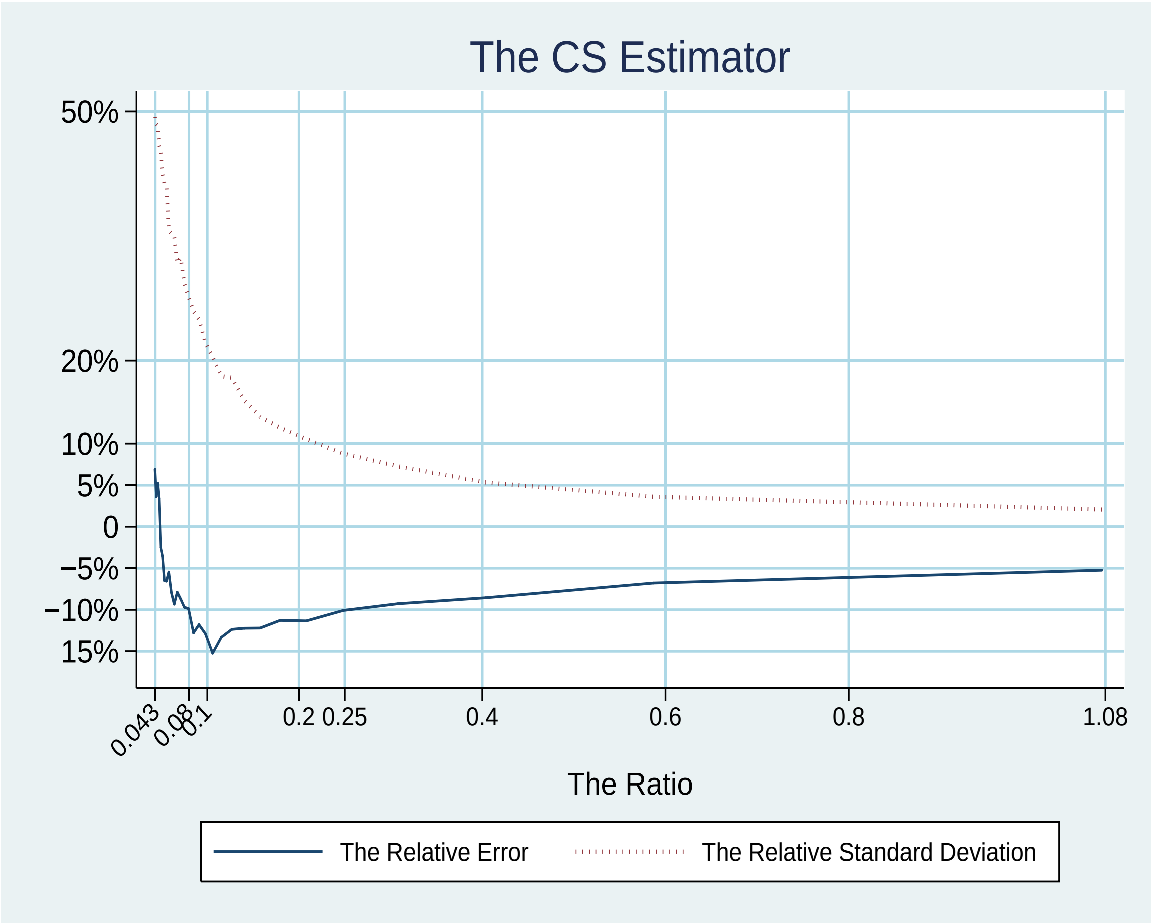

Figure 2.4: The Error of the CS Estimator and the Ratio

The figure shares the same settings as Figure (2.2)

2.7 Reuters D2000-1 USD/DEM Transaction Data

In this section, we use transaction data for Deutsche-Mark against US dollar transactions where such data include transaction prices and order flow from the Reuters D2000-1 system. The performance of the estimators can be tested in the most realistic environment. Based on conclusions drawn in earlier sections and in the existing literature (such as Goodhart et al. 1996 and Daníelsson and Payne 2002), we expect the unobserved mid-price returns to be autocorrelated, spreads to be time-varying, and the ratio between them to be similar to that in EUR/USD data. Furthermore, later in this section, we will show that the data are irregularly timed and that order flow is autocorrelated. It is also reasonable to believe that there are inventory control and asymmetric information (thereafter IC& AS) components in the spreads and that there is feedback trading. In other words, the data exhibit all the features that have been discussed in the previous chapter. Because we do not have data on spreads, we cannot compare the estimates to the true values. We can only discuss which one is more likely to be close to the true value, based on features of the data and the conclusions of the previous chapter.

We describe the data before performing the estimation. The data include all the interbank tick- by-tick quotes and order flow of the Deutsche-Mark against the US dollar from May 1 to August 31 1996 on Reuters Dealing 2000-1 (hereafter, D2000-1). There are 83 days and 256838 observations in total. The data set format is shown in Table (2.11). Trades on D2000-1 occur between two anonymous dealers: a calling dealer, who requires the quotes, and a quoting dealer. The quoting dealer provides both bid and ask prices to the calling dealer. The calling dealer must then decide to buy (make a positive order flow) or sell (make a negative order flow) dollars or reject the quote in seconds. If the transaction is made, the time and the direction will be recorded by the system. Two points should to be noted. First, traders can only observe their own trading records. Second, although both bid and ask prices (two series of exchange rates) are quoted by the dealer, only a bid or ask price that reflects the direction of an actual trade is in the dataset.

Let s be the logarithm of the nominal exchange rate of the US dollar against another currency (an increase representing an appreciation of the US dollar); let BS be the order flow equal to +1 when an agent buys dollars and -1 when an agent sells dollars; let t denote time.

Table 2.11: Data Format(USD/DEM)

Month	Day	Hour	Min	Sec	B/S	Ask	Bid
4	30	18	45	40	1	1.5326	.
4	30	18	46	23	-1	.	1.5326
4	30	18	47	56	1	1.5328	.
4	30	18	48	22	1	1.533	.
4	30	18	49	53	1	1.5332	.
4	30	18	51	0	-1	.	1.5327
4	30	18	52	34	1	1.5327	.
4	30	18	53	8	1	1.533	.
4	30	18	53	35	-1	.	1.5329
4	30	18	54	21	1	1.5329	.
4	30	18	54	27	-1	.	1.5333
4	30	18	55	10	-1	.	1.533

This table shows the format of the data on the Reuters D2000-1 system.
B/S represents trade directions. It is 1 if there is a buy order and is -1 if there is a sell order.

$$s_t = \begin{cases} A_t, & \textit{Buy order } (BS_t = 1) \\ B_t, & \textit{Sell order } (BS_t = -1) \end{cases} \tag{2.93}$$

where B is the logarithm of the bid price and A is the logarithm of the ask price, and $A \geq B$. The unobserved bid-ask spread (SP) is the difference between ask and bid prices:

$$SP_t = A_t - B_t$$

Table (2.12) reports the summary statistics on the returns of the USD/DEM rate. The returns of the transaction data exhibit features similar to those of the mid-price returns of the indicative data discussed in the previous section: the mean is close to zero; the standard deviation is smaller during peak trading time; returns are negatively autocorrelated; and the Box-Ljung tests reject the hypothesis of no 10-lag autocorrelation up to ten lags. Unlike the returns of the indicative data, the transaction returns do not exhibit weak autocorrelation during peak trading time.

Table (2.13) reports the summary statistics for order flow. In the first row, we can observe that mean order flow is approximately zero and that the fraction of buy orders is close to 50%, which suggests that the number of buy orders is roughly the same as the number of sell orders. The second row corresponds to the case when the last order is a buy order. The conditional mean of order flow, when the last order is a buy order, is greater than zero, and the fraction of buy orders is 3.56% above 50%. The third row shows

Table 2.12: Summary Statistics: USD/DEM Transaction Price Returns

Variable	Mean $\times 10^{-6}$	Std. Dev. $\times 10^{-3}$	ρ_1	$Q(10)$
Pooled	-1.11	0.386	-0.485	60361
0 - 1	4.07	0.5704	-0.488	269
1 - 2	-2.69	0.5783	-0.475	362
2 - 3	-0.0502	0.5459	-0.496	1173
3 - 4	-0.428	0.5006	-0.486	1223
4 - 5	-0.677	0.4535	-0.496	956
5 - 6	2.24	0.5827	-0.544	474
6 - 7	0.00821	0.4574	-0.451	551
7 - 8	1.24	0.4218	-0.474	1751
8 - 9	-0.612	0.4206	-0.501	4184
9 - 10	-0.309	0.3513	-0.490	5824
10 - 11	0.282	0.3245	-0.485	5690
11 - 12	-0.0344	0.3352	-0.479	4365
12 - 13	0.456	0.3321	-0.491	3893
13 - 14	-0.234	0.3626	-0.482	4268
14 - 15	-0.517	0.3457	-0.482	5814
15 - 16	-0.104	0.3378	-0.481	5813
16 - 17	0.0241	0.3822	-0.488	5338
17 - 18	0.222	0.4263	-0.497	3737
18 - 19	-3.4	0.4564	-0.473	1753
19 - 20	2.38	0.4187	-0.453	827
20 - 21	2.05	0.4005	-0.424	568
21 - 22	-0.733	0.426	-0.449	371
22 - 23	-1.58	0.5944	-0.479	314
23 - 24	-1.52	0.5931	-0.466	277

This table reports the summary statistics for the returns of USD/DEM transaction exchange rates on the Reuters D2000-1 system from 01/05/1996 to 02/09/1996.
ρ reports the coefficient the first-order autocorrelation of the returns.
$Q(10)$ is a tenth-order Box-Ljung statistic. The 5% critical value is 18.31.

a symmetric case when the last order is a sell order. Therefore one can conclude that an order is likely to be in the same direction as the previous one. In other words, order flow is positively autocorrelated in our data. This phenomenon can be explained by hot potato trading (Lyons 1997 and Evans and Lyons 2002).

Table 2.13: Summary Statistics Order Flow of USD/DEM Pair

Variable	Obs	Mean	Std. Dev.	Fraction of $BS_t = 1$
BS_t	256838	0.00072	1.000	50.04%
BS_t *when* $BS_{t-1} = 1$	128502	0.0713	0.997	53.56%
BS_t *when* $BS_{t-1} = -1$	128320	-0.0699	0.998	46.50%

This table gives the summary statistics for order flow of the USD/DEM pair on the Reuters D2000-1 system from 01/05/1996 to 02/09/1996.
BS represents the order flow (the trade indicator). It is 1 if there is a buy order and is -1 if there is a sell order.

Figure (2.5) shows the average daily time pattern of the number of trades in five minute intervals of the transaction data. The horizontal axis shows the time intervals, where 1 represents the interval from 00:00 to 00:05, and 288 represents the interval from 23:56 to 24:00. The number of trades in each interval is the 83-day average. The pattern is similar to that of the EUR/USD indicative data. In the afternoon (GMT), the number of trades in a five-minute interval exceeds 20 because both the US and European markets are open. During the night, the trading frequency is lower because only Asia markets are open at that time. There are roughly 11 trades in a five-minute interval on average, which is 14 less than the number of quotes in the EUR/USD data, which can be explained by two factors: first, some quotes were rejected; second, trading in 1996 was not as intensive as in 2003.

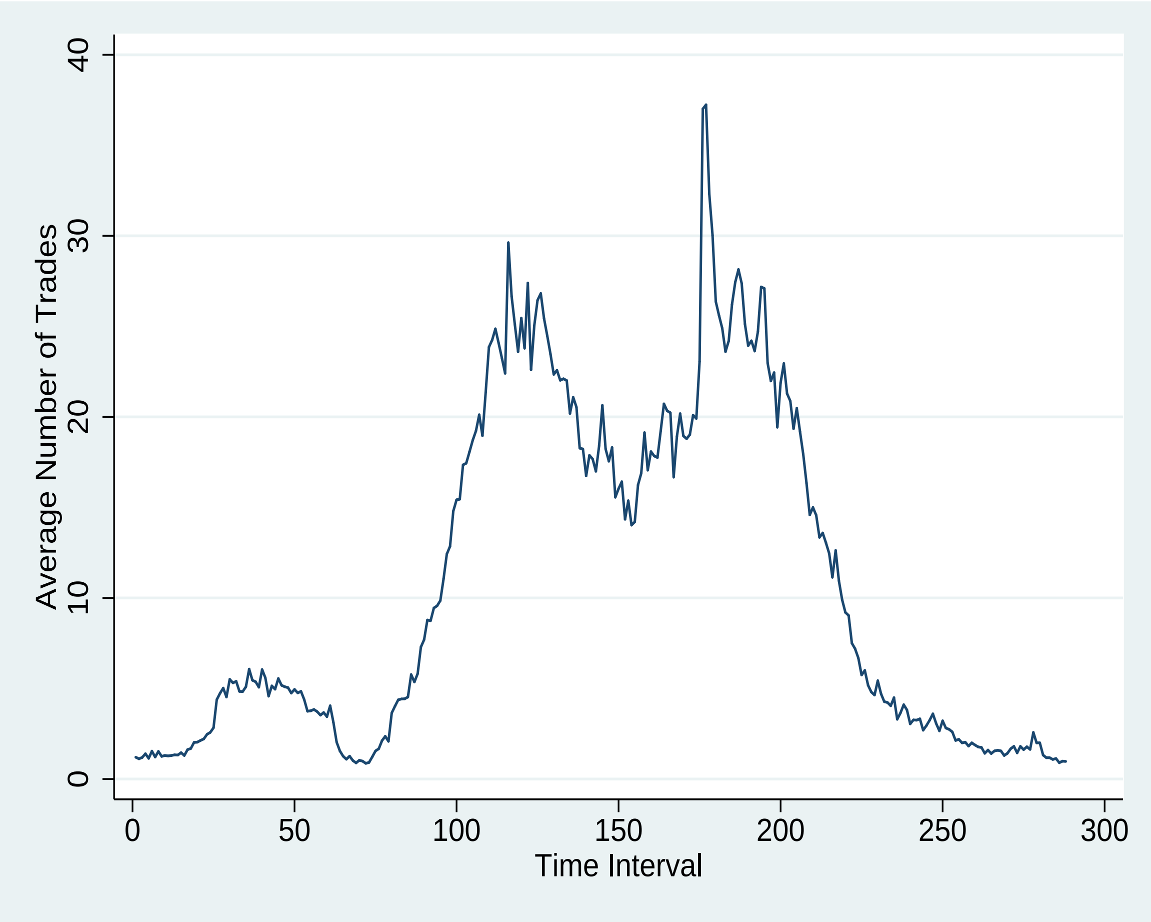

Figure 2.5: Average Number of Trades USD/DEM Transaction Data

This figure shows the average numbers of trades in five-minute intervals in a day. The data are recorded using Greenwich Mean Time (GMT). There are 83 days and 256838 observations in total. The horizontal axis shows the time intervals, where 1 represents the interval from 00:00 to 00:05 and 288 represents the interval from 23:56 to 24:00. The number of trades in each interval is the 83 days average.

We now turn to the estimates of the estimators. The results are shown in Table (2.14). In the tick-by-tick case, the HS estimator suggests that the spread is 0.0000794, while the Roll estimator yields a much higher result, namely, 0.000537. Unlike the estimates in simulation experiments, for longer time intervals, the estimates of the HS estimator are unstable and unpredictable across time intervals. The Roll estimator has relatively stable estimates when the time interval is shorter than four hours. The CS estimator suggests a negative spread in the five-minute case, and its estimate is an increasing function of the time interval. Generally speaking, the patterns of the estimates of the estimators exhibited in Table (2.14) are similar to those in the simulation experiments, except for those of the HS estimator. The reasons for this may be, first, that the bid-ask spread is relatively small compared with the mid-price volatility (i.e., the ratio is small) and second, that unlike the simulation experiments in which there are 1000 replications, the transaction data have only one series, so that the HS estimator does not converge to the true one. Furthermore, the time pattern of the spreads may also influence the estimation.

Table 2.14: Compare estimators USD/DEM pair

USD/DEM	24 hours	12 hours	4 hours	1 hour	30 mins	15 mins	5 mins	Tick
Roll$\times 10^{-3}$	3.35	0.374	-0.477	0.775	0.663	0.663	0.632	0.537
HS $\times 10^{-3}$	0.703	0.302	0.00444	0.0746	0.0558	0.0224	0.00512	0.0794
CS $\times 10^{-3}$	2.079	1.418	1.357	0.995	0.673	0.371	-0.032	N/A
Obs	83	163	477	1881	3738	7352	20261	256836

This table presents the estimates of the Roll, HS and CS estimators of the USD/DEM transaction data under various sampling frequencies. These estimators were introduced in the previous chapter.

Based on the conclusion of the previous chapter — that the HS estimator is the most accurate estimator in most circumstances — and features of the transaction data — autocorrelated mid-price returns, autocorrelated order flow, time-varying spreads, and feedback trading — it is reasonable to believe that the estimates of the HS estimator in the tick-by-tick case are closest to the true one.

In the previous chapter, possible errors of the estimators were analysed quantitatively. Thus, if relative parameters are known, one can obtain similar results from the estimators after adjusting for the errors. Some parameters, such as the autocorrelation of order flow, are known, and some are not. Table (2.15) lists the factors that may influence the performance of the estimators. All of these, which are discussed in the previous chapter, are unknown, except $(1-\delta)$, which is obtained by examining the properties of order flow[j].

As discussed in the previous chapter, the HS model is only influenced by feedback trading:

$$SP = \widehat{SP} - 2 \cdot Cov(\Delta M_{t-1}, BS_{t-1})$$

where $\widehat{SP}$ is the estimated spread of the HS estimator. The second term (the coefficient

[j]See Table (2.13)

Table 2.15: Factors That May Influence Estimations

Parameters	Meaning/Intuition	Value
$Cov\ (\Delta M_t, \Delta M_{t-1})$	covariance of mid-rice returns	unknown
$Cov\ (\Delta M_{t-1}, BS_{t-1})$	Feedback trading	unknown
$Cov\ (\Delta M_{t-2}, BS_{t-1})$	Two-period Feedback trading	unknown
σ_{sp}^2	variance of the spreads	unknown
$1-\delta$	the probability that adjacent trades have different direction	0.465
$1-\varrho$	the proportion of the transaction cost part of the spread	unknown
$(1-\varrho) \cdot \frac{SP}{2}$k	the value of the transaction cost	5.83×10^{-6}
The Ratio	the ratio of the mean of the spread to the standard deviation of mid-price returns	unknown

$(1-\delta)$ is obtained by examining the property of order flow. See Table (2.13)

for BS_{t-1}) in the HS estimator is not strongly influenced by feedback trading and other factors, as the error of the coefficient is not very large. However, as $Cov(\Delta M_{t-1}, BS_{t-1})$ is not known, one cannot evaluate the error of the estimator directly. Furthermore, because the Roll estimator is affected by all the above factors, unlike that in Section (1.6.8), the difference between the estimates of the HS and the Roll estimators is no longer $Cov(\Delta M_{t-1}, BS_{t-1})$. Therefore, at this stage, we cannot obtain an accurate estimate of the spread.

However, if we assume that the estimate of the HS estimator is the true spread, we can examine the errors of the Roll estimator. If the estimate of the HS estimator were the true spread, i.e. $\widehat{SP}_{HS} = 0.0000794$, the features of mid-price returns, order flow, and feedback trading can be obtained, and these are shown in Table (2.16).

The subscript HS indicates that the values are obtained based on the estimated spread of the HS estimator. The mid-price returns are strongly negatively correlated, which suggests that the Roll estimator overestimates the true spread. One-period feedback trading is positive. Because the HS estimator does not take feedback trading into account, one-period feedback trading is filtered when we use the estimated spread of the HS estimator to calculate the mid-prices. Two-period feedback trading is negative and net feedback trading in this case is positive, which suggests that the Roll estimator overestimates the true spread. δ is greater than 0.5, which suggests that order flow is positively autocorrelated, thus, the Roll estimator underestimates the true spread. $\varrho = 85.3\%$ suggests that the IC & AS components are dominant elements in the spread and that the Roll estimator underestimates the true spread.

As discussed in the previous section when there is feedback trading, the Roll estimator

Table 2.16: Features of Mid-Prices Obtained From The HS Estimator

Parameters	Value
$Cov\left(\widetilde{\Delta M}_{t,HS}, \widetilde{\Delta M}_{t-1,HS}\right)$	-7.15×10^{-8}
$Corr\left(\widetilde{\Delta M}_{t,HS}, \widetilde{\Delta M}_{t-1,HS}\right)$	-0.462
$Cov\left(\widetilde{\Delta M}_{t-1,HS}, \Delta BS_{t-1}\right)$	2.2×10^{-6}
$Corr\left(\widetilde{\Delta M}_{t-1,HS}, \Delta BS_{t-1}\right)$	0.0057
$Cov\left(\widetilde{\Delta M}_{t-1,HS}, \Delta BS_{t}\right)$	-4.1×10^{-5}
$Corr\left(\widetilde{\Delta M}_{t-1,HS}, \Delta BS_{t}\right)$	-0.106
$\sigma^2_{sp,HS}$	3.85×10^{-10}
$1-\delta$	0.465
$1-\varrho_{HS}$	14.7%
The Ratio	0.206

$Cov()$ is the covariance
$Corr()$ is the correlation coefficient
$(1-\delta)$ is obtained by examining the property of order flow, which is not influenced by the spread estimation. See Table (2.13)
The meaning and intuition of the parameters are the same as those in Table (2.15)

is biased. Because the effects of feedback trading may last longer than one period, here we assume that there is one-period and two-period feedback trading. When there is feedback trading ($E[\Delta M_{t-1} \cdot \frac{SP}{2}(BS_t - BS_{t-1})] \neq 0$, the true spread becomes

$$\begin{aligned} &Cov(\Delta s_t, \Delta s_{t-1}) = \tfrac{SP}{2} \cdot [Cov(\Delta M_{t-1}, BS_t) - Cov(\Delta M_{t-1}, BS_{t-1})] - \tfrac{SP^2}{4} \\ &SP = Cov(\Delta M_{t-1}, BS_t) - Cov(\Delta M_{t-1}, BS_{t-1}) \\ &\quad + \sqrt{[Cov(\Delta M_{t-1}, BS_t) - Cov(\Delta M_{t-1}, BS_{t-1})]^2 - 4Cov(\Delta s_t, \Delta s_{t-1})} \end{aligned} \tag{2.94}$$

Thus the error is given by,

$$\begin{aligned} Error = {} & 2 \cdot \sqrt{-Cov(\Delta s_t, \Delta s_{t-1})} - Cov(\Delta M_{t-1}, BS_t) + Cov(\Delta M_{t-1}, BS_{t-1}) \\ & - \sqrt{[Cov(\Delta M_{t-1}, BS_t) - Cov(\Delta M_{t-1}, BS_{t-1})]^2 - 4Cov(\Delta s_t, \Delta s_{t-1})} \end{aligned} \tag{2.95}$$

The error is positive when net feedback trading is positive and vice versa.

Equation (2.96) combines all errors of the Roll estimator discussed in the previous and present chapters, where SP is the true spread.

$$Cov(\Delta s_t, \Delta s_{t-1}) = Cov(\Delta M_t, \Delta M_{t-1}) - \frac{SP^2}{4} \cdot [4 \cdot (1-\delta)^2] \cdot [(1-\varrho)]$$

$$-\frac{\sigma_{sp}^2}{4} + \frac{SP}{2} \cdot [Cov(\Delta M_{t-1}, \Delta BS_t) - Cov(\Delta M_{t-1}, \Delta BS_{t-1})]$$

$$SP = \frac{-Feedback + \sqrt{Feedback^2 - 4 \cdot \Theta \cdot \Xi}}{\Xi} \tag{2.96}$$

$$Feedback = [Cov(\Delta M_{t-1}, \Delta BS_{t-1}) - Cov(\Delta M_{t-1}, \Delta BS_t)]$$

$$\Theta = Cov(\Delta s_t, \Delta s_{t-1}) - Cov(\Delta M_t, \Delta M_{t-1}) + \sigma_{sp}^2$$

$$\Xi = [4 \cdot (1-\delta)^2] \cdot [(1-\varrho)]$$

By substituting the parameters above into Equation (2.96), we can obtain the "true" spread if the estimate of the HS estimator is the true spread.

$$SP_{AdjustedRoll} = 6.41 \times 10^{-5} \tag{2.97}$$

The "true" spread is slightly less than the estimated spread of the HS estimator. This suggests that the estimate of the HS estimator is not very far from the true spread.

2.8 Conclusion

Simulation experiments are used to compare the performance of various bid-ask spread estimators and to identify the influence of a number of factors on the estimates.

The Roll estimator performs well in the ideal case. Other generations of the Roll estimator (e.g. Choi et al. 1988, George et al. 1991, Stoll 1989) can obtain accurate estimates under specific conditions. The HS estimator outperforms the Roll and CS estimators for short time intervals in most cases; the only exception was when feedback trading occurs, which might be because the HS model uses more information (the order flow) than the other two estimators. The CS estimator does not yield accurate estimates except in the ideal case. Both the Roll and HS estimators perform better for short time intervals. The CS estimator should not perform well for very short time intervals because of the magnitude of the standard deviation of the estimates.

This chapter emphasises the importance of the ratio of the spread to the volatility of mid-price returns, because it is the only factor that influences all the estimators. When the ratio is large, all estimators produce accurate estimates. The HS estimator has the widest working range of the ratio followed by the CS estimator, while the Roll estimator has the narrowest range.

Generally speaking, the HS model can produce accurate estimates in most cases. Unfortunately, information about the order flow is not always available. Future research should develop a method which can apply the HS model without information about the order flow. The feedback trading case suggests that one cannot use the mid-price returns to predict the order flow Hasbrouck (2004, 2009) uses the Gibbs sampler to infer the order flow and the result is satisfactory. However, his method requires heavy computing. Corwin and Schultz (2012) assume that the highest price of the day corresponds to a buy order and the lowest price corresponds to a sell order which is a useful assumption for order flow trends.

For transaction data, the tick-by-tick HS estimator (7.94×10^{-5}) might produce the best estimates of the true spread. The HS estimator suggests that the inventory control and asymmetric information parts contribute 85.3% of the spread. We calculate mid-prices using the HS spread. The mid-price returns exhibit strong negative autocorrelation. Two-period negative feedback trading is observed. Furthermore, we calculate the adjusted Roll estimator by assuming that the HS spread is the true estimator. The adjusted Roll estimator yields an estimate for the spread of 6.41×10^{-5}, which is not very different from the estimate of the HS estimator. Overall, we find that the estimate of the HS estimator is close to the true spread.

Chapter 3

Decomposing the bid-ask spread: a new indicator model for multi-dealer markets

3.1 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 chapter 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 model thereafter). 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. Because of the unavailability of spread estimates, there is little research analysing the intra-day spread pattern on the Reuters D2000-1 system, which is an important component of the foreign exchange market.

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. 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.

Though the HS model encompasses all other spread models, in practice, it is not widely used because it sometimes delivers ambiguous decomposition results: fractions sometimes are negative or larger than one. This is because the HS model matches better the structure of NYSE which has a single market maker for each stock and is not suitable for markets without market makers (e.g. Electronic Brokerage System, EBS) or for markets with more than one market maker (e.g. Reuters D2000-1). McGroarty et al. (2006) develop an alternative model for decomposing spread components specifically designed for the order-driven intra-dealer market. Their model avoids the problems related to the HS model, but it can only be applied when the spread data are available. In this chapter, we derive a model which can estimate and decompose spreads and which is also suitable for 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.

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. Unlike datasets from stock markets such as NYSE and NASDAQ, foreign exchange datasets provide information about inter-dealer transac-

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

tions.

Because spread information is not always available, the literature studying the intra-day spread pattern on the ReutersD2000-1 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. 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 adverse selection part are stable during the day. The components that can better explain the patterns of the intra-day spreads are the order processing cost part and the competition part. The U-shape of the spread might be caused by the U-shape or inverse U-shape of these determinant variables.

