ESSAYS ON THE APPLICATION OF THE MARKET MICROSTRUCTURE APPROACH IN EXCHANGE RATE ECONOMICS

Liang Ding

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill

2006

Approved by
Advisor: Stanley Black
Reader: Robert Connolly
Reader: Richard Froyen
Reader: William Parke

ABSTRACT

LIANG DING: Essays on the Application of the Market Microstructure Approach in Exchange Rate Economics

(Under the direction of Stanley Black)

ESSAY ONE: Exchange Rate Dynamics and Order Flow: An Alternative Model

Realizing and overcoming the disadvantages of the well-known model by Evans and Lyons (2002), this essay presents an alternative model to connect the exchange rate dynamics and foreign exchange market order flow from the perspective of the dealer's quoting behavior in a more realistic environment. Agreeing with previous works, the essay shows theoretically and empirically that order flow has significant positive effects on the short-run exchange rate change.

ESSAY TWO: Market Structure and Dealer's Quoting Behavior in the Foreign Exchange Market

Realizing that the structure of inter-dealer and customer foreign exchange markets is different in terms of market transparency and information asymmetry, this essay examines how the differences affect the same dealer's quotes in the two markets. It is shown that customer spreads are generally wider than inter-dealer ones, while their differential tends to fall with the rise in order size. Meanwhile, the same dealer's mid-quotes are found to be

equal to each other in the two markets. In contrast to the preceding models claiming that the order size has a positive effect on the spread, the impact is found to be ambiguous in our study. Supportive empirical evidence can be provided by the new data set applied in the essay.

ESSAY THREE: The Term Structure of the Forward Premium Puzzle

In contrast to many previous studies that examined the efficiency of the forward market in one particular horizon, this essay tests the unbiased hypothesis by using forward rates in the same period but with various maturities. The hypothesis continues to be rejected as before in the horizon of 1-week and 1-month, but is found to be accepted in the 1-day horizon. It is also found that the magnitude of the slope coefficient increases as the horizon rises. A model in the spirit of the term structure model of interest rates is proposed to reconcile the inconsistency, and the results of empirical testing support the model.

ACKNOWLEDGEMENTS

I thank Stanley Black, Robert Connolly, Patrick Conway, Richard Froyen and William Parke for extensive comments that helped clarify many of the issues contained in the dissertation, and I am grateful for William Parke's help of the data collection. I also benefit a lot from Richard Lyons and Michael Melvin about the FX market microstructure. Many thanks also to the participants in UNC seminars for stimulating discussions. Finally, special thanks to my wife Hui Sun and my parents for their continuous support and understanding.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

ESSAY ONE

EXCHANGE RATE DYNAMICS AND ORDER FLOW:

AN ALTERNATIVE MODEL

1. Introduction

Economists have spent decades seeking models to explain the movement of the exchange rate. Traditionally, most of these models were built from macro perspective, e.g. Balance of Payment Flow Approach, Purchasing Power Parity, Flexible price and sticky price monetary approach, portfolio balance model and general equilibrium model and so on. However, the poor empirical performance of the traditional exchange rate determination approaches puts these models in crisis, especially for high frequency data. In the classic papers by Meese and Rogoff (1983a, 1983b), they concluded that "the proportion of (monthly or quarterly) exchange rate changes that the current models can explain is essentially zero".

The exchange rates are generated in the FX market by dealers' bid-ask quotes. In this market, not all information is publicly available; different agents have different information sets; market participants might be irrational sometimes; and a change of trading mechanics can have a dramatic effect on the process of exchange rate formation. Due to the limitations of macro-basis models, the traditional approaches usually neglect those micro features of the actual FX market and might not be able to capture the exchange rate movements, especially in the short run.

The market microstructure approach, which has been extensively studied and applied in the finance field, provides a solid platform to incorporate the micro features of the FX market into the model and study the exchange rate formation process from the micro perspective. Just in the past several years, increasing attention has been attracted to this field and some economists are making significant advancement. Evans and Lyons (2002), for instance, designed a multiple-dealer simultaneous trading model to determine the exchange rate dynamics. Osler (2006) incorporated the microstructure approach into the traditional macro-based model by integrating the order flow with the Balance of Payments approach. Other major contributions have been made by Lyons (2001), Evans (2002), Bjonnes and Rime (1998) etc. Among them, the most well-known model is the one proposed in Evans and Lyons (2002). The model recognized that order flow conveys some information relevant to the exchange rate that is not publicly available, and dealers can use the information to adjust quotes. A more appealing fact is that with the increasing availability of transaction data, the empirical test of the new approach also provides encouraging results that order flow and the nominal exchange change rate are strongly positively correlated.

Although Evans and Lyons' model was successful to some extent, some aspects of the model are vulnerable to challenges. First, they assumed that every dealer quotes only one scalar price, which in practice is not true. Actually, as market makers, dealers quote a bid price for selling and an ask price for buying simultaneously. Second, they claimed that no arbitrage makes different dealers quote an identical price. In fact, this assumption is not true either: different dealers give different quotes all the time. Because of transaction cost and inadequate information disclosure in the FX market, the dealers' quotes can be different even with the chance of arbitrage existing. Third, they divided one trading period into three rounds, in which a dealer trades first only with customers, then only with other dealers, and finally with customers again. However, this model structure is not realistic at all; in fact, dealers don't have special trading time reserved only for either of them. Both types of trading take place simultaneously, and dealers cannot expect them to happen in the order that the model assumed. Fourth, to be compatible with the setup of trading rounds, they assumed that a dealer quotes only for inter-dealer market or customer market in each round. But in the real market, a dealer should give quotes for both markets at any time requested.

Accordingly, I relax these unrealistic assumptions in this essay and try to connect order flow and exchange rate dynamics in a more practical and realistic environment. Specifically, I first relax the assumption that dealers quote only one scalar price and derive the exchange rate as the mid-quote of bid-ask quotes. Second, I relax the assumption that dealer's quotes must be identical and allow heterogeneity in the way that each dealer's quoting problem is solved based on the specific information set they receive. Third, my model reflects not only an inventory effect but also an information effect of the order flow. Fourth, my model is applicable to explain the exchange rate dynamics in any period universally and no special timing set-up is needed. The last three features are new and become the main contributions to the literature.

The remainder of the essay is organized as follows: section 2 introduces the structure of the FX market; section 3 proposes the theoretical framework of the model. Empirical evidence will be provided in section 4. Finally conclusions are drawn in section 5.

2. The structure of the Foreign Exchange Market

Not knowing the current FX market structure would make it hard to understand the basis and structure of the model. It is necessary to introduce the structure of the FX market before I construct the model. In fact, the structure of the FX market has been changing constantly over the past two decades. I will first introduce the general features of the FX market, and then discuss its evolution of time, and finally compare the structure of the FX market with that of the traditional stock market to highlight its special features.

2.1 General features of the FX market structure

There are three major participants in the FX market: dealers, customers, and brokers. Dealers act as intermediates of the market. They quote prices, provide liquidity and seek profit by taking risks in the market. Dealers in the FX market include most commercial banks, some investment banking firms and other financial institutions that buy and sell foreign exchange for their customers and themselves. Customers are the ultimate end-users of currency, and they can be central banks, government, importers and exporters of goods, and financial institutions such as hedge funds. In the transactions with the dealers, customers can not quote price; what they can do is to take a dealer’s quote or leave it. Finally, brokers only match dealers in the inter-bank market without being a party to the transaction themselves and without taking positions. They just play roles that bring together a buying dealer and a selling dealer in return for a fee or commission.

The foreign exchange market can basically be divided into the inter-dealer (inter-bank) market and the customer market. In the inter-dealer market, dealers can contact each other directly, ask for quote, and make transactions. Since it is between two dealers, this style of trading is called direct trading or bilateral inter-dealer trading. Meanwhile, dealers can also contact brokers asking for other dealer’s quoting information. Brokers will provide them the best quote available, match sellers and buyers, and make the transaction take place. Since this style of trading is through brokers, it is called indirect trading or brokered inter-dealer trading. The customer market is the part of the FX market where customers buy and sell foreign currencies. Unlike other equity markets, customers do not trade with each other in the customer market, and they generally trade only with the dealers.

In practice, the exchange rate is not a single price; actually, it is a pair of prices: the ask price for selling and the bid price for buying. Usually, the ask price is larger than the bid price, and the gap between them is the spread. In empirical research, the observation of the exchange rate is usually taken as the middle point of the ask price and the bid price. In the FX market, dealers need to quote both prices. And also, since they trade in the inter-dealer market and the customer market simultaneously, the dealers need to give quotes in both markets. Therefore, dealers should give two quote pairs: the bid-ask price for other dealers and another pair for the customers.

In the inter-dealer market, any dealer can initiate the transaction and ask for another dealer's quotes. He also can be the passive side who receives the request and gives other dealers his quotes. In the customer market, it is always the customers that initiate the trading, and only dealers give quotes. In inter-dealer trading, the active side requests the quotes, revealing the size of order (the quantity he wants to trade), but without revealing the tendency (to buy or sell). In customer trading, the active side (always the customer) asks for the quotes, telling the passive side (always a dealer) the size of order, usually as well as the tendency.

2.2 Evolution of the FX market structure

With the development of technology, the microstructure of the FX market has been changing dramatically in the past two decades. Before the 1990s, the foreign exchange market was still operating in a very traditional way, which had not been changed a lot since 1930's. In the inter-dealer market, all brokers were real persons, and therefore also called "voice brokers". They collected information about different dealers' quotes, and announced the best ask and bid price available when dealers called them for price information. Since different dealers worked with different brokers, none of the brokers knew the quotes of all dealers. The best bid and ask prices provided by any broker were just the best among the dealers working with him, but might not be the best in the whole market. Meanwhile, to obtain brokering services, the dealers had to pay commissions based on transaction quantity. In the customer market, customers traded only with banks (dealers).

Very little information was disclosed for the customers and even the dealers in the FX market before the 1990's. The information of any transaction was observed only by buyer and seller. In the inter-dealer market, trades that were made directly between two dealers were only observed by these two, and other dealers knew nothing about it. In the customer market, only the bank that received the order knew the information of the transaction.

In 1992, the electronic broker[1] system was introduced to the inter-dealer market. Since then, traditional voice brokers have been replaced by electronic ones, and most of the turnover is now conducted through electronic order-matching systems. The application of the electronic broker system has changed the market structure dramatically. Compared to the voice broker, the dealers have better and more efficient services at a lower cost by employing the new system. Instead of paying commissions to brokers for every transaction, the dealers only need to pay the rental fee for the instrument, which reduces dealers' operation cost. The real time trading information is displayed on the screen of the system, so that dealers can

[1] Two electronic broking systems currently operating in the U.S. are Electronic Brokerage Systems (EBS) and Reuters dealing system 2000-2.

receive the information much faster than before. Furthermore, every dealer has access to the system, and the system collects more information than any voice brokers before.

Although the information structure for customers and direct inter-dealer trading did not change a lot after introduction of electronic system, indirect inter-dealer trading is more transparent than before. From the screen of Reuters D2000-2 system—one of the most popular electronic dealing systems currently used, which is displayed in figure 1.1 in appendix II, the dealers can observe other dealers' trading information, which was impossible before.

In the late 1990s, the internet became available as a trading channel for customers. Customers used to trade only with dealers. The internet trading platform makes it possible that customers can trade with each other. The mechanism for the internet trading between customers is similar to the electronic inter-dealer trading system. Meanwhile, some websites appeared which can show several dealers' quotes simultaneously in response to one customers' request. Customers can use these websites to find the best quote for themselves while they could only contact dealers individually before. However, for some reason, internet customer trading is still not popular, and its trading volume only accounts for a very small part (less than 10%) of the turnover.

2.3 Differences between the stock market[1] and the FX market

To see the structure of the FX market more clearly, I compare the FX market and the stock market in this subsection. The major differences are listed below:

1) The FX market is a dealer market, where orders for execution pass to an intermediary (dealer) for execution. The stock market is an auction market, where buyers enter competitive

[1] The stock market is referring to NYSE not NASDAQ. The latter is more like dealership market not the traditional stock market.

bids and sellers enter competitive offers at the same time.

2) The FX market is a quote-driven market: the price is created by quoting bid and ask prices in response to trading initiators. The stock market is an order-driven market: where price is determined by an order execution algorithm which automatically matches buy and sell orders on a price and time priority basis.

3) The FX market is decentralized across several locations, as opposed to centralized on an exchange as is the case of the stock market. Introduction of telecommunications allows decentralized trade of the asset foreign exchange. Banks also want to be present where the customers are, and because the exchange rate is the relative price of two assets in two different countries, it is natural to have a decentralized market.

4) There is continuous trading around the clock in the FX market, as opposed to only when called upon as in a stock market (regular trading hour). Given that customers are in different time zones and may have an interest in the same asset, trading must be continuous around the clock.

5) The FX market is relatively opaque, i.e. has low transparency compared to the stock market. The decentralized structure also makes it very difficult to regulate foreign exchange trading. These factors, together with the lack of regulatory disclosure requirements, mean that the foreign exchange markets are characterized by low transparency.

All these differences have economic consequences: compared to a centralized call market, which is more transparent, in the FX market information aggregation will typically be slower and noisier. In the next section I begin to construct a model from the perspective of a dealer's quoting behavior to explain the determination of the spread and the exchange rate.

3. Theoretical framework of the model

3.1 Assumptions and timing of the model

Motivated by Biais (1993), Ho and Stoll (1983) and Evans and Lyons (2001), the model in this essay describes a decentralized dealership market, in which dealers and customers are the only market participants[1]. I assume that the number of dealers active in the market is M, and all those dealers have identical utility functions, with constant absolute risk aversion parameter A:

$$U(x) = -e^{-Ax}$$

The model focuses on one particular dealer (dealer *i*) and describes his quoting behavior in one period. Since the FX market is continuously trading 24/7, the period is not restricted to one day or any other particular interval of time and any two times can be randomly picked as the beginning and ending points of the period. Within the period, T number of orders are assumed to be sent to the dealer, and thus this period can be divided into T sub-periods according to the arrival of the orders so that only one order is included in each sub-period. Since orders arrive totally randomly, the lengths of these sub-periods are not identical. The timing and dynamics of the whole period and sub-period are described as below:

1. Initial status at the beginning of the whole period.

At the beginning of the whole period, dealer *i* is assumed to be endowed with cash C_i, and he enters the market at a fixed cost $\overline{F}_i$. $\overline{F}_i$ accounts for the cost of an administrative structure and the cost to be connected to an information network (such as Reuters). To provide liquidity, the dealer must hold some position for trading, but apparently, excessive position would increase the dealer's risk. Therefore, the dealer always tries to keep the position to

[1] Since electronic brokering system has replaced "voice brokers", the role of brokers is just ignored.

some particular level, which can guarantee his normal transaction and minimize risk meanwhile. We can call this position "ideal level of inventory" and normalize the dealer's ideal level to zero. Thus I assume that unconditionally the dealer's inventory level is uniformly distributed within [-R, R]. Since the starting point of the period is randomly picked, the inventory at this point is assumed to be $\varepsilon_{0,i}$, where $\varepsilon_{0,i}$ is a number drawn from the uniform distribution.

At the beginning of the period, before new information is released, dealer *i*'s current quotes, which can be denoted as $p_{0,i}^{ante}$, should be the quotes he made in the previous period. If the middle point of his quotes in the last period is denoted as $p_{-1,i}$, then $p_{0,i}^{ante} = p_{-1,i}$.

2. Public information releasing

Public information is assumed to be released at the beginning of the period[1] and observed by every agent in the market. The public information refers to the information available for all market participants, for example, central bank announcements. It systematically changes market participants' expectation for the exchange rate, and the dealer adjusts his quote immediately to incorporate the information. If the change in payoff of currency implied by public information is denoted as r, and then dealer's ex post quote after incorporating the public information becomes $p_{0,i}^{post} = p_{0,i}^{ante} + r$, i.e. $p_{0,i}^{post} = p_{-1,i} + r$.

3. The first order arrives and the first sub-period starts.

In this stage, the first order comes to the dealer, and the first sub-period starts. In the sub-period, the dealer first needs to quote in response to the transaction request, and then, after seeing the quotes, the transaction initiator will decide to decline or accept. This result will

[1] Actually, the timing of public information release does not have to be the beginning of the period. Whenever it is released during the period, or in any sub-period, the information will be incorporated into the quote immediately. This assumption just gives readers a clearer image about the timing of the period.

help the dealer to update his belief about the currency value and his inventory level. Thus one sub-period can also be divided into three steps as below:

1) Order arrives

One order comes to the dealer in this sub-period. As introduced before, the transactions executed by dealer *i* can only be observed by the dealer himself and transaction initiators in the FX market, hence become his private information. The orders the dealer receives convey very important information about the future exchange rate in the sense that market agents' transaction decisions imply their beliefs of the true value of the currency, i.e. they tend to buy the currency if they believe it will appreciate and sell if depreciate. At this step, since no transaction occurs before the first sub-period in the whole period, the dealer *i*'s current inventory level is the same as that at the beginning of the whole period.

2) The dealer makes quote

In this step, the dealer is supposed to quote in response to the trading request in a very short time (i.e. several seconds). The dealer needs to give both ask and bid prices and his quotes should maximize his expected utility conditioned on the current information set. This optimization problem and finding the solution to it become the core issue in this essay.

3) The dealer updates his belief and inventory

After seeing the quotes, the transaction initiator decides to decline or accept. This result affects the dealer's expectation of the currency value for the quotes made in the next sub-period. Thus the quotes made before the decision can be called ex ante quotes and denoted as $p_{1,i}^{ante}$, and the belief of currency value after the decision can be called ex post prices and denoted as $p_{1,i}^{post}$, which becomes the dealer's base quote at the beginning of the next sub-period. Meanwhile, the dealer's inventory level will also be updated after the result is revealed and ex ante quote made in the next sub-period will be based on this new inventory level. Denote ex ante and ex post inventory as $I_{1,i}^{ante}$ and $I_{1,i}^{post}$ respectively. According to the set-up, $I_{1,i}^{ante} = \varepsilon_{0,i}$.

4. Other orders come and other sub-periods start

Next, the second, the third, … the T[th] order will come to the dealer and steps 1)-3) in previous stage will be repeated in each later sub-period. Thus we can have an exchange rate (mid-quote) dynamics equation in each sub-period. In each sub-period, the ex ante and ex post quotes can be denoted as $p_{t,i}^{ante}$ and $p_{t,i}^{post}$ where t=1, 2….T, and similarly, ex ante and ex post inventory in the t[he] sub-period can be denoted as $I_{t,i}^{ante}$ and $I_{t,i}^{post}$. Apparently the following relationship should be satisfied: $I_{t,i}^{ante} = I_{t-1,i}^{post}$.

5. The end of whole period

When the last order is processed in the last sub-period (sub-period T), the whole period ends. Since we have an exchange rate dynamics equation in each sub-period, integrating the changes of mid-quote in each sub-period can give us the dynamics equation for the whole period.

3.2 Optimal quotes in one sub-period

Apparently, the optimization problem to be solved is the quoting problem in the sub-period, i.e. step 2) of stage 3 and 4 as described above. Following Biais (1993), the dealer's process of selecting optimal bid-ask quotes is laid out by two steps. First, the bottom line for the dealer to make a quotation is that at least he should not be worse off after revealing his quotes no matter if they are accepted or not. The quotes making him indifferent between trading or not trading are reservation quotes, which the dealer would take as a benchmark. Second, the dealer apparently has an incentive to raise the ask (selling) price above or reduce the bid (buying) price below the reservation quotes to earn extra profit, however, such a behavior will decrease the probability that the quotes are accepted. Therefore, there must be a point, which maximizes his expected utility and becomes the optimal choice.

According to this framework, reservation quotes are about to be derived first. At the time of deciding the quotes in the t[th] sub-period, dealer *i* is endowed with cash $C_{t,i}$, pays operation cost $\overline{F}_i$ (assumed to be the same across sub-periods), and holds inventory level $I_{t,i}^{ante}$. Random variable z_t is the difference between current and terminal market rate in this sub-period, and therefore the terminal currency price will be $p_{t-1,i}^{post} + z_t$ at the end of this sub-period. (Ex post quote in the previous sub-period $p_{t-1,i}^{post}$ becomes the base quote at the beginning of this sub-period).

Now, a new market order comes: some agent contact the dealer *i* with potential order size of $Q_{t,i}$ without revealing the potential transaction direction (buy or sell)[1]. Superscript *t* refers to sub-period t and superscript *i* means that this order is dealer *i*'s private information. In response to the trading request, the dealer *i* quotes both ask and bid prices. Let's denote them as $p_{t-1,i}^{post} + a_{t,i}$ (ask price) and $p_{t-1,i}^{post} - b_{t,i}$ (bid price) respectively[2]. If the dealer *i*'s quotes are not accepted, this transaction does not take place, and the dealer *i*'s anticipated wealth at the end of the sub-period will be:

$$W(0) = C_{t,i} - \overline{F}_i + I_{t,i}^{ante}(p_{t-1,i}^{post} + z_t) \tag{1.1}$$

[1] It should be noted that essay 1 and 2 have different assumptions with regard to this issue. In essay 1, dealers are assumed not to know transaction directions when they quotes, which is the case when data used in essay 1 were collected. While the dealers know the direction in essay 2, which is also the case when data used in essay 2 were collected. This arrangement is made to be consistent with the data used in different essays.

[2] The ask and bid prices are designed this way just to make the prices symmetric so that we can determine the ask price or bid price alone and get the other easily.

If his quotes are accepted, dealer *i* sells $Q_{t,i}$ at the ask price, and his anticipated wealth at the end of the sub-period after the transaction becomes:

$$W(a_{t,i}) = C_{t,i} - \overline{F}_i + (I_{t,i}^{ante} - Q_{t,i})(p_{t-1,i}^{post} + z_t) + (p_{t-1,i}^{post} + a_{t,i})Q_{t,i} \tag{1.2}$$

If his quotes are accepted, dealer *i* buys $Q_{t,i}$ at the bid price, and his anticipated wealth at the end of the sub-period after the transaction becomes:

$$W(b_{t,i}) = C_{t,i} - \overline{F}_i + (I_{t,i}^{ante} + Q_{t,i})(p_{t-1,i}^{post} + z_t) - (p_{t-1,i}^{post} - b_{t,i})Q_{t,i} \tag{1.3}$$

Suppose the dealer *i*'s reservation quotes are denoted as $p_{t-1,i}^{post} + a_{t,i}^r$ and $p_{t-1,i}^{post} - b_{t,i}^r$, and those quotes should make three quoting results (not trading, buying and selling $Q_{t,i}$) indifferent to the dealer. Accordingly, reservation quotes will let the dealer's conditional expected utility of three cases equal to each other, i.e. equation (1.4) holds:

$$E(U(W(0)) \mid S_{t,i}) = E(U(W(a_{t,i}^r)) \mid S_{t,i}) = E(U(W(b_{t,i}^r)) \mid S_{t,i}) \tag{1.4}$$

where $S_{t,i}$ is the information set dealer *i* receives. The solutions to equation (1.4) are given in the following lemma, and the derivation process is given in appendix III.

Lemma 1.1: *Under our set of assumptions, the reservation selling and buying price in the FX markets for dealer i in sub-period t are* $p_{t-1,i}^{post} + a_{t,i}^r$ and $p_{t-1,i}^{post} - b_{t,i}^r$ *, where*

$$a_{t,i}^r = \frac{A\sigma^2}{2}(Q_{t,i} - 2I_{t,i}^{ante}) + E(z_t \mid S_{t,i})$$

$$b_{t,i}^r = \frac{A\sigma^2}{2}(Q_{t,i} + 2I_{t,i}^{ante}) - E(z_t \mid S_{t,i})$$

In lemma 1.1 and throughout the essay, σ^2 denotes the conditional variance of exchange rate. Strictly speaking, it should be updated based on the information set as well. However, in a very short period such that only one order is included, volatility update usually is not as frequent as exchange rate update.[1] I assume that the dealer takes constant conditional volatility in this essay, which will greatly simplify our later analysis.

Apparently, the dealer would not be satisfied with the reservation quotes. Next, the dealer's task is to maximize his expected utility by choosing the optimal bid-ask prices. The dealer's expected utility depends on the utility of each possible state and its corresponding probability. For each trade request[2], there are two possible states dealer *i* might end up with: trading or not trading. In the case of not trading, the utility will be $E(U(W_i(0)) | S_{t,i})$, while $E(U(W_i(a_{t,i})) | S_{t,i})$ is the utility in the case of trading. The reason that I use expected utility even for each specific state is because the dealer's anticipated change in exchange rate z_t is still a random variable and will not be realized until at the end of the sub-period. Let $\pi_{a,i}$ denote the probability that dealer *i*'s quote is accepted, then as usual, the total expected utility for dealer *i* is:

$$EU_{a,i} = \pi_{a,i} E(U(W(a_{t,i})) | S_{t,i}) + (1 - \pi_{a,i}) E(U(W(0)) | S_{t,i}) \tag{1.5}$$

The next question is how to determine the probability that dealer *i*'s quote is accepted ($\pi_{a,i}$). In the electronic inter-dealer dealing system (such as Reuters D2000-1 system), the real time information of the best bid-ask price is displayed on the screen, and every dealer can observe it easily and almost costlessly. If the dealer *i* is chosen to be bought from, it actually means his ask price is lower than any other competitor's. Accordingly, the probability of his price being accepted by is equivalent to the probability that his ask price is

[1] Actually, the GARCH estimation in essay 2 shows that in a short horizon such as 1-minute, the exchange rate volatility basically depends on a constant item, which suggests that the volatility is pretty stable in very short horizons. For more details about the predictability of the exchange rate volatility in short horizons, see Andersen and Bollerslev (1998).

[2] Here the request is referring to the buying request. Since the problem is symmetric, to simplify the process of solving the model, we just analyze the case of the buying request.

the lowest among all the dealers, which can be expressed as $\pi_{a,i} = P(a_{t,i} < a_{t,-i})$ where subscript $-i$ represents other dealers.

It is claimed that to all dealers the optimal quoting strategy should be decreasing in each dealer's inventory level. This claim results from the inventory effect. Since I normalize the ideal inventory level to zero, any inventory change caused by trading with other agents is actually a deviation from the preferred level. If the dealer's current inventory is higher than the preferred level, which suggests that the dealer holds more position than he expects, naturally, he wants to sell more and/or buy less to reduce the inventory level. To achieve this goal, one efficient way is to quote a lower price. Similarly, if the current inventory is lower than the preferred level, the dealer intends to increase the price in order to sell less and/or buy more, so that the inventory will return to the ideal level.

Since the dealer's optimal quoting strategy is decreasing in the inventory, the probability that dealer i's ask price is lower than others is equivalent to the probability that his inventory is larger than others, i.e. $\pi_{a,i} = P(a_{t,i} < a_{t,-i}) = P(I_{t,i}^{ante} > I_{t,-i}^{ante})$. From the perspective of the dealer, he certainly does not know other dealers' inventory distributions, and a reasonable assumption for him to make is that others have the same distribution as his. In addition, I assume these distributions are independent. If his distribution function is denoted as F_j , using these two assumptions gives the probability as below:

$$\pi_{a,i} = P(a_{t,i} < a_{t,i}) = P(I_{t,i}^{ante} > I_{t,-i}^{ante}) = \prod_{j \neq i}^{M} F_j(I_{t,i}^{ante}) = G(I_{t,i}^{ante}) \tag{1.6}$$

where $G(I_{t,i}^{ante})$ is just another notation of $\pi_{a,i}$. By adjusting equation (1.5) we can get equation (1.7)

$$EU = \pi_{a,i}[E(U(W(a_{t,i})) \mid S_{t,i}) - E(U(W(0)) \mid S_{t,i})] + E(U(W(0)) \mid S_{t,i}) \tag{1.7}$$

According to the definition of the reservation quote, we have:

$$E(U(W(a_{t,i}))\mid S_{t,i}) = E(U(W(0))\mid S_{t,i})e^{-A(a_{t,i}-a_{t,i}^r)Q_{t,i}} \tag{1.8}$$

Substituting equation (1.8) to equation (1.7) gives equation (1.9):

$$EU = \pi_{a,i}(e^{-A(a_{t,i}-a_{t,i}^r)Q_{t,i}} - 1)E(U(W(0))\mid S_{t,i}) + E(U(W(0))\mid S_{t,i}) \tag{1.9}$$

Since $E(U(W(0))\mid S_{t,i})$ does not depend on the choice of the ask price, it can be neglected from the objective function, and the aim of the dealer will be simplified as:

$$\text{Max } \pi_{a,i}(e^{-A(a_{t,i}-a_{t,i}^r)Q_{t,i}} - 1)$$

Using a Taylor expansion and neglecting the terms of second order derivatives, the objective function is:

$$\text{Max } \pi_{a,i}(A(a_{t,i} - a_{t,i}^r)Q_{t,i})$$

Following Biais (1993)'s general proof, the problem is solved, and the solutions are illustrated in the following proposition 1.1. The proof is given in appendix III.

