

Approaches to Deposit Pricing:

**A Study in the Determination
of Deposit Interest and
Bank Service Charges**

Juha Tarkka

Approaches to Deposit Pricing:

**A Study in the Determination
of Deposit Interest and
Bank Service Charges**

SUOMEN PANKKI
Bank of Finland



BANK OF FINLAND STUDIES E:2

ISBN 951-686-457-0
ISSN 1238-1691

Oy Trio-Offset Ab
Helsinki 1995

Preface

The present study is a theoretical inquiry into the problems of pricing bank deposits, especially chequing accounts and other similar accounts which can be used for making payments. I became interested in the subject already in the late 1980s, when it became apparent that the long era of tranquillity and repressed price competition in the Finnish bank deposit market would soon come to an end. The debate of that time on bank service charges, deposit interest rates and the taxation of interest earnings soon convinced me that not very much was known about the economics of deposit pricing. Not yet understanding the maxim "don't write what you don't know", I took up the line of research that has now culminated in this dissertation.

Several persons have given me advice and support, without which this study would have been impossible to complete. In particular, I want to thank Jouko Vilmunen, who plunged with me into the intricate world of nonlinear pricing. His contribution was extremely valuable. Special thanks also go to Seppo Honkapohja and Erkki Koskela, my official examiners. They made important comments on the preliminary drafts and helped me with their expert advice on the relevant literature. Discussions with Sixten Korkman (at the initial stage of the work), and later with Charles Goodhart, Moshe Kim, David Llewellyn, Matti Virén and Xavier Vives were also most useful. Their comments and suggestions are gratefully acknowledged.

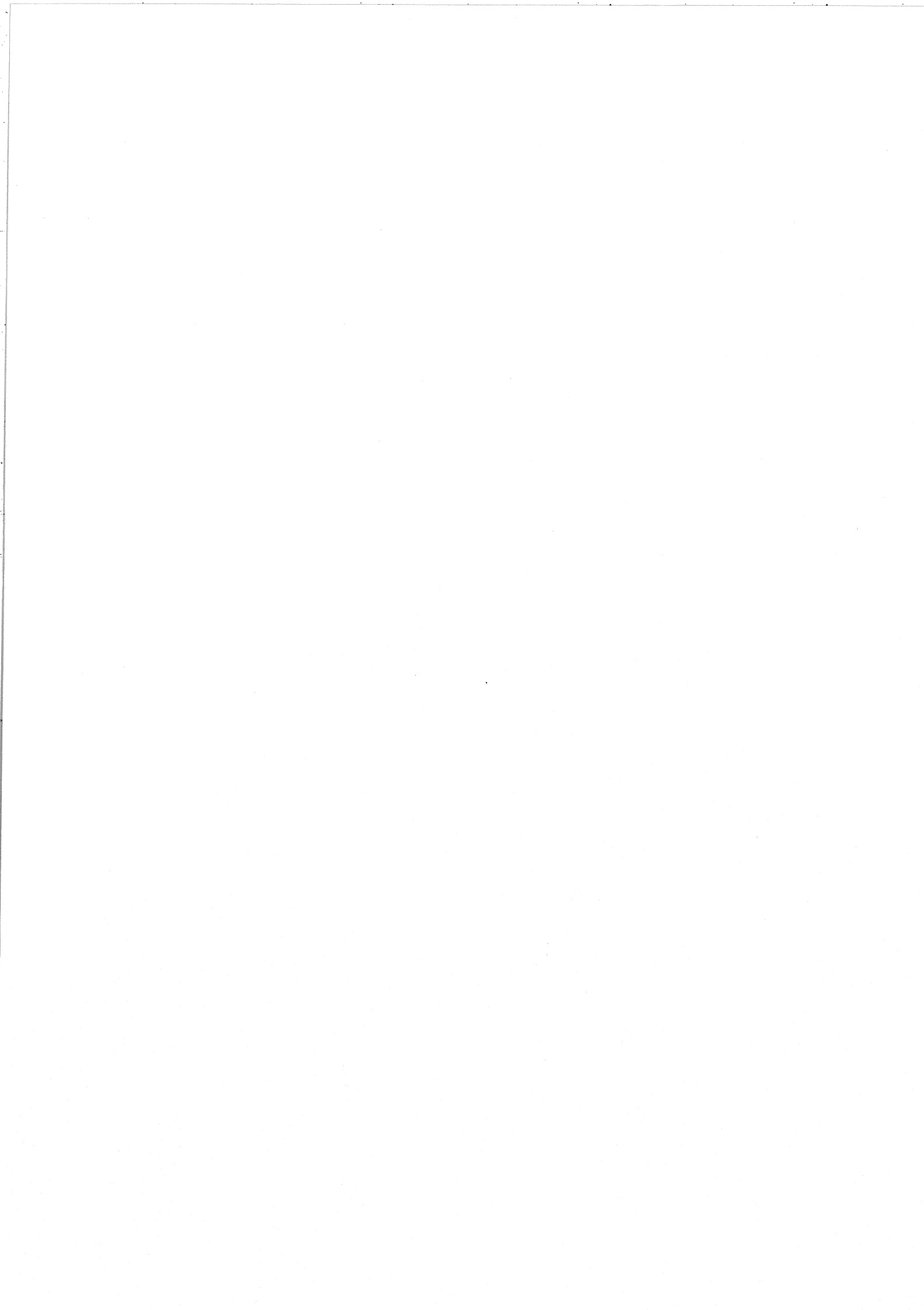
A prerequisite for this work was that my employer, the Bank of Finland, included the study in its research programme, which I highly appreciate. Heikki Koskenkylä and Pekka Ilmakunnas, the successive heads of the Bank's Research Department, deserve thanks for creating and maintaining the stimulating atmosphere of the Department, as well as for their friendly encouragement throughout the project.

Malcolm Waters was of invaluable help in putting style into and removing lapses in my English. Päivi Lindqvist took care of the final processing of the text and the mathematical formulas with her unique spirit and skill. Heli Tikkinen and Anneli Majava prepared the graphs. Anneli Heikkilä organized the printing. My sincere thanks are due to all of them.

Last but not least, I want to thank my wife Helena and my children Laura and Julia for their enduring patience and support.

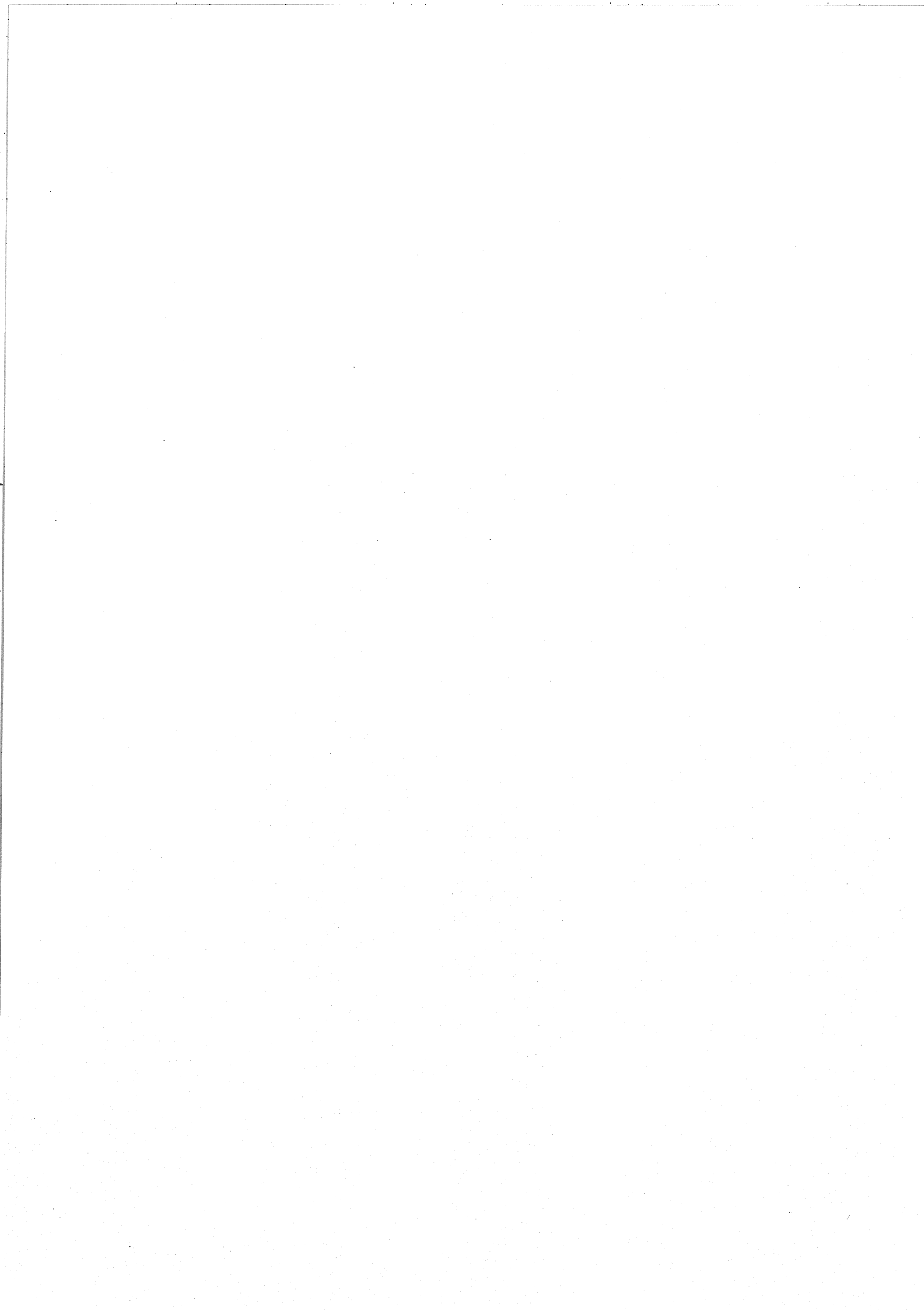
Helsinki, May 1995

Juha Tarkka



Contents

	Page
Preface	5
The Theory and Practice of Bank Deposit Pricing	9
I Risk Sharing in the Pricing of Payment Services	61
II Tax on Interest and the Pricing of Personal Demand Deposits	89
III Implicit Interest as Price Discrimination in the Bank Deposit Market	113
IV Switching Costs and Price Discrimination in the Bank Deposit Market	143



The Theory and Practice of Bank Deposit Pricing

Contents

1	Introduction	11
2	Price parameters of the deposit relationship	15
2.1	Dimensions of pricing	15
2.2	The tariff function	18
3	On the historical development of demand deposit pricing	22
3.1	The classical period of deposit pricing	22
3.2	The period of regulated banking	26
3.3	The new competition	30
4	Theories of deposit account pricing	35
4.1	The free competition model	35
4.2	The monopolistic approach	36
4.3	The implicit interest question	39
4.4	The multiproduct approach	43
5	Developing the theory of deposit pricing	45
5.1	Essay I: The risk sharing explanation	46
5.2	Essay II: The tax explanation	48
5.3	Essay III: The price discrimination explanation	49
5.4	Essay IV: Switching costs	51
6	Conclusions	52
	References	54
	Appendix	
	The deposit account terms offered by a Finnish bank in 1995	59

1 Introduction

Liquid deposits held in various kinds of so called transaction accounts¹ are an essential part of the payment systems of all developed economies; they are also important channels for financial intermediation as these deposits constitute a large part of the asset holdings of the private sector. For banks, the transaction account business typifies their dual role in the economy, as financial intermediaries, on the one hand, and as operators of the payment system, on the other.

The general aim of this study is to develop and consider alternative theories of price formation in the market for transaction accounts. This area of research covers both the pricing of payment services and the determination of the rate of return on privately created deposit money. It will be argued below that the state of theory is unsatisfactory in this area.

The obvious starting point for the analysis is provided by the neoclassical price theory, built on assumptions of efficient markets and competition. This has been applied to banking by a number of authors, including Johnson (1968), Black (1975), Saving (1979), Fama (1980) and Fischer (1983). Writing either before or during a period of almost worldwide deregulation of deposit pricing, these authors attempted to predict what kind of deposit pricing policy would emerge in a deregulated, competitive banking industry.