We organize the rest of the chapter as follows. In Section 2, we briefly introduce the HS model. In Section3, we present the MHS model and compare the results of the HS and the MHS models. In Section 4, we introduce time dummies in the MHS model in order to study the intra-day pattern of the spread. We use the MHS model to analyse USD/DEM transaction data collected from the Reuters D2000-1 system.

3.2 Theoretical Background

When the bid-ask spread is included in the data or can be estimated from the data, we can study the factors that influence its value. One such factor is 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 adverse selection cost is incurred when a dealer trades with an agent with better information. Third, inventory control costs arise when a dealer tries to keep an ideal inventory level. 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:

$$SP_t = (OP_t, IC_t, AS_t, COM_t) \tag{3.1}$$

or more specifically,

$$SP_t = OP_t + IC_t + AS_t + COM_t \tag{3.2}$$

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 adverse selection 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:

$$SP_t = \tau_1 + \sum \tau_i timedummy_i + \varepsilon_t \tag{3.3}$$

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 (3.3) will be called the spread description model (SD model) in this chapter.

The Huang and Stoll model (HS model) aims to estimate the bid-ask spread and to decompose its components. The HS model can be called a spread-estimating model.

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:

$$s_t = M_t + \frac{SP}{2} \cdot BS_t \tag{3.4}$$

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

$$BS = \begin{cases} 1 & buy\ order \\ -1 & sell\ order \end{cases} \tag{3.5}$$

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

$$\Delta s_t = \Delta M_t + \frac{SP}{2}(BS_t - BS_{t-1}) \tag{3.6}$$

where Δ is the first-order difference operator.

The mid-price depends on the degree of divergence from the ideal inventory level and the fundamental value of the asset. Formally, the mid-price is given by:

$$M_t = F_t + \beta \cdot \frac{SP}{2} \sum_{i=1}^{t-1} BS_i \tag{3.7}$$

where F_t is the fundamental, and $\frac{SP}{2}$ is the half spread. Taking the first-order difference of Equation (3.7) gives:

$$\Delta M_t = \Delta F_t + \beta \cdot \frac{SP}{2} BS_{t-1} \tag{3.8}$$

Equation (3.8) 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. Deviations from this conditional expectation therefore convey information about the fundamental value. Formally, the change in the fundamental value can be written as:

$$\begin{aligned} \Delta F_t &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} [E(BS_{t-1} | BS_{t-2})] \\ &= \alpha \tfrac{SP}{2} BS_{t-1} - \alpha \tfrac{SP}{2} (1 - 2\theta) BS_{t-2} \end{aligned} \tag{3.9}$$

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:

$$E(BS_{t-1} | BS_{t-2}) = (1 - 2\theta) BS_{t-2} \tag{3.10}$$

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

$$\Delta s_t = \frac{SP}{2} BS_t + (\alpha + \beta - 1) \frac{SP}{2} BS_{t-1} - \alpha \frac{SP}{2} (1 - 2\theta) BS_{t-2} + \varepsilon_t \tag{3.11}$$

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

$$BS_{t-1} = \vartheta BS_{t-2} + \varepsilon_t \tag{3.12}$$

We use the generalised method of moments 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 adverse selection 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.

3.3 Modified HS Model

In this section, we modify the HS model to fit the structure of a multiple-dealer market.

The HS model relies on the assumption that the probability that the trade direction reverses (θ) is higher than 0.5, which suggests that order flow is negatively correlated. This is a reasonable assumption for a single-dealer market. After receiving a buy (sell) order, to rebalance the inventory, a dealer would move the mid-price up (down) to attract a sell/buy order. Consequently, the order flow is negatively correlated because of the dealer's inventory control behaviour. If θ is less than 0.5, the weight of the adverse selection costs might be negative. This model also does not perform well when the probability is near 0.5, because the weight of adverse selection costs might be larger than one. McGroarty et al. (2006) suggest that this assumption is not suitable for "the electronic inter-dealer spot foreign exchange market" and introduce a new model to decompose the components of the spread that is specifically suited to the order-driven intra-dealer market. Their model avoids the ambiguous results of the HS model, but it can only be applied when spread data are available. Thus, both the HS and the McGroarty et al. (2006)) models are not appropriate neither for a quote-driven market nor when the spread is not available, and thus we need to build a new model.

The basic structure of the HS model is retained. The dealer who gives the bid-ask quotes decides the mid-price and the bid-ask spread. He chooses the mid-price according to the fundamental value of the asset and the current inventory level, which is equal to the stock of historic order flow. When there is no inventory control cost, the mid-price equals the fundamental. When the fundamental is not taken into account, the mid-price depends on the degree of divergence from the ideal inventory level, as in Ho and Stoll 1981). Formally, the mid-price is given by:

$$M_t = F_t + s_t^I \cdot I_t \tag{3.13}$$

where F_t is the fundamental, s_t^I is the movements of the price caused by inventory imbalance, and I_t is the sign of the inventory imbalance.

We assume each order has roughly the same size (Evans 2002), so that the sum of the trade indicators is an integer and a cardinal measure of inventory.

We discuss inventory control first. If the ideal inventory level is zero (Evans 2002), then when there are more sell orders than buy orders, the dealer tends to reduce the mid-price in order to increase demand for the asset and vice versa. When the fundamental does not change and there is a buy (sell) order in the previous period, the dealer has to increase (reduce) both the ask and the bid prices as much as possible to return the inventory level back to the ideal, assuming that a decrease in price increases demand. When there is a buy order, the new bid price should not be higher than the fundamental, which equals the mid-quote in the previous period. A sell order follows a similar guideline. Thus, the change in the mid-price per unit inventory imbalance should equal half of the spread caused by inventory control costs, and it should be a proportion of the total spread if there are other factors influencing the spread. Formally,

$$| s_t^I | = \frac{SP^I}{2} = \beta \cdot \frac{SP}{2} \tag{3.14}$$

where β is the weight of inventory control costs in the total spread. Because we assume that a reduction in price will increase demand, if the inventory level is higher than ideal, the dealer will raise the price. Thus, we can remove the absolute value signs from Equation (3.14).

$$s_t^I \cdot I_t = \frac{SP^I}{2} \cdot I_t = \beta \cdot \frac{SP}{2} \cdot I_t \tag{3.15}$$

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:

$$Z = k_1 \cdot X \tag{3.16}$$

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,

$$k_1 \sum_{i=1}^{t-1} BS_i \tag{3.17}$$

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 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). Lyons (1997) and Evans and Lyons (2002) emphasise that in the foreign exchange interdealer 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 3.7). 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 tolerable inventory before the most recent order is given by:

$$k_2 \sum_{i=1}^{t-2} BS_{i,d} \tag{3.18}$$

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:

$$
\begin{aligned}
I_{t,d} &= k_1 BS_{t-1,IN,d} + \sum_{i=1}^{t-2} BS_{i,IN,d} + \sum_{i=1}^{t-2} BS_{i,OUT,d} \\
&= k_1 BS_{t-1,IN,d} + k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN,d} + (1-k_2) \cdot \sum_{i=1}^{t-2} BS_{i,IN,d} + \sum_{i=1}^{t-2} BS_{i,OUT,d} \\
&= k_1 BS_{t-1,IN,d} + k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN,d} + 0 \\
&= k_1 BS_{t-1,IN,d} + k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN,d}
\end{aligned}
\tag{3.19}
$$

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 (3.17) 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.

For example, according to equations (3.15) and (3.19), dealer *d*'s quote increment after receiving an incoming order at period $t-1$ ($BS_{t-1,IN,d}$) is given by:

$$
\begin{aligned}
s^I_{t,d} \cdot I_{t,d} &= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-1,IN,d} + k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN,d} \right) \\
&= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-1} + \tfrac{k_1}{N} \cdot k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN} \right)
\end{aligned}
\tag{3.20}
$$

Similarly, if dealer *D*'s received an incoming order ($BS_{t-2} = BS_{t-2,IN,D}$) at period $t-2$, dealer *D*'s quote increment after receiving the incoming order is given by:

$$
\begin{aligned}
s^I_{t-1,D} \cdot I_{t-1,D} &= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-2,IN,D} + k_2 \cdot \sum_{i=1}^{t-3} BS_{i,IN,D} \right) \\
&= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-2} + \tfrac{k_1}{N} \cdot k_2 \cdot \sum_{i=1}^{t-3} BS_{i,IN} \right)
\end{aligned}
\tag{3.21}
$$

As mentioned earlier, the price increments are the same throughout the market, thus:

$$
\begin{aligned}
s_t^I \cdot I_t &= s_{t,D}^I \cdot I_{t,D} = \dots = s_{t,d}^I \cdot I_{t,d} \\
&= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-1} + \tfrac{k_1}{N} \cdot k_2 \cdot \sum_{i=1}^{t-2} BS_{i,IN} \right) \\
s_{t-1}^I \cdot I_{t-1} &= s_{t-1,D}^I \cdot I_{t-1,D} = \dots = s_{t-1,d}^I \cdot I_{t-1,D} \\
&= \beta \cdot \tfrac{SP}{2} \cdot \left(k_1 BS_{t-2} + \tfrac{k_1}{N} \cdot k_2 \cdot \sum_{i=1}^{t-3} BS_{i,IN} \right)
\end{aligned} \tag{3.22}
$$

Equations (3.22) suggest that, although the liquidity suppliers (the dealers who receive the order) and thus the owner of the unwanted inventory may be different in different periods, the inventory information will get into the quote in the market.

Therefore, we re-write Equation (3.7) as follows:

$$
M_t = F_t + \beta \cdot \frac{SP}{2} \left[k_1 BS_{t-1} + \frac{1}{N} \cdot k_1 \cdot k_2 \cdot \sum_{i=1}^{t-2} BS_i \right] \tag{3.23}
$$

Taking the first-order difference of Equation (3.23), we obtain that:

$$
\Delta M_t = \Delta F_t + \beta \cdot \frac{SP}{2} \cdot k_1 BS_{t-1} + \beta \cdot \frac{SP}{2} \cdot \left(\frac{k_1 k_2}{N} - 1 \right) \cdot BS_{t-2} \tag{3.24}
$$

Equation (3.24) 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.

Now we discuss the other factor that influences the mid-price, the fundamental value. Our setting is similar to the simple version of the HS model. When the market has asymmetric information, not all market participants know the fundamental value. Dealers can therefore use the spread to compensate their losses from trading with agents who have better information, and the incoming order is a source of information about the fundamental value. A buy order suggests that the fundamental value has gone up and thus the mid-quote will increase. Glosten and Milgrom (1985) suggest that if adverse selection is the only source of the spread, bid (ask) prices are the supremum (infimum) of the fundamental value. Therefore, when there is a buy order, the dealer tends to quote a new mid-price equal to the ask-price in the previous period. When there is a sell order, the dealer tends to quote a new mid-price equal to the bid-price in the previous period. Formally, the influence of adverse selection caused by an incoming order on the mid-prices is given by

$$|\Delta s_t^{AS}| = \frac{SP^{AS}}{2} \tag{3.25}$$

where Δs_t^{AS} is the change of the mid-price due to the adverse selection, SP^{AS} is the spread caused by asymmetric information. When there is not only adverse selection but also inventory control costs, order processing costs, and competition in the market, the spread caused by asymmetric information is a proportion of the total spread:

$$|\Delta s_t^{AS}| = \frac{SP^{AS}}{2} = \alpha \cdot \frac{SP}{2} \tag{3.26}$$

Similar to the HS model, the changes in the fundamental values are given by the formula,

$$\begin{aligned} \Delta F_t &= \Delta s_t^{AS} \cdot AS_{t-1} - \Delta s_t^{AS} \cdot [E(AS_{t-1}|AS_{t-2})] \\ &= \tfrac{SP^{AS}}{2} \cdot AS_{t-1} - \tfrac{SP^{AS}}{2} \cdot [E(AS_{t-1}|AS_{t-2})] \\ &= \alpha \cdot \tfrac{SP}{2} \cdot AS_{t-1} - \alpha(1-2\theta) \cdot \tfrac{SP}{2} \cdot AS_{t-2} \end{aligned} \tag{3.27}$$

where AS_{t-1} is the signed volume of the incoming order. As in the inventory control case, we assume all the trades have the same size so that the trade indicator is also the signed volume of the trade. Formally, the signed volume of the incoming order is given by:

$$AS_{t-1} = BS_{t-1};\ AS_{t-2} = BS_{t-2} \tag{3.28}$$

Equation (3.27) becomes

$$\Delta F_t = \frac{SP^{AS}}{2} \cdot BS_{t-1} = \alpha \cdot \frac{SP}{2} \cdot BS_{t-1} - \alpha(1-2\theta) \cdot \frac{SP}{2} \cdot BS_{t-1} \tag{3.29}$$

Both adverse selection and inventory control costs, which influence the mid-price, have been discussed. Substituting Equation (3.29) into Equation (3.24) gives the change in mid-price,

$$\begin{aligned} \Delta M_t &= \alpha \cdot \tfrac{SP}{2} \cdot BS_{t-1} - \alpha(1-2\theta) \cdot \tfrac{SP}{2} \cdot BS_{t-1} + \beta \cdot \tfrac{SP}{2} \cdot k_1 BS_{t-1} + \beta \cdot \tfrac{SP}{2} \cdot \left(\tfrac{k_1 k_2}{N} - 1\right) BS_{t-2} \\ &= (\alpha + k_1\beta) \cdot \tfrac{SP}{2} \cdot BS_{t-1} - \left[\alpha(1-2\theta) + \beta\left(1 - \tfrac{k_1 k_2}{N}\right)\right] \cdot \tfrac{SP}{2} \cdot BS_{t-2} \end{aligned} \tag{3.30}$$

Taking the order possessing cost into account, we can finally obtain the modified HS model. Substituting Equation (3.30) into Equation (3.6), we have,

$$\Delta s_t = \frac{SP}{2}BS_t + (\alpha + k_1\beta - 1)\frac{SP}{2}BS_{t-1} - \left[\alpha(1-2\theta) + \beta\left(1 - \frac{k_1k_2}{N}\right)\right]\frac{SP}{2}BS_{t-2} + \varepsilon_t \qquad (3.31)$$

The new model will be called the modified HS model (MHS model for short) in this chapter.

Equation (3.31) 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

$$\Delta s_t = \frac{SP}{2}BS_t + (\alpha + \beta - 1)\frac{SP}{2}BS_{t-1} - [\alpha(1-2\theta) + \beta]\frac{SP}{2}BS_{t-2} + \varepsilon_t \qquad (3.32)$$

The decomposition results show the lower bound of the inventory component and the upper bound of the asymmetric information component. Equation (3.32) 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.

3.4 The Comparison of the HS and MHS Models in the Reuters D2000-1 System

In this section, the performance of the HS and MHS models are compared using USD/DEM data. We first discuss how the MHS model fits the microstructure of the Reuters D2000-1 system. Then we compared the results of the HS and the MHS models empirically.

3.4.1 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 the quote 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. We assume there is only one trade per time no matter how intensive the trades are in the market. 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.

3.4.2 Evans and Lyons's Model and the MHS Model

Evans and Lyons (2002) develop a model to describe the price formation process of the Reuters D2000-1 system. Although their model does not include the bid-ask spread and thus is not directly related to the MHS model, their model is the workhorse of research based on the Reuters D2000-1 system. If we consider the price in Evans and Lyons's model (the EL model thereafter) as the mid-price (because it assumes there is no spread), the MHS model does share similar intuitions with the EL model.

The EL model has two basic assumptions: the quote price and the net inter-dealer order flow are observed by all participants in the inter-dealer market. There are three rounds. In the first round, dealers trade with the non-dealer public and collect private information from customer order flow. In the second round, dealers trade with each other and aggregate the private information. In the third round, dealers quote the price based on the aggregate information from the second round and trade with the public again. To maximise the welfare at the end of the third round, traders choose their quotes at the beginning of each round and choose their outgoing orders in the second round.

The MHS model decomposes the inter-dealer bid-ask spread, so we focus on the inter-dealer trading round (round two) of the EL model. At the beginning of the second round, dealer *d*'s quote is $M_{2,d}$ based on the public information, and because of the no-arbitrage condition, all dealers' quotes are the same M_2. Then all trades in the second round are at price M_2. Dealer *d*'s outgoing order in round two is given by:

$$T_{2,d} = c_{1,d} + D_{2,d} + E\left[T'_{2,d} \mid \Omega_{T_{2,d}}\right] \quad (3.33)$$

where $D_{2,d}$ is dealer *d*'s demand in round two; $c_{1,d}$ is customer order flow that dealer d received in round one; $E\left[T'_{2,d} \mid \Omega_{T_{2,d}}\right]$ is dealer *d*'s expectation of what he/she will receive from other dealers. Equation (3.33) suggests that dealer d uses the active method discussed in the previous section to manage his/her inventory. $c_{1,d}$ and $E\left[T'_{2,d} \mid \Omega_{T_{2,d}}\right]$ are the unwanted inventories and dealer d gets rid of them by one trade efficiently. Furthermore, when dealer *d* makes his/her order, he/she receives orders from others, which suggests that the latest incoming order brings unwanted inventory. These facts coincide with the intuition about inventory control in the MHS model. Incoming orders other than dealer *d*'s expectation ($E\left[T'_{2,d} \mid \Omega_{T_{2,d}}\right]$) bring private information about the price from other dealer. This is consistent with the asymmetric information feature of the MHS model. After trading in round two, dealers quote a new price M_3. The price increment is given by:

$$\Delta M = M_3 - M_2 = \alpha \sum_{d=1}^{N} T_{2,d} \tag{3.34}$$

where $\sum_{d=1}^{N} T_{2,d}$ is the net inter-dealer order flow observed by dealers in the market, and α is a coefficient. Equation (3.34) suggests that the net order flow influences the mid-price, which is the most important conclusion of the EL model. $\sum_{d=1}^{N} T_{2,d}$ includes inventory control and asymmetric information components. The MHS model shares similar interpretation, in that both the inventory control and asymmetric information moves the mid-price.

3.4.3 Empirical Result Reuters D2000-1 USD/DEM 1996

The transaction data of the USD/DEM pair from 1996.5.1 to 1996.9.2 are taken from Reuters D2000-1.

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.

Table (3.1) shows the results of the regressions of the HS and the LIMHS models. Table (3.2) gives the interpretations of the coefficients of these regressions.

γ_1 in Table (3.1) is the half spread, and thus 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 both models find that these factors explain 14.39% of the spread according to Table (3.2) . The difference between the HS and the LIMHS models is the interpretation of γ_3, which influence the decomposition of the spread. According to Table (3.2), the HS model suggests that the contributions of the inventory control cost and the asymmetric information cost to the spread are 19.42% and 66.18% respectively. The inventory control cost is not significantly different from zero. The LIMHS model suggests that the shares of the inventory control cost and the asymmetric information cost in the spread are -1.48% and 87.08% respectively. Similar to the HS model, the inventory control cost in the LIMHS model is also not significantly different from zero. Both the HS and the LIMHS models find that the asymmetric information cost is a dominant component of the spread. According to the EL 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 managing, 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 does. 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 3.1: 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*** (52.03)	-0.00572*** (-7.50)	-0.00186* (-2.44)	-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
0.0706*** (35.87)			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 LIMHS 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 3.2: Explanations of the Regressions

USD/DEM					
	spread	OP	IC	AS	θ
HS	0.0000794	14.39%	19.42%	66.18%	0.4647
	(1.53×10^{-6})	(0.0191)	(0.271)	(0.271)	(9.84×10^{-4})
LIMHS	0.0000794	14.39%	-1.48%	87.08%	0.4647
	(1.53×10^{-6})	(0.0191)	(0.72)	(0.0281)	(9.84×10^{-4})

This table present the results of the HS model (regressions 3.11 and 3.12) and the LIMHS model (regressions 3.31 and 3.12).
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 calculated from the results of regression (3.12)
Standard errors are shown in brackets

3.5 Positively Autocorrelated Order Flow

3.5.1 Sources of Positive Autocorrelation

Positive autocorrelation in order flow is always observed in financial markets. There could be two reasons for this.

The first is order-splitting, mentioned in Huang and Stoll (1997), which is the tendency of traders to split big orders into smaller ones executing at the same price. Kyle (1985) shows the strategic motivation of order splitting which is that because order flow includes dealers' private information, dealers try not to make a big order so as to avoid large impact on prices. Huang and Stoll show that larger orders bring big inventory pressure on dealers and thus have a great impact on inventory component of the spread. Using high-frequency data, researchers find evidence of order splitting, e.g. Chan and Lakonishok (1993, 1995), Chordia et al. (2002, 2005) and Lillo and Farmer (2005).

The second is hot-potato trading, modelled by Lyons (1997), which is the phenomenon that "inventory imbalances are passed from dealer to dealer, and this occurs independently of whether they offset the imbalance of the receiving dealer", which causes positive autocorrelation. Following Lyons (1997), theoretical papers such as Evans and Lyons (2002) and Viswanathan and Wang (2004) use this definition. The basic idea is that after accepting an income order, which move the inventory level away from the ideal one, a dealer tends to pass the unwanted order to another to bring the inventory back to the balance. Furthermore since orders reflect traders' private information about the fundamental value of the price, private information spreads with orders through hot-potato trading across the market. Berger et. al (2008) observe positively autocorrelated orders using EBS data.

The HS model assumes negative autocorrelation in order flow because after receiving a sell order, a dealer will reduce the mid-price to attract a buy order and to recover the inventory balance. Huang and Stoll (1997) suggest that because of splitting orders, order flow exhibits positive autocorrelation which may generate ambiguous results. To avoid positive autocorrelation, Huang and Stoll discuss two possible approaches (page 1019). The first approach is to aggregate a series of orders at the same price and direction together as one order. The second approach is to "propose a model of order submission by investors which could be used to separate demand side serial correlation in order flow from microstructure-induced serial correlation". Using approach one, Huang and Stoll obtain negatively autocorrelated order flow and less ambiguous decomposing results (table 6 in their paper). Furthermore, by categorising orders according to the sizes of orders, Huang and Stoll identify the various impacts of small, median and big orders on prices. Our data do not have information about order size, therefore, it is impossible for us to distinguish the order sizes. And also, the source of positive autocorrelation in our data is widely considered as type two: hot-potato trading, which does not very fit the purpose of Huang and Stoll's solution. MHS model, unlike the HS model, does not make assumption on the autocorrelation of order flow.

3.5.2 The MHS Model with Long Autocorrelated Order Flow

Tables 3.3 and 3.4 describe the autocorrelation of order flow. Because order flow is a binomial variable, the probit model is used. The results show that order flow exhibits long and positive autocorrelation. The first 32 lags are statistically significant different from zero.

The results of the main regression might be influenced by the autocorrelation of order flow. In this section, we run the main regression controlling for more lags of order flow. The results are shown in Table 3.5. The results are not very different from the original HS model, which suggests that the results are robust to autocorrelated order flow.

Table 3.3: Autocorrelation of Order Flow

$$BS_t = \phi\left(\beta + \sum_{i=1}^{32} \beta_i \cdot BS_{t-i}\right) + \varepsilon_t$$

BS_{t-1}	BS_{t-2}	BS_{t-3}	BS_{t-4}	BS_{t-5}	BS_{t-6}	BS_{t-7}
0.136***	0.107***	0.092***	0.077***	0.066	0.060***	0.057***
(27.06)	(21.28)	(18.24)	(15.21)	(13.01)	(11.9)	(11.38)
BS_{t-8}	BS_{t-9}	BS_{t-10}	BS_{t-11}	BS_{t-12}	BS_{t-13}	BS_{t-14}
0.041***	0.045***	0.038***	0.033***	0.035***	0.038***	0.034***
(8.17)	(8.93)	(7.54)	(6.5)	(6.92)	(7.5)	(6.67)
BS_{t-15}	BS_{t-16}	BS_{t-17}	BS_{t-18}	BS_{t-19}	BS_{t-20}	BS_{t-21}
0.030***	0.024***	0.013**	0.017**	0.024***	0.022***	0.017***
(5.99)	(4.81)	(2.64)	(3.44)	(4.77)	(4.29)	(3.42)
BS_{t-22}	BS_{t-23}	BS_{t-24}	BS_{t-25}	BS_{t-26}	BS_{t-27}	BS_{t-28}
0.015**	0.013**	0.022**	0.015***	0.023**	0.017**	0.012*
(2.94)	(2.63)	(4.26)	(3.01)	(4.57)	(3.39)	(2.44)
BS_{t-29}	BS_{t-30}	BS_{t-31}	BS_{t-32}	constant	Obs	Pseudo R^2
0.021***	0.010	0.017**	0.016**	-0.593***	256326	0.0143
(4.16)	(1.9)	(3.46)	(3.16)	(-59.87)		

Z-statistics is in the parenthesis

* Significant at 5% level
** Significant at 1% level
*** Significant at 0.1% level

Table 3.4: Autocorrelation of Order Flow: Marginal Effects

BS_{t-1}	BS_{t-2}	BS_{t-3}	BS_{t-4}	BS_{t-5}	BS_{t-6}	BS_{t-7}
0.054***	0.043***	0.037***	0.031***	0.026***	0.024***	0.023***
(27.1)	(21.3)	(18.25)	(15.22)	(13.01)	(11.9)	(11.39)
BS_{t-8}	BS_{t-9}	BS_{t-10}	BS_{t-11}	BS_{t-12}	BS_{t-13}	BS_{t-14}
0.016***	0.018***	0.015***	0.013***	0.014***	0.015***	0.013***
(8.17)	(8.93)	(7.54)	(6.5)	(6.92)	(7.5)	(6.67)
BS_{t-15}	BS_{t-16}	BS_{t-17}	BS_{t-18}	BS_{t-19}	BS_{t-20}	BS_{t-21}
0.012***	0.010***	0.005***	0.007***	0.010***	0.009***	0.007***
(5.99)	(4.81)	(2.64)	(3.44)	(4.77)	(4.29)	(3.42)
BS_{t-22}	BS_{t-23}	BS_{t-24}	BS_{t-25}	BS_{t-26}	BS_{t-27}	BS_{t-28}
0.006	0.005	0.009	0.006	0.009	0.007	0.005
(2.94)	(2.63)	(4.26)	(3.01)	(4.57)	(3.39)	(2.44)
BS_{t-29}	BS_{t-30}	BS_{t-31}	BS_{t-32}			
0.008	0.004	0.007	0.006			
(4.16)	(1.9)	(3.46)	(3.16)			

Z-statistics is in the parenthesis

* Significant at 5% level
** Significant at 1% level
*** Significant at 0.1% level

Table 3.5: Autocorrelation of Order Flow

$$\Delta s_t = \beta + \beta_1 \cdot BS_t + \beta_2 \cdot BS_{t-1} + \beta_3 \cdot BS_{t-2} + \sum_{i=3}^{32} \beta_i \cdot BS_{t-i} + \varepsilon_t$$

BS_t	BS_{t-1}	BS_{t-2}	BS_{t-3}	BS_{t-4}	BS_{t-5}	BS_{t-6}
0.403***	-0.0516***	-0.0115	-0.0172*	-0.0164*	-0.00193	-0.0182*
(52.71)	(-6.74)	(-1.50)	(-2.24)	(-2.14)	(-0.25)	(-2.37)
BS_{t-7}	BS_{t-8}	BS_{t-9}	BS_{t-10}	BS_{t-11}	BS_{t-12}	BS_{t-13}
-0.00436	-0.0203**	-0.0117	-0.00664	-0.0113	0.00496	-0.00970
(-0.57)	(-2.64)	(-1.52)	(-0.86)	(-1.47)	(0.65)	(-1.26)
BS_{t-14}	BS_{t-15}	BS_{t-16}	BS_{t-17}	BS_{t-18}	BS_{t-19}	BS_{t-20}
-0.00252	-0.00559	-0.00142	-0.00623	0.00761	-0.0215**	-0.00173
(-0.33)	(-0.73)	(-0.19)	(-0.81)	(0.99)	(-2.80)	(-0.23)
BS_{t-21}	BS_{t-22}	BS_{t-23}	BS_{t-24}	BS_{t-25}	BS_{t-26}	BS_{t-27}
0.00550	-0.00837	-0.00824	-0.00693	0.00262	0.00423	-0.00155
(0.72)	(-1.09)	(-1.07)	(-0.90)	(0.34)	(0.55)	(-0.20)
BS_{t-28}	BS_{t-29}	BS_{t-30}	BS_{t-31}	BS_{t-32}	Obs	R^2
-0.00980	0.00794	-0.0111	0.00301	-0.0110	256326	0.01
(-1.28)	(1.03)	(-1.44)	(0.39)	(-1.44)		

Z-statistics is in the parenthesis

* Significant at 5% level
** Significant at 1% level
*** Significant at 0.1% level

3.6 An Extension of the LIMHS Model: Intra-day Patterns of the Spread and Its Components

In this section, we study the intra-day pattern of the bid-ask spread. Most previous papers that focus on intra-day patterns have used spread data, but data on the USD/DEM pair do not include spread data. Thus, it is not possible to run the SD model (equation 3.3) directly. We could first estimate the spread using the LIMHS model and then run the SD model, but this process requires two steps. Instead, we extend the LIMHS model with time dummy variables (MHSD model thereafter). By introducing time dummies and interaction terms between time dummies and trade indictor variables (BS), we can capture the intra-day pattern of the spread as well as the intra-day patterns of the components of the spread in one single regression. This method uses less coding and computing time than the two-step method.