Proposition 1.1: *Under our set of assumptions, the solutions to the dealer's problem and the optimal bid and ask prices of dealer i in the FX market are* $(p_{t-1,i}^{post} - b_{t,i})$ *and* $(p_{t-1,i}^{post} + a_{t,i})$, *where*

$$a_{t,i} = a_{t,i}^r + A\sigma^2 \frac{R + I_{t,i}^{ante}}{M}$$

$$b_{t,i} = b_{t,i}^r + A\sigma^2 \frac{R - I_{t,i}^{ante}}{M}$$

Since our major concern is the exchange rate dynamics, which is the middle point of the bid-ask prices, the following equation gives ex ante exchange rate:

$$p_{t,i}^{ante} = \frac{P_{t-1,i}^{post} + a_{t,i} + (P_{t-1,i}^{post} - b_{t,i})}{2} = P_{t-1,i}^{post} - A\sigma^2(1 - \frac{1}{M})I_{t,i}^{ante} + E(z_t \mid S_{t,i}) \tag{1.10}$$

Apparently, the explicit form of ex ante quote in this sub-period depends on his ex ante expectation about exchange rate change, i.e. $E(z_t \mid S_{t,i})$. It should be noted that all the information arriving before the sub-period has been incorporated into his ex post quote in the previous sub-periods, and in this sub-period the only new information so far is a quote request. Although the dealer receives this request, he does not know the direction of potential transaction, so a reasonable assumption is that either direction is equally likely. Thus, this request actually does not tell the dealer anything useful about the future exchange rate, and expectation of future exchange rate change can be simply written as zero, i.e. $E(z_t \mid S_{t,i}) = 0$. Thus the ex ante quote in sub-period t can be written as:

$$p_{t,i}^{ante} = P_{t-1,i}^{post} - A\sigma^2(1 - \frac{1}{M})I_{t,i}^{ante}$$

After seeing the dealer's quotes, the transaction initiator makes his decision, and the dealer will know the result. If the quotes are declined, no transaction happens and the dealer still does not know whether it was a buy or sell order in the first place. In this case, no new information should be incorporated when the dealer updates his belief, and ex post quote is the same as the ex ante quote so that:

$$p_{t,i}^{post} = P_{t-1,i}^{post} - A\sigma^2(1 - \frac{1}{M})I_{t,i}^{ante}$$

Meanwhile, since no transaction happens, inventory does not change either so that:

$$I_{t,i}^{post} = I_{t,i}^{ante}$$

If the quotes are accepted, then the transaction is executed and the dealer definitely knows the direction of the transaction, from which he can extract useful information and map it to his expectations. Intuitively, the transaction initiator tending to buy $Q_{t,i}$ units of the currency implies that in his opinion the market exchange rate in the future must be higher than the current one. Also, the order size $Q_{t,i}$ should be proportional to the $E(z_t)$. Through a linear signal extraction process, the dealer factors information set $S_{t,i}$ into his own expectation of z_t as follows:

$S_{t,i}$ = {some agent buys or sells $Q_{t,i}$ units of the currency} and $E(z_t \mid S_{t,i}) = \beta_i \cdot Q_{t,i}$

where β_i is a dealer-specific constant coefficient mapping the information of the order into the dealer's expectation. Order size $Q_{t,i}$ here is assigned with a positive sign when it is a buy order and a negative sign when it is a sell order. Thus it applies to both cases of buying and selling so that one general equation would be enough to describe the ex post quote, which can be written as:

$$p_{t,i}^{post} = P_{t-1,i}^{post} - A\sigma^2(1-\frac{1}{M})I_{t,i}^{ante} + \beta_i Q_{t,i}$$

and ex post inventory becomes:

$$I_{t,i}^{post} = I_{t,i}^{ante} - Q_{t,i}$$

Using a transaction indicator D_t, which equals 1 when quotes are accepted and 0 when quotes are declined, a general way to express this update process in both cases of trading and not trading is:

$$p_{t,i}^{post} = P_{t-1,i}^{post} - A\sigma^2(1-\frac{1}{M})I_{t,i}^{ante} + \beta_i D_t Q_{t,i}$$

$$I_{t,i}^{post} = I_{t,i}^{ante} - D_t Q_{t,i}$$

Since

$$I_{t+1,i}^{ante} = I_{t,i}^{post}$$

therefore

$$I_{t+1,i}^{ante} = I_{t,i}^{ante} - D_t Q_{t,i}$$

Hence, the mid-quotes and inventory dynamics equations in this sub-period can be written as:

$$p_{t,i}^{post} = P_{t-1,i}^{post} - A\sigma^2(1-\frac{1}{M})I_{t,i}^{ante} + \beta_i D_t Q_{t,i}$$

$$I_{t+1,i}^{ante} = I_{t,i}^{post} = I_{t,i}^{ante} - D_t Q_{t,i}$$

3.3 Order flow and exchange rate dynamics

Now, let's jump out of the sub-periods and take a look at the whole period. The exchange rate change in the horizon of this whole period can be expressed as $p_{T,i}^{post} - P_{0,i}^{post}$. Based on the exchange rate and inventory dynamics obtained before, $p_{T,i}^{post} - P_{0,i}^{post}$ can be calculated as:

$$p_{T,i}^{post} - p_{0,i}^{post} = \sum_{t=1}^{T}(p_{t,i}^{post} - P_{t-1,i}^{post}) = \sum_{t=1}^{T}[-A\sigma^2(1-\frac{1}{M})I_{t,i}^{ante} + \beta_i D_t Q_{t,i}] \tag{1.11}$$

$$\text{where } I_{t+1,i}^{ante} = I_{t,i}^{ante} - D_t Q_{t,i} \tag{1.12}$$

Plugging the inventory updating equation (1.12) into equation (1.11) gets:

$$p_{T,i}^{post} - P_{0,i}^{post} = -A\sigma^2(1-\frac{1}{M})[T\varepsilon_{0,i} - \sum_{t=1}^{T}\sum_{j=1}^{t}(D_j Q_{j,i})] + \beta_i \sum_{t=1}^{T} D_t Q_{t,i}$$

It can also be written as:

$$p_{T,i}^{post} - P_{0,i}^{post} = A\sigma^2(1-\frac{1}{M})[\sum_{t=1}^{T}(D_t Q_{t,i}) + \sum_{t=1}^{T-1}\sum_{j=1}^{t} D_j Q_{j,i} - T\varepsilon_{0,i}] + \beta_i \sum_{t=1}^{T} D_t Q_{t,i} \tag{1.13}$$

According to its definition, order flow is the summation of signed transaction volume. Given the definition of the transaction indicator D_t and signed order size $Q_{t,t}$, $\sum_{t=1}^{T} D_t Q_{t,i}$ is exactly equal to the order flow cumulated during the whole period. If the order flow cumulated between the beginning of the whole period and sub-period t for dealer *i* is denoted as $O_{0-t,i}$, equation (1.13) can be rewritten as:

$$p_{T,i}^{post} - P_{0,i}^{post} = A\sigma^2(1-\frac{1}{M})[O_{0-T,i} + \sum_{t=1}^{T-1} O_{0-t,i} - T\varepsilon_{0,i}] + \beta_i O_{0-T,i} \tag{1.14}$$

The first item of equation (1.14) reflects the inventory effect of the order flow. If the dealer receives a negative order flow, i.e. the volume of buy orders is less than that of sell orders, his inventory builds up, which leads to higher inventory holding risk. Therefore, in order to reduce inventory, intuitively, the dealer needs to reduce both bid-ask prices to encourage buy orders and deter sell orders, which is consistently with suggestions given by equation (1.14). The second item of equation (1.14) ($\beta_i O_{0-T,i}$) reflects the information effect of the order flow. Positive order flow, i.e. the volume of buy orders is higher than that of sell orders, suggests that market participants expect the future exchange rate to appreciate. To reduce the potential loss caused by adverse selection, intuitively, the dealer needs to raise both bid-ask prices to deter buy orders and encourage sell orders, which is also supported by equation (1.14).

Finally, substituting $p_{0,i}^{post} = p_{0,i}^{ante} + r$ into equation (1.14) gives:

$$p_{T,i}^{post} = p_{0,i}^{ante} + r + [A\sigma^2(1-\frac{1}{M}) + \beta_i] \cdot O_{0-T,i} + [A\sigma^2(1-\frac{1}{M})](\sum_{t=1}^{T-1} O_{0-t,i} - T\varepsilon_{0,i}) \tag{1.15}$$

Equation (1.15) can also be shown in a simplified way as equation (1.16):

$$p_{T,i}^{post} = p_{0,i}^{ante} + r + (\alpha + \beta_i) \cdot O_{0-T,i} + \alpha \cdot (\sum_{t=1}^{T-1} O_{0-t,i} - T\varepsilon_{0,i}) \tag{1.16}$$

$$where\ \alpha = A\sigma^2(1-\frac{1}{M})$$

Equations (1.15) and (1.16) imply several important theoretical conclusions about the exchange rate dynamics. First, the order flow has a direct positive effect on the path of the exchange rate dynamics. Thus this causality relationship brought up first by Evans and Lyons (2002) is verified by my alternative model which relaxes several unrealistic assumptions and therefore is more practical than the pioneering models. Second, equation (1.16) demonstrates in a very understandable format that the exchange rate updating process is based on the public (r) and private information (order flow O). Third, since the private information each individual dealer receives and his initial inventory draw are different across dealers, various dealers' quotes are not identical. There must exist a distribution to describe the quotes across dealers, and apparently, this distribution depends on the distribution of the order flow and initial inventory draw across dealers. If the inventory draw continues to follow a uniform distribution as assumed before and the order flow is assumed to be normally distributed, then the quotes across dealers would follow an unknown distribution combining a normal distribution and a uniform distribution.

4. Empirical evidence

The data set[1] for the testing contains the daily measures of actual transactions for six spot markets over a four-month period, May 1 to August 31, 1996: Mark, Yen, Pound, French Franc, Krona, and Lira; all verses the Dollar. The data were collected from the Reuters Dealing 2000-1 system via an electronic feed customized for the purpose. Dealing 2000-1 was the most widely used electronic dealing system during the sample period. According to Reuters, over 90 percent of the world's direct inter-dealer transactions took place through the system[2]. All trades on this system take the form of bilateral electronic conversations. The conversation is initiated when a dealer uses the system to call another dealer to request a quote. Users are expected to provide a fast two-way quote with a tight spread, which is in turn dealt or declined quickly (i.e., within seconds). To settle disputes, Reuters keeps a

1 This data set is provided by Martin Evans, and the data were first used in his paper: Informational Integration and FX Trading, *The Journal of International Money and Finance*, 2002.

2 In 1996, inter-dealer transactions accounted for about two-thirds of total trading in major spot markets. This two-thirds from inter-dealer trading breaks into two transaction types: direct and brokered. Direct trading accounted for about half of inter-dealer trade and brokered trading accounted for the other half.

temporary record of all bilateral conversations, and this record is the source of the data. For these trades executed on D2000–1 the data set includes a bought/sold indicator that allows us to measure signed volume, i.e., order flow, directly. One drawback is that it is not possible to identify the size of individual transactions. Thus, the order flow variable in the data set is therefore measured as the difference between the number of buyer-initiated and seller-initiated trades during the 24 hours before 4:00 pm London time.[1] The daily exchange rates are the prices of the last transaction before 4pm London time each day.

Apparently, there are several inconsistencies between the data and the variables required to test equation (1.16). First, the order flow in equation (1.16) is total order flow including both customer and inter-dealer order flows, while the order flow in the data is only the inter-dealer order flow (Reuters Dealing 2000-1 is an inter-dealer trading system). Since the customer and inter-dealer order flow are supposed to have a consistent effect on the exchange rate dynamics, using only inter-dealer order flow would not generate fundamentally different results than the case using total order flow. Hence this drawback of the data would not undermine the efficiency of testing.

Another more critical inconsistency is that equation (1.16) describes an individual dealer's quoting behavior, which is linked to his own order flow, while the data are aggregate order flow across dealers. This gap can be bridged by the following arrangements.

The order flows received by various dealers within this period are different and must follow some distribution. It is reasonable to assume it to be a normal distribution so that each dealer's order flow is normally distributed around its mean value. And the mean value of all order flows can be approximated by its numerical average as below:

[1] Strictly speaking, order flow is summation of signed trading volumes, and the difference between the number of buyer-initiated and seller-initiated trades is not order flow. However, considering that orders in the inter-dealer trading often take standard sizes, it should be o.k. to take this measurement as the proxy of order flow.

$$\frac{\sum_{i=1}^{M} O_{0-T,i}}{M} = \overline{O}_{0-T}$$

where $\overline{O}_{0-T}$ is the average of all order flows. Thus, one particular dealer's order flow can be expressed as a random deviation ($\eta_{0-T,i}$) from the mean value of all order flows ($\overline{O}_{0-T}$), such that:

$$O_{0-T,i} = \overline{O}_{0-T} + \eta_{0-T,i}$$

where error item $\eta_{0-T,i}$ follows a standard normal distribution. Plugging the equation above into equation (1.16) gets:

$$p_{T,i}^{post} = p_{0,i}^{ante} + r + (\alpha + \beta_i) \cdot (\overline{O}_{0-T} + \eta_{0-T,i}) + \alpha \cdot (\sum_{t=1}^{T-1} O_{0-t,i} - T\varepsilon_{0,i}) \tag{1.17}$$

The equation above gives the dynamics of dealer *i*'s quotes within the whole period, while the daily exchange rates contained in the data set could be quoted by different dealers and should be the lowest rates at the end of each day. According to the rule of thumb, the lowest price is likely to deviate from the mean value of the quotes (which can be approximated by average quote) by three standard deviation of the quote distribution. Since the quote distribution across dealers depends on the distributions of the order flow and initial inventory draw, which are assumed to be a normal and a uniform distribution respectively across the periods, the standard deviation of the quote distribution should be stable across periods. Thus the dynamics of the lowest quote at the end of each day can be approximated by the dynamics of the average quote at that time. According to equation (1.17), the average quote can be derived as below:

$$\frac{\sum_{i=1}^{M} p_{T,i}^{post}}{M} = \frac{\sum_{i=1}^{M} p_{0,i}^{ante}}{M} + r + \overline{O}_{0-T}\frac{\sum_{i=1}^{M}(\alpha + \beta_i)}{M} + \frac{\sum_{i=1}^{M}(\alpha + \beta_i)\eta_{0-T,i}}{M} + \alpha\frac{\sum_{i=1}^{M}\sum_{t=1}^{T-1} O_{0-t,i}}{M} - \frac{\sum_{i=1}^{M}(\alpha\ T \cdot \varepsilon_{0,i})}{M}$$

If the average quotes at the beginning and the end of the period are denoted as $\overline{p}_0, \overline{p}_T$ respectively, we have:

$$\overline{p}_T = \overline{p}_0 + r + (\alpha + \frac{\sum_{i=1}^{M}\beta_i}{M}) \cdot \overline{O}_{0-T} + \alpha \cdot \sum_{t=1}^{T-1}\overline{O}_{0-t} + \alpha \cdot \frac{\sum_{i=1}^{M}\eta_{0-T,i}}{M} + \frac{\sum_{i=1}^{M}\beta_i \cdot \eta_{0-T,i}}{M} - \alpha T \frac{\sum_{i=1}^{M}\varepsilon_{0,i}}{M}$$

Putting $\frac{\sum_{i=1}^{M} O_{0-T,i}}{M} = \overline{O}_{0-T}$ back into the equation above gets:

$$\overline{p}_T = \overline{p}_0 + r + (\frac{\alpha}{M} + \frac{\sum_{i=1}^{M}\beta_i}{M^2}) \cdot (\sum_{i=1}^{M} O_{0-T,i}) + \alpha \cdot \sum_{t=1}^{T-1}\overline{O}_{0-t} + \alpha \cdot \frac{\sum_{i=1}^{M}\eta_{0-T,i}}{M} + \frac{\sum_{i=1}^{M}\beta_i \cdot \eta_{0-T,i}}{M} - \alpha T \frac{\sum_{i=1}^{M}\varepsilon_{0,i}}{M}$$

and finally we have:

$$\overline{p}_T - \overline{p}_0 = r + (\frac{\alpha}{M} + \frac{\sum_{i=1}^{M}\beta_i}{M^2}) \cdot (\sum_{i=1}^{M} O_{0-T,i}) + \alpha \cdot \sum_{t=1}^{T-1}\overline{O}_{0-t} + \alpha \cdot \frac{\sum_{i=1}^{M}\eta_{0-T,i}}{M} + \frac{\sum_{i=1}^{M}\beta_i \cdot \eta_{0-T,i}}{M} - \alpha T \frac{\sum_{i=1}^{M}\varepsilon_{0,i}}{M} \quad (1.18)$$

Equation (1.18) becomes the market aggregate version of equation (1.17), and the order flow in this equation is aggregate order flow across dealers. Last three items of equation (1.18) are random items with zero mean. (As noted before, $\eta_{0-T,i}$ is a random deviation from the mean value of order flow and follows standard normal distribution. $\varepsilon_{0,i}$ is the initial inventory draw and follows uniform distribution around zero). Therefore, the data are suitable to test the model based on equation (1.18).

Since α contains the exchange rate volatility, the model is tested by a GARCH-in-mean model, which can be specified as below:

$$\log p_t - \log p_{t-1} = \gamma_0 + \gamma_1 \sigma_t^2 + \gamma_2 O_t + \varepsilon_t \quad (1.19)$$

$$\sigma_t^2 = Var(\varepsilon_t) \quad (1.20)$$

$$\sigma_t^2 = \kappa_0 + \kappa_1 \sigma_{t-1}^2 + \kappa_2 \varepsilon_{t-1}^2 + \eta_t \quad (1.21)$$

where $\gamma_0, \gamma_1, \gamma_2$ $\kappa_0, \kappa_1, \kappa_2$ are constant coefficients, and η_t is a gaussian process. σ_t^2, O_t denote the exchange rate volatility and the order flow respectively. It would be complete that both customer and inter-dealer order flows are included in the model, however, due to the lack of customer order flow data, only inter-dealer order flow variable is applied in the empirical testing. As suggested in many earlier works such as Bollerslev & Melvin (1994) as well as Hartmann (1998), GARCH(1,1) specification is proper to describe the variance part. To make results comparable across currencies, I use the change of log spot rate instead of the spot rate return itself. In addition, since the change of log spot rate is in a much smaller magnitude than the variable of order size, the right side of the equation is multiplied by 10000 to estimate the model.

Table 1.1 in appendix I documents the estimation results of equations (1.19) through (1.21) for each currency. As predicted by the theoretical model, the estimated coefficient for order flow (γ_2) is found to be positive across currencies, and large t-statistics for those estimates suggest that the results are highly significant and very robust. Clearly, the order flow has a very significant positive effect on the dynamics of the exchange rate, which provides a strong supportive evidence for our theory. On the contrary, the coefficient of GARCH-in-mean item (γ_1) is not significantly different than zero due to their low t-statistics, which means that the exchange rate volatility does not have a significant impact on the exchange rate dynamics in our estimation. In the volatility part, all tested currencies except for Lira generate a very significant constant item (κ_0), while ignorable ARCH coefficient (κ_1) and GARCH coefficient, which implies that the error item for these currencies in our model (equation (1.19)) does not possess the features of autocorrelation and heteroscedacity. On the contrary, the Lira's GARCH coefficient is significantly positive while constant item is ignorable, which displays a significant heteroscedacity for the Lira's error item.

It should be noted that the European currencies data used in the testing are affected by the European Exchange Rate Mechanism (or ERM), which links European currency rates and allow them to fluctuate within a narrow margin of 2.25% on either side of the bilateral rates with the exception of the Italian Lira[1]. The linkage between them should make those currencies have generally consistent results, while the special arrangements made for the Lira might cause some inconsistency with other European currencies. The testing results displayed in table 1.1 do support such expectations. Generally, order flows have a significant effect on the exchange rate dynamics for all European currencies. Meanwhile, in contrast to the insignificance of the volatility part for other currencies, the Lira's GARCH item is highly significant, which can be explained by the special arrangement made for Lira in ERM.

Figures 1.2 through 1.7 in appendix II also illustrate the dynamics of the exchange rate and the aggregate order flow for each of the six currencies. As shown in these figures, the exchange rate and the order flow generally have the same direction of change each day for every currency, and the magnitude of the change in the exchange rate is much determined by the order flow. Graphically, it shows obviously that the order flow has a significant impact on the change of exchange rate.

5. Conclusions

This essay constructs a model to explain the exchange rate formation from the perspective of dealer's quoting behavior in the environment of the FX market. The dealer's

[1]Lira dropped out of the ERM in 1992 and rejoined it at the end of 1997 so that Lira data used in the testing were not subject to ERM.

problem in the FX market is described as choosing optimal quotes, which need to be both bid and ask price, to maximize his expected utility, conditioned on the information set he receives. The middle value of the bid-ask price is taken as the observation of exchange rate in the market.

Agreeing with the results of the existing literature, my model shows that the order flow has a significant effect on the short-run exchange rate dynamics in a more realistic and practical environment. The empirical testing in the essay also provides strong support for the theoretical conclusions found in the essay.

For the future, there are lots of further work can be done. First, one question in many people's minds is what drives order flow, and we don't see any model explaining the formation of order flow itself. Second, so far, this type of model can easily solve some of the well-known international monetary puzzles (like disconnect puzzle, volatility puzzle), but no model has been seen to solve the forward premium puzzle. Third, we can analyze the traditional macro topics like central bank intervention from the perspective of market microstructure too. Finally, of course, the emerging new data can let us do more accurate tests and estimation.

ESSAY TWO

MARKET STRUCTURE AND DEALER'S QUOTING BEHAVIOR IN THE FOREIGN EXCHANGE MARKET

1. Introduction

The foreign exchange market (FX market) is a decentralized dealership market, where market orders pass to an intermediary (dealer) for execution, and the price of currency is created by quoting bid and ask prices in response to trading initiators. Based on different initiators, the FX market can be divided into an inter-dealer (inter-bank) market in which dealers (banks) trade with each other, and a customer market in which customers trade with dealers. Accordingly, each dealer actually quotes two pairs of bid-ask prices in practice: one for inter-dealer trading and the other for customers. The questions arising are: does the dealer quote the same bid-ask prices in the two markets at any given point in time? Furthermore, are his spreads and mid-quotes the same in the two markets? If not, why are they different? How different are they? And, is there any empirical evidence to show the difference?

These questions are important in both academic and practical senses. Since the customer and inter-dealer FX markets are characterized by different structures, the dealer's quoting decision in both markets is a perfect case to examine the impact of market microstructure on the market participant's behavior. Moreover, since it is about the same dealer's simultaneous quotes in the two markets, results will not be jeopardized by the change of the agent's preference over time or the heterogeneity of various dealers. Practically, knowing dealers' quoting mechanisms is helpful for central banks to improve the efficiency of intervention and regulation[1]. And also, it will help dealers make a reliable prediction and evaluation on the effect of introducing new trading platforms or services to the FX market.

To the questions noted above, unfortunately, the current literature has not yet provided satisfying answers. On the theoretical level, many models have been devoted to the problem of bid-ask quote determination, but most of them are built originally for the traditional stock markets[2], which operate differently than the FX market. NYSE, for instance, is a centralized auction market, where buyers enter competitive bids and sellers enter competitive offers at the same time, and the price of stock is determined by an order execution algorithm automatically matching buy and sell orders on a price and time priority basis. Unlike the FX market, the stock markets do not have separate venues for inter-dealer and customer trading, so it is understandable that these models do not make such a differentiation, e.g. Ho and Stoll (1981), Biais (1993), Laux (1995), etc. On the other hand, there do exist some models built specifically for the FX market in the literature. However, these models either concentrate on one market alone, e.g. Black (1991) (customer) and Bessembinder (1994) (inter-dealer), or just ignore the distinctions between the two markets and assume dealers quote only one pair of bid-ask prices, e.g. Bollerslev and Melvin (1994).

On the empirical level, a lot of data sets have been used to examine the bid-ask price determinants in the FX market. Glassman (1987) uses daily quotes of a single Chicago futures dealer. Bollerslev and Domowitz (1993), Demos and Goodhart (1996), Huang and Masulis (1999) as well as Hartmann (1999) have access to continuous Reuters quotes.

[1] On the usefulness of the information approach to the Central Bank Intervention, see Vitale (1999).

[2] The stock markets in this essay refer to NYSE and AMEX, which are centralized auction markets. Another security market NASDAQ, on the contrary, is more like a decentralized dealership market.

Bessembinder (1994) and Jorion (1996) use daily data from the DRI data bank, which are Reuters quotes of a "representative" dealing bank at some time during the day. Disregarding the variation in the sources, all of these data are only quotes for inter-dealer trading, and the differences between inter-dealer and customer quotes have never been examined empirically before.

In order to answer the questions mentioned above, effort is made both theoretically and empirically in this essay. Following Biais (1993), a model is built viewing the dealer as a risk averse agent who chooses bid-ask prices to optimize his portfolio based on the information he receives. In contrast to the preceding work focusing on the stock market, the setup and assumptions of the model are made compatible with the environment and structure of the FX market. Moreover, the distinctions between the customer market and the inter-dealer market will be identified and incorporated into the model. The dealer's quoting problems are solved in both markets, and the solutions of optimal bid-ask quotes lead to the explicit form of spreads and mid-quotes, so that comparison can be made. Empirically, a new data set has been collected, which contains both customer and inter-dealer bid-ask quotes covering the same period and picked at the same frequency, to test several theoretical conclusions implied by our model.

This essay makes contributions to the literature in two major aspects. Theoretically, noticing that the structure of the customer and inter-dealer foreign exchange markets are different, the essay identifies those differences and incorporates them into a formal model. As the solutions to the dealer's quoting problem, the optimal bid-ask quotes in both markets show clearly how different they are as well as why. Empirically, by employing a data set characterized by several new features, some hypotheses that have never been examined by real data before have been tested in this essay. To my knowledge, the essay makes an empirical comparison between the inter-dealer and customer quotes for the first time in the literature. And also, our research finds evidence that contradicts the existing claim that people used to believe with regard to the relationship between the spread and the order size.

The rest of the essay is structured as follows: section 2 presents a model describing the dealer's quoting problem in the inter-dealer and customer FX markets respectively, and shows the explicit form of the optimal bid-ask prices. Section 3 introduces the new data set I collected and illustrates the empirical evidence of the model. Finally, section 4 concludes.

2. Theoretical framework of model

2.1 Setup and assumption

The model describes a decentralized dealership market, in which dealers and customers are the only market participants[1]. The FX market is divided into two sections: dealers trade with each other in the inter-dealer market, and customers trade only with dealers in the customer market.

I assume N dealers are active in the market, and each dealer's preference is represented by the regular utility function as below:

$$U(W) = -\exp(-A \cdot W)$$

where A is the degree of risk aversion and W represents wealth. To stay in the business, dealer i pays the fixed cost $\overline{C_i}$, which accounts for the operation expenses independent of the number of transactions, e.g. the fee of connection to the inter-dealer dealing system. Meanwhile, the rising scale of business increases operating cost, too. For instance, more staff

[1] Since electronic brokering system has replaced "voice brokers", we just ignore the role of brokers.

needs to be hired to handle an increasing number of transaction requests. Suppose the dealer assumes extra cost ω for handling one more market order, and this cost is assumed to be fixed relative to the size of the order.

As the provider of liquidity[1], the dealer must hold inventory to facilitate transactions. However, an excessive position leads to higher risk, so the dealer always tries to keep the position to some particular level, which guarantees normal transactions and minimizes risk. This particular level is called as the ideal level of inventory, and I just assume it is constant across periods and normalize it to zero for simplicity.

Our dynamic model focuses on one price adjustment period, in which the real interest rate is normalized to 0. The details about the timing of the model are as follows:

1) Initial status at the beginning of the period

At the beginning of the period, the dealer is endowed with cash K_i and holds inventory level I_i. For each particular dealer, he definitely knows his own inventory level, whereas inventory levels across the dealers are publicly unobservable and considered as random variables, which are characterized by the distribution function F_i. For simplicity, different dealers' I_i are assumed to be independently and identically distributed with uniform distribution within [-R, R][2].

Since the FX market is a continuously trading market, before any new information is released, the dealer's status at the beginning of the current period is actually the same as that at the end of the previous period. Hence, dealer *i*'s current quotes should be ones he made in

[1] The pricing of liquidity is important to explain the role of volatility in the determination of exchange rate bid-ask prices.