The results of the aforementioned authors constitute what I propose to call "the Simple Competitive Hypothesis" (SCH) of deposit pricing. Taken in its simplest form, the SCH states that the deposit interest rate should be equal to the rate of return earned by the bank from re-investing the deposited funds; similarly, the SCH would imply that the users of payment services should be charged according to the marginal cost principle. Actually, the SCH is not only a prediction, but is also offered as a norm against which to assess the efficiency of bank pricing practices; in this form, it is called "the Johnson Norm" in the literature (after Harry G. Johnson, who advocated it in his article of 1968).

As will be argued below, the predictions of the SCH are not well corroborated by factual observation. On the contrary, even without external constraints on their pricing behaviour, banks seem to have a

¹ The term transaction account refers to a banks account used for holding transaction balances. Typically, transaction accounts are debitable by cheque or plastic card. The chequing account is the standard example of a transaction account.

tendency to set an apparently "excessive" interest margin on deposits (setting deposit rates low in relation to the return they earn on the deposited funds), and using part of the revenue earned from this interest margin to cover costs generated by the provision of payment services to depositors. Thus payment services appear to be underpriced and an element of "cross subsidization" is present.

It is likely that the understanding of the nature of deposit banking would be greatly enhanced if a theory could be developed which could explain the salient features of banks' pricing behaviour better than the SCH does. It is the aim of the present study to contribute to this goal.

Of course, the SCH rests on certain very restrictive assumptions. The relaxation of these assumptions, one by one, especially in the areas of uncertainty, taxation and imperfect competition, is what constitutes the research strategy of the present study. All of the models which will be presented attempt to explain some features of commonly observed deposit pricing practices. In addition, they provide further insights into the nature and economics of the activities of deposit banks. These additional implications of the theories may be considered to be the potentially most valuable result of the analysis. In short, deposit pricing practices may reveal something essential about what makes banking special as compared with other service industries, and what makes bank deposits special as compared with other financial assets.

Thus, behind the proximate research problem stated above lie more fundamental analytical and policy issues which this study may help resolve. Some of the more readily obvious ones are the following.

First, seen from the perspective of pure monetary theory, the main interest of this study lies in the fundamental controversy about what explains the ubiquitous use of means of payment, which are clearly dominated by other assets in terms of expected return.² Another related problem for pure theory is why payment transmission and financial intermediation activities should be as complementary as they appear to be, judging from the long-standing ability of banks to dominate both activities as compared with more specialized institutions.

Another, more "applied" set of issues, to which the present study may contribute, relates to the economic efficiency of the banking industry and especially that of the payment system. Here, the question is whether the apparent underpricing of payment services is a sign of market inefficiency (as Johnson thought) and whether this encourages

² See Hicks (1935) for an early statement of the problem and Cowen and Kroszner (1987) for a survey.

excessive, wasteful use of resources in the payment system. If it does, and if the sources of this inefficiency are not easily removed by public policy, issues of the "second best" (ie correcting intervention) also arise.

In the context of efficiency, much attention has been focused on deposit interest margins from the related perspectives of competition policy and banking regulation. There is a well-established tradition in banking literature which treats deposit interest margins as evidence of monopolistic pricing. On the other hand, there are also studies which assume that, because of competition, any "excess" interest margin on deposits must be offset by corresponding "implicit interest" in the form of underpriced services rendered to depositors. A comprehensive analysis of both interest rate determination and service pricing is required before either of these approaches can be accepted. Of course, conclusions regarding how the conduct of banks coincides with public interest are at stake here.

The typical price structures in the transaction account market seem to give rise to cross subsidies: the deposit side of the market is "expensive" to the customer in terms of the interest income which is lost, while the payment services side of the market is "cheap". So in addition to the efficiency issue mentioned above, this invokes a third issue, the issue of equity: consumers with relatively large deposit balances and few transactions may suffer, while those with a lot of transactions relative to deposit balances benefit from the system.

Not only regulators but also monetary policy makers are interested in the bank deposit market. For example, the influence of money market interest rates on the rates paid on deposits are crucial for the interest elasticity of money (cf. Keeley and Zimmerman, 1986). Recent developments in monetary aggregation have stressed the importance of deposit interest margins as indicators of the "moneyness" of different types of bank deposits (see Barnett, 1980). The justification for this depends on how the deposit interest rates (and the liquidity of the accounts) are determined, an issue largely ignored in the aggregation literature.

Finally, as will be demonstrated in one of the theoretical essays included in this study, a full understanding of the effects of reserve requirements and the taxation of interest income cannot be achieved without analyzing these together in the context of a model of the market for bank services.

The analysis presented in this study sets out from the observation that banks provide multiple services to their deposit customers: deposit-taking and payment transmission services. These services are conceptually distinct, but possibly complementary in the sense that it

is economical for the customer to obtain these services from the same institution; also, it may be that they are complementary in the traditional sense that the marginal benefit from one increases with the use of the other. Recognizing the kinds of services used by depositors, it is immediately obvious that the transaction accounts business typifies the banks' activities in the deposit markets. These accounts constitute a vehicle for providing both deposit-taking and payment transmission services in a very flexible way, while also allowing for most sophisticated pricing policies, including tying, price discrimination and cross subsidies.

It should be clear that, although deposit-taking and payment services appear together in deposit banks' activities in general, and in their transaction account business in particular, these services are by no means necessarily conjoined. Deposit-taking can be carried on without providing any payment (or liquidity) services; the CD market provides an example of this. On the other hand, there are payment transmission services which in practice do not provide any useful investment outlet for the service users; postal money transmission services and some credit card programmes are often of this kind. It is a typical, and virtually unique, feature of transaction accounts that the customer is free to use the deposit and payment functions in variable, selected proportions (compare this with standard credit card services, for example).

The present study consists of this introductory essay and four theoretical chapters. Each chapter approaches the bank account pricing problem from a different angle. The introductory essay includes a general survey on the practice and development of deposit pricing in some countries, as well as a review of the most important theoretical studies on the subject. A brief summary and assessment of the following theoretical essay chapters is also provided. The three theoretical essays contain the main analytical contributions of the study.

The first two theoretical essays use models which assume perfect competition in the transaction deposit market. The first of these looks at bank deposits as a risk-sharing arrangement. From that perspective, the "underpricing" of payment services can be rationalized as insurance against uncertainty in individuals' demand for payment services. The second essay considers the effects of the taxability of deposit interest. It is shown that it may encourage banks to reward depositors with tax-free "implicit interest" in the form of cheap payment services instead of paying them taxable explicit interest. This has further, possibly adverse implications for the efficiency of the banking system.

Essays three and four use models of imperfect competition and approach deposit pricing as a price discrimination problem. The first of these considers the case of a monopoly bank and demonstrates how the optimal use of market power is reflected in the structure of the deposit interest and service charge schedules. The last essay generalizes these results to a case where there is some competition. It is assumed that the market imperfection which creates the opportunity for price discrimination is caused by (sunk) customer switching costs.

Before proceeding to the survey of the development of actual deposit pricing practices and the theories explaining these practices, it is necessary to clarify what dimensions the pricing of transactions accounts contains and how the price systems can be represented mathematically. This will be done in the next section.

2 Price parameters of the deposit relationship

2.1 Dimensions of pricing

The relationship between a bank and an account holding customer is actually far more complex than the conventional neoclassical model of the market of a single good is able to capture. In the conventional neoclassical model, the revenue from selling a given amount ("quantity") of a service is defined as the product of quantity and a scalar called "the price". In that analysis, price is considered to be separable from the size of each transaction, and the gross revenue accruing to the seller is independent of factors other than quantity and price. For example, the mere existence of a customer relationship is not assumed to generate revenues or costs when no services are used. Although the basic neoclassical framework has frequently been applied to banking, it is not very suitable for that industry in a number of respects. There are two main reasons why less restrictive models are required for the analysis of banking (and the analysis of deposit markets, in particular).

First, the bank renders the account holder several services rather than just one. At the minimum, one should distinguish between payment services and deposit taking services (financial intermediation). This is the choice of the present study. Even these categories are not completely homogenous internally. In more detailed analysis, payment

services could be classified into several subcategories, such as incoming and outgoing payments; credit and debit transfers; and paper-based and electronically-handled services. Similarly, the deposit-taking service is not a well defined, homogenous category, for money can be deposited under different terms and conditions. The accessibility of the deposit, in particular, may vary as the agreed terms may limit the amount which can be withdrawn during a given period, or the number of withdrawals that can be made, for example. Also, the deposit account may be denominated in different monetary units serving as a numeraire etc.

Second, the prices of bank services are not necessarily uniform. This means that different units of the same service can be sold at a different price, even in the same place and on the same day. When price differentiation is exercised, prices may be conditional on the objective characteristics of the customer (age, occupation etc.), or on the purchased quantities of services. The latter case is also called nonlinear pricing, and usually takes the form of quantity discounts or entry fees (monthly fixed charges etc.).

As explained by Wilson (1993), nonlinear pricing is a case of a pricing strategy called bundling. In the conventional definition, bundling is thought to occur when customers purchase several products from the same seller, and the pricing is such that it is cheaper to purchase the products together rather than separately. This definition can be extended to cover nonlinear pricing, however. Nonlinear pricing typically takes the form of quantity discounts and this may be interpreted as bundling successive purchases of the same good, making it cheaper to purchase the first and second unit of the good together rather than purchase the first unit twice, for example. Since banking is a multiproduct business, bundling of different goods is also possible, and actually quite common in reality.

The possibility of nonlinear pricing and price differentiation in banking stems from the fact that services rendered to holders of transaction accounts are not easily transferable. By this I mean that usually there is no useful second-hand market for these services. Non-transferability implies that price arbitrage among the purchasers of the depositors' services is impossible or at least very difficult. Since arbitrage is not possible, the "law of one price" does not necessarily hold.

This scope for price differentiation in the transaction deposit market is in clear contrast with the market for transferable certificates of deposit (CD). These certificates do not entitle their holders to receive any payment or liquidity services from the bank during the time to maturity of the deposit. They do in many cases have a second-

hand market, which makes it difficult to practice price discrimination in the CD market. Even when an organized market is not present, the mere possibility of second-hand trade may prevent discriminatory pricing of these deposits.

Non-transferability of demand is a necessary, but not sufficient condition for price discrimination to emerge, however. Typically, some form of imperfect competition is also required (cf. Philips (1983)). It is obvious that imperfect competition cannot be ruled out in the case of transaction deposit markets for different reasons which hinder the shifting of demand between different suppliers of the service. These reasons include spatial differentiation between banks (due to transportation or access costs), on the one hand, and sunk information and accounting costs involved in establishing new customer relationships, on the other (the costs of the latter category are often called switching costs; see eg Klemperer (1987)).

All in all, the transaction account can be seen as a flexible package of different financial intermediation (deposit) and transaction (payment) services, which allows the supplying bank to practice very sophisticated bundling and price discrimination, if it so wishes and if this is viable in the market. This being the case, it is not surprising that the real world pricing practices of banks actually do often display considerable complexity. At the extreme, bankers speak about "relationship banking", where the banks consider all services used by a customer together before setting the charges and interest rates on them.

Vittas et al. (1988, p. 163) consider the following dimensions of pricing policy in respect of bank services as being the most important:

- a) an implicit charge in the form of an interest rate spread;
- b) activity fees (charges) imposed on individual transactions, consisting of either flat fees or *ad valorem* fees or both;
- c) account maintenance fees and other flat fees.

It is worth noting that in drawing up the above list, Vittas et al. take the view that the interest rate on a deposit account is in a sense not really a price, but a determinant of a price; it is the spread (margin) between the deposit rate and the alternative cost of funds (for the depositor, or for the bank) which can be regarded as a price. This is in accordance with the idea of interest rate spreads as "user costs of money" (see eg Barnett (1979)). Thus, the selection of the appropriate benchmark rate of interest constitutes a major problem in the measurement of prices in banking.

To the classification given above one must add the observation that interest rate spreads and activity fees are not necessarily constants;

rather, the banks often apply nonlinear tariffs specifying interest rates and activity charges as functions of deposit balances and transaction volumes. The nonlinearities often appear as a conditional rebate or the waiving of some charges. For example, the following types of conditionalities and nonlinearities are found in the pricing of transaction services (Llewellyn and Drake (1993), pp. 10–11):

- d) transaction charges are waived if some agreed minimum or average deposit balance is maintained in the account;
- e) fixed quarterly (or monthly or annual) charges entitle the account holder to make a prescribed number of transactions free of charge;
- f) notional calculated transaction charges are reduced on the basis of the size of the average balance maintained in the transaction period.