The LIMHS model can also decompose components of the spread, so we can apply the MHSD model to study the intra-day patterns of the components of the spread. To incorporate the time dummy variables into the HS model, we can substitute Equation (3.3) into Equation (3.31) and control the time dummy variables in the intercept term, so that the new model is now given by:

$$\begin{aligned}\Delta s_t = &\tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \cdot BS_t + \tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \\ &\cdot(\beta + \alpha - 1) \cdot BS_{t-1} - \tfrac{1}{2}(\tau_1 + \textstyle\sum \tau_i \cdot timedummy_i) \cdot [\alpha(1-2\theta) + \beta] \cdot BS_{t-2} \\ &+ \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t\end{aligned} \tag{3.35}$$

Re-arranging the equation, we have,

$$\begin{aligned}\Delta s_t = &\tfrac{1}{2}[\tau_1 \cdot BS_t + \tau_1 \cdot (\beta + \gamma - 1) \cdot BS_{t-1} - \tau_1 \cdot [\alpha(1-2\theta) + \beta] \cdot BS_{t-2}] + \tfrac{1}{2} \cdot \Sigma[\tau_i \\ &\cdot timedummy_i \cdot BS_t + \tau_i \cdot (\beta + \gamma - 1) \cdot timedummy_i \cdot BS_{t-1} - \tau_i \\ &\cdot [\alpha(1-2\theta) + \beta] \cdot timedummy_i \cdot BS_{t-2}] + \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t \\ = &\tfrac{1}{2}[\tau_1 \cdot BS_t + \Phi_1 \cdot BS_{t-1} - \Lambda_1 \cdot BS_{t-2}] + \tfrac{1}{2} \cdot \Sigma[\tau_i \cdot timedummy_i \\ &\cdot BS_t + \Phi_i \cdot timedummy_i \cdot BS_{t-1} - \Lambda_i \cdot timedummy_i \cdot BS_{t-2}] \\ &+ \textstyle\sum \mu_i \cdot timedummy_i + \varepsilon_t\end{aligned} \tag{3.36}$$

where $\Phi_i = \tau_i \cdot (\beta + \alpha - 1)$ and $\Lambda_i = \tau_i \cdot [\alpha(1-2\theta) + \beta]$, then we can obtain all the parameters.

Regression equation (3.36) is MHSD model.

There are two groups of time dummies: (1) for studying the intra-day paten, 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 (2) for studying the week paten, we use 4 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 (**??**), (3.6) and (3.7) show coefficients of trade indicators (BS) and decomposition results at each time interval, which are calculated using the regression results.

Table (3.6) shows the intra-day pattern of the spread. We use 23 hour-dummies in regression (3.36). 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, adverse selection 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 adverse selection costs is between 70% and 90%, which coincides with the results in section (3.3). 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 adverse selection 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 flow in the interdealer market may include more fundamental information from end-users, and thus, dealers may face a greater asymmetric information cost.

Table (3.7) shows the week pattern of the spread. We use 4 day-name-dummies in regression (3.36). 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 (2006a), 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 (3.1) 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 3.6: 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
0 - 1	2.94	-0.167	-0.0823	0.463	5.68 %	-4.50 %	98.82 %
1 - 2	1.181	0.147	0.0224	0.417	-12.45 %	-24.72 %	137.17 %
2 - 3	0.905	0.552	-0.541	0.431	-60.99 %	43.64 %	117.35 %
3 - 4	4.44	-1.757	-0.0662	0.449	39.57 %	-5.27 %	65.70 %
4 - 5	7.98	-1.317	1.496	0.448	16.50 %	-30.63 %	114.12 %
5 - 6	1.751	-0.867	0.514	0.417	49.51 %	-45.22 %	95.71 %
6 - 7	3.92	-1.067	-0.119	0.450	27.22 %	-4.70 %	77.48 %
7 - 8	3.85	-0.877	0.115	0.463	22.78 %	-9.34 %	86.56 %
8 - 9	3.86	-0.557	-0.0668	0.472	14.43 %	-3.17 %	88.74 %
9 - 10	3.83	-0.567	0.0147	0.468	14.80 %	-6.18 %	91.37 %
10 - 11	3.86	-0.353	-0.423	0.476	9.15 %	7.00 %	83.85 %
11 - 12	4.16	-0.301	-0.39	0.470	7.24 %	4.06 %	88.70 %
12 - 13	4.2	-0.617	0.0698	0.467	14.69 %	-7.89 %	93.20 %
13 - 14	4.16	-0.657	-0.236	0.471	15.79 %	0.77 %	83.44 %
14 - 15	3.92	-0.567	-0.108	0.468	14.46 %	-2.91 %	88.45 %
15 - 16	3.62	-0.627	-0.12	0.468	17.32 %	-2.18 %	84.85 %
16 - 17	5.36	-0.907	-0.0649	0.463	16.92 %	-5.32 %	88.40 %
17 - 18	5.02	0.253	-1.524	0.460	-5.04 %	23.94 %	81.10 %
18 - 19	5.21	-0.937	-1.264	0.446	17.98 %	17.31 %	64.70 %
19 - 20	4.48	-0.365	-1.194	0.446	8.15 %	18.81 %	73.05 %
20 - 21	4.04	-0.547	1.186	0.473	13.54 %	-35.94 %	122.40 %
21 - 22	4.06	0.174	-0.0711	0.474	-4.29 %	-3.93 %	108.22 %
22 - 23	4.77	-0.947	-1.524	0.469	19.85 %	28.80 %	51.35 %
23 - 24	1.59	0.553	-0.0642	0.458	-34.78 %	-7.84 %	142.62 %

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 3.36) with 23 hour-dummies.
Settings in this table are the same as in Table (**??**)

Table 3.7: 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
Monday	3.92	-0.437	-0.23	0.467	11.15 %	-0.02 %	88.87 %
Tuesday	4	-0.475	-0.22	0.469	11.88 %	0.09 %	88.04 %
Wedsday	3.89	-0.644	-0.14	0.465	16.56 %	-2.44 %	85.88 %
Thursday	3.82	-0.669	0	0.458	17.51 %	-7.53 %	90.02 %
Friday	4.22	-0.606	-0.35	0.465	14.36 %	2.44 %	83.20 %

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 3.36) with 4 day-dummies.
Settings in this table are the same as Table (**??**)

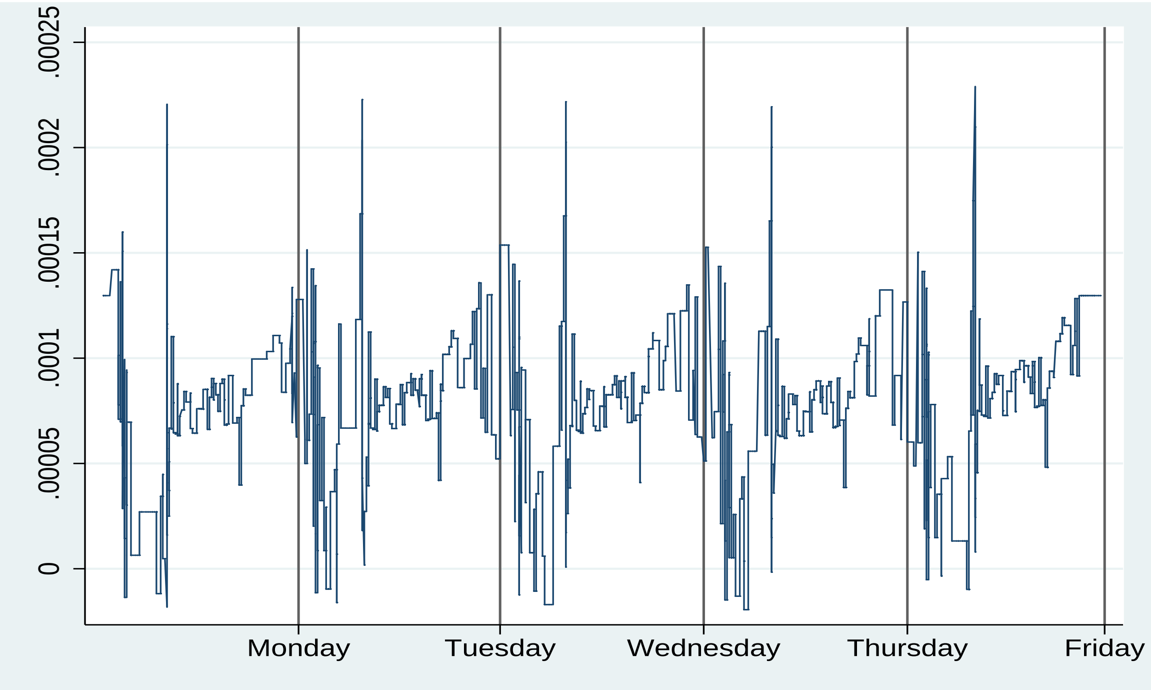

Figure 3.1: Week Spread USDDEM

This figure shows the time pattern of the spread in a whole week.
The spread is estimated by the MHSD model (Equation 3.36) with 11 five-minute-dummies, 23 hour-dummies and 4 day-dummies.

3.7 Conclusion

In this chapter, we have introduced a modified version of the spread-estimating and decomposition model in Huang and Stoll (1997). The new model is suitable for the multi-dealer market where dealers can place direct orders to manage their inventory and, in addition, it overcomes the drawback of the HS model which yields ambiguous results when it is used to decompose the components of the spread. For instance, the HS model sometimes assigns either negative or greater than one values when measuring the fraction of some part of the spread. 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 adverse selection part is the dominant component of the spread and the weight of the inventory control part is very low.

Incorporating time dummies, the modified HS model can be used to analyse the intraday pattern of the spread as well as the components of the spread. Spreads have a hill-shaped pattern during off-peak trading hours. Spreads are stable during peak trading hours and are lower after the closing of the European markets. The weights of the components of spreads do not change significantly during a day.

The autocorrelation of the order flow and the influence of price returns on the order flow (feedback trading) are not considered in this chapter, while these factors are of interest because they include the information of market makers' strategy of inventory control. In future research, one may use the factors to find market makers' allocation of the two methods of inventory control.

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

$$(1 - v)[(1 - v)I_a + vI_b] + v[(1 - v)I_b + vI_a] \tag{3.37}$$

dealer b's imbalance is given by

$$v[(1 - v)I_a + vI_b] + (1 - v)[(1 - v)I_b + vI_a] \tag{3.38}$$

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.

$$\begin{aligned} I_{a,n} &= \frac{[1-(1-2v)^{n-1}]I_b + [1+(1-2v)^{n-1}]I_a}{2} \\ I_{b,n} &= \frac{[1+(1-2v)^{n-1}]I_b + [1-(1-2v)^{n-1}]I_a}{2} \end{aligned} \tag{3.39}$$

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.

Chapter 4

Gambler's Fallacy and Hot Hand Fallacy in the Foreign Exchange Market

4.1 Introduction

The market microstructure theory (e.g. Evans and Lyons 2002) suggests that order flow conveys information and influence exchange rate returns. The reverse effect from exchange rate returns to order flow, called feedback trading, has also attracted much attention both at the theoretical (De Long et al. 1990, Barberis et al. 1998 and Rabin and Vayanos 2010) and the empirical level Daníelsson and Love (2006). One potential explanation of feedback trading is that it arises because of traders' behavioural biases. In this chapter, we follow this line of research by analysing the role of two such biases, the gambler's fallacy (GF) and the hot hand fallacy (HHF), in feedback trading. Research in both economics and psychology suggests that when agents predict the next value of a random series, they frequently exhibit these two types of biases.

Although there is much evidence from laboratory experiments regarding GF and HFF, evidence from the field is scarce. In particular, to date, research analysing index series (e.g., prices, returns, etc.) has focused on agents' reactions to the streak length (i.e., the number of consecutive observations with the same sign) but has ignored the strength of streaks.

The latter is measured by streak width, which is the distance between the current level and the point where a trend began.

In this chapter, we analyse the effects of GF and HHF, using data from the foreign exchange market. Financial markets such as the foreign exchange market are of particular interest because in such markets, these fallacies should be costly to traders, in contrast to lotteries, where outcomes are random. We identify the effects of the streak length and streak width separately. We find that professional traders in the foreign exchange market react to both the streak length and streak width. In particular, traders' behaviour exhibits a U-shaped pattern as the streak length increases and a hill-shaped pattern as streak width increases. In the final part of the chapter, we investigate whether observed trading behaviour derives from a psychological fallacy or reflects rational exploitation of patterns in the data.

The rest of the chapter is organised as follows. In section 2, we review related studies from both the economic and the psychology literatures. In section 3, we describe the raw data and use the modified HS model developed in the previous chapter to deal with raw data and estimate the spread; we also define the streak length and width formally. In section 4 we analyse the relationship among traders' behaviour (the order flow), streak length and width using the probit model. Our results of sections 3 and 4 have uncovered a statistically significant relationship between the trend and the behaviour of traders. In section 5, we examine whether this behaviour is consistent with rationality. We conclude in section 6.

4.1.1 Related Literature

Both the gambler's fallacy and the hot hand fallacy are cognitive biases arising when individuals predict the future outcomes of a random series.

Agents who suffer from GF expect a positive (negative) shock after a negative (positive) shock. Clotfelter and Cook (1993) find that lottery players exhibit GF in a numbers game: numbers which are already drawn become very quickly less popular among the players.

HHF is similar to GF, but reverses the direction of prediction. Agents who suffer from HHF expect a positive (negative) shock after a positive (negative) shock. Gilovich et al. (1985),) have demonstrated the presence of HHF among basketball players and fans who believe that a player who is having a streak of successes also has a high probability of scoring the next time though the evidence does not support these beliefs. Camerer (1989) and Brown and Sauer (1993) also find evidence on HHF in the basketball betting market. Guryan and Kearney (2008) provide evidence for the "lucky store effect" which refers to the tendency of gamblers to believe that there is a higher probability of winning if they buy a ticket from a store which has recently sold a winning ticket.

Though GF and HHF have contradictory effects, researchers find that both can be observed in the behaviour of the same individual. Sundali and Croson (2006) and Croson and Sundali (2005) identify both effects in the behaviour of the same single individual by using data collected from a casino. Psychologists and economists offer several explanations for this phenomenon.

4.1.1.1 Psychological theories

The psychology literature provides two explanations for the relationship between GF and HHF. One explanation suggests that both fallacies are cognitive biases caused by lack of knowledge of probability theory. The other explanation distinguishes two different types of expectations with respect to human performance and natural events. In human performance, individuals normally exhibit positive recency (HHF). In natural events, individual normally exhibit negative recency (GF).

Tversky and Kahneman (1974) suggests that GF, one of the representativeness heuristic biases, is caused by poor understanding of probability theory. Individuals may expect the small sample to have the statistical characteristics of a large sample. Therefore, when individuals toss a fair coin they attach a probability of more than 50% of getting a tail after observing a streak of heads. On the other hand, when the streak of heads becomes very long or when the frequency of heads is consistently larger than that of tails, individuals may reject the belief that the series is random and start to believe that the outcome of any tossing is positively related to the one before. Then GF becomes HHF. Rao (2009) has designed an experiment to study the relation between GF and HHF. The experiment suggests that GF normally happens after short streaks and HHF happens after long streaks. The experiment also identifies the transition from GF to HHF. Guryan and Kearney (2008) argue that the "lucky store"effect provides two challenges for the representativeness explanation of the GF and HHF. In the case of the "lucky store", firstly, there is no transformation from GF to HHF and, secondly, it does not rely on the streak which is necessary for the representativeness explanation.

Ayton and Fischer (2004) give an alternative explanation. They argue that the representativeness explanation cannot show why HHF is observed in basketball games but the GF is not and why the GF is observed in lottery games but HHF is not. In their experiments individuals suffer from GF when predicting a series of natural events such as the outcomes of tossing a coin or of a roulette wheel, and suffer from HHF when predicting a series of human activities such as the outcomes of basketball players' shooting or of the first serve of a tennis player.

4.1.1.2 Economic explanations

Economists are interested in GF and HHF because the two biases may help to explain puzzles or anomalies in the financial market. All theoretical work is developed to analyse the behaviour of traders in the stock market and there is no related work for the foreign exchange market. The Rabin and Vayanos (2010) (hereafter RV) model follows the suggestion of the psychological literature and tries to link the two fallacies. The RV model assumes that agents believe that random shocks are negatively related (GF). GF becomes HHF after a long streak because agents update their beliefs about the state of the world. Barberis et al. (1998) (hereafter BSV) assume that agents believe that random movements of a stock price has two regimes: reversing (which can be interpreted as GF) and trending (which can be interpreted as HHF). Agents continuously update their subjective probabilities (relative strength) of the two regimes. GF becomes HHF after a long streak an implication also of the RV model. The transformation is because the relative strength of GF and HHF changes as the streak increases.

In the RV model, agents can observe a signal which is generated by the state (a mean-reversing process) and an i.i.d shock. The state is not known by the agents. Agents also believe that the shock is negatively autocorrelated. Therefore agents use the weighted average of past shocks to predict the next shock and update their belief about the state according to the signal. The setting of the RV model can be interpreted through the second psychological explanation of GF and HHF. Shocks are the natural events and the states are related to human performance. Therefore the prediction process represents the agents'

GF effect and the updating process represents the agents' HHF effect. The prediction of the agents is a combination of these two effects. A key implication of the model is that the reversing probability associated with agents' beliefs is not monotonic in the streak length, due to the difference in the convergence speeds between the GF effect and the HHF effect. GF gets stronger during the first several periods of the streak and then HHF becomes dominant as the streak increases.

BSV develop a similar model based on the conservatism and the representativeness heuristics. Conservatism means individuals update their mind slowly when facing new information. Individuals believe that the random series is not really 'random', but follows two regimes: trending and reversing. The prediction is a weighted average of the predictions of the two regimes. In contrast to the probability of reversing in the RV model, the weights of regimes in BSV are monotonic responses to the feedback from the market. The model suggests individuals under-react to short term shocks and over-react to long term trends. The mean-reversing regime can be considered as the GF effect and the trending regime can be explained as the HHF effect.

Both the RV and the BSV models are supported by experimental evidence when compared to either the random walk model or the Bayesian learning model (Bloomfield and Hales 2002 and Asparouhova et al. 2009). However, overall, the RV model does a better job than the BSV model.

4.2 The Data and Series Construction

4.2.1 The Data

In this section, we describe the dataset and the methodology we employed for generating the series which we used in our analysis.

The data include all the interbank tick-by-tick prices and order flow of nine currencies against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 (hereafter D2000-1). We use the data for the Deutsche Mark (DEM) and the Japanese Yen (YEN).

The data set format is shown in Table (4.1). 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).

Although the Reuters D2000-1 system does not publish transaction prices, this does not affect our analysis. Traders in the market are aware of prices at all times for several reasons. First, traders in the market frequently ask for quotes to obtain the latest prices and to seek arbitrage opportunities. The no-arbitrage condition, on which Evans and Lyons (2002) model relies, requires that quotes of different market makers are identical at a given time. Second, users of the Reuters D2000-1 system have indicative FXFX data (real time indicative prices) for reference. Although FXFX data are not transaction data, they do reflect the basic tendencies of prices (Goodhart et al. 1996, Daníelsson and Payne 2002). Third, users of the Reuters D2000-2 system have data from Reuters D2000-1. Therefore, traders have many indirect methods of obtaining real-time price information, and it is reasonable to assume that transaction prices are known by everyone in the market.

Table 4.1: Data Format(USD/DEM)

Month	Day	Hour	Min	Sec	B/S	Ask	Bid
4	30	18	45	40	1	1.5326	.
4	30	18	46	23	0	.	1.5326
4	30	18	47	56	1	1.5328	.
4	30	18	48	22	1	1.533	.
4	30	18	49	53	1	1.5332	.
4	30	18	51	0	0	.	1.5327
4	30	18	52	34	1	1.5327	.
4	30	18	53	8	1	1.533	.
4	30	18	53	35	0	.	1.5329
4	30	18	54	21	1	1.5329	.
4	30	18	54	27	0	.	1.5333
4	30	18	55	10	0	.	1.533

This table shows the format of the data on the Reuters D2000-1 system.
B/S represents trade directions. It is 1 if there is a buy order and is 0 if there is a sell order.

4.2.2 Estimation of the Spread

Let s be the logarithm of the nominal exchange rate of the US dollar against another currency (an increase representing an appreciation of the US dollar), and let OF be the order flow which is +1 when agent buys dollars and 0 when an agent sells dollars, and t denote time.

$$s_t = \begin{cases} A_t, & \text{Buy order } (OF_t = 1) \\ B_t, & \text{Sell order } (OF_t = 0) \end{cases} \tag{4.1}$$

where B is the logarithm of bid price and A is the logarithm of ask price and $A \geq B$. The unobserved bid-ask spread (SP) is the difference between ask and bid prices,

$$SP_t = A_t - B_t$$

The presence of the spread affects the calculations of both the streak length and the exchange rate return. We therefore need to estimate and eliminate this effect.

The bid-ask spread together with the properties of the dataset imply that we only have partial knowledge of the two series of exchange rates. Therefore both the streak length and the exchange rate return at period t (k_t and Δs_t), if obtained directly from the data, are not well defined.

We define the mid-price at period t (M_t) as a half of the sum of the bid and the ask prices. More formally,

$$M_t = \frac{1}{2} \cdot (A_t + B_t) = \begin{cases} A_t - \frac{SP_t}{2}, & \text{Ask price is known } (OF_t = 1) \\ B_t + \frac{SP_t}{2}, & \text{Bid price is known } (OF_t = 0) \end{cases} \tag{4.2}$$

Assume the bid-ask spreads are fixed. Then the change in the mid-price is given by

$$\Delta M_t = \frac{1}{2} \cdot [(A_t - A_{t-1}) + (B_t - B_{t-1})] = A_t - A_{t-1} = B_t - B_{t-1} \tag{4.3}$$

where Δ is the first-difference operator. The mid-price returns can be obtained if both the transaction price returns and the corresponding spread are known. The spread is estimated by using the modified version of Huang and Stoll (1997) model (MHS model, from chapter four).

$$\Delta s_t = \alpha + \frac{SP_t}{2}(BS_t - BS_{t-1}) + \frac{SP_t}{2} \cdot \beta_1 \cdot BS_{t-1} - \frac{SP_t}{2} \cdot \beta_2 \cdot BS_{t-2} + \varepsilon_t \tag{4.4}$$

Similar to OF_t, BS_t is an index of trade direction. The coefficient of $(BS_t - BS_{t-1})$ (i.e. $\frac{SP_t}{2}$) is used to estimate the spread. The last two terms control for the adverse selection and inventory control effects. As discussed in Chapter three, the spread varies systematically across the day. Our estimate of the movement in the mid-price, M, is based on the estimate of the spread in the previous section.

$$BS_t = \begin{cases} 1 \;\; if \; OF_t = 1 \\ -1 \;\; if \; OF_t = 0 \end{cases} \tag{4.5}$$

$$\Delta s_t^{adj} = \Delta s_t - SP_t \cdot (OF_t - OF_{t-1}) \tag{4.6}$$

Thereafter, the subscript adj indicates that the variable is obtained based on mid-price return Δs_t^{adj} rather than the transaction price return Δs_t .

4.3 Streak Length and Width

If agents believe that a given series is autocorrelated, then the properties of past observations should have predictive power. In our analysis we will focus on two properties of the streak defined as a sequence of movements of the estimated mid-price of the dollar in the same direction over successive transactions.

The streak length k_t which denotes the number of periods that the series continues to move in the same direction. Thus the streak length should be a non-negative integer.

We also need a measure of the strength of the streak, and a candidate would be the distance between the current exchange rate and the one at the beginning of the streak. However, clearly the streak length and the distance are positive correlated. In order to avoid this problem we use the average streak width w_t, the quotient of streak distance and the streak length, and thus eliminate the influence of the streak length.

Let Δs_t denote the difference between s_t and s_{t-1}. Then we can formally express the streak length as

$$k_t = \begin{cases} k_{t-1} + 1, & sgn\,[\Delta s_t] = sgn\,[\Delta s_{t-1}] \\ 0, & sgn\,[\Delta s_t] \neq sgn\,[\Delta s_{t-1}] \end{cases} \tag{4.7}$$

$$sgn\left[\Delta s_t\right] = \begin{cases} 1, & \Delta s_t \geq 0 \\ 0, & \Delta s_t < 0 \end{cases}$$

The formal expression for the width is given by

$$w_t = \frac{|s_t - s_{t-k_t-1}|}{(k_t + 1)} \cdot 100 \tag{4.8}$$

If agents suffer from GF and HHF, their behaviour should be influenced by the streak length and the streak width. Lots of evidence related to these fallacies has been gathered from tossing coins experiments where, by definition, the average streak width is constant. Therefore, the literature has focused on the effects of the streak length on agents' behaviour and has ignored the average streak width. That is also fine when the evidence is related to scoring in a basketball game or gambling in a numbers game, but is not appropriate for the foreign exchange market where the shocks vary in size.

Figure (4.1) provides an example of the streak length, streak distance and average streak width. Focusing on the upward trend section of the figure, the horizontal right-angle side of the triangle represents the streak length and the vertical right-angle side of the triangle represents the streak distance. The average streak width is equal to $tan\theta$, that is the slope of the hypotenuse of the triangle and provides a measure of the strength of the shock. The steeper the hypotenuse is, the stronger the shocks are. The length of the segment AB is the average streak width of the trend before period t. The return of period t is given by Δs_t.