[2] Instead of the uniform distribution, it might be more appropriate to assume the inventory follows the normal distribution. However, the explicit form of normal distribution function is very hard to induce, if it is still possible. To obtain a relatively simple explicit form of solutions, I made this assumption. In fact, the fundamental relationships implied by the model will not be changed by just assuming another distribution.

the previous period[1]. If the middle point of his bid-ask quotes in the last period is denoted as p_{t-1}, then, before any information observed at the beginning of the current period t, his mid-quote should still be p_{t-1} too.

2) Public information releasing

I assume that the public information is released right after the beginning of the period and observed by every agent in the market[2]. Public information refers to the information available for all market participants, such as central bank announcements. Public information systematically changes the customers and dealers' expectations of exchange rates, and the dealer will adjust his quote immediately to incorporate the information. If the change in the exchange rate implied by the public information is denoted as r_t, then the dealer's quote becomes $p_{t-1} + r_t$. Let p^0 denote the dealer's mid-quote after releasing the public information, then $p^0 = p_{t-1} + r_t$. Since the mid-quote of the dealer's bid-ask prices is usually taken as the observation of the exchange rate, p^0 becomes the current market exchange rate and is observed by every agent in the market.

3) Market order arrives

Now, a market order comes. Some agent contacts the dealer with a potential order size of *Q* and asks for quotes. The transaction initiator could be another dealer in the inter-dealer market or one customer in the customer market. Suppose the initiator reveals a tendency of buying or selling when asking for quotes, so that the dealer knows the order size *Q* and the

---

[1] In the security market, the opening price of a new day and the closing price at the previous day are often different. But in the FX market, it is reasonable to make such an assumption, as the market is continuously trading.

[2] Another way to understand this assumption is separating the periods based on the release of public information.

direction of potential transaction before quoting[1]. This information is only observed by the dealer himself and therefore becomes his private information.

No matter whether dealers or customers initiate the transaction, their trading requests are assumed to be based on the difference between their own belief of the terminal exchange rate in this period and the current prevailing exchange rate (p^0). Furthermore, the sign of such difference determines the direction of transaction (buy or sell), and the order size is proportional to the magnitude of the difference.

4) Make the quote decision

As the response to the buying and selling request, the dealer needs to quote the ask price $p^0 + a_i$ and the bid price $p^0 - b_i$ respectively[2]. Knowing the information of order size Q and order direction, the dealer's objective is to choose a and b to maximize his expected utility in the inter-dealer and customer markets. The solutions to the problem will be his optimal quotes in both markets.

At this stage, I assume every dealer has to make quotes once being asked. In fact, it is possible that in some circumstances dealers refuse to quote, probably because they think the order is too large or the market is too volatile so that it is too risky to make quotations. However, due to the fierce competition among the dealers in the FX market today, very few would do so, as that will damage their reputation and scare customers away. In the model, I ignore the issue of whether to quote or not, and assume that dealers always make quotes once being asked.

2.2 Dealer's quoting problem and optimal quotes in the inter-dealer market

[1] Sometimes, dealers quote both bid-ask prices without knowing the tendency of the transaction. But the dealer from whom I collected the data knows the tendency before displaying his quote. To be consistent with the data source, I make such an assumption.

[2] Such a structure of the quotes makes the quoting problem symmetric and will facilitate our analysis later.

Following Biais (1993), the dealer's process of selecting optimal bid-ask quotes are laid out as two steps. First, the bottom line for the dealer to make a quotation is that at least he will not be worse off after revealing his quotes regardless if they are accepted or not. The quotes making him indifferent between trading and not trading are reservation quotes, which the dealer will take as a benchmark. Second, the dealer apparently has an incentive to raise the ask (selling) price above or reduce the bid (buying) price below the reservation quotes to earn extra profit; however, such a behavior will obviously decrease the probability that the quotes are accepted. Therefore, there must be a point which maximizes his expected utility and becomes his optimal choice.

Now, suppose another dealer sends a buying request with order size *Q*, and dealer *i* quotes the ask price $p^0 + a_i^d$, in which superscript *d* denotes the inter-dealer market. At this moment, before the dealer reveals his quotes, as assumed before, he holds cash K_i and inventory I_i.

If his quote is declined, and no transaction takes place, then the dealer's cash holding and inventory remain the same, however, the dealer still pays the fixed operation cost $\overline{C}_i$. Suppose the random variable $\tilde{z}$ is the difference between the current prevailing exchange rate p^0 and the exchange rate at the end of the period. The terminal value of the exchange rate will be $p^0 + \tilde{z}$, and the dealer's terminal wealth can be written as:

$$W(0) = K_i - \overline{C}_i + I_i(p^0 + \tilde{z}) \tag{2.1}$$

If his quote is accepted, and the transaction initiator buys Q at the ask price $p^0 + a_i^d$, then the dealer's inventory falls by *Q*, cash holding rises by $(p^0 + a_i^d)Q$, and he pays the extra operation cost ω in addition to $\overline{C}_i$. If $\tilde{z}$ is still defined as before, the dealer's terminal wealth after the transaction becomes:

$$W(a_i^d) = K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + \tilde{z}) + (p^0 + a_i^d)Q \tag{2.2}$$

Clearly the dealer's expected wealth depends on the expected value of $\tilde{z}$, and the dealer will certainly form his expectation about $\tilde{z}$ conditioned on the information he receives. According to our analysis before, the transaction initiator tending to buy Q units of the currency implies that in his opinion the market exchange rate in the future must be higher than the current one, i.e. $E(\tilde{z})$>0. Also, the order size Q should be proportional to $E(\tilde{z})$. Through a linear signal extraction process, the dealer factors the information set S^a into his own expectation of $\tilde{z}$ as follows:

S^a= *{some agent wants to buy Q units of the currency}* and $E(\tilde{z} \mid S^a) = \beta \cdot Q$

In the equation, the positive parameter β is used to measure how accurate the transaction initiator's expectation is in the dealer's opinion. Given the same order size, β is larger if the dealer believes that the agent asking for quotes is better informed[1]. In an extreme case, if the dealer thinks the agent is a noise trader, and his transaction request cannot reflect the true value of the exchange rate in the future at all, then β would be zero. In this sense, the value of β actually reflects the dealer's evaluation about how well informed the transaction initiator is.

According to its definition, the reservation quote should make quoting results (trading or not trading) indifferent, i.e. the dealer's conditional expected utility of the two cases equal to

[1] In the model of linear signal extraction, β is actually negatively related to the variance of expectation. If the agent is believed to be better informed, then his expectation variance will be smaller, so β is larger.

each other. Suppose dealer *i*'s reservation ask price in the inter-dealer market is denoted as $p^0 + a_i^{r,d}$, the following equation (2.3) should hold:

$$E(U(W(0)) \mid S^a) = E(U(W(a_i^{r,d})) \mid S^a) \tag{2.3}$$

Similarly, if another dealer sends selling request with the order size *Q*, the dealer quotes bid price $p^0 - b_i^d$. If his quote is declined, the terminal wealth is the same as shown in equation (2.1). If his quotes are accepted, the dealer's terminal wealth will be:

$$W\ (b_i^d) = K_i - \overline{C}_i - \omega\ + (I_i + Q)(p^0 + \tilde{z}) - (p^0 - b_i^d)Q \tag{2.4}$$

In the selling case, the dealer will receive information set S^b and make the expectation about $\tilde{z}$ as below:

S^b= {*some agent wants to sell Q units of the currency*} and $E(\tilde{z} \mid S^b) = -\beta \cdot Q$

The reservation bid price $p^0 - b_i^{r,d}$ will make equation (2.5) hold:

$$E(U(W(0)) \mid S^b) = E(U(W(b_i^{r,d})) \mid S^b) \tag{2.5}$$

The dealer's reservations quotes, which are the solutions to equations (2.3) and (2.5), are shown in lemma 2.1, and derivation process is attached in appendix III.

Lemma 2.1: *Under our set of assumptions, dealer i's reservation selling (ask) and buying (bid) prices in the inter-dealer market are* $(p^0 + a_i^{r,d})$ *and* $(p^0 - b_i^{r,d})$ *, where*

$$a_i^{r,d} = (\frac{A\,\sigma^2}{2} + \beta^d) \cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i$$

$$b_i^{r,d} = (\frac{A\sigma^2}{2} + \beta^d) \cdot Q + \frac{\omega}{Q} + A\sigma^2 \cdot I_i$$

Next, I derive the dealer's optimal deviation from the reservation quotes. For each buying request, the dealer revealing quotes leads to two possible results: accepted or not accepted. In the first case, the dealer's utility will be $E(U(W(a_i^d)) \mid S^a)$, while the utility is $E(U(W(0)) \mid S^a)$ in the second, where in both cases the wealth of portfolio is defined as before. The reason that expected utility is applied here even for each specific state is because the change in the exchange rate ($\tilde{z}$) is still a random variable and will not be realized until the end of the period. Let $\pi_{a,i}^d$ denote the probability that dealer *i*'s quote is accepted, then as usual, the total expected utility for dealer *i* can be written as:

$$EU = \pi_{a,i}^d E(U(W(a_i^d)) \mid S^a) + (1 - \pi_{a,i}^d) E(U(W(0)) \mid S^a) \tag{2.6}$$

Currently, the major fraction of turnover in the inter-dealer market is conducted through the electronic inter-dealer dealing system, such as Reuters D2000-2. The system displays the real time information of the best bid-ask price available on its screen, and every dealer can observe it easily and almost costlessly. If dealer *i* is chosen to be bought from, it usually means his ask price is lower than any other competitor's. Accordingly, the probability of other dealers accepting his quote is equivalent to the probability that his ask price is the lowest among all the dealers, which can be expressed as $\pi_{a,i}^d = P(a_i^d < a_{-i}^d)$ where subscript $-i$ represents all other dealers.

Now, if the dealer's current inventory is higher than the preferred level, i.e. $I>0$, it suggests that the dealer holds more position than he would like to. In order to reduce the inventory, quoting a lower price is an efficient strategy to encourage selling and deter buying. On the contrary, if the current inventory is lower than preferred level, i.e. $I<0$, the dealer intends to increase the price in order to sell less or buy more, so that the inventory will move up to the ideal level. Therefore, I claim that to all dealers the optimal quoting strategy should be decreasing in each dealer's inventory level *I*.

Since the dealer's optimal quoting strategy is decreasing in the inventory, the probability that the dealer's ask price is lower than the others is equivalent to the probability that his inventory is larger than the others, i.e. $\pi_{a,i}^{d} = P(a_i^d < a_{-i}^d) = P(I_i > I_{-i})$. Using the assumption that the dealers' inventory levels are random variables with the identical and independent uniform distributions F_i within [-R, R], the probability can be expressed as:

$$\pi_{a,i}^{d} = P(a_i^d < a_{-i}^d) = P(I_i > I_{-i}) = \prod_{j \neq i}^{N} F_j(I_i) = (\frac{I_i + R}{2R})^{N-1} \tag{2.7}$$

Now, go back to the original quoting problem. Rearranging equation (2.6) gives equation (2.8):

$$EU = \pi_{a,i}^{d}[E(U(W(a_i^d)) \mid S^a) - E(U(W(0)) \mid S^a)] + E(U(W(0)) \mid S^a) \tag{2.8}$$

According to the definition of the reservation quote, equation (2.9) must hold:

$$E(U(W(a_i^d)) \mid S^a) = E(U(W(0)) \mid S^a) e^{-A(a_i^d - a_i^{r,d})Q} \tag{2.9}$$

Substituting equation (2.9) to equation (2.8) gives:

$$EU = \pi_{a,i}^{d}(e^{-A(a_i^d - a_i^{r,d})Q} - 1)E(U(W(0)) \mid S^a) + E(U(W(0)) \mid S^a) \tag{2.10}$$

Since $E(U(W(0)) \mid S^a)$ does not depend on the choice of the ask price, it can be neglected from the objective function, and the aim of the dealer will be simplified as:

$$\underset{a_i^d}{Max}\, \pi_{a,i}^{d}(e^{-A(a_i^d - a_i^{r,d})Q} - 1) \tag{2.11}$$

Finally, the linear approximation of objective function (2.11) by a Taylor expansion gives:

$$\underset{a_i^d}{Max}\, \pi_{a,i}^{d}(A(a_i^d - a_i^{r,d})Q) \tag{2.12}$$

where $\pi^d_{a,i}$ is defined in equation (2.7). Objective function (2.12) demonstrates the dilemma that the dealer is facing when he chooses quotes. On one hand, a higher ask price (a_i^d) relative to the reservation quote ($a_i^{r,d}$) is justified by a higher value of the objective function and potential profit. On the other hand, it reduces the probability of the quote being accepted ($\pi^d_{a,i}$) for obvious reasons. Hence, there must be an optimal ask price that maximizes the objective, which will be the dealer's choice.

The analysis above describes the process of selecting the ask price when the dealer receives the buying request. Due to the symmetric setup of bid-ask quotes, the optimal bid price can be determined similarly.

Following Biais (1993)'s general proof, I derive the explicit form of dealer *i*'s optimal bid-ask quote in the inter-dealer market and show them in proposition 2.1.

Proposition 2.1: *Under our set of assumptions, the solutions to the dealer's problem and the optimal bid and ask prices of dealer i in the inter-dealer market are* $(p^0 - b_i^d)$ *and* $(p^0 + a_i^d)$*, where*

$$a_i^d = a_i^{r,d} + A\sigma^2 \cdot \frac{R + I_i}{N}$$

$$b_i^d = b_i^{r,d} + A\sigma^2 \cdot \frac{R - I_i}{N}$$

2.3 Dealer's quoting problem and optimal quotes in the customer market

The customer reservation quotes can be derived in the same way as in the inter-dealer market. In response to buying and selling requests from customers, dealer *i* quotes the ask price $p^0 + a_i^c$ and bid price $p^0 - b_i^c$ respectively. The terminal wealth of each state can also be expressed in equations (2.1), (2.2) and (2.4), except that the superscript *d* should be replaced by *c* to denote the customer market. Still, reservation quotes make equations (2.3) and (2.5) hold in the customer market. Solutions are illustrated in the following lemma 2.2.

Lemma 2.2: *Under our set of assumptions, dealer i's reservation selling (ask) and buying (bid) price in the customer markets for dealer i are* $(p^0 + a_i^{r,c})$ *and* $(p^0 - b_i^{r,c})$ *, where*

$$a_i^{r,c} = (\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i$$

$$b_i^{r,c} = (\frac{A\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} + A\sigma^2 \cdot I_i$$

The composition of the dealer's expected utility in the customer market is similar to the inter-dealer market. With each buying request from the customer, the dealer can end up trading or not trading, and his utilities are $E(U(W(a_i^c)|S^a)$ and $E(U(W(0)|S^a)$ respectively. If $\pi_{a,i}^c$ is the probability that the customer accepts dealer *i*'s quote, then the total expected utility can be written as:

$$EU = \pi_{a,i}^c E(U(W(a_i^c)|S^a) + (1-\pi_{a,i}^c)E(U(W(0))|S^a) \tag{2.13}$$

Reusing the technique I managed the expected utility with before, the quoting problem in the customer market can be simplified to:

$$\underset{a_i^c}{Max}\,\pi_{a,i}^c(e^{-A(a_i^c - a_i^{r,c})Q} - 1) \tag{2.14}$$

Disregarding the similarities, two significant differences between the customer and inter-dealer FX markets are identified in our model. The first one is embodied in the difference of information asymmetry. In the FX market, it is common sense that dealers are more likely to have superior information to customers, although it is not always true[1]. Therefore the dealer tends to believe other dealers are better informed than customers when he deals with trading requests from them. As a result, he believes that dealers trading requests reveal the tendency of the exchange rate change more accurately than the customer request. As indicated before, the parameter β reflects such a difference in our model. So, I claim that the value of β in the inter-dealer market should be greater than the customer market, i.e. $\beta^d > \beta^c$. In lemmas 2.1 and 2.2, the reservation quotes in both markets are exactly the same except for the superscript in parameter β, which embodies this difference.

The second significant difference results from different market transparency. In the inter-dealer market, all dealers have access to the electronic dealing system, which displays the real time information of the best available quotes, thus, it does not cost much effort or money for the dealer to obtain the lowest quote information. While in the customer market, there is no such platform for the customers to obtain the information about all the dealers' quotes[2], not to mention finding out the best one. Mathematically, this difference is reflected by the way of determining the probability of customers accepting the dealer's quote.

As noted before, in the inter-dealer market, dealer *i* is chosen to trade with mostly because his ask price is lower than those of his competitors. In the customer market, customers might not buy from dealer *i*, even his ask price is the lowest. Without doubts, the customer would like to trade at a price as low as possible when he is planning to buy. However, since the customer market is relatively less transparent than the inter-dealer

[1] The central bank is also considered as a customer, and it is hard to say that the Fed is less informed than dealers.

2 In the inter-dealer market, dealers can easily find the best price in a real time updated media (like Reuters D2000-2), at an ignorable cost relative to the huge turnover. But, customers don't have such a platform to get all dealers' quotes efficiently.

market, the customer has to pay significant costs to achieve this goal. The costs come from two sources: 1) gathering the information of all dealers' quotes and finding the best one, and 2) opening a new account in the bank that provides the best price, and transferring his money to the chosen one (except when the customer happens to hold an account in that bank). Customers will stick with dealer *i* unless his ask price exceeds another dealer's ask price by the cost of searching and switching.

In addition to the searching and transferring costs that hinder customers switching to other dealers, the dealer has other strategies to win customers' loyalty, even with inferior quotes in the market. The most-seen customer-attracting strategies include: free access to the market information, free research report, free training program, higher leverage rate and lower limit of minimum trade. I use one parameter δ_i to represent the non-price attractiveness of dealer *i*, and call it the "*attractiveness coefficient*". Considering these factors, the customer chooses dealer *i* only if his ask price is less than an other's plus the additional cost, which is an increasing function of the attractiveness coefficient (δ_i) and the searching and switching cost (*c*). Supposing the additional cost function is $\delta_i c$, the probability that the customer accepts dealer *i*'s quote can be written as equation (2.15):

$$\pi^c_{a,i} = P(a^c_i < a^c_{-i} + \delta_i c) \tag{2.15}$$

When I discuss the dealer's optimal quoting strategy in the inter-dealer market, I claim that his ask price should be decreasing in the inventory level due to the inventory effect. Based on it, the probability that the dealer's ask price is the lowest in the inter-dealer market is equivalent to the probability that his inventory is the largest. In the customer market, the inventory effects play similar roles, and the dealer's optimal ask price for customers is still decreasing in the current inventory level. But in the customer market, as illustrated in equation (2.15), the customer accepting the dealer's quote does not necessarily mean his quote is the lowest. On the contrary, if only the dealer's ask price is less than any other's ***plus*** the additional cost ($\delta_i c$), his quote will be accepted. Therefore, the dealer's current inventory does not have to be larger than any other's to make the price accepted; on the contrary, if only his inventory is just greater than any other's ***minus*** the effect of the additional costs mapped into the inventory level, his quote will be accepted. Suppose *f* is the function which maps the effect of extra cost into the inventory, equation (2.16) should hold. It is clear that *f* is an increasing function in the cost $\delta_i c$.

$$P(a_i^c < a_{-i}^c + \delta_i c) = P(I_i > I_{-i} - f(\delta_i c)) \tag{2.16}$$

Hence the probability of the customer accepting the dealer's quote can be written as:

$$\pi_{a,i}^c = P(a_i^c < a_{-i}^c + \delta_i c) = P(I_i > I_{-i} - f(\delta_i c)) \tag{2.17}$$

Now, we have dealer *i*'s quoting problem in the customer market as follows:

$$\text{Max}\, \pi_{a,i}^c (e^{-A(a_i^c - a_i^{r,c})Q} - 1)$$

where $\pi_{a,i}^c = P(a_i^c < a_{-i}^c + \delta_i c) = P(I_i > I_{-i} - f(\delta_i c))$. The solutions to the problem are displayed in the following proposition 2.2, and the proof is shown in appendix III:

Proposition 2.2: *Under our set of assumptions, the solutions to the dealer's problem and the optimal bid and ask price of dealer i in the customer market are* $(p^0 - b_i^c)$ *and* $(p^0 + a_i^c)$, *where:*

$$a_i^c = a_i^{r,c} + A\sigma^2 \cdot \frac{R + I_i}{N} + \delta_i c$$

$$b_i^c = b_i^{r,c} + A\sigma^2 \cdot \frac{R - I_i}{N} + \delta_i c$$

2.4 Bid-ask spreads and mid-quotes in both markets

So far, I have presented a theoretical framework to describe the dealer's quoting problem in both inter-dealer and customer FX markets, and propositions 2.1 and 2.2 display the explicit form of the optimal bid-ask price in these two markets respectively. Our further concerns are the spread and mid-quote of bid-ask prices. The spread is the difference between the bid and ask quotes, and the mid-quote is just the simple average number of a quote pair. Through simple mathematical derivations, the following proposition 2.3 illustrates the mid-quotes and spreads in both markets.

Proposition 2.3: *Under our set of assumptions, dealer i's mid-quote of bid-ask prices and spread in the inter-dealer market are* $p_{t,i}^d$ *and* S_i^d*, where:*

$$p_{t,i}^d = p_{t-1,i} + r_t - A\sigma^2(1-\frac{1}{N})\cdot I_i, \quad S_i^d = (A\sigma^2 + 2\beta^d)\cdot Q + \frac{2\omega}{Q} + A\sigma^2\cdot\frac{2R}{N}$$

The mid-quote of bid-ask prices and spread in the customer market are $p_{t,i}^c$ *and* S_i^c*, where:*

$$p_{t,i}^c = p_{t-1,i} + r_t - A\sigma^2(1-\frac{1}{N})\cdot I_i, \quad S_i^c = (A\sigma^2 + 2\beta^c)\cdot Q + \frac{2\omega}{Q} + A\sigma^2\cdot\frac{2R}{N} + 2\delta_i c$$

Several important relationships are implied by proposition 2.3.

Implication 1: *the customer mid-quote equals the inter-dealer mid-quote*

The current exchange rates ($p_{t,i}^d$, $p_{t,i}^c$) in both markets are formed as an adjustment on the basis of the exchange rate at the previous period ($p_{t-1,i}$). This adjustment is negatively affected by the inventory level, which embodies the inventory effect on the exchange rate dynamics. Item *r* embodies the public information integrated into the price.

Comparing the mid-quotes in both markets, we can find that they are exactly the same. As we all know, many dealers quote their customer rates based on inter-dealer ones, and this implication basically explains why. Intuitively, the mid-quote can be interpreted as the dealer's belief of the currency value, and this belief should be the same no matter in which market and with whom he is trading. This relationship is presumably so right that nobody has ever questioned it before in the literature, and nobody has tested this relationship with data either due to the lack of customer quotes data.

Implication 2: *the effect of order size on spread is ambiguous in the two markets*

The explicit form of the spread illustrates three components of the effect of order size on spread caused by three different factors: inventory holding risk, information asymmetry and the economy of scale. The coefficient $A\sigma^2$ indicates a positive relationship between the order size (Q) and the spread, with the intuition that the dealer needs to bear more inventory holding risk when accepting larger orders. Meanwhile, the coefficient 2β shows such a relationship too, in the sense that the dealer would widen the spread to deter a transaction with better informed agents and avoid increasing loss if the order size is larger. On the other hand, the item $2\omega/Q$ demonstrates the negative effect of order size on spread from the perspective of the economy of scale. Intuitively, the operation cost would be spread more thinly when the order size is larger. Hence, the overall effect should be the combination of these three mixed impacts, and therefore ambiguous.

In the literature, both inventory holding risk models and information asymmetry models predict that the order size should positively affect the spread, i.e. the larger order size leads to the wider spread. In contrast, my model shows that the effect is ambiguous. Empirically, no work has been devoted to testing this relationship before, so we have no clue about which one is right so far.

Implication 3: *the customer spread is generally wider than the inter-dealer spread, but their differential decreases with the rise in order size.*

To compare the spreads in the two markets, the inter-dealer spread is subtracted from the customer one, and the result is shown in equation (2.18):

$$S_i^c - S_i^d = 2\delta_i c + 2(\beta^c - \beta^d)\cdot Q \tag{2.18}$$

The first item of the differential $(2\delta_i c)$ says that the customer spread is generally wider than the inter-dealer one, as the extra cost paid by customers to trade at the best quote makes the margin possible for the dealer to do so. Moreover, if the dealer wins customers' high loyalty, i.e. the value of the attractive coefficient δ_i is high, he can even keep a wider customer spread without losing customers, and the difference with the inter-dealer spread would be bigger. This reflects a fact that in the FX market the dealers compete with each other not only in quotes but also in services, and keeping a good reputation and high customer loyalty is critical to earn higher profit. The coefficient of Q in equation (2.18) implies that the difference will decline with the rise in order size, which results directly from the claim of $\beta^d > \beta^c$. Intuitively, with the order size increasing, the information cost of trading with customers is not as much as with dealers, as customers are usually less informed than the latter.

In the literature, no formal models have been built to explain this difference, but there is some indirect evidence just suggesting that the customer spread is wider than the inter-dealer one. Yao (1997), for example, found that the customer trades account for only about 14% of total trading volume, but they represent 75% of the dealer's total profits. And also, the market survey by Braas and Bralver (1990) finds that most dealing rooms generate between 60 and 150 percent of total profits from the customer business. Other than these analyses, neither theoretical model nor empirical testing indicates the relationship before.

Implication 4: *Spreads in both markets are positively affected by the exchange rate volatility*

Proposition 2.3 clearly shows that the spreads in both markets have a positive relationship with the exchange rate volatility σ^2, i.e. the more volatile the exchange rate is, the wider the spreads should be. Intuitively, a higher volatility of the exchange rate leads to higher risk of holding the currency, and the dealer would widen the spread to cover the possibly increasing loss caused by higher market uncertainty. Many models theoretically show such a relationship in the literature (e.g. Stoll (1978), Ho and Stoll (1981), Black (1991), Biais (1993), Bollerslev and Melvin (1994), etc). Also, this relationship has been extensively examined and verified by several persons' findings. Bessembinder (1994), Bollerslev and Melvin (1994) as well as Hartmann (1999) all found a positive correlation between the spread and expected volatility measured by GARCH forecasts. Wei (1994) regresses the bid-ask spread on the realized exchange rate volatility measured by the standard deviation of the exchange rate and the anticipated volatility induced from the price of foreign exchange options respectively, and both regressions agree on a significant positive relationship between the spread and volatility. Those findings provide strong evidence to support the solutions of our model.

Implication 5: *spreads in both markets are negatively related to the number of dealers*

According to the proposition, we also can see that the spreads are negatively related to the number of dealers (*N*) in both markets, i.e. when more dealers are active in the market, the spreads will fall. Intuitively, the more dealers quote in the market, the more competition among them, and one of the efficient strategies to win the competition is to reduce the spread. In the literature, the competition between dealers and its impact on the spreads have been identified by Stoll (1978) and Biais (1993). Empirically, Huang and Masulis (1999) used the market data and found that the spreads decrease with a rise in the number of dealers, which is measured as the frequency of new quotes.

3. Empirical evidence

3.1 Data collection and description

The data used in this essay were collected from one of world's largest online foreign exchange dealers.[1] The dealer displays the customer and inter-bank rates simultaneously for several major currencies in a response to each individual transaction request on his quote window, providing a platform for us to collect data. To raise the efficiency of data collecting, I focused on the rate of the US dollar versus the Euro (USD/EUR), as it is currently the most frequently traded currency pair in the world.

A series of order sizes are generated randomly by computer and sent to the dealer at the interval of 1 minute. Corresponding to each sent order size, the dealer displays bid-ask quotes, which are recorded and become my data. This work was done during the period of July 7 to July 15, 2004, with the weekend excluded due to low transaction activities. About three hours were spent each day to do this job (usually 9am—12pm, which is supposed to be the busiest trading time every weekday). Thus, our data set consists of the order size of each transaction request, and the corresponding customer and inter-dealer bid-ask prices on the basis of a 1-minute interval. The mid-quotes of bid-ask prices and spreads in both markets can be easily obtained according to their definitions.

Compared to previously used data, our data set is characterized by several new features. First, it contains customer bid-ask quotes rather than just inter-dealer quotes. Before online

[1] The data and more details regarding data sources are available upon request.

foreign exchange dealing services became popular, the customer quotes were dealers' commercial secrets and publicly unavailable. Nowadays, to satisfy the increasing trading demands around the world and around the clock, the dealer offering online trading services shows the quotes publicly, which makes it possible to obtain the data for academic purpose.

Second, in addition to the customer bid-ask quotes, the data set also contains inter-dealer quotes, which make the empirical comparison between them possible. In order to do such a comparison, the data set should contain bid-ask quotes from both markets covering the same period and picked at the same frequency, and our data completely satisfy these requirements.