Finally, nonlinearities are quite common in the determination of interest rates on deposits. The deposit rate is then an increasing function of the balance on the account.

In addition to the pricing parameters which are directly connected to the bank account in question, banks may also bundle transaction account services together with other products. An instance of this is the preferential treatment of "good" depositors in loan markets. When practiced systematically and reliably, this kind of bundling serves a function in the deposit markets not unlike the direct interest rates paid on deposits (Hodgman (1961) is a classic on this topic). Similar bundling may occur, in the spirit of "relationship banking", when a linkage is established between different types of deposit account held by the customer. For example, service charges on a current account may be waived if the customer has sufficiently large savings deposits.

2.2 The tariff function

In the present study, bank account pricing is analyzed as if there were only two services, namely "payment services" and "deposit-taking". The prices of these services are called "service charges" and "deposit (interest) rates", respectively. Only passing reference is made to more detailed consideration of depositors' services. Nonlinearities in the pricing of the two services are taken into account, however. When this is done, the pricing of transaction deposits will often be described by a tariff function. A rather general form for the tariff function can be written as follows:

$$P = P(D, N, C). \quad (1)$$

Here P is the rate of net revenue of the bank from the account (ie net charges levied by the bank less interest paid by the bank). The tariff function specifies this net flow of revenue as a function of the stock of deposit balances D and the volume of payment services N (number of transactions per unit of time). The third argument of the tariff function is the (vector) of observable characteristics C of the customer which are used as a basis for pricing.

There is the question of the sign of the tariff. As defined here, the tariff measures the income flow as the bank's revenue. Therefore, the tariff is positive if the flow of deposit interest is smaller than the (opposite) flow of service charges, and *vice versa*. In principle, the customer characteristics must be included in the tariff function if first or third-degree price discrimination is to be allowed. However, the analytical need for including the argument C in the tariff functions is limited, for the following reasons. On the one hand, if there are no externalities in the use of bank services between customers with different values of C , then the pricing problem can be solved separately for all values of C (treating C as a constant). On the other hand, price discrimination on the basis of observable (extraneous) characteristics is in many cases disallowed by law or custom. In that case, too, the variable C is redundant.

The relationship of the tariff function to more familiar dimensions of account pricing can be illuminated by the following remarks.

1. The marginal interest rate on deposits and the marginal service charge on transactions can be defined with the aid of partial derivatives of the P function:

$$i_{\text{marg.}} = -\partial P(D, N, C)/\partial D \quad (\text{the marginal deposit rate}) \quad (2)$$

$$s_{\text{marg.}} = \partial P(D, N, C)/\partial N \quad (\text{the marginal service charge}). \quad (3)$$

The minus sign in the definition of the marginal deposit rate follows from the fact that the deposit interest is defined as flowing from the bank to the depositor, whereas the tariff is defined as flowing in the opposite direction.

If we denote the marginal rate of return that the bank earns from reinvesting the deposited funds by r , then the difference $r - i_{\text{marg.}}$ is the marginal interest spread on the deposit. Similarly, the difference

between the marginal deposit rate and the customers' opportunity cost of funds defines the rental price of deposits.

The case analyzed by standard neoclassical banking models corresponds to the following type of tariff function:

$$P = s \cdot N - i \cdot D. \quad (4)$$

Here the service charges and deposit rates are uniform, giving rise to a linear tariff.

2. The "intercept" of the tariff defines the account maintenance fee:

$$p = P(0, 0, C) \quad (\text{account maintenance fee}). \quad (5)$$

3. If the tariff function is not additively separable, the tariff cannot be unambiguously or meaningfully decomposed into interest payments, on the one hand, and service charges, on the other. However, it seems to be customary in the industry that, while service charges may be made conditional on deposit balances and the number of transactions, deposit interest rate is not usually a function of the use of transaction services. If this convention is accepted, one could use the following decomposition:

$$P = S - I, \quad (6)$$

where I is the flow of interest payments, defined as

$$I = -P(D, 0, C) - P(0, 0, C) \quad (7)$$

and S is the flow of service charges, defined as

$$S = P(D, N, C) - P(D, 0, C) - P(0, 0, C). \quad (8)$$

According to this decomposition, $i = I/D$ is the average interest rate on deposits. This is not, however, the only possible way to decompose the tariff.

4. Minimum balance requirements are a common feature of bank tariffs. Usually, they specify some level of deposit balances which must be maintained in the account for service charges to be waived. Minimum balance requirements can easily be obtained from the properties of the tariff function. Minimum balance requirements can be said to be used if there is a minimum balance function $D^*(C)$ such that

$$\partial P(D, N, C)/\partial N = 0 \text{ for all } N \text{ if } D > D^*(C) \quad (9)$$

and

$$\partial P(D, N, C)/\partial N > 0 \text{ for all } N \text{ if } D < D^*(C). \quad (10)$$

Finally, some caveats regarding the use of the tariff function must be made.

In some cases, banks offer their customers a choice of different tariffs. A typical example is one where the depositor has an option of paying a higher monthly fee which entitles him to a discount on the transaction-specific charges. Such kinds of optional tariffs create additional difficulties for analysis. As shown by Roberts (1979), a set of optional tariffs can be effectively represented by a single tariff consisting of the lowest-cost envelope of the alternative tariffs. This solution to the problem of optional tariffs presupposes, however, that the customer always chooses the tariff which is most economical for him. Under certainty, this is a natural assumption. This is not so under uncertainty: if there is "usage risk", ie if the customer is not sure in advance of the quantities he will purchase, he may sometimes choose the "wrong" tariff. Thus, in the case with usage risk, the use of a set of optional tariffs is generally more expensive for the consumer than the envelope of the tariffs would indicate (Wilson (1993), pp. 91-92).

Further, it should be noted that the tariff function described as here does not capture all nuances of deposit account pricing as observed in actual markets. This is due to the static formulation of the tariff function, which precludes the analysis of time-dependent aspects of deposit pricing. A common example of this is the distinction between minimum, end-of-month, and average account balances in the tariff schedules. The deposit variable D in the static tariff function is best thought to measure the average deposit balance over the period of analysis (say a month). However, it is not uncommon for banks to pay deposit interest on the minimum monthly balance, or specify the account balance requirements in terms of the minimum monthly balance. While some additional assumption may enable the analysis of such cases with the tariff function, basically they require more information on the time path of the deposit balance than is obtainable from the static formulation.

3 On the historical development of demand deposit pricing

In this section, the actual deposit pricing practices of banks in some countries are reviewed. The treatment is historical, which allows for a greater variety of regulatory regimes than would be possible by considering the recent situation alone. As will become evident below, quite significant changes have occurred in the deposit pricing practices of banks during the last century. Broadly speaking, three phases can be distinguished:

1. the classical period, extending from the last decades of the 19th century to the Great Depression;
2. the period of regulated banking, extending from the Great Depression to the end of the 1970s;
3. the new competitive period, from the end of 1970s to the beginning of the 1990s.

This periodization is very rough, of course, and it does not do full justice to the large differences which can be observed between the pricing practices of banks in different countries at any one time – differences which may be due to idiosyncracies in traditions, regulations, or banking structure. Unfortunately, these differences are hard to survey exhaustively; this is due to scantiness of information and the extreme complexity (multi-dimensionality) of the banks' deposit pricing decision.

However, an attempt will be made below to give a broad outline of the development of "general" pricing practices. This outline is based mainly on available information concerning Finnish, Swedish, U.S., British and German banking practices.

3.1 The classical period of deposit pricing

Broadly speaking, the classical period of deposit pricing is distinguished from the regulated period which followed it by the absence of strong intervention by the authorities in the determination of the terms applied to bank deposits. Thus, in the classical period, the deposit interest decision probably reflected "market forces" – not necessarily only competition, however, for local monopoly, collusion,

or more complex strategic considerations may have played a role as well. By and large, the same can be said about fees and "non-price terms" of the deposit relationship. When assessing the prevalent pricing practices in the classical period, one should bear in mind that money market interest rates or their near equivalents were not usually very high during the classical period. In such conditions, even modest interest rate spreads on deposits brought deposit rates close to zero.

As summarized by Korpisaari (1930), the outstanding features of deposit pricing in the late classical period were the following: in most countries, banks rendered (payment) services to their depositors without levying specific charges. In London and the U.S.A., large banks did not pay interest on chequable deposits and required the account holder to keep a certain minimum balance on the account. Incidentally, Britain and the U.S. were the countries where the custom of using bank deposits as money first became normal among private individuals, as well as among firms. In other countries, it was common to pay some interest on chequable deposits, the deposit rate margin varying from country to country.

Next, these general observations are supplemented by some, slightly more detailed evidence from a few different banking markets.

Before the First World War, Britain had probably the most well developed banking system in the world. Rae (1910) provides a review of British deposit pricing practices at the turn of the century. He distinguishes three methods or "systems" of pricing "current account" services which had become well-established in the British market by the beginning of this century: the London system, the Country system, and the Scottish system.

In the London system, current account balances yielded no interest; on the other hand, transaction charges were not usually levied on either credit or debit transactions. However, to ensure that the deposit relationship was not unprofitable, the bank required the customer to keep some minimum balance on his account. According to Gilbert and Sykes (1911), this minimum balance was "never definitely fixed, but (was) regulated very much by the good sense and proper feeling of the parties".

By contrast, it was normal for English country banks to pay interest on demand deposits (Rae cites representative interest rates of 2 per cent on demand deposits and 2.5 per cent on "money lodged for longer periods"). On the charges side, the Country system typically involved a semiannual maintenance fee. This could be made proportional to the total value of debits during the six months, "whatever they may consist of". Rae notes that the Country system was not completely "fair" in the sense that some customers may have

been cross-subsidized by others, depending on the particular nature of their transactions. The system had, however, "the merit of simplicity".

The Scottish system was different from the Country system in that the banks levied a specific charge on each transaction, debit or credit. According to Rae, some Scottish banks were working upon a scale indicating about 60 varieties of bank charges.

One aspect of the British banking practice repeatedly stressed by Rae and also by Gilbert and Sykes was the need to consider each customer relationship separately and to view the terms applied to current accounts as a part of a larger package, including the different kinds of credit services rendered to the customer. This kind of bundling was apparently a crucial feature of classical banking, suggesting that it may be futile to analyze bank deposit pricing without giving due attention to imperfect competition or other market imperfections which could explain the viability of such pricing policies.

One reason why very different pricing strategies could coexist in Britain may be geographic segmentation of the banking markets. Whatever the reason, the diversity of pricing practices can be taken as evidence against the simple competitive hypothesis. It suggests that even in an unregulated deposit market, different market conditions (and possibly different traditions) may give rise to different deposit pricing practices.

In the United States, the practice of paying some interest on chequable deposits was first introduced by trust companies. One explanation for this may be that trust company deposits were less active (ie had lower turnover and lower operative costs) than bank deposits. Yet, by the early 1920s, the practice of paying interest on demand deposit balances which were above a certain minimum was spreading to the commercial banks as well (Holdsworth (1921), pp. 203-204). By the end of 1926, the average interest rate paid by Federal Reserve member banks on demand deposits was 1.5 per cent. At the same time, the average rate paid on time deposits was 3.5 per cent and the discount rate was 4 per cent. The Federal Reserve Bulletin emphasizes, however, the diversity of demand deposit interest rates by noting that many demand deposit accounts did not carry any interest at all (FRB (1927), pp. 461-462).

Turning to service charges, American payment system in the classical period was (and still is) based on the use of the cheque. The costs of collecting cheques were large, owing to the size of the country and the large number of banks. Prior to the establishment of the Federal Reserve System in 1913, the banks were increasingly levying charges for collecting out-of-town cheques. These charges,

paid by the depositor for whom the collections were made, were often conditional on the distance to the bank on which the cheque was drawn. When banks accepted cheques "at par" ie without collecting charges, the customer was usually required to keep a "compensating balance" on the deposit account. In 1916, the Federal Reserve Board started to regulate the cheque collecting and clearing system of the country by requiring banks to collect cheques (of other member banks) "at par" (Holdsworth (1921), pp. 223–234).