Table (4.2) presents the summary statistics of the data. The streak length k is calculated from the series s using Equation (4.9) and then, using (4.6) the adjusted streak length k^{adj} is obtained. Formally, k^{adj} is given as follows.

$$k_t^{adj} = \begin{cases} k_{t-1}^{adj} + 1, & sgn\left[\Delta s_t^{adj}\right] = sgn\left[\Delta s_{t-1}^{adj}\right] \\ 0, & sgn\left[\Delta s_t^{adj}\right] \neq sgn\left[\Delta s_{t-1}^{adj}\right] \end{cases} \tag{4.9}$$

$$sgn\left[\Delta s_t^{adj}\right] = \begin{cases} 1, & \Delta s_t^{adj} \geq 0 \\ 0, & \Delta s_t^{adj} < 0 \end{cases}$$

The average value of SP is positive, which means that the existence of spreads magnifies exchange rate changes. The spread seems very small (the average SP is about one-tenth

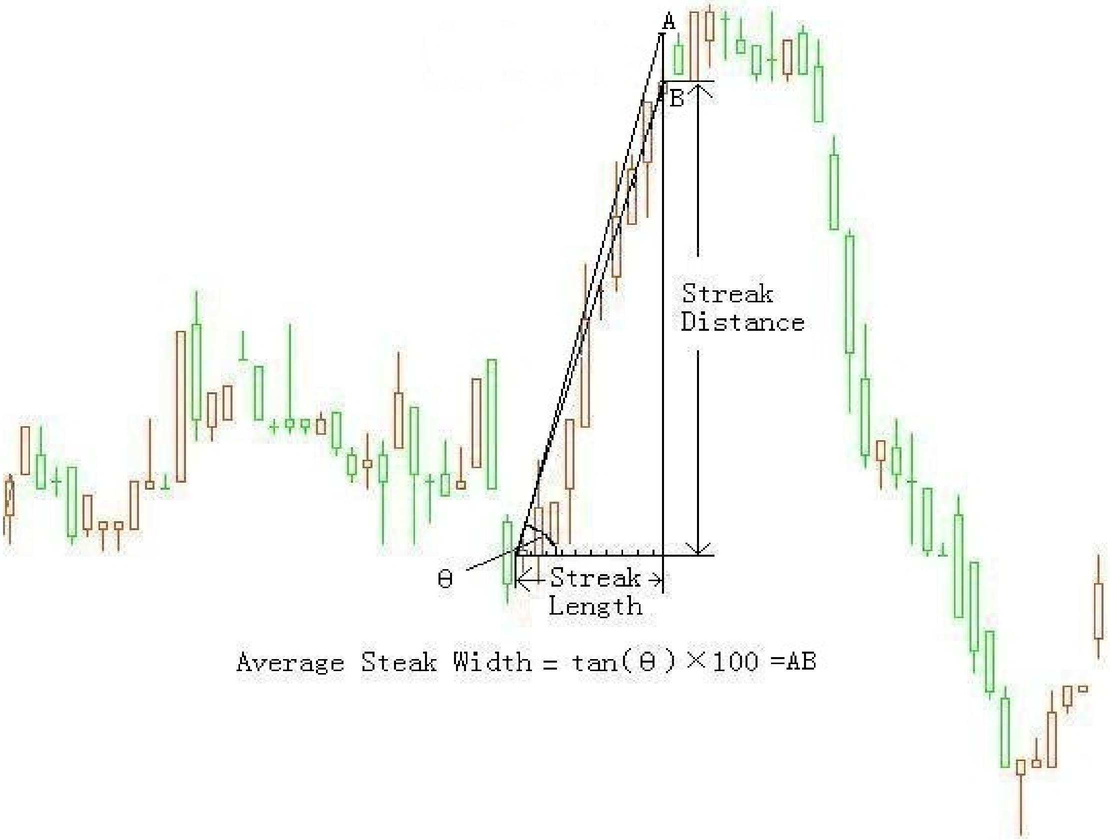

Figure 4.1: Steak Length, Streak Distance and Average Streak Width

This figure provides an example of the streak length, streak distance and average streak width. The exchange rate in the figure is indicative.

of the standard deviation Δs). The adjusted exchange rate series has either the same or a smaller average value than the original one, which is consistent with the analysis above. Figure (4.2) shows the histogram graph of the streak length. The average streak length is less than one, and the fraction of instances where $k^{adj} = 0$, is greater than 50% which suggests that the signs of exchange rate returns change very frequently. The average streak length becomes smaller after adjustment. Figure (4.3) shows the histogram graph of the average streak width. In most cases the average streak width lies in the interval $[-0.4, 0.4]$.

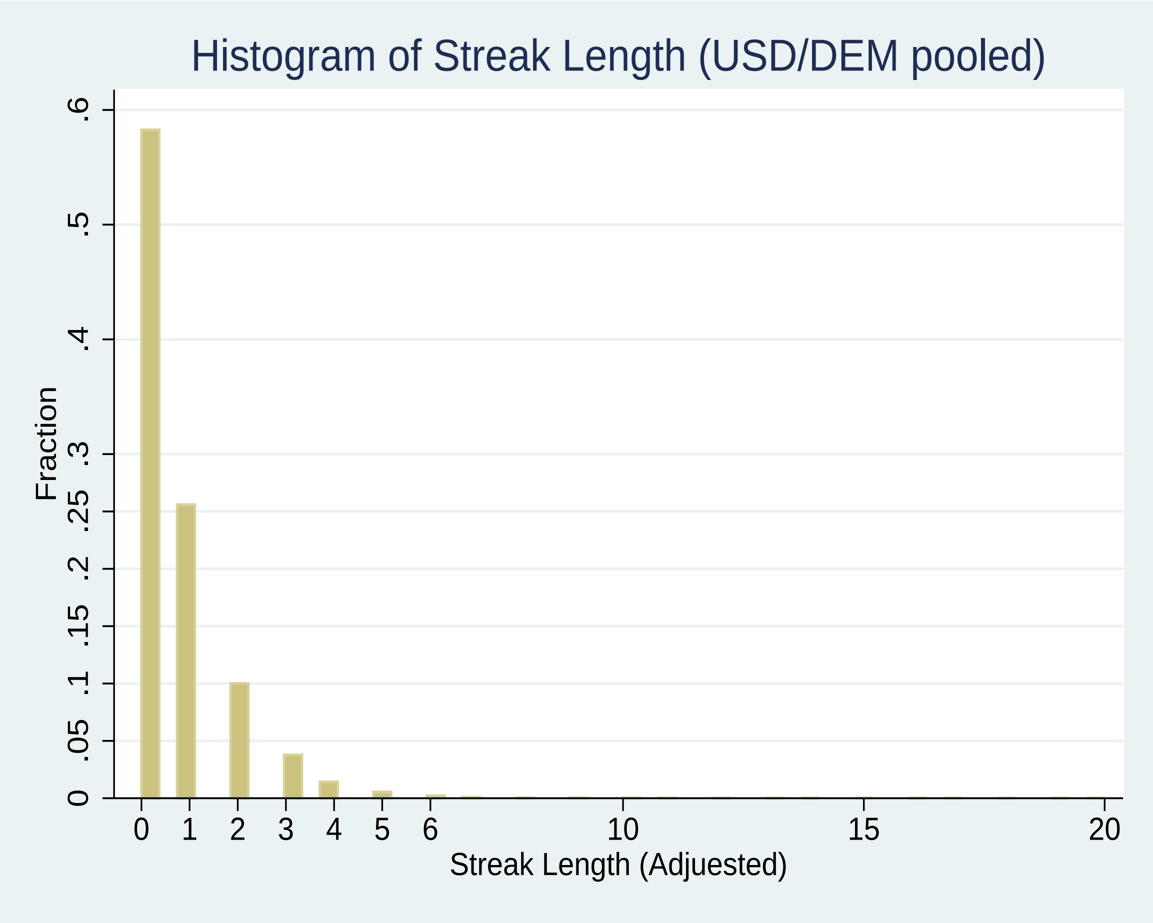

Figure 4.2: Histogram of Streak Length USD/DEM

This figure shows the histogram of the streak length. The USD/DEM transaction data on the Reuters D2000-1 system are used. The streak length is obtained from Equation (4.9)

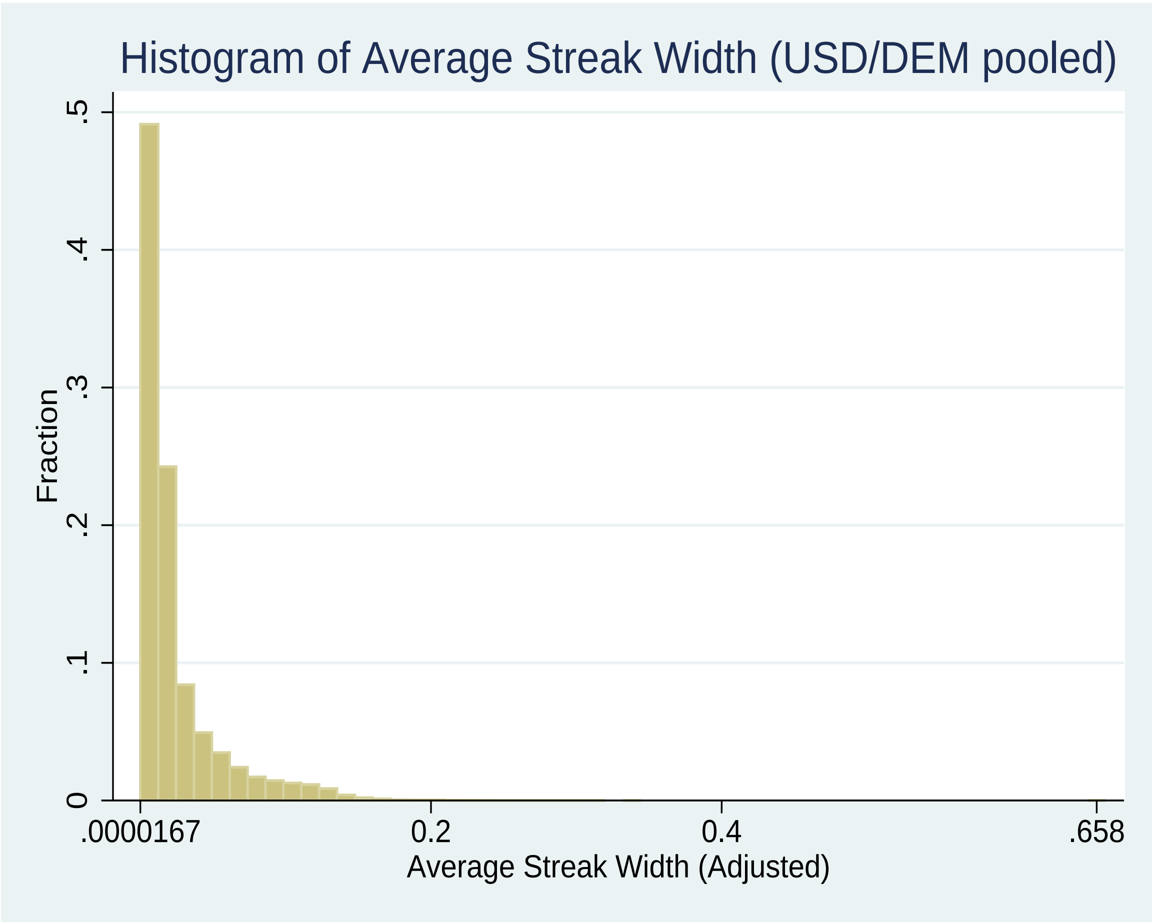

Figure 4.3: Histogram of Average Streak Width USD/DEM

This figure shows the histogram of the streak width. The USD/DEM transaction data on the Reuters D2000-1 system are used. The streak width is obtained from Equation (4.8)

Table 4.2: Data Description

	Obs	Mean	Std. Dev.	Min	Max
			USD/DEM		
Δs	256837	-1.18×10^{-7}	3.859×10^{-4}	-0.00656	0.00271
k	256837	0.759	1.151	0	21
Δs^{adj}	256838	-1.12×10^{-7}	3.85×10^{-4}	-0.00663	0.00271
k^{adj}	256838	0.700	1.0437	0	19
w^{adj}	256838	0.0235	0.0292	2.98×10^{-6}	0.663
			USD/YEN		
Δs^{adj}	151834	3.18×10^{-7}	6.08×10^{-4}	-0.00468	0.00683
k^{adj}	151834	0.717	1.0797	0	13
w^{adj}	151834	0.0382	0.0450	7.95×10^{-6}	0.648

This table shows the summary statistics of the transaction returns s, streak length k obtained from Equation (4.9), the average streak width obtained from Equation (4.8). The superscript adj represents the variables obtained from the series s^{adj} which is obtained from Equation (4.6).
In this chapter, we use USD/DEM and USD/YEN transaction data on the Reuters D2000-1 system.

4.4 Estimating Feedback Trading

4.4.1 The General Framework

The influence of the order flow on the quotes has been confirmed in the previous section. However, the effects of the quotes on the traders' decision (feedback trading), which is not mentioned in Evans and Lyons (2002), should not be ignored. Feedback trading may be caused by the gambler's fallacy and the hot hand fallacy (e.g. Rabin and Vayanos 2010). There is some evidence from the field that supports the RV model. A transition from GF to HHF as the streak becomes longer is evident in laboratory experiments (Rao 2009 and Asparouhova et al. 2009).

Our data provide us with an opportunity to study professional traders' reactions when they face an uncertain series, as in a laboratory experiment. Experiments designed to test the GF and HHF combine two basic elements: a series of uncertain numbers and a series of agents' reactions. Our data also have these two elements. The tick-by-tick data capture the causality: traders (agents) first observe quotes and then make decisions to buy or sell.

Suppose that order flow is a function of previous exchange rate movements. Then the general framework of the analysis can be written as

$$E(OF_t\,|\,x_t) = F(x_t, \beta) \tag{4.10}$$

where x is a vector that describes the past movements of the exchange rate and β is a vector of coefficients. Feedback trading can be confirmed if the coefficients in Equation (4.10) are significant. The dependent variable OF_t (order flow) is a binomial variable (values of OF

could be either 1 or 0). Given the value of vector x, the conditional expectation $E(OF_t\,|\,x_t)$ equals the conditional probability of traders making a buy order $Pr(OF_t = 1|\,x_t)$. Both the linear probability model and the probit method are used to estimate the model. Equation (4.10) can be re-written as

$$E(OF_t\,|\,x_t) = Pr(OF_t = 1|\,x_t) = F(x_t, \beta) \tag{4.11}$$

For probit model,

$$F(x_t, \beta) = \Phi(x_t' \beta) \tag{4.12}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

4.4.2 Comparison with the Independent and Identical Distribution Experiments

One may compare the trend continuing to the scenario of coin-tossing, which provides an independent and identical distribution experiment. We start with the simplest case of tossing a coin, and then establish the similarity between our analysis and the laboratory experiment step by step.

If we toss a fair coin several times, with each toss, whether the outcome is heads or tails is independent of the outcomes of other tosses. In other words, the event follows an independent and identical distribution, and the series of outcomes is random. Furthermore, the outcome is not influenced by the previous streak of heads or tails.

If we toss an unfair coin, for which the probability of the outcome being heads is, for example, 0.37, the outcome still follows an independent and identical distribution, and the series of outcomes is random. Furthermore, the outcome is not influenced by the previous streak of heads or tails, although a tail is more likely.

We may have two unfair coins, where the probability of tossing coin A and obtaining heads is, say, 0.37, and the probability of obtaining heads with coin B is, say, 0.63. Let us say that we start by tossing coin A, and If the outcome is heads, we use coin B the next time, whereas if the outcome is tails, we use coin A the next time. Under these conditions, the series of outcomes is no longer random but positively autocorrelated. We can predict the outcome on the next occasion based on the current outcome. However, most importantly, the outcome is not influenced by the previous streak of heads or tails. The predicated probability of heads next time should be 0.63 if the current outcome is heads, a pattern that is apparent after observing several rounds.

The scenarios above can be used as laboratory experiments to study whether individuals suffer from the gambler's fallacy or the hot hand fallacy. Because the streak length does not influence the outcome, if individuals' predictions are influenced by the streak length, one may conclude that they suffer from GF or HHF.

As we will show, the market data we use resemble to the data generated by an experiment based on the third scenario (if there were an experiment): the returns are positively autocorrelated. However, the probability of price reversal is not influenced by the streak length. If one is going to predict future returns, one should know the probability of a trend continuing. In this case, to maximise profits, the probability of following the trend should equal the probability of the trend continuing.

Therefore, although our evidence is not as perfect as laboratory evidence, it does have somewhat similar properties.

4.4.3 Relation Among the Exchange Rate, the Order Flow and the Streak Length

The analysis in this section is conducted in several steps. We start with simple regressions to estimate the effect of exchange rate returns on order flow. In the next step, instead of using exchange rate returns we use their direction by regressing order flow on linear polynomials of the streak length. Finally, we allow for non-linearities in the above relationship.

Our first step is to test for the presence of feedback trading in the market. If traders' behaviour is influenced by mid-price returns, one can conclude that there is feedback trading. The following regression can be used to test this hypothesis.

When $x_t' \beta = \beta_1 + \beta_2 \Delta s_t^{adj}$, Equation (4.11) expresses the basic relationship between order flow and exchange rate returns. Then the probit model is given by

$$OF_t = \Phi(\beta_1 + \beta_2 \Delta s_t^{adj}) + \varepsilon_t \tag{4.13}$$

If there is feedback trading, i.e. exchange rate returns do affect traders' decisions, the parameter β_2 in Equation (4.13) should be significant. The results are presented in Table

(4.3). β_2 is positive and significant for both currency pairs, which suggests the existence of positive feedback trading in the foreign exchange market.

Next, we seek evidence that would be consistent with GF and HHF. If the two fallacies cause feedback trading, then the streak length will affect traders' decisions. When $x_t' \beta = \beta_1 + \beta_2 \Delta s_t^{adj} + \beta_3 k_t^{adj} \Delta s_t^{adj}$, Equation (4.11) can be used to test the linear relationship between the streak length and order flow. The probit model is given by

$$OF_t = \Phi(\beta_1 + \beta_2 \Delta s_t^{adj} + \beta_3 k_t^{adj} \Delta s_t^{adj}) + \varepsilon_t \tag{4.14}$$

The sign of $x_t' \beta$ represents the propensity of a trader to make a buy (sell) order. When the sign is positive (negative), there is a greater probability that the trader buys (sells). There are four possible situations and the complexity of the analysis can be reduced through effective coefficient analysis: trade following the trend and trade against the trend.

A buy order, in the upward trend	A buy order, in the downward trend
A sell order, in the upward trend	A sell order, in the downward trend

The effective coefficient for Δs_t^{adj} is $\beta_2 + \beta_3 k_t^{adj}$, a polynomial of k_t^{adj}.A positive coefficient suggests that the probability that traders trade by following the trend is greater. The results of the regression are shown in Table (4.3). The coefficients are significant: $\beta_2 > 0$ and $\beta_3 < 0$. The longer the streak length, the higher the propensity of traders to trade against the trend.

Our next step is to use the direction of Δs_t^{adj} instead of Δs_t^{adj} itself. Because Δs_t^{adj} includes two types of information (direction and size), it is preferable to separate these two factors when we analyse the effects of GF and HHF. Let $sign_t$ be the sign of Δs_t^{adj} representing the direction of a shock. More precisely, $sign_t$ is given by

$$sign_t = \begin{cases} 1, & \Delta s_t^{adj} > 0 \\ sign_{t-1}, & \Delta s_t^{adj} = 0 \\ -1, & \Delta s_t^{adj} < 0 \end{cases}$$

Table 4.3: Relationship Between Order Flow and Streak Length with Δs_t^{adj}

OF_t is the dependent variable	Δs_t^{adj}	$k_t^{adj}\,\Delta s_t^{adj}$	constant	$Pseudo-R^2$	N
DEM	18.48**		0.000910	0.000	256838
	(2.88)		(0.37)		
DEM	184.4***	-503.7***	0.00142	0.009	256838
	(24.64)	(-35.55)	(0.57)		
YEN	12.47*		0.0266***	0.000	151834
	(2.36)		(8.27)		
YEN	114.5***	-300.2***	0.0277***	0.009	151834
	(18.75)	(-28.50)	(8.57)		

This table reports the relationship between order flow and the streak length with Δs_t^{adj}
We use the tick-by-tick data for the Deutsche Mark (DEM) and the Japanese Yen (YEN) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.
We use the probit model (Equation 4.10) for our analysis.
k^{adj} is the streak length which is defined by Equation (4.9)
Z-statistics is in the parenthesis. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

Then

$$x_t'\beta = \beta_1 + \beta_2\, sign_t^{adj} + \beta_3\, k_t^{adj}\, sign_t$$

The intuition behind Equation (4.15), used to examine the relationship of order flow, the streak length and the direction of exchange rate returns, is similar to that behind equation (4.14):

$$OF_t = \Phi(\beta_1 + \beta_2\, sign_t + \beta_3\, k_t^{adj}\, sign_t) + \varepsilon_t \tag{4.15}$$

Table (4.4) reports the results for Equation (4.15). The coefficient β_3 is significant and negative. Table (4.5) reports the average marginal effect of each coefficient. The first and second columns of the table suggest that when the streak length increases by 1, the probability of following the trend decreases by 13.7% (the USD/DEM pair) on average, and by 13.4% for the USD/YEN pair. Based on these results, Figures (4.4) and (4.5) show the predicted probabilities of trend-following behaviour of traders who trade on the USD/DEM currency pair in the case of a rise and fall in the exchange rate. The trend following behaviour is quite similar for upward and downward movements. The streak length is a discrete variable. Therefore, the points in the figure represent actual values, and the curve links the points smoothly. The figure suggests that when $k_t^{adj} = 0$, traders are more likely to follow the trend (with a probability above 50%), and when $k_t^{adj} \geq 1$, traders are more likely to trade against the trend (so that the probability of following the trend is less than 50%). We observe that the coefficients in Equation (4.15) are more significant than those in Equation (4.14), which suggests that Equation (4.15) is the better choice.

For the final step, we follow Rabin and Vayanos (2010) who suggest a non-linear relationship between order flow and the streak length (a transfer from GF to HHF). According to the RV model and experiments (Rao 2009 and Asparouhova et al. 2009), traders' attitudes toward the trend change with the streak length. More precisely, traders' probability of trading against the trend should be U-shaped, as the streak length increases. At the beginning of the trend their reaction is to trade against the trend. By introducing dummies for the streak length (Dk_i), this hypothesis can be tested.

$$x_t{}'\beta = OF_t = \beta_1 + \beta_2 sign_t + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t$$

where $Dk_i = 1$ if the streak length is j, and $Dk_i = 0$ otherwise. The subscripts of the dummies represent the streak length. Equation (4.16) shows the corresponding probit model.

$$OF_t = \phi\left(\beta_1 + \beta_2 sign_t + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t\right) + \varepsilon_t \tag{4.16}$$

The effective coefficient for $sign_t$ is $\left(\beta_2 + \sum_{i=1}^{5} \delta_i \, Dk_i\right)$, a polynomial of streak length dummies. β_2 is the intersection term. Compared with regression (4.15), regression (4.16) uses dummies for the streak length rather than the streak length itself, which allows the influence of the streak length on order flow to differ for different streak lengths. The coefficients for the dummies will then reflect the non-linear relationship between order flow and the streak length.

Table (4.4) and (4.5) report the results of the probit regression and the marginal effects of the coefficients. The results are similar to the previous regression: as the streak length increases, the probability of following the trend decreases. For example, compared with the case when the streak length is zero, the probability of following the trend is 24.7% lower when the streak length is one and 32.5% lower when the streak length is two. Furthermore, the decline in the probability of trend-following falls as the streak length increases: when the streak length increases from one to two, the probability of trend-following falls by 7.7%, and when the streak length increase from two to three, the probability of trend-following falls by 5.8%. Figure (4.6) reports the predicted probability of trend following against the streak length. We find that the slope of the trend decreases as the streak length increases. This phenomenon may be caused by the hot hand fallacy, which partially cancels the impact of the gambler's fallacy. The left part of a U-shaped curve is observed here.

We must note two factors.

First, because the data do not identify order flow coming from different traders, trend-following behaviour when the streak length is very short (less than or equal to one) can also be attributed to the sequential nature of order arrivals. Traders holding the same expectations with regard to exchange rate movements may place orders in the same direction but not at the same time, which causes two adjacent orders to be positively correlated. As a result of this effect, there are two circumstances when trend-following behaviour could be observed when the streak length is very short. 1) Suppose that in period t, the exchange rate falls and that the exchange rate was trending upward in period t-1. Thus, the streak length is zero in period t. According to the results of the regression, there could be a trend-following order (sell order) in period t. The order could be a (slightly late) reaction (gambler's fallacy) to the previous upward trend. However, the order is recorded as a trend-following order when the streak length is zero. 2) Suppose that the exchange rate rises when the streak length is zero in period t and that both traders, A and B, believe that the exchange rate will rise in the future. The buy order of trader A comes first. The buy order of trader B comes later than A's order when the exchange rate has started to rise. As a result, B's order is recorded as a trend-following order when the streak length is one.

Second, the significance of the streak length variables, which cannot be explained by the sequential nature of order arrivals, confirms the gambler's fallacy.

4.4.4 Relation Among the Exchange Rate, the Order Flow and the Average Streak Width

In this section, we analyse how the average streak width affects the exchange rate and order flow. When the sizes of shocks are not identical, one cannot distinguish effects due to the size of a shock from effects due to the persistence of the streak. This is why we now introduce the average streak width. Consider the following figures. In both of the two figures, the exchange rate rises continuously for eleven periods (the streak length k=11). In contrast to Figure (4.7), where the exchange rate rises from 1.6307 to 1.6319, in Figure (4.8), it rises from 1.6321 to 1.6355. Using the average streak width, we can account for such differences. The average streak width, which is captured by the slope of the curve, is larger in Figure (4.8) than in Figure (4.7).

The streak length and streak distance are positively related. Therefore, although no related literature examines how the streak width influences agents' behaviour, one can reasonably expect that the streak distance has effects similar to those of the streak length. In contrast to the streak length and streak distance, the average streak width is not an accumulated variable. A large shock may be strong enough to change the agent's prediction of the exchange rate in the future, even when the streak length is short. In other words, the average streak width is not influenced by the streak length but reflects the strength of shocks to the trend. Thus the average streak width is appropriate for analysing the effects of the strength of shocks.

The introduction of the average streak width, provides another explanation for the "lucky store effect" (noted by Guryan and Kearney 2008). They find that the "lucky store effect" occurs when a store sells a winning ticket in the absence of a streak. Therefore, it suggests that the representativeness bias may not explain this phenomenon. However, if the winning ticket is viewed as the streak width, an explanation can be found within the representativeness framework. Because the profits of a winning lottery ticket are very large (ranging from \$8,888 to \$51,200,200 in the dataset), the shock is sufficiently strong to induce HHF behaviour of the gambler. As in section (4.4.3), we study the relationship between exchange rates, order flow and the average streak width in two steps.

First, we consider the basic linear relationship.

$$x_t' \beta = \beta_1 + \beta_2 \, sign_t + \beta_3 \, w_t^{adj} \, sign_t$$

The probit model is given by

$$OF_t = \Phi(\beta_1 + \beta_2 \, sign_t + \beta_3 \, w_t^{adj} \, sign_t) + \varepsilon_t \tag{4.17}$$

The effective coefficient for $sign_t$ is $\beta_2 + \beta_3 \, w_t^{adj}$. The results are shown in Table (4.4). β_3 is positive and significant for both currency pairs. Figure (4.9) shows the predicted probability of following the trend. The probability rises nearly linearly as the average streak width rises. When the average streak width is small (large), traders tend to trade against (follow) the trend. When w_t^{adj} is larger than approximately 0.065, the probability is above 50%, and there is a higher probability that traders make orders that follow the trend. Table (4.5) presents the average marginal effect of each coefficient in the probit model. The average marginal effect implies that when the average streak width rises by 0.1, the probability of following the trend rises by 5.45%.

Second, as in the case of streak length, traders' responses to the strength of shocks may be non-linear. With the following specification, we can test for the presence of a non-linear relationship.

$$x_t' \beta = \beta_1 + \beta_2\, sign_t + \beta_3\, w_t^{adj}\, sign_t + \beta_4 \left(w_t^{adj}\right)^2 sign_t$$

The probit model is given by

$$OF_t = \Phi\left[\beta_1 + \beta_2\, sign_t + \beta_3\, w_t^{adj}\, sign_t + \beta_4 \left(w_t^{adj}\right)^2\right] sign_t + \varepsilon_t \tag{4.18}$$

Using Equation (4.18), we can examine the relationship between the order flow and streak width. The effective coefficient of $sign_t$ is $\beta_2 + \beta_3\, w_t^{adj} + \beta_4 \left(w_t^{adj}\right)^2$, a quadratic polynomial of w_t^{adj}. β_2 is the intercept term. Let ς be the effective coefficient of $sign_t$.

$$\varsigma = f(w_t) = \beta_2 + \beta_3\, w_t^{adj} + \beta_4 \left(w_t^{adj}\right)^2 \tag{4.19}$$

If traders suffer from the gambler's fallacy at the beginning and the hot hand fallacy later, β_3 should be significantly negative and β_4 should be significantly positive. The curve of the function ς should be U-shaped.