Third, our data set contains order sizes corresponding to each customer and inter-dealer quote, while no previous data set has[1]. Theoretically, both inventory risk models and information asymmetry models indicate a certain relationship between the spread and single transaction volume, but no empirical work has been done to test this relationship due to the lack of data.

Two major concerns about the quality and reliability of the data should be addressed here. First, instead of real transaction quotes, those quotes are just indicative quotes. Since our model is built to show the dealer's optimal quotes without dealing with the issues of whether the customer or another dealer accepts them or not, this disadvantage would not undermine the efficiency of our testing. Second, the typicality of the dealer and his quotes is another serious concern. Given the nearly perfect capital mobility between industrialized countries and the dominance of the electronic trading system in today's FX market, which is continuously traded 24/7, transaction costs in such a market become ignorable. Thus neither time nor geographic differences can sustain significant differences among the quotes of

[1] Some data sets contain the trading volume, but trading volumes and order size are different concepts: the former is the summation of transaction of quantity no matter sell or buy, while the latter is volume of single transaction.

various dealers. Meanwhile, fierce competition among the dealers, especially since online trading has become popular, drives their quotes closer to prevailing market rates. Therefore, I have reasons to believe that the quotes I obtained from this particular dealer can represent the market rates and are reliable for our test.

3. 2 Model testing and results

The last two implications indicated in section 2.4, especially the relationship between the spread and the volatility, have been examined extensively, while the first three have never been tested before. Therefore, our empirical exercise will focus on these three implications.

I start with implication 1 to test the relationship between customer and inter-dealer mid-quotes. The mid-quotes in both markets are computed as the average value of bid-ask prices. Two series of data will be tested in both markets: the mid-quotes at the interval of 1-minute as well as 5-minute. In order to test whether two mid-quotes are equal to each other exactly, we can regress the following equation to see whether $\alpha = 0, \beta = 1$:

$$p_t^c = \alpha + \beta \cdot p_t^d + \varepsilon_t \tag{2.19}$$

where p_t^c, p_t^d are the customer and inter-dealer mid-quotes respectively. However, it is well known that the spot exchange rate is a nonstationary process integrated by order 1 (I(1) process). The results of the unit root tests in table 2.1 in appendix I verify such a judgment. Both Dickey-Fuller and Phillips-Perron tests cannot reject the null hypothesis of unit root at a significance level above 99%. Therefore it might be inappropriate to regress the two series of mid-quotes directly.

To overcome this problem, the first order difference of the mid-quotes is regressed, instead of their levels, shown as the following equation:

$$p_{t+1}^c - p_t^c = \alpha + \beta(p_{t+1}^d - p_t^d) + \varepsilon_{t+1} \tag{2.20}$$

Apparently, the null hypothesis that $\alpha = 0, \beta = 1$ should also be accepted, if the customer mid-quote equals the inter-dealer one. The regression results of both equations (2.19) and (2.20) are summarized in table 2.2. According to the table, all regressions display extremely significant results about the coefficients of the equation, and the null hypothesis is accepted at a high significance level.

According to proposition 2.3, the differential between the customer and inter-dealer spread can be written as:

$$S_i^c - S_i^d = 2\delta_i c + 2(\beta^c - \beta^d) \cdot Q \tag{2.21}$$

Since equation (2.21) describes a simple linear relationship, and all variables contained in the equation are well measured, I just use OLS to test a linear model as below:

$$s^c - s^d = \alpha_0 + \alpha_1 \cdot Q + \varepsilon \tag{2.22}$$

Variable Q in the equation is measured by the logarithm of order size, and spreads in both markets are represented by the relative spreads, which are computed from customer and inter-dealer bid-ask prices at the 5-minute interval by the following formula:

$$s = \log(a) - \log(b) \tag{2.23}$$

where a and b are ask and bid prices respectively. Based on the analysis in section 2.4, the customer spread is generally greater than the inter-dealer spread, but their differential decreases with a rise in order size. If this analysis is right, it is expected to see a significant positive intercept and a negative coefficient of order size in the regression. Table 2.3 reports the regression results. A highly significant positive intercept is found, implying that the customer spread is in general greater than the inter-dealer one, while a significantly negative coefficient of order size says the difference will fall with the rise in order size. The results of R square and F-test indicate that this regression is very reliable.

Finally, to inspect the relationship between the spreads and the order sizes, the econometric model should include the spread as a dependent variable and the order size as an independent one. However, spreads are also affected by other factors which could jeopardize the estimation of order sizes' impacts if they are excluded from the model. To avoid this problem, these factors must be identified and incorporated to estimate the model properly.

Exchange rate volatility is one such factor and has been proven to affect spreads significantly in many theoretical and empirical studies. Numerous studies have demonstrated that the volatility of spot exchange rates change can be modeled as a GARCH process[1]. In this essay, the spot exchange rate return is modeled as MA(1)-GARCH (1,1) process[2] as below:

$$
\begin{aligned}
&10{,}000 \cdot \Delta M_t = \mu + \theta \varepsilon_{M,t-1} + \varepsilon_{M,t} \\
&\sigma^2_{M,t} = \varpi + \alpha \varepsilon^2_{M,t-1} + \beta \sigma^2_{M,t-1}, \\
&\varepsilon_{M,t} \mid I_{t-1} \sim N(0, \sigma^2_{M,t})
\end{aligned}
$$

In this model, M stands for the exchange rate and ΔM is the spot exchange rate change within the window period of data collection. I represents information set, and $\mu, \theta, \varpi, \alpha, \beta$ are the parameters to be estimated. The time t subscript refers to the place in the order of the series of quotes, so that $\hat{\sigma}^2_{M,t}$ provides an estimate of the exchange rate volatility. Since the magnitude of mid-quote fluctuations is very small within the 1-minute window, the exchange rate change ΔM is multiplied by 10,000 to enlarge the effects of dependent variables and make the estimated parameters not too small. In this estimation model, the observation of the

[1] See Bollerslev et al. (1992)

[2] Our model literally suggests that a GARCH-in-Mean specification would be the best fit to estimate the exchange rate volatility. However, according to the estimation results in essay 1, the GARCH-in-mean item was insignificant in the short horizons; therefore I just applied a simpler method GARCH to estimate the volatility.

exchange rate (M) is measured as the logarithm of the mid-quote of bid-ask prices. Given that a and b denote the ask and bid prices respectively, M is computed by the following formula:

$$midquote = M = \log(\frac{a+b}{2})$$

The GARCH model specified above is estimated by the maximum likelihood method, and estimation results are reported in table 2.6. According to the table, none of the parameters are significantly different than zero except for the constant item in innovation equation. This result suggests that the volatility of the spot exchange rate is basically determined by the constant item ϖ and independent of other variables. In other words, exchange rate fluctuation within the 1-minute interval is pretty stable, probably because this data collection window is too short to incorporate volatile factors.

To simplify model specification, a linear model is applied to expresses the dependent variable spread as a function of its determinants--order size and exchange rate volatility. Suppose that the symbol s_t denotes the spread, the order size is denoted by O_t and $\hat{\sigma}_t$ represents the exchange rate volatility, which is estimated by the GARCH model above, then we have the following estimation model:

$$s_t = \gamma_0 + \gamma_1 \hat{\sigma}_t + \gamma_2 O_t + \varepsilon_t \quad \varepsilon \sim N(0, \sigma_\varepsilon^2), t = 1,2....,T. \tag{2.24}$$

In the model above, $\gamma_0, \gamma_1, \gamma_2$ are constant parameters and ε_t is an error item following a normal distribution with zero mean. The variable O_t is taken as the logarithm of the original order sizes. Given that a_t and b_t denote the ask and bid prices respectively, the customer and inter-dealer spreads are also computed as the logarithm of the spread measured in pips, i.e., based on the following equation:

$$S_t = \log[(a_t - b_t) \cdot 10000] \tag{2.25}$$

The choice of estimation method is determined by the nature of the spread data. It is well known that high frequency financial data such as exchange rate spreads usually exhibit non-normality and high autocorrelation. Therefore, the OLS or the Maximum Likelihood Estimation is not an efficient or appropriate method to estimate the model. Therefore, GMM was used to estimate the model in this essay. A key advantage to GMM over other estimation procedures is that the statistical assumptions required for hypothesis testing are quite weak. Therefore, neither the autocorrelation of the data nor the non-normality of the residuals would jeopardize the estimation.

Let γ denote the vector of parameters in our model, and $M(\lambda)$ is a vector of moment conditions. Given a weighting matrix W, generally, GMM chooses the parameters which minimize the quadratic function $J(\hat{\gamma})$ as below:

$$J(\hat{\gamma}) = M(\gamma)'WM(\gamma) \tag{2.26}$$

Suppose the moment conditions are denoted by the matrix Z, and ε represents the vector of residuals, the moment conditions for our linear model can be written as:

$$T^{-1}Z'\varepsilon = 0 \tag{2.27}$$

Substitute the moment conditions into equation (2.26), the objective function becomes:

$$J(\hat{\gamma}) = (T^{-1}Z'\varepsilon)'W(T^{-1}Z'\varepsilon) \tag{2.28}$$

The weighting matrix W in the GMM objective function determines the relative importance of the various moment conditions. There are many approaches to estimate *W*, which can account for various forms of heteroskedasticity and/or serial correlation. In this essay, Newey and West, White, and Gallant weighting matrixes are applied for the purpose of robustness. Meanwhile, instrument variables are chosen from the explanatory variables themselves. The first instrument matrix employs one period lag of the explanatory variables (GMM 1) and the other one is the square of the explanatory variables (GMM2).

Three types of tests have been conducted to inspect whether the order size affects the spread significantly. First, the t statistic of the order size coefficient in the model is computed and the significance based on the t-distribution can be detected. Second, the likelihood ratio test (*LR* test) is conducted. In order to obtain the likelihood ratio, an unconstrained model (with the variable of order size in the model) is estimated first and the weighting matrix obtained in the testing is saved, and then the same weighting matrix is used to estimate a restricted model, which fixes the coefficient of order size as zero, i.e. the independent variables do not include order size. If the difference between the logarithm values of the two objective functions significantly deviates from zero, the order size is a significant determinant of spreads. Otherwise, the hypothesis that it has no impact on spreads is accepted. Given *LR* as the likelihood ratio, $J^r(\hat{\gamma})$ and $J^u(\hat{\gamma})$ are the values of the restricted and unrestricted likelihood functions after estimation respectively; then the likelihood ratio is computed as:

$$LR = 2 \cdot [\log J^r(\hat{\gamma}) - \log J^u(\hat{\gamma})] \tag{2.29}$$

Finally, the Wald test is used to check the significance of coefficients jointly. The Wald test is based on the restriction of the form $R\gamma = r$, where R is the parameter vector of restriction conditions, γ is the coefficient vector of our model, and r is a constant to be tested. In this testing, since we are concerned about whether the order size coefficient γ_2 is significantly different from zero, therefore R=[0,0,1] and r=0. Denote the variance of the coefficient vector γ as Σ such that $\mathrm{var}(R\gamma - r) = R\Sigma R'$, the Wald test statistic can be computed as:

$$W = (R\gamma - r)'(R\Sigma R')^{-1}(R\gamma - r) \tag{2.30}$$

Under the hypothesis that Z and ε are orthogonal, the asymptotic distribution of the objective function, *LR* statistic, and Wald statistics all follow the Chi-square distribution, therefore the acceptance or rejection decision can be made according to the critical values computed from the Chi-square distribution.

Tables 2.4 and 2.5 report the results of estimation of equation (2.24) for the inter-dealer and customer spreads respectively. In the estimation of inter-dealer spreads, it turns out that order size does not significantly affect the inter-dealer spreads, as shown in Table 2.5. In all estimations, no matter what instruments or density matrix are used, the parameter of order sizes (γ_2) is very close to zero. More importantly, t statistics are around -0.5, and likelihood ratio as well as Wald statistics are all less than 0.3. Given that the t statistic follows t distribution, and both the *LR* and Wald statistics follow Chi-square distribution, they are far below the critical value, which suggests the acceptance of the null hypothesis that γ_2 is zero. In other words, spreads are independent of order sizes in the inter-dealer foreign exchange market.

On the other hand, there is a different scenario in the customer market, as shown in table 2.4. All estimations show that order sizes affect customer spreads negatively at over 95% significance level, except for the GMM1 using NW and Gallant, which is still above 90% significant. In other words, the larger order size is, the narrower the spread is in the customer market.

4. Conclusions

To examine the dealer's quoting behavior under the different market structures, this essay develops a model viewing the dealer as a risk averse agent who optimizes his portfolio by choosing bid-ask prices. The differences between inter-dealer and customer markets are recognized and incorporated into the model, and the optimal bid-ask quotes are solved in both markets, so that the explicit forms of mid-quotes and spreads are obtained for further analysis.

Two major distinctions between the inter-dealer and customer FX markets are theoretically identified in this essay. The first one is embodied in the difference of market transparency. Compared to the inter-dealer market, where most transactions are conducted through the electronic dealing system (e.g. Reuters) on which dealers recognize the best quotes efficiently and almost costlessly, the customer market is relatively more opaque, and it costs regular customers much more to find the best price available and trade at it. This difference allows a margin for the dealer to keep the customer spread generally wider than the inter-dealer one. The second is embodied in the difference of information asymmetry. Since dealers are usually believed to be better informed than customers, if the order size is larger, the dealer is more inclined to raise the spread to deter adverse selection when trading with dealers than customers. Therefore, the differential between the customer and the inter-dealer spreads decreases with the rise in order size. In spite of these distinctions, as the dealer's belief of the true value of currency, his mid-quotes of bid-ask prices in the two markets are proved to be identical in the model.

To test the relationship implied by our model, a new data set has been collected from an online FX dealer based in Australia. This data set contains both inter-dealer and customer bid-ask quotes as well as order size covering the same period and picked at the same frequency. By employing the data set, the essay obtains convincing evidence supporting our theoretical conclusion that the customer spread is wider than the inter-dealer spread while the difference declines if the order size is larger. Our results also suggest that the customer mid-quote and the inter-dealer one are statistically the same variables. Meanwhile, the impact of order size on spread is found to be negative in the customer market while no effect is found in the inter-dealer market, which implies that the impact is ambiguous and stands against the conclusions of preceding models.

Although our work throws some light on the market microstructure of the FX markets and its impact on the dealer's quoting behavior, it also reveals more open areas to be explored. First, our study only involves one individual dealer and one currency pair. Whether the conclusions apply to other dealers and other currencies deserves further work. Second, I did not consider too much heterogeneity of various dealers when I built the model, while different dealers with various preferences should act differently in the FX market. How this factor affects their behavior is another important topic for further research. Third, the model solves the dealer's optimal quoting strategy without dealing with the issue that what is the final transaction price. A more complete and sophisticated model considering all dealers and whole market is needed to derive the equilibrium market rate. Finally, IT technology has been progressing rapidly in the past two decades, which has a huge effect on the FX market. What changes the new trading platform and services bring to the market participants and how their behavior adjusts in the new system should be another exciting research objective.

ESSAY THREE

THE TERM STRUCTURE OF THE FORWARD PREMIUM PUZZLE

1. Introduction

One of the most extensively examined topics in economics has been the efficiency of the forward market for foreign exchange. In particular, an enormous number of studies have shown the failure of the forward exchange rate to serve as an unbiased predictor of the future spot rate. This finding has been replicated in an extensive literature, including the initial studies by Bilson (1981) as well as Meese and Rogoff (1988). Furthermore, most empirical testing generates significant negative coefficients, instead of the theoretical value of unity, in the regression of the spot rate change on the forward premium. In the literature, this phenomenon is often referred to as the forward premium (or discount) puzzle (or anomaly).

A notable aspect of almost all published studies, however, is that these studies usually involve data with one particular maturity, mostly one month or one week. Naturally, an important question asks: would the results be consistent if we use the same exchange rate data but in various horizons? To answer the question, this study is initiated by employing the same period data to test the unbiasedness hypothesis for major currencies in various horizons, including one day, one week and one month.

A significant inconsistency has been found in the testing. The results of testing with the next day tomorrow maturity are found to differ strikingly from those obtained using longer horizons (1-week and 1-month). Specifically, in the horizon of 1-day, all the coefficients of the forward premium have correct signs (positive), and appear to be consistent with their theoretical value of unity across currencies. However, the testing in 1-week and 1- month horizons, just like the results reported in the pioneering studies, continues to generate significant negative coefficients. Meanwhile, it is also found that the slope coefficient tends to increase in magnitude as the horizon rises.

The previous efforts to reconcile the puzzle have been focusing on models' ability to explain why the slope coefficient is negative. Proposed theories claiming a certain element to be the cause of the puzzle are more convincing if it can be shown that in the circumstances that the puzzle does not exist, neither does the claimed cause. However, such circumstances have hardly been documented in previous literature, while this essay finds the testing in the 1-day horizon to be such a case and suggests that the puzzle is horizon-dependent, which provides a solid foundation to interrogate current models and direct further research. In these senses, these new findings are important to help us understand the old puzzle better.

The new empirical findings actually suggest a term structure of forward premium puzzle, while none of current standard models, at first glance, is compatible with the horizon-dependent property and able to offer an explanation for the apparently anomalous differences in the tests using a 1-day horizon versus 1-week or 1-month data. Given the fact that the forward premium is determined by forward rates, and the forward rates are primarily related to interest rate movements across countries, I extend a term structure model of interest rates to explain the term structure of forward premium puzzles.

Motivated by Bansal (1997), Backus *et al* (2001) and Ahn(2004), I build an exchange rate model based on an affine term structure model of interest rates. A preliminary inspection of data suggests that domestic and foreign yield rates are highly related across horizons, therefore each country's yield rates are assumed to be determined by one common factor (world factor) and one local factor, both with horizon-dependent coefficients. Then the forward premium can be expressed as a function of these latent factors based on Covered Interest Rate Parity. Meanwhile, the spot exchange return can be connected with these factors as well through pricing kernels (stochastic discount factors). Thus, the slope coefficient in the testing equation can be derived and an explanation for the new findings can be offered.

Theoretically the model shows that the specific value of the slope coefficient depends on the ratio of the common factor's impact on the risk premium over the difference between the sensitiveness of foreign and domestic yield rates to the common factor, while local factors have little effect. The model also shows that since the factor is more volatile in longer horizons, the risk premium for the common factor increases with the horizon as well. Thus, in a very short horizon like 1-day, the common factor is likely to contribute little risk to the spot exchange rates, so that the risk premium for the factor tends to be ignorable, and therefore the slope coefficient appears to be close to its theoretical value—one. On the contrary, in longer horizons like 1-week and 1-month, a significant risk premium is required for investors to assume exchange rate risk and the coefficient starts to deviate from one. Meanwhile, the investors in the country with higher sensitiveness to the common factor assume more exposure to exchange risk and require higher compensation, so that the numerator and denominator of the ratio always have opposite signs, which leads to the negative slope coefficient. An empirical exercise provides strong supportive evidence in favor of the theoretical explanations.

The contribution of this essay to the literature related to the forward premium puzzle is two-fold. First, it provides new empirical evidence that deepens the forward premium puzzle.

This evidence shows that the forward premium puzzle depends on the maturity of forward contracts. Second, the essay develops a framework based on a term structure model of the interest rate to explain the new puzzling phenomena from the perspective of factors' asymmetric impacts on domestic and foreign currency markets.

The rest of the essay is structured as below: section 2 reports new empirical evidence about the forward premium puzzle; section 3 examines the possibility of using current theories to explain new puzzles; section 4 presents a theoretical model; section 5 simplifies the model and offers explanations for what I found and section 6 tests the theoretical explanations empirically. Finally section 7 concludes.

2. Empirical findings

Suppose $f_{t,\tau}$ is the logarithm of the forward exchange rate with maturity τ at time t, and $s_{t+\tau}, s_t$ are the logarithms of spot rates at time $t+\tau$ and t respectively. A typical specification to test the efficiency of the forward market is:

$$s_{t+\tau} - s_t = \alpha + \beta(f_{t,\tau} - s_t) + \varepsilon_{t+\tau} \tag{3.1}$$

The previous estimates of equation (3.1) overwhelmingly reject the Unbiased Forward Rate Hypothesis (UFRH), i.e. α=0, β=1, across various time periods and currencies. More puzzling, the slope coefficient β is frequently found to be a significantly negative number. Since 1980s, a large number of papers have been devoted to reconciling the puzzle, e.g. Fama (1984), McCallum (1994), Cornell (1989), Bekaert and Hodrick (1991) and Lewis (1995). Among these works, it is noticeable that almost all data previously tested involve forward rates with one particular maturity, usually one week or one month. The curiosity about whether results are consistent for regressions in various horizons becomes my initial motivation to collect data and explore the puzzle from a new perspective.

The data employed in this essay are extracted from DataStream and contain dollar rates for several major currencies, including Canadian Dollar (CAD), Euro (EUR), and British Pound (GBP). Both spot and forward exchange rates were collected on daily basis for all these currencies, and the maturities of forward rates include next day tomorrow, 1-week, 1-month, 6-month and 12-month. CAD data cover a period between 10/27/1997 and 11/15/2004[1]; GBP data are between 12/29/97 and 2/11/2005; while EUR data start from 1/4/1999 and end at 11/15/2004. All rates on weekends are excluded as they are unavailable in DataStream either.

Given a 7-year coverage period for CAD, GBP and 5-year for EUR, only 7 or 5 independent observations are available for each currency to test UFRH based on one year maturity. Even for 6-month, the sample size of data is still too small to generate reliable estimations. The constraint on sample size forces us to focus only on the maturities of next day tomorrow, 1-week and 1-month. For 1-week testing, Mondays' rates are picked as the representatives of the matching week, while 1-month testing is based on the observations at the beginning of each month.

I use the data described above to estimate equation (3.1) through OLS for each currency in three horizons. Table 3.1 in appendix I reports the results of these regressions. The 3rd and 4th column of the table display the estimates for α, β with their t-statistics in parenthesis below, and null hypothesis for the t-statistics is $\alpha = 0, \beta = 1$. Surprisingly, the testing in 1-day horizon and longer-term (1-week and 1- month) obtain completely different results across

[1] Starting date of the data is the earliest date when next day tomorrow forward data are available on DataStream. Ending date is chosen just because I started this research late 2004 and collected CAD and EUR data on 11/15/2004, and GBP data later on 2/11/2005.

currencies. Specifically, 1-week or 1- month estimations, as usual, continue to generate significant negative results and reject UFRH consistently. In contrast, the estimates of β in the horizon of 1-day appear to be positive numbers, and furthermore, corresponding small t-statistics suggest the acceptance of UFRH for all the currencies I tested. Apparently, the forward premium puzzle depends on the maturity of forward contracts, and in this case, it exists in the horizon of 1-week or 1-month, but not 1-day.

Comparing the estimates of β between 1-week and 1-month horizons, I also find another interesting pattern: both estimates are negative, but the result of 1-month is always bigger than that of 1-week in magnitude. In other words, the longer the horizon is, the further away β deviates from its theoretical value--one.

Both these two findings have never been documented in the literature before. In order to check the robustness of these features, rolling regressions are conducted on the data. In the regressions, I pick a window period which is shorter than entire data coverage, starts from the first observation, regress the data within the window, get one result, then move the window one position forward and get another result, and this process continues until the end of window reaches the end of whole data coverage. Thus, multiple estimates can be obtained to see whether the new features reported before are consistent across periods. A 5-year window is picked for CAD and GBP, while a 3-year window for EUR due to Euro data's smaller sample size.

Table 3.2 documents rolling regression results. The first row of the table lists several continuous intervals in which most estimates of β fall. The number of the estimates lying in each interval is reported in the "frequency" row, and the "percentage" row records the ratio of total estimates in that interval. Based on the table, a majority of estimates in the 1-day horizon are found in positive intervals, specifically, 60.26% for CAD, 68.5% for EUR, and 100% for GBP appear to be positive results. The intervals with the largest percentage are [0, 1) and [1, 2), which contain 60.26% total estimates for CAD, 53% for EUR and 84% for GBP. Moreover, the further away from 1 the interval is, the fewer estimates it contains, which implies that the estimates in the horizon of 1-day converge to some point within a narrow interval around 1. In contrast, a majority of estimates in 1-week and 1-month horizons are still found to be negative numbers, and over 98% of total estimates are negative in almost all cases, and the only outlier, EUR in 1-week testing, is still 79.2%. According to the frequency distribution, the convergence of the estimates falls in the interval between -5 and -4 for CAD, less than -5 for EUR, between -4 and -3 for 1-week GBP, -3 to -2 for 1-month GBP.

Table 3.2 also demonstrates evidence to support the other new finding that the slope estimate is larger in magnitude in 1-month horizon than 1-week. CAD 1-week and 1-month testing both converge to the interval between -5 and -4, but compared to 34% of 1-week estimates falling in that interval, it contains 44.76% of total 1-month estimates. Similarly, both EUR 1-week and 1-month converge to the interval less than -5, which contains 59% of 1-month estimates but only 33.73% for 1-week. The only exception in our study is GBP, in which I did not find significant evidence to support this feature.

In summary, the two new findings reported before are proved to be robust phenomena based on the distribution of β displayed in table 3.2. An important question would be how to explain these new features. A solution is presented from both theoretical and empirical perspectives in the following sections.

3. Review of the existing theories

To explain the new empirical findings noted in the previous section, an attempt was made first to find out whether the existing theories of the forward premium puzzle have such a capability. Current explanations include: 1) time varying risk premium, 2) irrational expectations, 3) learning problem, 4) endogenous interest rates or monetary policy, 5) market microstructure explanations, 6) measurement error or misspecification, and 7) econometric artifact. Generally, all these theories tend to show that under certain conditions, the slope coefficient becomes negative. To be consistent with our findings, a successful theory should be able to show that the conditions causing the coefficient to be negative exist in the horizon of 1-week or 1-month, but not 1-day.

1) The attempt to solve the original puzzle began with the reconsideration of the conditions that UFRH is based on. Relaxing the assumption of risk neutrality, Fama (1984) reconciled the anomalous results by considering the existence of a risk premium. The argument is that risk-averse investors require compensation to assume risk, thus relationship $E_t\, S_{t+k} = F_{t,k}$ does not hold, and instead, we have:

$$E_t\, s_{t+k} = f_{t,k} + rp_t$$

where rp_t is the risk premium. It can be shown that the slope coefficient (β) being negative requires that the risk premium and expected exchange rate change are negatively correlated and the former has a higher variance than the latter. The real challenge for this theory, however, is to build a model that can generate the risk premium satisfying those conditions, which, unfortunately, as concluded by Engle (1996), appears to be unsuccessful in the current literature.

2) Another possible explanation of the puzzle can be offered by relaxing the other necessary condition on which the UFRH is based --rational expectations. The argument of this theory is that systematic errors in expectations introduce bias into the forward forecast error, which causes the coefficient to be negative. Gourinchas and Tornell (2004), for instance, introduced a particular distortion in the investors' belief about the future interest rates and assumed that the investors misperceive the relative importance of transitory and persistent interest rate shocks. They showed that the puzzle appears if the investors overestimate the importance of the transitory shocks relative to the persistent shocks, or equivalently, under-react to the interest rate innovations. According to this logic, our empirical findings require that the investors misperceive the interest rate innovations in longer horizons but not in short horizons like one day, which is likely but needs further investigation.

3) A learning process is another popular explanation of the puzzle (e.g. Lewis (1995)). This theory notices that when regime switches occur in the foreign exchange market, it takes time for agents to update their information and realize the change. The model claims that if over half of market agents make their expectation based on the old system, then the covariance of the spot exchange rate change and the forward premium becomes negative, which leads to the negative coefficient β. To be the solution to our problems, the theory should justify that on the weekly or monthly basis over half agents make expectation based on the old system, but on the daily basis less than half do so. Intuitively, if there is a regime switch, agents are more likely to realize the switch within one week or one month rather than one day. How come the learning problem exists in longer horizons but not in shorter ones? Apparently, this theory is questionable to become the answer to our problem.

4) McCallum (1994) proposed a solution to the puzzle from the perspective of monetary policy. The model argues that there is a monetary policy rule making the interest rate differential endogenous and adjusted to exchange rate changes, and given the parameters of the policy rule lying in some certain range, the slope coefficient β will be negative. To be a successful theory in our case, the model is supposed to show that this policy rule works on the horizon of one week or one month, but not on one day. This claim seems likely but requires further work to verify such a situation.

5) McBrady (2005) proposed a new explanation for the puzzle from the perspective of market microstructure of the FX market. The paper argues that the UIP is derived based on the arbitrage, while the arbitrage process is not as simple as it seems in the actual market. "Carry traders", the real world arbitrageurs who try to profit from differences in interest rates, do not borrow in one currency and lend in another as the standard UIP requires. They finance positions in the Repo markets, earning benchmark bond returns and paying interest at local Repo rates. And also, as the counterpart of this arbitrage transaction in the FX market, FX dealers can hedge currency risk and earn interest premiums for the currency they sell by pre-funding the deliveries they need to make two days later. Therefore, the standard regressions are mis-specified, and both interest rate differential and expected benchmark holding period returns should be included. Including the expected benchmark holding period returns, new empirical testing provides positive coefficients for the interest rate differentials. To explain what I found on this track, the carry traders and FX dealers are expected to behave differently in different horizons, and this claim requires further verification.