In Germany, some interest was commonly paid on cheque accounts in the classical period. Prior to the First World War at least, the interest rate on these accounts was usually 2–3 percentage points below the money market rate (the so-called Bankdiskont; this rate fluctuated between 4 and 6 per cent at the time). No commission was charged on cheque accounts which carried a constantly positive balance (Obst (1923) pp. 242–250).

In Finland and Sweden, banking practices in the classical period resembled the German model. Interest was paid on current accounts, albeit at a clearly lower rate than on time deposits. In 1871–1886, the interest rate on current account balances was usually 2 per cent in both countries, while the time deposits rate was about 3–4½ per cent. In the 1890s, deposit rates became more responsive to the discount rate quoted by the central bank. Yet, the spread between the current account rate and time deposit rates still seems to have been, as a rule, about 2 per cent in Finland and 3 per cent in Sweden (Korpisaari (1920); SÅS (1923)).

As regards direct charges on payment services in Finland during the classical period, no fees or charges seem to have been imposed on normal cheque business; the popularity of the cheque seems to have grown relatively late in the period, however. This was probably due to the widespread use of the postable cashier's draft (postremiss-växel, postilähetysvekseli), a particularly Swedish and Finnish instrument. The cashier's draft service was reportedly offered free of charge from 1910 on in Finland, the float constituting the banks' only revenue from this line of business. On the credit payment side, the payment order was used. For these transactions, the banks charged telephone costs or postage to the payer whenever the payment was transferred to an account in another bank (Korpisaari (1920)).

These brief observations clearly demonstrate the large diversity of pricing practices which prevailed in the largely unregulated environment of the classical period. In different markets, both interest-bearing and interest-free cheque accounts existed, as did both "free" and explicitly priced payment services. Nonlinear pricing seems to have been the rule rather than the exception in the British and U.S.

markets. Despite this diversity, some common features can be observed. In particular, interest rates paid on chequable accounts seem to have been relatively low. Thus it is evident that the SCH, according to which the absence of regulations generally leads to the elimination of the deposit interest spread, is not supported. Also, full cost-based service charges do not appear to have dominated in the unregulated banking market of the classical period.

3.2 The period of regulated banking

The classical banking period ended with the Great Depression. In many countries, the depression caused a debt crisis and severe problems with bank solvency. Bank failures were numerous, especially in the U.S.A., and debtors worldwide suffered from very high real interest rates. These problems triggered a wave of new regulations, the aim of which was to curtail the allegedly "excessive" competition for deposits. Ultimately, the goal of legislators and regulators was to find a way to enhance the profitability of banks and lower debtor interest rates. Reducing deposit costs by regulation or an officially supported bank cartel seemed to offer a solution to that end. The developments in the U.S. provide a case in point. In that country, the payment of interest on demand deposits was prohibited by law (the so-called Glass-Steagall Act) in 1933. At about the same time, an interest rate ceiling, the famous Regulation Q, was imposed on time deposits.

In other countries, similar developments occurred. The forms of regulation varied, however, and it might actually be more appropriate to speak about the following period as the period of repressed interest rate competition rather than that of "regulation", but the latter seems to be common parlance. While the U.S. restrictions on the payment of interest on chequable deposits were based on law, cartel arrangements and "recommendations" by central banks were prevalent in Europe. In Finland, for example, the banks concluded a deposit rate agreement in 1931 which put an end to virtually all interest rate competition for the next 60 years very short interruptions notwithstanding. This agreement was negotiated under some pressure by the central bank. Later, in the 1950s, the Finnish deposit rate cartel was reinforced by a law which exempted interest paid on the accounts covered by the cartel from personal income tax.

In many instances, the arrangements by which interest rate competition was reduced were quite informal or even tacit in nature. One well-known arrangement was the interest rate cartel of the London clearing banks, and there was also an informal agreement in

Sweden. In Germany, by contrast, where the interwar economic crisis was the most severe of all European countries, the banks' deposit rate agreement dating from 1928 had to be officially reinforced and extended by a special emergency decree which was issued in 1932. After the war, the legal regulation of deposit rates continued in Germany until 1967.

The Great Depression was followed by war and reconstruction. The regulations on deposit interest rates survived these events by decades, on both sides of the Atlantic. In the mid-1960s, every European central bank still had one form or another of direct interest rate controls (Bingham (1985), pp. 129-133; Gardener and Molyneux (1990), pp. 46-47). In the U.S., too, the deposit market regulations remained virtually untouched until the 1970s.

The pricing practices which became generally prevalent worldwide during the regulated banking period are summarized by Vittas et al. (1988) as follows:

"... retail banking institutions did not have to face difficult pricing decisions. They tended either by choice, or more likely by regulatory dictum, to pay no interest, or very low interest, on retail deposit balances. In return, they tended to offer payment services mostly free of charge or at much below cost... few services were subject to explicit charges."

This general description can be supplemented with the observation that, even when explicit charges on transactions were levied in principle, they were in many cases waived if sufficiently large deposit balances were maintained. This practice, inherited from the classical period, was used at least by the U.S. and British banks.

When price competition was suppressed, other means to attract profitable deposits were sought by banks. As is well known, price floors imposed by cartels or regulatory authorities generally create incentives for non-price competition (see eg Stigler (1968) for an analysis). Recall that a deposit rate ceiling is, in fact, a price floor on the rental rate on (the user cost of) deposits. It is therefore not at all surprising that various types of non-price competition proliferated during the period of regulated deposit rates.

In retail deposit banking, non-price competition has taken many different forms. In addition to tying lower service charges or "free" services to deposits, there has been quality competition as banks have sought to improve the convenience and accessibility of their services by expanding their network of branches and service outlets, or by relaxing various terms and conditions applied to deposit accounts

(such as withdrawal limits etc.); there has been much advertising; and, finally, price floors may have been at least partly evaded by the bundling of deposits with loans as analyzed by Hodgman (1961).

In countries where also lending rates were regulated (as in Finland and Sweden), and excess demand for credit thus prevailed, the incentives for tying deposit and loan markets were especially strong. This is due to the fact that, under conditions of credit rationing, depositors can be accorded "preferential treatment" with little cost to the bank, simply by changing the allocation of scarce credit. By tying loans to deposits, the interest rate regulations could in principle be entirely circumvented, at least in a frictionless world without uncertainty, reserve requirements or capital requirements. Even with some frictions, there is probably much scope for nonprice competition under interest rate regulation.

Actually, the distortionary, or potentially distortionary, effects of the regulations became increasingly more apparent during the decades of deposit rate regulation. This can be explained by gradually accelerating inflation and rising "free" interest rates. The effects of deposit rate regulations on interest rates or other deposit terms were probably not very great during the era of "cheap money" policy which prevailed in most industrial market economies from the late 1930s to the 1950s. During that period, monetary policy aimed at pegging market interest rates at a very low level. When both money market rates and lending rates were low, the distortion caused by deposit rate regulation was small. This situation changed, however, in the 1960s, and especially in the 1970s, when market interest rates rose to double-digit levels throughout the western industrialized world. When the difference between free market interest rates and regulated deposit rates increased, the margin a bank could earn on a given stock of deposits increased. As a result, the incentives to compete for depositors using non-price means increased as well; hence the growth of distortions.

Some attempts have been made to measure tariff remissions and other forms of non-price competition for deposits in the U.S.A. Barro and Santomero (1972) report results from a survey measuring the conditionality of bank service charges on the deposit balances held on chequing accounts. According to their study, the implicit return on demand deposits (in the form of remitted service charges) is significant, but not as high as the explicit rate paid on saving accounts, for example. Over the period 1950-1968, the average implicit rate on demand deposits was 1.71 per cent, while the average explicit rate on saving accounts in the savings and loan associations was 3.57 per cent. Both rates were increasing slowly during the period investigated.

Later, Startz (1979) noted that the implicit return calculated by Barro and Santomero showed remarkable sensitivity to other interest rates. This supports the view that there was strong non-price competition for deposits. Startz also studied a broader measure of non-price competition, ie the total sensitivity of deposit service costs to the prevailing level of market interest rates. According to this measure, too, there was evidence of active non-price competition. These studies suggest that service charge rebates and costly quality improvements were actively used by U.S. banks to circumvent the prohibition of the payment of interest on demand deposits stipulated by the Glass-Steagall Act as early as in the 1950s and 1960s. According to the results obtained by Startz, changes in implicit interest did not fully offset the effects of the prohibition of the payment of interest, however. Similar results can also be found in Becker (1975) and Santomero (1979). These results contradict the assumption made by B. Klein (1974), who tested (and found some support for) a demand-for-money model in which the implicit return on deposits "fully offset" the effects of regulation.

In the U.K., nonprice competition for deposits also intensified in the 1970s and the 1980s as money market rates rose. The evidence presented in Llewellyn and Drake (1993) reveals this development very clearly. First, in 1974, British banks even ceased calculating service charges for accounts which had a balance above a certain minimum, and in 1985, they finally moved to a "free if in credit" policy, imposing no charges at all for accounts with positive balances. At about the same time, however, the first signs of reviving interest rate competition for chequable deposits emerged in Britain (see next section below).

In Germany, banks (and other deposit-taking institutions, which are important in that country) imposed no transaction charges on cheque accounts (wage and salary accounts) of private persons until the beginning of the 1970s (Vittas et al. 1988). At that time, deposit markets were already formally deregulated, however.

To the knowledge of the present author, nonprice competition for deposits has not been empirically studied in the Nordic countries. A widely held view, however, attributes the growth of banks' branch networks and the early move into salary account-based customer banking in these countries from the 1960s to the 1980s to the nonprice competition occasioned by regulation; see also Koivisto (1962). Generalizing from the American research on the subject, it may be concluded that interest rate ceilings or prohibitions did not prevent competition in banking from taking place. As long as the general level of (market) rates of interest was low, the effects of interest rate

ceilings on service pricing or on deposit quality were probably not very strong. However, as interest rates gradually rose during the era of regulated deposit rates, the nonprice aspects of deposit pricing became increasingly important while explicit pricing of transaction services diminished in importance.

3.3 The new competition

The regulated deposit rate system did not accord well with the more active interest rate policies which were adopted in many countries in the late 1960s and especially in the 1970s. When inflation accelerated, money market rates also rose. As a result, strong pressures emerged to raise deposit rates as well. These pressures were strongest in the time deposit market, but they were also felt in demand deposits. On the one hand, balances on bank accounts shrank because the rates of return offered by alternative investment vehicles increased. This was reflected not only in the erosion of the market share of deposit banks in financial intermediation, but also in the gradual reduction of the share of transaction accounts in banks' funding. On the other hand, profits (or, rather quasi-rents) derived by banks from funds obtained through the transaction deposit market increased.

As signs of distortions grew stronger, and as the general attitude of policy makers became more suspicious of the benefits of interest rate regulation, the basis for trying to preserve the market conditions created in the 1930s was progressively undermined. As a result, deposit markets were ultimately deregulated (or cartels dismantled) in many countries. This process started gradually in the U.S. in the 1970s and prevailed in both the U.S. and many European markets in the 1980s.

International surveys of the transaction account pricing systems which emerged in the 1980s after deregulation are provided by Vittas et al. (1988) and Llewellyn and Drake (1993); see also BIS (1991). Additional sources on individual countries are cited below. In general, banks have attempted to introduce more explicit charges on transactions. It is often also mentioned that the pricing of alternative payment methods has been differentiated in order to encourage customers to favour the least cost-intensive methods of payment (such as ATMs and paperless transfers). However, despite the trend towards quoting explicit prices, price policies have not vindicated the predictions of the SCH. Two observations, in particular, support this claim.

First, significant margins remain between the rates paid on transaction deposits and the rates paid on time deposits and money market instruments. This is true even in those countries where deposit rates are now completely deregulated. In many cases, a tendency towards nonlinear deposit rate schedules can be observed, with large balances earning significantly higher interest. The observed interest differentials are too large to be plausibly explained by reference to costs of holding free reserves. So, a large part of banks' operating costs continues to be financed out of interest rate margins.

Second, explicit transaction charges, although often imposed in principle, are often waived in practice or at least reduced for "good" customers, ie those who maintain sufficiently large deposit balances. So, in many cases, posted charges may not actually be applied. This clearly contradicts with the predictions of the SCH.

Some notes on individual countries follow.

In the U.S.A., the easing of deposit rate regulations started already in the 1970s with the introduction of NOW (Negotiable Order of Withdrawal) accounts in the New England states between 1972-1976. These were chequable accounts which bare interest in the same way as time deposits. In 1981, NOW accounts were introduced nationally.