Table (4.4) reports the results of the model. β_3 is significant and positive, while β_4 is significant and negative. The even-lines in table (4.5) indicate the average marginal effects of each coefficient in the probit model. The results suggest that when the average streak width rises by 0.1, the linear term implies that the probability that traders follow the trend rises by 11.33%, and when the quadratic term rises by 0.1, this probability falls by 48.65%. $\beta_3 > 0$ and $\beta_4 < 0$ suggest that the function $\varsigma = f(w_t)$ yields a hill-shaped curve.

Figure (4.12) shows the shape of the predicted probability that traders of USD/DEM will follow the trend, implied by the probit model. When the probability is above 50%, traders tend to follow the trend, and vice versa when the probability is below 50%. The hill-shaped curve suggests that traders tend to trade against the trend when the trend is gentle

(w_t^{adj} < 0.05), follow the trend when the trend is sharper (w_t^{adj} < 0.185), and again trade against the trend when the trend shows greater persistence (w_t^{adj} > 0.185). One can also observe the interaction of the marginal GF effect and the HFF effect at various levels of the average streak width. There may be two levels of shock strength: weak ($0 \leq w_t^{adj} \leq 0.12$, in the case of USD/DEM) and strong ($w_t^{adj} \geq 0.12$). When a shock is weak, the slope of the curve is positive, and the marginal GF effect is weaker than the marginal HFF effect. When a shock is strong, the slope of the curve becomes negative, and the marginal GF effect is stronger than the marginal HFF effect.

4.4.5 Relation Among the Exchange Rate, the Order Flow, the Streak Length and the Average Streak Width

In previous sections, the streak length and the average streak width were studied separately. In this section, both these variables, which measure different aspects of the trend, are studied together.

While we have presented evidence that these two variables can influence traders' behaviour, the coefficients of the regressions could be biased due to neglect of possible correlations among the variables. Intuitively, traders could jointly consider all two aspects when making their decisions. Therefore, it is reasonable and necessary to examine the effects of both the variables in one regression.

Once again, using the probit model, we consider non-linear relationships between the order flow and the variables. Equation (4.20) captures the non-linear relationship.

$$OF_t = \Phi\left[\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \beta_4 \left(w_t^{adj}\right)^2 sign_t + \sum_{i=1}^{5} \delta_i\, Dk_i\, sign_t\right] + \varepsilon_t \tag{4.20}$$

The results and the marginal effects are shown in the last two columns of Tables (4.4) and (4.5). Although there are contradictions between the signs of the coefficients of the streak width term of the USD/DEM and the USD/YEN pairs, in the domain of the streak width, the slopes of the curves are negative. Traders' reactions to the streak length exhibit a U-shaped pattern and are the same as those obtained in the previous section.

Table 4.4: Relationship Between Order Flow and Streak Variables

	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN
$sign_t$	0.171*** (54.47)	0.205*** (50.53)	0.242*** (73.03)	0.272*** (62.94)	-0.0883*** (-27.75)	-0.0322*** (-7.63)	-0.106*** (-26.69)	-0.0276*** (-5.31)	0.260*** (53.96)	0.341*** (54.04)
$k_t^{adj}\, sign_t$	-0.343*** (-105.10)	-0.330*** (-81.73)								
$Dk_1\, sign_t$			-0.605*** (-101.93)	-0.567*** (-73.14)					-0.619*** (-103.01)	-0.59 *** (-75.18)
$Dk_2\, sign_t$			-0.800*** (-91.42)	-0.797*** (-70.50)					-0.876*** (-92.38)	-0.832 *** (-72.68)
$Dk_3\, sign_t$			-0.945*** (-67.05)	-0.938*** (-52.60)					-0.962*** (-67.92)	-0.978 *** (-54.50)
$Dk_4\, sign_t$			-1.054*** (-45.23)	-1.087*** (-37.23)					-1.071*** (-45.88)	-1.131 *** (-38.62)
$Dk_5\, sign_t$			-1.148*** (-37.15)	-1.168*** (-32.55)					-1.166*** (-37.67)	-1.216*** (-33.83)
$w_t^{adj}\, sign_t$					1.367*** (16.04)	0.313*** (4.38)	2.840*** (13.07)	0.0566 (0.31)	0.236 (0.106)	-1.408*** (-7.65)
$(w_t^{adj})^2\, sign_t$							-12.20*** (-7.33)	1.475 (-1.53)	-11.900*** (-7.01)	-0.496 (-0.51)
constant	0.00189 (0.75)	0.0264*** (8.04)	0.00218 (0.86)	0.0267*** (8.1)	0.00117 (0.47)	0.0266*** (8.28)	0.00113 (0.46)	0.0266*** (8.27)	0.00204 (0.8)	0.263*** (7.97)
$Pseudo - R^2$	0.05	0.049	0.059	0.057	0.002	0	0.002	0	0.06	0.0586
N	256838	151834	256838	151834	256838	151834	256838	151834	256838	151834

This table reports the relationship between order flow and streak variables: the streak length and streak width.
We use the tick-by-tick data for the Deutsche Mark (DEM) and the Japanese Yen (YEN) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.
We use the probit model (Equation 4.10) for our analysis.
$sign_t$ is the sign of the return (Δs_t^{adj}) which represents the direction of a shock, and is given by Equation (4.4.3).
Dk_i is the dummy of the streak length. $Dk_i = 1$ if the streak length is i, and $Dk_i = 0$ otherwise. The subscripts of the dummies represent the streak length.
w_t^{adj} is the average streak width which is defined by Equation (4.8)
Z-statistics is in the parenthesis. *** $p < 0.001$

4.5 Extensions

4.5.1 Time Spaces

One of the important properties of the ultra-high frequency data is that the time spaces between two observations are not identical. It is reasonable to believe that the time spaces can influence traders' behaviours. In this section the effects of time spaces are discussed.

Let T be the number of seconds between two orders. For example, the first trade in Table (4.1) occurred at 18:45:40, and the second trade occurred at 18:46:23. There are thus 43 seconds between the two trades, and accordingly, $T_2 = 43$. Table (4.6) presents the summary statistics for the time space.

The following regressions are considered.

$$OF_t = \Phi(\beta_1 + \beta_9\, T_t\, sign_t) + \varepsilon_t \tag{4.21}$$

Table 4.5: Marginal Effect Relationship Between Order Flow and Streak Variables

	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN
$sign_t$	0.0683***	0.0819***	0.0966***	0.108***	-0.0352***	-0.0129***	-0.0422***	-0.0110***	0.104***	0.136 ***
	(54.47)	(50.53)	(73.03)	(62.94)	(-27.75)	(-7.63)	(-26.69)	(-5.31)	(53.96)	(54.04)
$k_t^{adj}\, sign_t$	-0.137***	-0.132***								
	(-105.10)	(-81.73)								
$Dk_1\, sign_t$			-0.242***	-0.226***					-0.247***	-0.235***
			(-101.93)	(-73.14)					(-103.01)	(-75.18)
$Dk_2\, sign_t$			-0.319***	-0.318***					-0.325***	-0.332***
			(-91.42)	(-70.50)					(-92.38)	(-72.68)
$Dk_3\, sign_t$			-0.377***	-0.374***					-0.384***	-0.390***
			(-67.05)	(-52.60)					(-67.92)	(-54.50)
$Dk_4\, sign_t$			-0.420***	- 0.434***					-0.427***	-0.451***
			(-45.23)	(-37.23)					(-45.88)	(-38.62)
$Dk_5\, sign_t$			-0.458***	- 0.466***					-0.465***	-0.485 ***
			(-37.15)	(-32.55)					(-37.67)	(-33.83)
$w_t^{adj}\, sign_t$					0.545***	0.125***	1.133***	0.0226	0.0942	-0.561***
					(16.04)	(4.38)	(13.07)	(0.31)	(1.06)	(-7.65)
$(w_t^{adj})^2\, sign_t$							-4.865***	0.588	-4.747***	-0.198
							(-7.33)	(1.53)	(-7.01)	(-0.51)
$Probability$	0.500	0.511	0.500	0.511	0.500	0.511	0.500	0.511	0.500	0.511

This table reports the marginal effects of the regressions in Table (4.4).
$Probability$ shows the provability of a buy order when the streak length and width are zero. Other settings are the same as in Table (4.4).
Z-statistics is in the parenthesis. *** $p < 0.001$

Table 4.6: Data Description: Time Spaces

	Obs	Mean (seconds)	Std. Dev.	Min	Max
USD/DEM	256837	26.058	75.989	0	10560
USD/YEN	151833	43.997	103.384	0	10516

This table shows the summary statistics of the time spaces between trades.
We use the tick-by-tick data for the Deutsche Mark (DEM) and the Japanese Yen (YEN) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.

$$OF_t = \Phi\left[\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \beta_4 \left(w_t^{adj}\right)^2 sign_t + \sum_{i=1}^{5} \delta_i\, Dk_i\, sign_t + \beta_5 T_t sign_t\right] + \varepsilon_t \tag{4.22}$$

The term $T_t sign_t$ measures traders' propensity to trade against or follow the trend with respect to time spaces. If β_5 is positive, the longer the time space, the greater is the likelihood that a trader will follow the trend. If β_5 is negative, the probability that a trader will trade against the trend rises as the time space increases.

The first four columns in Tables (4.7) and (4.8) show the pooled analysis results together with the marginal effects. In the first and third rows of Table (4.7), the coefficient is positive and significant. In the second and fourth rows of Table (4.7), the coefficient is statistically significant in the USD/DEM case but not the USD/YEN case, which suggests that the longer the waiting time, the greater is the probability that a trader will sell US dollars during an upward trend and buy US dollars during a downward trend. However, the value of the coefficient is small which means that the time spaces do not have large effects on traders' decisions.

4.5.2 The Impact of Order Flow on Exchange Rate Returns

Figures 4.11 shows the basic relationship between exchange rate returns and order flow in tick-by-tick data. Arrow A in both figures represents feedback trading. Arrow B in both figures represents the effect of order flow on exchange rate returns. The other arrows in the figures represent lag effects, which are not the main focus of the analysis.

The trader first observes the exchange rate return and then makes a decision. Microstructure theory suggests that the order flow can influence the exchange rate return in the next period.

Equation (4.20) is used to examine whether there are gambler's fallacy or hot hand fallacy effects in the foreign exchange market.

As seen in figures (4.11), we have omitted the impact of order flow on exchange rate returns, which may bias in the coefficients. Therefore, it would be preferable to consider the following equation simultaneously with Equation (4.20):

$$\Delta s_t = \alpha_1 + \alpha_2 \, OF_{t-1} + \varepsilon_t \tag{4.23}$$

To facilitate the analysis, we substitute Equation (4.23) into Equation (4.20) and estimate the following equation rather than Equation (4.20).

$$\begin{aligned} OF_t = \Phi \Bigg[& \beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \beta_4 \left(w_t^{adj}\right)^2 sign_t \\ & + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t + \beta_5 OF_{t-1} sign_t \Bigg] + \varepsilon_t \end{aligned} \tag{4.24}$$

The results are shown in the last four columns of Tables (4.7) and (4.8). The coefficients for OF_{t-1} are significant and positive. This result suggests that traders tend to follow the direction of incoming orders and that a buy incoming order raises the probability that the next order will be a buy order by approximately 6%, which supports the hot potato trading hypothesis. The results for the other coefficients are similar to those of previous sections, which suggest that although the most recent order flow significantly influences traders' behaviour, it does not change traders' behaviour in relation to the trend.

Table 4.7: Relationship Between Order Flow and Streak Variables

	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN
$sign_t$			0.258*** (52.70)	0.341*** (52.83)			0.237*** (48.60)	0.324*** (50.69)
$Dk_1\ sign_t$			-0.619*** (-102.93)	-0.590*** (-75.04)			-0.635*** (-104.96)	-0.616*** (-77.53)
$Dk_2\ sign_t$			-0.816*** (-92.36)	-0.832*** (-72.64)			-0.795*** (-89.95)	-0.815*** (-71.45)
$Dk_3\ sign_t$			-0.962*** (-67.91)	-0.978*** (-54.47)			-0.927*** (-65.56)	-0.941*** (-52.66)
$Dk_4\ sign_t$			-1.071*** (-45.87)	-1.131*** (-38.59)			-1.026*** (-44.21)	-1.080*** (-37.08)
$Dk_5\ sign_t$			-1.166*** (-37.65)	-1.216*** (-33.82)			-1.109*** (-36.08)	-1.150*** (-32.17)
$(w_t^{adj})\ sign_t$			0.192 (0.85)	-1.408*** (-7.36)			0.448* (1.97)	-1.309*** (-6.76)
$(w_t^{adj})^2\ sign_t$			-11.76*** (-6.81)	-0.496 (-0.49)			-12.55*** (-7.25)	-0.702 (-0.68)
$T_t\ sign_t$	-0.000105*** (-3.37)	-0.0000424 (-1.44)	0.000114** (2.83)	-1.73×10^{-9} (-0.00)				
OF_{t-1}					0.177*** (35.77)	0.225*** (34.95)	0.179*** (34.01)	0.218*** (32.19)
constant	0.0009 (0.36)	0.0266*** (-8.27)	0.00205 (0.81)	0.0263*** (7.97)	-0.0877*** (-25.04)	-0.0883*** (-19.18)	-0.0874*** (-23.93)	-0.0849*** (-17.76)
$pseudo-R^2$	0	0	0.060	0.059	0.004	0.006	0.063	0.064
N	256838	151834	256838	151834	256822	151818	256822	151818

This table reports the results of the robust test of the relationship between order flow and streak variables controlling the time spaces and the past order flow.
We use the tick-by-tick data for the Deutsche Mark (DEM) and the Japanese Yen (YEN) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.
We use the probit model (Equation 4.10) for our analysis.
$sign_t$ is the sign of the return (Δs_t^{adj}) which represents the direction of a shock, and is given by Equation (4.4.3).
Dk_i is the dummy of the streak length. $Dk_i = 1$ if the streak length is i, and $Dk_i = 0$ otherwise. The subscripts of the dummies represent the streak length.
w_t^{adj} is the average streak width which is defined by Equation (4.8)
T is the time space between two trades.
OF_{t-1} is the past order flow. Z-statistics is in the parenthesis. *** $p < 0.001$

Table 4.8: Marginal Effect Relationship Between Order Flow and Streak Variables

	DEM	YEN	DEM	YEN	DEM	YEN	DEM	YEN
$sign_t$			0.103*** (52.70)	0.136*** (52.83)			0.0947*** (48.60)	0.129*** (50.69)
$Dk_1\ sign_t$			-0.247*** (-102.93)	-0.235*** (-75.04)			-0.253*** (-104.96)	-0.246*** (-77.53)
$Dk_2\ sign_t$			-0.326*** (-92.36)	-0.332*** (-72.64)			-0.317*** (-89.95)	-0.325*** (-71.45)
$Dk_3\ sign_t$			-0.384*** (-67.91)	-0.390*** (-54.47)			-0.370*** (-65.56)	-0.375*** (-52.66)
$Dk_4\ sign_t$			-0.427*** (-45.87)	-0.451*** (-38.59)			-0.409*** (-44.21)	-0.431*** (-37.08)
$Dk_5\ sign_t$			-0.465*** (-37.65)	-0.485*** (-33.82)			-0.442*** (-36.08)	-0.459*** (-32.17)
$(w_t^{adj})\ sign_t$			0.0767 (0.85)	-0.561*** (-7.36)			0.179* (1.97)	-0.522*** (-6.76)
$(w_t^{adj})^2\ sign_t$			-4.690*** (-6.81)	-0.198 (-0.49)			-5.005*** (-7.25)	-0.280 (-0.68)
$T_t\ sign_t$	-0.0000418*** (-3.37)	-0.0000169 (-1.44)	0.0000454** (2.83)	-6.91×10^{-10} (-0.00)				
OF_{t-1}					0.0706*** (35.87)	0.0897*** (35.1)	0.0712*** (34.10)	0.0869*** (32.31)
Probability	0.500	0.511	0.500	0.511	0.500	0.511	0.500	0.511
N	256838	151834	256838	151834	256822	151818	256822	151818

This table reports the marginal effects of the regressions in Table (4.7).
Probability shows the provability of a buy order when the streak length and width are zero. Other settings are the same as in Table (4.7).
Z-statistics is in the parenthesis. *** $p < 0.001$

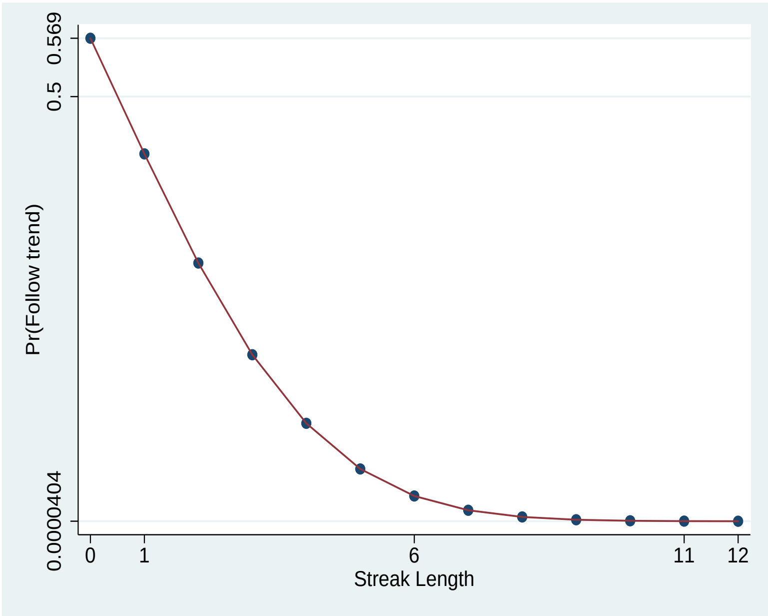

Figure 4.4: Predicted Probability in Upward Trends

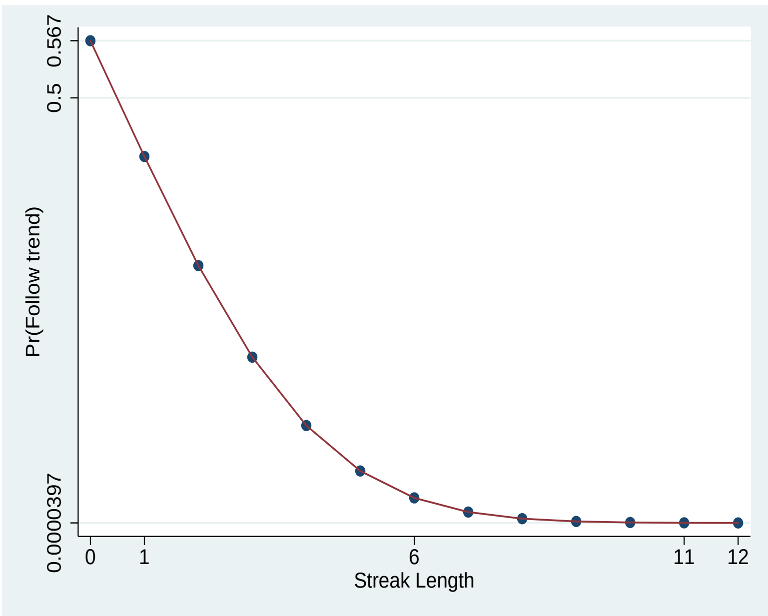

Figure 4.5: Predicted Probability in Downward Trends

This figure shows the linear relationship between the streak width and the predicted probability of following the trend. The probability is predicted by Regression (4.15). The USD/DEM transaction data on the Reuters D2000-1 system are used.

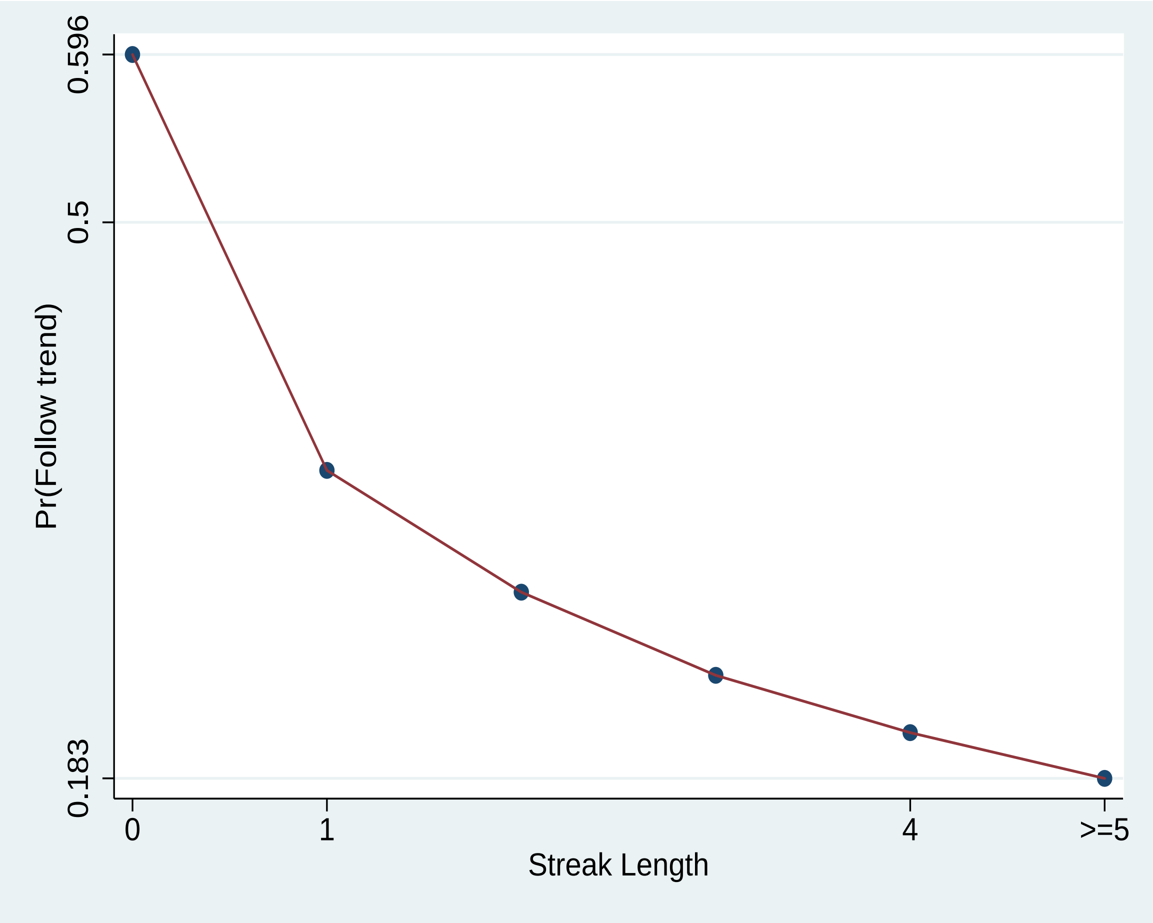

Figure 4.6: Predicted Probability

This figure shows the non-linear relationship between the streak width and the predicted probability of following the trend. The probability is predicted by Regression (4.16). The USD/DEM transaction data on the Reuters D2000-1 system are used.

Figure 4.7: Low Average Streak Width

This figure shows an example of long trend with a low average streak width. The exchange rate in the figure is indicative.
The average streak width is defined by Equation (4.8)

Figure 4.8: High Average Streak Width

This figure shows an example of long trend with a high average streak width. The exchange rate in the figure is indicative.
The average streak width is defined by Equation (4.8)

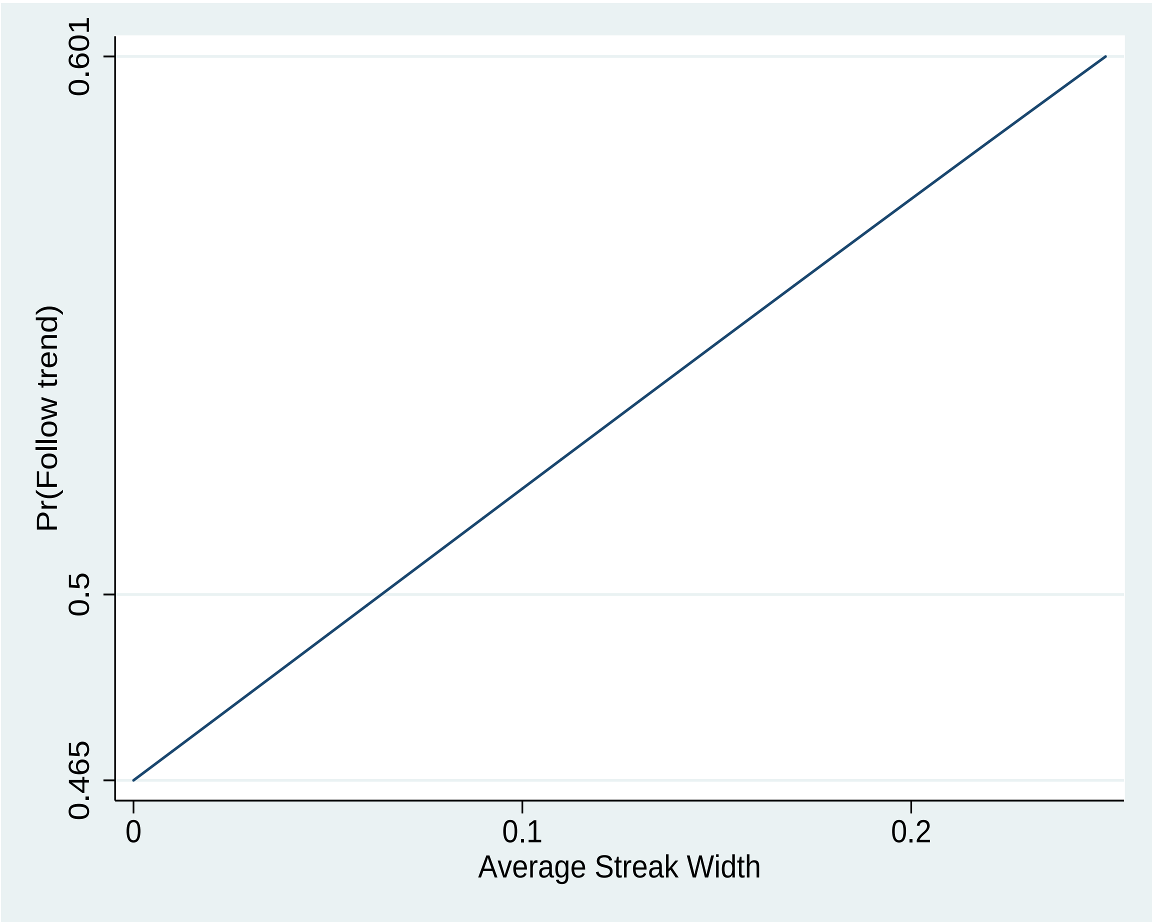

Figure 4.9: Predicted Probability

This figure shows the linear relationship between the streak width and the predicted probability of following the trend. The probability is predicted by Regression (4.17). The USD/DEM transaction data on the Reuters D2000-1 system are used.