6) Several papers such as Cornell (1989), Bekaert and Hodrick (1991) and Breuer (1996) also mention that data quality and measurement error could also be the factors contributing to

the negative result of β. Usually those problems involve the existence of spreads, end of month effects, delivery convention problems etc. It turns out that these factors can only explain the puzzle partially, and even after correction to these errors, β still appears to reject UFRH, although the results are less negative. Therefore, it is not likely that we can find answers to our problem on this track.

7) Recently, the development in time series econometrics adds a new perspective to the solutions of the puzzle. In addition to the well known fact that regressing two completely unrelated I(1) processes could generate significant results, spurious results can also be found in the regression of two completely unrelated fractionally integrated process denoted by I(d), $0<d<1$ (See Tsay and Chung (2000) and Marmol (1998)). What relates our problem to this econometric finding is the statistical property of the forward premium and changes in spot rates. According to Baillie and Bollerslev's estimations (Baillie and Bollerslev (1994)), the forward premium is a nonstationary fractionally integrated process with d value between 0.5 and 1. Meanwhile, many studies highlight the fact that the change in spot exchange rate follows a nearly I(0) process. So the question is when we regress a I(0) process on a I(d) process by OLS, which is the scenario for testing the unbiasedness hypothesis of forward rates, is it possible that the negative slope coefficients are just spurious effects of regression due to some particular econometric features that the data possess? Baillie and Bollerslev (2000) as well as Maynard and Phillips (2001) make some advancements in this direction, and draw a major conclusion that the forward puzzle might be just a statistical artifact caused by the unbalance of integration order in both sides of the testing equation.

Baillie and Bollerslev's conclusion is drawn from simulations. In their paper, the mean value of spot rates are generated by an Auto Regressive Integrated Moving Average model

(ARIMA(1,1,1)). A specification of Fractionally Integrated GARCH (FIGARCH(1,d,0)) with d=0.75 generates the variance of daily spot rates. Then, the forward rates are created by a consumption-based UIP equation, which contains the variance of spot rates, so that forward rates possess the property of long memory too. The model was calibrated by reasonable parameters and regressed repeatedly. The results of the simulation show that with shorter samples these regressions typically generate slope coefficient estimates that are very widely dispersed. Hence the empirical results are suggestive of slow convergence to the true parameter value of one. The anomaly might be viewed as a statistical artifact from having small sample size and persistent autocorrelation in the forward premium.

Maynard and Phillips confirm the evidence of non-stationary long-memory behavior in the forward premium based on several estimating models, including an Auto Regressive Fractionally Integrated Moving Average (ARFIMA) specification. This property implies an imbalance of integration order in both sides of the testing equation, which regresses spot return (I(0) process) on the forward premium (I(d) process). Furthermore, they derived a non-standard limiting distribution for the OLS estimator of coefficient β in terms of stochastic integrals of fractional Brownian motion, which shows that the estimated β converges to zero, but the distribution might be skewed to the left (negative region). Accordingly, they argue that the puzzle is caused by the unbalanced regression, and therefore the negative β might be just a statistical artifact.

Applying this theory in our empirical findings, the regression equation should be balanced in the testing of next day tomorrow forward rates, while unbalanced for 1-week or 1-month testing. In other words, the forward premium should be a I(0) process with the horizon of 1-day, but a I(d) process in 1-week or 1-month horizon. To verify the statement above, a regular ARFIMA model is employed to estimate the integration order d for forward premium series with various horizons. The general specification of ARFIMA (p, d, q) can be written as:

$$\phi(L)(1-L)^d y_t = \mu + \theta(L)\varepsilon_t$$

where y_t is the long memory series and ε_t is white noise, $\phi(L)$ is the autoregressive coefficients polynomial with *p* lags, $\theta(L)$ is a moving average coefficients polynomial with *q* lags, *d* is the partial integration order, μ is the unconditional mean value of the series y_t, and *L* is the lag operator.

Based on the Akaike Information Criteria, the specifications fitting the data best are found to be ARFIMA(0,d,0) and ARFIMA(1,d,0), which can be written as:

$$(1-L)^d (f_t - s_t) = \mu + \varepsilon_t$$

and

$$(1-L)^d (f_t - s_t) = \mu + \phi_1 (1-L)^d (f_{t-1} - s_{t-1}) + \varepsilon_t$$

The forward premium $(f_t - s_t)$ is the series we focus on and other symbols are defined as before. The models are estimated by the Maximum Likelihood Method, and the estimation statistics are computed based on Baillie (1996).

Table 3.3 reports the estimation results, and according to the table, integration order *d* of the 1-day forward premium is close to zero across currencies, which suggests the series to be a stationary *I(0)* process. On the other hand, *d* is over 0.5 for the 1-week premium, and even close to 1 for 1-month, which makes both series non-stationary *I(d)* processes. The estimation results are completely consistent with the hypothetic analysis, which makes this theory very promising to become a winner so far.

However, the distribution of estimated coefficients obtained in the rolling regressions raises the first objection to the theory. Disregarding the slow rate of convergence with small sample sizes, the estimated coefficient still converges to one in Baillie and Bollerslev (2000)'s simulation. In the non-standard limiting distribution derived by Maynard and Phillips (2001), the OLS estimator of β converges to zero, although with an asymmetric distribution skewed to the negative region. However, the truth, which can be easily seen from the distribution of the slope coefficients displayed in table 3.2, is that the estimated β in 1-week or1-month testing converges neither to zero, nor one; instead, it tends to be distributed around some negative number across currencies.

Furthermore, the theory actually implies that as long as two time series are characterized with econometric features mentioned before, the regression of these two series will obtain negative slope coefficients or make the distribution of estimated coefficients skewed to negative region. A simulation experiment is designed in this essay to test whether it is true. In the simulation, a I(0) process is created by a ARMA(1,1)-GARCH(1,1) model as the proxy of the change in the spot exchange rate, and a ARFIMA(0,d,0)-GARCH(1,1) specification generates a I(d) (0.5<d<=1) process representing the 1-week or 1- month forward premium.

Simulating an ARFIMA(0,d,0)-GARCH(1,1) process starts with the variance part, and the error item ε_t can be generated through a standard GARCH(1,1) model first. Then for the mean part, a time series y_t governed by ARFIMA(0,d,0) with mean value μ can be written as:

$$(1-L)^d(y_t-\mu)=\varepsilon_t$$

which can be rewritten as:

$$y_t = \sum_{k=0}^{\infty} \varphi(k)\varepsilon_{t-k} + \mu$$

where $\varphi(k) = \dfrac{\Gamma(k+d)}{\Gamma(d)\cdot\Gamma(k+1)}$ and $\Gamma(\cdot)$ is the gamma function. Thus, given the time series of the error item ε_t, y_t can be computed through the above equations based on the truncation number of 100. The coefficients of the specification are calibrated based on the parameters estimated from the real data of CAD/USD, and eventually two simulated processes are regressed by OLS to obtain the value of the slope.

This procedure is done repeatedly based on various sample sizes and values of d, and the distribution of estimated coefficients is displayed in each sub-figure of Figure 3.1. As illustrated in the figure, the coefficient consistently converges to zero no matter what value of d or sample size is. More importantly, although the estimates are distributed diversely when the sample size is small, there is no asymmetric distribution of β skewed to the negative region in any case I tested. The result puts another serious doubt on the conclusion that the puzzle is just a statistical artifact, and makes the econometric explanation to the puzzle questionable.

4. Theoretical framework

The stark difference between the testing results using short (one day) - versus medium-horizon (one week or one month) data becomes an old puzzle's new puzzle. The empirical findings in section 2 actually suggest the existence of a term structure of forward premiums. This horizon-dependent property possessed by the forward premium apparently results from the forward rates, which are associated with various maturities. Since the forward rates are primarily related to the interest rate movements across countries[1], the term structure of the forward rates must be also related to the term structure of interest rates. Considering a large number of models dealing with the term structure of interest rates in the literature, there is a possibility of extending those models to explain the term structure of forward premiums. In this sense, although some previous explanations have potential to reconcile this inconsistency, this essay follows the term structure track in the following sections.

Bansal (1997) as well as Backus *et al* (2001) started using the affine term structure model of interest rates to explain the forward premium puzzle. Motivated by these earlier works, I explored the possibility by building an exchange rate model based on this type of term structure model as well. Backus *et al* (2001) is the closest in the literature to my work, but several aspects distinguish my work from the prior work. First, Backus *et al*'s paper tries to explain why the slope coefficient is negative, while mine explains why the testing results are inconsistent in different horizons. Second, Backus *et al*'s interdependent model contains two related factors and assumes that the impact of the domestic factor on foreign yield rates and that of the foreign factor on domestic yield rates are the same, while my model contains one common factor and two local factors and does not have strict restrictions on the coefficients of the factors. Third, Backus *et al*'s paper assumes that the latent factors follow square root processes and uses GMM to estimate the model, while mine avoids the complication of assuming the dynamics of the latent factors, extracts the factors first from interest rate data and then uses a regular regression method to estimate the model. And finally, Backus *et al*'s interdependent model does not obtain successful empirical evidence to explain the negative slope coefficients, while my model is supported empirically in this essay.

[1] Actually banks price the forward contract based on the CIP, which specifically describes the relationship between the forward rates and the interest rates differential

The framework for the affine term structure models is set out in Duffie and Kan (1996) and summarized in Dai and Singleton (2000). Generally, these models argue that interest rates are determined by several latent state variables. A preliminary examination of interest rate data suggests that domestic and foreign yields are highly correlated in all horizons (1-day, 1-week and 1-month). Therefore I assume that each country's interest rate is determined by a common (world) factor[1] and an individual local factor. Coefficients of these factors represent yields' sensitiveness to each factor, and as a term structure model, they are horizon dependent. Thus, domestic and foreign nominal interest rates with maturity τ at time t $i_{\tau,t}, i^*_{\tau,t}$ can be written as:

$$i_{\tau,t} = \alpha_{\tau,1} \cdot Z_{1,t} + \alpha_{\tau,2} \cdot Z_{2,t} \tag{3.2}$$

$$i^*_{\tau,t} = \alpha^*_{\tau,1} \cdot Z_{1,t} + \alpha^*_{\tau,2} \cdot Z^*_{2,t} \tag{3.3}$$

where $Z_{1,t}$ is the common factor, $Z_{2,t}, Z^*_{2,t}$ are the domestic and foreign local factor respectively. $\alpha_{\tau,1}, \alpha_{\tau,2}$, and $\alpha^*_{\tau,1}, \alpha^*_{\tau,2}$ are horizon-dependent coefficients. When the horizon is extremely short, $\alpha_{\tau,1}$, $\alpha^*_{\tau,1}$ and $\alpha^*_{\tau,2}$ are assumed to become one and $\alpha_{\tau,2}$ equals ρ, i.e. domestic and foreign short rates i_t, i^*_t have the following specification:

$$i_t = Z_{1,t} + Z_{2,t} \tag{3.4}$$

$$i^*_t = \rho Z_{1,t} + Z^*_{2,t} \tag{3.5}$$

Suppose $R^*_{\tau,t}, R_{\tau,t}$ are the return rates of foreign and domestic risk free assets with maturity τ at time t, and $r^*_{\tau,t}, r_{\tau,t}$ are their logarithm forms, the logarithm format of Covered Interest Rate Parity (CIP) gives the equation below:

[1] It could contain multiple common factors, but factor analysis in the next section will show that one major common factor is sufficient to capture the dynamics shared by both countries' interest rates.

$$f_{\tau,t} - s_t = r^*_{\tau,t} - r_{\tau,t} = \log R^*_{\tau,t} - \log R_{\tau,t} \tag{3.6}$$

Given that nominal interest rates $i_{\tau,t}, i^*_{\tau,t}$ are measured as decimal numbers, which should usually be less than 0.1, return rates $R^*_{\tau,t}, R_{\tau,t}$ can be written as $R_{\tau,t} = 1 + i_{\tau,t}$, $R^*_{\tau,t} = 1 + i^*_{\tau,t}$. Using the approximation that $\log(1+x) \approx x$ if x is a small number (usually less than 0.1), we have:

$$\log R_{\tau,t} = \log(1 + i_{\tau,t}) \approx i_{\tau,t} \tag{3.7}$$

$$\log R^*_{\tau,t} = \log(1 + i^*_{\tau,t}) \approx i^*_{\tau,t}, \tag{3.8}$$

Plugging equations (3.7) and (3.8) into equation (3.6) gets:

$$f_{\tau,t} - s_t = i^*_{\tau,t} - i_{\tau,t} = (\alpha^*_{\tau,1} - \alpha_{\tau,1}) \cdot Z_{1,t} + \alpha^*_{\tau,2} \cdot Z^*_{2,t} - \alpha_{\tau,2} \cdot Z_{2,t} \tag{3.9}$$

Thus, the logarithm of the forward premium, the left hand side of equation (3.1), is decomposed as a linear function of the factors in equation (3.9). The right hand side of the equation—the change of logarithmic spot exchange rate needs to be connected to these factors too to derive an expression for β, while this connection can be made through pricing kernels (k). Based on the standard relationship between yield rates and the pricing kernels, i.e. $i_{t,\tau} = -\log(E_t\, k_{t+\tau})$, to have nominal interest rates specified as in equations (3.2) and (3.3), domestic and foreign pricing kernels would be:

$$-\log k_{t+\tau} = (\alpha_{\tau,1} + \lambda^2_{\tau,1}/2) \cdot Z_{1,t} + (\alpha_{\tau,2} + \lambda^2_{\tau,2}/2) \cdot Z_{2,t} + \lambda_{\tau,1} \cdot \sqrt{Z_{1,t}}\,\varepsilon_{1,t} + \lambda_{\tau,2} \cdot \sqrt{Z_{2,t}}\,\varepsilon_{2,t} \tag{3.10}$$

$$-\log k^*_{t+\tau} = (\alpha^*_{\tau,1} + \lambda^{*\,2}_{\tau,t}/2) \cdot Z_{1,t} + (\alpha^*_{\tau,2} + \lambda^{*\,2}_{\tau,2}/2) \cdot Z^*_{2,t} + \lambda^*_{\tau,1} \cdot \sqrt{Z_{1,t}}\,\varepsilon^*_{1,t} + \lambda^*_{\tau,2} \cdot \sqrt{Z^*_{2,t}}\,\varepsilon^*_{2,t} \tag{3.11}$$

Here $\varepsilon_{1,t}, \varepsilon_{2,t}$ $\varepsilon^*_{1,t}, \varepsilon^*_{2,t}$ are normal random variables with zero mean and unit variance and $\lambda_{\tau,1}, \lambda_{\tau,2}$ $\lambda^*_{\tau,1}, \lambda^*_{\tau,2}$ are horizon dependent constant coefficients for error items in equations (3.10) and (3.11). The derivation can be verified easily by taking the expected value on both sides of the pricing kernel equations. Meanwhile, both Bansal (1997) and Backus *et al* (2001) (Proposition 1) have proved that the change of the log spot exchange rate is the differential between the foreign and domestic log pricing kernels, such that

$$s_{t+\tau} - s_t = \log(k^*_{t+\tau}) - \log(k_{t+\tau}) \tag{3.12}$$

Therefore, plugging equations (3.10) and (3.11) into equation (3.12) gives the expected log spot exchange rate return:

$$E_t(s_{t+\tau} - s_t) = E_t \log(k^*_{t+\tau}) - E_t \log(k_{t+\tau}) = f_{\tau,t} - s_t + \frac{({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)}{2} \cdot Z_{1,t} + \frac{{\lambda^*_{\tau,2}}^2}{2} \cdot Z^*_{2,t} - \frac{{\lambda_{\tau,2}}^2}{2} \cdot Z_{2,t} \tag{3.13}$$

Apparently, the items after $f_{\tau,t} - s_t$ on the right hand-side of equation (3.13) represent the risk premium of the spot exchange rate. Combining equations (3.9) and (3.13) can give an OLS estimate of the slope coefficient of equation (3.1) as below:

$$\beta = 1 + \frac{({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)(\alpha^*_{\tau,1} - \alpha_{\tau,1}) \cdot V(Z_1) + {\lambda^*_{\tau,2}}^2 \cdot V(Z^*_2) + {\lambda_{\tau,2}}^2 \cdot V(Z_2)}{2(\alpha^*_{\tau,1} - \alpha_{\tau,1})^2 V(Z_1) + {\alpha^*_{\tau,2}}^2 \cdot V(Z^*_2) + {\alpha_{\tau,2}}^2 \cdot V(Z_2)} \tag{3.14}$$

It is obvious that equation (3.14) can generate both negative and positive β, and the key items are the impacts of the common factor on the risk premium and the interest differential, i.e. $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$. However, equation (3.14) is still not able to provide a clear image about the explanations for our findings, and more details about the factors are required.

5. Factor analysis and model simplification

In order to examine the properties of the factors and estimate the model, the factors that are specified in equation (3.2) through (3.5) need to be extracted from interest rate data. Various methods could be applied to conduct this exercise, and Principal Factor Analysis (PFA) is employed in this essay. Compared to other popular latent factor estimation techniques such as Markov Chain, Monte Carlo and Kalman filter, the PFA method avoids the complication of assuming the dynamics of the latent factors so that factors can be recovered regardless of the time series process the factors follow. In addition, the way of retrieving the factors by PFA is economically in line with the model.

Following the factor extraction method applied in Ahn (2004), the factor analysis model is set up as below:

$$\underbrace{Y_t}_{6\times T} = \underbrace{B}_{6\times 3} \cdot \underbrace{F_t}_{3\times T} + \underbrace{\varepsilon_t}_{6\times T} \tag{3.15}$$

$$\underbrace{Y_t}_{6\times T} = [\underbrace{i_{o/n,t}}_{T\times 1}, \underbrace{i_{1-week,t}}_{T\times 1}, \underbrace{i_{1-month,t}}_{T\times 1}, \underbrace{i^*_{o/n,t}}_{T\times 1}, \underbrace{i^*_{1-week,t}}_{T\times 1}, \underbrace{i^*_{1-month,t}}_{T\times 1}]'$$

$$\underbrace{F_t}_{3\times T} = [\underbrace{Z_{1,t}}_{T\times 1}, \underbrace{Z_{2,t}}_{T\times 1}, \underbrace{Z^*_{2,t}}_{T\times 1}]'$$

where Y_t is a vector of nominal interest rates formed by vertically stacking domestic (U.S.) and foreign overnight, 1-week and 1-month interest rates and its covariance matrix is $\Sigma_{6\times 6}$. B is the factor loading matrix, and F_t is a vector of the common, domestic and foreign local factors. T is the number of observations. The random factors F_t are assumed to have zero mean, i.e. $E(F_t) = 0$ and identity variance, i.e. $Cov(F_t) = I_3$.

There are three steps to extract the factors, and the first step is to obtain initial principal factors and the loading matrix without worrying about their economic implications. The covariance matrix of observable variables Y_t can be decomposed into the product of its eigenvalues and matching eigenvectors such that:

$$\Sigma_{6\times 6} = P \cdot Q \cdot P' \tag{3.16}$$

where Q is a diagonal matrix with eigenvalues $\Lambda_i, i = 1,2,...6$ and P is the matrix containing corresponding eigenvectors $\ell_i, i = 1,2,...6$. Meanwhile, based on equation (3.15) and the assumption about the variance matrix of factors, the covariance matrix of interest rates can also be written as:

$$\Sigma_{6\times 6} = B \cdot B'Cov(F) = B \cdot B' \tag{3.17}$$

Thus the equivalence of equations (3.16) and (3.17) gives the initial loading matrix as below:

$$B = P \cdot \sqrt{Q}$$

To be compatible with the model's three-factor set-up, I only take the three largest eigenvalues $\Lambda_1, \Lambda_2, \Lambda_3$ and their matching eigenvectors ℓ_1, ℓ_2, ℓ_3, so that the loading matrix becomes:

$$B = [\sqrt{\Lambda_1}\ell_1, \sqrt{\Lambda_2}\ell_2, \sqrt{\Lambda_3}\ell_3]$$

and initial extracted factors are:

$$F_t = [\frac{\ell_1}{\sqrt{\Lambda_1}}, \frac{\ell_2}{\sqrt{\Lambda_2}}, \frac{\ell_3}{\sqrt{\Lambda_3}}]' \cdot Y_t$$

Since the eigenvectors ℓ_1, ℓ_2, ℓ_3 are not unique, so that there could be an infinite number of extraction results, as the second step, the initial extraction should be rotated to obtain more economically meaningful factors, i.e. the common factor, domestic and foreign local factor. Suppose an orthonormal matrix *T* such that $TT' = I_3$, then $BF = BT \cdot T'F$, where the loading matrix after rotation B^R is BT and the factors after rotation F^R are $T'F$. An ideal rotation *T* should make a local factor of one country ineffective on the other country, such that $B^R[1:3,3] = 0$ and $B^R[4:6,2] = 0$, where [*i*:*k* , *j*] denotes the elements of the *i*th through *k*th row and *j*th column of loading matrix. Such a rotation matrix *T* can be obtained by solving the following optimization problem:

$$MinB^R[1:3,3]' \cdot B^R[1:3,3] + B^R[4:6,2]' \cdot B^R[4:6,2] \text{ subject to } TT' = I_3$$

The last step is to rescale the loading matrix and factors. By construction, the short rates should satisfy equations (3.4) and (3.5), i.e. the sensitiveness of the domestic short rate to the common factor and the domestic local factor should be one, and the sensitiveness of the foreign short rate to the foreign local factor should be one as well. I pick overnight interest rates as the proxy for the short rates, and rescale rotated loading matrix in the following way:

$$B^{R,R}[1:6,1] = B^R[1:6,1] / B^R[1,1],$$

$$B^{R,R}[1:6,2] = B^R[1:6,2] / B^R[1,2],$$

$$B^{R,R}[1:6,3] = B^R[1:6,3] / B^R[4,3].$$

Through these three steps, interest rates can be decomposed into a linear function of three factors (retrieved in the rotated factor matrix F^R) with loading matrix $B^{R,R}$. Interest rate data applied in this factor analysis are LIBOR overnight, 1-week and 1-month, which are extracted from DataStream as well with the same coverage period as the exchange rates data[1].

Table 3.4 documents the results of factor extraction. Panel A of the table displays the eigenvalues $\Lambda_1, \Lambda_2, \Lambda_3$ and the variances of each factor, and panel B reports the total explanatory power of the factors and the contribution of each factor to the domestic and foreign interest rates. Between the U.S. and each foreign country, only one principal factor's eigenvalue is found to be greater than 1, which is usually the criterion of determining the number of the common factors. Moreover, this factor's eigenvalue ranges from 6.8 to 17.34

[1] LIBOR overnight data are available only from 1/2/2001 in DataStream. So for EUR, the factor extraction exercise has been done only between 1/2/2001 to 11/15/2004, GBP is between 1/2/2001 to 2/15/2005, and for CAD, I found the overnight rate from the website of Bank of Canada as a proxy of LIBOR between 10/27/1997 to 1/2/2001 so that CAD testing is conducted between 10/27/1997 to 11/15/2004.

and dominates over the other two factors, whose eigenvalues are less than 0.5 or even smaller. According to panel B, the common factor is also able to capture most of the dynamics of yields in the two countries. It contributes over 97% to the total variance of US yields, over 90% for Canadian yields, over 54% for UK and around 66% for Euro yields. These results suggest that one common factor would be sufficient to be able to capture the dynamics shared by two countries' yield rates without loosing much explanatory power, which justifies our one-common-factor set-up of the model.

We have reasons to believe that the common factor actually refers to U.S. economic conditions and monetary policy, which are known to have profound and global effects on other countries' currency markets and therefore becomes a world factor. On the other hand, the contribution of this factor to foreign yields indicates how closely the two countries' monetary policy are related and how greatly each country's currency market is affected by the U.S.. According to table 3.4, Canada apparently appears to have a closer economic relationship with the U.S. than the Euro zone and UK, which is consistent with intuition.

More importantly, the common factor is found to be much more volatile than the other two local factors according to panel A of table 3.4, and the variance of the common factor is a much bigger magnitude than that of local factors. This discovery reveals an important direction to simplify the model. Recall expression of β (equation (3.14)) obtained in previous section:

$$\beta = 1 + \frac{({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)(\alpha^*_{\tau,1} - \alpha_{\tau,1}) \cdot V(Z_1) + {\lambda^*_{\tau,2}}^2 \cdot V(Z^*_2) + {\lambda_{\tau,2}}^2 \cdot V(Z_2)}{2(\alpha^*_{\tau,1} - \alpha_{\tau,1})^2 V(Z_1) + {\alpha^*_{\tau,2}}^2 \cdot V(Z^*_2) + {\alpha_{\tau,2}}^2 \cdot V(Z_2)}$$

It suggests that the volatility of each factor reflects the relative importance of the factor in determining the value of β. Dividing the variance of the common factor in both numerator and denominator of the second item in the above equation gets:

$$\beta = 1 + \frac{({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)(\alpha^*_{\tau,1} - \alpha_{\tau,1}) + {\lambda^*_{\tau,2}}^2 \cdot \frac{V(Z^*_2)}{V(Z_1)} + {\lambda_{\tau,2}}^2 \cdot \frac{V(Z_2)}{V(Z_1)}}{2(\alpha^*_{\tau,1} - \alpha_{\tau,1})^2 + {\alpha^*_{\tau,2}}^2 \cdot \frac{V(Z^*_2)}{V(Z_1)} + {\alpha_{\tau,2}}^2 \cdot \frac{V(Z_2)}{V(Z_1)}} \tag{3.18}$$

Due to the dominance of the common factor in terms of volatility, the ratio of the variance of local factors relative to the common factor are small, specifically, around 0.025 for CAD, 0.01 for Euro, 0.125 for GBP, and therefore become ignorable. If $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ is significantly different than zero, which will be verified empirically in the next section, equation (3.18) can be simplified as:

$$\beta = 1 + \frac{({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)}{2(\alpha^*_{\tau,1} - \alpha_{\tau,1})} \tag{3.19}$$

Equation (3.19) implies several important theoretical conclusions about the model. First, the value of β depends on the common factor's impacts on the exchange rate risk premium and the interest rate differential, i.e. $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$, while local factors have little effect. Second, whether UFRH is accepted depends on whether the premium compensated for the exchange rate risk resulting from the common factor is significantly different from zero, i.e. $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2) \neq 0$. The puzzle can not be seen if investors require little compensation for the risk brought about by the common factor, otherwise, the values of β would deviate from one and UFRH would be rejected. Third, since all coefficients in this ratio are horizon-dependent, therefore the value of β should have the same property as well, which is compatible with the term structure framework. Clearly, an appropriate choice of parameter values allows us to generate both negative and unity values for β .

Based on the model's theoretical implications discussed above, an explanation for the empirical findings reported in section 2 can be offered[1]. The most striking finding in section 2 is that the old forward premium puzzle does not exist in the 1-day horizon, which suggests that the risk premium for the common factor is ignorable in such a short horizon, i.e. $(\lambda_{\tau,1}^{*\,2} - \lambda_{\tau,1}^{2}) = 0$. Intuitively, because the factors determining the exchange rates are less volatile in shorter horizons, the spot exchange rate is less risky as the horizon decreases, and therefore a lower risk premium is required. If the horizon is short enough, the risk caused by the common factor is likely to become ignorable and investors require little risk premium for the common factor[2], thus, the puzzle cannot be seen in the 1-day horizon.

In the 1-week or 1-month testing, horizons are long enough that the risk premium has to be compensated i.e. $(\lambda_{\tau,1}^{*\,2} - \lambda_{\tau,1}^{2}) \neq 0$. Thus, the value of β certainly would deviate from its theoretical value—one. The question is why it always tends to be a negative number? Suppose the domestic yield rate is more (less) sensitive to the common factor than its foreign counterpart, so that $\alpha_{\tau,1}^{*} - \alpha_{\tau,1} < 0$ ($\alpha_{\tau,1}^{*} - \alpha_{\tau,1} > 0$); then domestic investors assume more (less) exposure to the risk caused by the common factor than foreign investors, and therefore certainly require a higher (lower) risk premium. According to equation (3.13), the risk premium for the common factor is $\frac{(\lambda_{\tau,1}^{*\,2} - \lambda_{\tau,1}^{2})}{2} \cdot z_{1,t}$. Since the spot exchange rates used in this essay are U.S. dollar rates, i.e. measured in units of foreign currency per unit of domestic

[1] It should be noted that this is just a tentative theoretical explanation, which is subject to empirical verification, and the empirical testing in the next section provides supportive evidence

[2] This does not mean that there is no risk in short horizons, and the risk might come from local factors and other variables that are not included in the model.

currency (USD), a higher (lower) risk premium for the domestic investor means a positive (negative) value for $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$, i.e. $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2 > 0$ ($({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2 < 0$). Thus, ${\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2$ and $\alpha^*_{\tau,1} - \alpha_{\tau,1}$ always have opposite signs, which leads to a negative slope coefficient in the testing. It should be noted that ${\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2$ and $\alpha^*_{\tau,1} - \alpha_{\tau,1}$ having opposite signs is just a necessary condition for the negative results; to be sufficient, the magnitude of $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ still needs to be higher than $\alpha^*_{\tau,1} - \alpha_{\tau,1}$, which cannot be seen directly from equation (3.19) but will be verified in the empirical testing.