Originally, NOW accounts were subject to interest rate ceilings in accordance with Regulation Q of the Federal Reserve. In 1983, however, it became possible to offer transaction accounts with no interest rate limits when new "Super NOW" accounts were allowed. These had a higher (regulated) minimum balance requirement than NOWs. After 1986, U.S. institutions were finally allowed to offer unrestricted interest rates on any NOW account. With the disappearance of the distinction between NOW and Super NOW accounts, the deregulation of the retail deposit market could be considered complete, at least as far as pricing was concerned.

After deregulation, the interest rate spread between money market rates and the new deregulated chequable deposits soon stabilized at about 2 per cent, varying however with the general level of interest rates (see Wenninger (1986); Davis, Korobow and Wenninger (1987)). The correlation of the Super NOW spread with the level of market interest rates has been found to be perhaps surprisingly strong, at least from the perspective of the SCH (see Diebold and Sharpe (1988), for example). Facing rising interest costs, banks attempted to raise the prices of payment services offered to holders of these new transaction accounts from the low levels associated with the traditional U.S. chequing account.

Turning to fees and service charges, the systems of pricing transaction services which emerged in the deregulated deposit market

were typically quite complex, often featuring monthly account maintenance fees and transaction charges, as well as minimum balance requirements which entitled the depositor to earn interest or avoid fees of both. According to several studies, almost all banks levied service charges or maintenance fees on personal and commercial chequing accounts; however, in most cases, these were waived for personal accounts if the minimum or average monthly balance was high enough. Monthly fees seem to have been slightly more frequently applied than activity charges. Only a minority of institutions levied per-cheque type charges on personal transaction accounts regardless of the account balance (cf. Rogowski (1984); Rose, Kolari and Riener (1985); and Davis and Korobow (1987) for details). The complexity of the tariffs which were developed by banks is striking, and the strong evidence of conditionality on the deposit balance makes it very difficult to explain fee structures exclusively on the basis of bank costs, as the SCH does.

In the United Kingdom, the Bank of England broke the clearing banks' long-standing interest rate cartel in 1971 by adopting a new policy favouring competition (the new policy was called Competition and Credit Control after the title of the paper in which it was announced). However, the custom of not paying interest on chequable deposits continued in Britain until the latter half of the 1980s. As a matter of fact, the deposit market moved further away from the pricing system predicted by the SCH as explicit fees became less and less important in the late 1970s and early 1980s. This has been attributed to higher money market rates, which intensified the competition for depositors conducted by non-interest-rate means.

The tradition of interest-free transaction accounts was finally broken in Britain in the late 1980s when building societies launched chequable, interest-bearing accounts for private customers. "High interest chequing accounts" first appeared in Britain in 1984. Typically, the new types of accounts offered "free banking" (ie no service charges for normal transactions), provided the account was kept in credit.

Thus it is evident that in Britain, too, some "implicit interest" continued to be paid on transaction accounts even after the onset of interest rate competition. Heffernan (1992) has measured the "interest equivalences" (essentially an implicit interest concept) of these free services from a data set extending from the latter half of 1985 to 1989. The results indicate that offering customers chequing facility for use in connection with a deposit account allows the bank to pay 0.4 to 1.2 percentage points (40 to 120 basis points) lower interest on that

account. While not very great, the effect is significant, suggesting clear cross-subsidization of payment services.

In Germany, deposit rates were deregulated in 1967. After deregulation, interest rates on time and savings deposits increased, while interest rates on cheque accounts remained at zero or very close to zero; see Deutsche Bundesbank (1967). In the 1970s explicit service charges were introduced. It is reported that, in Germany, service charges are seldom levied in practice if a large enough balance is maintained on the account (Llewellyn and Drake (1993)).

In Sweden, the informal regulation of deposit interest rates was relaxed in 1978, when the central bank announced that it did not oppose adjustments in bank deposit rates, within limits of prudence. Interest rate competition was slow to start, but by the late 1980s, Sweden was considered to be "the world leader in innovative pricing of transaction accounts, proceeding quite far in the direction of paying interest on deposit balances and levying economic charges for services and transactions" (Vittas et al. (1988)) However, even in Sweden, interest rates on transaction accounts are typically not uniform: higher rates are paid for larger balances. There have been attempts by the authorities to reduce or even eliminate some "basic" charges (cf. Hörngren (1989); Lindblom (1990)).

In Finland, the banks' long-standing deposit rate cartel became obsolete in 1989, when the tax rules concerning deposit interest were changed. However, even the new tax rules effectively limited interest competition in respect of transaction accounts, owing to tax reliefs granted for specific, regulated accounts. Significant interest rate competition in the market for personal transaction accounts started only in 1991, when tax rates on interest income were drastically reduced (Soltila and Jokinen (1991)). In the the 1990s, Finnish banks have so far offered two types of transaction accounts: tax-free accounts bearing a low, regulated interest rate and a "high-yielding" account, bearing taxable interest. Typically, the latter type of account has been associated with an interest rate schedule according to which the interest rate payable rises in line with the size of the minimum monthly balance (Halonen (1992) and (1993)).

On the service charges side, a step towards a more market-oriented approach was taken in 1985, when Finnish banks terminated their cartel agreement in regard to tariffs and fees. This did not, however, lead to reductions in service charges imposed on transaction accounts; these had been small in any case. Presumably the cartel had been mostly aimed at limiting competition in bank services which are less complementary to deposit-taking.

Subsequently, Finnish banks started to raise their transaction charges in 1988 at a time when the onset of deposit rate competition was already in sight. Another factor was the almost explosive growth in the numbers of bank-intermediated payments in the mid-1980s (see Suominen and Tarkka (1991)). By the early 1990s, all banks had a rather sophisticated tariff of charges for transaction accounts. Different prices were charged for different types of transactions, with preference given to self-service and automated activities. Often the charges were waived or reduced, however, if the average or minimum monthly balance in the account exceeded some pre-announced threshold. Since 1992, some banks have been offering "service packages", giving customers the option of choosing lower transaction-specific charges in return for a fixed monthly charge (see Suomen kuluttajaliitto (1992); Suomen Pankkiyhdistys (1992); and Siikala (1993) for details). An example of the terms of a representative private Finnish bank account is given in the appendix to this chapter.

As is shown by the brief international survey presented above, the latter half of the 1980s and the early 1990s have seen a revival of price competition in the transaction accounts market in many countries. However, the conditions under which this has happened differ in many respects from those of the classical period of banking in the 1920s and before. Banking technology has advanced dramatically because of computerization and better communications, and the use of bank accounts has become more common among the population. Moreover, one important difference is that the general level of interest rates (in money markets, for example) was much higher in the 1980s than in the beginning of the century. Hence interest rate spreads, which did not give rise to significant deposit rates in the first decades of this century, certainly did so in the 1980s. Loosely speaking, one can say that the same "degree" of deposit rate competition was thus much more conspicuous in terms of the resulting deposit rates in the "new competitive" years than in the "classical period".

The pricing systems which have emerged as a result of the revival of price competition appear to be at least as varied and complicated as in the classical period, and have not eliminated interest rate spreads from the deposit market. Banks have paid much more attention to their fee structure, but despite this it is not obvious that service charges only approximate costs (marginal or average) of transaction services. The many instances in which "good" depositors are provided with transaction services free of charge serve as cases in point.

4 Theories of deposit account pricing

4.1 The free competition model

It is useful to start the discussion of theories of deposit pricing by considering a class of models which are best labelled free competition models. The reason is that these models and the predictions which are derived from them serve as useful benchmarks with which other, "imperfecionist" models can be compared. The free competition models are concerned with the question of what would happen if deposit pricing were not regulated at all and competition in the deposit market were perfect in other respects, too, ie unfettered by monopoly or any frictions. When the "simple competitive hypothesis" of deposit pricing is put forward, it is usually done on the basis of this class of models.

The research interest in the pricing of bank deposits started in the late 1960s and was originally motivated by the then renewed interest in the foundations of monetary theory. In a pioneering contribution, Kareken (1967) criticized the standard analyses of the money supply process for "implicit theorizing". The standard analysis is, of course, the money multiplier approach based on the fractional reserve model of banking. Claiming that this approach had focused on what is only a special case by tacitly assuming a regulated demand deposit rate, Kareken set out to develop a more general money supply theory with competitive determination of the demand deposit rate.

Kareken's model is driven by convex intermediation costs which he imposes both on the lending side and the deposit side on the bank's balance sheet. The intermediation costs are specified as "real" factor costs, which are due to labour and capital services used in the bank's operations. The costs are functions of the amount of funds raised and placed in the deposit and loan markets, respectively. The competitive equilibrium is derived as the point at which the bank's marginal cost of deposits equals the marginal return on the bank's assets.

In Kareken's model, the deposit rate of interest deviates from the marginal return on the bank's assets by the amount of the marginal costs of servicing deposits. Obviously, this result rests on the assumption that the services rendered to depositors are strictly proportional in quantity to the stock of deposits. The proportionality must hold for each account. From this it follows that these services

need not be charged separately but the marginal cost can in fact be deducted from the interest paid out. Unfortunately, this assumption cannot, for obvious reasons, be accepted as more than a very crude approximation. Transaction account pricing cannot be analyzed in all of its dimensions unless the quantities of payment services and the stock of deposit balances can vary independently of each other.³

In later studies using the free competition approach, the multi-product nature of deposit accounts has been taken into consideration and pricing models have been developed which motivate separate service charges. This leads to the general result that, in a frictionless competitive equilibrium, the depositors are paid the equivalent of the marginal return on the bank's assets. Together with this goes the prediction that payment services will be priced separately according to their marginal cost. This is the essence of the Simple Competitive Hypothesis (SCH) of bank account pricing. It appears, with only minor variations, in the papers by Johnson (1968), Black (1975), Saving (1979), Fama (1980), and Fischer (1983).

The SCH pricing structure is not only a prediction, but several economists have advocated this pricing system in a normative sense, too. This argument can, of course, be ultimately reduced to the well-known welfare implications of marginal cost pricing. The normative idea that bank account pricing should ideally obey the predictions of the SCH has been called "the Johnson norm" after Harry G. Johnson, who proposed it at a time when deposit pricing was still regulated in the U.S. From a normative perspective, lower service charges and lower deposit rates than implied by the SCH would cause wasteful use of resources in the production of payment services, and an unjustified transfer of income from those customers with large balances and few transactions to those who have less money in their accounts but a lot of transactions (see Johnson (1968)).

4.2 The monopolistic approach

The most obvious alternative among the possible explanations of price-cost margins is, of course, the market power of price-making firms. In the 1970s, Klein (1971) and Monti (1972) followed this route

³ However, the assumption of the proportionality of payment services to deposit balances seems to survive in several important papers on the production microeconomics of banking, as well as in the monetary aggregation literature. This is probably the most natural way one can justify the models of Hancock (1985) and Barnett (1980), for example.

and developed the monopolistic model of deposit rate formation. Later, the model was extended by Dermine (1984) to include, among other things, the effects of credit risk and "menu costs" of adjusting deposit rates.

Basically, the "Monti-Klein" approach to deposit pricing is analogous to the standard neoclassical model of the profit-maximizing firm facing a downward-sloping demand function for its output. The deposit market is assumed to be characterized by product differentiation, and this is thought to give rise to smooth demand functions for deposits. Since the demand elasticity in the deposit market is assumed to be finite, the bank has some leeway as regards the deposit rates it offers to pay. Higher deposit rates attract more deposits, but are of course more costly to the bank. Thus, there is a tradeoff between scale and the deposit rate margin in the model.

The monopolistic models of deposit banking explain the determination of the interest margin on deposits by the so-called "inverse elasticity rule" familiar from monopoly theory. Briefly explained, the analysis proceeds as follows. There is a competitive securities market which determines the exogenous opportunity cost for funds. Denote this by r . Then, if the demand for deposits function is given by $i = i(D)$, the profit of the bank from its deposit taking activity is maximized when the marginal revenue from additional deposits equals the marginal opportunity cost of funds, ie

$$r = i + i'(D) \cdot D \quad (11)$$

or, if we define the own interest elasticity of deposit demand as e ,

$$i = r / (1 + 1/e). \quad (12)$$

In the Klein-Monti model, the deposit rate decision appears quite independent of the bank's asset management in general and its lending decisions in particular. Dermine (1984) extended the Klein-Monti model and incorporated credit risk in the analysis. He pointed out that the "posted" deposit rates are not directly relevant for the (profit-maximizing) equilibrium conditions when there is uncertainty about the bank's ability to fulfil its obligations. Rather, the equilibrium conditions such as (12) define the expected returns from deposits, and the posted returns must be somewhat higher, depending on the probability of bank failure.

