ETLA The Research Institute of the Finnish Economy

Antti Suvanto

FOREIGN EXCHANGE DEALING

Essays on the Microstructure of the Foreign Exchange Market

A19

Helsinki 1993

ETLA ELINKEINOELÄMÄN TUTKIMUSLAITOS
The Research Institute of the Finnish Economy
Lönnrotinkatu 4 B 00120 Helsinki Finland

Sarja A 19 Series

Antti Suvanto

FOREIGN EXCHANGE DEALING

Essays on the Microstructure of the Foreign Exchange Market

ETLA, The Research Institute of the Finnish Economy
Publisher: Taloustieto Oy

Helsinki 1993

ISBN 951-9206-98-1
ISSN 0356-7435

Printed in Tampere by Tammer-Paino Oy 1993

Suvanto, Antti, *Foreign Exchange Dealing: Essays on the Microstructure of the Foreign Exchange Market.* The Research Institute of the Finnish Economy (Elinkeinoelämän tutkimuslaitos), ETLA, Helsinki 1993, 132 p. (A, ISSN 0356-7435; No 19). ISBN 951-9206-98-1.

ABSTRACT: The study examines the intra-day pricing and position adjustment behaviour of an individual foreign exchange dealer. The analysis in Chapters Two and Three rests on two key assumptions, the price sensitivity of customers' buy and sell orders and profit maximization on the part of the dealers subject to a closed position constraint for the end of the trading day. In Chapter Two, the dealer operates in monopolistic conditions, whereas in Chapter Three the model is opened up to allow for customer flows and interbank transactions. Special attention is paid to the adjustment of two-way prices to changes in the position and to the determination of the bid-ask spread. In Chapter Four the price-making aspect of dealer behaviour is disregarded and the emphasis is shifted to the costs of producing dealer services. The analysis is analogous to that of the transactions demand for money under stochastic cash flow. Chapter Five extends the analysis to an *N*-currency world by considering the problem of detecting profitable arbitrage chains when there is a positive bid-ask spread and when the currencies are quoted against each other simultaneously by a large number of dealers who do not know how other dealers are quoting at the same moment. The motivation of the study, as well as the summary of the four analytical essays, are presented in the introductory essay.

KEY WORDS: Foreign exchange market, financial market microstructure, market, market maker, interbank market, foreign exchange position, arbitrage

Suvanto, Antti, *Foreign Exchange Dealing: Essays on the Microstructure of the Foreign Exchange Market.* The Research Institute of the Finnish Economy (Elinkeinoelämän tutkimuslaitos), ETLA, Helsinki 1993, 132 s. (A, ISSN 0356-7435; No 19). ISBN 951-9206-98-1.

TIIVISTELMÄ: Tutkimuksessa tarkastellaan yksittäisen valuuttadealerin hinnoittelukäyttäytymistä ja position kontrollia kaupankäyntipäivän aikana. Toisen ja kolmannen luvun analyysi nojautuu kahteen keskeiseen oletukseen, asiakkaiden osto- ja myyntipyyntöjen hintajoustavuuteen sekä dealerin voitonmaksimointiin rajoituksella, että valuuttapositio suljetaan päivän lopussa. Toisessa luvussa dealerin oletetaan olevan monopoliasemassa markkinoilla, mutta kolmannessa luvussa malli avataan kilpailulle sallimalla asiakassiirtymät ja pankkien väliset markkinat. Huomiota kiinnitetään erityisesti kaksisuuntaisten hintojen riippuvuuteen valuuttaposition kehityksestä sekä osto- ja myyntikurssien välisen eron määräytymiseen. Neljännessä luvussa sivuutetaan hinnoitteluongelma ja tarkastellaan valuuttakaupan kustannuksia. Ongelma esitetään analogisena rahan transaktiokysyntäongelman kanssa stokastisen kassavirran tapauksessa. Viidennessä luvussa analyysi laajennetaan usean valuutan olosuhteisiin kiinnittämällä huomiota vaikeuteen löytää voittoa tuottavat arbitraasiketjut tilanteessa, jossa osto- ja myyntikurssin välinen ero on positiivinen ja jossa useat dealerit antavat kaksisuuntaisia hintanoteerauksia tietämättä, miten toiset noteeraavat samalla hetkellä. Tutkimusaiheen perustelut samoin kuin yhteenveto on esitetty johdantoluvussa.

ASIASANAT: Valuuttamarkkinat, rahoitusmarkkinoiden mikrorakenne, markkinatakaaja, pankkien väliset markkinat, valuuttapositio, arbitraasi

ACKNOWLEDGEMENTS

The history of this monograph dates back to early 1980s when I was involved in empirical research on the international operations of Finnish commercial banks. This project, which was published as a monograph in the series of the Research Institute of the Finnish Economy (ETLA) in 1983, left behind a series of preliminary notes and working papers for further development on a suitable future occasion. Such an occasion arose in 1982 when I had an opportunity, with the financial support of the Cooperative Banks Research Foundation, to spend the spring term at the C.V. Starr School of Economics of New York University. Thereafter, similar opportunities never really arose, as I repeatedly found the more worldly affairs of applied research, banking and later central banking much more exciting.

This is not meant to imply that the monograph has been produced entirely in odd hours of late evenings or during weekends. I am grateful to my present and previous employers, who were generous enough to allow me to devote periods of one to two months for full-time research. In this respect, I would especially like to express my gratitude to Pentti Vartia of ETLA, Lauri Siltala of the Savings Banks Association, and to Sirkka Hämäläinen and Johnny Åkerholm of the Bank of Finland, who all encouraged me, and even put pressure on me, to escape from my day-to-day responsibilities in order to complete the thesis.

Over the years, I have discussed various topics of the monograph with a large number of colleagues and friends. I am especially grateful to Kari Alho, Pentti Haaparanta and Urho Lempinen, who all contributed to the early stage of the work, although at that stage I did not even plan to develop it into a monograph. At a later stage, Pentti Kouri and Seppo Salo, the official examiners of the thesis, contributed in a decisive way to the monograph's final form. I owe a special debt of gratitude to both of them. I appreciate the discussions with Gavin Bingham, whose insight and knowledge have taught me a lot over the years. Various parts of the monograph have been presented in a number of academic seminars and workshops both in Finland and elsewhere. All these occasions have been enjoyable, encouraging and useful for the conduct of the study.

I am grateful to my many friends both at ETLA and at the Bank of Finland who during various phases of the work assisted me in technical matters. Tuula Ratapalo typed the earlier versions of the monograph with skill and care, while Marita Castrén took care of the final processing of the text. Arja Virtanen drew the diagrams. Marketta Rautio has provided invaluable secretarial support. John Rogers and Malcolm Waters have both participated, at different stages, in polishing my English.

Although the study is as far from practical dealing as possible, I have enjoyed and benefitted from discussions with those friends of mine in Helsinki, London, Luxembourg and New York who are or have been involved in foreign exchange dealing in practice. When I started the study I had hardly any connections with the profession at all. When, at a later stage of my career, I became involved in banking, I found it easy and straightforward to talk with the dealers and to comprehend their work. I also learned what the difference is between theory and practice. I want to express my gratitude to my dealer friends and my respect for their work.

Although the work started ten years ago, it has not been a ten-year project. Nonetheless, it has been a ten-year burden for my family. The few short periods when I had an opportunity to conduct full-time research were periods of full-time absenteeism from the family, while in between the project was too often a good excuse to escape family responsibilities. For this reason, I owe a deep debt of gratitude to my wife Ritva and to my son Aki for their patience and understanding.

November 1993
Watermael-Boitsfort, Brussels

Antti Suvanto

CONTENTS

Chapter One

INTRODUCTION

The foreign exchange market is an arrangement which facilitates international monetary exchange in a world where there is no single universally accepted means of payment (Einzig, 1969, p. 15, and Hirsch, 1969, p. 198). The market exists to serve the currency conversion needs of the inhabitants of different countries engaged in economic contracts with each other. Merchants involved in exporting and importing, investors diversifying their portfolios internationally, speculators seeking to profit from exchange rate fluctuations and tourists travelling abroad, all constantly face situations where they need to convert one currency into another.

A distinct feature of the foreign exchange market is that it is almost entirely intermediated. The ultimate buyers and sellers of currencies seldom meet face to face. Instead, they place buy and sell orders with commercial banks, whose foreign exchange departments stand ready to trade on the basis of incoming orders on immediate demand.

The buy and sell orders received by any single bank do not usually match in any short period. For many banks the flow of buy orders for a given currency may systematically exceed the flow of sell orders. This mismatch of market orders gives rise to the interbank market, where individual banks can acquire the currencies currently oversold or sell the currencies currently overbought.

In addition to the trade between banks and their customers and the interbank business, central banks also participate in the foreign exchange market. Their primary interest is to ensure that their currencies trade in a stable fashion *vis-à-vis* other currencies. The practices may vary from country to country, depending on monetary and exchange rate policy considerations. In addition to intervention, central banks may, and indeed do, use the foreign exchange market for other purposes as well. They may transact business for other central banks, or they may simply speculate in currencies other than their own (*cf.* Sender, 1985).

Although the central banks' participation may from time to time constitute a significant proportion of the turnover in the foreign exchange market, normally the volume of trade arising from official operations is small in relation to the regular flow of international payments and interbank business. Hence, despite the important macroeconomic role central banks may play in creating a stable environment for international monetary exchange, they are not the key actors in the foreign exchange market. Perhaps the most important role played by national authorities in the international money game is that, being sovereign issuers of national monies, the governments of the industrialized countries and of many less developed countries have committed themselves to the free convertibility of their currencies (*cf.* McKinnon, 1979).

The value of world trade in goods in 1989 was estimated to be about USD 3,000 billion a year (GATT, 1990). Taking into account the trade in services, interest payments and non-bank transactions in the capital account, the value of non-bank transactions of a cross-border nature was probably four or five times this amount. In addition, a large number of trade-related currency conversions take place without being recorded as flows through the balance of payments or being registered in any official statistics. These include foreign exchange transactions in two currencies outside their respective home countries. Hence, the order of magnitude of the currency conversions of the non-bank public may have amounted to USD 15,000 billion meet in 1989, or USD 40 to 60 billion a day. This is the amount the foreign exchange market handles each day in order to meet non-bank customer demand.

Altogether, the total turnover in the foreign exchange market, adjusted for interbank double counting, was estimated to be USD 600 to 700 billion a day in April 1989 (BIS, 1990). Given that the market is open 24 hours a day, this means that some USD 20 to 25 billion changes hand in any single hour. Only around 10 per cent of this trade is directly linked with customers' needs, while the remaining 90 per cent represents professional trading between banks.

The object of trade in the foreign exchange market is money, which is "the most fluid and most easily movable of all commodities" (Hirsch, 1969, p. 197). This is another way of saying that the transaction costs of trading money are low. Low transaction costs imply that the effective price paid by the buyer is close to the effective price received by the seller. In the foreign exchange market this means that the *bid-ask spread* is very small. Under competitive conditions and without any artificial obstacles to foreign exchange transactions, the spread should not be larger than the marginal cost of trading in currencies. When this requirement is fulfilled, the foreign exchange market is *functionally* efficient, using the terminology introduced by Tobin (1984).

Another important aspect of the efficiency of the foreign exchange market is related to information and arbitrage. According to the most widely used definition, a market is said to be efficient, if it is impossible to predict future price movements on the basis of publicly available information (Samuelson, 1965, and Fama, 1970). To this one may want to add a further requirement that in an efficient market it is, on average, impossible to gain from spatial price differentials which may prevail across marketplaces. When these requirements are fulfilled the market is *information-arbitrage* efficient, again using Tobin's terminology. This type of efficiency is not entirely independent of functional efficiency, because information costs depend on transaction costs. In the foreign exchange market, in particular, transacting is one way of acquiring information. Low transaction costs increase the volume of trade and improve the quality of information. Indeed, it will be argued below that the impressive size of the interbank market is, to a large extent, understandable in terms of the exchange of information.

When the bid-ask spread is very small and the exchange rate quotations are practically equal everywhere, firms and households can regard their domestic money balances as if they represented international means of payment (McKinnon, 1979, pp. 4-13). Under these conditions, international monetary exchange does not in any essential way differ from monetary exchange in general. This is the theme the present study endeavours to explore.

1 Aims and Scope of the Study

Transaction costs, imperfect information and other economic frictions are facts of life. But they are also important theoretical concepts, without which many economic phenomena and institutional arrangements could not be properly understood. The aim of the present study is to apply microeconomic tools to the analysis of the foreign exchange market, which is interpreted as an institutional arrangement understandable in terms of transaction costs and imperfect information. The focus of the analysis is on the behaviour of those agents who actually make the market, *i.e.*, foreign exchange dealers in commercial banks.

Transaction costs play a role in three respects. First, the very existence of dealers presupposes that they are able to provide currency conversion services to customers at a lower cost than would be possible under alternative arrangements. Secondly, customers' transaction costs depend on their distance from dealers, which gives rise to local markets. Thirdly, exchange dealing is a resource-using activity, which draws attention to the costs of producing dealer services.

Imperfect information plays a role in a number of respects. First, customers' buy and sell orders arrive at a stochastic rate, which means that dealers have to cope with transactions uncertainty. This is the basic feature of the dealership market. Secondly, dealers face instantaneous price uncertainty, because they do not know what prices their competitors are quoting and dealing at the same time as they are proposing their own prices. In addition, they face temporal price uncertainty, because they do not know at what level exchange rates will settle in the future. This, in turn, depends on changes in and expectations about the macroeconomic environment, which determines the stochastic processes generating nonbank customers' buy and sell orders.

The sensitivity of customers to price quotations and the profit motive of dealers are the key behavioural assumptions used in the study. These two assumptions are sufficient for constructing an analytic model of the behaviour of an individual dealer. Small variations in the assumptions specifying the environment in which the dealer operates allow us to trace the sources of the above-mentioned efficiency properties of the foreign exchange market. Inter-dealer competition limits the scope within which each single dealer can exercise his price-making power and increases the importance of interbank transactions. As a result of the latter, large local transactions shocks lead to small global price reactions. Major transactions shocks are diffused through the interbank market not only spatially across financial centres, but also temporally across time-zones.

The analysis proceeds step by step. Although each separate essay examines the dealer's behaviour in specific circumstances, the essays are closely related and complementary. It may be useful, already at this stage, to warn the reader about the limitations of the study and to notice that there are many areas the study does not attempt to cover.

The emphasis is on the medium-of-exchange function of money. Therefore, the analysis is confined to spot transactions alone[1]. This simplification appears justified in the sense that other international financial operations in which commercial banks are involved, such as deposit dealing and forward transactions, normally lead to spot transactions at some stage. The simplification is not, however, entirely innocent, because these other contracts affect the dealer's future currency flows in a predictable way, which may affect the dealer's behaviour in advance.

Despite much talk about how dealers speculate in currencies, the present study focusses on the less exciting side of their daily work, their behaviour as market makers, providing liquidity services to customers. This means that dealers are assumed to be interested primarily in maximizing their regular trading income, and not particularly disposed to taking risks. In fact, fairly strong assumptions are made about the risk aversion of dealers. In the formal models this is taken into account by specifying the *closed position* target for the end of the day. There is ample room for intra-day and overnight speculation in the present framework, but it is not an integral part of the analysis.

Finally, the study is silent on practically all macroeconomic issues. Although dealers actually set prices, they have very little or no influence on the average levels of exchange rates. These are determined by the macroeconomic structure that generates customers' buy and sell orders. Dealers may be quick to react to any new information about changes in the macroeconomic environment, but their price-setting behaviour *as such* does not feed back to this environment. In fact, the formal models are designed to be consistent with any macroeconomic structure which generates a continuous, though stochastic, flow of two-way market orders from the

---

[1] In practice, the foreign exchange market is defined to cover spot and forward transactions, as well as currency options and futures. According to a number of surveys, roughly two-thirds of total foreign exchange turnover represents spot transactions, while the share of futures and options trade is very small (Bell and Kettel, 1983, Tygier, 1983, and BIS, 1990).

non-dealer public.[2] Dealers need not respect any fundamentals in the sense that any one of them would gain by setting exchange rates at those levels which they believe to be the *true* price. In other words, *fundamental-valuation* efficiency, using Tobin's terminology, is not an issue when the market is examined from the micro perspective.

Because the study aims at improving our grasp of the way the foreign exchange market is organized, or at least one important aspect of it, in the following we will discuss very briefly the foreign exchange market as a mechanism for international monetary exchange. We then go on to explain the principles of the dealership market, which is the starting point for much of the subsequent analysis. The overview and summary are presented at the end of the introductory essay.

2 International Monetary Exchange

Exchange is one of the most fundamental concepts in economics. It is the source of the division of labour and a prerequisite for any analysis of markets and prices. As such, it is an abstract notion and leaves open the question of how exchange is organized. The possibilities range from bilateral barter to monetary exchange. The social benefits of the latter arrangement, which stem from circumventing the *double coincidence of wants*, are well established in the literature.

International trade is just one form of exchange, the benefits of which are equally well established. While much of the literature has been written as if cross-border exchange were barter trade between nations, the analysis of multilateral trade requires some description of the payments mechanism and hence of the use of money in one form or another.

International monetary exchange is characterized by the absence of a single universally accepted medium of exchange. Although some national monies may be more internationally acceptable than others, no currency is generally good for all purposes in all geographical regions. Monetary history shows that this has never been the case (Einzig, 1970).

[2] An analytic description of the foreign exchange market as a market process requires a macroeconomic theory which is capable of explaining the supply of and the demand for currencies in flow terms; *cf.* McKinnon (1979, pp. 13-20) and Kouri (1983, pp. 116-124).

For foreign trade to be based on monetary exchange in a multi-currency world, it is necessary that there are buyers (importers) who are able to make payments in foreign currencies and/or sellers (exporters) who are willing to accept foreign money in exchange for goods sold abroad. As a large proportion of the cash payments of exporting firms and a large proportion of the cash receipts of importing firms are in their domestic currencies, either some of the former have to accept currencies they do not need, or some of the latter have to pay in currencies they do not earn. Both groups have a common interest in converting one currency into another. Therefore, the basic issue is how these diverse needs can meet (*cf.* Chacholiades, 1978, p. 5, Riehl and Rodriguez, 1983, pp. 11-13).

At a very general level, we are back in a situation which forms the starting point for the analysis of exchange economies, where there are a large number of individuals, each having an endowment of *goods*. These goods may be of little or no value to each individual taken separately, but they may represent substantial values for the group as a whole. In this kind of a situation, the *Walrasian auctioneer* steps in and announces price vectors until the excess demands disappear and the goods are delivered.

For many purposes it is perfectly adequate to characterize the foreign exchange market as being coordinated by the Walrasian auctioneer, who collects information about the currency conversion needs of individuals and quotes prices until everybody is happy and the conversions can take place. But in the same way as money becomes unnecessary in exchange economies with full information and zero transaction costs, foreign exchange transactions would become unnecessary, if this costless clearing arrangement existed. Therefore, any analysis of the foreign exchange market, as that of monetary exchange, must start from those imperfections that prevent the Walrasian auctioneer from fulfilling the task of coordinating individual excess demands.

In monetary theory the origin of money is explained as a result of a learning process by which individuals learn to save on transaction costs. Compared to bilateral barter, the transaction costs are lowered when traders first learn to gather in certain physical locations (fairs) in order to exchange commodities. Later, they start to accept intermediate commodities, which they may not need, but which may increase the probability of an acceptable exchange on some future occasion. This process leads to monetary exchange once one of the intermediate commodities acquires a dominant role, *i.e.*, becomes a universally accepted means of payment.[3]

In principle, the same model should be applicable to the analysis of the evolution of international monetary exchange.[4] For example, it seems obvious that fulfilling the double coincidence of conversion needs would become easier if the individuals involved in exporting and importing learned to gather in *bourses* in order to exchange currencies. It is also conceivable that, in the absence of suitable counter-offers, some traders would be willing to accept currencies other than those ultimately needed because it might increase the probability of an acceptable conversion in the near future. Frequent indirect converversions of this type might gradually lead to the emergence of a widely accepted international medium of exchange, a *vehicle currency*.

However, as long as national monies exist and there is a *preferred monetary habitat* where domestic currency is used for domestic transactions, a vehicle currency would not free traders from the double coincidence of conversion needs, although in many cases it would make matters much easier. For example, an importer having to make payments to many countries would need to find only one counterpart who is willing to take the importer's currency in exchange for the vehicle currency. On the other hand, sellers might not always be willing to accept the vehicle currency in payment, because for them it would imply extra costs to find buyers for that currency.[5]

These observations suggest that there are limits as to how far the standard story about the emergence of a universally accepted means of payment is applicable to the international payments mechanism. There is, however, another solution. Since the fundamental problem is that traders with mutual interests may not be aware of each other, or it may be costly for them to meet, a profit opportunity exists for any third party to intervene

---

[3] For the theory of the origin and evolution of money, see Clower (1969), Brunner and Meltzer (1971), Niehans (1971), Jones (1976) and Nagatani (1977).

[4] As a matter of fact, it seems that the history of money is to a large extent the same as the history of international trade; *cf.* Einzig (1970).

[5] Swoboda (1968) presents a model, based on the theory of the transactions demand for money, which shows how commercial traders can save in transaction costs if they are able to use one currency, instead of many, for international payments. However, he does not discuss how the currency conversion needs become coordinated.

to enable the currency holders to carry out the desired transaction. Some individuals may find it profitable to specialize entirely in this kind of intermediation activity.[6]

The availability of intermediation services frees importers and exporters from the double coincidence of wants. This is now shifted on to specialized money traders.[7] This saves costs because pooling customers' buy and sell orders allows economies of scale to be reaped in matching counterbalancing orders at acceptable prices. The specialists may meet each other in bourses and act as *brokers* trading for their customers' account, or they may hold inventories of currencies and act as *dealers* trading for their own account. The trade carried on by specialists with each other gives intermediation its universal nature by bringing all participants indirectly together in a single foreign exchange market.

Universal intermediation plays the same role in the international economy as money plays in a national economy, or in the imaginary one-currency world economy. Money does not only bridge the gap between sales and purchases, but it also represents immediately disposable purchasing power for any purpose whatsoever. Universal intermediation facilitates the mismatch of the currency composition of payments and receipts, and it also provides commercial customers with means to use their national monies for any unspecified purpose in other currency areas.

3 On the Dealership Market

The above discussion used some analogies with monetary theory to sketch a general framework for the intermediated foreign exchange market. In this section the focus will be on the intermediation activity itself. Key ideas are

[6] This is just one example of a situation, well-known in microeconomics, where a middleman can profit from bringing the ultimate transactors together; *cf.* Malinvaud (1972, pp. 155-156) and Hirshleifer (1973).

[7] Chrystal (1984) applies Jones's (1976) evolutionary model to the evolution of a vehicle currency in the foreign exchange market. In his model foreign exchange dealers pool customer orders and spend time in searching amongst other dealers for a complementary trade. A vehicle currency emerges as each dealer seeks to structure his trade in order to minimize these search costs.

borrowed from the microstructure literature dealing with the analysis of market maker behaviour in the organized securities market.[8]

Security dealers are specialists who differ from brokers in that, rather than matching individual buy and sell orders directly, they stand ready to buy and sell on immediate demand by trading for their own account. Their presence saves customers from waiting, and hence lowers the costs of making transactions. In the words of Demsetz (1968), security dealers produce immediacy or liquidity services. The same applies to dealers in the foreign exchange market.

The dealership market can be illustrated by means of a simple diagram. In *Figure 1.1* the *DD*-curve represents the flow demand for foreign currency (the flow supply of domestic currency). The *SS*-curve represents the flow supply of dollars (the flow demand for domestic currency). The curves refer to the currency conversion needs of the residents of different countries during a given short period.[9] The exchange rate s is defined as the price of one unit of foreign currency (*dollar*) in terms of domestic currency (*marks*).

If those who need to convert one currency into another could communicate without costs, the two curves would, in each short period, determine the equilibrium exchange rate s^* and the equilibrium volume of trade q^*. Because this costless communication is not possible, commercial customers place their orders with foreign exchange dealers, who stand ready to trade on immediate demand.

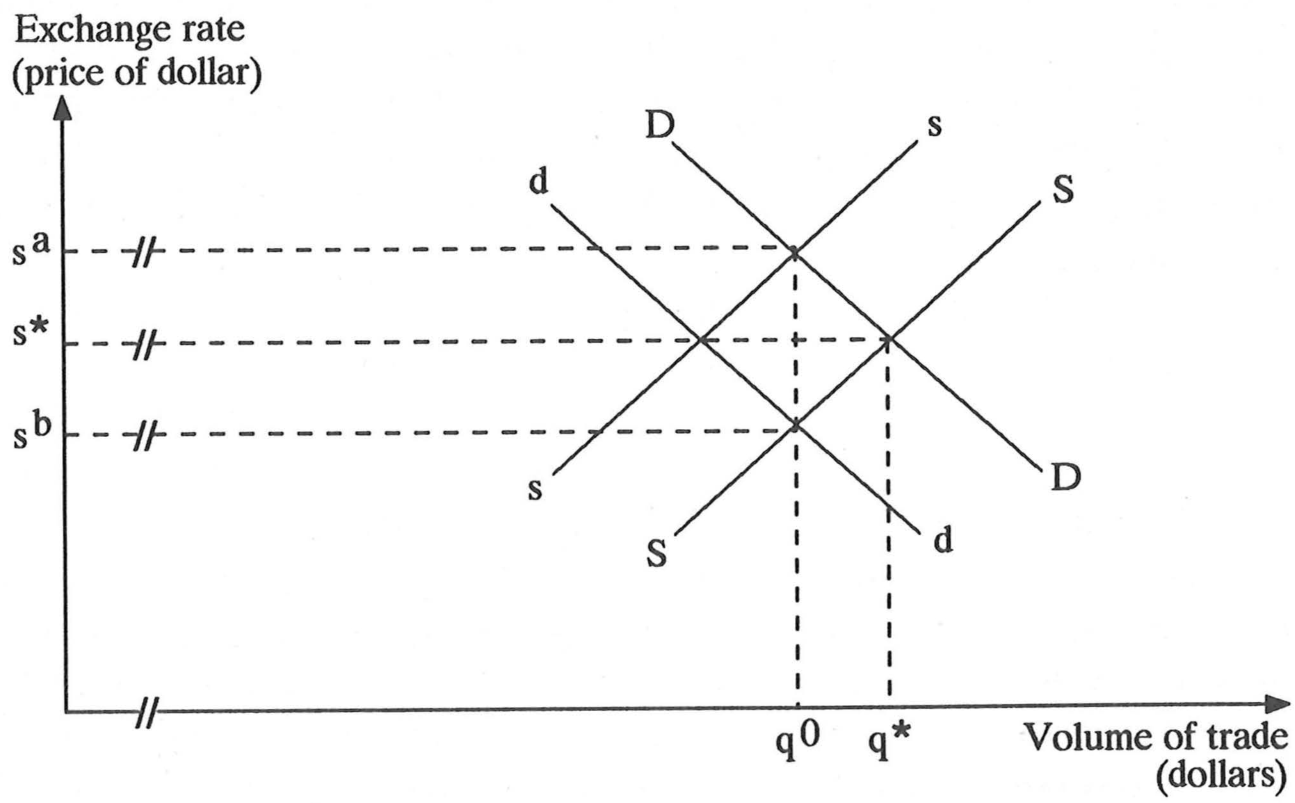
The dealers do not purchase dollars (or marks) in order to hold them or to spend them on goods, but rather to sell them back to customers at a higher price. Therefore, the dealers' supply curve (*ss*-curve) lies everywhere above the customers' supply curve by a given distance. Similarly, the dealers' demand curve (*dd*-curve) lies everywhere below the customers' demand curve by the same distance.

---

[8] Baumol (1958) was perhaps the first to draw attention to the important role stock dealers play in the securities market. The most frequently cited reference, however, is Demsetz (1968). More recent contributions include Tinic (1972), Garman (1976), Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1980, 1981), Zabel (1981), Glosten and Milgrom (1985), Grossman and Miller (1988), Glosten (1989), Dennert (1989), and Hagerty (1991). For surveys, see Cohen, Maier, Schwartz and Whitcomb (1979), and Stoll (1985).

[9] This illustration is borrowed from Demsetz (1968). For an application to the foreign exchange market, see Levich (1979).

Figure 1.1 The Dealership Market: An Illustration



The distance between the buying price and the selling price is called the *bid-ask spread.* For a given spread, the equilibrium selling price s^a, the *ask-rate,* is determined by the intercept of the dealers' supply curve and the customers' demand curve. This is the price at which the dealers are willing to exchange dollars for marks. Similarly, the equilibrium purchase price s^b, the *bid-rate,* is determined by the intercept of the dealers' demand curve and the customers' supply curve. This is the price the dealers are prepared to pay in order to buy dollars in exchange for marks. When transactions are made at these prices, the incoming buy and sell orders match as long as the structure which generates the customer orders remains constant.

Two further aspects have to be taken into account in interpreting the diagram. First, the *DD*-curve and the *SS*-curve do not represent always-present market orders. Rather, they represent the average arrival rates of incoming orders generated by a given macroeconomic environment. In any short period, the actual rate of incoming orders may differ from the average one. Secondly, as market makers, the dealers quote prices upon request and generally stand ready to transact at these prices if the customer decides to buy or sell. Therefore, the *DD*- and *SS*-curves correspond to the dealers' expectations about their average levels, and the *dd*- and *ss*-curves are set accordingly. Equilibrium prevails when expectations are correct on average, but they need not be correct in any short time interval. This is where transactions uncertainty enters the picture. Price uncertainty stems from the fact the dealers' expectations about the average flow of orders may prove to be wrong. This may happen because new information hits the market unexpectedly, or because successive customer orders begin to depart from the expected path.

Figure 1.1 also shows that the average volume of currency conversions q^o depends negatively on the bid-ask spread. This results from the sensitivity of customers to the levels of exchange rates. Using the industry analogy, it means that the demand for dealer services depends negatively on its price, which is represented by the spread.

While the demand side of the market for dealer services is straightforward, the supply side is more difficult to handle analytically. It requires a theory of the behaviour of an individual dealer, including the identification of relevant costs associated with dealing activity. Leaving aside the fixed costs related to the capital equipment and highly specific human skills, the major variable cost arises from the opportunity cost of holding currencies (including trading balances in domestic currecy). These are needed for the dealers to be able to make transactions on immediate demand.

Finally, the specification of the market conditions under which the dealers operate is very important, because they determine to what extent each individual dealer can exercise his potential price-setting power. The decisive factors in this respect are the absence of exchange controls, the degree of competitition between the dealers, and the functioning of the interbank market.

4 Overview and Summary

Pricing Behaviour and Position Adjustment

Figure 1.1 provides a useful starting point for the analysis of the pricing behaviour of an individual foreign exchange dealer. In *Chapter Two* it is assumed that the dealer operates in isolated circumstances and is able to fully exploit the monopoly power created by the price sensitivity of the customers' buy and sell orders. The dealer is assumed to maximize his trading income subject to the constraint that his expected foreign exchange position is closed at the end of the trading day or meets some other fixed target. Trading income is the revenue generated by a sequence of successive purchases and sales. A closed position is a state in which the dealer's holdings of foreign currencies are hedged against any unexpected change in the processes governing the customers' average buy and sell orders.[10]

Dynamic programming techniques are applied to derive the optimal pricing rule in two types of situation. In the first situation the dealer faces pure transactions uncertainty and prices are adjusted sequentially in response to any net sales or net purchases that arise because of the stochastic flow of customer orders. In the second situation the dealer has *private* information on the sequence of customer orders during the day. A quotation involves two numbers, the ask-rate and the bid-rate, both of whicvh are adjusted.

According to the results, the mid-rate is adjusted to steer the expected position towards the target, whereas the spread is always chosen to maximize the return on the equilibrium volume of one-period trade. In other words, the spread reflects the dealer's monopoly (pricemaking) power, whereas the dynamic adjustment of the mid-rate follows a kind of *tatonnement* rule in real time.[11] This implies that earlier net sales (a short position) lead to an upward adjustment of prices while earlier net purchases (a long position) lead to a downward adjustment of prices during the remaining part of the trading day.

A quick return to the closed position is non-optimal, implying that the dealer wants to avoid large swings in quotations from one short period to the next. In the case of serially uncorrelated transactions disturbances, the pricing rule implies that, after an unexpected shift in the position, the

---

[10] Formally, the position is closed if assets denominated in a given currency are equal to the liabilities denominated in the same currency. Because two currencies are involved in an exchange transaction, two positions are being affected. From the dealers point of view, the closed (squared) position can be interpreted as a requirement for zero net sales over some relatively short time period, say, a trading day. The open position can be long, in which case the dealer has overbought the given currency, or short, in which case this currency has been oversold; *cf.* Einzig (1969, pp. 25-26), Hudson (1979, p. 44), and Riehl and Rodriguez (1983, pp. 51-54).

[11] Note that, although the dealer's job is to make decisions on prices, the position target makes the price adjustment follow the *Walrasian* principle: $dp/dt = f(D-S)$, $f' > 0$, where $D-S$ refers to the excess demand that has already occurred (realized net sales). Unlike in the case of an auction system, $f(0)$ need not be equal to zero in any short trading period; *cf.* Arrow (1959) and Zabel (1981).

expected, or planned, quotation will be constant during the remainder of the day. This implies that the expected quotation does not differ from the latest announced quotation and there is no reason for a customer to postpone the transaction in anticipation of better prices later on. In other words, the optimal pricing rule is consistent with information-arbitrage efficiency, even in the very short run.