As the horizon increases, naturally, the common factor becomes more volatile, and the spot exchange rate is more risky. Therefore, a higher risk premium is required, and the magnitude of $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ increases, which causes β to increase in magnitude as well. This can be considered as the explanation for the phenomena that the 1-month result is greater than 1-week in magnitude.

6. Empirical testing

The simplified model suggests that the slope coefficient we are concerned about depends on the ratio $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2(\alpha^*_{\tau,1} - \alpha_{\tau,1})$. To be consistent with the testing results in the 1-day horizon, the value of the ratio should be zero at the short horizons, which requires both an insignificant compensation for the exchange rate risk caused by the common factor (i.e. $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2) = 0$) and an asymmetric impact of the common factor on domestic and foreign yield rates (i.e. $(\alpha^*_{\tau,1} - \alpha_{\tau,1}) \neq 0$). In 1-week or 1-month testing, to generate a significant negative slope coefficient, $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ are expected to possess opposite signs and the magnitude of the first item is greater than that of latter. The task in this section is to inspect whether coefficients $\alpha^*_{\tau,1}, \alpha_{\tau,1}, \lambda^{*\,2}_{\tau,1}, \lambda_{\tau,1}^2$ are consistent with these theoretical expectations.

According to the tests described above, estimating the items $(\lambda^{*\,2}_{\tau,1} - \lambda_{\tau,1}^2)/2$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ is sufficient to conduct this test without requiring an estimation strategy to obtain each individual value of those coefficients. As shown in equation (3.9), $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ is the difference between domestic and foreign yields' sensitiveness to the common factor. Therefore it can be estimated through the equation below:

$$f_{\tau,t} - s_t = \theta_{\tau,1} \cdot Z_{1,t} + \theta_{\tau,2} \cdot Z^*_{2,t} + \theta_{\tau,3} \cdot Z_{2,t} + \varepsilon_t \tag{3.20}$$

where $f_{\tau,t} - s_t$ is the logarithm of forward premium, and $Z_{1,t}, Z^*_{2,t}, Z_{2,t}$ are the factors that have been extracted in the previous section. $\theta_{\tau,1}, \theta_{\tau,2}, \theta_{\tau,3}$ are horizon-dependent coefficients. Obviously, $\theta_{\tau,1}$ is an unbiased estimate of $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$.

Moving the forward premium to the left side in equation (3.13) gives:

$$E_t\, s_{t+\tau} - f_{\tau,t} = \frac{(\lambda^{*\,2}_{\tau,1} - \lambda_{\tau,1}^2)}{2} \cdot Z_{1,t} + \frac{\lambda^{*\,2}_{\tau,2}}{2} \cdot Z^*_{2,t} - \frac{\lambda_{\tau,2}^2}{2} \cdot Z_{2,t}$$

Thus, the other key item in the test $(\lambda^{*\,2}_{\tau,1} - \lambda_{\tau,1}^2)/2$ can be obtained by estimating the following specification:

$$s_{t+\tau} - f_{\tau,t} = \gamma_{\tau,1} \cdot Z_{1,t} + \gamma_{\tau,2} \cdot Z^*_{2,t} + \gamma_{\tau,3} \cdot Z_{2,t} + \eta_t \tag{3.21}$$

where $\gamma_{\tau,1}, \gamma_{\tau,2}, \gamma_{\tau,3}$ are horizon-dependent coefficients, and apparently coefficient $\gamma_{\tau,1}$ is an estimate of $(\lambda^{*\,2}_{\tau,1} - \lambda_{\tau,1}^2)/2$.

In the 1-day horizon, we only need to test whether $(\lambda^{*\,2}_{\tau,1} - \lambda_{\tau,1}^2)/2 = 0$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1}) \neq 0$ hold simultaneously, so a t-test can be used directly when equations (3.20) and (3.21) are estimated. To test whether $\beta < 0$ in 1-week and 1-month horizons, β can be rewritten as:

$$\beta = \frac{(\alpha_{\tau,1}^{*} - \alpha_{\tau,1}) + (\lambda_{\tau,1}^{*}{}^{2} - \lambda_{\tau,1}{}^{2})/2}{(\alpha_{\tau,1}^{*} - \alpha_{\tau,1})} \tag{3.22}$$

Thus, I can conduct a joint t-test with regard to the sign of the product of the numerator and denominator in equation (3.22) to find out whether they have opposite signs.

This empirical estimation strategy is applied to test all currencies with various maturities, and all estimation results are reported in table 3.5. The t-statistics under the estimates are calculated based on the hypothesis that the coefficients equal zero. According to the table, the two countries' yield rates do react differently to the common factor, i.e. $(\alpha_{\tau,1}^{*} - \alpha_{\tau,1}) \neq 0$ or $\alpha_{\tau,1}^{*} \neq \alpha_{\tau,1}$, and evidence can be found based on the high significance level implied by t-statistics associated with the estimates of $(\alpha_{\tau,1}^{*} - \alpha_{\tau,1})$. The results suggest that Canada, Euro and the UK's 1-day yield rates are more sensitive to the common factor than the U.S. counterpart. Meanwhile, the compensation for the spot exchange rate risk caused by the common factor is proved to be insignificant across currencies according to low t-statistics for the estimate of $(\lambda_{\tau,1}^{*}{}^{2} - \lambda_{\tau,1}{}^{2})/2$. This result is totally consistent with our theoretical prediction and provides a strong support for the explanation that in a short horizon like one day, spot exchange rate risk premium for the common factor is not significant so that the forward premium puzzle does not appear to exist.

With regard to the results in longer horizons, according to the table, U.S. yield rates with 1-week and 1-month maturities are more sensitive to the common factor rather than their Canada, Euro and UK counterparts, which is totally opposite to the situation in the 1-day horizon. It suggests that the common factor still has significantly different impacts on domestic and foreign yield rates so that $(\alpha^*_{\tau,1} - \alpha_{\tau,1}) \neq 0$ in both horizons across currencies. Meanwhile, unlike the results in the 1-day horizon, the risk premium for the common factor ($({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$) is not ignorable any more, instead, it is highly significant in both horizons.

Table 3.5 also demonstrates that $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ do have opposite signs across currencies in 1-week and 1-month horizons. Specifically, each of CAD, EUR and GBP versus USD has negative $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ and positive $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$. These results support the theoretical claim that the investors in a country exposed to more risk of the common factor require a higher risk premium. More importantly, the magnitude of the risk premium is proved to be larger than that of the interest rate differential in the table, so that all estimates of $(\alpha^*_{\tau,1} - \alpha_{\tau,1}) + ({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ and $({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ have the same signs, and at the same time, $(\alpha^*_{\tau,1} - \alpha_{\tau,1}) + ({\lambda^*_{\tau,1}}^2 - {\lambda_{\tau,1}}^2)/2$ and $(\alpha^*_{\tau,1} - \alpha_{\tau,1})$ always have opposite signs for all the currencies examined in both horizons. Such a judgment can also be supported by the t-statistics in the opposite signs test with over 95% significance level. Thus both the numerator and denominator in equation (3.22) are shown to be significant and have opposite signs, which can certainly generate negative results in the testing and also is consistent with our pre-test expectations. It should be noted that the interdependent specification in Backus *et al* (2001) was rejected empirically by monthly data of Deutsch Mark/USD because the foreign price of risk ($\lambda^*_{\tau,1}$) relative to domestic counterpart ($\lambda_{\tau,1}$) could not generate negative slope coefficients. In contrast, the estimation in my essay obtains the right relationship between domestic and foreign price of risk so that negative slope coefficient can be replicated.

According to the table, the risk premium for the common factor in the 1-month horizon, as predicted, is higher in magnitude than the 1-week. Meanwhile, the difference between the common factor's impact on foreign and domestic yield rates does not change as much as the risk premium in these two horizons. This feature can be found in all currencies except for GBP. As a result, the absolute value of the slope coefficient is higher in the 1-month horizon than 1-week, except for GBP, which is consistent with empirical findings reported in section 2.

The table also suggests that either domestic or foreign local factors are significant to explain the spot exchange rate risk premium. While according to the columns of $\theta_{\tau,2}$ $\theta_{\tau,3}$, the US local factor is significant to explain the forward premium or interest rate differential for the currency pair of CAD/USD and EUR/USD. In contrast, the UK local factor appears to have significant explanatory power for the forward premium of GBP/USD. Finally, implied slope coefficients are calculated based on the estimation results according to equation (3.19) and displayed in table 3.5. The results are not exactly equal to the actual slope coefficients obtained in section 2 but appear to be consistent with the signs and relative magnitude of the actual coefficients.

7. Conclusions

This essay tests the unbiasedness hypothesis of forward rates in various horizons, and finds that the results of the testing with next day tomorrow maturity are strikingly different from those obtained using longer horizons (1-week and 1-month). Basically, the unbiasedness hypothesis is accepted on the daily basis, but rejected on the weekly or monthly basis. It is also found that the magnitude of the slope coefficient increases as horizon rises.

Realizing that current standard models do not show much potential to explain the new findings, I extend a term structure model of interest rates to reconcile the new puzzling anomalies. The solution of the model suggests that the value of the slope coefficient depends on how differently the common factor affects the domestic and foreign security market as well as the foreign exchange market.

This result provides a potential explanation for the term structure puzzle found in this essay. In a short horizon like one day, the risk premium becomes ignorable due to little risk brought by the common factor, thus UIP holds perfectly and there is no puzzle to be seen. In longer horizons, a significant risk premium has to be paid for investors to assume the risk of the common factor, and it always has the opposite sign with the common factor's effect on the interest rate differential; therefore the slope coefficient tends to be negative and the puzzle starts to show up.

Although this essay reveals exciting new evidence about the old forward premium puzzle and provides promising explanations, two limitations of this research deserve further attention. First, with regard to the term structure property, the model just displays the compatibility with the feature, but does not give an explicit theoretical result about how the coefficients of the factors are affected by horizons exactly. A more sophisticated term structure model should be built from scratch to solve this problem.

Second, due to the constraints of sample sizes, this essay only considers three different horizons, which basically focus on short and medium terms. In future, more horizons should be examined, from short term like 1-day to the long terms like 5-year or 10 year, so that we can have a more complete and clearer image about how the result of testing changes along with horizons. Hopefully we can approach the ultimate truth about this old puzzle.

APPENDEX I: TABLES

Table1.1: Regression results of exchange rate change on order flow

	γ_0	γ_1	γ_2	κ_0	κ_1	κ_2
Mark / Dollar	-7.3485 (-1.66)	0.010 (0.96)	0.2194* (14.38)	297.9253* (3.65)	0.4372 (1.85)	2.53e-19 (0.00)
Yen / Dollar	-7.426 (-0.44)	0.002980 (0.17)	0.2775* (9.27)	792.4960* (7.58)	0.1696 (0.80)	7.43e-23 (0.00)
Pound / Dollar	-8.7409 (-1.59)	5.89e-7 (0.00)	0.3530* (6.83)	734.4876* (5.20)	0.00 (0,00)	0.000849 (0.01)
French Franc / Dollar	78.0472 (0.24)	-0.0563 (-0.22)	0.5194* (8.64)	1247* (4.02)	0.0539 (0.26)	9.09e-23 (0.00)
Swedish Krona / Dollar	-1.6037* (-2.18)	-2.62e-6 (-0.00)	0.7502* (8.76)	678.5285* (33.01)	4.52e-23 (0.00)	0.0000673 (0.009)
Lira / Dollar	-239.896 (-0.1)	0.6372 (0.09)	1.0637* (7.60)	1.3600 (0.08)	0.003592 (0.09)	0.9926* (44.71)

*: Over 95% significant

Table 2.1: Unit root test for mid-quote series

Mid-quote series	Dickey-Fuller test				Phillips-Perron test			
	Test statistic	10% critical value	5% critical value	1% critical value	Test statistic	10% critical value	5% critical value	1% critical value
Customer 1-minute	-1.189	-1.62	-1.950	-2.580	-1.533	-1.620	-1.950	-2.580
Inter-dealer 1- minute	-1.189	-1.62	-1.950	-2.580	-1.364	-1.620	-1.950	-2.580
Customer 5-minute	-0.708	-1.61	-1.950	-2.593	-0.321	-1.614	-1.950	-2.593
Inter-dealer 5-minute	-0.707	-1.61	-1.950	-2.593	-0.320	-1.614	-1.950	-2.593

Notes: both tests have no constant, no trend and no drift. The test statistic is Z(t).
DF test has no lag, while PP test has lag 6 for 1-minute series and lag 4 for 5-minute.

Table 2.2: Customer and inter-dealer mid-quote regression

Interval of data	Regression	constant	coefficient	R square	D-W
1-minute	p_t^c on p_t^d	0.001950 (1.1154)	0.997632** (460.8950)	0.9955	1.6466
	$p_{t+1}^c - p_t^c$ on $p_{t+1}^d - p_t^d$	-0.000000 (-0.1432)	0.967013** (105.5955)	0.9206	2.9395
5-minute	p_t^c on p_t^d	0.000002 (0.0373)	1.000036** (13321.3621)	1.0000	1.9180
	$p_{t+1}^c - p_t^c$ on $p_{t+1}^d - p_t^d$	0.000000 (0.0453)	1.005824** (87.3383)	0.9758	2.9018

Notes: inter-day gaps and large outliers have been eliminated in these regressions
*: over 95% significant
**: over 99% significant

Table 2.3: Difference between customer and inter-dealer spread

Intercept α_0	coefficient of order size α_1	Adjusted R^2	F	Prob>F
.0257655** (62.25)	-.0012653** (-33.94)	0.8570	1152.12	0.0000

*: over 95% significant
**: over 99% significant

Table 2.4: Spread, order size and volatility in customer market

Density matrix	GMM1 (instruments: square of explanatory variables)					GMM2 (instruments: lag 1 of explanatory variables)				
	γ_0	γ_1	γ_2	LR	W	γ_0	γ_1	γ_2	LR	W
White	4.9360 (59.96)	0.0098 (0.50)	-0.0081 (-2.47)	5.98	6.08	4.9544 (16.21)	0.0082 (0.10)	-0.0092 (-2.72)	7.12	7.40
N-W	4.9360 (41.44)	0.0098 (0.42)	-0.0081 (-1.80)	3.12	3.24	4.9544 (12.85)	0.0082 (0.08)	-0.0092 (-1.90)	3.34	3.61
Gallant	4.9360 (41.92)	0.0098 (0.42)	-0.0081 (-1.83)	3.21	3.35	4.9544 (12.75)	0.0082 (0.08)	-0.0092 (-1.95)	3.56	3.80

Notes: White, N-W, and Gallant represent various density matrixes. t statistics are in parenthesis, LR stands for the Likelihood Ratio test and W is the Wald test statistics for γ_2.

Table 2.5: Spread, order size and volatility in inter-dealer market

Density matrix	GMM1 (instruments: square of explanatory variables)					GMM2 (instruments: lag 1 of explanatory variables)				
	γ_0	γ_1	γ_2	LR	W	γ_0	γ_1	γ_2	LR	W
White	2.7490 (102.8)	0.0066 (1.33)	-0.0007 (-0.48)	0.19	0.24	2.9163 (65.85)	-0.0383 (-1.07)	-0.0010 (-0.60)	0.29	0.36
N-W	2.7490 (101.4)	0.0066 (1.44)	-0.0008 (-0.51)	0.21	0.27	2.9163 (61.61)	-0.0383 (-1.69)	-0.0010 (-0.63)	0.30	0.39
Gallant	2.7490 (100.3)	0.0066 (1.45)	-0.0008 (-0.50)	0.20	0.25	2.9163 (60.04)	-0.0383 (-1.61)	-0.0010 (-0.61)	0.29	0.37

Notes: White, N-W, and Gallant represent various density matrixes. t statistics are in parenthesis, LR stands for the Likelihood Ratio test and W is the Wald test statistics for γ_2.

Table 2.6: Exchange rate volatility estimation

$$10{,}000 \cdot \Delta M_t = \mu + \theta \varepsilon_{M,t-1} + \varepsilon_{M,t}$$
$$\sigma^2_{M,t} = \varpi + \alpha \varepsilon^2_{M,t-1} + \beta \sigma^2_{M,t-1},$$
$$\varepsilon_{M,t} \mid I_{t-1} \sim N(0, \sigma^2_{M,t})$$

μ	θ	ϖ	α	β
0.14529 (0.41553)	-0.37618 (0.18904)	15.248 (13.645)	0.11463 (0.13535)	0.0002 (0.89617)

Notes: standard errors are reported in parentheses.

Table 3.1: Forward premium puzzle regression results

Currency	Forward Term	constant α	coefficient β	R square	D-W
CAD/USD	Next day tomorrow	-0.000071 (-0.73032)	0.581041 (-1.5333)	0.0025	1.9998
	One week	-0.000198 (-0.40655)	-4.737422 (-2.3239)	0.0141	2.0609
	One month	-0.000988 (-0.32388)	-6.715255 (-2.4595)	0.0807	2.0526
EUR/USD	Next day tomorrow	0.000054 (0.337972)	1.33454 (0.4211)	0.0088	1.9776
	One week	0.000448 (0.542884)	-9.347055 (-2.5505)	0.0179	2.0271
	One month	0.002635 (0.482796)	-12.315631 (-2.1214)	0.1219	2.5981
GBP/USD	Next day tomorrow	-0.00005 (-0.4133)	1.497391 (0.3154)	0.0010	1.8853
	One week	-0.0009 (-1.0497)	-2.845243 (-2.1314)	0.0093	2.0300
	One month	-0.0070 (-1.3525)	-4.042574 (-2.0743)	0.0796	2.4674

Notes: the numbers in parenthesis are t-statistics based on the hypothesis that $\alpha = 0, \beta = 1$. D-W is the Dubin-Watson statistics.

Table 3.2: Distributions of slope coefficients

			<-5	[-5, -4)	[-4, -3)	[-3, -2)	[-2, -1)	[-1, 0)	[0, 1)	[1, 2)	[2, 3)	>3
CAD	1-day	frequency	0	28	27	34	47	79	282	44	0	0
		percentage	0	5.1756	4.9908	6.285	8.688	14.6	52.13	8.133	0	0
	1-week	frequency	106	188	116	33	21	70	5	1	0	0
		percentage	19.63	34.815	21.481	6.111	3.889	12.96	0.926	0.185	0	0
	1-month	frequency	141	308	159	68	1	0	0	11	0	0
		percentage	20.49	44.767	23.11	9.884	0.145	0	0	1.599	0	0
EUR	1-day	frequency	0	0	3	37	63	58	78	323	48	141
		percentage	0	0	0.3995	4.9268	8.3888	7.723	10.386	43.009	6.3915	18.775
	1-week	frequency	253	130	63	50	63	35	18	29	51	58
		percentage	33.733	17.333	8.4	6.6667	8.4	4.6667	2.4	3.8667	6.8	7.7333
	1-month	frequency	569	178	130	56	19	0	1	2	8	0
		percentage	59.086	18.484	13.499	5.8152	1.973	0	0.1038	0.2077	0.8307	0
GBP	1-day	frequency	0	0	0	0	0	0	0	472	88	0
		percentage	0	0	0	0	0	0	0	84.286	15.714	0
	1-week	frequency	36	180	204	101	38	0	0	0	0	0
		percentage	6.4401	32.2	36.494	18.068	6.7979	0	0	0	0	0
	1-month	frequency	3	45	233	312	119	1	0	0	0	0
		percentage	0.42076	6.3114	32.679	43.759	16.69	0.1402	0	0	0	0

Notes: the frequency is the number of the slope coefficients falling in the interval indicated on the top of the table, the percentage is the ratio of the number in that interval over the total number of the slope coefficients obtained in the rolling regression.

Table 3.3: Estimation of integration order in forward premium

		ARFIMA(0,d,0)			ARFIMA(1,d,0)		
currency	term	d	Log likelihood	AIC	d	Log likelihood	AIC
CAD	Next day	0.0442 (2.77)	12014.12	-13.049	0.0757 (3.23)	12015.99	-13.050
	One week	0.7102 (8.45)	3184.5818	-17.391	0.8333 (3.47)	3191.10454	-17.421
	One month	0.9256 (8.37)	400.5768	-13.511	1.0224 (4.99)	400.7009	-13.481
EUR	Next day	0.0985 (6.32)	10637.145	-13.874	0.1458 (6.65)	10643.104	-13.881
	One week	0.725 (8.49)	2580.8987	-16.910	0.8507 (3.45)	2591.9786	-16.976
	One month	1.2103 (1.95)	344.3182	-13.692	1.2285 (1.22)	344.3248	-13.652
JPY	Next day	0.2140 (18.9)	13849.443	-15.043	0.2692 (19.0)	13879.603	-15.075
	One week	0.5424 (14.3)	2855.1424	-15.548	0.6472 (8.14)	2862.1319	-15.581
	One month	0.9033 (2.13)	381.6219	-12.654	1.1701 (1.13)	383.2917	-12.676

Table 3.4: Factor analysis results

Panel A

	U.S.: Canada			U.S.: Euro			U.S.: UK		
	$Z_{1,t}$	$Z_{2,t}$	$Z^*_{2,t}$	$Z_{1,t}$	$Z_{2,t}$	$Z^*_{2,t}$	$Z_{1,t}$	$Z_{2,t}$	$Z^*_{2,t}$
eigenvalues	17.34	0.3615	0.0360	7.24	0.3770	0.01454	6.8542	0.40484	0.03165
variance	4.0018	0.1125	0.1113	1.70	0.03761	0.00161	1.6212	0.00614	0.27975

Panel B:

	Common factor	U.S. local factor	Foreign local factor	total	Common factor	U.S. local factor	Foreign local factor	total
		U.S. yields				Canadian yields		
overnight	0.96895	0.02725	0.00378	0.999979	0.921	0.008	0.069	0.998
1-week	0.99089	0.00225	0.00538	0.998521	0.934	0.004	0.061	0.999
1-month	0.98765	0.0041	0.00656	0.998308	0.943	0.003	0.051	0.998
		U.S. yields				U.K. yields		
overnight	0.98716	0.003741	0.005301	0.9962	0.54816	0.005113	0.44679	0.9999
1-week	0.98954	0.005049	0.004325	0.99891	0.69904	0.079643	0.221	0.99968
1-month	0.98748	0.007146	0.002707	0.99733	0.98797	0.004306	0.005656	0.99793
		U.S. yields				Euro yields		
overnight	0.975	0.021585	0.000181	0.99676	0.65553	0.34262	0.00169	0.99984
1-week	0.97716	0.021032	0.000826	0.99902	0.66372	0.32615	0.00899	0.99885
1-month	0.9779	0.018971	0.000737	0.99761	0.66834	0.31588	0.01482	0.99904

Table 3.5: Factor model empirical testing results

		$\theta_{\tau,1}$ $(\alpha^*_{\tau,1}-\alpha_{\tau,1})$	$\theta_{\tau,2}$ $(\alpha^*_{\tau,2})$	$\theta_{\tau,3}$ $(\alpha_{\tau,2})$	R^2	F-test	$\gamma_{\tau,1}$ $(\frac{\lambda^{*\,2}_{\tau,1}-\lambda^{2}_{\tau,1}}{2})$	$\gamma_{\tau,2}$ $(\frac{\lambda^{*\,2}_{\tau,2}}{2})$	$\gamma_{\tau,3}$ $(\frac{\lambda^{2}_{\tau,2}}{2})$	R^2	F-test	opposite signs test	$\hat{\beta}$
CAD	1-day	0.789947 (5.212123)	1.112384 (1.23091)	-3.09553 -(3.40685)	0.0217	19.7061	2.618507 (1.465145)	5.647033 (0.5270)	15.099878 (1.417231)	0.0024	2.2116	1.4687	4.3148
	1-week	-0.43488 (-61.5412)	0.108978 (2.309745)	1.455411 (34.60409)	0.9293	177.7976	2.742976 (2.125958)	5.00554 (0.65182)	15.87696 (1.843016)	0.0210	4.3257	-1.6828	-5.3074
	1-month	-0.44349 (-24.698752)	1.460859 (13.953775)	-0.05554 (-0.32411)	0.9343	28.2388	4.458189 (2.347754)	2.974111 (0.267956)	9.946184 (0.54974)	0.0954	3.0626	-1.7756	-9.0525
EUR	1-day	0.716615 (3.312555)	-1.501483 (-1.032695)	2.004097 (0.285816)	0.0119	5.9882	7.338287 (1.340169)	8.521877 (0.23102)	133.396222 (0.751602)	0.0024	1.2043	0.4905	11.2402
	1-week	-0.375571 (-6.762754)	1.995699 (5.215754)	0.721701 (0.401803)	0.9631	99.3770	6.632264 (1.765485)	16.897331 (0.652846)	13.173972 (0.108428)	0.0183	2.1132	-1.9879	-16.6623
	1-month	-0.516334 (-2.694458)	2.443058 (2.372815)	-1.34258 (-0.17957)	0.9088	26.8119	10.348239 (1.841327)	19.460225 (0.406224)	98.883119 (0.922917)	0.0668	2.9579	-1.9552	-19.0418
GBP	1-day	0.855447 (10.021813)	-0.768869 (-1.956209)	-0.124391 (-0.43386)	0.0892	47.5692	4.075726 (0.884984)	31.067068 (1.465024)	6.293299 (0.406633)	0.0029	1.5429	0.8459	5.7644
	1-week	-0.667405 (-5.563809)	0.855305 (1.613700)	0.389688 (0.951099)	0.8616	122.5178	4.312084 (1.956162)	21.808972 (1.446685)	4.640375 (0.398198)	0.0170	1.7757	-1.8396	-5.4609
	1-month	-1.247282 (-2.872228)	1.075038 (0.765846)	0.412867 (0.379721)	0.7630	42.8246	5.244626 (2.038287)	0.869579 (0.037268)	11.564122 (0.632642)	0.0314	1.9948	-1.9556	-3.2048

Notes: 1) All the numbers in parenthesis are the t-statistics calculated on the basis of the null hypothesis that coefficients equal zero

2) The degree of freedoms for F-test are df1=3 df2=infinity, in this case 10% critical value is 2.08.

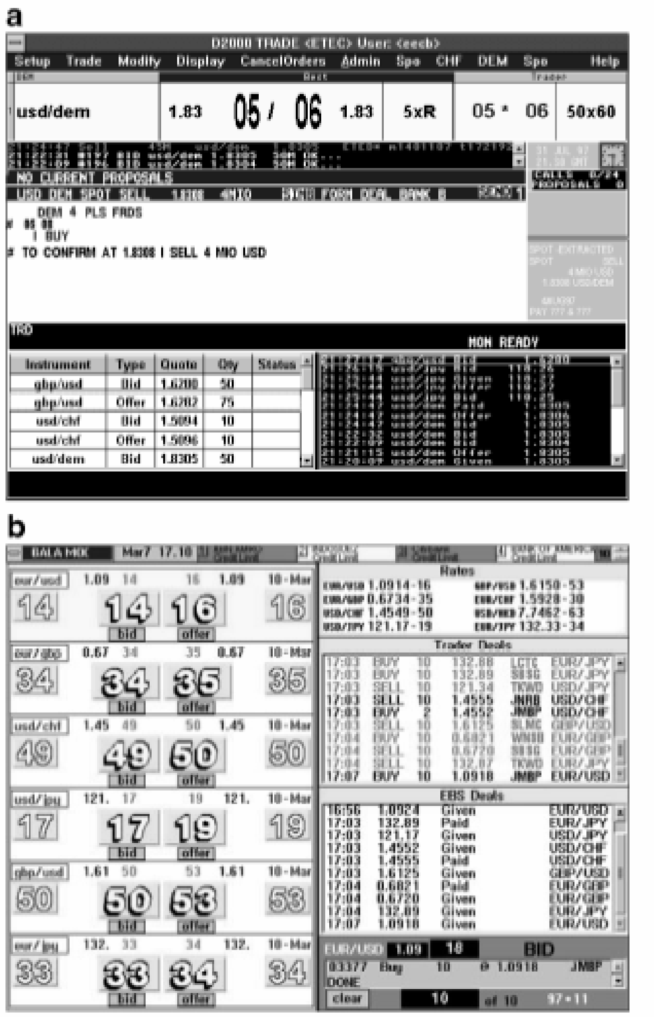
3) For t-test, one way critical value at 10% significance level is 1.38, and 5% is 1.64.