More recently, Hannan (1991) has proposed a variant of the Klein-Monti model as a theoretical foundation for the application of the Structure-Conduct-Performance paradigm to empirical studies of the banking industry. Studies which have interpreted interest rate spreads as evidence of market power are indeed numerous (see eg Gilbert (1984)).

It is clear that, as an explanation of the existence of interest spreads, the monopolistic model has a lot of appeal, not least because of its simplicity. However, the monopolistic model as used in the literature is not fully satisfactory, at least as far as the treatment of payment services is considered. This is important because it is very likely that the product differentiation on which the monopolistic models are based is essentially due to the characteristics of the services component of the deposit relationship. This can be seen by considering different types of deposits. If product differentiation and imperfect competition are important factors in the deposit markets, their effect should be strongest in the transaction deposit markets, while the CD market, where services play a very small role, is presumably much more competitive.

In most of the monopolistic models, payment services are not explicitly taken into consideration (Monti, Dermine, Hannan, for example). Thus these models offer no theory of service charges either. They can be interpreted as making the implicit assumption that transaction services (all deposit services, in fact) are produced in constant proportion to deposit balances. Thus, they can be considered as variants of the Kareken (1967) model, with some market power added. Under the strict proportionality assumption, service charges do not have a meaningful, independent role in the bank's tariff policy (as was mentioned in the previous section). If less strict complementarity in consumption or production were allowed, separate pricing of deposits and transaction services would be possible in a meaningful sense, and the deposit rate decision would become much more complicated.

M. Klein (1974) includes a discussion of the determination of service charges in an unregulated banking environment. Klein claims that in the absence of the prohibition of the payment of interest, banks would pay a profit-maximizing, monopolistically determined rate of return on deposits, and set average service charges (per dollar of deposits) equal to the average cost of providing these services. This claim is valid, however, only in the special case where there is perfect competition in payment services. The assumption of a perfectly competitive services market is convenient, but it is an extreme assumption and does not fit very well in a model which is otherwise

built on product differentiation and imperfect competition (for deposit funds).

Another problem in the monopolistic models is the assumption of linear pricing. All of the published models have assumed *a priori* that bank "pricing" (which in these models means interest rate determination) is linear. When scrutinized more closely, however, linear pricing policy seems likely to be suboptimal. It is obvious from consideration of the nature of the deposit relationship that the services which banks render to the holders of transaction accounts are not easily transferable. An enterprise selling such products under imperfect competition generally gains from nonlinear pricing. This is indeed the case in banking, as the development of account pricing practices in the 1980s shows.

To conclude this critical assessment of the monopolistic models of bank account pricing, doubt must be cast on the type of imperfection assumed in these models. While the possibility of imperfect competition certainly cannot be ruled out in the analysis of banking, the type of imperfection which is assumed in the Klein-Monti and Hannan models (deposit differentiation in the sense of quality differences) is not necessarily the most plausible one. Below, the other alternative of switching costs is argued for (see Chapter 5 of this study).

4.3 The implicit interest question

An important issue which became topical during the period of deposit rate regulation was whether banks circumvented the prohibition of the payment of interest on demand deposits, partly or entirely, by paying "implicit interest". A few ways in which used implicit interest can be paid were mentioned in Sections 2 and 3.2 above. The method most frequently referred to is the provision of payment services to depositors either free of charge or at prices below cost. A number of studies have sought to develop a theory of implicit interest, or measure it. In the early studies at least, the motivation came mainly from the wish to improve empirical models of the demand for money by measuring the "own rate of return" on money as accurately as possible. Some of the empirical studies were discussed briefly in Section 3.2. We now turn to the theoretical problems in this area.

Implicit interest literature studies the equilibrium which emerges in the bank deposit market if price competition for deposits is repressed by regulation, cartel, or by a tax system which discriminates against explicit interest payments and favours implicit interest. The problem

appears in its simplest form if one assumes that depositors demand payment services which in the unregulated (nondistorted) equilibrium would give rise to a flow of service charges. In such a situation, explicit interest could be replaced by remissions of service charges, made conditional on deposit balances. Then, the interest rate ceiling could in principle be completely ineffective, if the underlying service charge flow is large enough. The pioneering empirical study of Barro and Santomero (1972) was based on this idea. In an extremely simplified form, this "theory" of implicit interest is also suggested in Klein's (1971) study on deposit rate determination as a way in which that study could be applied to the then regulated U.S. deposit market.

In more recent studies, the implicit interest problem has been conceived in a slightly different fashion, however. There are essentially two alternative specifications. The first of them might be labelled "the quantity-setting model". It does not consider the determination of service charges; it is merely assumed that there is a flow of free services which is delivered to depositors in a quantity which is set by the bank. In Startz (1983), for example, deposit rate regulation induces the bank to offer implicit interest in the form of free services rendered in some given proportion to deposit balances and determined by competition.

Formally, this way of conceiving the implicit interest problem is virtually equivalent to the general models of non-price competition by nonbank firms (Stigler (1968); Schmalensee (1976)). The most important result from the general models of nonprice competition is that, if free entry is not allowed, nonprice competition can be either more or less costly to the banks (in terms of equilibrium profits) than normal price competition. This depends on the shape (concavity of convexity) of the cost functions of the competing firms. In the case of free entry, profits are competed away by definition, and so this will be of no consequence for profitability. Thus the free entry, nonprice competition equilibrium differs from the standard price competition equilibrium in terms of the equilibrium outputs of the goods. The regulated good is usually underproduced, the "free" good is overproduced and customer welfare is reduced compared with the unregulated equilibrium.

The second way of looking at the implicit interest problem has been to use what might be labelled "the price-setting model". The banks' decision is taken with respect to the price of transaction services, and the depositor makes an independent decision concerning the use of transaction services. The important thing is that the ratio of transaction services to deposit balances is beyond the control of the bank. The central papers are Mitchell (1979, 1988) and also Merris

(1985).⁴ They analyze the determination of the service charge per cheque in a monopolistic situation where the interest rate on chequable deposits is regulated. The downward-sloping demands for bank services are either specified *ad hoc* (Mitchell (1979)) or derived from an extended Baumol–Tobin money demand framework (Merris (1985); Mitchell (1988)).

The results obtained by Mitchell (and discussed by Merris) suggest that the service charge can, in principle, be either an increasing or decreasing function of the regulated deposit rate. Explicit and implicit interest may thus be either complements or substitutes, in contrast with the perfectly competitive equilibrium where an increase in the deposit rate must reduce the amount of "implicit interest". The possibility of complementarity between explicit and implicit interest arises, if the (negative) effect of the deposit rate on the ratio of transactions to deposit balances is large enough. An increase in the deposit rate will then cause a decrease in the relative benefit of service underpricing, and the bank may find it optimal to offset this by lowering the service charge.

Besides outright regulation of deposit interest, there may be other reasons for nonprice competition in the deposit markets. In particular, one must take into account the discriminatory effects of the tax system. In most fiscal systems, interest income from bank deposits is taxable. Now, depending on the tax treatment of service charges, this may give rise to a distortion which favours nonprice competition (implicit interest) and punishes explicit interest. This distortion arises if a) the benefit of "underpriced" bank services is not taxable and b) service charges are not deductible as expenses in taxation. It may then be in the mutual interest of the bank and the depositor to pay part of the deposit interest in kind, ie in the form of "free" services, although depositors, in the absence of taxes, would of course prefer to be paid in money.

The conditions a) and b) above are both necessary conditions for the existence of the tax distortion. So neutrality with respect to explicit and implicit interest payments is achieved if either a) or b) is not fulfilled. In most countries, both conditions are, in fact, fulfilled in personal taxation. However, in corporate (profit) taxation, bank charges

⁴ A related analysis is presented by Shaffer (1985). He studies cases where depositor heterogeneity (with respect to the velocity of deposit balances) may give rise to a competitive equilibrium where one depositor group is effectively subsidizing the transactions of the other group.

are typically deductible as expenses, resulting in tax neutrality with respect to pricing in that segment of the deposit market.⁵

Previous studies which have analyzed the effects of the tax distortion have used the "quantity-setting model", ie the assumption that the bank supplies "free" services in some fixed proportion to deposits. Orgler and Taggart (1983) analyzed the determination of the deposit rate when this proportion is exogenously determined outside the model. Assuming that depositors have a valuation function $V(s)$, where s is the amount of free services, the equilibrium deposit rate can be solved from the equation relating the "total" after-tax return i on deposits to the after-tax return $r(1-t)$ which the depositor might obtain from the securities market:

$$i(1-t) + V(s) = r(1-t). \quad (13)$$

Solving for i yields

$$i = r - \frac{V(s)}{(1-t)}. \quad (14)$$

This implies that the deregulated deposit rate should be a decreasing function of the tax rate.

The problem with the Orgler and Taggart model is of course the exogeneity of s , the free services content of deposits. The question of how s is determined in equilibrium was taken up by Walsh (1983). He considered the issue of implicit interest in a more general model, where the proportion of free services to deposits is endogenously determined. In Walsh's model, the pricing decisions of banks are taken with respect to two parameters: the deposit rate and the ratio of free services to deposit balances. Walsh shows that, in this kind of framework, competing banks generally end up paying both implicit and explicit interest to their depositors. Further, the level of implicit interest is independent of the level of (nondeposit) interest rates.

In the course of the present study, it will be shown that relaxing the restrictive assumptions used by Walsh will lead to considerably different results (see Essay 2 below). The limitations of the Walsh model are essentially due to the "quantity-setting model" of the tariff

⁵ In principle at least, the taxation of value added may add some further complications (distortions), depending on the integration of the banking sector in the VAT system. To my knowledge, these complications have not, however, been treated in the bank pricing literature.

function. The way in which the tariff is specified is therefore not innocuous in this case.

4.4 The multiproduct approach

The multiproduct approach to the pricing of transaction accounts can best be defined as simultaneous analysis of the determination of both deposit rates and service charges, taking into account that these services are neither produced nor used in fixed proportions but can be varied by the "consumer" (the depositor) independently of each other.

According to this definition, many of the free competition models of bank pricing reviewed in Section 4.1 above also belong to the multiproduct approach. Obviously, however, the multiproduct approach does not require the assumptions of perfectly competitive, frictionless markets. Indeed, if the multiproduct approach is preferred, while recognizing the stylized facts of deposit pricing, richer models than the free competition model should be developed. Some important attempts to do that are surveyed in this section.

An early piece which clearly uses the multiproduct approach and tries to explain the observed pricing patterns in the deposit markets (the market for transaction accounts in particular) is a chapter by Paul Cootner in the book *Retail Banking in the Electronic Age* (Baxter, Cootner and Scott (1977), pp. 16–35). That chapter may well be the first analysis of the deposit pricing problem from the perspective of price discrimination.

Cootner's model is built on assumptions of free entry and heterogenous customers. Banks offer their customers deposit accounts which include some payment services and a deposit facility. All customers use the same amount of payment services (this could be just a normalization), but are heterogenous with respect to the amount of deposits they are willing to hold. The customers are also heterogenous with respect to the price they are willing to pay for bank services. This is possible because customers are not perfectly mobile, but are instead to some extent attached to particular bank service points.⁶ Differences in reservation prices create an incentive for price discrimination.

The driving assumption in Cootner's model is that the stock of deposits held by a customer (at the given deposit rate) will perfectly

⁶ Cootner suggests that this could be justified by a spatial model or other ways of specifying access costs. Owing to the informal treatment of the issue, the model is somewhat vague on this point, however.

reveal his reservation price for bank services. The correlation in the population of reservation prices with the demand for deposit balances is due to a single underlying variable, such as the opportunity cost of time (real wage). Moreover, Cootner assumes that the deposit balances are an increasing, linear function of reservation prices. Under this assumption, the banks are able to capture all of the consumer surpluses through setting the deposit interest rate at a level suitably below the "market rate of interest". The result is equivalent to perfect (or first-degree) price discrimination.