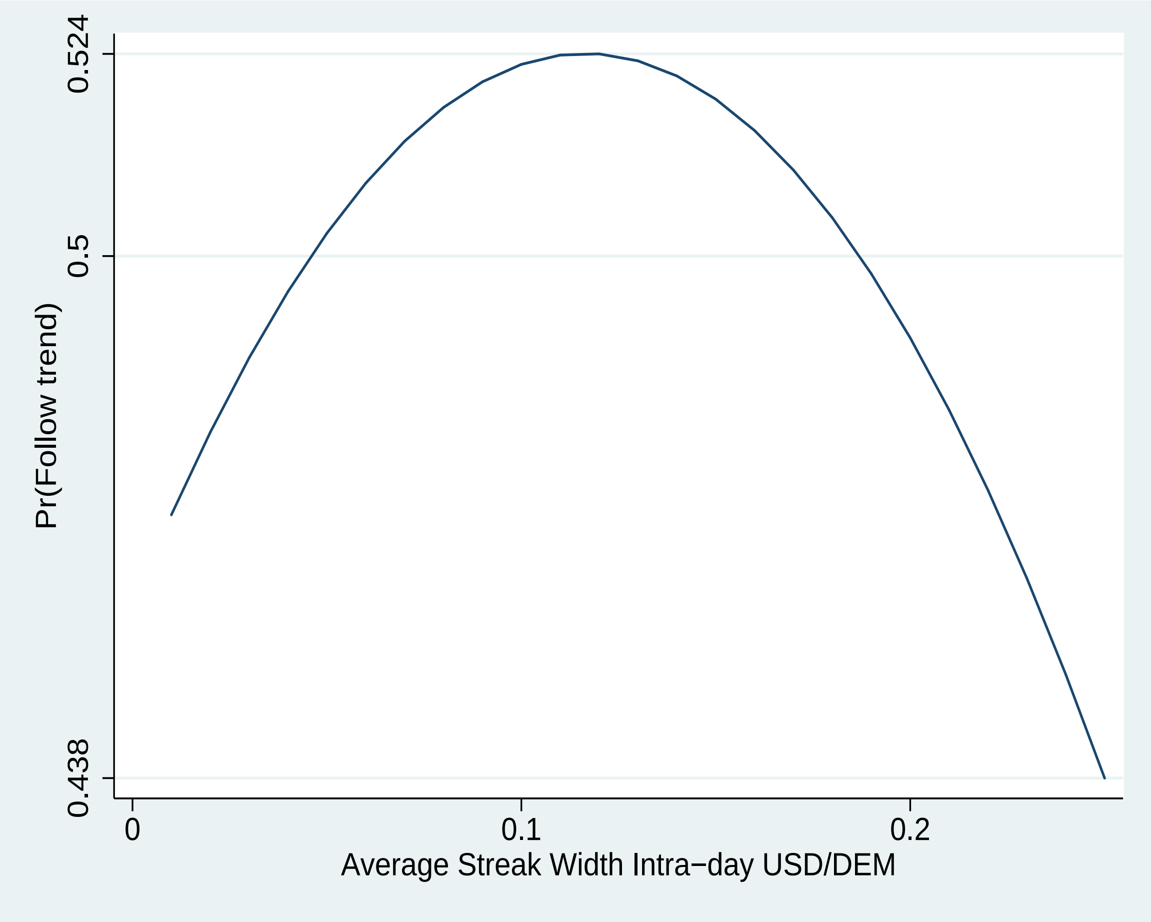

Figure 4.10: Predicted Probability

This figure shows the quadratic relationship between the streak width and the predicted probability of following the trend. The probability is predicted by Regression (4.18). The USD/DEM transaction data on the Reuters D2000-1 system are used.

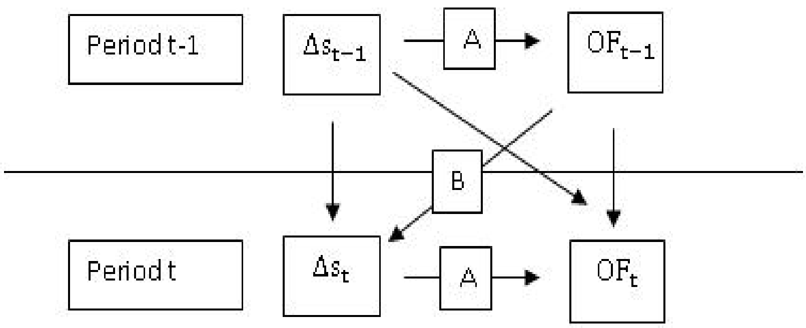

Figure 4.11: Structure of Tick-by-tick Data

This figure shows the basic relationship between exchange rate returns and order flow in tick-by-tick data. Arrow A in both figures represents feedback trading. Arrow B in both figures represents the effect of order flow on exchange rate returns. Other arrows in the figures are the lag effects which are not part of the main focus of the analysis.

4.6 Rational or Irrational

In the previous section, using the probit model, we found that traders tend to trade against the trend. A still more interesting question is whether this behaviour is rational or irrational.

One may divide the question into three sub-questions. A) Is the trend the reason (at least one of the reasons) for the behavioural changes? B) Does the trend carry information about future price movements? C) Do traders use information about the trend correctly (is their behaviour profitable)? In this section, we discuss these questions.

Question A is the most important one of the three. By using probit models and the RDD, we evaluate traders' behavioural changes caused by the trend, controlling for other factors such as trading strategies and private information, which are unknown to us.

The answers to questions B and C are supplementary to the discussion of question A. After obtaining an answer to question B, we may discuss how traders utilise the information. Are their reactions correct, or do they simply react to false information?

4.6.1 Discussion of Question B

In the previous section, we provided evidence related to what happens around the point where the mid-price return is zero; trend breaking and trend continuance do not include information about future price returns. In this section, we study the pooled case, which includes all observations rather than just the ones close to zero. In other words, we discuss whether trend breaking and trend continuance carry information about future price returns in general.

Table (4.9) provides summary statistics relevant to this question. The third column of Table (4.9) shows that the true probability that the streak continues in the next period is always less than 0.5 and is even lower when the price is part of a trend. It suggests that trend continuing/breaking does include information about future price movements. Therefore, trading against the trend is in general more rational than trading with the trend.

The third column also shows that the probability that the streak continues is very stable and approximately 0.37 when the price is part of a trend, which is roughly independent of the streak length. The second column shows the probability that traders will follow the trend indicated by the probit model in the previous section. The fourth column shows the t-values of a test of which the null hypothesis is that the probability of the trend continuing is the same as the probability of following the trend. The results suggest that traders' expectation on the probability of trend continuing is significantly different from the true one. Although this is in the correct direction, the probability falls as streak length increases. This suggests that traders overreact to streak length.

Table 4.9: Trend Following and Trend Continuing

Streak length	Probability of following the trend	True probability of streak continuing (TC)	T-values
0	0.596	0.465 (0.00131)	100.6
1	0.358	0.396 (0.00188)	20.2
2	0.289	0.374 (0.00295)	28.8
3	0.241	0.375 (0.00483)	27.7
4	0.209	0.362 (0.00783)	19.5
>=5	0.183	0.396 (0.0103)	20.7

This table reports the predicted probability of trend following behaviour and the true probability of streak continuing. We use the tick-by-tick data for the Deutsche Mark (DEM) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.
The first column is predicted by the following probit model $OF_t = \phi(\beta_1 + \beta_2 sign_t + \sum_{i=1}^{5} \delta_i Dk_i \, sign_t) + \varepsilon_t$
The second column is calculated from the true data. TC is a binomial variable which is 1 if a trend continues and is 0 if the trend breaks. Therefore, the mean of TC is the probability of the trend continuing. The standard deviations of TC is presented in brackets
T-values shows the t-values of the test of which the null hypothesis is that probability of the trend continuing is the same as the probability of following the trend.

4.6.2 Discussion of Question C

Traders' reactions to streak length could be either irrational or consistent with rationality. In this section, we examine whether trader reactions provide evidence of a psychological fallacy. In the previous section, we found that streak length conveys information about future price returns. In that case, use of streak length to predict future price returns may be rational.

We study the relationship between order flow and future price returns, using the following regressions. If the relationship is statistically significant and positive, it is possible that traders can predict future returns correctly.

$$\Delta s^{adj}_{t+1} = \beta_1 + \beta_2 OF_t + \varepsilon_t$$
$$sign(\Delta s^{adj}_{t+1}) = \phi(\beta_1 + \beta_2 OF_t) + \varepsilon_t \tag{4.25}$$

where $sign(\Delta s^{adj}_{t+1})$, the sign of the future price return, takes a value of 1 if the future return is positive and 0 otherwise. Table (4.9) presents the results of the regressions. In both regressions, β_2 is statistically significant and positive, which suggests a positive relationship between order flow and future returns. This could be because investors exploit patterns in the data to make profitable decisions. However, because order flow include information about future prices, a dealer could change the price based on the order flow he/she received. If dealers adjust prices in response to order flow (as in the Huang-Stoll model of spreads), this positive correlation would appear, even if investors were not making profitable decisions. Based on these regressions, we cannot distinguish the two explanations.

Table 4.10: The Impacts of the Order Flow on Future Returns

	Δs^{adj}_{t+1}	$sign(\Delta s^{adj}_{t+1})$
OF_t	0.0000340*** (44.86)	0.232*** (92.76)
constant	-0.000000133 (-0.18)	0.00249 (1.00)
N	256822	256822

This table shows the results of the following regressions,
$\Delta s^{adj}_{t+1} = \beta_1 + \beta_2 OF_t + \varepsilon_t$
$sign(\Delta s^{adj}_{t+1}) = \phi(\beta_1 + \beta_2 OF_t) + \varepsilon_t$
The second regression is estimated using the probit model. We use the tick-by-tick data for the Deutsche Mark (DEM) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.
Δs^{adj}_{t+1} is the return at period t+1
$sign(\Delta s^{adj}_{t+1})$ is the sign of the return at period t+1
OF_t is the order flow.

To test whether traders are exploiting past data, we examine whether future returns are negatively correlated with streak length in the same way as with order flow. Thus, we consider the following regressions:

$$\Delta s^{adj}_{t+1} = \beta_1 + \beta_2 \Delta s^{adj}_t + \sum_{i=1}^{5} \delta_i Dk_i \Delta s^{adj}_t + \varepsilon_t$$
$$sign(\Delta s^{adj}_{t+1}) = \Phi(\beta_1 + \beta_2 sign_t + \beta_3 w^{adj}_t sign_t + \sum_{i=1}^{5} \delta_i Dk_i\, sign_t) + \varepsilon_t \tag{4.26}$$

The basic structure of the regressions is similar to that of regressions (4.16) and (4.20). In this section, the future return and its sign are dependent variables. If traders predict future returns using the streak length and width, the signs of the coefficients should be the same as those of the corresponding coefficients in regressions (4.16) and (4.20). By comparing the coefficients in regression (4.26) and regressions (4.16) and (4.20), we can distinguish the effects of the streak length on future returns and order flow.

The results are shown in columns (1), (2), (4) and (5) of Table (4.11). All the coefficients for the streak length are negative and significantly different from zero, which suggests (a) that the streak length can be used to predict future returns and (b) that future returns are negatively correlated with the streak length. All the coefficients for the streak length dummies have the same signs as those in regressions (4.16) and (4.20). The coefficient declines as the streak length increases, except when the streak length exceeds four, which suggests that the relationship between future returns and the streak length is U-shaped. The pattern of the relationship between future returns and the streak length is similar to the pattern of the relationship between the order flow and streak length. The signs of the streak width terms in regressions (4.26) are the opposite of those in regression (4.20): the relationship between future returns and streak width is negative, and the relationship between order flow and streak width is hill-shaped. However, in the domain of streak width, the slopes of both curves are negative, i.e., the curves have similar shapes. Therefore, traders may be aware of the pattern of the relationship between future returns and the streak length and thus use the streak length to forecast future returns. However, the coefficients in regressions (4.26) are roughly one-third as large as those in regressions (4.16) and (4.20), which suggests that the power to forecast future returns is not as great as traders' believe.

Because order flow influence the movements of exchange rates (inventory control and asymmetric information costs), the pattern of the relationship between mid-prices and the streak length may reflect traders' behavioural responses to the streak length. Therefore, it is of interest to examine the relationship between the "underlying mid-price", which is not affected by order flow, and the streak length. We use the residual of the HS model as the underlying mid-price. The HS model suggests that the transaction price is a combination of the impact of order flow and shocks. The shocks are independent of order flow.

$$s_t = \frac{SP}{2}BS_t + (\alpha + \beta - 1)\frac{SP}{2}BS_{t-1} - \alpha\frac{SP}{2}(1 - 2\theta)BS_{t-2} + \varepsilon_t \tag{4.27}$$

The residual of the HS model ε_t, is the underlying mid-price which is not influenced by order flow.

$$s_t^{residual} = \varepsilon_t \tag{4.28}$$

To study the relationship between the underlying mid-price and the streak length, we use the following regressions which are similar to regressions (4.26), except that the dependent variables are the underlying mid-price and its sign.

$$\begin{aligned} \Delta s_{t+1}^{residual} &= \beta_1 + \beta_2 \Delta s_t + \sum_{i=1}^{5} \delta_i \, Dk_i \Delta s_t + \varepsilon_t \\ sign(\Delta s_{t+1}^{residual}) &= \Phi(\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t) + \varepsilon_t \end{aligned} \tag{4.29}$$

The results are shown in columns (3) and (6) of Table (4.11). Although the signs of the coefficients are the same as those in regressions (4.26), the coefficients are about half as large as those in regressions (4.26), and are one sixth as big as those in regressions (4.20). The finding suggests that the relationship between the underlying mid-price and the streak length is similar to that between order flow and the streak length. Furthermore, by controlling for the impact of order flow, we find that the relationships between the mid-price and the streak length and streak width are much weaker.

To sum up, the answer to Question C is that the streak length can, to some degree, predict future returns. However, the relationship is weak. Traders react to the direction of the relationship, but overestimate the strength of the relationship.

4.7 Conclusion

The literature suggests that people may suffer from the gamblers' fallacy and the hot hand fallacy when they predict future values of uncertain series. Although there is laboratory and field evidence (from casinos and lotteries) of such fallacies, field evidence from financial markets is rare. This chapter studies professional traders' behaviour in the foreign exchange market. We find that traders are unlikely to be rational and that the gamblers' fallacy and the hot hand fallacy can help explain the phenomenon of trading with/against the trend.

Unlike previous literature, which focuses on dual-outcome events such as heads and tails, win or lose, and score or miss, this chapter decomposes streaks into two parts: streak

Table 4.11: The Trend and Future Returns

	Δs^{adj}_{t+1}	Δs^{adj}_{t+1}	$\Delta s^{residual}_{t+1}$		$sign(\Delta s^{adj}_{t+1})$	$sign(\Delta s^{adj}_{t+1})$	$sign(\Delta s^{residual}_{t+1})$
Δs^{adj}_t		-0.348*** (-183.05)	-0.354*** (-186.56)	$sign_t$	-0.023*** (-13.12)	0.980*** (23.55)	-0.126*** (-30.08)
$Dk_1 \cdot \Delta s^{adj}_t$		-0.516*** (-106.48)	-0.492*** (-101.71)	$Dk_1 \cdot sign_t$		-0.239*** (-39.97)	-0.108*** (-18.01)
$Dk_2 \cdot \Delta s^{adj}_t$		-0.622*** (-82.60)	-0.594*** (-79.06)	$Dk_2 \cdot sign_t$		-0.322*** (-37.59)	-0.165*** (-19.12)
$Dk_3 \cdot \Delta s^{adj}_t$		-0.673*** (-54.26)	-0.643*** (-51.95)	$Dk_3 \cdot sign_t$		-0.335*** (-25.25)	-0.170*** (-12.72)
$Dk_4 \cdot \Delta s^{adj}_t$		-0.702*** (-33.81)	-0.667*** (-32.22)	$Dk_4 \cdot sign_t$		-0.380*** (-17.91)	-0.190*** (-8.88)
$Dk_5 \cdot \Delta s^{adj}_t$		-0.668*** (-25.43)	-0.634*** (-24.20)	$Dk_5 \cdot sign_t$		-0.300*** (-11.11)	-0.105*** (-3.87)
$w^{adj}_t \cdot sign_t$	-0.00427*** (-231.04)			$w^{adj}_t \cdot sign_t$	-5.720*** (-62.82)	-6.574*** (-72.20)	-5.616*** (-59.27)
constant	-3.29×10^{-7} (-0.48)	3.11×10^{-7} (0.48)	4.08×10^{-7} (0.63)	constant	0.00264 (1.06)	0.00281 (1.12)	0.00178 (0.7)
N	256822	256822	256822	N	256838	256838	256838

This table shows the results of the following regressions,

$$\Delta s^{adj}_{t+1} = \beta_1 + \beta_2 \Delta s^{adj}_t + \sum_{i=1}^{5} \delta_i Dk_i \Delta s^{adj}_t + \varepsilon_t$$

$$\Delta s^{residual}_{t+1} = \beta_1 + \beta_2 \Delta s^{adj}_t + \sum_{i=1}^{5} \delta_i Dk_i \Delta s^{adj}_t + \varepsilon_t$$

$$sign(\Delta s^{adj}_{t+1}) = \Phi(\beta_1 + \beta_2 sign_t + \beta_3 w^{adj}_t sign_t + \sum_{i=1}^{5} \delta_i Dk_i sign_t) + \varepsilon_t$$

$$sign(\Delta s^{residual}_{t+1}) = \Phi(\beta_1 + \beta_2 sign_t + \beta_3 w^{adj}_t sign_t + \sum_{i=1}^{5} \delta_i Dk_i sign_t) + \varepsilon_t$$

The last two regressions are estimated using the probit model. We use the tick-by-tick data for the Deutsche Mark (DEM) against the US dollar from May 1 to August 31 1996 on the Reuters Dealing 2000-1 system.

Δs^{adj}_{t+1} is the return at period t+1

$sign(\Delta s^{adj}_{t+1})$ is the sign of the return at period t+1 $sign_t$ is the sign of the return (Δs^{adj}_t) which represents the direction of a shock, and is given by Equation (4.4.3).

Dk_i is the dummy of the streak length. $Dk_i = 1$ if the streak length is i, and $Dk_i = 0$ otherwise. The subscripts of the dummies represent the streak length.

w^{adj}_t is the average streak width which is defined by Equation (4.8)

OF_t is the order flow.

length, which is the length of a trend based on the directions of price returns (up or down, as with other dual-outcome events); and streak width, which concerns the strength of a trend.

We find that traders' reactions to streak length are U-shaped: when a trend is short, traders tend to trade against the trend, and as a trend lengthens, traders' marginal tendency to trade against the trend decreases. Traders' reactions to streak width are captured by a curve with a negative slope, which suggests that traders follow the trend when streak width is small and trade against the trend when streak width is large.

Several factors, which might affect the conclusion, such as the time intervals between trades, past incoming orders and traders' rational expectation, are taken into account.

Technical analysis is widely used by traders. The literature suggests that the use of technical analysis may reflect a behavioural bias. In this chapter, we find that behavioural biases can be used to explain traders' behaviour. In future research, one may study the relationship between technical analysis and fallacies, shedding light on the utility of the former.

Appendix: The Regression Discontinuity Design

Introduction

This section provides a brief introduction of the regression discontinuity design (RDD).

This section is based on the literature review of Lee and Lemieux (2010). The RDD was first used by Thistlethwaite and Campbell (1960) to estimate the impact of a national scholarship programme on students' career choices. The basic idea of the paper is that borderline students, some of whom obtain the scholarship while others do not, share roughly the same abilities and other unobserved characteristics; therefore, one can identify the effect of the scholarship on different career choices between the two groups of students. The RDD has become very popular in recent years because, if some weak and testable assumptions are satisfied, it can be used in identification problems without potential problems related to omitted variables. The key assumption of the RDD is that assignment is random within a small neighbourhood around the cut-off.

The Basic Structure of the RDD

Assume that there are many individuals and that each is allocated a number (an assignment). Let X be the assignment variable. Individuals whose assignment variable is at or beyond a threshold c (i.e. $X \geq c$) receive the treatment (D). Let $Y_i(1)$ be individual i's outcome with the treatment ($D = 1$) and let $Y_i(0)$ be individual i's outcome without the treatment ($D = 0$). Then the treatment effect for i is $Y_i(1) - Y_i(0)$. Because no individual simultaneously both receives and does not receive the treatment, we usually focus on the average treatment effect. The average outcome for individuals who receive and do not receive the treatment can be written as $E[Y_i(1)]$ and $E[Y_i(0)]$, respectively. Then the average outcomes, conditional on the assignment X, are $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$. Although we can only observe $E[Y_i(1)|X]$ when $X \geq c$ and $E[Y_i(0)|X]$ when $X < c$, our results will be unaffected if we assume that both $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$ are continuous in the domain of X, as in Figure (4.12). Our focus of concern is the average treatment effect at the cut-off c, which is the difference between $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$ when $X = c$, given by:

$$\begin{aligned} &E[Y_i(1) - Y_i(0)|X = c] \\ &= B - A \\ &= \lim_{\varepsilon \downarrow 0} E[Y_i \,|\, X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i \,|\, X_i = c - \varepsilon] \end{aligned} \tag{4.30}$$

To this point, we have not imposed any assumptions on the allocation of the assignment. If X is allocated purely randomly among individuals, the curves of $E[Y_i(1)|X]$ and $E[Y_i(0)|X]$ become flat, as in Figure (4.13). When X is irrelevant to the outcomes, the process is actually a randomised experiment. In the RDD case, X is normally related to the outcome as in Figure (4.12). If the assignment is random around the cut-off (local randomisation), the RDD is as good as a randomised experiment. Furthermore, this assumption, although plausible, is testable. We will develop these issues in the following sections.

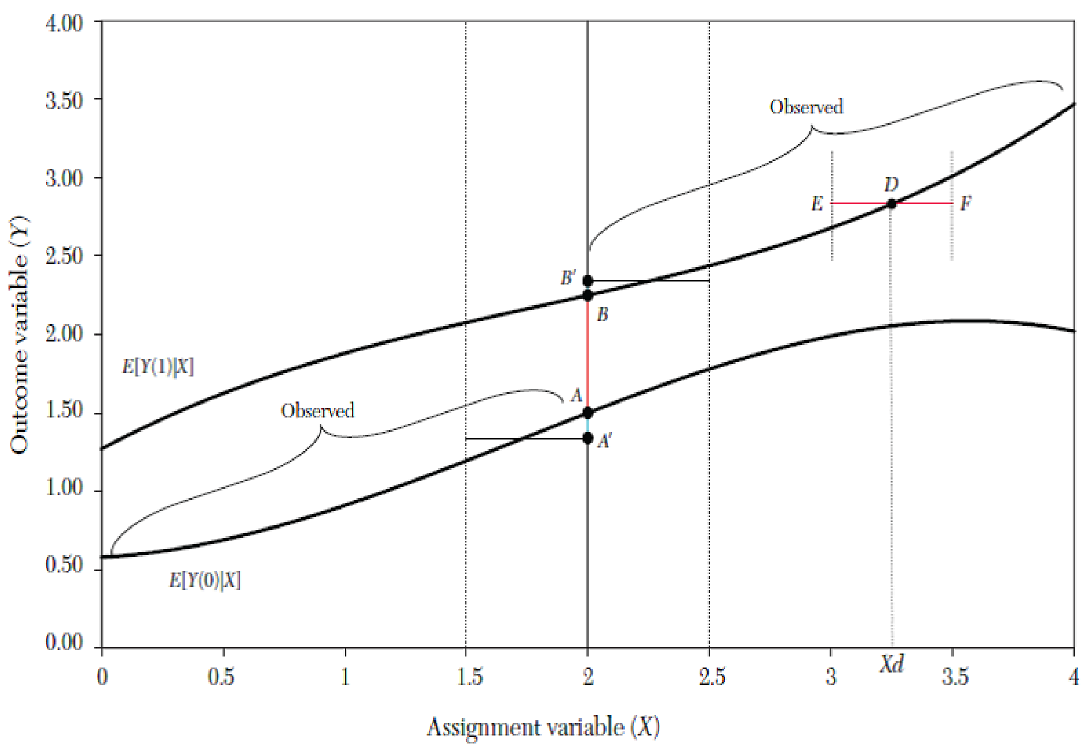

Figure 4.12: Nolinear RD

This figure is from Lee and Lemieux (2010)

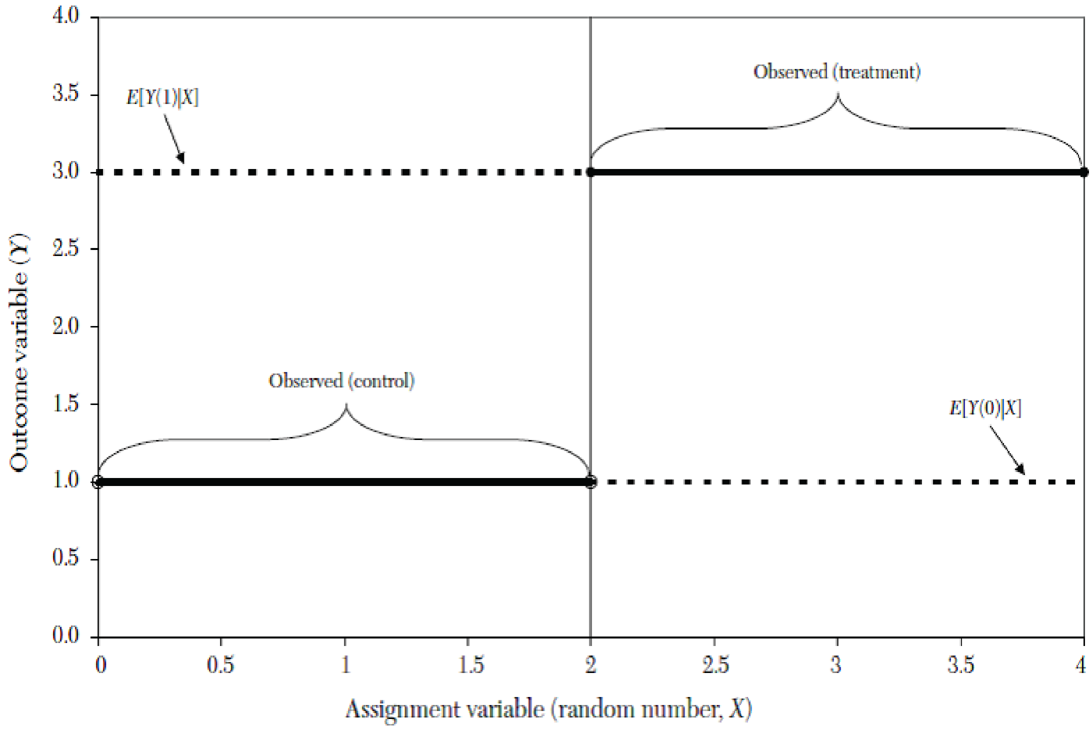

Figure 4.13: Randomised Experiment as a RD Design

This figure is from Lee and Lemieux (2010)

Comparing Randomised Experiments (RE) and the RDD

Following Lee (2008), we compare randomised experiments and the RDD in this section to clarify the analogy.

Let W be a random unobserved variable, which could be the underling type of individual, that influences outcomes Y. Let V be a variable that is determined prior to the assignment. In a random experiment, the probability of receiving the treatment is irrelevant to the type of individual one is. Formally,

$$Pr[D = 1|W = w] = p_0 \text{ for all } w \text{ in the domain of } W$$

Therefore the distribution of W is irrelevant to the treatment. Formally,

$$\text{RE A: } Pr[W \leq w|D = 1] = Pr[W \leq w|D = 0] = Pr[W \leq w]$$

The average treatment effect is also irrelevant to the assignment, as in Figure (4.13). Formally,

$$\text{RE B:} \quad \begin{aligned} &[Yi(W)|X \text{ and } D = 1] - E[Yi(W)|X \text{ and } D = 0] \\ &= E[Yi(W)|D = 1] - E[Yi(W)|D = 0] \end{aligned}$$

Furthermore, because V is pre-determined, the distribution of V is irrelevant to the treatment. Formally,

$$\text{RE C: } Pr[V \leq v|D = 1] = Pr[V \leq v|D = 0] \text{ for all } v$$

The equation above ensures that individuals in both the control and treatment groups share the same characteristics, apart from the treatment itself.