Of course, under transactions uncertainty the quotations will never be constant, but the price fluctuations are substantially restrained, compared with a situation where the dealer would immediately try to undo the open position that has arisen because of earlier net sales. Recall that in order to be able to buy and sell on immediate demand the dealer has to hold inventories of currencies. The dampening effect on the price fluctuations implied by the optimal pricing rule hence reflects the *buffer stock* property of inventories.

In the case where the dealer has private information on the sequence of future customer orders, this information is fully reflected in the *shadow price* for the position constraint. This is constant for given information, although the spread and the mid-rate may fluctuate according to the volume of trade in each period. It is, however, essential that information is private, because otherwise the customers would anticipate the dealer's future quotations, and therefore the sequence of customer orders would alter until predictable price changes disappear. While transactions uncertainty has only a minor effect on the volatility of price quotations, the arrival of new information on future customer orders affects the quotations immediately. The shadow price and the mid-rate both jump promptly and fully in the direction that depends on whether it is customer puchases (prices jump up) or customer sales (down) which are expected to rise at some later moment.

Although the formal analysis is carried out for a one-day horizon, the results are directly applicable to a horizon of any length, as long as an assumption of a closed overnight position target is maintained. New information on future customer net demand affects all price quotations already today. Speculation with open overnight positions can be incorporated into the model, although the payoff of such an exercise appears to be small. It can be handled more easily in other frameworks which do not require any formalization of intraday pricing behaviour.

Although the circumstances referred to above are far from those characterizing the near-perfect foreign exchange market in reality, the results of *Chapter Two* are important for the analysis of more complicated and more realistic situations. These are analyzed in *Chapter Three*.

The price-setting power of a single dealer can be justified by transaction and information costs, or by artificial obstacles such as exchange controls, which prevent customers from trading outside their own location. When these costs are sufficiently low, the buy and sell orders of customers received by an individual dealer become increasingly price sensitive.[12] The same holds if in the same local market there are many dealers who compete with each other. It is easily seen that increased price sensitivity implies a narrower spread, *i.e.*, the price of dealer services becomes lower. In other words, interdealer competition reduces the price-setting power of each individual dealer in the same manner as monopoly power is reduced by the existence of close substitutes in standard microeconomic analysis.

In the above case the fact that customers are able to shop around by asking for quotations from more than one dealer signals information about market conditions and tends to equalize quotations across dealers. But the dealers can be in direct contact and also make deals with each other. These deals are called *interbank* transactions. A dealer can place a buy or sell order with some other dealer, in which case he operates in the role of a *customer* in relation to his counterpart, who is a *market maker*.

Transacting in the role of a customer is equivalent to being on the 'wrong side' of the market, which implies that the customer-dealer has to accept a less advantageous price than in the case where he himself would be able to make a market. For example, if the dealer decides to buy some currency in the interbank market, he has to pay the market maker's ask price, but he is not generally able to resell the currency immediately at a higher price. Hence, the motive for this kind of transaction, in general, is position adjustment, not trading income as such. Transactions the dealer undertakes in the role of a customer with a market maker are called *cover*

---

[12] While some monopoly power appears necessary for price makers, customers may learn to use other markets for currency conversion purposes if the spread in the local market is large relative to the spread elsewhere. The *customer flows* of this kind create a competitive situation between the providers of dealer services in different local markets; *cf.* Phelps and Winter (1969).

transactions. The question then is: under which conditions is such a cover transaction profitable?

In the static case, the answer is straightforward and follows from the interpretation of the Lagrange coefficient for the end-of-day position constraint as a shadow price. It depends negatively on the required net sales (positively on the required net purchases). If the amount of required net sales is large enough and the shadow price goes below the average (or the best observed) bid-price in the interbank market, then it is advantageous for the dealer to cover at least a part of his long position in the interbank market by a cover sale rather than to make a sufficiently large downward adjustment in his own price. Similarly, a cover purchase is profitable if the position is sufficiently short and the shadow price exceeds the average (or the lowest observed) ask-price in the interbank market. As a result, the price quoted by the dealer does not differ much from those quoted by other dealers on average.

In the same way as an individual dealer can place orders in the interbank market, he himself may receive orders from other dealers. This has two implications. First, the buy and sell orders received by any single dealer become increasingly price-sensitive on average, because the population of potential customers now includes other dealers who are likely to be better-informed than non-dealer commercial customers. Secondly, transactions uncertainty tends to increase, because the size of the average deal in the interbank market is generally much larger than the size of a typical customer order (*cf.* the distinction between wholesale and retail transactions). While increased price-sensitivity of buy and sell orders tends to reduce the spread, greater price uncertainty has the opposite effect. In particular, if the dealer does not want to see large fluctuations in the foreign exchange position arising from other dealers' behaviour, he can take defensive action by broadening the spread in order to reduce the interbank order flow.

Recall that the optimal pricing rule in the case of an isolated local market had a dampening effect on the intra-day fluctuations in quotations. Allowing for the interbank market restrains these fluctuations even further, because price adjustments are less dependent on local transaction shocks. Large net sales or purchases by any one dealer give rise to follow-up interbank transactions for position adjustment purposes, whereby the initial shock is diffused globally. In other words, large local transaction shocks lead to small global price reactions.

Although in the integrated market the monopoly power of an individual dealer is reduced, if not eliminated entirely, each dealer remains a price-maker and has to decide upon the price without knowing how others are quoting at the same moment. Therefore, the quotations need not be exactly the same everywhere. Hence, nothing guarantees that the quotations are always consistent in the sense that all ask-rates are higher than the highest bid-rate. In other words, arbitrage profit opportunities may arise occasionally. This does not contradict the notion of an efficient market in the information-arbitrage sense, because inconsistency only occurs occasionally, lasts only momentarily, and any act of arbitrage is based on private, not publicly available, information.

This gives each dealer a further incentive to be informed on outside quotations, which integrates the local markets even more tightly. While any one dealer may be anxious to take home any arbitrage profits which arise from inconsistent quotations, each of them wants to avoid situations where his quotations could offer riskless profit opportunities to others. A defensive attitude of this kind is likely to lead to the broadening of the spread, especially in times of great uncertainty. This situation is parallel to the presence of informed insiders in the securities market. Because of information asymmetries, a stock dealer is unable to distinguish between insiders and liquidity traders. Because the dealer loses on average in trade with insiders, he has to compensate for that loss by applying a spread which is wider than the marginal cost of providing dealer services (*cf.* Glosten and Milgrom, 1985, Kyle, 1985, Glosten, 1989, Dennert, 1989, and Hagerty, 1991).

In a dynamic case the dealer has simultaneously to decide the pricing rule and the strategy for cover transactions in the interbank market. As in the monopoly case, the shadow price is constant for given information on the day and it jumps immediately as new information arrives. For given information the quoted price remains within a given range in relation to the average quotation. The price is adjusted when the average quotation changes, although not necessarily by the full amount.

The reults can be applied to a situation where all operators know that an important piece of news will be announced later in the day. If all participants have similar expectations concerning the contents of the announcement, this effect will already be fully reflected in prices in advance. There will be no price effects when the announcement is made, as long as its contents do not differ from expectations. On the other hand, the price reaction will be prompt, if the announcement contains important surprises. If there are differences in opinion concerning the contents of the forthcoming announcement, some participants will be disappointed at the time of the announcement and observe their positions to move in an unwanted direction, which can be reversed only at a cost. A defensive strategy applicable to such situations would be to widen the spread prior to the announcement. This would reduce the risk that the dealer would receive orders from other dealers who have a different opinion and who may even have some inside information on the contents of the forthcoming news.

The fact that at any particular moment the prices quoted in different local markets tend to be equalized through interbank transactions and customer flows is equivalent to saying that the local markets are integrated spatially. The fact that different local markets are located in different time-zones implies that the operating hours of different dealers differ. Some dealers are closing business at the same time as others are opening. In so far as the business hours overlap, the markets in different time-zones become spatially integrated. This tends to draw the quotations in two time-zones close to each other in the same way as spatial integration draws quotations close to each other across the local markets at any particular moment. Any regular pattern in the exchange of positions across the time-zones during hours when these overlap is part of the information set and is fully reflected in the shadow price.

Transactions Demand for Foreign Exchange

While the previous chapters focus entirely on the pricing behaviour of an individual foreign exchange dealer, in *Chapter Four* the price-making aspect is disregarded altogether and the emphasis shifted to the costs of producing dealer services. A major cost arises from holdings of trading currencies. These are assumed to be non-interest bearing transactions balances held with banks both at home and abroad.

The problem is analogous to the analysis of the transactions demand for money under stochastic cash flow. It is to be noted, however, that the demand for money in the dealer's case is the demand for a composite of transactions balances denominated in a number of currencies, including domestic money. The formal analysis is based on the stochastic cash management model of Miller and Orr (1966). Transactions uncertainty is characterized as a symmetric Bernoulli process. If at some moment the dealer finds himself in a situation of having sold out of one of the trading currencies, he can repurchase it with the currencies overbought in the interbank market.

Taking into account the cost of such wholesale transactions the dealer's decision problem is to minimize the total costs consisting of the opportunity cost of holding a non-interest bearing trading portfolio and the expected cost of wholesale transactions. The expected trading income from customer trade is given, because it is assumed that the dealer's price-making power is non-existent.

Not surprisingly, the results resemble those of Miller and Orr. The size of the trading portfolio is determined by a *cubic-root rule* and depends positively on the average volume of customer orders and the cost of a wholesale transaction, and negatively on the rate of interest. In addition, after each wholesale transaction the composition of trading balances is restored to the state that maximizes the expected time duration between two wholesale transactions.

The cubic-root rule implies, among other things, that there are likely to be significant economies of scale in the clearing of international payments by foreign exchange dealers.[13] This is reflected in the spread, which tends to be narrower for the currencies for which the market is deep, as compared to currencies for which the volume of trade is small or irregular (Swoboda, 1969).

The economies-of-scale result is, in principle, in conflict with the assumption of the competitive market structure. While nothing conclusive can be said about the possible tendency toward a natural monopoly in terms of the very short-term partial equilibrium model of *Chapter Four*, a number of factors can be mentioned which are likely to work against it. Extending the time-horizon would bring other cost elements into the picture, for which

---

[13] The economies of scale associated with commercial banks' foreign exchange holdings was the subject of several empirical studies in the 1960s and 1970s (e.g. Heller, 1969, and Officer, 1976). The problem with these studies, however, is the lack of data on the working balances of commercial banks (nostro accounts). As noted by McKinnon (1978, p. 12), since the mid-1960s the recorded data on claims on foreign banks of commercial banks has been overshadowed by the huge rise in Eurocurrency deposits.

reason the marginal cost curve may well be increasing after some point. Moreover, the inherent risk associated with dealing activity is likely to reduce the scope for economies of scale beyond some limit.

As long as the interest parity holds, it does not matter which interest rate, domestic or foreign, is used to measure the opportunity cost of holding trading balances. From the trading point of view, the currencies are *complements*, not substitutes. Currency substitution enters the picture only if the dealer is ready to speculate on short-term exchange rate movements. This can be discussed in terms of the transactions demand model, but it is somewhat outside the main line of the argument, because speculative positions can be taken simply by making a wholesale transaction without waiting until the successive customer deals force the dealer into the interbank market.

Consistent Exchange Rates and the Vehicle Currency

In *Chapter Five* the analysis is extended to an N-currency world. The starting point is Chacholiades's (1971) study on the determination of the necessary and sufficient conditions for the establishment of consistent spot exchange rates across financial centres. The purpose is to demonstrate the extent and nature of the information problem required for the consistency of exchange rates. Consistency is defined as the absence of profitable arbitrage opportunities. In the case of two currencies the exchange rates are consistent when no ask-rate is lower than the highest bid-rate. In the case of N currencies the situation is much more complicated. Arbitrage profit may be reaped by successive purchases and sales of a number of currencies even though the exchange rates are consistent for any single pair of currencies. Indeed, it turns out that the absence of arbitrage profit opportunities in any subset of potential potential arbitrage chains does not guarantee that all arbitrage opportunities are absent.

The fact that each dealer quotes prices without knowing how other dealers are quoting at the same moment implies that arbitrage profit opportunities may always arise.[14] Observing a chain of inconsistent

[14] If markets were always perfectly arbitraged, there would never be any profits to be made from this activity and there would be no incentives for dealers to be informed of each other's quotations. In this respect, less-than-perfect arbitrage is essential for the efficient working of the foreign exchange market; *cf.* Grossman and Stiglitz (1976, pp. 247-248).

quotations by pure chance or by systematic detection is not sufficient. The arbitrageur must also be able to react instantaneously to such an opportunity. The act of arbitrage itself helps eliminate the inconsistency in quotations, because the dealer quoting low finds his position becoming excessively short, while the dealer quoting high finds his position becoming excessively long, leading to a corrective price adjustment.

To check for all possible arbitrage profit opportunities requires, in principle, the acquisition and processing of an enormous amount of information. This would especially be the case if all currencies were quoted bilaterally against each other by all market makers. Although the computational problem associated with the detection of profitable arbitrage opportunities may be overcome, the problem of acquiring the necessary price data remains.

The magnitude of the information problem is greatly reduced when there is one central currency against which all other currencies are quoted. Since the spreads in bilateral quotations differ, depending on the average volume of trade, this unit of account, the *numeraire currency*, cannot be chosen arbitrarily. Since, in the competitive market, the spread reflects the marginal cost of producing dealer services and at the same time is the price of these services, a narrow spread indicates low transaction costs. The currency that can serve as a numeraire must be the one for which the market is the deepest and for which the transaction costs are the lowest. In this case the proportionate spread in the direct quotation for a given currency (against the numeraire) is smaller than the proportionate spread in quotations between this currency and any other non-numeraire currency.

When such a numeraire can be found and when all cross rates are based on the quotations against it, the amount of information required to observe inconsistent quotations is relatively small and manageable. This is because the market participants only need to compare quotations in different centres against the numeraire. Without it, short arbitrage chains could not guarantee that arbitrage opportunities through longer chains (successive sales and purchases of more than, say, three currencies with more than three market makers) do not exist. With the numeraire, an arbitrage opportunity, if it exists, can be observed as a pairwise inconsistency in direct quotations against the numeraire by two market makers.

The solution to the information problem in the manner sketched above gives rise to a *vehicle currency* and hence to a truly international medium of exchange, although the use of such a medium is confined only to currency trade between dealers. The reason is that usually it is as cheap to buy a third currency by first buying the numeraire currency and using the proceeds to buy the currency required. In addition, the numeraire currency is used in arbitrage transactions, when such opportunities arise. Frequent indirect purchases via the vehicle currency increase the volume of trade in this currency, which tends to reinforce its role as the vehicle currency.

Chapter Two

PRICE ADJUSTMENT AND THE POSITION CONSTRAINT

This essay focusses on the pricing and position-adjustment behaviour of an individual foreign exchange dealer. The dealer is assumed to operate in an isolated market and hence to have a short-term monopoly in relation to his customers. This means that he is able to influence the average flows of buy and sell orders of customers by changing prices.

Although monopoly conditions are far from those characterizing the foreign exchange market in reality, these rather extreme assumptions provide a useful starting point for the analysis of the more complicated and more realistic situations dealt with in *Chapter Three*. The aim of this chapter is to provide a tractable analysis of one of the most important aspects of the dealer's behaviour, his pricing decisions, using as simple techniques as possible.

The dealer is assumed to maximize the expected trading income subject to stochastic price-dependent customer orders and to the constraint that the expected foreign exchange position at the end of the trading day is closed or meets some other fixed target. The formal analysis is carried out by solving a simple dynamic programming problem in which the quotation is used as a control variable and the customers' buy and sell orders determine the evolution of the position and hence form the system constraint of the problem.

It will be shown that the bid-ask spread is independent of the state of the system, *i.e.,* the foreign exchange position, and it is always chosen to maximize the return on the one-period equilibrium volume of trade, whereas the mid-rate is adjusted to control the evolution of the expected position. Whatever the position may be at any particular moment as a result of transactions uncertainty, the dealer will always quote in such a way that an open position is expected to disappear gradually during the remainder of the day. This means that he is willing to accept an open position in the course of the day and is reluctant to make large changes in quotations, even if the position moves off the desired path. The reason for the reluctance to accept large price adjustments is that they are generally revenue-reducing, *e.g.* selling at a low price the currency which a few moments ago was bought in at a high price.

The results repeat many features which are familiar from the finance literature on dealer behaviour in organized securities markets.[1] Indeed, the basic formulation of the problem is borrowed from one such analysis, *viz.* that of Zabel (1981). His results implied a state-dependent price-adjustment rule, whereas the spread was independent of the position. Ho and Stoll (1981) analyzed the optimal pricing behaviour of a specialist at the New York Stock Exchange. In addition to transactions uncertainty, they took into account return uncertainty and used a Poisson jump process to describe the arrivals of buy and sell orders. The results were similar to those of Zabel as regards the nature of the price adjustment and also in the sense that in his study the spread is independent of the dealer's inventory of securities as well as of the degree of transactions uncertainty. On the other hand, the degree of return uncertainty affected the spread positively. Amihud and Mendelson (1980) showed a similar price-adjustment rule, but in their study the spread depended positively on the deviation of the security inventory from its preferred level, a result that followed from the assumption of the presence of predetermined limits to the permissible inventory levels.

Similarities between the results of the present study and those of the finance literature are to be expected, because the dealership market has the same characteristics irrespective of the particular empirical reference. While the insights and techniques have been borrowed from the microstructure literature, the following analysis is carried out with much fewer technical and institutional assumptions. Return uncertainty, limit orders, market participants' perceptions of the *true* price of an asset and asymmetric information, which play an important role in the analysis of stock dealer behaviour, are less important in the analysis of the spot market for foreign exchange and they are not formally incorporated in

[1] The behaviour of stock dealers at the New York Stock Exchange has been the most frequently cited empirical reference in these studies. The assumption of a monopoly dealer is, in fact, realistic in their case, because they specialize in trading in particular shares; see Demsetz (1968), Ho and Stoll (1981), and Zabel (1981).

the present study. It is to be noted, however, that price uncertainty enters the picture through the end-of-day position constraint, which takes care of the fact that the dealer's position does not move very far from the desired and relatively safe path. Price uncertainty becomes more important in *Chapter Three*, where inter-dealer competition will be introduced.

The most significant and, perhaps, the least innocent abstraction in the analysis is the assumption that the dealer is not liquidity-constrained. Although, it is acknowledged that he must have inventories of currencies in order to be able to buy or sell on immediate demand, these are not explicitly taken into account. The focus is thus entirely on pricing behaviour. Inventory management will be the subject of *Chapter Four*.

1 One-Period Model

In order to introduce the notation and to present the structure of the problem, the dealer's pricing behaviour is first analyzed in a simple one-period case. It will be seen that the analysis is, in fact, a fairly straightforward application of standard microeconomic theory.

In accordance with the discussion of the dealership market in *Chapter One*, the price-dependent arrival rates of customer orders are written as follows (*cf. Figure 1.1.*):

$$p(t) = a(t) - cs^a(t) \tag{2.1}$$

$$q(t) = b(t) + cs^b(t). \tag{2.2}$$

The first equation expresses the customer demand (arriving sell orders) as a function of the ask-rate $s^a(t)$. The second equation shows the customer supply (arriving buy orders) as a function of the bid-rate $s^b(t)$ quoted by the dealer at the beginning of a short period t. The length of the period will be discussed below. The ask- and bid-rates are defined as the price of one unit of foreign currency (*dollar*) in terms of domestic money (*marks*). The amounts bought, $q(t)$, and sold, $p(t)$, by the dealer have the dimension of dollars per discrete unit of time. This time span will in the following be called the *trading session*.

The shift variables $a(t)$ and $b(t)$ are assumed to be known for the time being. Assuming that $a(t) - b(t)>0$ guarantees that in equilibrium both the price and the volume of trade are positive. Note that the demand is high when $a(t)$ is high. The supply is high when $b(t)$ is high. The parameter $c>0$ stands for customer sensitivity to exchange rates. Nothing is lost in generality if it is assumed to be equal on both the buy and the sell side.[2]

It is convenient to redefine the ask-rate and the bid-rate in terms of the *mid-rate*, $s(t)=[s^a(t)+s^b(t)]/2$, and the *half-spread*, $z(t)=[s^a(t) - s^b(t)]/2$.[3] The trading income per session, defined in terms of domestic money, is equal to the value of sales minus the value of purchases, or

$$(2.3) \qquad R(t) = [\, s(t)+z(t)]p(t) - [\, s(t) - z(t)]q(t)$$

$$= z(t)[\, p(t)+q(t)] + s(t)[\, p(t) - q(t)].$$

From the latter expression it is seen that trading income is equal to the spread times the volume of trade[4], adjusted for any cash inflow or outflow arising from net sales or purchases of the foreign currency. Redefining the parameters, the trading income can be expressed in the following form:

$$(2.4) \qquad R(t) = \alpha(t)s(t)+\beta(t)z(t) - \delta[\, s^2(t)+z^2(t)],$$

where $\alpha(t)=a(t) - b(t)>0$, $\beta(t)=a(t)+b(t)>0$, and $\delta=2c>0$.

Maximizing the expected revenue in terms of domestic currency without any constraints would lead to substantial net sales of foreign currency. This is not possible over successive trading sessions, assuming that the dealer intends to stay in the market, because the source of his income is two-way trade, buying foreign currency from the customers and selling it back at a higher price. This observation gives an important

---

2 *Cf.* Suvanto (1982), where the price sensitivities differ on each side of the market.

3 The half-spread is a measure of the transaction cost. A simultaneous purchase and sale of one dollar involves *two* transactions and costs the amount of the full spread.

4 See *Figure 1.1*. Note that the spread is $2z(t)$ and the average volume of trade is $[p(t)+q(t)]/2$.

reason why the dealer has to be concerned about net sales; *i.e.*, changes in his foreign exchange position. The change in the position, in turn, is equal to net sales during the trading session:

$$(2.5) \qquad x(t) - x(t+1) = p(t) - q(t) = \alpha(t) - \delta s(t),$$

where $x(t)$ is the foreign exchange inventory at the beginning and $x(t+1)$ at the end of the trading session. Note that net sales are measured in terms of flows of foreign currency over the trading session. Hence the foreign exchange position is a stock variable representing the dealer's inventory of dollars at a given moment.

Define the equilibrium quotation [$s^o(t), z^o(t)$] as that combination of $s(t)$ and $z(t)$ that will maximize the trading income subject to the constraint that expected net sales are equal to zero. It is obtained by solving the following problem:

$$(2.6) \qquad \max_{s(t),\, z(t)} \{\alpha(t)s(t) + \beta(t)z(t) - \delta[\, s^2(t) + z^2(t)]\}$$

$$\text{s.t. } \alpha(t) - \delta s(t) = 0,$$

which gives

$$(2.7) \qquad s^o(t) = \alpha(t)/\delta, \quad z^o(t) = \beta(t)/2\delta.$$

It is seen that in periods of high demand (high $a(t)$, therefore both $\alpha(t)$ and $\beta(t)$ are high) the equilibrium mid-rate is high and the spread is wide. Similarly, in periods of high supply (high $b(t)$, therefore $\alpha(t)$ is low and $\beta(t)$ is high) the equilibrium mid-rate is low but again the equilibrium spread is wide. This result brings out the distinction between the demand for and the supply of foreign exchange by customers, on the one hand, and the demand for dealer services, on the other. Whereas the mid-rate is adjusted to balance customer orders, the spread is the price of the dealer's liquidity services. It increases with the volume of trade (and hence with $\beta(t)$), reflecting the dealer's monopoly power in the market.

At quotation [$s^o(t), z^o(t)$] the trading income becomes

(2.8) $$R^o(t) = \beta^2(t)/4\delta,$$

which is equal to the equilibrium spread $2z^o(t)$ times the equilibrium volume of trade; *cf.* equation (2.3). The latter is obtained by dividing the sum of expected sales and purchases at quotation $[s^o(t), z^o(t)]$ by two and it is equal to $\beta(t)/4$.

Assuming next that the dealer has a given inventory $x(t)$ of dollars and that he wants to go to some target level $x(t+1)=x^*$ during the trading session implies that the dealer has to announce a quotation that differs from the equilibrium one in order to attract net sales by the amount $x(t) - x^*$. Maximizing $R(t)$ with respect to $s(t)$ and $z(t)$ and subject to the constraint that $x(t)-x^*-\alpha+\delta s(t)=0$ leads to the following state-dependent quotation:

(2.9) $$s(t) = s^o(t) - (1/\delta)[\, x(t) - x^*], \quad z(t) = z^o(t).$$

Only the mid-rate is state-dependent. It is below the equilibrium rate, if the current position is above the target, and above it, if the current position falls short of the target. The spread, in turn, is independent of both the initial and the target position and it is always chosen to maximize the return on the one-period equilibrium volume of trade. This result will appear throughout the subsequent analysis.

The *Lagrange coefficient* associated with the net sales constraint is

(2.10) $$\mu(t) = s(t) - (2/\delta)\ [\, x(t) - x^*].$$

It is one of the key variables and is interpreted as the shadow price of one unit of foreign currency, which is the price the dealer would be willing to pay or offer for one unit of foreign currency in order to reduce the position constraint marginally. This can be seen by inserting the quotation (2.9) into the revenue function (2.4) and differentiating it with respect to $[x(t)-x^*]$.

Assume next that buy and sell orders are stochastic. The interpretation is that, as a market maker, the dealer sets prices in advance and the customers then decide whether to buy or sell at these prices and in what amounts during a short trading session. A simple way to introduce stochasticity is to assume that the shift variables $a(t)$ and $b(t)$ are random variables with a constant mean: $a(t)=a+u(t)$ and $b(t)=b+v(t)$, where $u(t)$ and $v(t)$ are independent and have zero expectation.

As a result, net sales are now stochastic:

$$x(t) - x(t+1) = \alpha - \delta s(t) - w(t), \tag{2.11}$$

where $\alpha=a-b>0$, and $w(t)=v(t)-u(t)$. The expected trading income remains unchanged, except that $\alpha(t)$ and $\beta(t)$ in equation (2.4) are replaced by α and β, respectively, $\beta=a+b>0$. Maximizing the expected trading income subject to the constraint that expected net sales are zero; *i.e.*, $x(t)-\mathbf{E}_t\{x(t+1)\}=0$, gives the following *equilibrium* (zero net sales) quotation:

$$s^o = \alpha/\delta, \quad z^o = \beta/2\delta. \tag{2.12}$$

This quotation is used as a reference price in much of the subsequent discussion. The corresponding expected trading income is denoted by R^o.

Several extensions are possible even within the framework of this simple one-period model. For instance, one can incorporate the dealer's transaction costs into the analysis. Assuming a constant cost per dollar bought or sold will widen the spread, although not by the full amount. It does not, however, affect the mid-rate rule. The extra cost is thus shared by the dealer and his customers. On the other hand, assuming that customers bear a constant transaction cost per dollar transacted will lead to a narrower spread. Again the mid-rate rule remains unchanged and the extra cost is shared by both counterparts.[5] Because these extensions do not affect the structure of the problem, they are ignored below.

---

[5] These examples are discussed in Suvanto (1982).

2 *Sequential Pricing Decisions under Transactions Uncertainty*

The target level x^* for the dealer's dollar holdings at the end of the trading period can be interpreted as the closed position. The interpretation is formally correct if it is assumed that the dealer has borrowed an amount x^* of dollars, which appears on the liability side of his balance sheet. In order to facilitate settlements, the same amount has been placed as liquid demand deposits with foreign commercial banks (*nostro accounts*) and therefore appears on the asset side of the balance sheet. These trading balances, in turn, are constantly fluctuating.[6] The amount of borrowing and the average trading balances are fixed over a longer period. Therefore, the net interest costs are fixed and do not affect the dealer's short-run pricing decisions.

Although the position may be closed at the beginning and at the end of the trading day, the intraday position, denoted by $x(t)-x^*$, fluctuates throughout the day as a result of deals done during successive dealing sessions. In this interpretation, $x(t)-x^*>0$ is a *long position*; *i.e.*, the dealer has overbought dollars and the dollar-denominated assets exceed liabilities. Similarly, $x(t)-x^*<0$ is a *short position*; *i.e.*, the dealer has oversold dollars and the dollar-denominated assets fall short of liabilities. A closed position target can be justified by risk considerations.

With an open position the dealer may lose or gain if something happens that will change the average arrival rates of customers' buy and sell orders in such a way that it would call for a new equilibrium mid-rate. Therefore, the assumption of a closed position target involves a substantial degree of risk aversion. The formal analysis does not, however, depend on how the target position x^* is interpreted, as long as it is fixed.

In the following the stochastic version of the one-period model is extended to a situation where the position is allowed to fluctuate, but the quotation can be changed from session to session in order to control the expected position. The closed position constraint does not apply to each

---

[6] As a mirror image, the dealer's trading balances in domestic currency are fluctuating as well, although in the opposite direction.

trading session separately, but only to the last session of a sequence of sessions, which together make a *trading day*. Following Zabel (1981), a trading day is divided into T trading sessions, which are of equal length.[7] The length of the day and the number of sessions within the day are exogenous.

Assume that the dealer opens the day with a given position $x(0) - x^*$ and he wants to be in a closed position at the end of the day. At the beginning of each trading sesssion $(t, t+1)$, the dealer is free to announce a new quotation at which he stands ready to trade on the basis of incoming buy and sell orders during that short period. His objective is to maximize the expected trading income during the day subject to the system constraint as well as to the constraint that the expected position at the end of the day is closed. The system constraint, or the state equation, describes how the position evolves in discrete time.

This formulation of the problem leads to a straightforward application of dynamic programming techniques (cf. Bertsekas, 1981, *Ch. 2*). The dealer's problem is to find a feedback controller of the form $\{ f_i [x(i)] \; ; \; i = 0,1, \ldots, T\text{-}1\}$, which states that if at moment t the state is $x(t)$, then the control $f_t [x(t)]$ should be applied. The control $f_t : \mathbb{R} \mapsto \mathbb{R}^2$ is a mapping from the state space (the position) into the control space (the quotation) such that $f_t [x(t)] = [s(t), z(t)]$. The state equation is of the form $x(t+1) = x(t) - [\, p(t) - q(t) - w(t)]$.

Assuming a finite horizon and perfect information on the current state, the dynamic programming algorithm can be written as follows:

(2.13) $$J_t [x(t)] = \underset{s(t),\, z(t)}{\mathbf{Max}}\ \mathbf{E}_t \{R(t) + J_{t+1} [x(t+1)]\} \qquad \text{(value function)}$$

(2.14) $$J_T [x(T)] = 0 \qquad \text{(terminal value)}$$

(2.15) $$x(t+1) = x(t) - \alpha + \delta s(t) + w(t) \qquad \text{(system constraint)}$$

(2.16) $$\mathbf{E}_t\, x(T) = x^* \qquad \text{(terminal state constraint),}$$

---

[7] The assumption that the trading periods are of equal length is not essential in itself. What matters for the subsequent analysis is that the number of trading periods per day is given and that the degree of transactions uncertainty is the same for each period.

where $t = 0, 1, 2, ..., T-1$, and $\mathbf{E}_t$ stands for the expectation at moment t. The one-period revenue function $R(t)$ in equation (2.13), written with stochastic disturbances, is

$$(2.17) \qquad R(t) = \alpha s(t)+\beta z(t) - \delta[s^2(t)+z^2(t)]+z(t)[u(t)+v(t)] + s(t)[u(t) - v(t)].$$

The value function $J_t [x(t)]$ gives the expected revenue from moment t until the end of the day when optimal control is applied at each future moment $t+1, t+2, ..., T-1$. No cost is attached to the possible open position at the end of the day, which makes the terminal value $J_T [(x(T)]$ equal to zero. This assumption can be dropped without affecting the results if the horizon is extended indefinitely over many trading days, provided that the requirement of the closed expected position at the end of each day is maintained (see *Appendix*).