4) Opposite sign test for daily data is a t-test for $\theta_{\tau,1}\cdot\gamma_{\tau,1}<0$, while weekly and monthly tests inspect whether $(\theta_{\tau,1}+\gamma_{\tau,1})\cdot\gamma_{\tau,1}<0$.

5) Implied slope coefficient $\hat{\beta}$ is calculated based on equation (3.19)

APPENDIX II: FIGURES

Figure 1.1: Electronic trading screens



Notes to the figure: Electronic trading screens. (a) Reuters Dealing 2000.This screen shows the Reuters Dealing 2000 system. The part in the middle contains the D2000-1 system for direct bilateral trading, and the top section is the D2000-2 electronic broker. The dealer may choose the contents of the screen. The dealer may choose which exchange rates to display in the electronic broker and whether to display the best prices in the market (column marked best) and/or the best available to him (from credit-approved banks only). From the D2000-1 part we can see that the dealer has been contacted for a quote for USD 4 million against DEM. The dealer replies with the quote “05 08”, which is understood to be bid 1.8305 and ask 1.8308. The contacting dealer responds with “I BUY,” and the system automatically fills in the line “TO CONFIRM AT 1.8308. . . .”In the lower right corner of the screen, the dealer can see the price and direction of the last trades through the D2000-2 system. (b) EBS. The left half of the EBS screen shows the bid and offer (ask) prices. The dealer chooses which exchange rates to display (the base currency is written first). The prices shown are either the best prices in the market or the best available (from credit-approved banks only). The upper part of the right half of the screen shows the dealer’s own trade. The lower part shows the price and direction of all trades through the system for selected exchange rates. “Given” means that it was traded at the bid price, and “paid” means it was traded at the ask price. The intuition is that the limit order dealer is “given” the base currency (buys).

Figure 1.2: Inter-dealer order flow and exchange rate for DM/USD
(Solid line is order flow and dash line is exchange rate, the same as below)

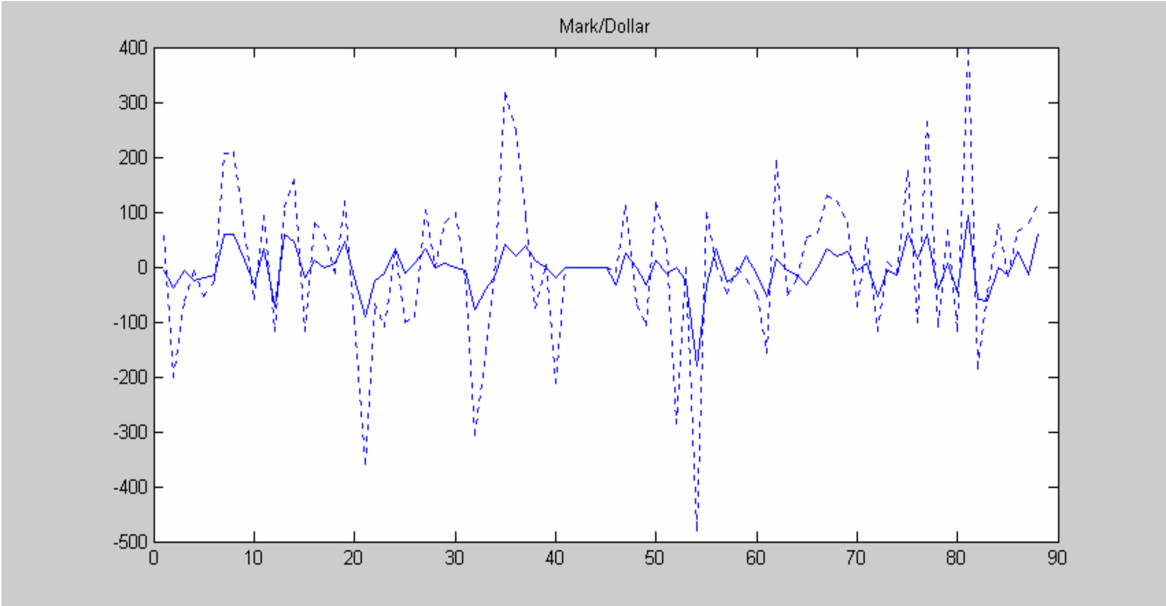


Figure 1.3: Inter-dealer order flow and exchange rate for Yen/USD

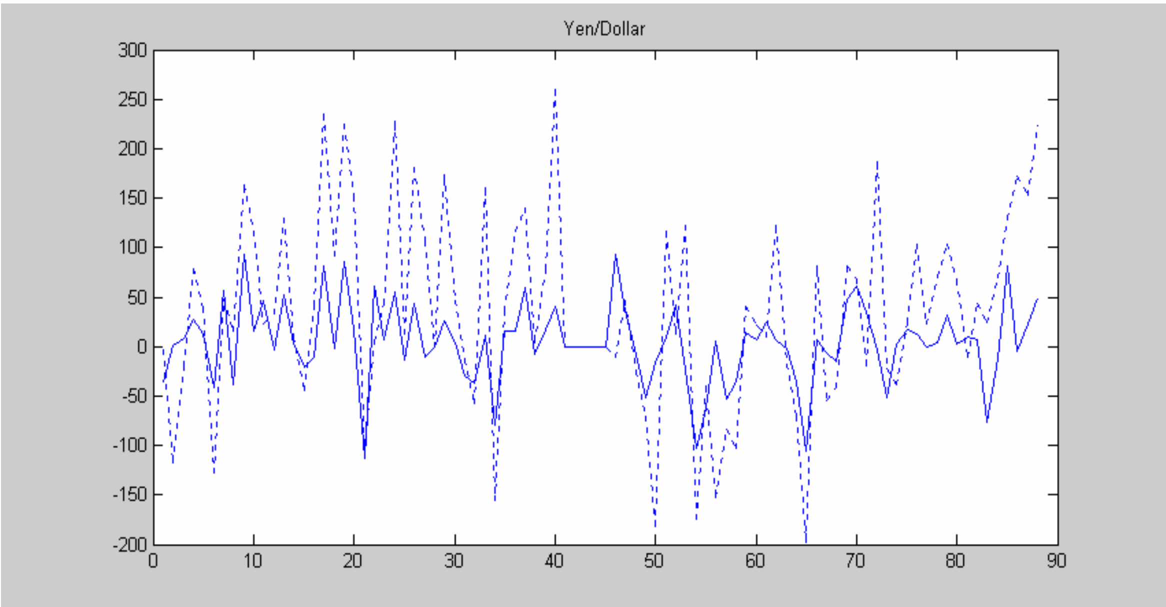


Figure 1.4: Inter-dealer order flow and exchange rate for Pound/USD

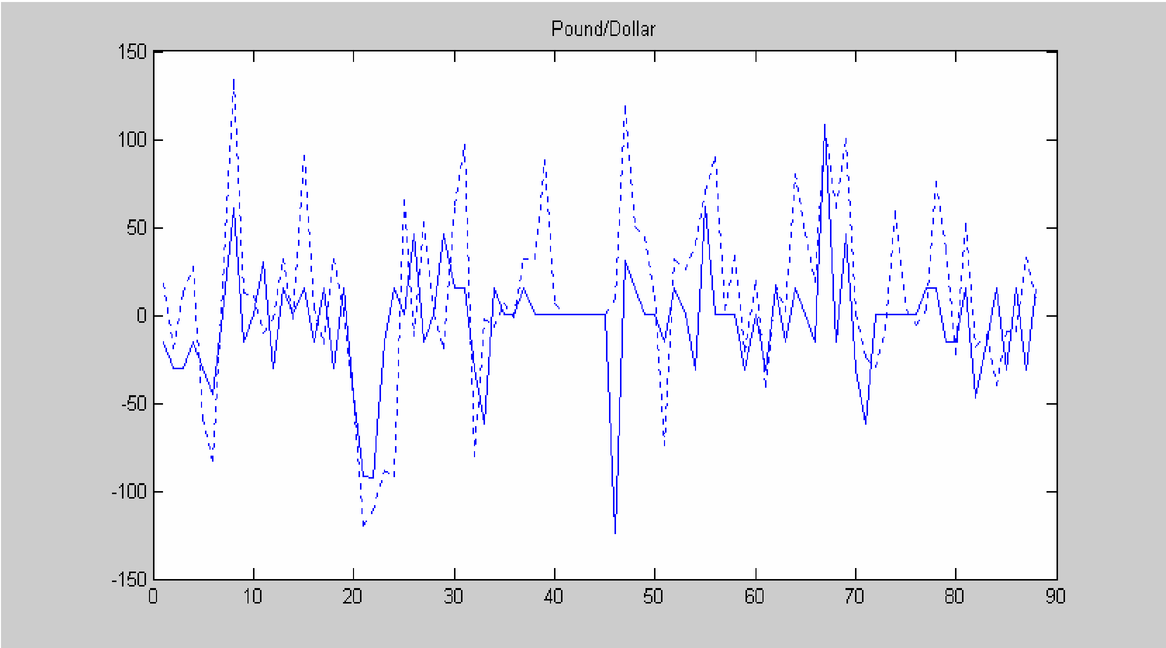


Figure 1.5: Inter-dealer order flow and exchange rate for French Franc/USD

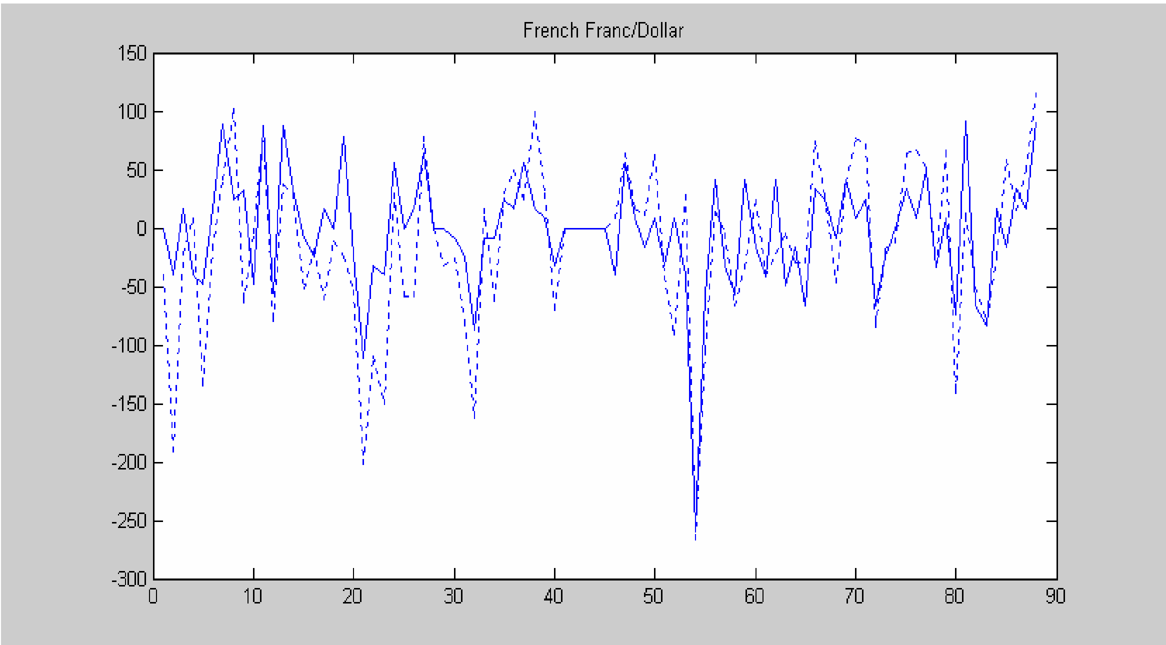


Figure 1.6: Inter-dealer order flow and exchange rate for Krona/USD

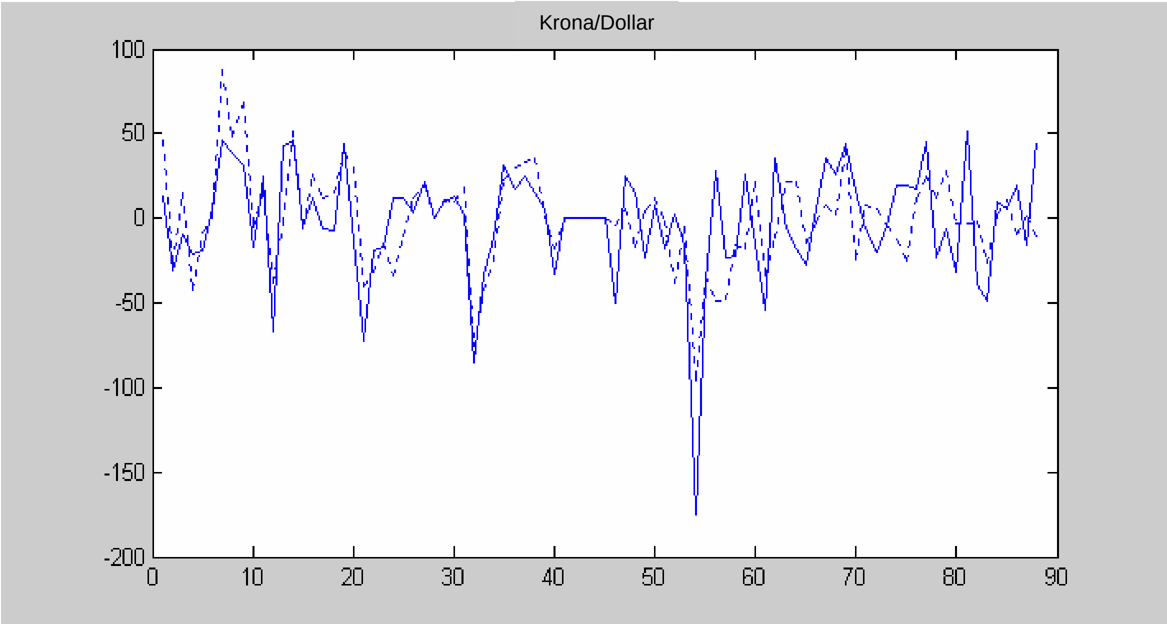


Figure 1.7: Inter-dealer order flow and exchange rate for Lira/USD

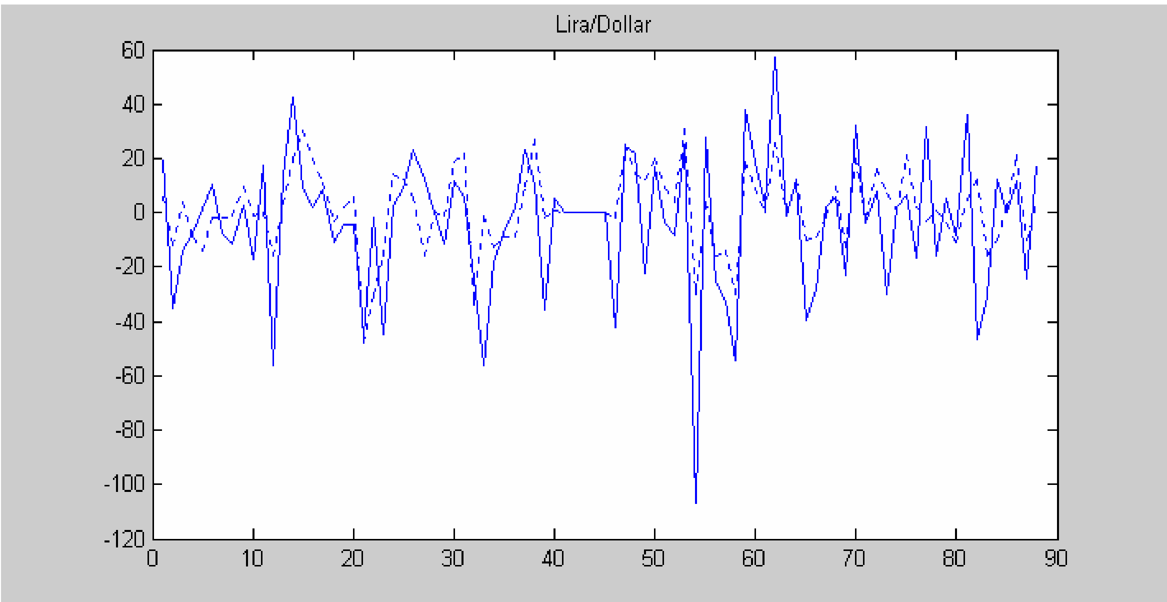
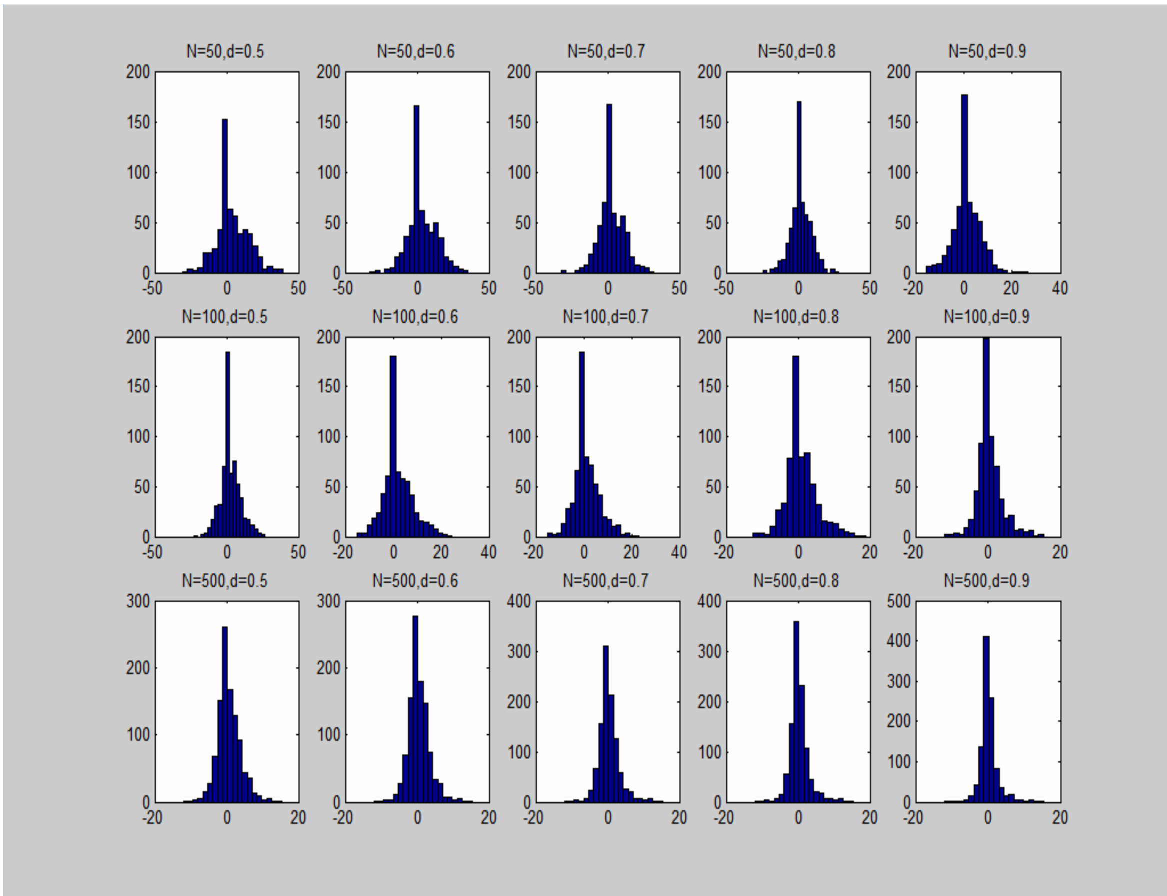


Figure 3.1: Distribution of simulation results



Notes: N is sample size, d is the integration order of series

APPENDIX III: PROOFS

Lemma 1.1: *Under our set of assumptions, the reservation selling and buying price in the FX markets for dealer i in sub-period t are* $p_{t-1,i}^{post} + a_{t,i}^{r}$ and $p_{t-1,i}^{post} - b_{t,i}^{r}$ *, where*

$$a_{t,i}^{r} = \frac{A\sigma^2}{2}(Q_{t,i} - 2I_{t,i}^{ante}) + E(z_t \mid S_{t,i})$$

$$b_{t,i}^{r} = \frac{A\sigma^2}{2}(Q_{t,i} + 2I_{t,i}^{ante}) - E(z_t \mid S_{t,i})$$

Proof:

Plugging the total wealth without trading into the utility function gives the expected utility in the case of not trading:

$$E(U(W(0)) \mid S_i) = E(-e^{-A(C_i - \overline{F}_i + I_{t,i}^{ante}(P_{t-1,i}^{post} + z_t))} \mid S_{t,i}) \qquad \text{(A-1.1)}$$

It is well known that if a random variable *x* is normally distributed with N(μ, σ^2) and the utility function is $U(x) = -e^{-Ax}$, then $E[U(x)] = -e^{-A(\mu - A\sigma^2/2)}$. Using this property, equation (A-1.1) can be written as:

$$E(-e^{-A(C_i - \overline{F}_i + I_{t,i}^{ante}(P_{t-1,i}^{post} + z_t)} \mid S_{t,i}) = -e^{-A[C_i - \overline{F}_i + I_{t,i}^{ante}(P_{t-1,i}^{post} + E(z_t \mid S_{t,i}))] + \frac{1}{2}AI_{t,i}^{ante^2}\sigma^2} \qquad \text{(A-1.2)}$$

where σ^2 is the conditional variance of price based on the current information. If the dealer makes the transaction, his wealth is $W(a_i)$. Substituting it into the utility function obtains his expected utility as below:

$$E(U(W(a_i)) \mid S_{t,i}) = E(-e^{-A[C_i - \overline{F}_i + (I_{t,i}^{ante} - Q_{t,i})(p_{t-1,i}^{post} + z_t) + (P_{t-1,i}^{post} + a_{t,i})Q_{t,i}]} \mid S_{t,i}) \qquad \text{(A-1.3)}$$

Adjusting the items in the exponential part, we have equation (A-1.4)

$$\begin{aligned} &E(-e^{-A[C_i - \overline{F}_i + (I_{t,i}^{ante} - Q_{t,i})(P_{t-1,i}^{post} + z_t) + (a_{t,i} + P_{t-1,i}^{post})Q_{t,i}]} \mid S_{t,i}) = \\ &E(-e^{-A[C_i - \overline{F}_i + I_{t,i}^{ante} P_{t-1,i}^{post} + a_{t,i}Q_{t,i} + (I_{t,i}^{ante} - Q_{t,i}) z_t]} \mid S_{t,i}) \end{aligned} \qquad \text{(A-1.4)}$$

Similarly, using property we mentioned before, the dealer's expected utility is:

$$E(-e^{-A[C_i-\overline{F}_i+I_{t,i}^{ante}P_{t-1,i}^{post}+a_{t,i}Q_{t,i}+(I_{t,i}^{ante}-Q_{t,i})z_t]} \mid S_{t,i}) = \\ -e^{-A[C_i-\overline{F}_i+I_{t,i}^{ante}P_{t-1,i}^{post}+a_{t,i}Q_{t,i}+(I_{t,i}^{ante}-Q_{t,i})E(z_t|S_{t,i})]+\frac{1}{2}A^2(I_{t,i}^{ante}-Q_{t,i})^2\sigma^2} \tag{A-1.5}$$

According to the definition of reservation price, the reservation ask price $a_{t,i}^r$ should make equation (A-1.6) hold:

$$E(U(W(0)) \mid S_i) = E(U(W(a_{t,i}^r)) \mid S_i) \tag{A-1.6}$$

Substituting equations (A-1.2) and (A-1.5) into (A-1.6) gives the following equation (A-1.7):

$$-A[C_i-\overline{F}_i+I_{t,i}^{ante}(P_{t-1,i}^{post}+E(z_t \mid S_{t,i}))]+\frac{1}{2}A^2 I_{t,i}^{ante2}\sigma^2 = \\ -A[C_i-\overline{F}_i+I_{t,i}^{ante}P_{t-1,i}^{post}+a_{t,i}^r Q_{t,i}+(I_{t,i}^{ante}-Q_{t,i})E(z_t \mid S_{t,i})]+\frac{1}{2}A^2(I_{t,i}^{ante}-Q_{t,i})^2\sigma^2 \tag{A-1.7}$$

The solution of equation (A-1.7) is the reservation ask price a_i^r:

$$a_{t,i}^r = \frac{A\sigma^2}{2}(Q_{t,i}-2I_{t,i}^{ante})+E(z_t \mid S_{t,i})$$

The reservation bid price $b_{t,i}^r$ can be solved similarly.

Proposition 1.1: *Under our set of assumptions, the solutions to the dealer's problem and the optimal bid and ask price of dealer i in the FX market are* $(p_{t-1,i}^{post}-b_{t,i})$ *and* $(p_{t-1,i}^{post}+a_{t,i})$, *where*

$$a_{t,i} = a_{t,i}^r + A\sigma^2\frac{R+I_{t,i}^{ante}}{M}$$

$$b_{t,i} = b_{t,i}^r + A\sigma^2\frac{R-I_{t,i}^{ante}}{M}$$

Proof:

Equation (A-1.8) is the first order condition for the dealer's problem

$$\frac{\partial \pi_{a,i}}{\partial a_i}(a_{t,i} - a^r_{t,i}) + \pi_{a,i} = 0 \qquad \text{(A-1.8)}$$

Clearly equation (A-1.9) holds:

$$\frac{\partial \pi_{a,i}}{\partial a_{t,i}} = \frac{d\pi_{a,i}}{dI^{ante}_{t,i}} \cdot \frac{dI^{ante}_{t,i}}{da_{t,i}} \qquad \text{(A-1.9)}$$

Plugging equation (A-1.9) into equation (A-1.8) gets equation (A-1.10):

$$\frac{d\pi_{a,i}}{dI^{ante}_{t,i}} \frac{dI^{ante}_{t,i}}{da_{t,i}}(a_{t,i} - a^r_{t,i}) + \pi_{a,i} = 0 \qquad \text{(A-1.10)}$$

Dividing equation (A-1.10) by $\frac{dI^{ante}_{t,i}}{da_{t,i}}$ gives:

$$\frac{d\pi_{a,i}}{dI^{ante}_{t,i}}(a_{t,i} - a^r_{t,i}) + \pi_{a,i} \frac{da_{t,i}}{dI^{ante}_{t,i}} = 0 \qquad \text{(A-1.11)}$$

Then rewrite (A-1.11) as (A-1.12):

$$\frac{d\pi_{a,i}}{dI^{ante}_{t,i}} a^r_{t,i} = \frac{d(\pi_{a,i} \cdot a_{t,i})}{dI^{ante}_{t,i}} \qquad \text{(A-1.12)}$$

According to the discussion in the determination of probability $\pi_{a,i}$, we have

$$\pi_{a,i} = P(a_{t,i} < a_{t,-i}) = P(I^{ante}_{t,i} > I^{ante}_{t,-i}) = \prod_{j \neq i}^{M} F_j(I^{ante}_{t,i}) = G(I^{ante}_{t,i})$$

Denote the derivative of $G(I^{ante}_{t,i})$ to $I^{ante}_{t,i}$ as $g(I^{ante}_{t,i})$, i.e.: $\frac{d\pi_{a,i}}{dI^{ante}_{t,i}} = g(I^{ante}_{t,i})$

$I^{ante}_{t,i}$ is uniformly distributed within [-R, R], so the solution to the derivative function (A-1.12) is

$$\pi_{a,i} \cdot a_{t,i} = \int_{-R}^{I^{ante}_{t,i}} g(x) \cdot a^r_{t,i} dx + D \quad \textit{where D is constant.}$$

Now, suppose $I^{ante}_{t,i}$ = - R, the inventory is at the low bound, then any other dealer's inventory is higher than dealer i's, and their quotes must be lower than his. It is impossible that dealer i's price is the lowest in the market, which means $\pi_{a,i} = 0$. Therefore, the right hand side of the equation equals zero when $I_{t,i}^{ante}$ = - R, which implies D=0. So the solution is:

$$\pi_{a,i} \cdot a_{t,i} = \int_{-R}^{I_{t,i}^{ante}} g(x) \cdot a_{t,i}^{r} dx$$

then

$$\pi_{a,i} \cdot a_{t,i} = \int_{-R}^{I_{t,i}^{ante}} a_{t,i}^{r} dG(x) = [a_{t,i}^{r} \cdot G(x)] |_{-R}^{I_{t,i}^{ante}} - \int_{-R}^{I_{t,i}^{ante}} G(x) da_{t,i}^{r}$$

Plugging $\int_{-R}^{I_{t,i}^{ante}} G(x) da_{t,i}^{r} = -A\sigma^2 \int_{-R}^{I_{t,i}^{ante}} G(x)dx$ and $[a_{t,i}^{r} \cdot G(x)] |_{-R}^{I_{t,i}^{ante}} = a_{t,i}^{r} \cdot \pi_{a,i}$ into the equation above gives

$$\pi_{a,i} \cdot a_{t,i} = a_{t,i}^{r} \cdot \pi_{a,i} + A\sigma^2 \int_{-R}^{I_{t,i}^{ante}} G(x)dx$$

Dividing the equation above by $\pi_{a,i}$ gives the solution:

$$a_{t,i} = a_{t,i}^{r} + A\sigma^2 \frac{\int_{-R}^{I_{t,i}^{ante}} G(x)dx}{\pi_{a,i}^{d}} \quad \text{or} \quad a_{t,i} = a_{t,i}^{r} + A\sigma^2 \frac{\int_{-R}^{I_{t,i}^{ante}} G(x)dx}{G(I_{t,i}^{ante})} \qquad \text{(A-1.13)}$$

Next, if the inventory level I_i is i.i.d with uniform distribution [-R, R], then its distribution function $F(x)$ will follow $F(x) = \frac{x+R}{2R}$, and clearly equation (A-1.14) holds:

$$G(x) = \prod_{j \neq i}^{M} F_j(x) = (\frac{x+R}{2R})^{M-1} \qquad \text{(A-1.14)}$$

therefore:

$$\int_{-R}^{I_{t,i}^{ante}} G(x)dx = \int_{-R}^{I_{t,i}^{ante}} (\frac{x+R}{2R})^{M-1} dx = \frac{2R}{M} (\frac{I_{t,i}^{ante} + R}{2R})^{M} \qquad \text{(A-1.15)}$$

and also,

$$G(I_{t,i}^{ante}) = \prod_{j \neq i}^{M} F_j(I_{t,i}^{ante}) = (\frac{I_{t,i}^{ante} + R}{2R})^{M-1} \qquad \text{(A-1.16)}$$

So

$$\frac{\int_{-R}^{I_{t,i}^{ante}} G(x)dx}{G(I_{t,i}^{ante})} = \frac{R + I_{t,i}^{ante}}{M} \tag{A-1.17}$$

Finally, plugging (A-1.17) into (A-1.13) gives the solution:

$$a_{t,i} = a_{t,i}^{r} + A\sigma^2 \frac{R + I_{t,i}^{ante}}{M}$$

Optimal bid price can be derived in a similar way.