We now turn to the RDD. In addition to the outcomes Y, V and W influence the assignment variable X. Assume that the distribution of X is continuous on W and that the probability of receiving the treatment lies between 0 and 1 ($0 < Pr[X = c|W = w] < 1$). Then no one can be sure of receiving the treatment. Lee (2008) shows that if the above assumption is satisfied, one can obtain conclusions analogous to those of a random experiment. The distribution of W is continuous in x at the cut-off. Formally,

$$\text{RDD A: } Pr[W \leq w|X = x] \text{ is continuous in } x \text{ at } x = c \text{ for all } w$$

As in Figure (4.12), the treatment effect is given by:

$$\text{RDD B: } \lim_{\varepsilon \downarrow 0} E[Y_i \,|\, X_i = c + \varepsilon] - \lim_{\varepsilon \uparrow 0} E[Y_i \,|\, X_i = c - \varepsilon]$$

Furthermore, the distribution of V is continuous in x at the cut-off. Formally,

$$\text{RDD C: } Pr[V \leq v|X = x] \text{ is continuous in } x \text{ at } x = c \text{ for all } v$$

This assumption ensures that individuals can only influence the assignment not manipulate the treatment precisely. This assumption is not hard to approximate in the real world. For example, if a scholarship is awarded if a student's score is above or equal to 70%, a student could influence his/her score by studying harder. However, a student cannot obtain the exact score he/she plans to achieve because the judgement of the marker and other factors are beyond the control. Thus, borderline students cannot be sure whether they will be awarded the scholarship.

When the assumption is satisfied (or when the RDD is valid), "the treatment is as good as randomly assigned around the cut-off" (Lee and Lemieux 2010). Thus the RDD is "in important ways … more similar to a randomised experiment" (Lee and Lemieux 2010) than other evaluation strategies such as matching on observables or instrumental variables.

The Validity of the RDD

Another advantage of the RDD is that the assumptions of the RDD are testable.

As noted in Lee (2008), one can test "RDD C" to evaluation the imprecise manipulation assumption. "RDD C" can be easily tested by running an RDD of the pre-determined variables against the assignment variable.

McCrary (2008) introduces a method to test the imprecise manipulation assumption by examining the continuity of the density function of the assignment variable.

The Impact of the Trend: a Regression Discontinuity Design

A statistically significant relationship does not necessarily suggest a causal relationship. Therefore, one cannot necessarily conclude that the trend influences traders' behaviour based on the results of the probit models in previous sections. Furthermore, the foreign exchange market is highly complex, and thus it is conceivable that the analysis has failed to control for some important factors affecting the market. Thus, the probit models may suffer from bias due to the omission of unobserved or unavailable variables. In this section, we employ the regression discontinuity design (RDD) to address these issues. The regression discontinuity design (RDD) is a popular method that can be used to identify causality. If the testable assumptions of the RDD are satisfied, the RDD is as good as a randomised experiment (Lee and Lemieux 2010), and there is no need to control for other variables. By using the RDD, we can evaluate only the behavioural changes of traders caused by the trend, ignoring other factors such as trading strategies and private information, which are unknown to us.

We are interested in whether the trend influences traders' behaviour. If the price follows an upward trend, a positive mid-price return suggests that the trend is continuing (one of the streak length dummies in Equation 4.16 is equal to 1), and vice versa (all streak length dummies in Equation (4.16) are equal to 0). If the price follows a downward trend, a negative mid-price return suggests that the trend is continuing (one of the streak length dummies in Equation 4.16 is equal to 1) and vice versa (all streak length dummies in Equation 4.16 are equal to 0). Therefore, we use the current mid-price return as the assignment variable, and the threshold is set to zero. Let the order flow be the dependent variable. The treatment is whether the trend is continuing. The treatment in the RDD shares a similar intuition with the one we obtain from the streak length dummies in probit model (4.16). Therefore, the treatment effect measured by the RDD is similar to the coefficients of the streak length dummies in probit model (4.16).

Assume that the current price return contains some information about future returns (for example, autocorrelated mid-price returns). It is reasonable to believe that current price returns around zero (for example -0.000001 and 0.000001) all include very similar information about future returns. In an upward trend, the only difference between -0.000001 and 0.000001 is that the former breaks the trend and the latter maintains it. Therefore, the difference between the behaviours reacting to -0.000001 and 0.000001 is caused by the trend breaking/continuing.

A typical RDD analysis includes several procedures: first, draw a figure that represents the outcome and assignment variables, in this case, the order flow and the mid-price returns, respectively; second, run regressions of the outcome and assignment variables, using linear and higher order polynomial specifications; third, check the validity of the RDD: whether individuals can completely manipulate the outcome and whether the covariates are continuous at the threshold.

The Regression Discontinuity Design

In this section, we use the sharp RDD to evaluate the impact of the trend on order flow, following the procedures listed above. We use USD/DEM data. We do not include observations with corresponding returns equal to zero. The main reason for this is that when a return is equal to zero, the order flow is unchanged with certainty; in other words, the order flow is fully explained by the one that precedes it; therefore, we need not include it in the regressions. Furthermore, a trader may split a large order into smaller orders to avoid a large price impact. This behaviour will result in side orders of similar size with zero price returns.

We divide the sample into two groups: a) where the last order is a buy order; b) where the last order is a sell order. The two cases are symmetric. Therefore, for brevity, we report the results of case (b) only.

Figure (4.14) shows the treatment effect graphically when the last order is a sell order. The vertical axis represents the probability of a buy order. The horizontal axis represents the current mid-price return. The vertical line at the point where the return is zero splits the graph into two parts. If the last return was positive/negative, and the current return falls on the right part of the graph, the trend is continuing/breaking; if the current return falls on the left part of the graph, the trend breaks/continues. According to the figure, when the current return approaches zero from either the left or right, the probability of a buy order increases sharply. It is of greater interest that there is a significant decline when the current return crosses the threshold from the left side to the right side. Furthermore, if the last return was positive, the fall suggests that the fact that the trend is continuing reduces the traders' tendency to place buy orders. If the last return is negative, then crossing the threshold from the right to the left indicates that the trend is continuing and that there is a significantly higher probability of buy orders. Regardless of the direction of the last return, the figure shows that a continuing trend reduces traders' probability of trend following.

To estimate the impact of a trend on order flow, we run the following regressions. The coefficient τ captures the impact of the trend on traders' behaviour,

$$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \varepsilon_t \tag{4.31}$$

$$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \beta_3 (\Delta s_t)^2 + \beta_4 (\Delta s_t)^2 D + \varepsilon_t \tag{4.32}$$

$$\begin{aligned} OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \beta_3 (\Delta s_t)^2 + \beta_4 (\Delta s_t)^2 D \\ + \beta_5 (\Delta s_t)^3 + \beta_6 (\Delta s_t)^3 D + \varepsilon_t \end{aligned} \tag{4.33}$$

$$D = \begin{cases} 1 \; if \;\; \Delta s_t > 0 \\ 0 \; if \;\; \Delta s_t < 0 \end{cases}$$

$$-h \leq \Delta s_t \leq h$$

where D is the treatment dummy. D equal to one suggests that the trend is continuing in an upward direction and breaking in a downward direction, and is similar to the coefficients for the streak length dummies in the probit model (4.16). Δs_t is the mid-price return in period t, and h is the bandwidth of the regression. The RDD requires sub-samples close to the threshold rather than the whole sample. Technically, we run regressions using the samples that are within a bandwidth of the threshold. The choice of bandwidth is very important and affects the trade-off between error and noise. The coefficients of the regressions would be quite noisy if h were too small, and the error would be large if h were too large. The high-order polynomial models can also be influenced by observations far away from the threshold. Therefore, sub-samples are also preferred for high order polynomial models.

Imbens and Kalyanaraman (2012) suggest a data-driven method to calculate the optimal bandwidth.

$$\tilde{h}_{opt} = C_K \cdot \left(\frac{\hat{\sigma}_{-}^{2}(c) + \hat{\sigma}_{+}^{2}(c)}{\hat{f}(c) \cdot (\hat{m}_{+}^{(2)}(c) - \hat{m}_{-}^{(2)}(c))^2} \right)^{1/5} \cdot N^{1/5} \tag{4.34}$$

Following the suggestion of Lee and Lemieux (2010), the Akaike information criterion (AIC) of model selection is used to determine the order of the polynomial. The AIC is given by,

$$AIC = N \ln(MSE) + 2p \tag{4.35}$$

where N is the number of observations and MSE is the mean squared error of the regression. p is the number of coefficients.

Table (4.12) shows the results of the regressions when the last order is a sell order. The columns represent various bandwidths. The optimal bandwidth obtained from Equation (4.34) is 0.000330, and the optimal order of the polynomial, according to the AIC, is 2 for small bandwidths and 3 for large bandwidths. Generally, the parametric regressions of the RDD confirm the graphic results. One finds that there is a significant fall when the return is just past the cut-off point — zero. The jump is significantly different from zero. If the price is part of an upward trend, a negative current return that is just less than zero suggests that the trend breaks; and a positive current return that is just greater than zero suggests that the trend continues. The probability of a buy order falls by 6% if the trend continues, which is similar to the finding in the previous section: traders have a propensity to trade against the trend. If the price is part of a downward trend, a negative current return that is just less than zero suggests that the trend will continue; and a positive current return that is just greater than zero suggests that the trend will break. The probability of a buy order falls by 6% if the trend breaks, which is similar to the finding in the previous section: traders have a propensity to trade against the trend, and the impact is approximately 7%. The difference between the RDD and the probit model may be due to unobserved variables, which introduce biases into the probit model.

Figure (4.15) presents the pattern of the coefficient τ against the bandwidth. For small values of the bandwidth, the coefficient is unstable but becomes stable for large values of the bandwidth.

Table 4.12: The Treatment Effect

Bandwidth	0.000330	0.000165	0.000660	0.001	0.002
Linear	-0.0715*** (0.0061)	-0.0502*** (0.0072)	-0.0495*** (0.0051)	-0.0210*** (0.0044)	-0.00863* (.0041)
Quadratic	-0.0686*** (0.0099)	-0.0369** (0.014)	-0.0711*** (0.0072)	-0.0511*** (0.0061)	-0.0391*** (0.0052)
Cubic	N/A	N/A	-0.0902*** (0.098)	-0.0819*** (0.0081)	-0.0593*** (0.0064)
AIC	Quadratic	Quadratic	cubic	Cubic	Cubic

The table reports the results of the following regressions under various bandwidths.
The optimal bandwidth, which is obtained by Equation (4.34), is 0.000330
Linear represents the coefficients of τ of the linear model:
$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \varepsilon_t$
Quadratic represents the coefficients of τ of the quadratic model:
$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \beta_3 (\Delta s_t)^2 + \beta_4 (\Delta s_t)^2 D + \varepsilon_t$
Cubic represents the coefficients of τ of the cubic model:
$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \beta_3 (\Delta s_t)^2 + \beta_4 (\Delta s_t)^2 D + \beta_5 (\Delta s_t)^3 D + \beta_6 (\Delta s_t)^3 D + \varepsilon_t$
AIC reports the preferred model according to Akaike information criterion
Standard deviations are in brackets *** Significant at 0.1% level

Validity Tests

The RDD assumes the allocation to be random. In other words, agents cannot perfectly control the result of the the assignment variable, and thus the latter is locally random around the threshold, implying that the estimate is unbiased. If some agents can control the assignment variable, then they may do so beyond or below the threshold to obtain the benefit of the treatment, which would cause a jump in the density of the assignment variable at the threshold. Therefore, one can test this assumption by examining the distribution of the assignment variable. If there are no jumps in the density of the assignment variable, one can conclude that agents do not have perfect control. McCrary (2008) provides a test of this assumption.

Figure (4.16) shows that the density of the current mid-price returns is continuous. Furthermore, the hypothesis that there is a jump in the density is not statistically significant. This suggests that traders cannot perfectly manipulate the assignment variable. Thus, the assumption that control is imperfect is confirmed.

Another assumption of the RDD is that the covariates other than the assignment variable are continuous at the threshold. If other covariates jump at the threshold, one cannot identify the impact of the assignment variable, and the RDD procedure is undermined.

Here we examine two covariates: the last mid-price return and the future mid-price return. In high frequency data, the past price return could impact the order flow. The future price return is the most important variable for a trading strategy. Regardless of the strategies traders employ, the future price return is the variable they wish to know. Therefore, assuming that the future price return is partially known by some traders implies that the future price return influences the behaviour of traders.

Figure (4.17) shows the continuity test for one of the covariates: the last mid-price return. The optimal bandwidth is 0.000781 and the optimal order of the polynomial is 3. The results of the regression suggest that there is no significant discontinuity around the cut-off point. Therefore, the assumption of the RDD is satisfied.

Figure (4.18) shows the continuity test for the other covariate: the future mid-price return. The optimal bandwidth is 0.000542, and the optimal order of the polynomial is 3. The results of the regression suggest that there is no significant discontinuity around the cut-off point. Therefore, once again, the assumption of the RDD is satisfied. Figure 5 also suggests that there is no information about future price movements just around the current price movement.

Table (4.13) reports the results of the RDD, including the covariates mentioned above. The treatment effect is roughly the same as that without the covariates, which suggests that our results are robust.

Table 4.13: The Treatment Effect with Covariates

Bandwidth	0.000330	0.000165	0.000660
Linear	-0.0675*** (0.0060)	-0.0485*** (0.0072)	-0.0451*** (0.0050)

This table reports the results of the following regression under various bandwidths.
The optimal bandwidth, which is obtained by Equation (4.34), is 0.000330
Linear represents the coefficients of τ of the linear model:
$OF_t = \beta_0 + \tau D + \beta_1 \Delta s_t + \beta_2 \Delta s_t D + \beta_3 \Delta s_{t-1} + \beta_4 \Delta s_{t+1} \varepsilon_t$
Standard deviations are in brackets *** Significant at 0.1% level

To sum up, we have tested the validity of the RDD. The results of the RDD suggest that a continuing trend will reduce the probability that traders will follow the trend by approximately 6%.

The RDD assumes that all the covariates are continuous around the cut-off point. We tested the continuity of two covariates: past and future returns. There are also unobserved variables, such as trading strategies and private information about future returns. Jumps in unobserved variables may undermine the power of the RDD. If private information about future returns jumps near the cut-off point, then future returns should also jump near the cut-off point. If future returns are continuous around the cut-off point, the correct trading strategies should be continuous around the cut-off point; otherwise, there will be a loss. Given that the future return is continuous around the cut-off point, unobserved variables related to future returns, such as trading strategies and private information, are likely to be continuous.

Because the change of trend status around the cut-off point (where the return is just greater or less than zero) does not include information about future returns, under the assumption that unobserved variables are continuous, we can conclude that traders' behaviour is unlikely to be rational.

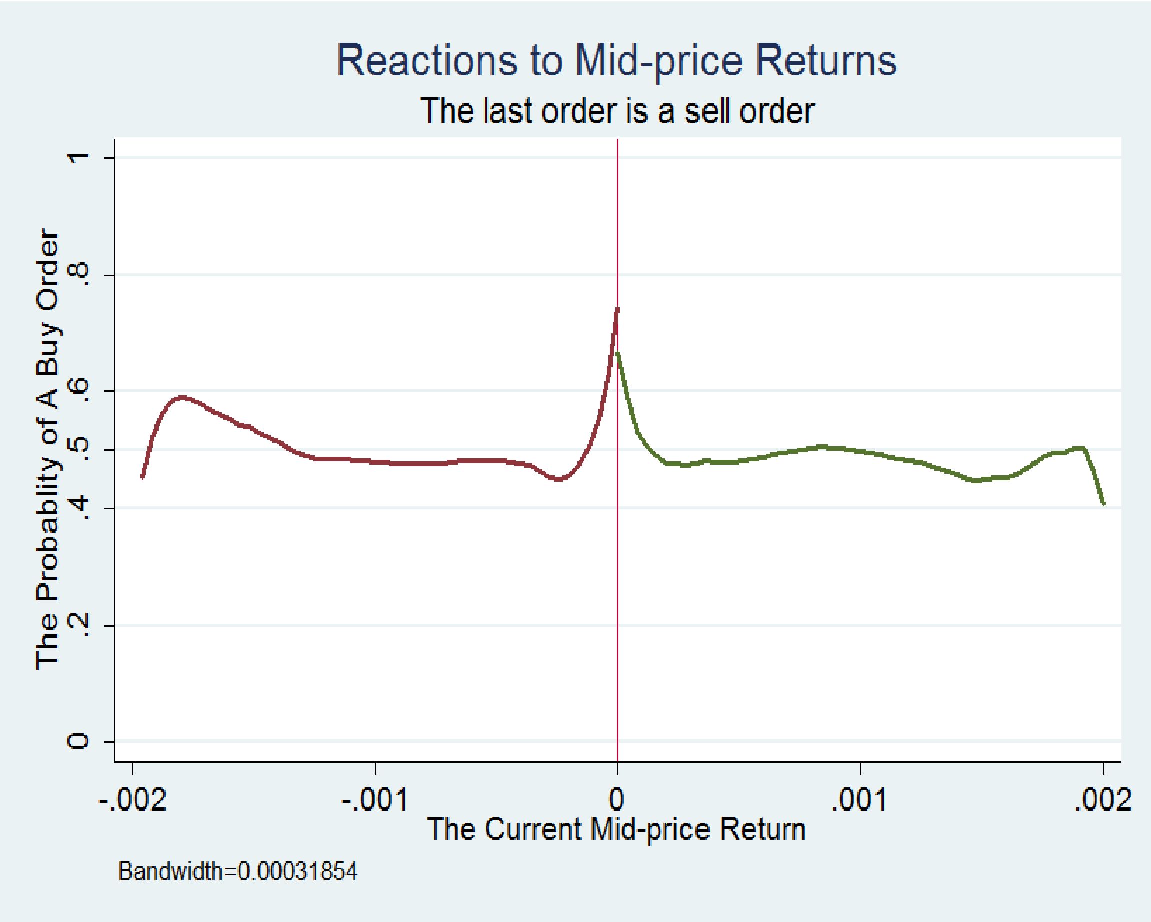

Figure 4.14: Reactions to Mid-price Returns

This figure shows the treatment effect when the last order is a sell order. The vertical axis represents the probability of a buy order. The horizontal axis represents the current mid-price return. The vertical line at the point where the return is zero splits the graph into two parts. If the last return was positive/negative and the current return falls on the right part, the trend is continuing/ breaking; if the current return falls on the left part, the trend breaks/ continues.
The USD/DEM transaction data on the Reuters D2000-1 system are used.
The figure is generated by the STATA code from Nichols (2011)

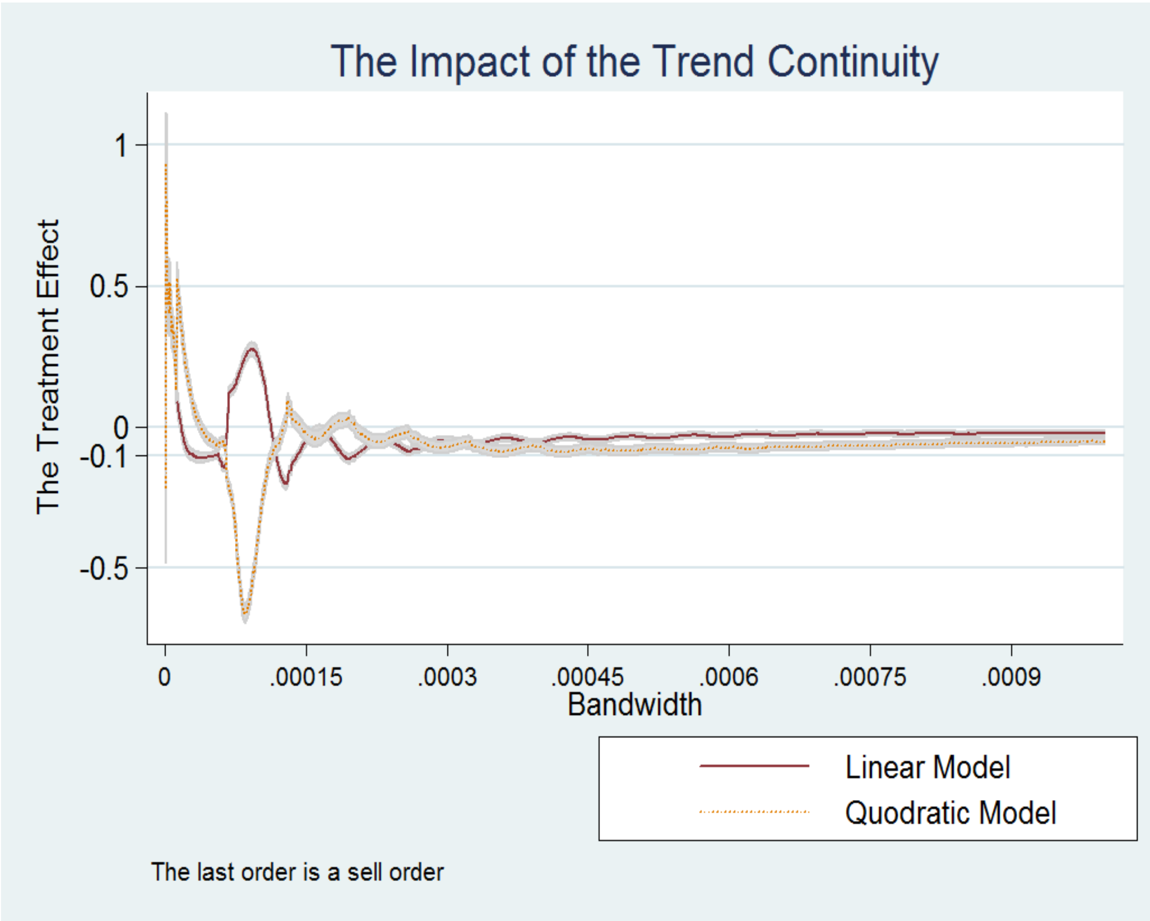

Figure 4.15: The Treatments Effect with Various Bandwidth

This figure shows the pattern of the coefficient τ (the treatment effect) against the bandwidth. The vertical axis represents the treatment effects. The horizontal axis represents bandwidth. The USD/DEM transaction data on the Reuters D2000-1 system are used.

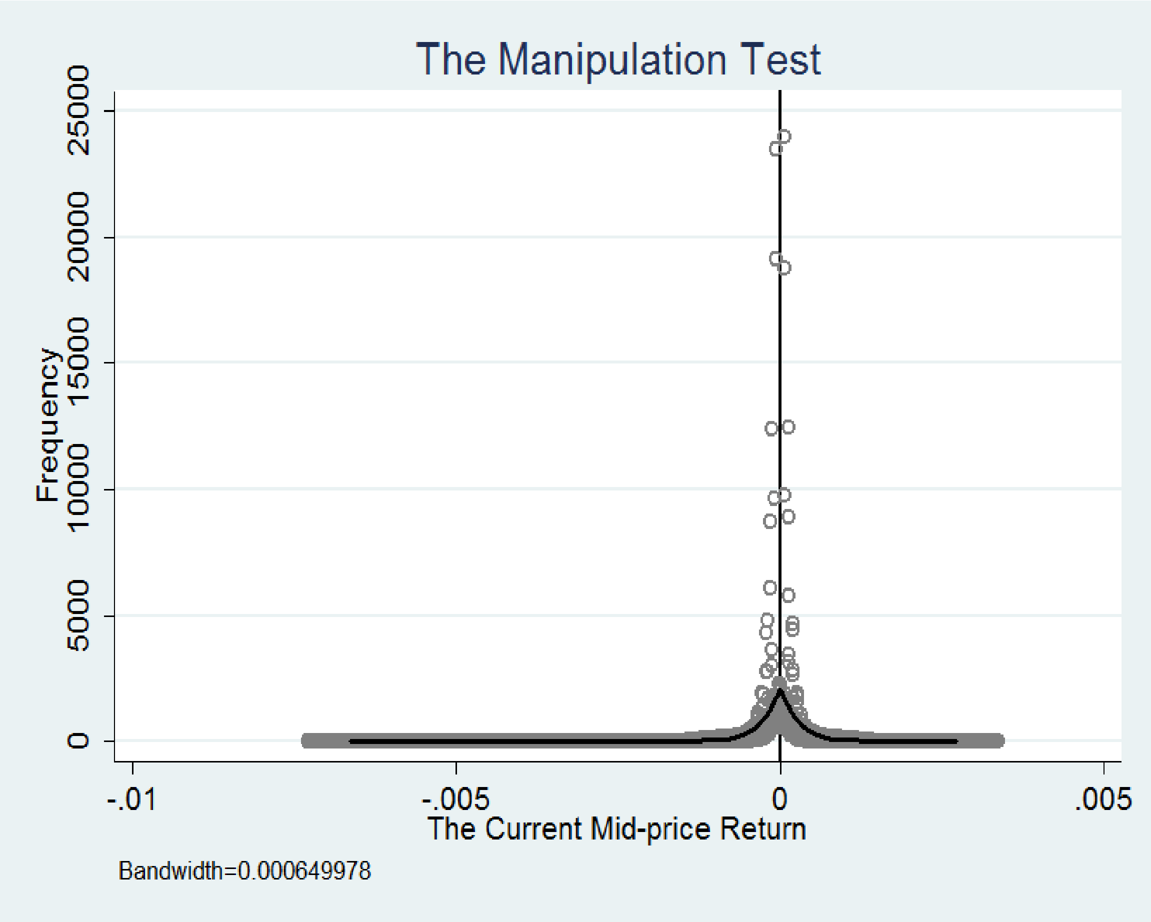

Figure 4.16: The Validity Test: Density

This figure shows the density of the current mid-price return. The vertical axis is the number of observations. The horizontal axis is the values of the returns.
The USD/DEM transaction data on the Reuters D2000-1 system are used.