Following the standard dynamic programming techniques, the sequential pricing rule is derived by solving first the optimal quotation for the last period $(T-1, T)$ and proceeding recursively backward in time. The solution is straightforward, although strenuous, and it is presented in the *Appendix*. The resulting feedback controller, or the sequence of functions expressing the optimal quotation for each period $(t, t+1)$ as a function of the current state, has the form:

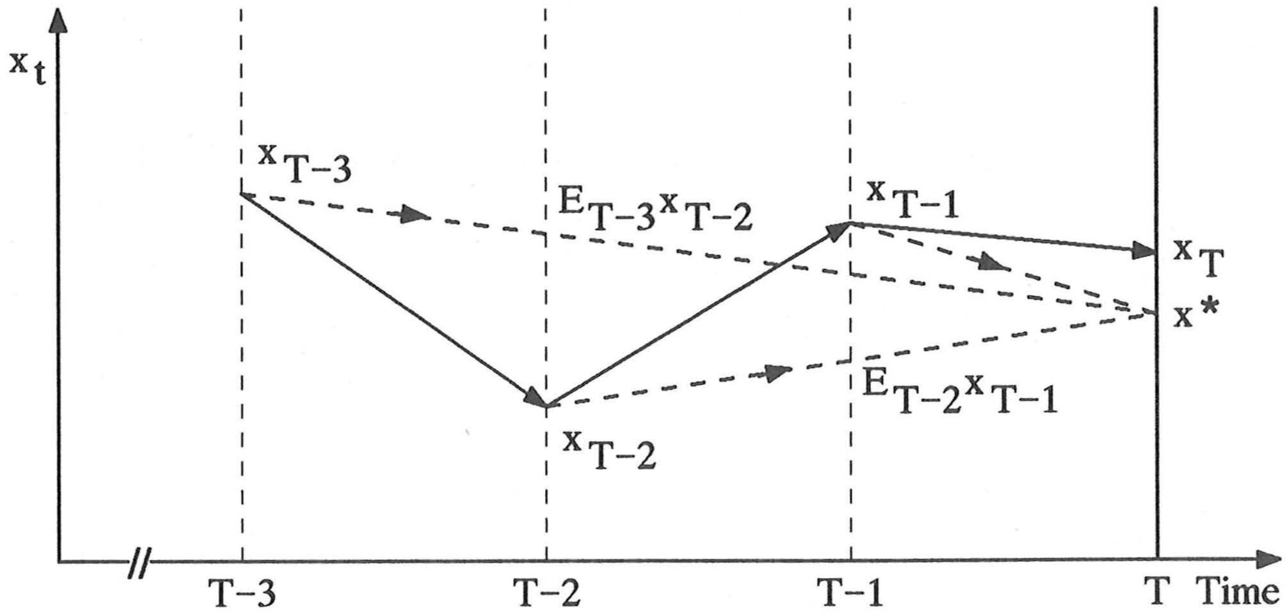
$$(2.18) \qquad s(t) = s^o - (1/\delta) [x(t) - x^*] /(T-t), \quad z(t) = z^o,$$

where s^o and z^o are equal to the one-period equilibrium quotation derived above; *cf.* equation (2.12). As in the one-period case, only the mid-rate is adjusted to steer the position. The spread remains constant at the level that maximizes the one-period return on the equilibrium volume of trade.

Inserting the quotation (2.18) into the net sales equation (2.11) and noting that $\alpha - \delta s^o = 0$ gives the following change (increase) in the foreign exchange position in session $(t, t+1)$:

$$(2.19) \qquad x(t+1) - x(t) = - [x(t) - x^*] /(T-t) + w(t).$$

Figure 2.1 Expected and Actual Position



Note: x^* is the end-of-day target position, x_t is the realized position and E_t is an expectation (a plan) at moment t.

The first term on the right-hand side expresses the planned, or expected, change in the position. Hence, according to the optimal pricing rule (2.18), the dealer quotes in such a way that he can expect to close a fraction $1/(T\text{-}t)$ of the eventual open position both during the current session and in each of the remaining sessions of the trading day. This result is illustrated in *Figure 2.1*.

Constant planned net sales for each of the remaining trading sessions imply a constant planned quotation. The actual quotation is adjusted in proportion to the latest realized transactions shock:

$$s(t+1) - s(t) = -(1/\delta)\, w(t)/(T\text{-}t\text{-}1). \tag{2.20}$$

Unexpected net purchases, $w(t)>0$, lead to a downward adjustment of the quotation, and *vice versa.* Note that the coefficient of proportionality increases as t approaches $T\text{-}1$ (*cf. Section 4* below).

The fact that the planned or expected price is constant for the remainder of the day corresponds to the familiar market efficiency in competitive models (*cf.* Amihud and Mendelson, 1980, p. 32). The fact that intra-day quotations are martingales implies, *inter alia*, that customers have no reason to postpone a transaction in anticipation of getting better bids or offers at some later moment.[8]

3 New Information and Jumps in Quotations

It is to be noted that the constancy of the planned quotation and the constancy of the spread result from the assumption that transactions uncertainty is characterized by serially uncorrelated disturbances of the customers' buy and sell orders. There are no jumps in the shift variables α and β describing the average flows of customer orders. This is equivalent to saying that no new information hits the market that changes the dealer's perceptions about the average arrival rates of future customer orders. If that happens in the middle of the day, the equilibrium mid-rate, and most likely also the spread, will react immediately. For instance, if the demand curve jumps up, the equilibrium mid-rate also jumps up (because α goes up) and the spread widens (because β goes up).

It is instructive to make a clear distinction between stochastic transaction disturbances and jumps in the model parameters. For that purpose let us forget, for a while, the stochasticity of customer orders and assume that the future flows of customer orders is known to the dealer, but not to customers. The sequential problem can now be reformulated in the following form:

$$(2.21) \qquad \underset{s(i),\, z(i)}{\mathbf{Max}} \sum_{i=0}^{T-1} \{\alpha(i)s(i) + \beta(i)z(i) - \delta[\, s^2(i) + z^2(i)]\}$$

[8] The empirical evidence on the behaviour of exchange rates within a trading day is scarce. A study by Wasserfallen (1989) using intra-day data (drawn at five-minute intervals) on one single bank's quotations seems to confirm the weak-form efficiency; *i.e.,* the absence of serial correlation inside a a day. Other studies with hourly data on indicative market quotations (keyed by different banks into electronic informations systems) show significant negative first-order autocorrelation around the occasion of large jumps in prices; see Ito and Roley (1986), and Goodhart and Giugale (1988). Negative autocorrelation appears to be rather common in indicative market quotations at ultra-high frequencies, but disappears after ten-minute intervals; see Goodhart and Fiuglioli (1991). Finally, Müller *et al.* (1991) analyze continuous 'tick-by-tick' data over a three-year period (taken from Reuters' page FXFX) which shows significant intra-day heteroscedasticity, which has a regular 'seasonal' pattern.

$$\text{s.t.} \quad \sum_{i=t}^{T-1} [\alpha(i) - \delta\, s(i)] = x(t) - x^*,$$

where i is now the index of time. In other words, the dealer sets the sequence of prices in such a way that the daily trading income is maximized subject to the constraint that, at each moment t, the position will be closed during the remainder of the day. The problem is essentially the same as above. The system constraint and the terminal state constraint are combined into one expression. The daily revenue is expressed explicitly as the sum of one-period revenues, because all the relevant information on the trading day is incorporated in the parameters $\alpha(i)$ and $\beta(i)$.

The first-order conditions for the optimal quotations are solved from

$$\alpha(i) - 2\delta\, s(i) + \delta\mu(t) = 0, \qquad \beta(i) - 2\delta\, z(i) = 0, \tag{2.22}$$

where $\mu(t)$ is the Lagrange coefficient associated with the position constraint at moment t. The spread depends only on each period's volume of trade $\beta(i)$, $z(i) = \beta(i)/2\delta$, whereas the mid-rate depends on each period's excess demand $\alpha(i)$, as well as on the position constraint. The mid-rate is obtained from the first equation as a function of the Lagrange coefficient:

$$s(i) = [\alpha(i)\,/\delta + \mu(t)]\,/2. \tag{2.23}$$

Inserting this into the position constraint, adding up over the remaining trading sessions and solving for $\mu(t)$ gives

$$\mu(t) = [\delta(T-t)]^{-1}\, \{ \sum_{i=t}^{T-1} \alpha(i) - 2\, [\, x(t) - x^*]\}. \tag{2.24}$$

It is obvious from (2.24) that the shadow price $\mu(t)$ contains all the relevant information on the remainder of the day. Through equation (2.23) it is transmitted to quoted prices. If there is no new information, $\mu(t)$ will remain constant for the remainder of the day. That $\mu(t+1) = \mu(t)$

can be shown by shifting (2.24) forward and inserting it into the net sales in period $(t, t+1)$; *i.e.*, $x(t) - x(t+1) = \alpha(t) - \delta s(t)$.

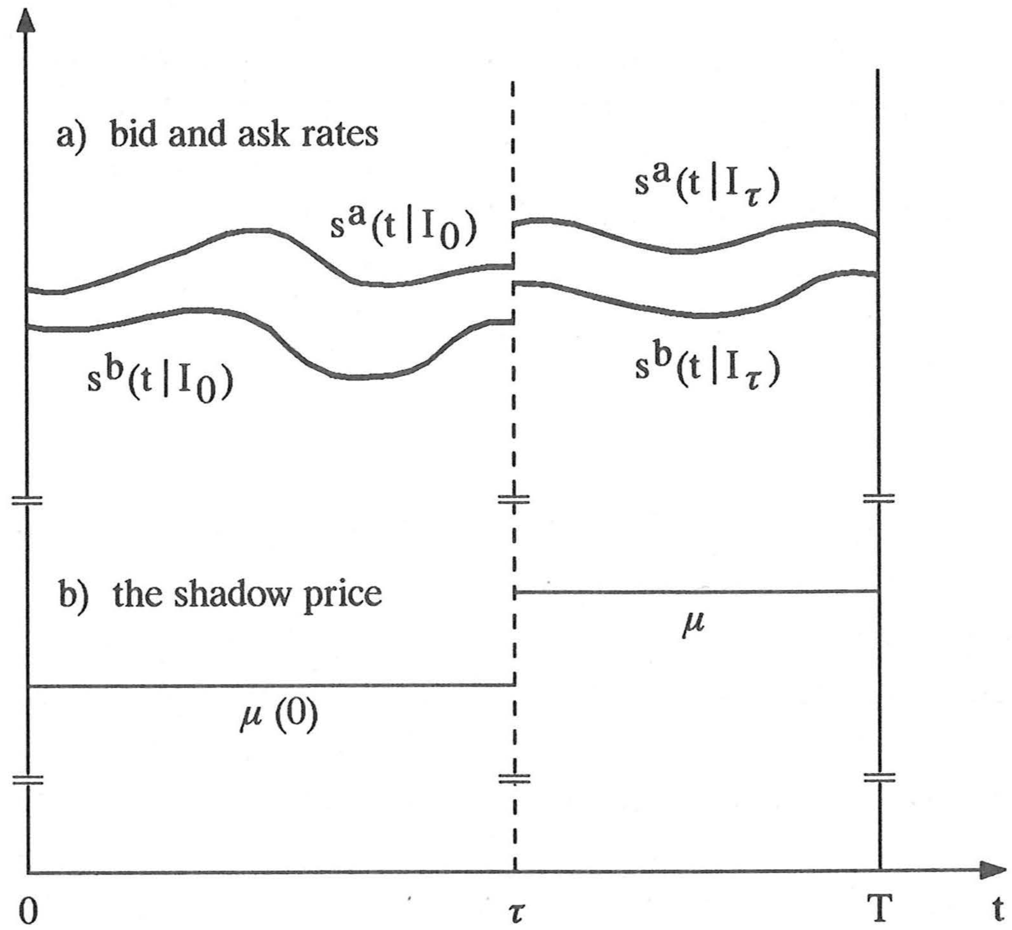
What is interesting in this result is that the mid-rate and the spread do change, although the shadow price remains constant. In periods of high customer demand, the mid-rate is high and also the spread is wide, implying that the ask-price rises more than the bid-price. Correspondingly, in periods of high customer supply the mid-rate is low, but the spread again is wide, implying that the bid-rate declines more than the ask-rate.

It is also obvious from equation (2.24) that $\mu(t)$ reacts to new information. Assume that at moment t the dealer learns that at some later time $i > t$ during the remainder of the day, there will be high demand. From eq. (2.24) it is seen that $\mu(t)$ will jump up immediately, as too will the mid-rate according to equation (2.23). The jump does not depend on whether high demand will materialize very soon or at some later period. An equal change in any of the remaining $\alpha(i)$'s affects the shadow price by an equal amount. The spread, however, will rise only when the increase in demand actually takes place.

As the mid-rate rises and remains at a higher level throughout the remainder of the day, the increase in the volume of trade will be less than what is implied by the increase in $\alpha(i)$'s, because a higher price reduces other customers' demand.

Figure 2.2 illustrates the implications of the model for the behaviour of the exchange rate during the day. The curves $s^a(t \mid \mathbf{I}_0)$ and $s^b(t \mid \mathbf{I}_0)$ illustrate the expected time paths of the ask- and bid-rates conditional on information $\mathbf{I}_0$, available at time $t=0$. Neither the mid-rate nor the spread need be constant. They depend on the dealer's information on the timing of customer orders within the day. At time τ, new information hits the market: both the ask- and bid-prices jump and thereafter move along the new time paths $s^a(t \mid \mathbf{I}_\tau)$ and $s^b(t \mid \mathbf{I}_\tau)$. This example highlights the distinction between anticipated changes and unanticipated jumps. The distinction becomes even clearer in the lower panel of *Figure 2.2*, in which the time path of the shadow price of foreign exchange associated with the position target is depicted. At time $t=0$, the shadow price is equal to $\mu(0)$. If there is no new information, it will remain constant. At time τ, however, new information hits the market, $\mu(0)$ jumps to $\mu(\tau)$ and remains in the new position unless any further news

Figure 2.2   New Information and Jumps in Quotations



arrives during the remainder of the day.

In the above model it is important, however, that the information of the sequence of $\alpha(i)$'s and $\beta(i)$'s is the dealer's private information. If the customers had the same information about the trading day, they would choose a different timing for their purchases and sales. As a result, the values of $\alpha(i)$ and $\beta(i)$ would alter until all the predictable changes in the quotations would disappear. If this happens then $\alpha(i)=\alpha$ and $\beta(i)=\beta$ for each i. The latter situation would lead us back to the stochastic version of the model discussed above, implying a constant spread and a constant planned quotation.

The distinction between the stochasticity of the arrivals of customer orders and the unexpected news is also important for understanding some real world phenomena. As will be shown below, one result of the optimal pricing strategy is the dampening of short-term price fluctuations. On the other hand, the dealer's reaction to new information is a prompt change in prices that reduces the amount of currency flows, which could potentially take place if the customers were better informed than the dealer. As noted by Begg (1989, pp. 28-29), by avoiding taking positions himself and by immediately adjusting prices to the level that chokes off most of the potential currency flows, the dealer restores the business to smaller-sized and more balanced trade in both directions and returns to servicing the currency conversion needs of liquidity-oriented commercial customers.

4 *Price and Position Variability*

Let us forget the jumps in prices arising from the arrival of new information and return to the sequential pricing model under transactions uncertainty. The results of this model bring out two important microeconomic principles which are present in the dealership market. First, the price adjustment follows the *Walrasian tatonnement* process in the sense that realized excess demand and excess supply call for a change in prices in the equilibriating direction. In other respects the market is non-Walrasian, because no market clearing is required for each market period. Secondly, the gradual adjustment towards a given end-of-day position target implies that the fluctuations in quotations become smaller than they would be if the dealer always attempted to keep the position closed, or if there were an auction mechanism that would continuously equilibriate the sales and purchases of customers. This result follows directly from the dealer's willingness to accept open positions within the day. An open position is analogous to inventory holdings. The price-smoothening effect of the optimal pricing rule is hence one reflection of the *buffer stock* property of inventories.

The price-smoothening effect of the optimal pricing rule can be seen more clearly by writing the two random variables, the dealer's position $x(t)-x^*$ and the mid-rate $s(t)$, in terms of stochastic net purchases $w(t)=v(t)-u(t)$ and by examining their variances. The variances are then compared with the price and position uncertainty under an alternative pricing rule according to which the dealer would always quote in such a way that the expected position is closed at the end of each trading session.

Starting from the beginning-of-day position $x(0)$-x^*, equation (2.19) can be used to generate

$$(2.25) \qquad x(t) - x^* = [\, x(0) - x^*]\,(T\text{-}t)/T + (T\text{-}t) \sum_{i=0}^{t-1} [\, w(i)\,/(T\text{-}1\text{-}i)],$$

$$t=1,\ 2,\ \ldots,\ T\text{-}1,$$

$$x(T) - x^* = w(T\text{-}1).$$

The last term in the summation multiplied by T-t is always equal to the latest disturbance. With t=T, the optimal pricing rule nets out the effects of all earlier shocks; only the latest transaction shock affects the end-of-day position. If the initial position is inherited from the previous day's last transaction shock, it is also a random variable and has the same variance as any of the one-period disturbances σ_w^2. Because the successive disturbances are assumed to be independent, the *ex ante* variance of the position for an arbitrarily chosen t becomes

$$(2.26) \quad \mathbf{Var}\,\{\, x(t) - x^*\} = \sigma_w^2\,(T\text{-}t)^2 \sum_{i=0}^{t} 1/(T\text{-}i)^2\,, \quad t = 0,\ 1,\ \ldots,\ T\text{-}1,$$

$$\mathbf{Var}\,\{\, x(0) - x^*\} = \mathbf{Var}\,\{\, x(T) - x^*\} = \sigma_w^2.$$

With t=0, the variance is σ_w^2, after which it first increases monotonically until it starts to diminish around the middle of the day (*Table 2.1*).

The mid-rate expressed in terms of random disturbances is obtained by inserting the right-hand side of (2.25) into the optimal quotation (2.18), which gives

$$(2.27) \qquad s(t) = s^o - [\, x(0) - x^*]\,/\delta T - (1/\delta) \sum_{i=0}^{t-1} [\, w(i)\,/(T\text{-}1\text{-}i)],$$

$$t=1,\ 2,\ \ldots,\ T\text{-}1.$$

The deviation of the mid-rate from the equilibrium rate is $s(t) - s^o$. Its *ex ante* variance depends on the one-period transactions uncertainty σ_w^2 in the following fashion:

$$\text{(2.28)} \qquad \mathbf{Var}\ \{\ s(t)\ -\ so\} = (\sigma_w/\delta)^2 \sum_{i=0}^{t} 1/(T\text{-}i)^2, \qquad t=0,\ 1,\ \ldots,\ T\text{-}1.$$

The variance increases monotonically with the period of the trading day.

Under the alternative rule the position would depend only on the latest transaction disturbance $w(t\text{-}1)$. Position uncertainty would remain constant at σ_w^2 throughout the day and be below the position variance prevailing under the pricing rule (2.18), except for the beginning and the end of the day. The mid-rate under this alternative rule would depend only on the previous period's transactions shock, $s(t)=s^o - (1/\delta)w(t\text{-}1)$. The *ex ante* variance of $s(t) - s^o$ is $(1/\delta)^2\, \sigma 4_w^{\ 2}$, which is always greater than the variance shown by equation (2.28). The numerical examples of *Table 2.1* illustrate this point. They show that the price variance (2.28) is generally less than one-tenth of the price variance under the alternative rule.

A third rule might be one whereby the dealer would always keep his quotation at the equilibrium level (s^o, z^o), implying zero price variance. Position uncertainty would, however, follow a random walk with the variance increasing cumulatively and reaching the value $T\sigma_w^2$ at the end of the day; *i.e.*, almost four times the maximum position uncertainty under the optimal rule. An auction mechanism would eliminate the need to keep inventories, but price fluctuations would be larger, because all stochastic customer orders would affect the transaction prices with full force in each period.

In the above examples the *length* of the trading day has been measured in terms of the number of trading sessions T, which has been assumed to be exogenously given. The results are therefore open to many interpretations depending on how the length of a trading session is defined in terms of real time units (minutes, hours, *etc.*). It also raises the question about whether T is exogenous.

Assume first that the length of the trading session is fixed at d and the length of the business day is equal to one. Linking two or more (N) business days together and assigning the closed position target to the end of the last day in the sequence is equivalent to allowing $T=N/d$ to increase. With the one-period transactions uncertainty σ_w^2 fixed, this

Table 2.1 Price and Position Uncertainty

T=40			T=60			T=80			T=100		
t	Var(s)[a)]	Var(x)[b)]	t	Var(s)[a)]	Var(x)[b)]	t	Var(s)[a)]	Var(x)[b)]	t	Var(s)[a)]	Var(x)[b)]
0	0.001	1.000	0	0.000	1.000	0	0.000	1.000	0	0.000	1.000
4	0.003	4.506	6	0.002	6.306	8	0.002	8.106	10	0.001	9.906
8	0.007	7.223	12	0.005	10.422	16	0.003	13.621	20	0.003	16.822
12	0.012	9.150	18	0.008	13.348	24	0.006	17.547	30	0.004	21.747
16	0.018	10.286	24	0.012	15.084	32	0.009	19.883	40	0.007	24.683
20	0.027	10.633	30	0.017	15.631	40	0.013	20.629	50	0.010	25.629
24	0.040	10.190	36	0.026	14.987	48	0.019	19.785	60	0.015	24.586
28	0.062	8.959	42	0.041	13.154	56	0.030	17.352	70	0.024	21.551
32	0.108	6.941	48	0.070	10.134	64	0.052	13.330	80	0.041	16.529
36	0.259	4.146	54	0.165	5.933	72	0.121	7.726	90	0.095	9.522
39	1.620	1.620	59	1.628	1.628	79	1.633	1.633	99	1.635	1.635
40		1.000	60		1.000	80		1.000	100		1.000

Explanations: a) The *ex ante* price variance is calculated from eq. (2.28) and is expressed as multiples of $(1/\delta)^2\ \sigma_w^2$. b) The *ex ante* position variance is based on eq. (2.26) and is expresssed as multiples of the one-period transactions uncertainty σ_w^2.

implies greater position uncertainty and larger required inventories. In addition, it would imply that the dealer has to be able and willing to carry, possibly large, open overnight positions. Assume, for instance, that d=1/40, and that two business days are combined to make one trading day, implying that T=80. From *Table 2.1* it can be seen that the maximum position uncertainty almost doubles, compared to the case where the position would be closed at the end of each business day (T=40).

Assume next that the length of the trading day matches the business day, but allow the length of the trading session to vary.[9] Increasing the number of trading sessions within the day raises the numerical value of T. In this case the maximum position uncertainty does not necessarily rise. Rather, it is likely to decline, though very little for large values of T. The exact result depends on the relationship between the one-period transactions uncertainty and the length of the session. It seems plausible to assume that the average volume of daily trade is independent of the number of trading sessions within the day. In that case the one-period transactions uncertainty measured in terms of the variance is more or less proportional to the length of the trading session[10], as is the average volume of one-period trade. Using the numerical examples of *Table 2.1*, assume that the one-period transactions uncertainty for d=1/80 is σ_w^2. For d=1/40 it would be $2\sigma_w^2$, and the maximum position uncertainty in the middle of the day would be only marginally higher than for d=1/80 (21.3 instead of 20.6, both in multiples of σ_w^2).

5 *Expected Revenue under Alternative Rules*

The above results show that by accepting open positions within some limits the dealer provides customers with a useful service. Not only does he save customers from waiting, but he also reduces the intra-day fluctuations of the exchange rates.

But there is another, and from the dealer's point of view more important, reason that makes the sequential pricing rule with the end-of-day position constraint superior to the alternative of keeping the

[9] Two interpretations are possible as regards the *length* of the trading session. One is that the dealer first announces a quotation and waits a short while until customers arrive. The other is that a customer arrives first and asks for the quotation without revealing whether he is a potential buyer or seller, for which reason the dealer has to announce a two-way price. In the latter case, which is perhaps more relevant for the foreign exchange market, the trading period is very short. Although *ex ante* the dealer does not know in which direction the deal will be made, *ex post* a transaction (if any) is made only on one side of the market. A new quotation will be announced when the next call arrives.

[10] Note that with a constant quotation the position is a random walk, and therefore its variance increases in proportion with time.

expected position closed throughout the day, *viz.* the expected revenue. This can be shown by examining the properties of the value function (2.13).

As shown in the *Appendix*, the expected revenue, or the monetary value of the optimal quotation at moment t, is

$$(2.29) \qquad J_t\,[x(t)] = (\alpha/\delta)[\,x(t) - x^*] - [\delta(T-t)]^{-1}\,[\,x(t) - x^*]^2$$

$$+(T-t)R^o - (1/\delta)W_t\,\sigma_w^2,$$

where $W_{T-1} = 0$ and $W_t = 1+(1/2)+(1/3)+\ \dots\ +[1/(T-t-1)]$, for $t = 0, 1, 2, \dots, T-2$. The sum of the first three terms is equal to $\alpha s(t)+\beta z^o - \delta[\,s^2(t) + z^{o2}]$ times $T-t$, which is equal to the trading income that would be earned between moment t and the end of the day in the case of no transactions uncertainty. It would also be equal to the expected income if at moment t the dealer decided to keep his quotation constant at $[s(t),z^o]$, no matter what happened to the position between moments t and T. But as shown above, the quotation is always adjusted when the position departs from the desired path. This uncertainty affects the expected revenue negatively, the reason being that any unexpected change in the position calls for an adjustment in the quotation in the opposite direction. For example, if there has been unplanned net purchases, the resulting long position will be sold out at somewhat lower prices, which contributes negatively to the income stream.

For comparison of the expected revenue under alternative pricing rules, assume that at the beginning of the day the position is closed, $x(0) - x^*=0$. The expected daily revenue associated with the equilibrium quotation is hence

$$(2.30) \qquad J_0\,(x^*) = TR^o - (1/\delta)W_0\,\sigma_w^2\,.$$

The second rule of aiming at the closed position after each transaction shock gives the following expected revenue:

$$(2.31) \qquad I_0\,(x^*) = TR^o - (1/\delta)T\sigma_w^2\,,$$

which is much smaller than J_0 (x^*), because $T>W_0=1+(1/2)+ \ldots + [1/(T-1)]$. The superiority of the optimal rule over the constant position target increases as the number of trading periods in a day becomes larger. The differences are illustrated in *Table 2.2.* The third alternative, included in the table, is the case of a constant equilibrium quotation (s^o,z^o) throughout the day, which would yield an expected revenue of size TR^o. This exceeds J_0 (x^*), but it would imply rapidly increasing position uncertainty. These observations suggest that, given a choice, the dealer chooses as short trading sessions as possible. A practical interpretation would be that the a new quotation is chosen each time the dealer receives a call from a customer, although the change in the quotation between two calls is, in the absence of new information, likely to be very small.

Because transactions uncertainty reduces the expected daily revenue, the latter may become negative. It turns out, however, that even for a comparatively large degree of transactions uncertainty the trading income remains positive. Let the standard deviation of the transaction shock be proportional to the average volume of one-period trade, $\sigma_w=k\beta/4$, and recall that the equilibrium trading income per period is equal to the spread times the average volume of trade, $R^o=\beta^2/4\delta$; *cf.* equation (2.8) above. Using this information in (2.30), it can be shown that J_0 $(x^*)>0$ if $T - (k^2W_0/4)>0$, or if $k<2\sqrt{(T/W_0)}$. For those values of T presented in *Table 2.2*, the critical value of k, for which J_0 (x^*) is positive, varies between 6 and 9. In a similar fashion, it can be shown that the expected revenue I_0 (x^*) under the alternative rule (2.31) of a constant position target remains positive as long as $k<2$.

Differences in position uncertainty between the alternative rules naturally imply different required inventories. Because position uncertainty is greatest in the case of a constant quotation throughout the day and smallest in the case of a constant closed position target throughout the day, the differences in expected net revenues are, in fact, not as large as indicated above. The costs of holding inventories are not explicitly taken into account in the above analysis, but they are introduced later in *Chapter Four*, although in a different context.

Nevertheless, the above results show that there is a trade-off between expected return and position uncertainty. If position uncertainty and hence the required inventories have to be reduced from the level

Table 2.2 Expected Revenue and Position Uncertainty

	T=40	T=60	T=80	T=100
Expected daily revenue[a]		(multiples of $(1/\delta)^2\sigma_u^2$)		
End-of-day target	- 4.285	- 4.687	- 4.972	- 5.194
Constant position target	- 40	- 60	- 80	- 100
Constant quotation	0	0	0	0
Maximum position variability		(multiples of σ_u^2)		
End-of-day target[b]	10.633	15.631	20.629	25.629
Constant position target[c]	1	1	1	1
Constant quotation[d]	40	60	80	100

Explanations : a) Deviation from TR^*. b) Maximum *ex ante* variance reached at $t = T/2$. c) Constant *ex ante* variance throughout the day. d) Maximum at the end of the day.

implied by the optimal pricing rule (2.18), it leads to lower average trading income and to more volatile prices. On the other hand, raising the average trading income by smaller and/or less frequent price adjustments implies much greater position uncertainty, which increases towards the end of the day. Greater position uncertainty at the end of the day increases the risk that the eventual open position has to be sold the following day at a loss. In this sense imposing a closed position target for the end of the day is equal to assuming a considerable degree of risk aversion.

6 *Extending the Horizon*

Extending the horizon over many trading days, or even indefinitely, does not change the optimal pricing rule at all, as long as the requirement that the expected position must be closed at the end of each trading day is maintained. The value function (2.13) at the end of the first day is no longer equal to zero. It is the discounted value of expected future trading incomes, but being constant it does not affect the current-day quotations (see *Appendix*). In other words the trading days are separable when solving the optimization problem.[11]

However, this extension brings out two interesting points, which deserve brief mention. One concerns the relationship between today's last quotation and tomorrow's first quotation, which may provide profit opportunities for informed customers. The other shows that in some situations the dealer himself may face the temptation to speculate deliberately with regard to open overnight positions.

It was seen above that the optimal pricing rule implies a constant expected quotation for the rest of the day and therefore eliminates the possibility of systematic intra-day speculation by customers. It was also seen that the price variance increases with the period of the day, which means that deviations from the equilibrium quotation are more likely toward the end of the day than at the beginning. Yet, the opening quotation $s(T+0)$ on the following day is known to be close to s^o; *cf.* *Table 2.1*. If the absolute value of $s(T-1) - s(T+0)$ is greater than the spread $2z^o$, an informed customer can make a profit by buying today and selling tomorrow, if today's closing price is lower than tomorrow's opening price, or *vice versa*, if today's closing price is higher.

The revenue-maximizing spread reduces the probability of this kind of an action, but it does not entirely eliminate it. A numerical example may again serve as an illustration. As above, let the standard deviation of the one-period transactions shock be proportional to the one-period average volume of trade, $\sigma_w=k\beta/4$. Assume that the distribution of the last quotation of the day is normal. Its standard deviation, denoted by σ_s can then be expressed in terms of the above-mentioned parameters; *cf.* equation (2.28). It becomes approximately equal to $\sigma_s =(k\beta/ 4\delta) \sqrt{1.6}=0.316\ k\beta/\delta$, where the number 1.6 is taken from *Table 2.1*. The probability **Prob** $\{s(T-1) - s^o>m\sigma_s\}$ can be read from statistical tables for each m. The critical values of m can be calculated from equation $m\sigma_s=2z^o$. Because $2z^o=\beta/\delta$, $m=1/0.316k$. When the degree of

---

[11] The only inter-day effect in the case of a fixed target for each separate day is the one-period transactions uncertainty, which makes the opening position dependent on the preceding day's last transaction shock.

transactions uncertainty is small ($k \leq 1$), the probability that $s(T-1)$ deviates from s^o by more than the size of the spread is negligible, less than one-third of a percentage point. However, it increases rapidly as transactions uncertainty rises, reaching 11 per cent for $k=2$ and 42 per cent for $k=4$.

As these examples show, informed customers, who know the dealer's pricing rule and are willing to speculate, do pose a problem. A defensive reaction on the part of the dealer against such insider-behaviour might be to widen the spread, although this would imply a smaller expected revenue.[12]

The strict independence of future quotations from those of today breaks down if the dealer is willing to carry open overnight positions and, in addition, is able to predict the changes in average customer orders in the coming days. As long as the customers are not able to make the same prediction, the dealer can exploit the profit opportunity created by his superior information (*cf. Section 3* for jumps within a given trading day).

To illustrate the point, assume that today at some moment t the dealer, but not the customers, learns that there will be higher demand tomorrow, implying that tomorrow's equilibrium mid-rate will rise and the spread will widen; *i.e.*, α and β will jump to α^1 and β^1, respectively. Maintaining the closed-position constraint for the end of tomorrow, but removing it from the end of today, is equivalent to saying that, in the face of the new *private* information, the dealer stands ready to carry an open overnight position between the two trading days. If the dealer wants to fully exploit the profit opportunity, the desired end-of-day position for today becomes endogenous. Let us denote this by $x^*(T)$. It is also the expected opening position for tomorrow. The planned quotation and the expected daily revenue for tomorrow are:

$$s^1(0) = (\alpha^1/\delta) - (1/\delta T)[\, x^*(T) - x^*] \tag{2.32}$$

[12] The risks of dealing with customers who have potential inside information are well known in the literature; *e.g.* Glosten and Milgrom (1985), Kyle (1985), Glosten 1989, and Dornbusch and Bechman (1985). Normally, it is assumed to arise from the possibility that the informed *insiders* learn the news before the dealers. In the present example the market maker's own pricing rule implies a similar risk. In either case a wider spread can be used to reduce or compensate for such a risk.

$$(2.33) \qquad J_0^{\,1}\,[x^*(T)] = (\alpha^1/\delta)[\,x^*(T) - x^*] - (1/\delta T)[\,x^*(T) - x^*]^2 + constant.$$

Both depend on tomorrow's expected opening position $x^*(T)$, which will be inherited from today. The constant term in the latter equation is equal to the expected daily revenue for a standard (non-speculative) day (*cf.* eq. (2.30)).

Assume, for simplicity, that the position at moment t happened to be closed, $x(t)=x^*$. The mid-rate and the expected trading income for the remainder of today are (*cf.* equations (A.6) and A.7) in the *Appendix*):

$$(2.34) \qquad s(t) = (\alpha/\delta) - [\delta(T-t)]^{-1}\,[\,x^* - x^*(T)]$$

$$(2.35) \qquad J_t\,(x^*) = (\alpha/\delta)[\,x^* - x^*(T)] - [\delta(T-t)]^{-1}\,[\,x^* - x^*(T)]^2 + constant,$$

where the constant depends on t (cf. eq. (A.7) in the *Appendix*).