In Cootner's model, the bank obtains average revenues which are greater than marginal cost. This implies that, in a free entry equilibrium, banks must be operating on the declining part of their average cost curve. Otherwise, profits would be generated. The number of banks in the market (ie the service capacity of the industry) is assumed to adjust until a zero-profit equilibrium is achieved.

The price system which emerges in Cootner's model is the perfectly discriminating two-part tariff. In the terminology of two-part tariffs, the interest rate spread constitutes the "entry fee" component of the tariff, the burden of which happens to coincide with the customers' willingness to pay for bank services in excess of the marginal costs of service. The service charge can be interpreted as the variable part of the tariff, and it will coincide with the cost of serving the marginal (most mobile) customers.

One of the limitations of Cootner's path-breaking model, in addition to the occasional vagueness of the exposition by modern standards, is that the incentive compatibility issue is not tackled, although incentive compatibility is absolutely central to tariff design. Instead, the problem is avoided by implicitly assuming that the demands for deposits are exogenous so that incentives do not matter. Another related limitation is the assumption of a perfect linear correlation between deposit holdings and reservation prices. However, the model clearly demonstrates that some of the stylized facts of deposit pricing may be explained by price discrimination. Unfortunately, this direction of research has received only limited attention so far.

The Cootner model was ahead of its time in the sense that it treated both the deposit rate and service charges as free, market-determined prices. Yet, much of the subsequent work on the theory of bank service pricing continued to assume an exogenous interest rate on transaction accounts: the contributions of Mitchell (1979, 1988), Shaffer (1984) and Merris (1985), discussed above in the context of the "implicit interest" question, belong to this category.

The multiproduct approach reappeared in the literature in the work of Whitesell (1988, 1989, 1992). Whitesell considers the price setting problem of a bank which is in a monopoly position *vis-à-vis* its depositors. The deposit rates and service charges are assumed to be parametric constants. The demand elasticities which determine the bank's optimal pricing policy relate to the competition from alternative means of payment, either currency, or both currency and credit cards.

Whitesell presents a model in which individuals have an exogenous size distribution of payments and use different means of payment to execute payments, depending on the size of the payment in question. In the monopolistic equilibrium, the bank sets service charges below service costs. The reason is that, by providing a cheap service, the bank is able to attract deposits from funds which would otherwise be kept in the form of currency. This is profitable, because the optimal deposit rate may be below the rate the bank earns on reinvesting the funds.

Whitesell's model is somewhat problematic, because the extreme simplicity of it generally rules out the existence of interior solutions with respect to optimal prices. As explained in Whitesell (1992), the model requires exogenous limits to the deposit rate and/or the service charge to explain the kind of price structures observed in practice. Another problem with the model is the assumed linearity of the tariff, which is generally suboptimal for a monopolistic firm. As is well known, it is optimal for a discriminating monopolist to price according to marginal cost, if fixed (intramarginal) charges can be used to extract all of the consumer surplus from customers (this kind of first degree discrimination is always possible when customers have identical demands and the products sold are nontransferable).

5 Developing the theory of deposit pricing

In the present study, three ways to rationalize the observed special features of bank deposit pricing are examined in the context of explicit microeconomic models. The first explanation, which is based on viewing deposit accounts as offering insurance against transaction risk, is apparently original to the present study (though the first exposition of it was in Tarkka (1989)). The other two approaches considered have also received some attention in the earlier literature. These are the tax explanation considered by Walsh (1983), in particular, and the

price discrimination/imperfect competition explanation of Cootner (1977) and Whitesell (1992). Here, these approaches are employed without the restrictive assumptions concerning the structure of the tariff (linearity) which have been made in the previous studies. Instead, the general nonlinear form of deposit rate and service charge schedules is allowed, thus modifying the previous result in many important respects.

5.1 Essay I: The risk sharing explanation

The purpose of the first essay is to analyze the pricing of transaction deposit accounts as arrangements for pooling transaction cost risk among depositors. In particular, the aim is to demonstrate that the risk sharing aspect of deposit pricing can be used to explain why banks cover part of the costs arising from the provision of transaction services indirectly, by levying interest margins on deposits, rather than directly, through service charges on transactions.

The general framework of analysis is as follows. Depositors are subject to idiosyncratic risk concerning the volume (number) or transactions they need to make. Transactions are costly; if they are made through a bank account (which is the assumption made in this paper), the bank incurs the transaction costs in the first place. The question is how are these transaction services priced? The market for deposit accounts is assumed to be competitive. According to the Simple Competitive Hypothesis outlined above, several authors have predicted that the equilibrium interest rate and service charge should be set equal to the money market rate and the marginal service cost, respectively. It turns out, however, that this may not be the result when depositors are risk averse.

Two variants of the deposit pricing problem are considered. First, a simple benchmark case is presented in which the risks are exogenous and the deposit balance is sufficient information for estimating the expected transaction costs. In that case, perfect insurance against transaction cost uncertainty is possible and the optimal price system is very simple: transactions services are provided free of charge and the deposit rate is adjusted so that the costs of producing transactions services can be covered from the spread between deposit and lending rates.⁷

⁷ The case with exogenous risks has been analyzed earlier in Tarkka (1989), under the restrictive assumption of a linear tariff structure.

After analyzing the case with exogenous risks, the issue of moral hazard is taken up. Moral hazard arises because, with full insurance against transaction cost risk, the customers' incentives to avoid excessive (uneconomical) use of banks' transaction services are reduced. One may doubt whether the price elasticity of the use of transaction services use is very great. However, the moral hazard argument is important because H.G. Johnson's famous case for explicit pricing of transactions was based on it, as too is the banks' case justifying their attempts at explicit pricing of transactions.

Unfortunately, the optimal risk sharing contract under moral hazard is very difficult to derive except under the simplest of assumptions concerning the nature of uncertainty involved (see Mirrlees (1974) and Jewitt (1988), for example). A unique optimal contract may not exist at all; moreover, the standard first-order approach of finding out what the optimal nonlinear contract looks like is applicable only under rather strict conditions. For that reason, the case with moral hazard is solved in this study under the assumption that the tariff is linear, consisting only of a uniform deposit rate and a uniform service charge.

It turns out that the presence of moral hazard may justify the use of positive service charges; transaction cost risk is then shared between the bank and the depositors. The part of the risk borne by depositors is, quite naturally, the higher, the stronger is the moral hazard effect on the demand for transaction services. Since the optimal service charge will generally cover only part of the costs of producing transaction services, the competitive equilibrium may feature a positive interest rate spread to cover the remaining transaction costs.

The risk sharing aspect of deposit contracts has been well established in the literature since Diamond and Dybvig (1983). The growing body of work which started from Diamond and Dybvig analyzes the problems of making long-term investments from funds collected as short-term liabilities. A major issue here is the stability of banks as institutions. However, these studies have implications for deposit rate determination, too. From the price determination point of view, the aim of the Diamond-Dybvig tradition can be seen as to analyze the pricing of the early withdrawal option, and the uncertainty which is being pooled has been in the time preferences and the saving behaviour of individuals. The result of the analysis has typically been an interest rate spread which corresponds to the costs of holding reserves to meet the withdrawal requirements of the depositors who turn out to be short term. Thus, the tradeoff between service charges and interest spreads has fallen outside the scope of the earlier studies.

The contribution of the present study is to modify the risk sharing approach to analyzing banks as producers of transaction services. The uncertainty pooled by banks is not with respect to deposit maturities, as in Diamond and Dybvig, but with respect to transaction costs caused by depositors' (exogenous) real activity.

5.2 Essay II: The tax explanation

This essay deals with the effects of income taxation on competitive bank deposit markets. The analysis extends the model by Walsh (1983) by allowing a completely free tariff structure, instead of the simple one assumed by Walsh (a uniform deposit rate and a given ratio of free services to deposits).⁸

As in Walsh, the starting point of the essay is the observation that, in many fiscal systems, explicit interest is taxable but the benefit of free services provided by a bank to its depositors is not. While this rule may be odd (probably no other class of debtors besides banks enjoys a similar advantage), it has important implications for the pricing of bank services, as well as for the resulting demand for them. The analysis is performed with a "production" model of banking in which the services produced by the bank enter the depositor's utility function as complements to other consumption.

It turns out that it will never be optimal in the model to levy service charges on a depositor who is earning positive interest on his deposit. The rational (tax-minimizing) tariff will be nonlinear. It has the property that, for some ratio of bank services to deposit balances, both the service charge and the deposit rate will be zero. With lower deposits in relation to services, there will be service charges but no deposit interest, and vice versa.

What kind of the equilibrium in the deposit market is generated by the model depends on how the tax rate on interest income is related to the reserve requirements imposed on banks. Probably the most relevant case arises when the tax rate on interest income is higher than the reserve requirement. In the resulting equilibrium, the depositor will then choose a deposit balance such that neither service charges nor deposit interest are paid. Other equilibria are less interesting special cases. The equilibrium thus resembles the so-called London system of deposit pricing, which is based on the interest-free minimum balance

⁸ The determination of an optimal linear tariff, consisting of the deposit rate and a uniform service charge, is analyzed in Tarkka (1988).

requirement, entitling the depositor to use bank services without charge.

The results are, then, different from the price system assumed by Walsh. In particular, the requirement of a minimum balance for interest to be paid turns out to be a feature of the efficient pricing system which is neglected in the Walsh model. Therefore, the result of Walsh that the bank will generally pay some deposit interest is seen to be an artifact created by the assumption of a uniform deposit rate.

It is shown in the essay that the price system which is optimal for banks actually includes a hidden subsidy on the production of transaction services by banks. This subsidy is increasing in the tax rate on interest income, but decreasing in the reserve requirement. Moreover, as long as the reserve requirement is below the tax rate, increasing the reserve requirement improves the allocational efficiency of the equilibrium (in the sense that the marginal opportunity costs of services are equalized for the bank and the depositor). This is in sharp contrast with the conventional idea that reserve requirements, taken separately, are harmful from the efficiency point of view.

Finally, the essay includes an adaptation of the model to the traditional Finnish system in which the interest on bank deposits is tax-free if the interest rate is below some prescribed minimum. It is shown that this system of allowing some tax-free interest actually increases the distortionary tax subsidy. The higher is the ceiling set for tax-free interest rates, relative to market rates, the greater is the subsidy on bank services.

As a caveat, it should be noted that the kind of tax effects which are analyzed here are not present in corporate taxation, where bank service charges are deductible in taxation. The deductibility removes the asymmetry between explicit and implicit interest rates. Therefore, the tax-based model cannot be used to explain the pricing of corporate transaction accounts. This, among other factors, clearly implies that attention should be paid to other aspects of the deposit pricing problem as well.

5.3 Essay III: The price discrimination explanation

This essay applies the theory of nonlinear pricing of multiple products to the problem of determining the terms of chequable bank accounts. The profit-maximizing interest rate and service charge schedule is characterized for a monopoly bank with a heterogeneous clientele. The

characteristics of the customers are not directly observable, or at least are not allowed to be used as a basis for price discrimination. From the bank's point of view, this is a tariff design problem in order to achieve best second degree price discrimination.

The customers' demand for the two types of banking service considered (deposits and transaction services) is derived from a cost-minimization problem. Depositor-customers minimize a cost function which describes how large internal "liquidity costs" they incur in earning and spending a given income flow (cf. Feenstra (1987) for a discussion on the concept of liquidity costs). The customers are heterogeneous with respect to the income level. Liquidity costs can be reduced by holding deposits and by using transaction services produced by the bank, and the structure of the problem is such that the demands for deposits and bank services are increasing functions of the individuals' income level.

This way of formulating the problem is more general than the earlier imperfect competition models of bank account pricing. The most important advantage is that the possibility of nonlinear tariffs is taken into account. This is crucial because uniform pricing is generally not optimal under imperfect competition when demands are not perfectly transferable between customers. Another aspect which makes this model more general than the previous ones is the use of the explicit multiproduct framework, which is very flexible and may cover a variety of different transaction technologies, and institutional arrangements, which determine the demand for bank services.

In accordance with the properties of optimal nonlinear tariffs, the analysis shows that the marginal prices paid by the customers having the highest income should equal marginal costs. In the usual language of bank account pricing, this prediction of the model means that the largest depositors should be paid the market rate of interest (at the margin); similarly, the marginal service charges charged to the heaviest users of payment services should be equal to the marginal cost of producing the service. For those customers with smaller demands for the two bank services, the marginal prices deviate from marginal costs. Since all customers are charged on the basis of the same nonlinear tariff, the "distorted" marginal prices paid by small customers also have an intramarginal effect on the prices paid by the biggest customers.