Lemma 2.1: *Under our set of assumptions, dealer i's reservation selling (ask) and buying (bid) price in the inter-dealer market are* $(p^0 + a_i^{r,d})$ *and* $(p^0 - b_i^{r,d})$ *, where*

$$a_i^{r,d} = (\frac{A\,\sigma^2}{2} + \beta^d) \cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i$$

$$b_i^{r,d} = (\frac{A\sigma^2}{2} + \beta^d) \cdot Q + \frac{\omega}{Q} + A\sigma^2 \cdot I_i$$

Proof:

If the dealer's quote is declined and no transaction takes place, the dealer's terminal wealth will be:

$$W_i(0) = K_i - \overline{C}_i + I_i(p^0 + \tilde{z}) \tag{A-2.1}$$

Plug (A-2.1) into the dealer's utility function, his expected utility will be:

$$E(U(W(0)) \mid S^a) = E\{-\exp[-A(K_i - \overline{C}_i + I_i(p^0 + \tilde{z}))] \mid S^a\} \tag{A-2.2}$$

It is well known that if a random variable x is normally distributed with $N(\mu, \sigma^2)$ and the utility function is $U(x) = -e^{-Ax}$ then $E[U(x)] = -e^{-A(\mu - A\sigma^2/2)}$. According to this property, equation (A-2.2) can be written as:

$$E\{-\exp[-A(K_i - \overline{C}_i + I_i(p^0 + \widetilde{z}))] \mid S^a\} = \\ -\exp[-A(K_i - \overline{C}_i + I_i(p^0 + E(\widetilde{z} \mid S^a)) + \frac{1}{2}A^2 I_i^2 \sigma^2] \tag{A-2.3}$$

where σ^2 is the conditional variance of price based on current information. If another dealer buys Q at the ask price, the anticipated wealth will be:

$$W(a_i^d) = K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + \widetilde{z}) + (p^0 + a_i^d)Q \tag{A-2.4}$$

Then his expected utility is:

$$E(U(W(a_i^d)) \mid S^a) = E\{-\exp[-A(K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + \widetilde{z}) + (p^0 + a_i^d)Q)] \mid S^a\} \tag{A-2.5}$$

Similarly, using the property mentioned before, the dealer's expected utility is:

$$E(U(W(a_i^d)) \mid S^a) = -\exp\{-A[K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + E(\widetilde{z} \mid S^a)) + (p^0 + a_i^d)Q] + \frac{1}{2}A^2(I_i - Q)^2\sigma^2\} \tag{A-2.6}$$

According to its definition, the reservation ask price $a_i^{r,d}$ should make equation (A-2.7) hold:

$$E(U(W(0)) \mid S^a) = E(U(W(a_i^{r,d})) \mid S^a) \tag{A-2.7}$$

The combination of equations (A-2.2) and (A-2.3) gives:

$$E(U(W(0)) \mid S^a) = -\exp[-A(K_i - \overline{C}_i + I_i(p^0 + E(\widetilde{z} \mid S^a)) + \frac{1}{2}A^2 I_i^2 \sigma^2] \tag{A-2.8}$$

Plugging equations (A-2.6) and (A-2.8) into (A-2.7) gets:

$$E(U(W(0)) \mid S^a) = -\exp[-A(K_i - \overline{C}_i + I_i(p^0 + E(\widetilde{z} \mid S^a)) + \frac{1}{2}A^2 I_i^2 \sigma^2] = \\ E(U(W(a_i^d)) \mid S^a) = -\exp\{-A[K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + E(\widetilde{z} \mid S^a)) + (p^0 + a_i^{r,d})Q] + \frac{1}{2}A^2(I_i - Q)^2\sigma^2\}$$

therefore:

$$-\exp[-A(K_i - \overline{C}_i + I_i(p^0 + E(\widetilde{z} \mid S^a)) + \frac{1}{2}A^2 I_i^2 \sigma^2] = \\ -\exp\{-A[K_i - \overline{C}_i - \omega_i + (I_i - Q)(p^0 + E(\widetilde{z} \mid S^a)) + (p^0 + a_i^{r,d})Q] + \frac{1}{2}A^2(I_i - Q)^2\sigma^2\}$$

therefore:

$$-A(K_i - \overline{C_i} + I_i(p^0 + E(\tilde{z} \mid S^a)) + \frac{1}{2}A^2 I_i^2 \sigma^2 =$$

$$-A[K_i - \overline{C_i} - \omega_i + (I_i - Q)(p^0 + E(\tilde{z} \mid S^a)) + (p^0 + a_i^{r,d})Q] + \frac{1}{2}A^2 (I_i - Q)^2 \sigma^2$$

Eliminating the same items contained in both sides of the equation above gives:

$$0 = -A[-\omega_i - QE(\tilde{z} \mid S^a) + a_i^{r,d} Q] + \frac{1}{2}A^2(-2I_i Q + Q^2)\sigma^2$$

This equation can also be written as:

$$A\omega_i + AQE(\tilde{z} \mid S^a) - Aa_i^{r,d} Q - A^2 I_i Q\sigma^2 + \frac{1}{2}A^2 Q^2 \sigma^2 = 0$$

Divided by A on both sides of the equation obtains:

$$\omega_i + QE(\tilde{z} \mid S^a) - a_i^{r,d} Q - A\, I_i Q\sigma^2 + \frac{1}{2}A\, Q^2 \sigma^2 = 0$$

Rearranging the equation gets:

$$a_i^{r,d} Q = \omega_i + QE(\tilde{z} \mid S^a) - A\, I_i Q\sigma^2 + \frac{1}{2}A\, Q^2 \sigma^2$$

Divided by Q on both sides gets:

$$a_i^{r,d} = \frac{\omega_i}{Q} + E(\tilde{z} \mid S^a) - A\, I_i \sigma^2 + \frac{1}{2}A\, Q\, \sigma^2$$

Finally, plugging $E(\tilde{z} \mid S^a) = \beta^d \cdot Q$ into the equation above gets:

$$a_i^{r,d} = \frac{\omega_i}{Q} + \beta^d Q - A\, I_i \sigma^2 + \frac{1}{2}A\, Q\, \sigma^2$$

Hence the solution is:

$$a_i^{r,d} = (\frac{A\,\sigma^2}{2} + \beta^d) \cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i$$

Similarly, reservation bid price can also be solved this way.

Proposition 2.2: *Under our set of assumptions, the solution to the dealer's problem and the optimal bid and ask price of dealer i in the inter-dealer market are* $(p^0 - b_i^c)$ *and* $(p^0 + a_i^c)$, *where*

$$a_i^c = a_i^{r,c} + A\sigma^2 \cdot \frac{R + I_i}{N} + \delta_i c$$

$$b_i^c = b_i^{r,c} + A\sigma^2 \cdot \frac{R - I_i}{N} + \delta_i c$$

Proof:

The dealer's quoting problem in the customer market is:

$$\text{Max}\, \pi_{a,i}^c (e^{-A(a_i^c - a_i^{r,c})Q} - 1)$$

where $\pi_{a,i}^c = P(a_i^c < a_{-i}^c + \delta_i c) = P(I_i > I_{-i} - f(\delta_i c))$. Using the technique solving the inter-dealer quoting problem as shown in proof of proposition 1, the solution to the customer quoting problem can be written as:

$$a_i^c = a_i^{r,c} + A\sigma^2 \cdot \frac{R + I_i + f(\delta_i c)}{N} \qquad \text{(A-2.9)}$$

According to its definition, the function *f* maps the effect of the additional cost ($\delta_i c$) onto the inventory level, i.e. inventory increasing by $f(\delta_i c)$ would lead to an increase in ask price by $\delta_i c$. Since inventory level and optimal quote are linearly related as shown in equation (A-2.9), we therefore have the following relationship:

$$\frac{f(\delta_i c)}{\delta_i c} = \frac{\partial I_i}{\partial a_i^c} \qquad \text{(A-2.10)}$$

Equation (A-2.9) is the preliminary solution of the problem and it indicates a linear relationship between the inventory and the optimal ask price such that:

$$\frac{\partial a_i^c}{\partial I_i} = A\sigma^2 \cdot \frac{1}{N} \qquad \text{(A-2.11)}$$

therefore

$$\frac{f(\delta_i c)}{\delta_i c} = \frac{\partial I_i}{\partial a_i^c} = N / A\sigma^2 \tag{A-2.12}$$

so

$$f(\delta_i c) = \frac{N}{A\sigma^2} \cdot \delta_i c \tag{A-2.13}$$

Plugging (A-2.12) into equation (A-2.13) obtains the final solution as below:

$$a_i^c = a_i^{r,c} + A\sigma^2 \cdot \frac{R+I_i}{N} + \delta_i c$$

Similarly, customer optimal bid price can be solved and shown as below:

$$b_i^c = b_i^{r,c} + A\sigma^2 \cdot \frac{R-I_i}{N} + \delta_i c$$

Proposition 2.3: *Under our set of assumptions, dealer i's mid-quote of bid-ask prices and spread in the inter-dealer market are* $P_{t,i}^d$ *and* S_i^d *, where*

$$p_{t,i}^d = p_{t-1,i} + r_t - A\sigma^2(1-\frac{1}{N}) \cdot I_i, \quad S_i^d = (A\sigma^2 + 2\beta^d) \cdot Q + \frac{2\omega}{Q} + A\sigma^2 \cdot \frac{2R}{N}$$

The mid-quote of bid-ask prices and spread in the customer market are $P_{t,i}^c$ *and* S_i^c *, where*

$$p_{t,i}^c = p_{t-1,i} + r_t - A\sigma^2(1-\frac{1}{N}) \cdot I_i, \quad S_i^c = (A\sigma^2 + 2\beta^c) \cdot Q + \frac{2\omega}{Q} + A\sigma^2 \cdot \frac{2R}{N} + 2\delta_i c$$

Proof:

According to proposition 2.1, the dealer i's optimal ask and bid price in the inter-dealer market are $(p^0 + a_i^d)$ and $(p^0 - b_i^d)$, where:

$$a_i^d = a_i^{r,d} + A\sigma^2 \cdot \frac{R+I_i}{N}, \quad b_i^d = b_i^{r,d} + A\sigma^2 \cdot \frac{R-I_i}{N}$$

and $a_i^{r,d}, b_i^{r,d}$ are given in lemma 1 as below:

$$a_i^{r,d} = (\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i \,,\quad b_i^{r,d} = (\frac{A\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} + A\sigma^2 \cdot I_i$$

Plugging the reservation quotes and $p^0 = p_{t-1} + r_t$ into optimal inter-dealer quotes gets:

$$p^0 + a_i^d = (p_{t-1} + r_t) + [(\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I_i}{N}) \tag{A-2.14}$$

$$p^0 - b_i^d = (p_{t-1} + r_t) - [(\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i] - (A\sigma^2 \frac{R-I_i}{N}) \tag{A-2.15}$$

Mid-quote is mid point of bid-ask prices, so that inter-dealer mid-quote $p_{t,i}^d$ can be written

as:

$$p_{t,i}^d = \frac{1}{2}(p^0 + a_i^d + p^0 - b_i^d)$$

Plugging equations (A-2.14) and (A-2.15) into the equation above gets:

$$\begin{aligned}
p_{t,i}^d &= \frac{1}{2}\{[(p_{t-1} + r_t) + (\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I}{N}) + \\
&(p_{t-1} + r_t) - [(\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i] - (A\sigma^2 \frac{R-I}{N})\} \\
&= \frac{1}{2}\{2(p_{t-1} + r_t) - 2A\sigma^2 I_i + 2\frac{I_i}{N} A\sigma^2\} \\
&= p_{t-1,i} + r_t - A\sigma^2(1 - \frac{1}{N})\cdot I_i
\end{aligned}$$

So, the inter-dealer mid-quote is:

$$p_{t,i}^d = p_{t-1,i} + r_t - A\sigma^2(1 - \frac{1}{N})\cdot I_i$$

Spread is the difference between ask and bid prices and can be express as below:

$$S_i^d = (p^0 + a_i^d) - (p^0 - b_i^d) = a_i^d + b_i^d$$

Plugging equations (A-2.13) and (A-2.14) into the equation above gets:

$$\begin{aligned}
S_i^d &= \{[(\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I_i}{N})\} + \{[(\frac{A\,\sigma^2}{2} + \beta^d)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i] + (A\sigma^2 \frac{R-I_i}{N})\} \\
&= (A\sigma^2 + 2\beta^d)\cdot Q + \frac{2\omega}{Q} + A\sigma^2 \cdot \frac{2R}{N}
\end{aligned}$$

So, the inter-dealer spread can be written as:

$$S_i^d = (A\sigma^2 + 2\beta^d)\cdot Q + \frac{2\omega}{Q} + A\sigma^2 \cdot \frac{2R}{N}$$

According to proposition 2.2, the customer optimal bid-ask prices are:

$$a_i^c = a_i^{r,c} + A\sigma^2 \cdot \frac{R+I_i}{N} + \delta_i c$$

$$b_i^c = b_i^{r,c} + A\sigma^2 \cdot \frac{R-I_i}{N} + \delta_i c$$

where $$a_i^{r,c} = (\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} - A\sigma^2 \cdot I_i$$

$$b_i^{r,c} = (\frac{A\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} + A\sigma^2 \cdot I_i$$

Therefore, complete customer bid-ask quotes can be written as:

$$p^0 + a_i^c = (p_{t-1} + r_t) + (a_i^{r,c} + A\sigma^2 \frac{R+I_i}{N} + \delta_i c) = (p_{t-1} + r_t) + [(\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I_i}{N}) + \delta_i c$$

$$p^0 - b_i^c = (p_{t-1} + r_t) - (b_i^{r,c} + A\sigma^2 \frac{R-I_i}{N} + \delta_i c) = (p_{t-1} + r_t) - [(\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i] - (A\sigma^2 \frac{R-I_i}{N}) - \delta_i c$$

Based on the two equations above, customer mid-quote can be derived as below:

$$\begin{aligned}
p_{t,i}^c &= \frac{1}{2}(p^0 + a_i^c + p^0 - b_i^c) = \frac{1}{2}\{[(p_{t-1} + r_t) + (\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I}{N}) + \delta_i c + \\
&(p_{t-1} + r_t) - [(\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i\,] - (A\sigma^2 \frac{R-I}{N}) - \delta_i c\} \\
&= \frac{1}{2}\{2(p_{t-1} + r_t) - 2A\sigma^2 + 2\frac{I_i}{N} A\sigma^2\} \\
&= p_{t-1,i} + r_t - A\sigma^2(1 - \frac{1}{N})\cdot I_i
\end{aligned}$$

And the customer spread can be calculated as:

$$\begin{aligned}
S_i^c &= (p^0 + a_i^c) - (p^0 - b_i^c) = a_i^c + b_i^c = \{[(\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} - A\sigma^2 I_i] + (A\sigma^2 \frac{R+I_i}{N}) + \delta_i c\} + \\
&\{[(\frac{A\,\sigma^2}{2} + \beta^c)\cdot Q + \frac{\omega}{Q} + A\sigma^2 I_i] + (A\sigma^2 \frac{R-I_i}{N}) + \delta_i c\} \\
&= (A\sigma^2 + 2\beta^d)\cdot Q + \frac{2\omega}{Q} + A\sigma^2 \cdot \frac{2R}{N} + 2\delta_i c
\end{aligned}$$

REFRERNCES

Andersen, T. and T. Bollerslev, 1998, "Answering the skeptics: yes, standard volatility models do provide accurate forecasts", *International Economic Review*, Vol.39, No.4, pp.885-905.

Andersen, T., and T. Bollerslev, 2002, "Micro Effects of Macro Announcements: Real-time Price discovery in Foreign Exchange", *NEBR working paper* 8959, May 2002.

Ahn, Dong-Hyun, 2004, "Common factors and local factors: implication for term structures and exchange rates", *Journal of Financial and Quantitative Analysis*, Vol39, No 1. March 2004, pp.69-102.

Backus, David, Foresi, Silverio and Telmer, Chris I. 2001, "Affine term structure models and the forward premium anomaly", *Journal of finance*, pp.279-304.

Baillie and T. Bollerslev, 1994, "The long memory of the forward premium", *Journal of International Money and Finance*, Elsevier, vol. 13(5), pp.565-571.

Baillie and T. Bollerslev, 2000, "The forward premium anomaly is not as bad as you think", *Journal of International Money and Finance*, Elsevier, vol. 19(4), pp.471-488.

Bansal, Ravi, 1997, "An exploration of the forward premium puzzle in currency markets", *Review of Financial Studies,* 10, pp.369-403.

Bekaert, G. and Hodrick, R., 1993, "On biases in the measurement of foreign exchange risk premiums", *Journal of International Money and Finance,* 12, pp.115-138.

Bessembinder, Hendrik 1994, "Bid-ask spreads in the inter-bank foreign exchange markets", *Journal of Financial Economics*, 35(3), pp. 317-348.

Biais, Bruno, 1993, "Price formation and equilibrium liquidity in fragmented and centralized markets", *Journal of Finance,* Vol 484, No.1 pp.157—185

Bilson, John, 1981, "The speculative efficiency hypothesis", *Journal of Business*, Vol. 54 pp.433-451.

Bjonnes, G., and D. Rime, 1998, "FX trading ... live: impact of new trading environments", typescript, Norwegian School of Management, December.

Black, Stanley W., 1991, "Transactions costs and vehicle currencies", *Journal of International Money and Finance*, Elsevier, vol. 10(4), pp. 512-526, 12.

Bollerslev, Tim and I. Domowitz 1993, "Trading patterns and prices in the inter-bank foreign exchange market", *Journal of Finance*, 48, pp.1421-43.

Bollerslev, Tim & Melvin, Michael, 1994, "Bid--ask spreads and volatility in the foreign exchange market: An empirical analysis", *Journal of International Economics*, Elsevier, vol. 36(3), pp. 355-372, 5.

Braas, A., and C.N. Bralver, 1990, "An Analysis of trading profits: How trading rooms really make their money", *Journal of Applied Corporate Finance*, 2 Winter, pp.85-90.

Brandnt, M., J. Cochrane, and P. Santa-Clara, 2003, "International risk sharing is better than you think", unpublished manuscript, UCLA.

Chinn, Menzie, D. and Meredith, Guy, 2004, "Monetary policy and long-horizon unconvered interest parity", *IMF Staff papers*, Vol 51, No 3 pp. 409-430.

Cornell, B., 1989, "The impact of data errors on measurement of the foreign exchange risk premium", *Journal of International Money and Finance*, Volume 8, Issue 1, March 1989, pp.147-157

Demos, Antonis A.; Goodhart, Charles A. E., 1996, "The interaction between the frequency of market quotations, spread and volatility in the foreign exchange markets", *Applied Economics*, March v. 28, iss. 3, pp. 377-86

Duffie, Darrell, and Rui, Kan, 1996, "A yield –factor model of interest rates", *Mathematical Finance,* 6, pp. 379-406.

Dai, Q., and K. Singleton, 2000, "Specification analysis of affine term structure models", *Journal of Finance*, 55, pp. 1943-1978.

Engel, Charles, 1996, "The forward discount anomaly and the risk premium: A survey of recent evidence", *Journal of empirical Finance* 3, pp. 123-191.

Evans, M., 2002, "FX trading and Exchange Rate Dynamics", *Journal of Finance*, December 2002, v. 57, iss. 6, pp. 2405-47

Evans, M., Lyons, R., 2002, "Informational Integration and FX Trading", *Journal of International Money and Finance*, v. 21, pp. 807-831

Evans, M., Lyons, R., 2002, "Order flow and exchange-rate dynamics", *Journal of Political Economy,* 110, pp. 170-180.

Fama, Eugene, 1984, "Forward and spot exchange rates", *Journal of Monetary Economics,* 14, pp. 319-338.

Faust, J. J. Rogers and J. Wright, "Exchange Rate Forecasting: The Errors We've Really Made," paper presented at conference on "Empirical Exchange Rate Models," University of Wisconsin, September 28-29, 2001.

Flood, M., 1994, "Market structure and inefficiency in the foreign exchange market", *Journal of International Money and Finance*, 13: pp. 131-158.

Flood, R., and M. Taylor, 1996, "Exchange rate economics: What's wrong with the conventional macro approach?", in Frankel, J., G. Galli, and A. Giovannini,(eds.) *The Microstructure of Foreign Exchange Markets* (Chicago: The University of Chicago Press), 261-294.

Frankel, J., and K. Froot, 1987, "Using survey data to test standard propositions regarding exchange rate expectations", *American Economic Review*, 77: pp. 133-153.

Frankel, J., G. Galli, and A. Giovannini, 1996, Introduction in *The Microstructure of Foreign Exchange Markets* (Chicago: The University of Chicago Press), 1-15.

Frankel, J., and A. Rose, 1995, "Empirical research on nominal exchange rates", in G.Grossman and K. Rogoff (eds.), *Handbook of International Economics*, Elsevier Science: Amsterdam, pp. 1689-1729.

Garman, Mark B., 1978, "The pricing of supershares", *Journal of Financial Economics*, March 1978, v. 6, iss. 1, pp. 3-10

Glassman Debra, 1987, "Exchange rate risk and transactions costs: evidence from bid-Ask spreads". *Journal of International Money and Finance*, Vol. 6, 1987, pp. 479-490.

Gourinchas, Pierre-Olivier and Tornell, Aaron, 2004, "Exchange rate puzzles and distorted beliefs", *Journal of International Economics*, 64, pp. 303-333.

Hartmann, Philipp, 1998, "Do Reuters spreads reflect currencies' differences in global trading activity?", *Journal of International Money and Finance*, October, v. 17, iss. 5, pp. 757-84

Hartmann, Philipp, 1999, "Trading volumes and transaction costs in the foreign exchange market", *Journal of Banking and Finance*, Vol. 23, pp. 801-824.

Hau, H., 1998, "Competitive entry and endogenous risk in the foreign exchange market", *Review of Financial Studies*, 11: pp. 757-788.

Ho, T., and H. Stoll, 1981, "Optimal dealer pricing under transactions and return uncertainty", *Journal of Financial Economics*, 9, pp. 47-73.

Ho, Thomas S. Y., and Hans R. Stoll, 1983, "The dynamics of dealer markets under competition", *Journal of Finance,* 38, pp. 1053−1074.

Huang. D Roger. and Ronald W. Masulis, 1999, "FX spread and dealer competition across the 24-Hour trading day", *The review of financial studies*, Vol 12 No. 1 pp. 61—93

Fan, M. and Lyons, 2000, “Customer-dealer trading in the foreign exchange market”, typescript, UC, Berkeley, July

Jorion, Philippe, 1996, “Risk and turnover in the foreign exchange market. The microstructure of foreign exchange markets”, pp. 19-37, National Bureau of Economic Research Conference Report series. Chicago and London: University of Chicago Press,

Laux, Paul A., 1995, “Dealer market structure, outside competition, and the bid-Ask spread”, *Journal of Economic Dynamics and Control*, May 1995, v. 19, iss. 4, pp. 683-710

Lewis, Karen K., 1995, “Puzzles in international financial markets”, in *Handbook of International Economics*, edited by Gene Grossman and Kenneth Rogoff, pp. 1913-71.

Lyons, R., 1995, “Tests of microstructural hypotheses in the foreign exchange market”, *Journal of Financial Economics,* 39, pp. 321-351.

Lyons, R., 2001, *The Microstructure Approach to Exchange Rates*, MIT Press.

McBrady, Matthew 2005, “Carry traders, dealers and the forward discount anomaly”, manuscript, University of Virginia, Darden Graduate school of Business.

MacCallum, Bennett T. 1994, “A reconsideration of the uncovered interest parity relationship”, *Journal of Monetary Economics*, Vol 33, pp. 105-132.

Marmol, Francesc, 1998, “Spurious regression theory with nonstationary fractionally integrated processes”, *Journal of Econometrics*, Volume 84, issue 2, pp. 233-250

Mark, N., 1995, “Exchange rates and fundamentals: evidence on long-horizon predictability”, *American Economic Review*, 85: pp. 201-218.

Maynard, Alex and Phillips, Peter, 2001 “Rethinking an old empirical puzzle: econometric evidence on the forward discount anomaly”, *Journal of Applied Econometrics*, Vol. 16, No. 6, pp. 671-708.

Meese, R., 1990, “Currency fluctuations in the post-Bretton Woods era”, *Journal of Economic Perspectives*, 4, pp. 117-134.

Meese, R., and K. Rogoff, 1983a, “Empirical exchange rate models of the seventies”, *Journal of International Economics*, 14: pp. 3-24.

Meese, R., and K. Rogoff, 1983b, “The out-of-sample failure of empirical exchange rate models”, in J. Frenkel (ed.), *Exchange Rate and International Macroeconomics*, University of Chicago Press: Chicago.

Meese, R., and K. Rogoff, 1988, "Was it real? The exchange rate-interest differential relation of the modern floating-rate period", *Journal of Finance*, 43: pp. 933-948.

Obstfeld, M., and K. Rogoff, 1995, "Exchange rate dynamics redux", *Journal of Political Economy*, 103: pp. 624-660.

Obstfeld, M., and K. Rogoff, 2000, "The six major puzzles in international macroeconomics: is there a common cause?", NBER working paper 7777.

Osler, C., 1998, "Short-term speculators and the puzzling behavior of exchange rates", *Journal of International Economics*, 45: pp. 37-57.

Osler, C., 2006, "Macro Lessons from Microstructure", *International Journal of Finance and Economics*, January, pp. 55-80

Payne, R. 1999, "Informed trade in spot foreign exchange markets: An empirical investigation", typescript, London School of Economics, January.

Stoll, Hans R., 1978, "The supply of dealer services in securities markets", *Journal of Finance,* Sept. 1978, v. 33, iss. 4, pp. 1133-51

Tsay, Wen-Jen and Chung, Ching-Fan, 2000, "The spurious regression of fractionally integrated processes", *Journal of Econometrics*, 96(1), pp. 155-182.

Vitale, Paolo, 1999, "Sterilized central bank intervention in the foreign exchange market", *Journal of International Economics,* 1999, pp. 245*–267*.

Wei, Shang-Jin, 1994, "Anticipations of foreign exchange volatility and bid-Ask spreads", *NBER Working Paper* No. w4737

Yao, J. 1998, "Market making in the inter-bank foreign exchange market", New York University Salomon Center Working Paper #S-98-3.