The target position $x^*(T)$ for the end of today is obtained by maximizing the sum of the expected trading income for the remainder of today and the expected income for tomorrow. Omitting the discount factor, differentiating $J_t\,(x^*)+J_0^{\,1}\,[x^*(T)]$ with respect to $[\,x^*(T) - x^*]$, and setting the partial derivative equal to zero yields the following result:

$$(2.36) \qquad x^*(T) - x^* = T(T-t)\,[2(2T-t)]^{-1}\,(\alpha^1 - \alpha).$$

Today's target position will be long for $\alpha^1>\alpha$. Because the position at moment t happened to be closed, the price has to jump up immediately, implying net purchases for the remainder of the day. The resulting long position at the end of the day will then be sold tomorrow at a somewhat higher price, which, however, will be lower than it would be if the day had started with a closed position. A lower price is needed in order to generate the net sales required to undo the inherited long position.

The dealer's speculation with regard to an open overnight position is stabilizing in the sense that the average quotations today and tomorrow are closer to each other than they would be in the case of a separate closed position target for each day. In a similar manner as above, the result is based on the assumption that the change in tomorrow's demand is the dealer's private information. If that information is also known to customers, they may shift planned purchases from tomorrow to today, or postpone planned sales from today to tomorrow, which tends to equalize the quotations because customer orders are distributed more equally between the two days.[13]

7 *A Note on Risk Aversion*

In all the examples examined so far the dealer has been assumed to maximize expected revenue subject to certain constraints, which in the terminology of the standard risk-return analysis is equivalent to assuming risk neutrality. However, it was pointed out that this is not strictly true, because the end-of-day position target can be interpreted in terms of risk aversion. It was also seen that, despite the risk-neutrality assumption in the formal maximization problem, transactions uncertainty affected the expected revenue negatively as a direct consequence of the fixed position target. Moreover, it was shown that in addition to position uncertainty, there is an insider risk related to the possibility of informed customers - as opposed to genuine liquidity-oriented customers - speculating with regard to predictable changes in optimal quotations between today and tomorrow, the risk of which could be reduced by widening the spread and accepting lower expected revenue.

Although these remarks should make it clear that proper attention has been given to risk considerations, nothing prevents us from adding yet another dimension of risk aversion to the basic one-period model and assuming that instead of expected revenue the dealer maximizes expected utility subject to a given position constraint. Assume, for simplicity, that the degrees of uncertainty are of equal size, $\sigma_u^2 = \sigma_v^2$. In accordance with the *mean-variance* approach, define the expected utility as being equal to the expected revenue adjusted for the risk factor (*cf.* eq. (2.17))

---

[13] Trading days can also be interconnected by overlapping time-zones. In *Chapter Three* it will be shown how local markets become interconnected and the exchange rates equalized across local markets and across time-zones, even if the local markets are isolated in the sense that customers are able to trade only in their local markets and there is only one monopolist dealer in each local market.

$$(2.37) \qquad \mathbf{E}_t \{U(t)\} = \mathbf{E}_t \{R(t)\} - \eta \, \mathbf{Var} \{ R(t) \}$$

$$= \alpha s(t) + \beta z(t) - \delta[s^2(t)+z^2(t)] - \eta[s^2(t)+z^2(t)]\sigma_w^2,$$

where $\sigma_w^2 = \mathbf{E}_t \{[u(t) \pm v(t)]^2\}$ and η is the risk aversion coefficient, $\eta>0$.

The effect of the risk term in the objective function can be seen by maximizing $\mathbf{E}_t\, U(t)$ with respect to $s(t)$ and $z(t)$, subject to a given net sales requirement. The optimal quotation is now

$$(2.38) \qquad s(t) = s^o - (1/\delta)[x(t) - x^*], \quad z = \beta[2(\delta+2\eta\sigma_w^2)]^{-1}.$$

The mid-rate remains unchanged, compared to the basic model, but the spread is altered. It is still constant, but smaller than in the case of risk neutrality. Because the structure of the problem does not change, the result generalizes directly to the sequential pricing decision.

The result that the spread becomes smaller because of risk aversion may appear surprising. Recall, however, that in the present context the risk-neutral spread z^o is the price charged by a monopolist dealer for his services. If risk aversion reduces the spread, it also reduces expected profits, brings in more customers and increases the average volume of trade. Transactions uncertainty per trading session remains unchanged in absolute terms, but decreases in relation to the average volume of trade.

Applying the same mean-variance approach, but relaxing the assumption of symmetry by allowing the degrees of uncertainty to differ on the two sides of the market, yields another interesting result in the sense that the spread now becomes dependent on the position. In order to demonstrate this, assume for simplicity that customer sales are certain, $\sigma_v^2=0$, while customer purchases are stochastic, $\sigma_u^2>0$. The variance of the expected revenue is $[s(t)+z(t)]^2 \sigma_u^2$. Inserting this into the expected utility function and maximizing $\mathbf{E}_t\, U(t)$ with respect to $s(t)$ and $z(t)$, one obtains for the spread

$$(2.39) \qquad z(t) = [2(\delta+\eta \, \sigma_u^2)]^{-1} \{\beta - 2\eta\sigma_u^2 \{ s^o - (1/\delta)[x(t) - x^*]\}.$$

The mid-rate does not change, because it is governed only by the required net-sales target. It is, however, evident that the spread depends on the required net sales positively. Recall that the mid-rate depends on the desired net sales negatively. Taken together, these results imply that for position adjustment purposes the required change in the ask-rate, $s(t) + z(t)$, will be smaller than in the bid-rate, $s(t) - z(t)$. In other words, the price instrument is used more actively on that side of the market where it is more efficient or more predictable.

In a similar fashion it can be shown that assuming greater uncertainty on the sellers' side ($\sigma_v^2>0$) than on the buyers' side ($\sigma_u^2=0$) makes the spread negatively dependent on the required net sales. The adjustements in the ask-rate are larger than in the bid-rate, because it has a more predictable effect on the position.

8 Concluding Remarks

The above analysis has been based on as few behavioural and environmental assumptions as possible. The three behavioural assumptions were the price sensitivity of customer orders, the profit motive of the dealer and the latter's concern about his foreign exchange position. The environmental assumptions were the presence of a single foreign exchange dealer operating under isolated circumstances, the division of time into distinct trading days, transactions uncertainty arising from the stochasticity of customer orders and the possibility of new information.

The assumption of a monopolist market structure helps one to concentrate on the essential features of pricing behaviour. This has been the focus of the analysis so far. In the case of pure transactions uncertainty, the monopoly assumption affects only the bid-ask spread, which is always chosen in such a way that it maximizes the expected return on the equilibrium volume of one-period trade, whereas the mid-rate is adjusted for the position adjustment purposes. If the dealer has private information on the size and the timing of forthcoming market orders, this information affects the mid-rate in advance, whereas the spread is adjusted when the transactions actually take place.

The distinction between the spread and the mid-rate makes explicit the two separate aspects related to foreign exchange dealing. First, it can be looked upon as business activity, the purpose of which is to generate profit for the operator. The spread and the volume of trade are the two sources of this profit. The second aspect, settlement of non-bank customers' international payments, is in fact a consequence of the dealer's business activity. Mid-rate adjustment is important for achieving this outcome.

The buy and sell orders of genuine liquidity-oriented commercial customers are the dealer's bread and butter. It is in the dealer's interest to guarantee that the two-way orders match on the average. Hence it is the customer orders that determine the average levels of exchange rates. It was shown above that aiming at a closed position at the end of the day, rather than continuously, brings about the average matching of customer orders, reduces intra-day price fluctuations and raises the dealer's expected revenue.

Although the buy and sell orders are not planned so that they balance in each short period, the price adjustment follows the *Walrasian* rule in the sense that realized excess demand calls for an increase in prices. However, only the latest transactions shock affects the current quotation, and even then its effect is only partial, because it need not be undone in a single dealing session.

An important implication of the results is that, although the dealer is prepared for frequent price revisions, he is reluctant to make large adjustments in prices. The reason for this is simple. If he has overbought dollars in the current session, he does not rush to sell them back at too low a price, provided that there is enough time left to undo the position gradually and provided that no new information hits the market. New information is, of course, possible, in which case it is in the dealer's interest to react immediately and promptly.

Transactions uncertainty and the possibility of new information imply that the dealer's income is inevitably a risky one. With one exception, the above analysis has been left carried out assuming risk-neutrality in the sense that the dealer maximizes expected revenue rather than risk-adjusted expected utility. That the expected revenue is risky does not matter *per se* as long as it is in proper relation to invested resources. The resources side has so far been left out of the analysis, although it has been made clear that the major resource consists of inventories of currencies.

Despite the absence of risk aversion in the objective function, transactions uncertainty matters for the expected revenue. The reason for this is that some fluctuations in quotations become inevitable and they are generally revenue-reducing. Transactions uncertainty matters for another reason as well. It determines the degree of position uncertainty, which has two consequences. First, the required inventories are likely to be positively dependent on the degree of position uncertainty. Secondly, accepting open positions implies an exchange rate risk. The latter risk exists even though the dealer is the sole price maker in the market, because the average customer orders may shift leaving the dealer with a position which he can get rid of only with a loss. In the very short run, this risk may be indistinguishable from transactions uncertainty, but the risk increases if open positions are held for longer periods, and especially if there are discontinuities in business preventing the dealer from being in constant touch with the market. This is the major justification for the closed-position target for the end of the business day.

Appendix

The dealer's optimal pricing rule is derived by solving the dynamic programming algorithm as described by equations (2.13) to (2.16). The solution is obtained by starting with the last trading session *(T-1, T)* and proceeding recursively backwards in time.

For the last period the problem is the simple one-period case already discussed in the main text:

$$\text{(A.1)} \qquad \max_{s(T\text{-}1),\, z(T\text{-}1)} \{\alpha\, s(T\text{-}1)+\beta\, z(T\text{-}1) - \delta\, [s^2(T\text{-}1) + z^2(T\text{-}1)\}$$

$$\text{s.t. } x(T\text{-}1) - \mathbf{E}_{T\text{-}1}\{x(T)\} - \alpha+\delta\, s(T\text{-}1)] = 0, \text{ and}$$

$$\mathbf{E}_{T\text{-}1}\{x(T)\} = x^*$$

Define the Lagrangian

$$\text{(A.2)} \qquad L(T\text{-}1) = \alpha\, s(T\text{-}1)+\beta\, z(T\text{-}1) - \delta\, [\, s^2(T\text{-}1) + z^2(T\text{-}1)]$$

$$+ \mu(T\text{-}1)\, [\, x(T\text{-}1) - x^* - \alpha + \delta\, s(T\text{-}1)]$$

and differentiate it with respect to the control variables $s(T\text{-}1)$ and $z(T\text{-}1)$ and the Lagrange multiplier $\mu(T\text{-}1)$ and set the partial derivatives equal to zero to obtain the first-order conditions

$$\text{(A.3)} \qquad \alpha - 2\delta\, s(T\text{-}1) - \delta\mu(T\text{-}1) = 0, \qquad \beta - 2\delta\, z(T\text{-}1) = 0,$$

from which the optimal quotation and the Lagrange coefficient are obtained:

$$\text{(A.4)} \qquad s(T\text{-}1) = s^o - (1/\delta)[\, x(T\text{-}1) - x^*], \qquad z(T\text{-}1) = z^o$$

$$\mu(T\text{-}1) = s^o - (2/\delta)[\, x(T\text{-}1) - x^*],$$

where $s^o = \alpha/\delta$ is the equilibrium mid-rate that equalizes expected sales and purchases and $z^o=\beta/2\delta$ is the revenue-maximizing spread; *cf.* eq. (2.12) in the text. The Hessian matrix of the second-order partial derivatives is always negative definite, which guarantees that the second-order conditions for a maximum are satisfied. This is seen also from the fact that for a given spread the expected revenue function is concave (parabola opening downwards).

Inserting the optimal last-session quotation (A.4) into the expected revenue function gives the value of the last-session problem as a function of the position:

(A.5) $$J_{T-1}\, x(T-1) = (\alpha/\delta)[\, x(T-1) - x^*] - (1/\delta)[\, x(T-1) - x^*]^2 + R^o\,,$$

where R^o is the equilibrium one-period revenue; *i.e.*, the maximum expected revenue subject to zero expected net sales.

Assume that the general solution for the optimal quotation at an arbitrary moment $t=0,\ 1,\ ...,\ T\text{-}1$ is

(A.6) $$s(t) = s^o - [\delta(T\text{-}t)]^{-1}\,[\, x(t) - x^*], \qquad z(t) = z^o,$$

and that the value function is

(A.7) $$J_t\,[x(t)] = (\alpha/\delta)[\, x(t) - x^*] - [\delta(T\text{-}t)]^{-1}\,[\, x(t) - x^*)^2$$
$$+ (T\text{-}t)R^o - (1/\delta)W_t\,\sigma_w^{\;2}\,,$$

where $W_t = 1+(1/2)+(1/3)+ \ldots +[1/(T\text{-}t\text{-}1)]$ for $t=0,\ 1,\ ...,\ T\text{-}2$ and $W_t=0$ for $t=T\text{-}1$.

If this is the case, then the optimal quotation at moment $t\text{-}1$ is obtained by solving

(A.8) $$J_{t-1}\,[x(t\text{-}1)] = \mathbf{Max}_{s(t-1),\, z(t-1)} \quad \{\alpha\, s(t\text{-}1)+\beta\, z(t\text{-}1)$$
$$- \delta\,[\, s^2(t\text{-}1) + z^2(t\text{-}1) + \mathrm{E}_{t-1}\,\{J_t\,[x(t)]\}\},$$

For the solution, write first the expected value of (A.7) in the explicit form

(A.9) $$\mathrm{E}_{t-1}\,\{J_t\,[x(t)]\} = (\alpha/\delta)[\, x(t\text{-}1) - x^* - \alpha+\delta\, s(t\text{-}1)]$$
$$- [\delta(T\text{-}t)]^{-1}\,[\, x(t\text{-}1) - x^* - \alpha+\delta\, s(t\text{-}1)]^2$$
$$+ (T\text{-}t)R^o - [\delta(T\text{-}t)]^{-1}\,\sigma_w^{\;2} - (1/\delta)\sigma_w^{\;2}.$$

It is immediately seen that $W_t + (T\text{-}t)^{-1} = W_{t-1}$. Solving the problem expressed in (A.8) leads to the following quotation:

(A.10) $$s(t\text{-}1) = s^o - [\delta(T\text{-}t+1)]^{-1}\,[\, x(t\text{-}1) - x^*], \qquad z(t\text{-}1) = z^o,$$

At this quotation the value function (A.8) is

(A.11) $$J_{t-1}\,[x(t-1)] = (\alpha/\delta)[\,x(t-1) - x^*] - [\delta(T-t+1)]^{-1}$$

$$\times[\,x(t-1) - x^*]^2 + (T-t+1)R^o - (1/\delta)W_{t-1}\,\sigma_w^{\,2}.$$

This completes the proof that the optimal quotation is, indeed, the one expressed in (A.6).

The procedure can be extended to two (or more) days. In the two-day case the second day is solved first. The dynamic programming problem is written as

(A.12) $$J_{2T}\,[x(2T)] = 0$$

(A.13) $$J_{T+t}\,[x(T+t)] = \underset{s(T-3),\, z(T-3)}{\mathbf{Max}} \quad \{\alpha\, s(T+t)+\beta\, z(T+t)$$

$$- \delta\,[\,s^2(T+t) + z^2(T+t)] + \mathrm{E}_{T+t}\,\{J_{T+t+1}\,[x(T-2)]\}\},$$

A(14) $$x(T+t+1) = x(T+t) - \alpha+\delta s(T+t)+w(T+t)$$

(A.15) $$\mathrm{E}_{T+t}\,\{x(2T)\} = x^*,$$

where $t=0,\ 1,\ 2,\ \ldots,\ T-1$.

The solution is the same as above. The value of the problem at the beginning of the second day is

(A.16) $$J_{T+0}\,[x(T+0)] = (\alpha\,/\delta)[\,x(T+0) - x^*]$$

$$- (1/\delta)[\,x(T+0) - x^*)^2 + \mathrm{T}R^o - (1/\delta)W_0\,\sigma_w^{\,2}.$$

The starting position in the second day is inherited from the first day, $x(T+0) = x(T)$. The requirement that the expected value of $x(T)$ must be equal to x^* is maintained. The problem for the first day changes only as regards the terminal value $J_T\,[x(T)]$; *cf.* equation (2.14). It is no longer equal to zero, but is replaced by (A.16) discounted by a discount factor. The same procedure can be repeated arbitrarily many times and even indefinitely. The trading days are separable in optimization as long as the requirement that $\mathrm{E}_{NT-t}\,\{x(NT)\} = x^*$ is maintained for all future trading days.

Chapter Three

INTERBANK MARKET AND COMPETITION

This essay concludes the analysis of the pricing behaviour of an individual foreign exchange dealer. The main aim is to show how foreign exchange markets become globally integrated and efficiently organized through the profit-maximizing behaviour of individual foreign exchange dealers.

In *Chapter Two* dealer behaviour was analyzed under isolated circumstances in the sense that domestic currency was exchanged for foreign currency (currencies) in only one marketplace. It was further assumed that there was only one dealer who made the local market by standing ready to trade on the basis of incoming orders on immediate demand. Being the sole market maker for the domestic currency, the dealer could enjoy a monopoly position *vis-à-vis* his commercial customers. The implication of this was that the dealer could exercise his price-making power for both profit-maximizing and position-adjustment purposes.

Although the circumstances referred to above are far from those characterizing the near-perfect foreign exchange market in reality, the results of the previous chapter are applicable to more complicated and more realistic situations, such as customer flows and inter-dealer competition. In the following we break down the isolation by assuming that there are a large number of foreign exchange dealers servicing their local customers but also transacting with each other. From the point of view of any individual dealer the interbank market performs the role of a wholesale facility in which the dealer can cover positions which have resulted from stochastic customer deals. Interbank transactions also serve as an integrating device in the sense that the quotations in different local markets are drawn close to each other.

Because in the interbank market any one dealer behaves not only as a market maker but also as a customer in relation to other dealers, this tends to increase the price sensitivity of the demand and supply schedules faced by each individual dealer. As a consequence, the spread narrows, reducing the dispersion of quotations even further. In other words, interbank transactions bring competition into the picture, which constrains the price-making power of each individual dealer. The same would happen if non-dealer customers were able to 'shop around' by requesting prices from a large number of market makers.

Although the price-making power of each individual dealer may, under competition and the global integration of the foreign exchange markets, become fairly limited, dealers remain price makers in the sense that each time they receive a call from a customer, who may be a dealer customer, they announce a two-way price and stand ready to trade at that price without knowing in advance whether the customer is a buyer or a seller, and without knowing how other dealers are quoting at the same moment. Quoting a price that is out of line with prices quoted by others may imply that the deal is lost to competitors, or, even worse, it may give competitors an opportunity to reap quick profits by an act of arbitrage. It is therefore essential that each dealer remains in constant touch with the market and attempts to guess how others are quoting at the same time as he is making his own price.

Despite the fact that the emphasis in this chapter is shifted to the interbank market, it should be borne in mind that it is customer orders and transactions uncertainty, together with each dealer's concern about his foreign exchange position, which keep the market alive. Although speculative or arbitrage profits may arise occasionally, it is the spread and the volume of trade which are the source of the regular trading income for an individual dealer as well as for the dealer community as a whole.[1] When placing an order in the interbank market, an individual dealer is on the 'wrong side of the market' as compared with a situation where he himself is the market maker. Operating on the wrong side may sometimes be desirable because an open position resulting from a customer deal (or a sequence of such deals) can thereby be covered.

When the spread, thanks to competition, is very small and the dealers are in continuous touch with each other, a transactions shock originating in any of the local markets is quickly transmitted to other markets through interbank transactions. If the price quoted by an individual dealer

---

[1] In practice, the bid-ask spread need not be the only source of income. The common understanding is that very short-term speculation on intra-day price movements is not only prevalent, but also profitable (Goodhart, 1988, pp. 454-457).

depends on his exchange position and if the position varies from one trader to another, a single transactions shock has a negligible effect on price quotations on the average. In a similar way as transactions shocks are transmitted spatially between the local markets, they are transmitted temporally between overlapping time-zones. This happens as those dealers who are closing business for the day attempt to sell the remaining open positions to dealers who operate in another time-zone where business is just opening.

As in the previous chapter, the average level of the exchange rate is determined by the macroeconomic structure that generates the non-dealer customers' average buy and sell orders. The dealer community need not play any role in this respect, despite the fact that in each individual transaction there is a dealer who makes an explicit decision upon the price. Indeed, each dealer is free to ignore the fundamentals altogether as long as he tries to avoid open overnight positions and reacts to, say, earlier net purchases of a given currency by quoting a lower price and/or by searching out from amongst other market makers a counterparty who is ready to pay an attractive price for his excess balances. This kind of behaviour brings about a global market, which is efficient in the sense that at each moment the quotations are practically equal everywhere, the spread is very small and there is no possibility to predict in a systematic way the direction in which future quotations will move. No *continuous market clearing* is needed to achieve these properties, which approximate those of a perfect market.

1 Introducing Interbank Transactions

Assume that the global foreign exchange market is composed of a large number of local markets with one foreign exchange dealer in each. Assume further that transaction and information costs prevent non-dealer customers from trading outside their local market. The latter assumption, which will be relaxed later on, implies that each dealer can control the expected buy and sell orders of his local customers by changing quotations. To that extent the results of earlier sections are directly applicable. The complication that arises is that we now allow for dealers to send buy and sell orders to each other.

Any one dealer can, whenever he so desires, place an order in the interbank market. Note that in an interbank transaction the roles of the two counterparties differ. The one who takes the initiative for a transaction by requesting a quotation acts in the role of a customer, while his counterparty, who makes a price, acts in the role of a market maker. In order to distinguish between different types of transactions, let us call the interbank transactions the dealer makes on his own initiative *cover transactions*.

For the time being, it is assumed that the share of any one dealer in total trade is so small that he can disregard any systematic reactions by others to his own actions. Quantitative characteristics, especially those describing the customers' average buy and sell orders, may differ from one local market to the next, but qualitatively the situation is the same in each local market. As above, each dealer maximizes his trading income during the day subject to price-dependent customer orders and to the constraint that the expected foreign exchange position at the end of the day is closed (or at some other well-defined target level).

Let us first consider a dealer who is able to initiate a transaction with other dealers, but who never receives orders from other dealers. The justification for this, admittedly artificial, assumption might be, for instance, the small size of the local market, or that the dealer applies too wide a spread.

Omitting the stochastic terms, the expected buy and sell orders of the customers in a local market j are written as follows:

$$p_j(t) = a_j(t) - c[s_j(t) + z_j(t)] \tag{3.1}$$

$$q_j(t) = b_j(t) + c[s_j(t) - z_j(t)], \tag{3.2}$$

where $s_j(t)+z_j(t)$ is the ask-rate and $s_j(t)-z_j(t)$ the bid-rate quoted by the dealer. The shift parameters $a_j(t)$ and $b_j(t)$ take into account the local characteristics of the market j and c describes the price sensitivity of customer orders.

Let the dealer's initial position be equal to $x_j(t)$ and define the *closed* position at the end of the period as $x_j(t+1) = 0$. The size of a possible cover sale in the interbank market is denoted by $P_j(t) - Q_j(t)$, with $P_j(t) > 0$ representing a sale and $Q_j(t) > 0$ representing a purchase. When $P_j(t) = Q_j(t) = 0$, no cover transaction is made. Note that, due to the positive spread in the interbank market, a simultaneous sale and purchase in the interbank market is not profitable, except in rare cases when quotations are inconsistent (see below).

Although the dealer may not know exactly what prices other dealers will quote if he places an order with them, it is assumed that, being in constant touch with the market, he knows at least the average bid- and ask-rate quoted in the outside markets.[2] He can take this information into account when making his own price and deciding upon a possible wholesale transaction. Hence, he knows that he can buy foreign exchange in the interbank market at a price that most likely is not higher than $S^a(t)$ (the average outside ask-rate) and sell foreign exchange there at a price that most likely is not lower than $S^b(t)$ (the average outside bid-rate).

The dealer's net revenue is the expected trading income from customer deals adjusted for the cash outflow (inflow) of a possible cover purchase (sale):

$$(3.3) \qquad R_j(t) = \alpha_j(t)s_j(t) + \beta_j(t)z_j(t) - \delta[s_j^2(t) + z_j^2(t)], \\ + S^b(t)P_j(t) - S^a(t)Q_j(t),$$

where $\alpha_j(t) = a_j(t)\text{-}b_j(t)$, $\beta_j(t) = a_j(t)+b_j(t)$, δ=2c. The first three terms indicate the net revenue from customer deals and the remaining terms take into account the contribution of a cover transaction.

The dealer is assumed to maximize the one-period net revenue subject to the constraint that at the end of the period the position is closed; *i.e.*, $x_j(t) - P_j(t) + Q_j(t) - p_j(t) + q_j(t) = x_j(t+1)=0$. The corresponding Lagrangian

$$(3.4) \quad L_j(t) = R_j(t) + \mu_j(t)[x_j(t) - \alpha_j(t) + \delta s_j(t) - P_j(t) + Q_j(t)]$$

is discontinuous and therefore non-differentiable with respect to $P_j(t)$ and $Q_j(t)$ at those points where either a cover sale or a cover purchase becomes

---

[2] This is a somewaht tricky assumption, because if he knows the average he should know something about the distribution and be able to find a better price than the average. In practice, all market participants are fully informed about the indicative prices, which, though frequently revised, are continously displayed on *Reuters* and *Telerate* (*cf.* Goodhart and Demos, 1990a, pp. 333-335). For the purposes of the subsequent analysis one could regard these indicative prices as representing estimates of the average transaction prices for each moment.

profitable. Taking this into account, the differentiation of (3.4) with respect to the four decision variables, $s_j\ (t)$, $z_j\ (t)$, $P_j\ (t)$ and $Q_j\ (t)$, gives the following first-order conditions:

$$\begin{array}{lll}
(3.5) & \text{(i)} & \alpha_j\ (t) - 2\delta s_j\ (t) + \delta\mu_j\ (t) = 0 \\
 & \text{(ii)} & \beta_j\ (t) - 2\delta z_j\ (t) = 0 \\
 & \text{(iii)} & \mu_j\ (t) = S^b \text{ when } P_j\ (t) > 0,\ Q_j\ (t) = 0 \\
 & & \mu_j\ (t) = S^a \text{ when } P_j\ (t) = 0,\ Q_j\ (t) > 0 \\
 & & S^b(t) < \mu_j\ (t) < S^a(t) \text{ when } P_j\ (t) = Q_j\ (t) = 0.
\end{array}$$

The spread, $z_j\ (t) = \beta_j\ (t)/2\delta$, depends only on the local market characteristics, but is independent of the position. The mid-rate is state-dependent and depends on the size of a possible cover transaction, in addition to the initial position, as follows:

$$(3.6) \qquad s_j\ (t) = s^o{}_j\ (t) - (1/\delta)[x_j\ (t) - P_j\ (t) + Q_j\ (t)].$$

The corresponding shadow price is

$$(3.7) \qquad \mu_j\ (t) = s_j\ (t) - (2/\delta)[x_j\ (t) - P_j\ (t) + Q_j\ (t)].$$

Because $S^a > S^b$, simultaneous cover purchases and sales are out of the question. Whether the cover transaction is a sale, a purchase, or neither depends on the initial position, as well as on the outside prices, as follows:

$$\begin{array}{ll}
(3.8) & P_j\ (t) > 0,\ Q_j\ (t) = 0 \text{ if } x_j\ (t) > (\delta/2)[s^o{}_j\ (t) - S^b(t)] \equiv x^u{}_j\ (t) \\
 & P_j\ (t) = 0,\ Q_j\ (t) > 0 \text{ if } x_j\ (t) < (\delta/2)[s^o{}_j\ (t) - S^a(t)] \equiv x^l{}_j\ (t) \\
 & P_j\ (t) = Q_j\ (t) = 0 \text{ if } x^l{}_j\ (t) \leq x_j\ (t) \leq x^u{}_j\ (t),
\end{array}$$

where $x^u{}_j\ (t)$ is that level of an open (*long*) position above which a cover sale becomes profitable. The reason for this is the fact that at $x_j\ (t) > x^u{}_j\ (t)$ the shadow price of foreign exchange would become lower than the bid-price available in the interbank market. Similarly, $x^l{}_j\ (t)$ is the size of the open (*short*) position below which a cover purchase becomes profitable, because for $x_j\ (t) < x^l{}_j\ (t)$ the shadow price would become higher than the ask-price available in the interbank market. These position limits result from the availability of outside quotations and they determine the range for the prices quoted to customers. In terms of the mid-rate, this is

$$(3.9) \qquad [s^{o}_{j}(t) + S^{b}(t)]/2 \leq s_{j}(t) \leq [s^{o}_{j}(t) + S^{a}(t)]/2.$$

Figure 3.1 illustrates the result. Both the mid-rate (ss-line) and the shadow price ($\mu\mu$-line) are depicted as a function of the initial position that has to be squared during one trading session. Within the range $[S^{b}(t), S^{a}(t)]$ the two lines are downward sloping, with the $\mu\mu$-line declining twice as fast as the ss-line. The limits $x^{l}_{j}(t)$ and $x^{u}_{j}(t)$ are determined at those points where the $\mu\mu$-line intercepts the bid- and ask-rates available in the interbank market. Beyond these points the $\mu\mu$-line becomes horizontal. The corresponding limits to the mid-rate are obtained from the ss-line at these position levels. If the initial position is above $x^{u}_{j}(t)$, it is adjusted by a cover sale of the size $x_{j}(t) - x^{u}_{j}(t)$ and the remaining $x^{u}_{j}(t)$ is closed by net sales to customers. If $x_{j}(t)$ is below $x^{l}_{j}(t)$, the position is adjusted by a cover purchase of the size $x^{l}_{j}(t) - x_{j}(t)$ and the remaining $x^{l}_{j}(t)$ is squared by net purchases from commercial customers.

As mentioned above, the price available in the interbank market is not well defined. At each moment, there are several dealers quoting prices without full knowledge of how others are quoting at the same moment. Therefore, the prices quoted by individual dealers may at any moment t differ across dealers. In *Figure 3.1* this is taken into account by drawing (non-observed) frequency distributions of the bid- and ask-prices in the market.[3]

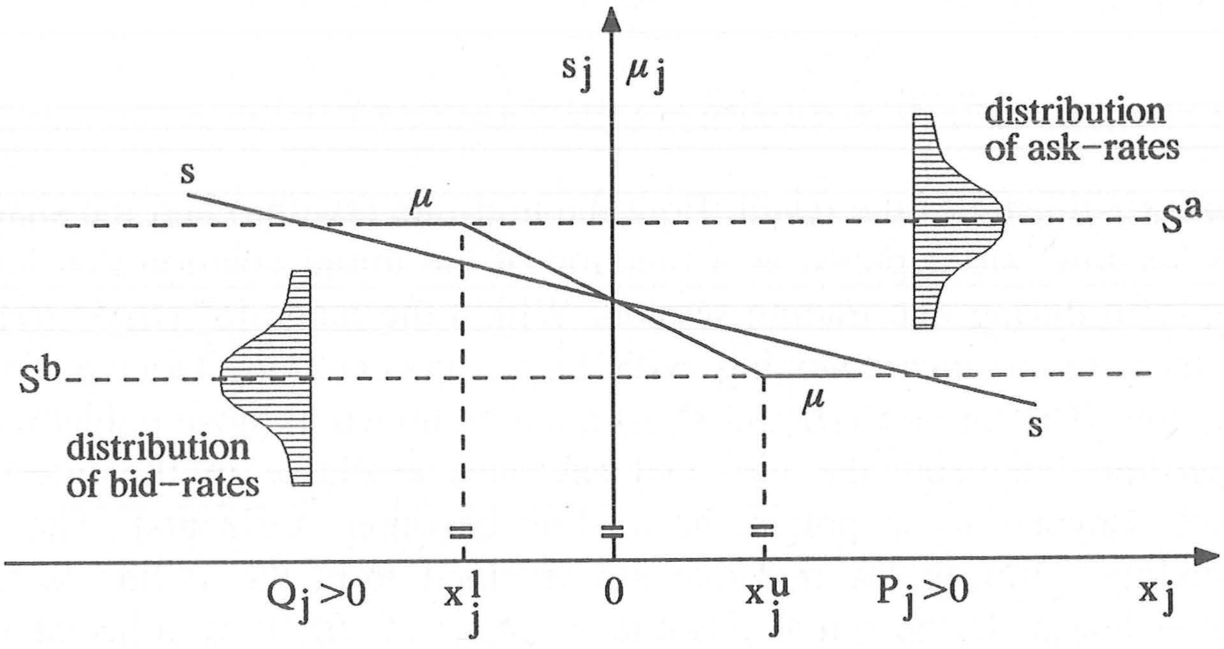
Above it was assumed that $S^{b}(t)$ and $S^{a}(t)$ represent indicative rates which are common knowledge to all dealers. If better prices are possible, it is in the dealer's interest in each situation to try to observe the best one of the available prices by requesting a binding quote from a number of counterparties. When such a request proves successful, the dealer concludes a cover deal before the position reaches either of the limits derived above.