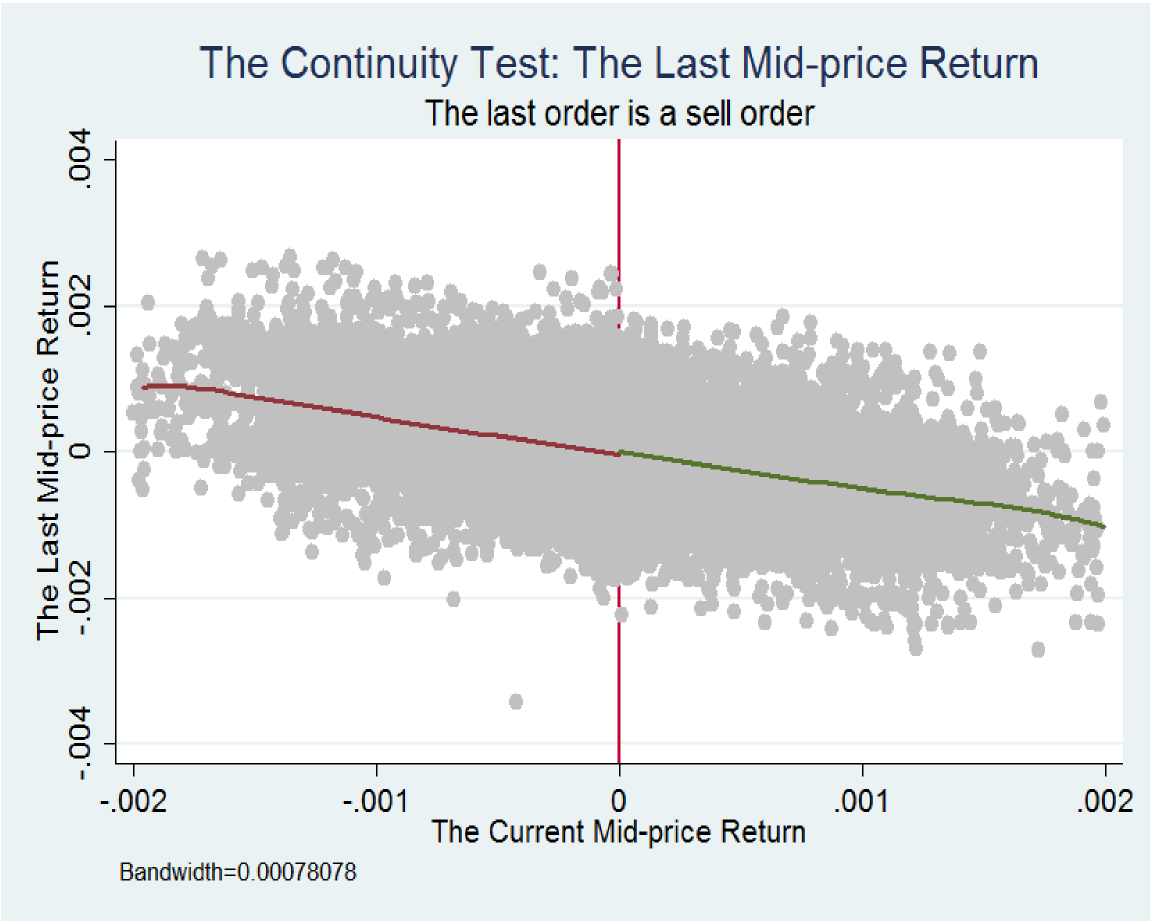

Figure 4.17: The Validity Test: Continuous the Last Return

This figure shows the treatment effect of the current return and the last return. The vertical axis represents the last mid-price return. The horizontal axis represents the current mid-price return. The vertical line at the point where the return is zero splits the graph into two parts.
The USD/DEM transaction data on the Reuters D2000-1 system are used.
This figure is generated by the STATA code from Nichols (2011)

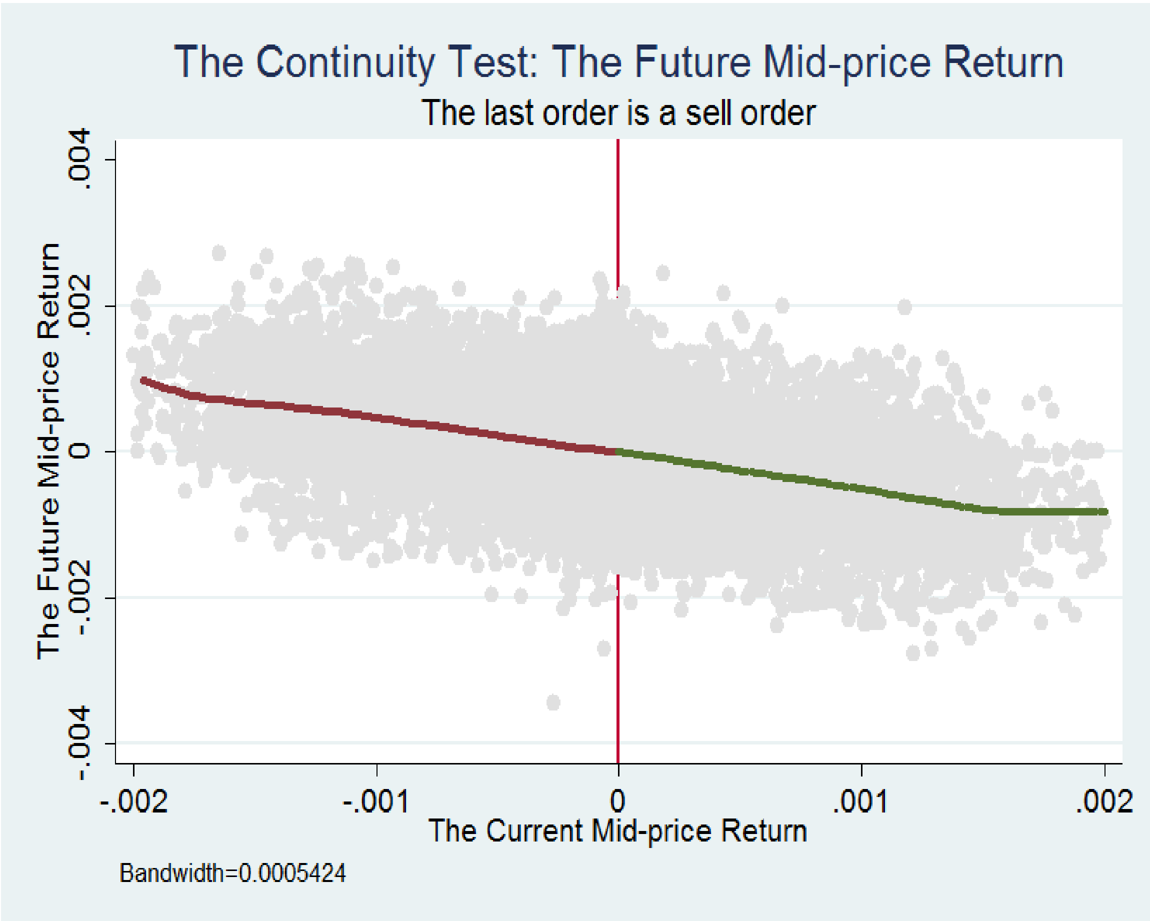

Figure 4.18: The Validity Test: Continuous the Future Return

This figure shows the treatment effect of the current return and the future return. The vertical axis represents the future mid-price return. The horizontal axis represents the current mid-price return. The vertical line at the point where the return is zero splits the graph into two parts.
The USD/DEM transaction data on the Reuters D2000-1 system are used.
This figure is generated by the STATA code from Nichols (2011)

Appendix: Robustness

In this section, we will show that the spread estimation does not have a big impact on the main results in this chapter. In the previous sections of this chapter, the mid-prices and other variables are calculated according to the estimated spread of the HS estimator, which is 7.94×10^{-5}. In this section, we run the main regression in this chapter using variables according to various values of the spread. Firstly, we check the case that the spread is zero, in other words, we use the transaction prices directly. Secondly, we check the case that the spreads is 1.665×10^{-4}, which is the highest possible value of the spread. The results are shown in Table (4.14). We find that the coefficients and the marginal effects in the table are very similar to the main results of this chapter, which suggests that our results are robust to the spread estimation.

Furthermore, in the previous chapter, we observe the autocorrelation of order flow, which might influence the results of the main regression in this chapter. In this section, we run the regression controlling for higher order autocorrelation. Tables 4.15 and 4.16 reports the results. We find that the coefficients and the marginal effects in the table are very similar to the main results of this chapter, which suggests that our results are robust to autocorrelated order flow.

Table 4.14: Robustness: various spreads

USDDEM	spread=0	Marginal Effects	spread=1.665×10^{-4}	Marginal Effects
$sign$	0.596*** (134.99)	0.238*** (134.99)	0.0351*** (8.28)	0.0140*** (8.28)
$DK_1 \cdot sign$	-0.604*** (-98.96)	-0.241*** (-98.96)	-0.567*** (-93.08)	-0.226*** (-93.08)
$DK_2 \cdot sign$	-0.772*** (-90.87)	-0.308*** (-90.87)	-0.725*** (-79.65)	-0.289*** (-79.65)
$DK_3 \cdot sign$	-0.887*** (-70.43)	-0.354*** (-70.43)	-0.891*** (-59.62)	-0.355*** (-59.62)
$DK_4 \cdot sign$	-0.944*** (-49.3)	-0.376*** (-49.3)	-1.00507*** (-39.93)	-0.401*** (-39.93)
$DK_5 \cdot sign$	-1.000*** (-45.56)	-0.399*** (-45.56)	-1.0882*** (-32.64)	-0.434*** (-32.63)
$w \cdot sign$	-3.0507*** (-34.03)	-1.217*** (-34.03)	-1.126*** (-12.67)	-0.449*** (-12.67)
contant	0.000918 (0.36)		0.00266 (1.04)	
Probability	0.500		0.500	

This table reports the results and marginal effects for the probit model as follows.

$$OF_t = \Phi\left[\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t\right] + \varepsilon_t$$

Variables are calculated accroding to the estimated bid-ask spread respectively. Standard deviations are in brackets *** Significant at 0.1% level

Table 4.15: Robustness: Autocorrelated Order Flow

OF		Marginal Effects
$DK_1 \cdot sign$	-0.684*** (-109.76)	-0.273*** (-109.76)
$DK_2 \cdot sign$	-0.907*** (-99.64)	-0.362*** (-99.64)
$DK_3 \cdot sign$	-1.024*** (-71.93)	-0.409*** (-71.93)
$DK_4 \cdot sign$	-1.104*** (-47.57)	-0.440*** (-47.57)
$DK_5 \cdot sign$	-1.137*** (-37.06)	-0.454*** (-37.06)
$w^{adj} \cdot sign$	-1.281*** (-14.26)	-0.511*** (-14.26)
$sign$	0.293*** (68.23)	0.117*** (68.23)
OF		
L.1	0.121*** (22.63)	0.0483*** (22.66)
L.2	0.272*** (50.91)	0.108*** (51.23)
L.3	0.128*** (24.46)	0.0510*** (24.50)
L.4	0.0796*** (15.30)	0.0317*** (15.30)
L.5	0.0584*** (11.25)	0.0233*** (11.25)
L.6	0.0543*** (10.47)	0.0217*** (10.47)
L.7	0.0545*** (10.51)	0.0218*** (10.51)
L.8	0.0393*** (7.57)	0.0157*** (7.57)
L.9	0.0397*** (7.65)	0.0159*** (7.65)
L.10	0.0352*** (6.77)	0.0140*** (6.78)
L.11	0.0288*** (5.55)	0.0115*** (5.55)
L.12	0.0323*** (6.22)	0.0129*** (6.22)

This table reports the results and marginal effects for the probit model as follows.

$$OF_t = \Phi\left[\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \sum_{i=1}^{5} \delta_i\, Dk_i\, sign_t + \sum_{i=1}^{3} 2\alpha_i\, OF_{t-i}\right] + \varepsilon_t$$

Standard deviations are in brackets *** Significant at 0.1% level

Table 4.16: Robustness: Autocorrelated Order Flow continuous

OF		Marginal Effects
L.13	0.0355*** (6.84)	0.0142*** (6.84)
L.14	0.0327*** (6.30)	0.0131*** (6.30)
L.15	0.0283*** (5.46)	0.0113*** (5.46)
L.16	0.0209*** (4.03)	0.00834*** (4.03)
L.17	0.0120* (2.31)	0.00480* (2.31)
L.18	0.0165** (3.17)	0.00656** (3.17)
L.19	0.0222*** (4.27)	0.00885*** (4.27)
L.20	0.0178*** (3.42)	0.00709*** (3.42)
L.21	0.0149** (2.86)	0.00594** (2.86)
L.22	0.0147** (2.83)	0.00586** (2.83)
L.23	0.0114* (2.19)	0.00453* (2.19)
L.24	0.0205*** (3.94)	0.00816*** (3.94)
L.25	0.0134* (2.57)	0.00533* (2.57)
L.26	0.0236*** (4.54)	0.00940*** (4.54)
L.27	0.0157** (3.02)	0.00626** (3.02)
L.28	0.00863 (1.66)	0.00344 (1.66)
L.29	0.0187*** (3.60)	0.00745*** (3.60)
L.30	0.00915 (1.76)	0.00365 (1.76)
L.31	0.0137** (2.65)	0.00549** (2.65)
L.32	0.0144** (2.79)	0.00576** (2.79)

This table reports the results and marginal effects for the probit model as follows.

$$OF_t = \Phi\left[\beta_1 + \beta_2 sign_t + \beta_3 w_t^{adj} sign_t + \sum_{i=1}^{5} \delta_i \, Dk_i \, sign_t + \sum_{i=1}^{3} 2\alpha_i \, OF_{t-i}\right] + \varepsilon_t$$

Standard deviations are in brackets *** Significant at 0.1% level

Chapter 5

Summary and Further Research

5.1 Summary

The thesis contains three essays related to the microstructure of the foreign exchange market. The first three chapters examine issues related to the spread and in particular its estimation and decomposition while the third chapter analyses the behaviour of professional traders.

In Chapter Two, three bid-ask spread estimators are compared using simulated data and pure market data. Huang and Stoll's estimator performs better than Roll's and Corwin's estimators, and in most cases it can estimate the spread accurately. When positive (negative) feedback trading exists, Huang and Stoll's estimator overestimates (underestimates) the spread. Roll's and Corwin's estimators provide accurate estimations only when the mean of the spread is much larger than the standard deviation of the spread.

In Chapter Three, we replicate the work in Daníelsson and Payne (2002) and describe features of mid-prices and spreads using the EUR/USD indicative data in 2003. The basic findings are similar: the spread is time-varying and the mid-price returns are negatively autocorrelated. Then we study three months USD/DEM transaction data on the Reuters D2000-1 system. The Roll, Huang and Stoll (HS), and Corwin and Schultz (CS) estimators are used to estimate the spread. According to the conclusion in Chapter Two, the tick-by-tick HS estimator (7.94×10^{-5}) might be the one that produces the best estimates of the true spread.

In Chapter Four, we introduce a new indicator spread estimating and decomposition model which is designed specifically for multi-dealer markets where dealers can either adjust the mid-price or place direct orders to control the inventory. The new model will not give ambiguous results (e.g. negative or larger than one fractions) like Huang and Stoll's model when it decomposes the components of the spread. With time dummies, the new model can be used to study the intra-day pattern of the spread.

In Chapter Five, we study professional traders' behaviour in the foreign exchange market. The tick-by-tick transaction data in the foreign exchange market are used, because they include information of price returns and traders' reactions to these price returns. The price trends are decomposed in to two parts: the streak length, the streak width, a decomposition which is more general than those in previous literature where only dual-outcome cases are considered. We find that the probability of following the trend will go down as the streak length or the streak width become larger. The probability will rise when the streak length is very large. We find that traders are unlikely to be rational and that the gamblers' fallacy and the hot hand fallacy can help explain the phenomenon of trading with/against the trend.

5.2 Further Research

The new spread estimating and decomposing model relies on information of price returns and order flows. However, the order flows are not always available. If it is reasonable to assume that the high (low) price in a time interval is normally at the ask (bid) price then order flow series can be deduced using that information.

Combining the fact that technical analysis is widely used in the financial market and the finding of the thesis which suggests that trader's reactions to the trend are irrational, it would be interesting in future work to study the relationship between the technical analysis and the fallacies.

The underlying price generating process of the exchange rate is unknown and is complicated. In chapter three, we have only controlled for rational expectations of further price returns. In further research, other factors such as news shocks and hot potato trading could be considered.

Another possible extension would be duration analysis. Are the fallacies taking place because traders have to make decisions in a very short time? Will the fallacies vanish if traders have a relatively long time to make their decisions? In laboratory experiments the underlying price generating process and the time that takes traders to make decision can be controlled. It would be fruitful to consider experimental and field evidence together.

Bibliography

Admati, A. and P. Pfleiderer (1988). A theory of intraday patterns: volume and price variability. *Review of Financial Studies 1*(1), 3–40. 3

Amihud, Y. and H. Mendelson (1980). Dealership market : Market-making with inventory. *Journal of Financial Economics 8*(1), 31–53. 3, 17

Anand, A. and A. K. Karagozoglu (2006). Relative performance of bid–ask spread estimators: Futures market evidence. *Journal of International Financial Markets, Institutions and Money 16*(3), 231 – 245. 7, 15, 19

ap Gwilym, O. and S. Thomas (2002). An empirical comparison of quoted and implied bid-ask spreads on futures contracts. *Journal of International Financial Markets, Institutions and Money 12*(1), 81–99. 7, 15, 19

Asparouhova, E. N., M. G. Hertzel, and M. L. Lemmon (2009). Inference from streaks of random outcomes: Experimental evidence on beliefs in regime shifting and the law of small numbers. *Management Science*. 140, 147

Ayton, P. and I. Fischer (2004). The hot hand fallacy and the gambler's fallacy: two faces of subjective randomness? *Memory and Cognition 32*, 1369–78. 11, 133

Bandi, F. M. and J. R. Russell (2006). Separating microstructure noise from volatility. *Journal of Financial Economics 79*(3), 655–692. 14

Banti, C., K. Phylaktis, and L. Sarno (2012). Global liquidity risk in the foreign exchange market. *Journal of International Money and Finance 31*(2), 267 – 291. 14

Barberis, N., A. Shleifer, and R. Vishny (1998). A model of investor sentiment. *Journal of Financial Economics 49*(3), 307–343. 10, 130, 133

Baxter, M. and A. C. Stockman (1989). Business cycles and the exchange-rate regime: Some international evidence. *Journal of Monetary Economics 23*, 377–400. 10

Bessembinder, H. (1994). Bid-ask spreads in the interbank foreign exchange markets. *Journal of Financial Economics 35*(3), 317–348. 5, 16, 17, 40

Bjønnes, G. H. and D. Rime (2005). Dealer behavior and trading systems in foreign exchange markets. *Journal of Financial Economics 75*(3), 571–605. 4, 107, 116

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. 4, 101

Bollerslev, T. and I. Domowitz (1993). Trading patterns and prices in the interbank foreign exchange market. *The Journal of Finance 48*(4), 1421–1443. 4, 123

Bollerslev, T. and M. Melvin (1994). Bid ask spreads and volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics 36*, 355–372. 5, 16

Brown, W. O. and R. D. Sauer (1993). Does the basketball market believe in the hot hand? comment. *The American Economic Review 83*(5), 1377–1386. 12, 132

Camerer, C. F. (1989). Does the basketball market believe in the 'hot hand,'? *The American Economic Review 79*(5), 1257–1261. 12, 132

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. 5, 100

Chan, L. K. and J. Lakonishok (1993). Institutional trades and intraday stock price behavior. *Journal of Financial Economics 33*(2), 173 – 199. 118

Chan, L. K. and J. Lakonishok (1995). The behavior of stock prices around institutional trades. *The Journal of Finance 50*(4), 1147–1174. 118

Choi, J. Y., D. Salandro, and K. Shastri (1988). On the estimation of bid-ask spreads: Theory and evidence. *The Journal of Financial and Quantitative Analysis 23*(2), 219–230. 6, 17, 18, 19, 25, 72, 97

Chordia, T., R. Roll, and A. Subrahmanyam (2002). Order imbalance, liquidity, and market returns. *Journal of Financial Economics 65*(1), 111 – 130. 118

Chordia, T., R. Roll, and A. Subrahmanyam (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics 76*(2), 271 – 292. 118

Clotfelter, C. T. and P. J. Cook (1991, July). The "gambler's fallacy" in lottery play. *NBER Working Paper Series* (3769). 12

Clotfelter, C. T. and P. J. Cook (1993). The "gambler's fallacy" in lottery play. *Management Science 39*(12), 1521–1525. 131

Corwin, S. A. and P. Schultz, P. (2012). A simple way to estimate bid-ask spreads from daily high and low prices. *The Journal of Finance 67*(2), 719–760. 6, 7, 15, 18, 30, 33, 86, 98

Croson, R. and J. Sundali (2005). The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of Risk and Uncertainty 30*(3), 195–209. 12, 132

Daníelsson, J. and R. Love (2006). Feedback trading. *International Journal of Finance & Economics 11*(1), 35–53. 40, 130

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. 5, 16, 40, 58, 79, 90, 100, 135, 191

De Grauwe, P. and M. Grimaldi (2005). Heterogeneity of agents, transaction costs and the exchange rate. *Journal of Economic Dynamics and Control 29*, 691–719. 10

De Grauwe, P. and M. Grimaldi (2006). Exchange rate puzzles: A tale of switching attractors. *European Economic Review 50*, 1–33. 10

De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Positive feedback investment strategies and destabilizing rational speculation. *The Journal of Finance 45*(2), 379–395. 40, 130

Ding, L. (2009). Bid–ask spread and order size in the foreign exchange market: an empirical investigation. *International Journal of Finance & Economics 14*(1), 98–105. 4

Easley, D. and M. O'Hara (1988). Contracts and asymmetric information in the theory of the firm. *Journal of Economic Behavior & Organization 9*(3), 229 – 246. 3

Evans, M. D. D. (1998). The microstructure of foreign exchange dynamics. 115

Evans, M. D. D. (2002). Fx trading and exchange rate dynamics. *The Journal of Finance 57*(6), 2405–2447. 105

Evans, M. D. D. and R. K. Lyons (2002). Order flow and exchange rate dynamics. *The Journal of Political Economy 110*(1), 170–180. 8, 92, 106, 107, 113, 118, 130, 135, 140

Flood, R. P. and A. K. Rose (1995). Fixing exchange rates a virtual quest for fundamentals. *Journal of Monetary Economics 36*, 3–37. 10

Frankel, J. A. and K. A. Froot (1986). Understanding the us dollar in the eighties: The expectations of chartists and fundamentalists. *Economic Record special issue*, 24–38. 10

Frankel, J. A. and K. A. Froot (1990). The rationality of the foreign exchange rate. chartists, fundamentalists and trading in the foreign exchange market. *American Economic Review 80*. 10

Frankel, J. A. and K. A. Froot (1991). Chartists, fundamentalists and the demand for dollars. *NBER Working Paper* (r1655). 10

Garman, M. B. and M. J. Klass (1980). On the estimation of security price volatilities from historical data. *The Journal of Business 53*(1), 67–78. 31

George, T. J., G. Kaul, and M. Nimalendran (1991). Estimation of the bid-ask spread and its components: A new approach. *The Review of Financial Studies 4*(4), 623–656. 7, 17, 18, 25, 97

Gilovich, T., R. Vallone, and A. Tversky (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology 17*, 295–314. 12, 132

Glosten, L. and P. Milgrom (1985). Bid ask and transaction prices in a specialist market with heterogeneously informed trades. *Journal of Financial Economics 14*(1), 71–100. 3, 4, 6, 17, 110

Glosten, L. R. and L. E. Harris (1988). Estimating the components of the bid/ask spread. *Journal of Financial Economics 21*(1), 123–142. 6, 7, 17, 18

Goodhart, C., T. Ito, and R. Payne (1996). One day in june 1993: A study of the working of reuters d2000-2 electronic foreign exchange trading system. In J. Frankel, G. Galli, and A. Giovannini (Eds.), *The Microstrttcture of Foreign Exchange Markets*. Chicago: University of Chicago Press. 5, 16, 40, 58, 90, 135

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*. 115

Goodhart, C. and R. Payne (1996). Microstructural dynamics in a foreign exchange electronic broking system. *Journal of International Money and Finance 15*(6), 829 – 852. 5, 16, 40

Goodhart, C. A. E. and L. Figliuoli (1991). Every minute counts in financial markets. *Journal of International Money and Finance 10*(1), 23–52. 10

Goyenko, R. Y., C. W. Holden, and C. A. Trzcinka (2009). Do liquidity measures measure liquidity? *Journal of Financial Economics 92*(2), 153 – 181. 6, 7, 15

Guryan, J. and M. S. Kearney (2008). Gambling at lucky stores: Empirical evidence from state lottery sales. *American Economic Review 98*, 458–473. 13, 132, 133, 149

Harris, L. (1990). Statistical properties of the roll serial covariance bid/ask spread estimator. *The Journal of Finance 45*(2), 579–590. 26

Hartmann, P. (1999). Trading volumes and transaction costs in the foreign exchange market: Evidence from daily dollar-yen spot data. *Journal of Banking & Finance 23*(5), 801–824. 5, 17

Hasbrouck, J. (1991). Measuring the information content of stock trades. *The Journal of Finance 46*(1), 179–207. 40

Hasbrouck, J. (2004). Liquidity in the futures pits: Inferring market dynamics from incomplete data. *The Journal of Financial and Quantitative Analysis 39*(2), 305–326. 7, 18, 98

Hasbrouck, J. (2009). Trading costs and returns for u.s. equities: Estimating effective costs from daily data. *The Journal of Finance 64*(3), 1445–1477. 7, 18, 98

Ho, T. S. Y. and H. R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9*(1), 47 – 73. 3, 17, 105

Ho, T. S. Y. and H. R. Stoll (1983). The dynamics of dealer markets under competition. *The Journal of Finance 38*(4), pp. 1053–1074. 3

Holden, C. W. (2009). New low-frequency spread measures. *Journal of Financial Markets 12*(4), 778 – 813. 6, 15

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. 4, 5, 16, 100, 123

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. i, 1, 6, 7, 15, 17, 18, 28, 99, 100, 118, 127, 137

Huber, J., M. Kirchler, and T. Stöckl (2008). The hot hand belief and the gambler's fallacy in investment decisions under risk. *Theory and Decision*. 11

Imbens, G. and K. Kalyanaraman (2012, July 1, 2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959. 177

Ito, T. and Y. Hashimoto (2006a). Intra-day seasonality in activities of the foreign exchange markets evidence from the electronic broking system. 124

Ito, T. and Y. Hashimoto (2006b). Microstructure of the yen/dollar foreign exchange market: Patterns of intra-day activity revealed in the electronic broking system. *Journal of the Japanese and International Economics 20*, 637–664. 5, 16

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*(6), 1315–1335. 3, 6, 17, 118

Lee, D. S. (2008). Randomized experiments from non-random selection in u.s. house elections. *Journal of Econometrics 142*(2), 675–697. 172, 173, 174

Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of Economic Literature 48*(2), 281–355. 170, 172, 173, 174, 175, 177

Lesmond, D. A., J. P. Ogden, and C. A. Trzcinka (1999). A new estimate of transaction costs. *The Review of Financial Studies 12*(5), pp. 1113–1141. 6

Lillo, F., S. M. and J. D. Farmer (2005). Theory for long memory in supply and demand. *Physical Review E*. 118

Lyons, R. K. (1995). Tests of microstructural hypotheses in the foreign exchange market. *Journal of Financial Economics 39*(2-3), 321–351. 4, 7, 16, 17, 18, 40, 115

Lyons, R. K. (1997). A simultaneous trade model of the foreign exchange hot potato. *Journal of International Economics 42*(3-4), 275–298. 92, 107, 118

Mancini, L., A. Ranaldo, and J. Wrampelmeyer (2013). Liquidity in the foreign exchange market: Measurement, commonality, and risk premiums. *The Journal of Finance 68*(5), 1805–1841. 14

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics 142*(2), 698–714. 174, 179

McGroarty, F., O. ap Gwilym, and S. Thomas (2006). Microstructure effects, bid-ask spreads and volatility in the spot foreign exchange market pre and post-emu. *Global Finance Journal 17*(1), 23–49. 15, 100, 104

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. 16

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. 4, 5, 100, 123

Nichols, A. (2011). rd 2.0: Revised stata module for regression discontinuity estimation. 182, 185, 186

Nofsinger, J. R. and R. W. Sias (1999). Herding and feedback trading by institutional and individual investors. *The Journal of Finance 54*(6), 2263–2295. 40

O'Hara, M. (1995). *Market Microstructure Theory*. Cambridge, MA.: Basil Blackwell. 1

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *The Journal of Business 53*(1), 61–65. 31, 36, 38

Rabin, M. and D. Vayanos (2010). The gambler's and hot-hand fallacies: Theory and applications. *Review of Economic Studies 77*(2), 730–778. 11, 130, 133, 140, 147

Rao, J. M. (2009). When the gambler's fallacy becomes the hot hand fallacy: An experiment with experts and novices. 11, 132, 140, 147

Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance 39*(4), 1127–1139. 6, 8, 15, 17, 18, 19, 23

Stoll, H. R. (1978). The supply of dealer services in securities markets. *The Journal of Finance 33*(4), pp. 1133–1151. 3, 4, 16, 17

Stoll, H. R. (1989). Inferring the components of the bid-ask spread: Theory and empirical tests. *The Journal of Finance 44*(1), 115–134. 6, 7, 17, 18, 19, 26, 97

Sundali, J. and R. Croson (2006). Biases in casino betting: The hot hand and the gambler's fallacy. *Judgment and Decision Making 1*(1), 1–12. 12, 132

Thistlethwaite, D. L. and D. T. Campbell (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology 51*(6), 309–317. 170

Tversky, A. and D. Kahneman (1974). Judgment under uncertainty: Heuristics and biases. *Science 185*(4157), 1124–1131. 132

Viswanathan, S. and J. J. D. Wang (2004). Inter- dealer trading in financial markets. *The Journal of Business 77*(4), 987–1040. 100, 118