The spread, $z_{j}(t) = \beta_{j}(t)/2\delta$, is always chosen to maximize the return on the equilibrium volume of customer trade (*cf. Chapter Two*). It is

---

[3] It is possible that he can find an ask-rate that is lower than some bid-rate, implying an arbitrage profit opportunity. We return to this situation in *Section 3.*

Figure 3.1 Cover Transactions in the Interbank Market

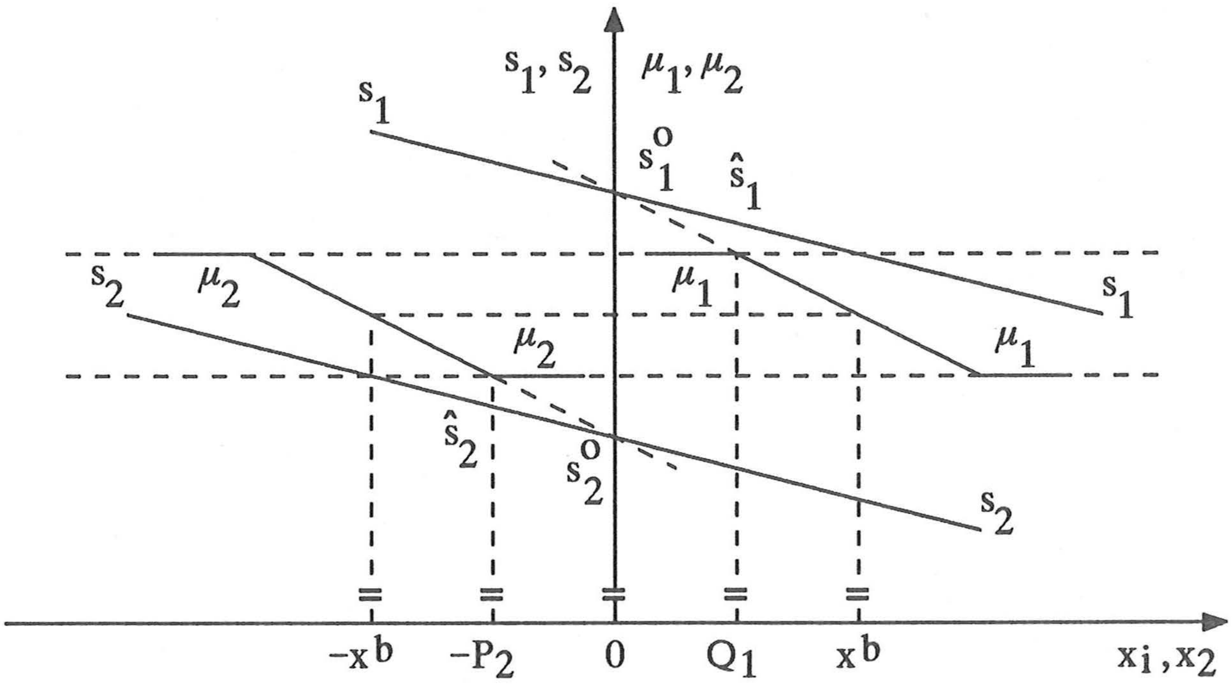


independent of the initial position as well as of the outside quotations. Because it depends on the characteristics of the local market, it may differ from one market to another. This result follows from the assumption that dealer *j* operates in an isolated environment in the sense that he does not receive interbank orders from other dealers and that local customers are unable to 'shop around'. These assumptions will be relaxed in *Section 2*.

The most important result, however, is the integration of local markets brought about by the cover facility in the interbank market. The possibility of placing interbank orders with other dealers tends to draw local quotations closer to the global average. At the same time the local characteristics become less important as far as the mid-rate is concerned.

Consider, for example, two local markets, each having only a small share of global trade. Let us assume that in the first market (market *1*) one currency is relatively scarce in the sense that the local customers' demand is high relative to what they supply. In the second market (market *2*) this currency is relatively abundant. *Figure 3.2* illustrates the point. Assuming that initially the dealers in the two markets have a closed position and assuming full isolation, they would quote the mid-rates s_1^o and s_2^o, respectively, the latter being much below the former because of the differences in local characteristics. Having the possibility to trade in the interbank market,

Figure 3.2 Interbank Transactions and Integration of Local Markets



dealer *1* finds it profitable to place a sell order in this market in order to purchase an amount Q_1 at the average ask-rate (the shadow price is equal to S^a). The size of the purchase would be larger if the dealer were able to receive an even lower ask-price. By lowering his own quotation somewhat to $\hat{s}_1$ or below, he can sell the resulting long position to his customers and make a higher profit compared to the case of complete isolation. Similarly, dealer *2* finds it profitable to sell an amount $-P_2$ in the interbank market at the average bid-rate (the shadow price is equal to S^b). At a bid higher than S^b the sale would be even larger. As a result dealer *2*'s quotation is drawn upwards to $\hat{s}_2$, or above, if he is able to observe a higher bid-price in the interbank market.

This is the mechanism by which local foreign exchange markets become integrated, even though some of them may be isolated from each other. The two dealers need not meet directly, nor through a broker or some other middleman; rather, they meet through the interbank market made up of other dealers standing ready to trade with dealers *1* and *2* on immediate demand. If these two dealers were symmetrically related to the *average market* and if they were able to meet directly, they would find it even more profitable to bypass the interbank market altogether and exchange an amount x^b bilaterally at a price which is somewhere between the average ask-rate and the average bid-rate quoted by the other dealers. As a result, their quotations would move even closer to each other. The bilateral exchange between the two dealers can, of course, be arranged by a broker. If the situation described above is temporary, it may be costly for the two potential partners to find each other, in which case an opportunity arises for a broker. Compared with two interbank transactions, a service provided by a broker is advantageous to both counterparties if the brokerage fee paid by each is less than half of the spread prevailing in the interbank market.[4]

2 *Customer Flows and Competition*

Although the two dealers in the above example could place orders in the interbank market, both remained in a relatively isolated environment and, in addition, both could fully exploit the monopoly power in relation to local customers. We can go further and relax the assumption that transaction and information costs prevent local customers from trading outside their local market. Instead of prohibitively high transaction costs, we can assume that local customers face greater transaction costs when trading outside than with the local dealer. The customers may also differ in this respect. In fact, we can assume a continuum of customers ordered according to their relative transaction costs. At one end, there are well-informed customers with low transaction costs. These customers choose to go outside once they observe a quotation that is only marginally better than the one quoted by the local dealer. At the other end, there are those customers for whom it is hardly ever possible to approach any of the outside dealers when they need to convert one currency into another. Allowing *customer flows* of this kind introduces inter-dealer competition into the model.

Though admittedly vaguely defined, these plausible economic frictions leave each local dealer with some freedom to make his own prices and hence to control the expected net sales by his own decisions. Accordingly, we can continue analyzing the behaviour of an individual dealer *as*

---

[4] This is one justification for the use of broker services in the foreign exchange market. These are rather common in practice, especially in cases when larger-than-normal amounts are being exchanged. According to the BIS Survey, 20 to 50 per cent of the total net turnover in the foreign exchange market in April 1989 was arranged through brokers (BIS, 1990, p. 10).

if he had a short-term monopoly.[5] As a market maker, the dealer announces his quotation upon request and stands ready to trade on the basis of any incoming order without knowing in advance whether the customer is a potential buyer or a seller and without knowing whether the same customer is asking quotations from other dealers at the same moment. By these assumptions the possibility of price discrimination is excluded.

For the sake of symmetry, we should allow for the possibility that the same dealers themselves may receive orders from outside and also that the other dealers are eager to find the best available prices for such transactions. These assumptions imply that equations (3.1) and (3.2), which describe the customers' buy and sell orders in the local market j, have to be modified.

Let us write the expected buy and sell orders received by a single local dealer j during a short period as follows:

$$p_j(t) = a_j(t) - cs_j^a(t) - d[s_j^a(t) - S^a(t)] \tag{3.10}$$

$$q_j(t) = b_j(t) + cs_j^b(t) + d[s_j^b(t) - S^b(t)], \tag{3.11}$$

where $[s_j^b(t), s_j^a(t)]$ is the price quoted by dealer j and $[S^b(t), S^a(t)]$ is the average quotation across dealers. The constant terms $a_j(t)$ and $b_j(t)$ take into account the local characteristics of the market j, whereas the price sensitivity parameters c and d are assumed to be equal across all local markets, *i.e.*, for all j. The parameter $d>0$ describes how sensitive the customers are to small price differentials, while the parameter $c>0$ shows how sensitive the non-dealer customers are, in the aggregate, to a uniform change in prices everywhere.[6] If the customers' information and transaction costs of trading outside the local market are low, then d is large relative to c. It will be even larger if the dealer does not regard interbank orders from the other dealers as purely random events, but instead recognizes the fact that he is more likely to receive orders from outside if his quotation differs from the global average.

---

[5] This assumption is similar to that of Phelps and Winter (1970), who analyzed atomistic competition with slow diffusion of information on prices between local marketplaces.

[6] This is seen by adding equations (4.4) and (4.5), respectively, over all N markets, $i=1, ..., N$, and taking the average.

As above, the dealer's expected trading income per period from customer trade adjusted for a possible cover transaction is

$$(3.12) \qquad R_j\,(t) = s_j^a(t)p_j\,(t) - s_j^b(t)q_j\,(t) + S^b(t)P_j\,(t) - S^a(t)Q_j\,(t)$$

$$= [\alpha_j\,(t) + \eta S(t)]s_j\,(t) + [\beta_j\,(t) + \eta Z(t)]z_j\,(t) - (\delta+\eta)[s_j^2(t) + z_j^2(t)]$$

$$+ [S(t) - Z(t)]P_j\,(t) - [S(t) + Z(t)]Q_j\,(t),$$

where $\alpha_j\,(t) = a_j\,(t) - b_j\,(t)$, $\beta_j\,(t) = a_j\,(t) + b_j\,(t)$, δ=2c, η=2d, $s_j\,(t) = [s_j^a\,(t) + s_j^b\,(t)]/2$, $z_j\,(t) = [s_j^a\,(t) - s_j^b\,(t)]/2$, $S(t) = [S^a(t) + S^b]/2$ and $Z(t) = [S^a(t) - S^b(t)]/2$. It is seen that the equation is formally similar to equation (27); only the parameters and their interpretation are different.

Assuming that the dealer maximizes the one-period revenue subject to the condition that the end-of-period position is closed leads to the following problem:

$$(3.13) \qquad \underset{s_j\,,\, zj\,,\, P_j\, -\, Qj}{\textbf{Max}} \quad R_j\,(t);$$

$$\text{s.t. } x_j\,(t) - [P_j\,(t) - Q_j\,(t)] - [\alpha_j\,(t) + \eta S(t) - (\delta+\eta)s_j\,(t)] = 0,$$

where the last term (within the brackets) of the system constraint is equal to net sales to customers associated with the mid-rate $s_j\,(t)$. The resulting Lagrangian is non-differentiable with respect to $P_j\,(t) - Q_j\,(t)$ at those points where this becomes either positive or negative. Applying the average interbank prices $[S^b(t), S^a(t)]$ to cover transactions, the first-order conditions for the maximum can be written as follows:

$$\begin{array}{lll}
(3.14) & \text{(i)} & s_j\,(t) = [\alpha_j\,(t) + \eta S(t)]/2(\delta+\eta) + (1/2)\mu_j\,(t) \\
 & \text{(ii)} & z_j\,(t) = [\beta_j\,(t) + \eta Z(t)]/2(\delta+\eta) \\
 & \text{(iii)} & \mu_j\,(t) = S^b(t), \ \text{when } P_j\,(t) > 0 \\
 & & \mu_j\,(t) = S^a(t), \ \text{when } Q_j\,(t) > 0 \\
 & & S^b\,(t) \le \mu_j\,(t) \le S^a(t), \ \text{when } P_j\,(t) = Q_j\,(t) = 0.
\end{array}$$

When $P_j\,(t) = Q_j\,(t) = 0$, the mid-rate is obtained directly from the system constraint, (3.15) $$s_j(t) = [\alpha_j(t) + \eta S(t) - x_j(t)]/(\delta+\eta).$$

The corresponding $\mu_j(t)$ is

(3.16) $$\mu_j(t) = [\alpha_j(t) + \eta S(t) - 2x_j(t)]/(\delta+\eta).$$

It is seen that the local quotation $[s_j(t), z_j(t)]$ now depends directly on quotations elsewhere $[S(t), Z(t)]$, a relationship which follows from the fact that customers (including dealer customers) are sensitive to price differentials across market makers. The spread is no longer entirely dependent on local conditions or on the local dealer's price making power.

Recall that local quotations were also dependent on the prices quoted elsewhere in the previous example, but in that case the result was entirely due to the fact that the local dealer could on his own initiative adjust his position by making cover transactions in the interbank market. In the present case, with customer flows and with the possibility of interbank orders arriving from other dealers, local quotations are drawn even closer to each other; *i.e.*, the spread narrows, as does the dispersion of prices. In the limiting case, as $\eta \rightarrow \infty$, the mid-rate becomes the same everywhere, $s_j(t) \rightarrow S(t)$ and the spread approaches zero, $z_j(t) \rightarrow Z(t)/2 \rightarrow 0$. The latter result follows from the fact that $z_j(t) \rightarrow Z(t)/2$ must hold for each j, which is possible only if $Z(t)$ approaches zero.

The conditions (3.14) determine whether or not a cover transaction is profitable. The answer depends on the initial position and on what price offers the dealer obtains in the interbank market. If the initial position $x_j(t)$ is such that without a cover transaction $\mu_j(t)$ would go above S^a, a cover purchase $Q_j(t) > 0$ is profitable until $x_j(t) + Q_j(t)$ reaches the point $x^l_j(t)$ at which $\mu_j(t) = S^a(t)$ (or the lowest observed ask-rate). Similarly, if the initial position would, without a cover transaction, bring $\mu_j(t)$ below $S^b(t)$, a cover sale $P_j(t) > 0$ becomes profitable until $x_j(t) - P_j(t)$ reaches the point $x^u_j(t)$ at which $\mu_j(t) = S^b(t)$ (or the highest observed bid-rate). Between these limits no cover transaction is profitable.

The problem is formally equivalent to the previous example. Therefore, *Figure 3.1* serves to illustrate the result in the present case as well. Because $\delta+\eta > \delta$, the ss-curve and the $\mu\mu$-curve both decline less steeply, and therefore the position, which is not covered in the interbank market, is allowed to fluctuate in a wider range:

$$(3.17) \qquad x^l_j(t) \equiv (1/2)(\delta+\eta)[s^o_j(t) - S^a(t)] \leq x_j(t)$$

$$\leq (1/2)(\delta+\eta)[s^o_j(t) - S^b(t)] \equiv x^u_j,$$

where $s^o_j(t) \equiv [\alpha_j(t) + \eta S]/(\delta+\eta)$ is the equilibrium mid-rate that would balance dealer j's customer orders (non-dealer and dealer alike) on the average. Note that it is possible that $s^o_j \geq S^a(t)$, in which case a closed initial position $x_j(t) = 0$ would imply an immediate cover purchase and a downward adjustment of the price, a process by which foreign currency flows from the interbank market to dealer j's customers; *cf. Figure 3.2.* Or *vice versa,* $s^o_j(t) \leq S^b(t)$ would imply an immediate cover sale and an upward adjustment of the price.

The less steep $\mu\mu$ -curve should reduce the likelihood of profitable cover transactions, because any open position can be easily unwound by marginal changes in quoted prices. On the other hand, greater price sensitivity reduces the size of the average spread and the dispersion of quotations, which increases the likelihood of profitable cover transactions in the interbank market. In the limiting case as $\eta \to \infty$ the spread becomes zero and $s_j(t)$ cannot deviate from $S(t)$, implying that it does not matter whether an open position is covered in customer trade or in the interbank market.

The price differentials and the resulting interbank transactions disappear in aggregation across all market makers. The global average price is

$$(3.18) \qquad S(t) = \mathrm{A}(t)/\delta - (1/\delta)X(t), \qquad Z = \mathrm{B}(t)/(2\delta+\eta),$$

where $\mathrm{A}(t) = \Sigma_n \alpha_n(t)/N$, $\mathrm{B}(t) = \Sigma_n \beta_n(t)/N$, and $X(t) = \Sigma_n x_n(t)/N$, the summation being taken over all N dealers $n=1, 2, ..., N$. The aggregated open position $X(t)$ is inherited from earlier global transaction shocks and is likely to be comparatively small in proportion to the global volume of trade. The effect of any local transaction shock on the global average price is small, because individual positions can be traded in the interbank market, whereby local shocks are diffused globally. Because the extent of genuine currency conversion needs by the non-bank public is likely to be limited in any short period, it is the interbank trade which grows most rapidly as the spread becomes increasingly narrow.

One consequence of having a narrower spread and larger volume of interbank trade is that non-bank customers benefit less and less from shopping around. In the limiting case, with zero spread and prices exactly the same everywhere, it would not matter which dealer a customer places his order with. However, interbank transactions would at the same time lose their economic meaning, because then it would not matter on which side of the market each dealer operates. In fact, the whole notion of the dealership market in producing liquidity services would lose its *raison d'être*.[7]

3 *Arbitrage and Uncertainty*

As long as there are economic frictions that make customer orders less than infinitely elastic with respect to price differentials and as long as the dealers do not have complete information about how others are quoting at the same moment, each single dealer is left with some price-making power, which can be used for position adjustment purposes. However, because each dealer quotes prices independently of, and simultaneously with, the other dealers, the quotations are bound to differ across the market makers at any particular moment. This implies the possibility of arbitrage profit opportunities from time to time.[8]

An arbitrage opportunity arises if one dealer quotes an ask-rate that is lower than the bid-rate quoted by some other dealer at the same moment. In this situation, any third party who is quick enough to observe the pair of inconsistent quotations has an opportunity to make an immediate riskless profit by buying from the former and simultaneously selling to the latter.

---

[7] Zero spread is equivalent to complete disappearance of monopoly profits arising from the dealers' capacity to exploit price-sensitive customer orders. The spread remains positive when the costs of producing dealer services are taken into account, even if the customers are infinitely sensitive to price differentials; see *Chapter Four*.

[8] "If the markets were perfectly arbitraged all the time, there are never any profits to be made from the activity of arbitrage. But then how do arbitragers make money, particularly if there are costs associated with obtaining information about whether markets are already perfectly arbitraged?", *cf.* Grossman and Stiglitz, 1976, pp. 247-248).

The arbitrager can be any third dealer or a well-informed non-dealer customer.

The two dealers who quoted inconsistently in the first place will lose if an arbitrage operation moves their positions far off the desired path. At the next moment they either have to restore the position in the interbank market at unfavourable prices or to make a relatively large adjustment to their customer quotation. In each case the quotation will change in the direction that will eliminate, or at least reduce, the probability of inconsistent quotations immediately after the event.

The possibility of an arbitrage opportunity was already illustrated in *Figure 3.1*. It shows the frequency distributions of the ask-rate and the bid-rate in the interbank market. As long as the two frequency distributions overlap, there is always a positive probability for an arbitrage opportunity to arise; *i.e.*, that at moment t two dealers m and n, without knowing each other's prices, quote in such a way that $s_m^a(t) < s_n^b(t)$.

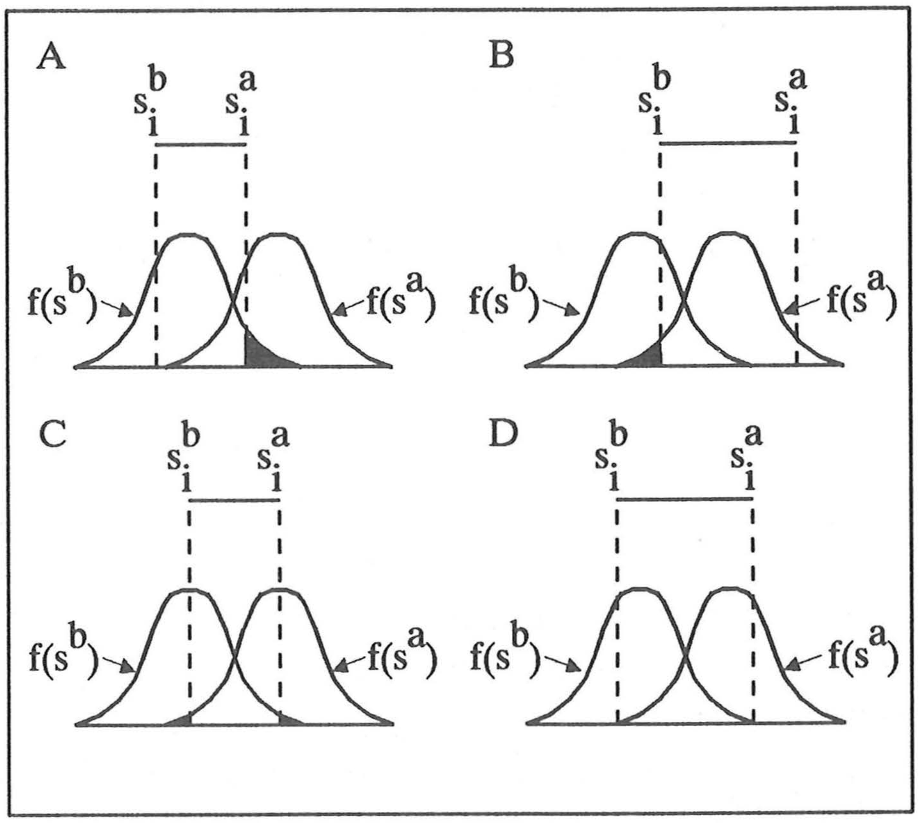
Figure 3.3 illustrates the point still further. Its four panels depict the frequency distributions of the ask-rate and the bid-rate quoted in the global market as well as the price quoted by a local dealer j at the same moment. The shaded area in each panel illustrates the probability of an inconsistent quotation for a given two-way price $(s_j^b(t), s_j^a(t))$ quoted by dealer j. The larger the shaded area, the larger is the probability that there is another market participant who is able to observe an inconsistent quotation and to make a profit at the cost of dealer j. A comparison between panels C and D shows that, for a given mid-rate, the wider the spread, the smaller is the probability of it being used in arbitrage between inconsistent prices.[9]
A comparison between panels A and C, or B and D, shows that this probability is positively related to the deviation of the mid-rate in dealer j's quotation from the global average price.

These observations suggest that imposing a cost for the risk of inconsistency will affect the individual dealer's spread positively and will draw the local quotations even closer to the global average. This can be shown formally by postulating the following cost function:

---

[9] This situation is parallel to the presence of informed *insiders* in the securities market. Because of information asymmetries, a stock dealer is unable to distinguish between insiders and *liquidity traders*. Because the dealer loses on the average in trade with insiders, he has to compensate for the loss by applying a spread which is wider than the marginal cost of providing the dealer services; see Glosten and Milgrom (1985), Kyle (1985), Glosten (1989), Dennert (1989) and Hagerty (1991).

Figure 3.3 Probability of Inconsistent Quotations



Note: In each panel *f(.)* shows the frequency distribution of the ask-rates and bid-rates quoted by the dealers at a particular moment. The sum of the shaded areas in each panel illustrates the probability distribution of inconcistency for a given price (s_i^b, s_i^a) quoted by dealer *i*.

(3.19) $$C_j(t) = c_1 z_j(t) + c_2 [s_j(t) - S(t)]^2, \quad c_1 > 0, c_2 > 0,$$

where the cost attached to the inconsistency risk is assumed to be linear with respect to z_j and quadratic with respect to $s_j(t) - S(t)$. Maximizing the net revenue function $R_j(t) - C_j(t)$, where $Rj(t)$ is defined in eq. (3.12), with respect to $s_j(t)$, $z_j(t)$ and $Pj(t) - Qj(t)$ and subject to the position constraint $x_j(t) - [\alpha_j(t) + \eta S(t) - (\delta+\eta)s_j(t)] - [(P_j(t) - Q_j(t)]$ gives the following first-order conditions:

$$
\begin{aligned}
(3.20)\quad &\text{(i)} && s_j(t) = [\alpha_j(t) + (\eta + 2c_2)\, S(t) + (\delta+\eta)\mu^c_j(t)\,]/2(\delta+\eta+c_2)\\
&\text{(ii)} && z_j(t) = [\beta_j(t) + \eta Z(t) + c_1\,]/[2(\delta+\eta)]\\
&\text{(iii)} && \mu^c_j(t) = S^b(t) \text{ when } P_j(t) > 0\\
& && \mu^c_j(t) = S^a(t) \text{ when } Q_j(t) > 0\\
& && S^b(t) < \mu^c_j(t) < S^a(t) \text{ when } Q_j(t) = P_j(t) = 0,
\end{aligned}
$$

where μ^c_j is the shadow price for the position constraint. The superscript c is added to distinguish it from μ_j in the previous case considered; *cf.* eq. (3.16).

It is seen that the inconsistency risk makes the spread wider, as expected. It widens as c_1 increases. However, with a very large spread, the dealer would lose all his customers.

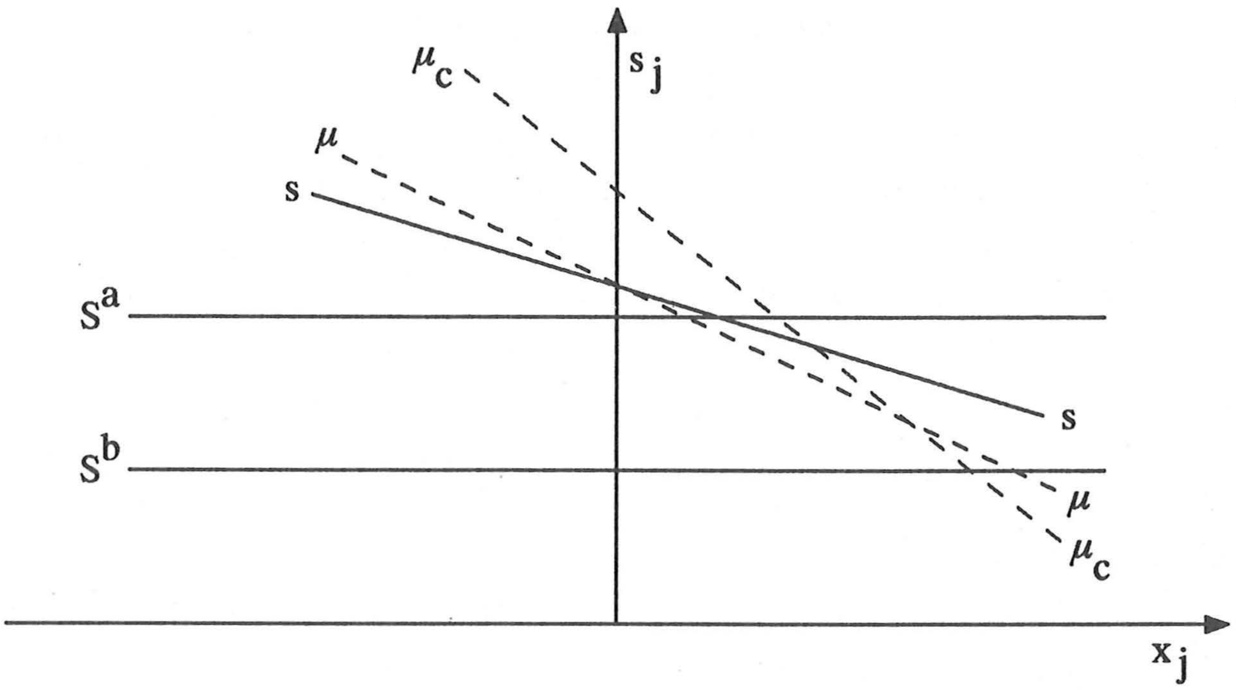
Compared to the previous example, the mid-rate rule does not change. When no cover transaction is made, $s_j(t)$ is obtained directly from the position constraint; *cf.* eq. (3.15). What does, however, change is the shadow price of the position constraint and, therefore, the limits to the position at which either a cover purchase or a cover sale becomes profitable will also alter. The shadow price $\mu^c_j(t)$ is solved from the condition (iii) above and it is equal to

$$(3.21)\qquad \mu^c_j(t) = \mu_j(t) + [2c_2/(\delta+\eta)]\,[s_j(t) - S(t)].$$

Because both $s_j(t)$ and $\mu_j(t)$ depend on $x_j(t)$ negatively, $\mu^c_j(t)$ declines much more rapidly than $\mu_j(t)$ as $x_j(t)$ increases, implying that the range within which the position is not squared by cover transactions diminishes.

Figure 3.4 illustrates the result in the case where the characteristics of the local market j would imply a price level above the global average; *i.e.*, $\alpha_j > \Sigma_n\, \alpha_n$. It shows the mid-rate rule (*ss*-line) as well as two lines for the shadow price, one for the case where the inconsistency risk is not taken into account ($\mu\mu$-line) and the other for the case when this risk matters ($\mu_d\mu_c$-line). As shown above, the latter line declines more rapidly than the former and the two lines intercept at the point where $s_j(t) = S(t)$. The range $(x^l_j\,,\, x^u_j\,)$ within which a cover transaction is not profitable becomes narrower and is shifted to the right. This implies that the quotations of dealer j are drawn even closer to the global average and their fluctuations become limited still further.

Figure 3.4 Pricing Rule Under Inconsistency Risk



In the limiting case as $c_2 \rightarrow \infty$, the $\mu_d\mu_c$-line becomes vertical, implying that the mid-rate of the local quotation never deviates from the global average. In this case any position resulting from a customer deal would be immediately covered by a transaction in the opposite direction in the interbank market. This would generate income only if the local spread was larger than the average, and even in that case (assuming that there are any customers left) the income would be much smaller than if at least a part of the position could be sold to customers.

Assuming that all dealers behave accordingly, the cost attached to the inconsistency risk will reduce the dispersion of price quotations between local markets, which in itself tends to reduce this risk. If, in addition, all dealers add a precautionary premium to the spread, the frequency distributions of the ask-rate and the bid-rate in *Figure 3.3* will become narrower and the distance between the means will increase. However, as long as the dealers quote prices without full knowledge of how others are quoting at the same moment, and as long as the spread is relatively narrow, the probability that occasional arbitrage profit opportunities will arise remains positive.

4 *Dynamic Extension and the Integration across Time-Zones*

The analysis of *Section 2* shows how the local foreign exchange markets become globally integrated as a result of customer flows and interdealer transactions. A small local transaction shock may be absorbed by the local dealer's position with only minimal subsequent price reactions. On the other hand, a position resulting from a large local transaction shock can always be reversed in the interbank market, whereby its effects will be diffused globally. As the capacity of the interbank market to absorb such shocks is large compared to that of any single dealer, the price reactions across time-zones will remain small.

In the following the analysis is extended to a dynamic framework. For this we use the jump version of the model presented in Section 5. Let the customer orders during a short trading session $(t, t+1)$ be represented by eqs. (3.10) and (3.11). The trading day is composed of a sequence of such periods, $t=0,\dots,T-1$. The dealer j maximizes his daily trading income adjusted for possible cover transactions subject to the constraint that any existing position $x(t)$ has to be reversed during the remainder of the day. The problem is written as follows:

$$(3.22)\qquad \mathbf{Max} \sum_{i=0}^{T-1} \{[\alpha_j(i) + \eta S(i)]s_j(i) + [\beta_j(i) + \eta Z(i)]\, z(i) - (\delta + \eta)[s_j^2(i) + z_j^2(i)] + S^b(i)P_j(i) - S^a(i)Q_j(i)\}$$

$$\text{s.t.}\quad x_j(t) - \sum_{i=t}^{T-1} [\alpha_j(i) + \eta S(i) - (\delta+\eta)s_j(i) + P_j(i) - Q_j(i)] = 0,$$

where the variables and the parameters are the same as defined in Section 2. The choice variables are $s_j(i)$, $z_j(i)$, $P_j(i)$ and $Q_j(i)$, $i = 0,\dots,T-1$.

Following the already familiar procedures, the solutions for the mid-rate and the spread, obtained from the first-order conditions, are

$$(3.23) \qquad s_j(i) = [\alpha_j(i) + \eta S(i)]/2(\delta+\eta) + (1/2)\mu_j(t),$$

$$(3.24) \qquad z_j(i) = [\beta_j(i) + \eta Z(i)]/2(\delta+\eta),$$

where $\mu_j(t)$ is the shadow price at moment t associated with the end-of-day position constraint. Insert eq. (3.23) into the position constraint and solve for $\mu_j(t)$ to obtain

$$(3.25) \qquad \mu_j(t) = \{\sum_{i=t}^{T-1} [\alpha_j(i) + \eta S(i)] + 2\sum_{i=t}^{T-1} [P_j(i) - Q_j(i)] - 2x_j(t)\} /(\delta+\eta)(T-t).$$

The mid-rate and the spread may fluctuate over the day, depending on the sequence of customers' excess demand and the volume of customer trade, as long as these are dealer j's private information. In the case of the mid-rate, the price at moment t depends on the whole sequence of $\alpha_j(i)$, $i>t$, whereas the spread is adjusted only in those trading sessions when the change in the volume of trade $\beta_j(i)$ actually takes place. The mid-rate and the spread also depend on the prices quoted by other dealers, *i.e.*, on $S(i)$ and $Z(i)$, although the sequence of future global average prices can hardly be regarded as dealer j's private information.

The shadow price $\mu_j(t)$ is constant for given information about the day in the present case as well. Although the mid-rate may fluctuate over the day, its fluctuations are limited both because on some occasions it is profitable to adjust the position by undertaking cover transactions in the interbank market and because the customers (including other dealers) will react to price differentials. As above, no cover transaction is profitable as long as $\mu_j(t)$ remains between $S^a(t)$ and $S^b(t)$, *i.e.*, between the average (or the best observed) bid- and ask-prices quoted in the interbank market. These limits together with equation (3.23) determine the limits within which the mid-rate is allowed to fluctuate, that is

$$(3.26) \qquad [s^o_j(t) + S^b(t)]/2 < s_j(t) < [s^o_j(t) + S^a(t)]/2,$$

where $s^o_j(t) = [\alpha_j(t) + \eta S(t)]/(\delta+\eta)$ is the one-period equilibrium mid-rate that would imply zero net sales to customers.

The shadow price - and along with it the mid-rate - will jump immediately each time dealer j receives new information on the customers' net demand during the remainder of the day. Assume, for instance, that at moment t dealer j learns that there will be more sell orders coming from certain customers at some later moment in the day, $i > t$. The shadow price and the mid-rate will jump upwards immediately and will remain at a higher level throughout the remainder of the day. The upward movement in the price will imply more sales to and less purchases from the other customers. Note that the equilibrium one-period mid-rate $s^o_j\ (t)$ does not change, although the quoted price $s_j\ (t)$ will jump. It will not, however, jump above the upper limit indicated in inequality (3.26), because beyond that point it will be advantageous for the dealer to initiate a position in the interbank market by making a cover sale.

An unexpected change in customer orders or in the average interbank quote both affect dealer j's quote, although only after the event. Assume that a quote $(s_j\ (t),\ z_j\ (t))$ has already been made and the deals concluded. If either the net sales to customers, $\alpha_j\ (t)$, or the prices quoted by other dealers, $S(t)$, proved to be higher than expected, the dealer would find his position $x(t+1)$ to be excessively short, thus forcing him to raise the price subsequently and/or to cover the position by a purchase in the interbank market. It is seen that transactions uncertainty (related to the stochasticity of customer orders) and price uncertainty (related to the prices quoted by competitors) both have similar effects.

The above results can be applied to a situation where all operators know that an important piece of news will be announced later in the day. If all participants have similar expectations concerning the contents of the announcement, this effect will already be fully reflected in prices in advance. There will be no price effects when the announcement is made, as long as its contents do not differ from expectations. On the other hand, the price reaction will be prompt, if the announcement contains important surprises.

It is, of course, possible that there are differences in opinion concerning the contents of the forthcoming announcement. If this is the case, some participants will be disappointed at the time of the announcement and observe their positions to move in an unwanted direction, which can be reversed only at a cost. A defensive strategy applicable for such situations would be to widen the spread prior to the announcement. This would reduce the risk that the dealer would receive orders from other dealers who have a different opinion and who may even have some inside information on the contents of the forthcoming news.

We have already seen in *Section 2* how at any particular moment the prices quoted in different local markets tend to be equalized through interdealer transactions and customer flows. This means that the local markets are integrated spatially. The fact that different local markets are located in different time-zones implies that the operating hours of different dealers differ. Some dealers are closing business at the same time as others are opening. In so far as the business hours overlap, the markets in different time-zones become spatially integrated. This tends to draw the quotations in two time-zones close to each other in the same way as spatial integration draws quotations close to each other across the local markets at any particular moment.

In the same way as local transaction shocks are transmitted spatially within a given time-zone with only negligible effects on the global average price, aggregate transactions shocks are transmitted from one time-zone to another. Assume, for example, that the dealers in time-zone *1* have, on the aggregate, overbought a large amount of a given currency late in the afternoon. The aggregate position is therefore excessively long, which would call for a large reduction in prices if the position had to be unwound through customer trade before the end of the day. Assuming that time-zone 2 has already opened, the dealers in time-zone *1* are likely to place interbank orders in the opening zone in order to square their positions, as a result of which the reduction in prices is much smaller or there is none at all. The dealers in the opening zone find their position to be long, but they need to reduce their prices only marginally, because they have plenty of time to steer their position in normal customer trade and in the interbank market during the day. Any regular pattern in the exchange of positions across the time-zones during hours when these overlap is taken into account and is fully reflected in the shadow price.

5 Concluding Remarks

The above analysis contains variations on one single theme which was set out in the previous chapter, *i.e.*, the price making behaviour of a foreign exchange dealer who maximizes his net revenue knowing that customer orders are price-sensitive and recognizing the fact that, in order to remain a market maker, he must aim at balancing purchases and sales of foreign exchange in the longer run, although the long run may be as short as one day. These few behavioural assumptions suffice to generate theoretical predictions which are of empirical relevance and broadly consistent with observable facts.

The results are consistent with the efficient market hypothesis in the sense that there is no possibility that any of the customers could predict the dealer's forthcoming quotations on the basis of the dealer's past pricing behaviour. Under pure transactions uncertainty and in the absence of new information, the dealer is reluctant to make a large adjustment to prices, because frequent price revisions are generally revenue-reducing. The necessity to reverse a position that has arisen unexpectedly forces the dealer to change a price in the 'wrong' direction. This result holds equally for a monopoly dealer and for a dealer operating in a competitive environment.

The reluctance of the dealer to make frequent price adjustments does not apply to situations where new information hits the market. Whenever this happens, the quoted price will react immediately. Any new information on forthcoming customer orders will cause a prompt change in the quoted price. Price adjustments are, however, limited by the sensitivity of customers to small price differentials. Because dealers can send orders to each other in the interbank market, the sensitivity of incoming orders to small price differentials is likely to be very large. As a result, the prices quoted by different dealers cannot deviate much from each other.

The price making power of each dealer is reflected in the spread. In monopolist conditions the spread is wide, because the dealer attempts to extract the full rent which can be reaped because of the price-sensitivity of liquidity-oriented commercial customers. The spread becomes narrower when customers have a possibility to shop around by requesting quotations from more than one market maker, and approaches zero (or the marginal cost of producing dealer services) once interdealer transactions are allowed for. Under pure transactions uncertainty, the spread tends to be constant. This need not be the case if the dealer has private information on customer orders and if he has some monopoly power in price making. In this case the spread widens as the volume of customer trade increases. Customer flows and interdealer transactions, however, greatly reduce the scope for changes in the spread for this reason. The empirical implication is that the spread tends to be large when markets are poorly organized and when there are only few market makers and customers tend to do business with a single bank. These circumstances are likely to promote collusive behaviour. As a result, the share of interbank transactions in the total volume of trade is small. Under competitive conditions, when the spread is very small, prices quoted by all dealers are practically equal and the share of interbank transactions in the total volume of trade is very large.

Price uncertainty is always present in foreign exchange dealing, although in some cases it may be difficult to distinguish between price and transactions uncertainty. Price uncertainty would be small if transactions uncertainty was its only source and if order flows were purely stochastic, apart from the average price sensitivity of customer orders. Any new information on forthcoming customer orders, whether privately or commonly observed, increases price uncertainty. The intraday volatility of exchange rate quotations stems from two sources, one representing normal order flow (customer orders and their cover operations) and the other the arrival of new information. The empirical implication is that the return distribution at ultra-high frequency is likely to exhibit leptokurtosis.

In addition to temporal price uncertainty, instantaneous price uncertainty is present because the dealer has to make a binding two-way price without knowing what prices his competitors are quoting at the same moment. If the price quoted by one dealer happens to deviate from the prices made by others, the position of the former may move in an unwanted direction, which has to be unwound at a loss. This prevents the spread from diminishing indefinitely. In times of increasing uncertainty the spread tends to widen.

If dealers have uniform expectations concerning future events, this information should already be embodied in prices. If dealers have different views, or different information, as regards the contents of a forthcoming news announcement, some dealers will experience losses at the moment when the new information is released. Widening the spread is a defensive action to avoid losses arising from heterogenous and asymmetric information.

Although the time horizon in the dynamic variations of the model is constrained to one trading day, the main results can be generalized to longer horizons as well. The formal analysis applies when allowance is made for open overnight positions, as long as the end-of-day position target is fixed. In the interdependent market the effects of local transactions shocks and arrivals of new information are quickly diffused spatially across dealers through interdealer transactions in a given time-zone. Overlapping time-zones extend spatial integration temporally as positions are exchanged across zones.

Chapter Four

TRANSACTIONS DEMAND FOR INTERNATIONAL MEANS OF PAYMENT

The notion of the transactions demand for foreign exchange dates back to at least the 1950s and to the then ongoing discussion on the so-called dollar shortage. At a relatively late stage of this macroeconomic debate, Yeager (1959) raised an important microeconomic question which is relevant to our general theme regarding the behaviour of foreign exchange dealers. He argued that the whole issue was misplaced and that the debate on the inadequacy of international liquidity had arisen because of inconsistent attempts by central banks to avoid deflationary monetary policies and, at the same time, to keep exchange rates fixed. Yeager's opinion was that under free market conditions, without official pegging of exchange rates, traders and dealers can choose the optimal size of their foreign exchange holdings and can quickly and cheaply transform balances in one currency into balances in another currency simply by change of ownership.

Nine years later, when the dollar shortage problem was already history, Heller (1968) returned to this issue and made an important distinction between the precautionary demand for international reserves by central banks, on the one hand, and the transactions demand for foreign exchange by commercial banks, on the other hand. Central banks hold reserves in order to be able to support the value of their currencies in times of temporary balance of payments deficits, but they do not directly participate in making international payments between the residents of different countries. In reality, almost all international payments are made through commercial banks, and therefore it is their foreign exchange holdings which are the most important from the point of view of the transactions demand for the international means of payments.

Heller, however, did not separate the transactions demand issue from the question of the adequacy of international liquidity. On the contrary, he presented empirical data showing that the ratio of the commercial banks' foreign assets to the value of imports had exhibited a rising trend since 1951, which he interpreted as evidence of the improved adequacy of international reserves. This conclusion was even further strengthened by taking into account the possibility of economies of scale in the handling of foreign exchange transactions by commercial banks.[1] An immediate reaction to Heller's article came from Willet (1969), who argued that Heller's way of measuring the adequacy of international reserves for transactions purposes by the commercial banks' foreign exchange holdings was beside the point. Because commercial banks are optimizing agents and therefore hold their foreign exchange balances at the optimal level, the whole question of the possible inadequacy of their foreign exchange holdings cannot be raised. Being optimal, these balances are by definition neither adequate nor inadequate.

This last-mentioned point raised by Willet has microeconomic significance and forms the starting point for the subsequent analysis. We take into account the fact that practically all international payments are made through commercial banks and concentrate on examining the determinants of optimal foreign exchange holdings by commercial banks. More specifically, our interest is the demand for money for transactions purposes by the foreign exchange dealers of commercial banks. In this respect, the present analysis is complementary to the two preceding essays. Instead of pricing behaviour, the emphasis is now shifted entirely to the costs of producing dealer services.

A major cost arises from inventory holdings, although this aspect was ignored in the preceding chapters. These inventories are non-interest-bearing money balances. An important aspect that distinguishes money traders from other money holders is that the former must have balances denominated in many currencies. This is not because of diversification by a risk-averse dealer in the face of exchange rate uncertainty, but simply because being able to buy and sell on immediate demand requires that the dealer possesses immediately saleable working balances in many currencies.

---

[1] Referring to the square-root rule of Baumol (1952) and Tobin (1956), Heller calculated the ratio of the commercial banks' foreign assets to the square-root of the value of imports. Because this ratio had shown an increasing trend in the 1950s and 1960s, he concluded that the adequacy of international means of payments had improved dramatically during that period.

If the amount of *trading balances* is very small, or if its composition is one-sided, then the probability is high that at some moment he will find himself in a situation where he is short of funds in one currency while having excess balances in some other. In such a situation he has either to borrow more of the sold-out currency or to buy it from other dealers with the overbought currency. We assume that the latter alternative, which Yeager would have called simply "a change of ownership", is cheaper and will always be used provided that other conditions remain unchanged. On the other hand, if trading balances are very large, the probability of becoming short of funds in any of the currencies may be small, but much of the dealer's resources would remain idle and, as a result, the cost in the form of forgone interest income would be high.

Restoring the position by a wholesale purchase from some other dealer is not, in general, profitable. This is because in such a situation the dealer operates on the 'wrong side' of the market. He has to pay more for the currency purchased than he would pay if he could buy the same amount from customers. Therefore, we can conclude that the dealer wants to avoid wholesale transactions, but this gain has to be weighed against the higher opportunity cost of holding non-interest-bearing balances. Optimization between these costs determines the transactions demand for currencies.

Once the problem has been formulated as above, it becomes apparent that the transactions demand for a multitude of monies by a foreign exchange dealer does not in any essential way differ from the transactions demand for money in general. Accordingly, the formal models developed for the analysis of transactions demand for money should be directly applicable.

1 Simple Transactions Uncertainty Model

Consider a single foreign exchange dealer who has invested a given amount of resources in non-interest-bearing demand deposits with commercial banks both at home and abroad. The size of these balances measured in domestic currency is

$$H = sD^* + D \,, \tag{4.1}$$

where D^* denotes the balances in foreign currency (*dollars*) and D those in domestic currency (*marks*), and where s is the conversion rate for translating foreign currency into domestic-currency terms.[2] The opportunity cost of holding the trading balances is rH, where r is the rate of interest. It represents the opportunity cost of holding resources in non-interest-bearing form, or the cost of debt if the dealer operates with borrowed capital.

In order to cover the operating costs as well as the costs of holding resources in non-interest-bearing form, the dealer applies a positive spread to the transactions he makes with customers. A lower price is applied to customer sales (dealer purchases) and a higher price to customer purchases (dealer sales). The former, the *bid-rate*, is denoted by s^b, and the latter, the *ask-rate*, by s^a, with $s^a > s^b$. We assume that the dealer operates in a competitive environment and, for fear of losing customers on one or the other side of the market, has to accept the quotation given by the market. In addition to the spread, the dealer's trading income depends on the volume of two-way trade.

As regards transactions uncertainty, let us assume that the sequence of customer (retail) transactions follows a simple symmetric *Bernoulli process* (*cf.* Miller and Orr, 1966). In each fraction of a day, *a minute,* the dealer either sells m dollars with a probability of $\pi=1/2$ or buys m dollars with a probability of $1-\pi=1/2$. The expected change in the dollar balances over N days is $\mu_N=0$ and its variance is $\sigma_N^2=Nm^2t$, where t denotes the number of minutes in the day. The variance becomes infinite as N grows indefinitely, which means that, given a finite amount of initial money balances, either dollars or marks are sold off with a probability of one at some time in the future. When dollars are sold out, the dealer can immediately exchange marks for dollars in the wholesale market. Similarly, when marks are sold out, he can immediately exchange dollars for marks in the market.

Given these assumptions, the dealer's expected trading income per day is exogenously determined and is equal to $(s^a-s^b)q$, where $q=mt/2$ is the average retail trade turnover in a day (average of customer sales

---

[2] It is the foreign exchange dealers of commercial banks who are directly responsible for their foreign exchange operations, and as the banks' treasury departments typically operate as profit centres, we can for analytical purposes separate this activity from the other activities of commercial banks and regard dealers as if they were private profit-maximizing agents.

and purchases). The costs incurred are the forgone interest income from holding an amount H in non-interest-bearing form, the expected cost of wholesale transactions and the possible costs related to retail transactions. The latter are assumed to depend on the average volume of autonomous transactions and hence to be independent of the dealer's decisions.[3] The first item mentioned is determined once H has been chosen, while the expected cost of wholesale transactions depends on the cost and the expected frequency of such induced transactions. Hence the genuine decision problem of the dealer is reduced to the choice of the amount of the trading balances and the choice of the composition of these balances after each wholesale transaction.

Let us assume that the dealer has a target composition for the trading balances; for instance, he wants to have a proportion K/H in dollars and a proportion $(H\text{-}K)/H$ in marks after a wholesale transaction. Hence, when dollars are sold out, the dealer will convert an amount K of marks into dollars and when marks are sold out, he will sell an amount $(H\text{-}K)/s$ of dollars in exchange for marks; *cf. Figure 4.1.*

When H and K are given and when the process starts from the target composition, then the expected value of the time duration (measured in numbers of steps) between two successive wholesale transactions is (*cf.* Miller and Orr, 1966, p. 422)

$$d = (K/sm)[(H\text{-}K)/sm]. \tag{4.2}$$

The inverse of this multiplied by the number of minutes in a day gives the expected number of wholesale transactions per day[4]

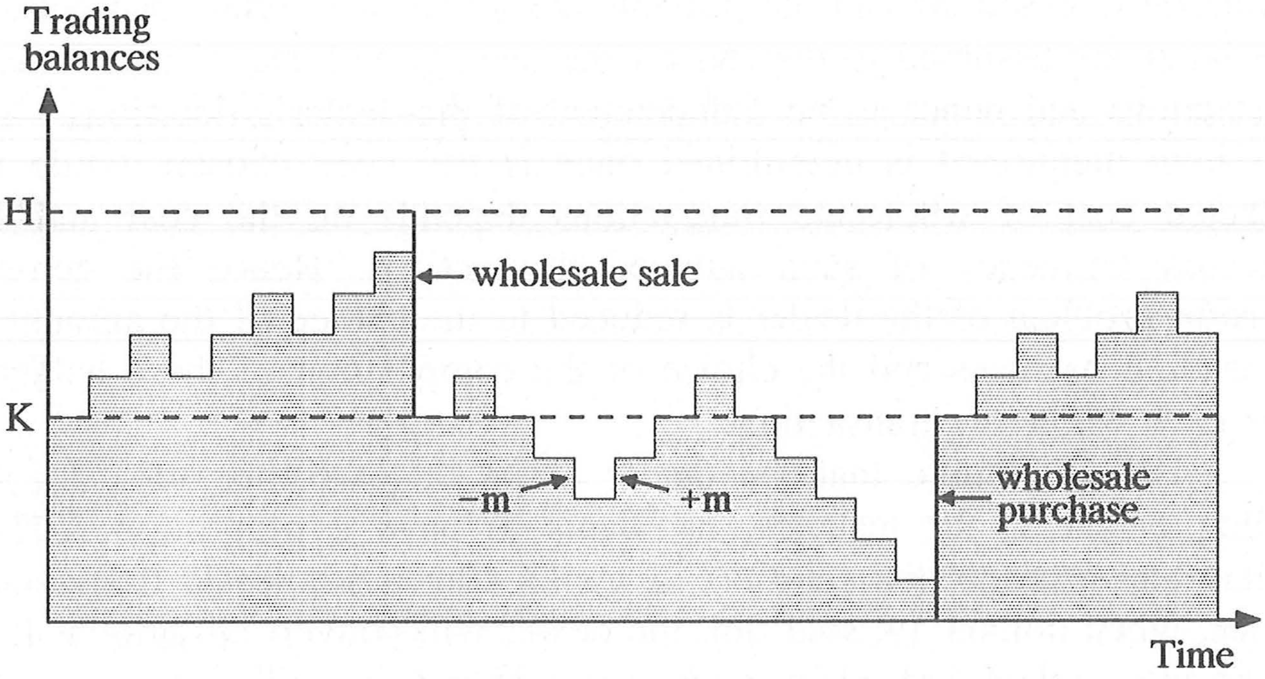
$$n = [(sm)^2 t]/[K(H\text{-}K)]. \tag{4.3}$$

---

3 The transaction costs for retail transactions in the form of paperwork may be considerable, but they are to large extent independent of the amount transacted. This cost may depend on the degree of foreign exchange control.

4 Note that the expected value of $1/x$, where x is a random variable, is not exactly the inverse of the expected value of x, even though it is a good approximation. For this reason, we call N the expected number of transactions and not the mean number of transactions.

Figure 4.1 Transactions Uncertainty and Wholesale Transactions



Multiplying this by the cost of a wholesale transaction c gives the expected cost of the wholesale transactions in a day. For the time being, it is assumed that this cost is a brokerage fee and is independent of the size of the wholesale transaction. The opportunity cost of holding a trading portfolio of the size H is rH, where r is the daily rate of interest.

The dealer's objective is to maximize the return on his capital. In this case the expected trading income and the transaction costs from retail sales and purchases do not depend on H and K, and therefore the dealer can fulfill his objective by minimizing the following expected cost function with a suitable choice of H and K:

(4.4) $$C = rH + c(sm)^2 t/[K(H-K)].$$

Differentiation of (4.4) with respect to K and H gives the first-order conditions

(4.5) $$\partial C/\partial K = -\, c(sm)^2 t(H-2K)/[K^2(H-K)^2] = 0$$

(4.6) $$\partial C/\partial H = r - c(sm)^2 t/[K(H-K)^2] = 0.$$

The first condition gives $K=H/2$, which implies that the decision on the composition of trading balances is separable from the decision concerning the amount of total currency holdings. With $K=H/2$, both $\partial^2C/\partial H^2$ and $\partial^2C/\partial K^2$ are positive, implying that (4.5) and (4.6) give the minimum cost. Inserting $K=H/2$ into the second condition and solving for H gives the optimal amount of the trading balances:

$$H = 2[c(sm)^2t/r]^{1/3}. \tag{4.7}$$

It was shown by Miller and Orr (1966, pp. 422-423) that if x, generated by a symmetric Bernoulli process, wanders within two barriers, 0 and H, and is always returned to K after a passage at either of the barriers, the steady state distribution of x is of discrete triangular form with base H and mode K. The mean of such a distribution is $(H+K)/3$. It follows that with $K=H/2$ the mean balances of both the domestic and foreign currency are therefore equal to $H/2$. This can be interpreted to represent the dealer's (stock) demand for the currencies, $D=sD^*=H/2$. If the volume of trade expands, or if the interest rate declines, the demand for both currencies increases in equal proportion, implying that the two currencies are complements.

As seen from equation (4.7), the transactions demand for trading currencies is based on the cubic-root rule, which is similar to that derived by Miller and Orr for the demand for cash by a business firm.[5] Also, the implications are similar. For example, the average amounts of trading balances are homogeneous of degree one in nominal variables; *i.e.*, if all prices measured in domestic currency rise by one per cent, s and c will rise by the same amount, as will H. The amount of trading balances depends negatively on the rate of interest with an elasticity of -1/3. Similarly, the result implies that there are likely to be significant economies of scale in foreign exchange dealing in the sense that optimal trading balances rise less than proportionately as the demand for dealer services (average volume of customer trade) increases. Note that the average volume of trade is $mt/2$, which can increase if the average size

[5] The major difference is that the target level in the present case turns out to be one-half of the upper threshold, whereas in Miller and Orr (1966, p. 423) it was one-third. The difference is due to the fact that here the opportunity cost covers total trading balances, which are also equal to the upper threshold. In Miller and Orr the opportunity cost concerned money balances, that is one-third of the upper threshold on the average.

of a retail transaction m increases or if the arrival of customer orders t becomes more frequent; *i.e.* the process intensity increases. The economies-of-scale result holds as long as m and t are proportionally related to each other. Thus the claim by the early authors that there are significant economies of scale in the commercial banks' foreign exchange holdings receives support from this result.

2 *Proportional Transaction Cost*

It is well known in the literature on the transactions demand for money that the economies-of-scale result as well as the one-target result of Miller and Orr are, in fact, special cases and depend on the assumption of a fixed lump-sum cost of a transfer between cash balances and interest-bearing assets. The general rule, which allows for both a fixed and a proportional cost, leads to a two-target/two-threshold policy. The targets and the thresholds are all obtained simultaneously as a solution to a dynamic programming problem (*cf.* Milbourne, 1983, and Penttinen, 1983). If the transfer cost is strictly proportional, the solution is of a two-threshold form, implying that only small adjustments are made to cash balances by always returning to either the upper or the lower thre-s-hold (Eppen and Fama, 1969). Before moving on to such a general problem, we maintain the assumption of a simple one-target policy and analyze a case where the transaction cost is assumed to be proportional.

A fixed cost associated with a wholesale transaction may be an appropriate assumption in the case of a dealer who always uses a broker in undoing a position. Brokers are intermediaries who bring buyers and sellers together and are paid commissions for their efforts. The commissions are fixed for relatively large amounts compared with the average size of a retail transaction. On the other hand, if the dealer typically adjusts his position by trading with other dealers, then an assumption of a proportional cost appears more appropriate.

Recall that when entering the market as a customer the dealer operates on the wrong side and therefore pays a higher price for a purchase than he would pay if he were buying the same amount as a market maker from commercial customers. Here we assume that the rates are consistent in the sense that the dealer's bid-rate is lower than the ask-rate quoted by others; *cf. Chapter Three*. A wholesale sale represents a cash inflow, whereas a wholesale purchase represents a cash outflow in domestic currency. Let us incorporate these observations into the formal model.

We continue assuming a symmetric Bernoulli process as a description of transactions uncertainty. In addition, we maintain the assumption of a fixed target composition to which trading balances are returned after each wholesale transaction. From equation (4.3) it is seen that for a given H the lower is the target composition K/H, the more frequent are wholesale purchases while wholesale sales are accordingly less frequent. Similarly, the expected number of wholesale sales increases and that of purchases decreases, as K approaches the upper threshold H. In the above example, the target level $K=H/2$ minimized the expected number and therefore the expected cost of wholesale transactions, because a fixed brokerage fee was associated with each such deal independently of its size or direction. This need no longer be the case when there is a two-way price available at which currencies can be bought or sold in the interbank market.

For the symmetric Bernoulli process starting from K, the probability that the first passage is at zero (wholesale purchase) is $(H-K)/H$, and the probability that the first passage is at H (wholesale sale) is K/H (*cf.* Miller and Orr, 1966, p. 423). The expected number of passages is n, and each time it occurs the process restarts from K. The expected number of wholesale purchases is therefore $n(H-K)/H$ and that of wholesale sales is nK/H. The size of the purchase is K/s dollars and the price offered in the market is S^a. The size of a wholesale sale is $(H-K)/s$ dollars and the price bid by the market is S^b, $S^b<S^a$. The expected net cash outflow arising from wholesale transactions is

$$(4.8) \qquad C_w = n\{[(H-K)/H](K/s)S^a - (K/H)[(H-K)/s]S^b\}$$

$$= (S^a-S^b)sm^2t/H,$$

where n is taken from equation (4.3). It is readily seen that the cost of wholesale transactions over a longer period is independent of K.

The total cost of holding trading balances by an amount H is C_w+rH. Minimization with respect to H yields

$$H = [(S^a - S^b) s m^2 t / r]^{1/2}, \tag{4.9}$$

which resembles the square-root rule of Baumol (1952) and Tobin(1956). Trading balances depend negatively on the rate of interest with an elasticity of -1/2 and positively on the degree of transactions uncertainty m^2t with an elasticity of 1/2. The size of the trading portfolio is proportional to the average size of a customer order m. Thus, the economies-of-scale result is no longer as pronounced as above. It holds only for the frequency t of the arrivals of customer orders. Note that H is homogenous of degree one with respect to the exchange rate. In other words, if S^a, S^b and s all rise by one per cent, the desired trading balances increase by the same percentage.

The spread applied in the wholesale market appears under the square root. Using the terminology introduced by Demsetz (1968) it can be interpreted to represent the cost of transacting; *cf.* also Hirschleifer (1973) and Levich (1979, pp. 7-10). The interpretation follows from the fact that, if in an organized market it is possible to undo a purchase immediately, it requires a sale at a lower price. Because two transactions are involved, the cost per transaction is one-half of the difference between the two prices, the half-spread in the present case. It can readily be seen from (4.9) that trading balances would not be needed at all if S^a-$S^b \rightarrow 0$; *i.e.*, when the cost of transacting becomes equal to zero. If this were the case, each retail transaction would be immediately covered in the interbank market.

The most interesting result, however, is the fact that the target level is not determined by the model, although it was still assumed that there is one fixed-target which will be maintained over some longer period. This has a perfectly intuitive explanation because any wholesale transaction implies that the dealer has already accumulated net sales or purchases from customers by an equivalent amount. Because of the positive bid-ask spread and the 'wrong' prices applied to wholesale transactions, the latter are generally revenue-reducing.

As an illustration, let us calculate the cash flow arising from both retail and wholesale transactions during a particular short period. Assume that starting from K a wholesale purchase becomes necessary after d minutes. There have been d retail transactions altogether, of which x sales and y purchases, $x+y=d$. The size of each retail transaction is m

dollars and the net sales are equal to $K/s=m(x-y)$ dollars. These two equations imply that the dollar sales mx to customers are equal to $(1/2)\times(dm+K/s)$ and the dollar purchases my are accordingly $(1/2)(dm-K/s)$. The cash flow generated by this sequence of retail transactions is hence equal to $(K/s)(s^a+s^b)/2+dm(s^a-s^b)/2$, where s^b and s^a are the dealer's bid-rate and ask-rate, respectively. The first term is equal to the net sales of dollars valued at the mid-rate and the second term is equal to the spread times the average volume of trade over a period of length d. Although trading income is greater than the average revenue for a period of that duration, the net revenue after the wholesale transaction of size K/s is less, assuming that the dealer's mid-rate is below the interbank market's ask-rate, $S^a-(s^a+s^b)/2>0$.

In a similar manner it can be shown that starting from K and hitting the upper threshold H in a sequence of d retail transactions implies sales of dollars by an amount $(1/2)[dm-(H-K)/s]$ and purchases by an amount $(1/2)[dm+(H-K)/s]$. Trading income is equal to $-[(H-K)/s]\times(s^a+s^b)/2+dm(s^a-s^b)/2$. Selling an amount $(H-K)/s$ in the market at price S^b reduces net income, assuming that the bid-rate in the interbank market is below the dealer's mid-rate, $S^b-(s^a+s^b)/2<0$.

These examples are valid for any sequence of retail transactions that *ex post* has led to a wholesale transaction. Starting at K, the outcome is, of course, not known. The expected trading income over an average duration d adjusted for the expected cost of an wholesale transaction can, however, be expressed as a function of H and K, noting that the probability of the first passage being at zero is $(H-K)/H$ and the probability of the first passage being at H is K/H. It is

$$(4.10)\qquad \begin{aligned} R_d &= [(H-K)/H][(K/s)(s^a+s^b)/2+dm(s^a-s^b)/2 - S^a(K/s)] \\ &\quad + (K/H)\{[-(H-K)/s](s^a+s^b)/2+dm(s^a-s^b)/2+ S^b[(H-K)/s)]\} \\ &= dm(s^a-s^b)/2 - (S^a-S^b)[K(H-K)/Hs]. \end{aligned}$$

Over the average duration, trading income from customer trade is always equal to the average income, or the income that would be generated if sales and purchases happened to be in balance. Wholesale transactions always reduce expected net income. Multiplying (4.10) by t/d and using equation (4.2) for d yields the average trading income per day,

$$R_{day} = mt(s^a\text{-}s^b)/2 - C_w, \tag{4.11}$$

which is independent of K. Although the expected one-duration cost arising from wholesale transactions depends on K, over a longer period K does not matter as long as it remains constant.

Although the dealer may transact in the interbank market and take prices from there, there may still be a fixed cost associated with each such deal. It need not be an explicit broker commission but rather a cost arising from acquiring information on prices as well as from handling the transaction.

Combining the two cost elements in the model leads to the following cost function C_w+c_on+rH, where c_o is the fixed cost of transaction. Because C_w does not depend on K, differentiation with respect to K leads back to equation (4.5), implying a target composition $K=H/2$. In other words, for a given H the expected number of wholesale transactions will again be minimized. Differentiating with respect to H and taking into account that $K=H/2$ yields the following first-order condition:

$$rH^3 - (S^a\text{-}S^b)sm^2tH - 8c_o(sm)^2t = 0. \tag{4.12}$$

It has only one positive root corresponding to that level of $\hat{H}$, which also gives the minimum cost. It is above the optimal trading balance obtained in (4.9), which is obvious because the fixed transaction cost in addition to the proportional one reduces the desired number of wholesale transactions, implying a need for larger trading balances.[6]

Note that (4.12) reduces to (4.7) when $S^a\text{-}S^b$ approaches zero. In that case the desired level of trading balances would be smaller and the economies of scale again more pronounced than in the case where the dealer faces a positive spread when operating in the wholesale market. The latter observation is interesting because it indicates that investing in

---

[6] Formally this is seen from (4.12) by noting that $(S^a\text{-}S^b)sm^2t/r=H^2$, where H is the optimal trading balance under the proportional transaction cost as obtained in equation (4.9). Because $\hat{H}^3 - \hat{H}^2H>0$, H must be smaller than $\hat{H}$.

information (e.g. by making more calls to other dealers), *i.e.*, by accepting greater c_o, may pay off in the form of better prices for wholesale transactions on the average.

3 *Two Return Points and Two Thresholds*

In all of the above examples it was *assumed* that there is only one return point to which the composition of the trading portfolio is returned after each wholesale transaction regardless of its direction. In the presence of a fixed cost for a wholesale transaction the return point is in the middle at $K=H/2$, whereas in the case of a proportional cost it is indeterminate, which means that the target level has to be chosen on grounds other than those related to economizing transactions costs. In the case of a foreign exchange dealer such a target could be justified by risk considerations, for example, by assuming that the position will be closed after each wholesale transaction. For such a decision $K=H/2$ is, in fact, a good choice, as will be shown later on.

However, as mentioned above, the one-target policy is not generally optimal for cash management purposes under transactions uncertainty. The same conclusion can be assumed to also hold for the demand for trading balances by foreign exchange dealers. Especially in the case of a strictly proportional transaction cost, it is quite conceivable that instead of undoing earlier net sales to customers by a single wholesale purchase the dealer decides to wait and see whether, by chance, the position will be restored autonomously by future customer orders. If this happens, then it pays off because the prices are better in customer trade than in dealing on the wrong side of the market. If it does not happen then it does not matter whether the wholesale purchases are made frequently in small amounts or less frequently in large amounts, because the purchases are nonetheless determined only by the net customer orders.

It is shown in the *Appendix* that this is, indeed, the case. There it is assumed that whenever the process hits the lower threshold (dollars are sold off) the position is returned to $k_1=K_1/H$ by a wholesale purchase of size K_1/s dollars. Similarly, whenever the passage is at the upper threshold H, a position $k_2=K_2/H$ is restored by a wholesale sale of $(H-K_2)/s$ dollars. Assuming that the relative frequencies of the two types of wholesale transactions correspond to the steady state probabilities of their occurrence, one can calculate the average number of such transactions as well as their net effect on the dealer's cash flow. This procedure allows us to derive the cost-minimizing return points and the size of the trading portfolio in a way that makes the results directly comparable to those of the previous examples.

In the absence of a proportional cost, the return points become identical at $k_1=k_2=1/2$, which corresponds to our first example; *cf.* equation (4.7). In the case of a strictly proportional cost, the results change rather significantly as regards the return-points but not as regards the other conclusions. The return points turn out to be as close to their respective thresholds as possible; *i.e.*, $k_1=sm/H$ and $k_2=(H-sm)/H$. This implies that the desired size of the trading portfolio is smaller than in the one-target case. The saving in holding costs is quite significant (about 30 per cent). The other properties of the demand function are maintained.

Combining a fixed cost and a proportional cost in the same model leads to an intermediate case where the lower return point is somewhere between zero and $H/2$ and the upper return point somewhere between $H/2$ and H. An increase in either of the costs, c_o or S^a-S^b, increases the size of the desired trading portfolio. These results are consistent with those of Milbourne (1983), although they are derived in a different way.

Allowing for two return-points does not change any of the qualitative results mentioned above. In particular, the economies of scale remain significant, at least as long as there is an important fixed cost associated with a wholesale transaction. They are more significant in an active market where the customer orders arrive at a more frequent rate than in an inactive market. The average size of a customer order is not important to scale economies, except in the case of a fixed transaction cost. In each case, trading balances are sensitive to the rate of interest.

4 Currency Substitution and Price Uncertainty

According to the currency substitution approach that acquired popularity in the late 1970s, domestic and foreign currencies are substitutes and their demand depends, *inter alia*, on the expected change in exchange rates (*e.g.*, Calvo and Rodriquez, 1977, Bilson, 1978, and Miles, 1978). In particular, the negative dependency of the demand for domestic money balances on the expected rate of depreciation is emphasized, which may have powerful implications as regards the autonomy of domestic monetary policy.

The advocates of the currency substitution approach have not been very specific either as regards the motives for holding an internationally diversified currency portfolio or as regards the definition of currencies. In general, it would appear that they are willing to define the currency portfolio broadly and include in it both non-interest-bearing demand deposits and interest-bearing time deposits. The motives for holding an internationally diversified currency portfolio seem to be the same as the motives for holding domestic currency in the standard money demand analysis. Thus the transactions demand has not been forgotten. For instance, Miles (1978, p. 428) writes: "anyone who consistently makes purchases from foreign countries has at least the same transactions motive for demanding foreign currency balances as for demanding domestic currency balances. Importers and exporters, businessmen who travel abroad, tourists, and residents of border areas all have incentives to diversify their currency balances."

Interestingly enough, none of these authors has mentioned commercial banks or foreign exchange dealers as being agents who may have a motive to hold liquid resources in a number of different currencies. Apparently, this is due to the old tradition in monetary theory of regarding only the money balances held by the nonbank public as money. Thus, because foreign exchange held by commercial banks appears as a liability of commercial banks in other countries, these balances cancel out when all commercial banks of the world are aggregated together. A notable exception, however, is McKinnon (1979, p. 182), who in his book on international monetary exchange separates foreign exchange dealers from the domestic nonbank public as holders of domestic money and introduces the currency substitution aspect by stating that, in the face of an expected depreciation, the dealers want to hold lower domestic currency balances "in order to raise their marginal product in clearing international payments".

Against this background, it is interesting to note that in the light of the models analyzed above the two currencies appear, at least at first glance, to be complements rather than substitutes. This follows from our strict concentration on the transactions motive and our definition of the dealer's trading balances as non-interest-bearing assets, the return on which is realized in the form of trading income, rather than in the form of speculative capital gains. Both currencies (or a large number of currencies in a multicurrency context) are needed for the dealer to be able to buy and sell on immediate demand.

The reason why the whole issue of currency substitution does not arise in the above models is that the holding cost has consistently been assumed to be independent of the composition of the trading portfolio. The assumption is justified if the interest parity holds exactly, *i.e.*, if the differential between domestic and foreign interest rates is equal to the expected rate of depreciation of the domestic currency. In this case, the decision on the allocation in an interest-bearing portfolio between domestic and foreign assets, or debt if the dealer operates with borrowed capital, is separable from the decision on the average composition of the trading portfolio. In order to destroy this separability, we need some assumptions that link together the composition of the trading portfolio and the way in which this capital is financed.

Assume that the dealer has financed his trading balances by borrowing an amount L^* in foreign currency and an amount L in domestic currency at daily interest rates r^* and r, respectively. The foreign currency is initially placed as non-interest-bearing demand deposits (*nostro accounts*) with a foreign bank (*correspondent bank*). Similarly, the domestic currency is initially placed as demand deposits in a domestic bank.[7] These amounts are the resources with which the dealer has to operate. Initially, the dealer's trading balances in foreign currency D^* exactly match the foreign-currency denominated liabilities, implying a closed position. Although the dealer's horizon may be very short, say, one day, we assume that the decision upon the amounts borrowed is made in advance. It follows that the liabilities are fixed for a given period, whereas the composition of trading assets is constantly fluctuating as a result of transactions uncertainty.

---

[7] This can, of course, be an amount allotted to the dealer by his own bank. This means that the bank makes a decision on the liquidity needs arising from foreign exchange dealing activity and calculates its price on the basis of the opportunity cost, *i.e.*. the market rate of interest.

Assume further that in order to safeguard the balance sheet against unexpected changes in the exchange rate during the day, the dealer restores a closed position after each wholesale transaction. When the position is closed the amount of dollars in the trading portfolio K/s is equal to the dollar-denominated debt L^*.

Let s=1 initially and let $\hat{e}$ stand for the rate of depreciation of the domestic currency (appreciation of the dollar) expected by the dealer. This may or may not be equal to the interest rate differential r-r^*, for which reason the expected holding cost may depend on the liability structure. Taking the differential holding cost into account and assuming a fixed cost c for a wholesale transaction leads to the following cost function:

(4.13) $$C = cn + (r^* + \hat{e})K + r(H-K),$$

where n is defined in equation (4.3), H is the size of the trading portfolio $L+L^*$ and K is the return level; *i.e.*, the domestic currency equivalent of dollars after each wholesale transaction. Differentiating with respect to K and H and setting the partial derivatives equal to zero yields:

(4.14) $$-cm^2t(H-2K)/[K^2(H-K)^2] + (r^* + \hat{e} - r) = 0$$

(4.15) $$-cm^2t/[K(H-K)^2] + r = 0.$$

Solve first equation (4.15) for $(H-K)^2 = (cm^2t)/rK$ and insert this into (4.14) to obtain the ratio

(4.16) $$K/H = r/(r + r^* + \hat{e}).$$

Either of the two conditions can be used to solve for H,

(4.17) $$H = (r + r^* + \hat{e})[r(r^* + \hat{e})]^{-2/3} (cm^2t)^{1/3}.$$

It can immediately be seen that whenever the interest rate parity holds r-r^*=$\hat{e}$, the return-point K/H=1/2, and the size of trading balances is equal to that obtained in the simple transactions uncertainty model of *Section 1*; *cf.* eq. (4.7). In that case the holding cost is the same for both currencies. The return-point approaches zero as the expected rate of depreciation of the domestic currency increases without being fully reflected in the interest rate differential. Looking at the situation on the liability side, this means that less dollars are borrowed because the cost of borrowing is higher. Similarly, if the domestic currency is expected to appreciate, the return-point $K/H>1/2$ and approaches the upper threshold $K/H=1$ as $\hat{e}\to -r^*$. The expected appreciation of the domestic currency makes borrowing in the foreign currency relatively cheap and therefore more of it is borrowed.

Multiplying H by K/H gives the dealer's demand for dollar-denominated debt

$$(4.18) \qquad L^* = [rcm^2t/(r^*+\hat{e})^2]^{1/3}.$$

Similarly, the dealer's demand for mark-denominated debt is obtained by subtracting L^* from H, which yields

$$(4.19) \qquad L = [(r^*+\hat{e})cm^2t/r^2]^{1/3}.$$

For both currencies the elasticity is -2/3 with respect to the own cost of borrowing in their own currency and +1/3 with respect to the cost of borrowing in the other currency. These equations also determine the size of the planned wholesale transactions, $L^*=K$ for a sale and $L=H-K$ for a purchase.

Recall that in the case of a symmetric Bernoulli process the mean dollar balances are equal to $(H+K)/3$. This is equal to K only when $K=H/2$; *i.e.* when the interest rate parity holds exactly. When $r-r^*\neq\hat{e}$, the average dollar holdings on the asset side differ from the dollar-denominated debt, which means that the foreign exchange position is open on the average. As an illustration, assume that the dealer expects the depreciation of the domestic currency to exceed the amount reflected in the interest rate differential, $\hat{e}>r-r^*>0$. The return-point will be set below $H/2$, implying relatively more frequent wholesale purchases than sales of the appreciating foreign currency. Although a closed position is restored after each wholesale transaction, the average dollar balances will exceed the dollar-denominated debt, implying a long position in the the appreciating foreign currency. In a similar way, if the foreign currency is expected to depreciate, the position will be short on the average.

The answer to our question of whether or not there is currency substitution in foreign exchange dealing is thus affirmative. The conclusion applies to the net position in foreign currency and is valid only in situations when the dealer takes a view that is different from that of the market expectation reflected in the interest rate differential. If all dealers share the same expectation and behave accordingly, the initial conditions must change in a way that restores the interest rate parity. Either s has to jump immediately to $1+\hat{e}$ or interest rates have to move to restore the parity.

The above analysis has been carried out without any explicit reference to exchange rate uncertainty. It was mentioned only in connection with the definition of the closed position at return-point K. The expected rate of depreciation as well as its deviation from the interest rate parity were treated as if they were certain. Normally, a certainty-equivalence assumption of this kind would lead to a corner solution, according to which all assets are held in the form which yields the highest return or all debt is held in the form which carries the lowest interest. In the above example, the liability diversification resulted from the optimization between the holding cost and the trading income adjusted for the cost of wholesale transactions. The corner solutions were obtained only in two extreme cases where either the expected rate of depreciation of the domestic currency was infinitely rapid or its expected rate of appreciation approached the foreign rate of interest.

The question of how exchange rate uncertainty affects the size of the trading portfolio and its composition cannot be unambiguously answered in terms of the models discussed so far. As regards the within-the-day variablity of exchange rates, there is no obvious way how this source of uncertainty can be separated from the transactions uncertainty which arises from the stochasticity of customer orders. Any change in the exchange rate, when the position happens to be open, will bring a capital gain or loss momentarily, but over time capital gains should compensate for losses.

Assuming, as above, that the trading assets are backed by liabilities denominated in the two currencies, a possible capital gain or loss is realized at the end of the short planning horizon, when the decision on new borrowing is made. Assume that $\mathbf{E}(\hat{e})=0$ and $\mathbf{Var}(\hat{e})=\sigma^2$. The expected position is $\hat{e}[(H+K)/3\text{-}K=0$. Although the expected capital gain is zero, its variance is positive, as long as the exchange rate is uncertain. Risk-aversion can now be introduced by assuming that instead of minimizing the expected cost of producing dealer services the dealer maximizes the expected utility of the daily revenue; *i.e.* the expected revenue adjusted for the daily exchange rate uncertainty. The expected utility is written as follows:

$$\text{(4.20)} \qquad \mathbf{E}(U) = (s^a\text{-}s^b)mt/2 - C_w - rH - \eta\sigma^2[(H\text{-}2K)/3]^2,$$

where the first term is the expected trading income from customer deals, C_w is the expected cost of wholesale transactions, rH is the holding cost and the last term is equal to the risk-aversion coefficient times the variance of the expected capital gain.

Introducing the costs arising from an open position in the manner described above does not affect any of the earlier results qualitatively, except in the case of a proportional transaction cost for a wholesale transaction. In that case the return-level is no longer indeterminate. When interest rate parity holds it is equal to $H/2$.

5 *Additional Implications*

The above results are based on partial equilibrium analysis, for which reason any implications concerning aggregate outcomes have to be interpreted with great caution. The underlying aggregate assumption as regards the market structure has been that of a competitive market. Under competitive conditions, the trading income of any one dealer adjusted for the wholesale transactions should be equal to the marginal cost of producing dealer services. This cost increases with the rate of interest, implying that a higher interest rate should affect the spread positively.

On the other hand, the economies of scale in foreign exchange dealing imply that the marginal cost of producing dealer services declines with the volume of trade. As a consequence, the spread should be narrower the larger is the volume of trade. This conforms to certain observable facts. For example, the spread tends to be narrower in quotations for the major currencies than in those for the minor ones (Levich, 1979, pp. 25-627, McKinnon, 1979, pp. 11 and 21-26, and Glassman, 1987). In addition, the spread is generally smaller in wholesale trade than in retail trade, where the average size of a transaction is smaller.

The economies-of-scale result, however, is in disagreement with the assumption of a competitive market. Without further assumptions, it would imply a tendency towards natural monopoly. Because all dealers are not identical as regards their share of global customer trade, then the dealer whose share in the market is largest should be able to systematically apply a narrower spread than his competitors, thus increasing his market share still further. While no formal answer to this question is possible within the framework of the present study, a number of factors can be mentioned which are likely to work against such a tendency.

First, extending the time horizon would bring other cost elements into the picture for which the cost function may, after some point, well be an increasing function of the volume of trade. Secondly, covering a position in the wholesale market may become increasingly difficult if the size of a wholesale transaction increases beyond some limit, because there may simply be no counterparty who would be willing to buy that position. Thirdly, the technical assumption that the quotations are the same everywhere need not be taken literally. Indeed, as emphasized in *Chapter Three*, the foreign exchange market can be characterized as highly competitive, even if some price-making power is left with the dealers. The spreads can differ from one local market to another, as long as the non-dealer customers are not infinitely sensitive to small quotation differentials between the markets and as long as the mid-rates are sufficiently close to each other so that arbitrage profit opportunities do not arise, except occasionally. Finally, the inherent risk associated with dealing activity is likely to reduce the scope for economies of scale. For example, the bank's management may set limits on the open position each individual dealer within the bank is allowed to take.[8]

---

[8] It is not uncommon in international bank's that the bank management sets limits on the open positions their dealing departments are allowed to take. Even within the dealing department, the chief dealer controls the amounts overbought or oversold by each individual trader.

This suggests that, instead of natural monopoly, the structure of the foreign exchange market is likely to attain a form where there are a few financial centres with a relatively small number of international banks servicing primarily customer banks, which are scattered all over the world and trade with their local non-dealer customers. Although the inter-dealer transactions serve the function of a wholesale market, especially for smaller banks, the fact that it sometimes may be difficult to find counterparties for large wholesale deals explains why there is room for brokers to step in and intermediate between wholesale buyers and sellers.

It was mentioned in *Chapter One* that more than 90 per cent of the total USD 700 billion daily turnover in the foreign exchange market represents interbank transactions. In the above examples the ratio of wholesale transactions to the total volume of trade is much smaller. Take, for instance, the case with a fixed transaction cost and let $s=1$. The gross volume of autonomous customer trade is mt, while the gross amount of induced wholesale transactions is equal to $nH/2$. The ratio of the latter to the former is $2[r/(S^a\text{-}S^b)t]^{1/2}$, where n and H are given by equations (4.3) and (4.9), respectively. With a reasonable assumption of 0.1 basis points for the relative spread and 0.025 basis points for the daily interest rate (corresponding to compounded 9.5 per cent per annum), the ratio becomes equal to $1/\sqrt{t}$. It is 0.14 for t=50, 0.10 for t=100 and 0.045 for t=500. With these figures, the induced interbank transactions $m\sqrt{t}$ would represent only 5 to 12 per cent of total trade (autonomous customer deals and induced wholesale transactions together). With a relative spread of 0.05 basis points, these proportions would rise to the range of 9 to 17 per cent.

These figures, however, underestimate the relative size of the interbank market, because the dealer himself may receive orders from other dealers. Let x, y, and z stand for autonomous non-bank customer orders, autonomous interbank orders and induced interbank orders, respectively. It follows that $mt=x+y$ and $z=m\sqrt{t}$. In a steady state, the dealer can be expected to receive on the average the same amount of autonomous orders from the interbank market as he himself is using for position adjustment purposes in the market, implying that $y=z=m\sqrt{t}$. The ratio of total interbank transactions to the gross volume of total trade is hence $2/(2+\sqrt{t})$, which makes 22 per cent for t=50, 17 per cent for t=100 and 8 per cent for t=500. The proportions are somewhat larger than above, but nonetheless small relative to the observed significance of the interbank trade.

In conclusion, it can be stated that, although interbank transactions are important for inventory management purposes, transactions demand and induced wholesale transactions do not as such explain the huge size of the interbank market. Therefore, a majority of interbank transactions must be related to motives other than a pure transactions motive. The results of *Chapter Three* can be used to fill the gap. As the quotations are constantly fluctuating over time and as there may be small differences between prices quoted in different markets even at any particular moment, dealers have an incentive to invest in information by keeping in constant touch with the market. Requesting prices from other dealers and engaging in transactions with them, even if there is no need to do so for position adjustment purposes, is a valuable source of information. Good information pays off over time because it is easier to find the best prices in the event that there is an immediate need to adjust the position by undertaking a wholesale transaction.

Appendix

Assume that after each wholesale purchase the position is restored to $k_1=K_1/H$ and after each wholesale sale to $k_2=K_2/H$, where $k_2 \geq k_1$. In the discrete case these target levels must be at least one step above or below their respective thresholds; *i.e.*, $K_1>sm$ and $K_2<H-sm$, where sm is the size of a unit step. It follows that $k_2-k_1<1$.

Let π_j stand for the probability that if at moment j, a wholesale deal is done, it is a purchase. The probability of it being a sale is hence $1-\pi_j$. The probability that the next passage is zero (the following wholesale deal is a purchase) is

(B.1) $$\pi_{j+1} = \pi_j(1-k_1) + (1-\pi_j)(1-k_2).$$

The probability that the next passage will instead be at H (the following wholesale deal is a sale) is accordingly

(B.2) $$1-\pi_{j+1} = \pi_j k_1 + (1-\pi_j)k_2.$$

The steady state probabilities of wholesale sales and purchases become

(B.3) $$\pi^* = (1-k_2)/(1+k_1-k_2) \quad \text{(purchases)}$$

(B.4) $$1-\pi^* = k_1/(1+k_1-k_2) \quad \text{(sales).}$$

These belong to the interval (0,1), because $0<k_1 \leq k_2<1$.

The interpretation of the steady state probabilities is that, if the process has been going on for a sufficiently long time, then the relative frequency of wholesale purchases and sales corresponds to these probabilities. The number of wholesale transactions, in turn, depends on the average duration between the two passages at either of the limits. The expected duration for a given H, when the process is restarted from the lower return point K_1, is (*cf.* eq.(5.2))

(B.5) $$d_1 = k_1(1-k_1)\, H^2/(sm)^2.$$

When the process is restarted from the upper return point K_2, the expected duration is

(B.6) $$d_2 = k_2(1-k_2)\, H^2/(sm)^2.$$

The average duration weighted by the steady state probabilities is

(B.7) $$d = [\pi^* k_1(1-k_1) - (1-\pi^*)k_2(1-k_2)]\ H^2/(sm)^2$$

$$= k_1(1-k_2)(1-k_1+k_2)/(1+k_1-k_2)\ H^2/(sm)^2.$$

Differentiating with respect to k_1 and k_2, setting partial derivatives equal to zero and solving for k_1 and k_2 gives two solutions. One is $k_1=k_2=1/2$, in which case the average duration is $(H/2sm)^2$, which solution gives the maximum expected duration between two wholesale deals. The other solution is $k_1=0$ and $k_2=1$, but these are not allowed. Therefore, the closest case will be the one in which $K_1=sm$ and $K_2=H-sm$, in which case the average duration will be $d=(H-sm)/sm$. This is always smaller than $(H/2sm)^2$, except for $H/sm=2$, in which case they are equal and each retail transaction is always automatically covered in the interbank market.

The average number of wholesale transactions per day is approximately equal to the inverse of d multiplied by the number *minutes* in a day,

(B.8) $$n = [(sm)^2t/H^2]\ (1+k_1-k_2)/k_1(1-k_2)(1-k_1+k_2).$$

The relative frequency of wholesale purchases is π^*, and each time one occurs the dealer pays an amount S^aK_1/s. The relative frequency of wholesale sales is $1-\pi^*$, and each occurrence brings home an amount $S^b(H-K_2)/s$ in domestic currency. The expected daily cost of wholesale transactions is hence

(B.9) $$C_w = n[\pi^* S^a K_1/s - (1-\pi^*)S^b(H-K_2)/s]$$

$$= (S^a-S^b)sm^2t/H(1-k_1+k_2).$$

Note that this cost increases monotonically with k_1 and decreases monotonically with k_2. Given the constraints on k_1 and k_2, the cost for a given H is minimized when k_1 obtains the lowest allowable value $k_1=sm/H$ and when k_2 receives its highest allowable value $k_2=(H-sm)/H$. Minimizing the cost function C_w+rH with respect to H and letting k_1 and k_2 take the above-mentioned values gives the following expression for the desired trading balances:

(B.10) $$H = [(S^a-S^b)sm^2t/2r]^{1/2} + sm,$$

which is smaller than H in equation (5.9).

Combining the proportional and the fixed cost elements into the same model leads to the following cost function:

(B.11) $$C = C_w + c_o(sm)^2t/\mathrm{H}^2(1+k_1-k_2)/k_1(1-k_2)(1-k_1+k_2) + rH,$$

where C_w is defined in (B.9) and the second term has been obtained by multiplying the expected number of wholesale transactions (equation B.8) by the fixed cost c_o associated with each such deal. Given the symmetric process and the equal holding cost r for both currencies, the return points are symmetrically related to their respective thresholds, k_1=1-k_2. Setting $A=(S^a-S^b)sm^2t$ and $B=c_o(sm)^2t$, the cost function can be rewritten as follows:

$$C = (AHk_1+2B)/[2H^2k_1(1-k_1)] + rH. \tag{B.12}$$

The necessary conditions for the minimum turn out to be

$$AHk_1^2 + 4Bk_1 - 2B = 0 \tag{B.13}$$

$$2rk_1(1-k_1)H^3 - Ak_1H - 4B = 0. \tag{B.14}$$

The positive root of (B.13) is

$$k_1 = \{[2B(2B+AH)]^{1/2} - 2B\}/AH. \tag{B.15}$$

Subtracting 1/2-k_1 yields $\{AH-2[2B(2B+AH)]^{1/2}+4B]\}/AH=[(AH+2B)^{1/2}+(2B)^{1/2}]^2/AH>0$, implying that the lower return point is somewhere between zero and $H/2$. Because $k_2=1-k_1$, it implies that the upper return point is somewhere between $H/2$ and H. From B.13) it is seen that $k_1\rightarrow 1/2$ as the proportional transaction cost (S^a-S^b) approaches zero, $A\rightarrow 0$. From (B.14) it is seen that $H^3=8B/r$ when $k_1=1/2$, which corresponds to equation (5.7) in the text. On the other hand, as the fixed transaction cost c_o approaches zero, $B\rightarrow 0$, and k_1 approaches its lowest allowable limit sm/H, in which case we return to equation (B.10).

Chapter Five

CONSISTENT EXCHANGE RATES AND THE ROLE OF THE VEHICLE CURRENCY

This essay extends the analysis to an N-currency world. The starting point is Chacholiades's (1971 and 1978) study on the sufficiency of three-point arbitrage to insure the consistency of cross exchange rates across financial centres. Consistency is defined as the absence of profitable arbitrage opportunities. The purpose of the essay is to demonstrate the extent and nature of the information problem required for the consistency of exchange rates across market makers.

Foreign exchange dealers of commercial banks are identified as the principal *arbitrageurs*. They are in the best position to be informed on prices quoted by other market makers and therefore in the best position to search for profit opportunities through arbitrage. Of course, their non-bank customers are interested in buying the currency they need as cheaply as possible, but as long as dealers arbitrage efficiently, customers have to choose from quotations which are already consistent. This contributes to the highly competitive nature of the foreign exchange market but as such it is not sufficient to ensure consistency.

The fact that each dealer quotes prices without knowing how others are quoting at the same moment implies that arbitrage profit opportunities may occasionally arise. This possibility was briefly discussed in *Chapter Three*, where it was interpreted as one of the determinants of the bid-ask spread. Observing a pair of inconsistent quotations in the case of two currencies is not a particularly demanding task technically, although it may require good luck and quick reaction. Technical problems arise when there are a large number of currencies and their rates against each other are simultaneously and independently quoted by a large number of foreign exchange dealers. Comparing the bilateral exchange rates across the market makers may not suffice, because it is possible that there exists an arbitrage opportunity which can be realized by simultaneous sales and purchases of many currencies. To check for all possible arbitrage profit opportunities requires, in principle, the processing of an enormous amount of information.

Chacholiades examined the conditions for the establishment of consistent spot exchange rates across financial centres and proved that *three-point arbitrage* is all that is necessary to eliminate all arbitrage profits. This is, according to him, the fundamental reason why *K-point arbitrage* (K>3) is a rare phenomenon in the real world, if it takes place at all. The result is valid, but only for the case where the bid-ask spread is zero. As shown in Salo and Suvanto (1993), the result does not hold in the general case when all cross exchange rates are quoted by a number of dealers operating in a number of financial centres and when each single quotation is defined as a two-valued price consisting of a *bid-rate* as well as an *ask-rate*.

By allowing for a positive *bid-ask spread*, the consistency criteria, *i.e.*, the conditions which eliminate all profitable arbitrage opportunities, appear in the form of inequalities rather than in the form of equalities. The implication is that even though prices need not be exactly the same everywhere, it may still not be possible to make riskless profits by simultaneous sales and purchases of various currencies. More importantly, checking for the absence of an arbitrage profit opportunity in any subset of possible arbitrage chains does not warrant the conclusion that all arbitrage opportunities are absent.

This implies that when there are many currencies and when exchange rates are quoted by a large number of dealers, the amount of information needed for potential arbitrageurs to be able to check for all possible arbitrage opportunities quickly explodes as the number of quotations grows. Salo and Suvanto (1993) propose a technical solution, based on a linear programming model, for the detection of a potential arbitrage profit opportunity. It shows that the computational difficulties associated with the processing of the huge amount of information involved can be overcome.

Although the processing of information, once acquired, can be made manageable, the acquisition of information remains a formidable task. If each market maker quotes all exchange rates bilaterally against all other currencies, full information on quotations of a given centre would require a matrix in which each element would include both a bid-rate and an ask-rate. A natural way to economize in the transmission and use of information on bilateral price ratios would be to choose a common *numeraire*, in terms of which all prices are expressed. This would reduce the dimension of information from matrix to vector form. Hence, one would expect that substantial savings in information costs could be achieved if dealers in all centres could agree upon a common numeraire. However, when each bilateral exchange rate quotation consists of two real numbers, it turns out that the unit of account cannot be chosen arbitrarily. Such a solution requires that the bid-ask spread in the quotation of each currency *vis-à-vis* the reference currency must be smaller than the bid-ask spread in the quotation of this currency *vis-à-vis* any other currency.

But the bid-ask spread is a market-determined price for dealer services. According to the results of *Chapters Three* and *Four*, the spread tends to be smaller, the more sensitive the (dealer) customers are to small price differentials and the higher is the volume of trade in this currency. Therefore, the currency chosen to serve as the numeraire currency must be the one for which the market is the most active. If such a currency can be found, then information on all *cross-rates* can be transmitted through a list of *direct quotations* against the numeraire currency. One consequence of this is that when any one dealer has to make a wholesale transaction for position adjustment purposes, he can always choose an indirect way of buying first the numeraire currency and using the proceeds to buy the currency required. Frequent indirect transactions of this kind also give the numeraire currency the role of a *vehicle currency*.

When the numeraire problem has been solved according to principles explained above and when the role of the vehicle currency has been established, it turns out that *two-point arbitrage* through the vehicle currency is sufficient to ensure the consistency of exchange rate quotations across market makers. Although the dimension of the information problem is greatly reduced by the introduction of the numeraire, it remains large enough for it to be practically impossible for any single dealer to be able to check with all potential counterparties whether or not the proposed prices are in fact consistent at each moment. Therefore, nothing prevents arbitrage opportunities from arising occasionally, but these situations are bound to be short-lived. In the words of Chacholiades

(1971, p. 88): "Somewhere, someone is missing the boat, and the trick is to find the rates that are out of line and act quickly."

1 *The Consistency Criteria*

In line with the microstructure approach of the previous chapters, we define the consistency criteria across *market makers* and not across *centres*. Let us adopt the following notation:

$s^a(i,j;\ m)$ = the ask-price of currency j in terms of currency i as quoted by dealer m;

$s^b(i,j;\ m)$ = the bid-price of currency j in terms of currency i as quoted by dealer m;

$\rho(i,j;\ m)$ = the bid-ask spread in proportion to the ask-price; *i.e.*, $\rho(i,j;\ m) = [s^a(i,j;\ m) - s^b(i,j;\ m)] \,/\, s^a(i,j;\ m)$.

A single quotation is denoted by a pair $[s^b(i,j;\ m),\ s^a(i,j;\ m)]$. All ask- and bid-prices are positive real numbers. There are N currencies, $i,j = 1,\dots,N$, and M foreign exchange dealers, $m=1,\dots,M$.

The following four groups of conditions define the consistency of quotations by any single dealer m (*cf.* Salo and Suvanto, 1993):

C1 $\quad s^a(i,i;\ m) = s^b(i,i;\ m) = 1 \quad \forall\ i,\ m;$

C2 $\quad s^b(j,i;\ m) = 1/s^a(i,j;\ m) \Leftrightarrow s^a(i,j;\ m)\, s^b(j,i;\ m) = 1 \quad \forall\ i,j,m;$

C3 $\quad s^a(i,j;\ m) > s^b(i,j;\ m) \Leftrightarrow \rho(i,j;\ m) > 0$
$\quad \Rightarrow\ s^a(i,j;\ m)\, s^a(j,i;\ m) > 1 \Leftrightarrow s^b(i,j;\ m)\, s^b(j,i;\ m) < 1 \quad \forall\ i{\neq}j;$

C4 $\quad s^a(i,j;\ m) \leq s^a(i,k;\ m)\, s^a(k,j;\ m)$
$\quad \Leftrightarrow s^b(j,i;\ m) \geq s^b(j,k;\ m)\, s^b(k,i;\ m) \quad \forall\ i,j,k,m$

The first group of criteria C1 states that any currency i is exchanged for itself at terms of trade equal to one and group C2 that in any single quotation the ask-price of j in terms of i is equal to the inverse of the bid-price of i in terms of j. The criteria C3 state that a positive bid-ask spread is associated with each single quotation, implying that simultaneous buying and selling of currency i with a single market maker is costly. The criteria C4 state that for a given list of quotations by a single market maker it is always at least as cheap to buy j directly with i as to buy it indirectly via some third currency k.

Let T^K stand for the set of all K-point arbitrage chains, $K=2,...,N$. A particular K-point chain is specified by the $2K$-vector $t=(i_1,...,i_K; m_1,...,m_K)$, where $i_1,...,i_K$ is the ordered chain of currencies and $m_1,...,m_K$ are the indices for the corresponding foreign exchange dealers. One transcation in an arbitrage chain is defined by the triplet $(i_j, i_{j+1}; m_j)$ (with $i_{K+1}=i_1$, indicating that the chain ends in the starting currency). The absence of any arbitrage profit opportunities across market makers is then defined as follows:

C5 $\quad s^a(i_1,i_2; m_1)\ s^a(i_2,i_3; m_2)\ ...\ s^a(i_K,i_1; m_K) \geq 1 \quad \forall\ t \in T^K, \quad K= 2,...,N$

The left-hand side defines a K-point arbitrage chain. It expresses the price in terms of the starting currency i_1 of an indirect purchase of the same currency by first purchasing i_2 with i_1 from dealer m_1, using the proceeds to buy i_3 from dealer m_2, and continuing until currency i_K is exchanged for i_1 with dealer m_K. C5 is a complete list of all potential arbitrage chains. If the price is ≥ 1 for each chain, no arbitrage profit is possible. Expressed in terms of bid-rates, the sign of the inequality changes. This follows from C2, which states that the inverse of the buying price of i in terms of j is by definition equal to the selling price of j in terms of i. This allows us to concentrate, for the time being, on the ask-rates alone.

Chapter Three dealt with the case where there were only two currencies quoted by a large number of dealers. For any two currencies the list of potential (two-point) arbitrage chains is $s^a(i,j; m_1)\ s^a(j,i; m_2)$. By C2 the absence of an arbitrage profit opportunity is guaranteed if there is no bid-rate that is higher than the highest ask-rate quoted for currency j in terms of i.

2 *Chacholiades' Theorem Revisited*

Chacholiades acknowledged the inherent complexity of observing K-point arbitrage opportunities when K is large, but he did not accept the view that computational difficulty as such would deter arbitrageurs from attempting to gain easy profits in an age where the use of computers is widespread. Instead, he searched for a more fundamental reason why K-point arbitrage is a rare phenomenon in the real world. For this he proved a theorem according to which three-point arbitrage is all that is needed to eliminate all profitable arbitrage opportunities. In the following we shall show that this result does not hold in the general case, once allowance is made for a positive bid-ask spread. Indeed, each single arbitrage chain has to be checked separately in order to make sure that no arbitrage chain is profitable.

Proposition 1. *Consistency criteria C1 - C5 are necessary and sufficient to exclude all profitable arbitrage opportunities among M foreign exchange dealers and N currencies.*

Proof. Criteria C1 and C2 can be omitted as definitions. Criteria C4 and C5 are strengthened to strict inequalities. Denoting

(5.1) $$c(i,j;\ m) = \log s^a(i,j;\ m),$$

the remaining criteria C3 to C5 can be written as a linear inequality system

(5.2) $$c(i_1,i_2;\ m_1) + c(i_2,i_3;\ m2) + \dots + c(i_k,i_1;\ m_k) > 0$$

$$t \in T^k\ ,\ k = 2,\ \dots\ ,\ N\ .$$

The proof for sufficiency is trivial. By definition no arbitrage opportunity is present, if the strict inequality holds for each single row in system (5.2). For necessity it is sufficient to show that for any one single arbitrage chain $t^a \in T^k$ there are ask-prices satisfying the conditions (5.2), except for chain t^a; *i.e.*, that the following system

$$c(i_1,i_2;\ m_1) + c(i_2,i_3;\ m2) + \dots + c(i_k,i_1;\ m_k) > 0$$

$$t \neq t^a,\ \ t \in T^k\ ,\ k = 2,\ \dots\ ,\ N$$

(5.3)

$$c(i_1,i_2;\ m_1) + c(i_2,i_3;\ m2) + \dots + c(i_k,i_1;\ m_k) > 0$$

$$t \in t^a$$

has a solution. The full proof that such a solution exists is presented in Salo and Suvanto (1993).[1]

The implication of the above result is that, if any single arbitrage chain is left unchecked while all other chains are found nonprofitable, this one chain may be profitable. More generally, checking for the absence of any subset of possible arbitrage chains, such as all three-point chains, does not warrant the conclusion that all arbitrage opportunities are absent.

As an illustration, let us assume that we have only four currencies, $N=1,\dots,4$, and any number M of dealers, and that all two-point and three-point arbitrage profit opportunities are absent. The absence of profitable three-point arbitrage chains implies that

(5.4) $$[s^a(1,2;\ m_1)\ s^a(2,3;\ m_2)\ s^a(3,1;\ m_3)] \times [s^a(3,4;\ m_4)\ s^a(4,1;\ m_5)\ s^a(1,3;\ m_6)] \geq 1$$

for any combination of six dealers. Rearranging the terms gives

---

[1] In Salo and Suvanto (1993) the problem is formulated as a travelling salesman problem: starting from one city we visit K-1 other cities exactly once and finally return to the city we started from while minimizing the length of the tour. In the present case currencies are cities, each pair of cities is connected by M different routes (foreign exchange dealers) and the tour length is the price in terms of the starting currency the arbitrageur pays if he makes an indirect purchase of this currency through simultaneous sales and purhases of other currencies.

$$(5.5) \qquad [s^a(i_1,i_2;\ m_1)\ s^a(i_2,i_3;\ m_2)\ s^a(i_3,i_4;\ m_4)\ s^a(i_4,i_1;\ m_5)] \times [s^a(i_3,i_1;\ m_3)\ s^a(i_1,i_3;\ m_6)] \geq 1.$$

The two-point chain in the latter brackets is ≥1 by assumption, but the four-point chain in the former brackets may be <1. The argument can be reversed. Three-point arbitrage opportunity may be present even if neither two-point nor four-point arbitrage profit is possible.

3 *Information Problem*

Although the absence of short arbitrage opportunities, such as three-point arbitrage, does not necessarily guarantee that all arbitrage profit opportunities would be eliminated, it is possible that two-point or three-point arbitrage would keep exchange rates quoted by different dealers very close to each other, so that for most practical purposes they could be regarded as equal. But if it is true, as Chacholiades claims, that the computational difficulty associated with observing longer arbitrage profit opportunities should not deter attempts to make easy profits, then we should assume that dealers actively seek out these opportunities and exploit them whenever they arise.

Let us, therefore, examine the dimension of the information problem more closely, while noting that two dimensions are involved in this problem: the difficulty related to the acquisition of data on prices, and the computational difficulty of detecting arbitrage profit opportunities from such data.

Assume that a potential arbitrageur wants to check all arbitrage profit opportunities when all N currencies are quoted bilaterally against each other by all M dealers. To check all two-point opportunities, the arbitrageur first has to choose $s^a(i,j;\ m)$, which can be done in $MN(N$-1) alternative ways and then to find $s^a(j,i;\ n)=1/s^b(i,j;\ n)$ from amongst the remaining M-1 market makers ($n \neq m$). The number of possible combinations is $N(N$-1)$M(M$-1). Because each two-point chain can be ordered in two different ways, only half of these chains are independent. In order to check all three-point chains, the arbitrageur must first choose $s^a(i,j;\ m)$, as above, from $MN(N$-1) alternatives, then choose the second quotation $s^a(j,h;\ n)$ from (N-2)(M-1) alternatives ($h \neq i;\ n \neq m$) and finally choose

Table 5.1 Number of Potential Arbitrage Chains

Number of dealers M	Number of currencies N	Number of combinations Two-point	Three-point
10	10	4 050	172 000
	20	17 100	1 641 600
50	10	110 250	28 224 000
	20	465 500	268 128 000
100	10	445 500	232 848 000
	20	1 881 000	22 212 056 000

$s^a(h,i;\ p)$ from any of the remaining M-2 market makers ($p \neq n,m$). Because any three-point chain can be written in three different orders, the number of independent combinations is $N(N\text{-}1)(N\text{-}2)M(M\text{-}1)(M\text{-}2)/3$. In the general case, the number of independent combinations of potential K-point arbitrage chains is $[N!/(N\text{-}K)!][M!/(M\text{-}K)!]/K$.

As shown in *Table 5.1* the number of combinations, each of which may represent a profitable arbitrage chain, increases explosively as the number of currencies and the number of market makers increase. This is true even for relatively short arbitrage chains, such as three-point chains.

Despite the huge number of combinations to be checked, the computational task related to the detection of profitable arbitrage chains may well be manageable in the age of computers. Salo and Suvanto (1993) propose a procedure, based on network-flow methods, by which one is able to discover a profitable arbitrage chain provided that at least one such opportunity exists. In formulating the problem, the computational difficulty is greatly reduced by first picking the cheapest ask-prices for each bilateral cross exchange rate and using this subset of data for detecting a profitable arbitrage chain, if any exists.

4 *Choice of the Numeraire*

Above we made a counterfactual assumption that foreign exchange dealers receive quotations from others in the form of a matrix in which each element is a single quotation $[s^b(i,j;\ m),\ s^a(i,j;\ m)]$. This appears to be a rather uneconomical way of transmitting information. We also demonstrated that to search for all potential profit opportunities, even if the arbitrage chains are short, involves great difficulty in acquiring the information on price data, even though the computational difficulties could be overcome. Therefore, market participants have a strong common interest in finding arrangements enabling them to economize in the use and transmission of information.

A practical way to economize in information on the matrix of bilateral prices is to choose one of the *goods*, in the present case one of the currencies, as a *numeraire*, which is a unit of account in terms of which the prices of all other currencies are expressed.[2]

Let us try to apply this procedure to the quotation matrix. We assume that dealers have agreed that all of them quote all currencies, $i=1,2,...,N$, *vis-à-vis* an arbitrarily chosen common numeraire currency, say, currency N. Thus dealer m announces a vector of direct quotations against this numeraire: $[s^b(i,N;\ m),\ s^a(i,N;\ m)]$, $i=1,...,N-1$. These are assumed to satisfy the criteria C1 through C3. What then are the cross exchange rates quoted by dealer m that are consistent with this vector of quotations?

For example, the price of k in terms of j can be chosen in four different ways:

(a) $$s^b(j,N;\ m)\ /s^b(k,N;\ m) = s^b(j,N;\ m)\ s^a(N,k;\ m)$$

---

[2] Full information on all bilateral rates of exchange for N *goods* requires a (NxN) matrix, $N(N-1)/2$ elements of which contain independent information. Choosing one good for a numeraire makes it possible to present the same information using a vector containing $N-1$ elements; *cf.* McKinnon (1979, p. 38).

(b) $s^b(j,N;\ m)\ /s^a(k,N;\ m) = s^b(j,N;\ m)\ s^b(N,k;\ m) = [s^b(j,k;\ (m)]$

(c) $s^a(j,N;\ m)\ /s^b(k,N;\ m) = s^a(j,N;\ m)\ s^a(N,k;\ m) = [s^a(j,k;\ m)]$

(d) $s^a(j,N;\ m)\ /s^a(k,N;\ m) = s^a(j,N;\ m)\ s^b(N,k;\ m).$

Because, by criteria C2, $s^a(j,N;\ m)>s^b(j,N;\ m)$ and $s^a(N,k;\ m)>s^b(N,k;\ m)$, it is seen that of the four alternatives (b) is the lowest and (c) is the highest.

Assume that dealer *m* receives a sell order from a customer who wants to buy a given amount of *k* with *j* and that the direct quotations against the numeraire are already given. In this situation it is in the dealer's interest to quote a price for *k* that will yield him as much *j* as possible (*cf.* Hudson, 1979, p. 40). Therefore, the ask-price of *k* in terms of *j* is equal to (c). With a higher price, the customer would refuse to buy, because he would have an option of first buying *N* with *j* and then exchanging *N* for *k*; *i.e.*, by using the chain (c).

Assume then that the dealer receives a buy order from a customer who wants to sell a given amount of *k* for *j*. It is in the dealer's interest to quote a bid price for *k* that enables him to give away as little of *j* as possible. Therefore, the bid-price of *k* in terms of *j* is equal to (b). With a lower price, the customer would refuse to sell because he would have the option of first selling *k* for *N* and then selling *N* for *j*; *i.e.*, by using the chain (b).

The proportionate bid-ask spread is consequently

$$(5.6)\quad \begin{aligned}\rho(j,k;\ m) &= 1 - s^b(j,N;\ m)\ s^b(N,k;\ m)/\ [s^a(j,N;\ m)\ s^a(N,k;\ m)]\\ &= \rho(j,N;\ m) + \rho(N,k;\ m) - \rho(j,N;\ m)\ \rho(N,k;\ m).\end{aligned}$$

Because the proportionate bid-ask spread is a very small number (generally below 0.005; *i.e.*, below 0.5 percent of the ask-rate), the cross-term $\rho(j,N;\ m)\ \rho(N,k;\ m)\approx 0$. For the same reason the proportionate bid-ask spread is practically equal irrespective of the dimension, $\rho(j,k;\ m) \approx \rho(k,j\ ;m)$. Taking these two aspects into account, equation (5.6) can be written as

(5.7) $$\rho(j,k;\ m) \approx \rho(j,N;\ m) + \rho(k,N;\ m).$$

As the bid-ask spread has to be positive, it follows from equation (5.7) that $\rho(j,k;\ m)$ must be larger than either $\rho(j,N;\ m)$ or $\rho(k,N\ ;\ m)$ alone. This must hold for all j and k in any quotation by any market maker m, or

(5.8) $$\rho(i,N;\ m) = \mathbf{Min}_{j}\ \rho(i,j;\ m) \quad \forall\ i.$$

So far the numeraire currency N has been chosen arbitrarily. However, in our earlier analysis the bid-ask spread was a market-determined price for dealer services. As argued in *Chapters Three* and *Four*, under competitive conditions the spread is the smaller, the more price-sensitive and the better informed are the customers and the more active is the market in terms of trading volumes. This seems to be in contradiction to the arbitrarily chosen numeraire. Assume, for instance, that there are two currencies, *1* and *2*, for which the market is deep enough to make the competitive bid-ask spread smaller than the spread that is derived from the arbitrarily chosen numeraire quotations for these currencies; *i.e.*, $\rho(1,2;\ m)<\rho(1,N;\ m)+\rho(2,N;\ m)$ is true for the prices made by at least one dealer m. This would imply that these currencies will be quoted independently of the numeraire quotations, which in turn would imply a loss of the saving in information costs achieved through the use of a common numeraire. The more these kinds of cases can be found, the more questionable any arbitrarily chosen numeraire becomes.

We conclude that a necessary condition for the establishment of a common numeraire is that it is possible to find such a currency (the Nth currency) for which the conditions expressed by equation (5.8) are in force and are backed by market forces.[3] When this kind of numeraire currency can be found and when cross exchange rates are determined in

---

[3] This does not mean that the bid-ask spread would remain constant once established. As shown in earlier chapters, the spreads may widen in reaction to growing uncertainty. For example, some dealers may for defensive reasons want to widen their spreads in order to avoid dealing or to ensure a reasonable profit if deals are made, while other dealers may behave offensively and narrow their spreads in order to obtain more business (*cf.* Hudson, 1979, pp. 48-49). However, over a somewhat longer period the proportionate spreads tend to show a reasonable degree of stability.

accordance with the principles explained above, then anyone who wants to buy k with j can do so either directly or indirectly via the numeraire currency. Both ways are equally cheap. On the other hand, buying k with j either directly or through N is always cheaper than buying it through some third currency h.

5 *Role of the Vehicle Currency*

The implication of the above reasoning is that when it is possible to agree upon a common numeraire currency, then this currency can also be used as a *vehicle currency*, which means that currencies can be bought and sold indirectly through this currency. Frequent indirect transactions of this kind make the market for this currency deeper than it would be otherwise, thus supporting the smaller bid-ask spread, which is necessary for maintaining this role.

If, as above, we define arbitrage as a chain of simultaneous sales and purchases of currencies starting from one currency and returning to this same currency, and if information on exchange rates is transmitted as a vector of numeraire currency quotations, then the question arises as to whether this information alone is sufficient to reveal arbitrage profit opportunities whenever they occur. Next, we shall show that this is indeed the case and that two-point arbitrage alone is sufficient to keep exchange rates consistent.

Proposition 2. *If the depth of the market allows for the choice of a common numeraire currency and if cross exchange rates are determined on the basis of the numeraire currency quotations and if there is an arbitrage opportunity, there is at least one two-point arbitrage profit opportunity through the numeraire currency N,*

Proof. Assume first that there is a profit opportunity through two-point arbitrage between currencies i and j quoted by dealers m and n, or

$$s^a(i,j;\ m)\ s^a(j,i;\ n) < 1, \qquad i,\ j \neq N. \tag{5.9}$$

This can be rewritten as follows:

$$(5.10) \qquad s^a(i,N;\ m)\ s^a(N,j;\ m)\ s^a(j,N;\ n)\ s^a(N,i;\ n) < 1,$$

or after rearranging the terms,

$$(5.11) \qquad [s^a(i,N;\ m)\ s^a(N,i;\ n)]\ [s^a(j,N;\ n)\ s^a(N,j;\ m)] < 1.$$

As the whole expression is <1, at least one of the expressions within the brackets must be <1. Assume next that there are three currencies i, j and k, which are quoted by dealers m, n and p in such a way that three-point arbitrage becomes profitable, or

$$(5.12) \qquad s^a(i,j;\ m)\ s^a(j,k;\ n)\ s^a(k,i;\ p) < 1, \quad i,\ j,\ k \neq N.$$

This can be rewritten as follows:

$$(5.13)\ s^a(i,N;\ m)\ s^a(N,j;\ m)\ s^a(j,N;\ n)\ s^a(N,k;\ n)\ s^a(k,N;\ p)\ s^a(N,i;\ p)<1,$$

and after rearranging the terms

$$(5.14) \qquad \begin{aligned} &[s^a(i,N;\ m)\ s^a(N,i;\ p)]\ [s^a(j,N;\ n)\ s^a(N,j;\ m)] \\ &\times [s^a(k,N;\ p)\ s^a(N,k;\ n)] < 1. \end{aligned}$$

As the whole expression is <1, there must be at least one two-point arbitrage chain which is <1. In a similar way, any given K-point arbitrage chain which is <1 can be decomposed into a sequence of two-point chains via the numeraire currency, at least one of which must be <1. This completes the proof for the general case.

According to this result, longer than two-point arbitrage chains are redundant in ensuring consistent exchange rates between market makers in the case where information is transmitted through the numeraire currency quotations. The result does not exclude the possibility of even larger profits through three-point or even longer arbitrage chains, but these can be reaped by two or more simultaneous acts of two-point arbitrage. For example, if the first two two-point chains in the expression

Table 5.2 Information Problem with the Vehicle Currency

Number of dealers M	Number of currencies N	Number of combinations $M(N\text{-}1)(M\text{-}1)$
10	10	810
	20	1 710
50	10	22 050
	20	46 550
100	10	89 100
	20	188 100

(5.14) are both <1, then a three-point chain, say, $s^a(j,N;\ m)\ s^a(N,i;\ m)\ s^a(i,N;\ m)$, would give the same profit as the two two-point acts. Because the dealers receive direct information only on quotations against the numeraire currency, it seems natural to assume that any single two-point profit opportunity is already observed before anybody even has time to search for possible longer-chain, or multiple two-point, profit opportunities.

The dimension of the informational problem in observing two-point arbitrage profit opportunities from a list of numeraire quotations is greatly reduced compared to the general case. To check for all potential profit opportunities in the case of N currencies and M dealers, one has first to choose one price $s^a(i,N;\ m)$ from $M(N\text{-}1)$ alternatives ($i{\neq}N$) and then to choose the second quotation $s^a(N,i;n)=1/s^b(i,N;n)$ from each of the remaining M-1 alternatives ($p{\neq}m$). Searching for arbitrage profits merely means that a potential arbitrageur checks from amongst the lists of direct quotations against the numeraire whether some dealer m quotes a bid-rate for some currency i that is higher than the ask-rate for the same currency quoted by some other dealer n.

The number of possible combinations is hence $M(N\text{-}1)(M\text{-}1)$. *Table 5.2* demonstrates the dramatic reduction in the dimension of the information problem, compared with the case of the full set of bilateral quotations by each dealer. Notwithstanding this, the information problem remains substantial when the number of dealers is large.

6 Universality of the Vehicle Currency

The intermediary role of the Nth currency in the above analysis should be interpreted with caution. First, the analysis is confined only to the interbank market and it takes into account only two aspects of the vehicle currency: its role as a numeraire and as a medium of exchange (*cf.* Magee and Rao, 1980, pp. 368-369). Of these, the numeraire role is, perhaps, more fundamental and requires a transaction cost explanation as a background. Secondly, the results do not imply the universality of the vehicle currency in the sense that all interbank transactions would be made through the Nth currency as an intermediary. What they show is that all or most arbitrage transactions are likely to be made through the vehicle currency. But the interbank trade in currencies is not constrained to pure arbitrage transactions alone.

As discussed in earlier chapters, dealers frequently trade with each other in order to adjust their positions. For example, one dealer may have overbought currency i and oversold currency j and is therefore looking for an opportunity to exchange i for j at the best possible price available. In some situations, it may be cheaper to use the indirect channel (two transactions with two counterparties) than to make a direct transaction (one transaction with one counterparty).

To show this, let us assume that the vehicle currency quotations for i and j by market makers m and p are related to each other in the following way:

$$(5.15) \qquad s^b(i,N;p) < s^a(i,N;p) = s^b(i,N;m) < s^a(i,N;m), \text{ and}$$
$$s^b(j,N;p) < s^a(j,N;p) = s^b(j,N;m) < s^a(j,N;m).$$

It can be easily checked that these quotations satisfy the consistency criteria, implying that a pure arbitrage profit opportunity is absent, although some quotations are just on the edge of a such an opportunity (buying price of one market maker being equal to the selling price of another). What these quotations show is that dealer m makes the price of the Nth currency relatively high, whereas dealer p makes it relatively cheap.

Four trading alternatives can now be written

(5.16) (i) $s^a(i,j;\ p) = s^a(i,N;\ p)\ s^a(N,j;\ p),$

(ii) $s^a(i,j;\ m) = s^a(i,N;\ m)\ s^a(N,j;\ m)$

(iii) $s^a(i,j;\ p,m) = s^a(i,N;\ p)\ s^a(N,j;\ m)$

(iv) $s^a(i,j;\ m,p) = s^a(i,N;\ m)\ s^a(N,j;\ p);$

where (i) and (ii) represent direct trading alternatives, the cross-rates being determined on the basis of numeraire quotations. The latter two represent indirect trade with two dealers p and m. Applying the quotations as shown by (5.16) and noting that the bid-rate is obtained by inverting the respective ask-rate, *i.e.*, $s^b(i,N;\ p)=1/s^a(N,i;\ m)$, it is seen that the cheapest way of buying j is to use the indirect alternative (iii) of first exchanging i for N with dealer p and using the proceeds to buy j from dealer m.

This demonstrates that the Nth currency can serve a useful purpose even when no arbitrage profit opportunities exist. This result does not require that some quotations are already on the arbitrage limit. The sufficient condition for a profitable indirect purchase is

(5.17) $$s^a(i,j;\ p,m) < \begin{matrix} s^a(i,j;\ m) \\ s^a(i,j;\ p) \\ s^a(i,j;\ m,p). \end{matrix}$$

This condition is satisfied when

(5.18) $$s^a(i,N;\ p) < s^a(i,N;\ m), \quad s^b(j,N;\ p) < s^b(j,N;\ m).$$

Given the fact that the dealers announce their quotations independently, situations where indirect wholesale transactions become profitable are likely to be frequent and systematically exploited. However, it by no means guarantees that the numeraire currency would be the only single vehicle currency in the world. Direct transactions and occasional use of currencies other than the Nth one in the intermediary role are likely to occur in a world where exchange rates are in continuous motion and no dealer has complete information on quotations by all other dealers at each moment of time (*cf.* McKinnon, 1979, pp. 38-39, and Chrystal, 1980).

Finally, it may be worth bearing in mind that the vehicle currency role in the above analysis is ultimately based on lower transaction costs, which in turn stem from the larger volume of trade with this currency, taking both retail transactions and inter-dealer transactions into account (*cf.* Krugman, 1980, p. 523). The vehicle currency role in itself increases the volume of trade and thus reinforces its own role and makes it more persistent. However, interbank transactions do not make the whole market, and if payment habits and the relative importance of various currencies in international trade change, new vehicle currencies may emerge and old ones decline.

REFERENCES

Amihud, Y. and H. Mendelson (1980), 'Dealership Market: Market Making with Inventory'. *Journal of Financial Economics* **8**, 31-53.

Arrow, K. (1959), 'Toward a Theory of Price Adjustment'. In Abramovitz, M. (ed.), *The Allocation of Economic Resources*. Stanford University Press.

Baumol, W.J. (1952), 'Transactions Demand for Cash: An Inventory Theoretic Approach'. *Quarterly Journal of Economics* **75**, 545-556.

Baumol, W. (1965), *The Stock Market and Economic Efficiency*. New York: Fordham University Press.

BIS (1990), *Survey of Foreign Exchange Market Activity*. Basle: Bank for International Settlements.

Begg, D. (1989), 'Floating Exchange Rates in Theory and Practice. *Oxford Review of Economic Policy* **5**, 24-39.

Bell, S. and B. Kettel (1983), *Foreign Exchange Handbook*. London: Graham and Trotman.

Bertsekas, D.P. (1976), *Dynamic Programming and Stochastic Control*. New York: Academic Press.

Brunner, K. and A. Meltzer (1971), 'The Use of Money: Money in the Theory of an Exchange economy'. *American Economic Review* **61**, 784-805.

Chacholiades, M. (1971), 'The Sufficiency of Three-Point Arbitrage to Insure Consistent Cross Rates of Exchange'. *Southern Economic Journal* **38**, 86-88.

Chacholiades, M. (1978), *International Monetary Theory and Policy*. New York: McGraw-Hill.

Chrystal, A. (1984), 'On the Theory of International Money'. In Black, J. and G. Dorrance (eds.), *Problems of International Finance*, New York: St. Martin's Press.

Clower, R.W. (1969), 'Introduction'. In Clower, R.W. (ed.), *Monetary Theory*, Harmondsworth: Penguin Books.

Cohen, K., S. Maier, R. Schwartz and D. Whitcomb (1979), 'Market Makers and the Market Structure: A Review of Recent Literature'. *Journal of Financial and Quantitative Analysis* **14**, 813-835.

Demsetz, H. (1968), 'The Cost of Transacting'. *Quarterly Journal of Economics* **82**, 33-53.

Dennert, J. (1989), 'Price Competition Between Market Makers'. London School of Economics, LSE Financial Market Group Discussion Paper No.52.

Dornbusch, R. and C. Bechman (1985), 'The Bid-Ask Spread in the Black Market for Dollars in Brazil'. *Journal of Money, Credit, and Banking* **17**, 517-520.

Einzig, P. (1969), *A Textbook on Foreign Exchange.* London: Macmillan (Second Edition).

Einzig, P. (1970), *The History of Foreign Exchange.* London: Macmillan (Second Edition).

Eppen, G. and E. Fama (1969), 'Cash Balance and Simple Dynamic Portfolio Problems with Proportional Costs'. *International Economic Review* **10**, 119-133.

Fama, E. (1970), 'Efficient Capital Markets: A Review of Theory and Empirical Work'. *Journal of Finance* **25**, 383-417.

Garman, M. (1976), 'Market Microstructure'. *Journal of Financial Economics*, vol. 3, pp.

GATT (1990), *World Trade 1990*. Geneva.

Glosten, L. (1989), 'Insider Trading, Liquidity and the Risk of the Monopolist Specialist'. *Journal of Business* **62**, 211-235.

Glosten, L. and P. Milgrom (1985), 'Bid, Ask and Transaction Prices in a Specialist Market with Heterogenously Informed Traders'. *Journal of Financial Economics* **14**, 71-100.

Goodhart, C. (1988), 'The Foreign Exchange Market: A Random Walk with a Dragging Anchor'. *Economica* **55**, 437-460.

Goodhart, C. and D. Demos (1990), 'Reuter Screen Images of the Foreign Exchange Market: The Deutsche Mark/Dollar Spot Rate'. *Journal of International Securities Market*, 333-348 (Winter).

Goodhart, C. and L. Fiuglioli (1991), 'Every Minute Counts in Financial Markets'. *Journal of International Money and Finance* **10**, 23-52.

Goodhart, C. and M. Giugale (1988), 'From Hour to Hour in the Foreign Exchange Market'. London School of Economics (*mimeo*).

Grossman, S. and M. Miller (1988), 'Liquidity and Market Structure'. *Journal of Finance* **43**, 617-637.

Grossman, S.J. and J. Stiglitz (1976), 'Information and Competitive Price Systems'. *American Economic Review* **66**, 246-253.

Hagerty, K. (1991), 'Equilibrium Bid-Ask Spreads in Markets with Multiple Assets', *Review of Economic Studies* **58**, 237-257.

Heller, W.R. (1968), 'The Transactions Demand For International Means of Payments'. *Journal of Political Economy* **76**, 141-145.

Hirsch, F. (1969), *Money International.* Harmondsworth: Penguin Books (Revised Edition).

Hirshleifer, J. (1973), 'Exchange Theory: The Missing Chapter'. *Western Economic Journal* **11**, 129-146.

Ho, T. and H.R. Stoll (1980), 'On Dealer Markets under Competition'. *Journal of Finance* **35** 259-267.

Ho, T. and H.R. Stoll (1981), 'Optimal Dealer Pricing under Transactions and Return Uncertainty'. *Journal of Financial Economics* **9**.

Hudson, N.R.L (1979), *Money and Exchange Dealing in International Banking*. London: Macmillan.

Ito, I. and V.V. Roley (1986), 'News for the U.S. and Japan: Which Moves the Yen/-Dollar Exchange Rate?'. NBER Working Paper.

Jones, R.A. (1976), 'The Origin and Development of Media of Exchange'. *Journal of Political Economy* **84**, 757-775.

Kouri, P.J.K. (1983), 'Balance of Payments and the Foreign Exchange Market: A Dynamic Partial Equilibrium Model'. In Bhandari, J.S. and B.H.Putnan (eds.), *Economic Interdependence and Flexible Exchange Rates*. Cambridge, Mass.: The MIT Press.

Krugman, P. (1980), 'Vehicle Currencies and the Structure of International Exchange'. *Journal of Money, Credit, and Banking* **12**, 513-526.

Kyle, A. (1985), 'Continuous Auctions and Insider Trading'. *Econometrica* **53**, 1315-1335.

Levich, R.M. (1979), *The International Money Market: An Assessment of Forecasting Techniques and Market Efficiency*. Greenwich, Conn.: Jai Press.

Magee, S.P. and R.K.S. Rao (1980), 'Vehicle and Non-Vehicle Currencies in International Trade'. *American Economic Review* **70** 368-373 (Papers and Proceedings).

Malinvaud,E. (1972), *Lectures on Microeconomic Theory*. Amsterdam: North-Holland.

Milbourne, R. (1983), 'Optimal Money Holding under Uncertainty'. *International Economic Review* **24**, 685-698.

Miller, M.H. and D. Orr (1966), 'A Model of the Demand for Money by Firms'. *Quarterly Journal of Economics* **80**, 413-435.

McKinnon, R. (1979), *Money in International Exchange: The Convertible Currency System*. Oxford: Oxford University Press.

Müller, U., M. Dacorogna, R. Olsen, O. Pictet, M. Schwartz and C. Morgenegg (1991), 'Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Change Scaling Law, and Intraday Analysis'. *Journal of Banking and Finance* **14**, 1189-1208.

Nagatani, K. (1977), *Monetary Theory*. Amsterdam: North-Holland.

Niehans, J. (1971), 'Money and Barter in General Equilibrium with Transactions Costs'. *American Economic Review*

Officer, L.H. (1976), 'The Demand for International Liquidity: A Test of the Square-Root Law'. *Journal of Money, Credit, and Banking* **8**, 325 337.

Phelps, E.S. and S.G. Winter (1969), 'Optimal Price Policy under Atomistic Competition'. In Phelps, E.S. (ed.), *Microeconomic Foundations of Employment and Inflation Theory*. London: Macmillan, 309-337.

Riehl, H. and R.M. Rodrigues (1983), *Foreign Exchange and Money Markets*. New York: McGraw-Hill.

Salo, S. and A. Suvanto (1993), 'Consistency of Cross Exchange Rates Across Market Makers'. Helsinki School of Economics Working Papers F-306.

Samuelson, P. (1965), 'Proof that Properly Anticipated Prices Fluctuate Randomly'. *Industrial Management Review* **6**, 41-49.

Sender, H. (1985), 'Games Central Banks Play with Currencies'. *Institutional Investor* 120-129 (November).

Stoll, R. (1978), 'The Supply of Dealer Services in Securities Markets'. *Journal of Finance* **33**.

Stoll, R. (1985), 'Alternative Views of Market Making'. In Amihud, Y. *et al.* (eds.), *Market Making and the Changing Structures of Securities Industries*. Lexington Mass.

Suvanto, A. (1982), 'Optimal Quotations and Foreign Exchange Dealing: A Dealer with Short-Term Monopoly'. The Research Institute of the Finnish Economy (ETLA), Discussion Paper No. 114.

Swoboda, A. (1968), *The Euro-Dollar Market: An Interpretation*. Princeton University, Essays in International Finance No. 64.

Swoboda, A. (1969), 'Vehicle Currencies and the Foreign Exchange Market: The Case of the Dollar'. In Aliber, R. (ed.), *The International Market for Foreign Exchange*. New York: Praeger.

Tinic, S.M. (1972), 'The Economics of Liquidity Services'. *Quarterly Journal of Economics* **86**, 77-93.

Tobin J. (1956), 'The Interest Elasticity of the Transactions Demand for Cash'. *Review of Economics and Statistics* **38** 241-247.

Tobin, J. (1984), 'The Efficiency of Financial Markets'. *Lloyds Bank Review* No. 153, 1-15.

Tygier, C. (1983), *Basic Handbook of Foreign Exchange: A Guide to Foreign Exchange Dealing*. London: Euromoney Publications.

Wasserfallen, W. (1989), 'Flexible Exchange Rates: A Closer Look'. *Journal of Monetary Economics* **23**, 511-512.

Willett, T.D. (1969), 'Adequacy of International Means of Payments'. *Review of Economics and Statistics* **51** 373-374.

Yeager, L.B. (1959), 'The Misconceived Problem of International Liquidity'. *Journal of Finance* **14** 347-360.

Zabel, E. (1981), 'Competitive Price Adjustment without Market Clearing'. *Econometrica* **49**, 1201-1221.