Virtually all previous work on nonlinear multiproduct pricing has been confined to cases where all products are priced above their respective marginal costs (in the interior points of the tariff, that is). The neglect of more general cases has been due to the complexity of the sufficient conditions required for the tariffs to be globally optimal

(see Wilson (1993)). However, this essay presents conditions under which the optimal tariff is cross-subsidizing in the sense that the marginal price of one service is below the marginal cost, while the other is priced above marginal cost. These conditions are required if the theory of optimal price discrimination is to be useful in explaining the interest rate spreads on deposits and the underpricing of transaction services.

A necessary condition for cross subsidies to emerge is that the increase in income has an asymmetric effect on the marginal benefits from the use of the two bank services. The payment services will be "underpriced" in the profit-maximizing optimum if the demand for both bank services is increasing in the income of the customer, but income has a negative partial effect on the marginal benefit of the payment services. It is further shown in the paper that these effects are indeed present in a slightly extended Baumol–Tobin type model of the demand for money.

The intuition behind the cross-subsidizing tariff is that the bank wants to extract the maximum possible amount of consumer surplus from its best customers, without losing the poorer ones. The way to accomplish this is to make the product valued highly by the best customers relatively expensive (deposits), while making the product valued highly by poor ones relatively cheap (transactions).

5.4 Essay IV: Switching costs

In the chapter summarized in the previous section, the price discrimination problem was treated from the point of view of a monopoly bank. The fourth theoretical essay of this study is an extension of the price discrimination model of tariff design to the case where competition limits the bank's ability to extract consumer surplus from its deposit customers.

Competition is assumed to be imperfect. (Under perfect competition, price discrimination is generally uninteresting; see Shaffer (1984), however). The imperfect competition model which is adopted is the switching cost model of Klemperer (1987), with some modifications. It is assumed that depositors can only bank at one bank at a time. Further, the depositors incur a once-and-for-all sunk cost (switching cost) each time they change their banking relationship to a new bank. The switching costs are completely revealed to the bank by the depositor's conduct (the deposit balances and the use of payment services). For simplicity, the liquidity cost function of the depositors is assumed to be quadratic.

The main result of the analysis presented in this note is that the main results, concerning the possibility of cross subsidies in particular, are also valid in this slightly more general (and, it is hoped, more realistic) specification of the market structure. Some minor differences emerge, however. These mostly concern the way in which the switching cost function (specifying the switching costs as a function of the income level) regulates the shape of the optimal tariff. In particular, it turns out that the direction of cross subsidies (if these are present) may be reversed along the income distribution of customers. Another difference between this and the monopoly model is that the "best" customers no longer necessarily pay marginal prices which are equal to marginal costs.

6 Conclusions

The various results which have been developed and presented in this study were summarized in the previous section, chapter by chapter. Taken together, the overall conclusion is that several, not obviously unrealistic factors may explain the deviations of the actual deposit pricing practices from the benchmark case provided by the "Simple Competitive Hypothesis" of Johnson (1968) and others. These special factors – risk sharing, taxation, and price discrimination – shaping the pricing systems used in banking should not be considered as competing theories, but rather as complementary forces. The problem is not, then, how to discriminate between these explanations, but rather how to integrate them into a tractable general model.

The fact that several explanations can be found for the apparent anomalies observed in deposit pricing undermines the SCH model of banking. In the light of the present study, it is by no means unnatural for deposit rates on transaction accounts to be significantly below the opportunity cost of funds to the bank or for transaction services may be priced below the marginal cost of producing them, even in the absence of interest rate regulation.

Second, the power of the "Johnson Norm", which is the normative aspect of the SCH, may not be all that strong. Generally, allocative efficiency in situations involving moral hazard does not imply marginal cost pricing (see eg Newbery and Stiglitz (1981), ch. 15). Here, the risk sharing model of deposit pricing shows how deviation from the SCH pricing system may be compatible with a perfectly competitive market with no external frictions of any kind. Actually, in that model, it is the presence of service charges rather than their

absence which is a sign of poorly functioning markets. This is because service charges can be seen as a way of reducing moral hazard problems in the use of banks' transaction services, by means of a kind of co-insurance arrangement.

As shown in the last two essays of this study, price discrimination may explain some of the salient features of deposit pricing – particularly tariff nonlinearities and even cross subsidization. If used to earn monopoly profits, these devices are of course dubious from the welfare point of view. However, as the literature on increasing returns to scale has demonstrated, some (Ramseyian) price discrimination may belong to the second-best optimum in markets where there are significant fixed costs of production, but government subsidies to finance them are ruled out. Moreover, as noted by Wilson (1993), the most efficient way to implement the necessary price discrimination is often by nonlinear tariffs. In this study, fixed costs are not considered. Generally, however, it is clear that the welfare implications of price discrimination in deposit pricing depend on whether it is used to earn monopoly profits or just to finance fixed costs in the banking industry.

From the point of view of public policy, it would be interesting to be able to judge whether imperfect competition or tax asymmetries distort the bank deposit market to a significant degree. The distortions caused by the tax system may be assessed by a study of the structure of the tax system as such, with the help of the kind of analysis developed in essay II of this study. On the other hand, the degree of monopolistic conduct in banking is clearly an empirical problem, requiring careful measurement. The issue may become increasingly topical in the future, as technological innovations alter the prerequisites for efficient competition for deposit customers. In any case, empirical studies on the deposit market should take into account the complicated, multiproduct nature of that market and the price systems prevailing in it.

References

- Barnett, W. (1979) **The User Cost of Money**. Economics Letters.
- Barnett, W. (1980) **Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory**. Journal of Econometrics.
- Barro, R. – Santomero, A. (1972) **Household Money Holdings and the Demand Deposit Rate**. Journal of Money, Credit, and Banking 4, 397–413.
- Baxter, W. – Cootner, P. – Scott, K. (1977) **Retail Banking in the Electronic Age. The Law and Economics of Electronic Funds Transfer**. Allanheld & Osmun & Co., Montclair, N.J.
- Becker, W. (1975) **The Determinants of the U.S. Currency-Demand Deposit Ratio**. Journal of Finance 30, 57–74.
- Bingham, T.R.G. (1985) **Banking and Monetary Policy**. OECD, Paris.
- BIS (1991) **Statistics on Payment Systems in Eleven Developed Countries**. Bank for International Settlements, Basle.
- Black, F. (1975) **Bank Funds Management in an Efficient Market**. Journal of Financial Economics 2, 323–329.
- Cowen, T. – Kroszner, R. (1987) **The Development of New Monetary Economics**. Journal of Political Economy 95, 567–590.
- Davis, R. – Korobow, L. (1987) **The Pricing of Consumer Deposit Products – The Non-rate Dimensions**. Federal Reserve Bank of New York, Quarterly Review, Vol. 11, No. 4, 14–18.
- Davis, R. – Korobow, L. – Wenninger, J. (1987) **Bankers on Pricing Consumer Deposits**. Federal Reserve Bank of New York, Quarterly Review, Vol. 11, No. 4, 6–13.
- Dermine, J. (1984) **Pricing Policies of Financial Intermediaries**. Springer-Verlag, Heidelberg.
- Deutsche Bundesbank (1967) **Erste Ergebnisse der Erhebungen über Bankzinsen**. Monatsberichte der Deutschen Bundesbank (Oktober 1967), 46–51.
- Diamond, D. – Dybvig, P. (1983) **Bank Runs, Deposit Insurance, and Liquidity**. Journal of Political Economy 91, 401–419.
- Diebold, F. – Sharpe, S. (1988) **Post-Deregulation Deposit Rate Pricing: The Multivariate Dynamics**. Federal Reserve Board, Finance and Economics Discussion Series No. 8.
- Fama, E. (1980) **Banking in the Theory of Finance**. Journal of Monetary Economics 6, 39–57.

- Feenstra, R. (1986) **Functional Equivalence between Liquidity Costs and the Utility of Money.** *Journal of Monetary Economics*, 17, 271–291.
- Fischer, S. (1983) **A Framework for Monetary and Banking Analysis.** *Economic Journal* (Conference Papers Supplement) 1–16.
- FRB (1927) **Federal Reserve Bulletin**, July 1927.
- Gardener, E. – Molyneux, P. (1990) **Changes in Western European Banking.** An International Banker's Guide. Routledge, London.
- Gilbart, J. – Sykes, E. (1911) **Gilbart on Banking.** The History, Principles and Practice of Banking. G. Bell & Sons, Ltd. London.
- Gilbert, R. (1984) **Studies of Bank Market Structure and Competition: A Review and Evaluation.** *Journal of Money, Credit, and Banking* 16, 617–645.
- Halonen, U. (1992) **Yksityishenkilöille tarkoitettut tilit.** Lokakuu 1992. Suomen Pankkiyhdistys, Helsinki.
- Halonen, U. (1993) **Yksityishenkilöille tarkoitettut tilit.** Lokakuu 1993. Suomen Pankkiyhdistys, Helsinki.
- Hancock, D. (1985) **The Financial Firm: Production with Monetary and Nonmonetary Goods.** *Journal of Political Economy* 93, 859–880.
- Hannan, T. (1991) **Foundations of the Structure-Conduct-Performance Paradigm in Banking.** *Journal of Money, Credit, and Banking*.
- Heffernan, S. (1992) **A Computation of Interest Equivalences for Nonprice Characteristics of Bank Products.** *Journal of Money, Credit, and Banking*, Vol. 24, No. 2, 161–172.
- Hicks, J. (1935) **A Suggestion for Simplifying the Theory of Money.** *Economica* 2 (N.S.), 1–19.
- Hodgman, D. (1961) **The Deposit Relationship and Commercial Bank Investment Behavior.** *Review of Economics and Statistics*.
- Holdsworth, J. (1921) **Money and Banking.** D. Appleton & Co, New York.
- Hörngren, L. (1989) **Kostnadsutveckling och konkurrens i banksektorn.** Statens offentliga utredningar SOU 1989:16, Finansdepartementet, Stockholm.
- Jewitt, I. (1988) **Justifying the First-Order Approach to Principal-Agent Models.** *Econometrica*, Vol. 56, No. 5.
- Johnson, H. (1968) **Problems of Efficiency in Monetary Management.** *Journal of Political Economy*, 76, 971–990.
- Kareken, J. (1967) **Commercial Banks and the Supply of Money: A Market-Determinend Demand Deposit Rate.** *Federal Reserve Bulletin* (October 1967).

- Keeley, M. – Zimmerman, G. (1986) **Deposit Rate Deregulation and the Demand for Transactions Media**. Federal Reserve Bank of San Francisco Economic Review, (Summer 1986), 47–61.
- Klein, B. (1974) **Competitive Interest Payments on Banks' Deposits and the Long Run Demand for Money**. American Economic Review Vol. 64 No. 4, 931–945.
- Klein, M. (1971) **A Theory of the Banking Firm**. Journal of Money, Credit, and Banking 3, 205–218.
- Klein, M. (1974) **Deposit Interest Prohibition, Transaction Costs, and Payment Patterns: A Theoretical Analysis**. Metroeconomica, Gennaio-Dicembre 1974.
- Klemperer, P. (1987) **Markets with Consumer Switching Costs**. Quarterly Journal of Economics 101, 375–394.
- Koivisto, M. (1962) **Pankkikilpailusta Suomessa**. Kansantaloudeellinen Aikakauskirja 58, 234–247.
- Korpisaari, P. (1920) **Suomen pankit. Niiden kehitys, rakenne ja toimintamuodot**. Taloustieteellisiä tutkimuksia No. XXV, Kansantaloudellinen Yhdistys, Helsinki.
- Korpisaari, P. (1930) **Raha ja pankit**. WSOY, Porvoo.
- Lindblom, Ted (1990) **Pricing Strategies on Payment Services**. Institute of European Finance Research Papers in Banking and Finance RP 90/16. University College of North Wales, Bangor.
- Llewellyn, D. – Drake, L. (1993) **The Economics of Bank Charges for Personal Customers**. Loughborough University Banking Centre Research Monograph No 9.
- Merris, R. (1985) **Explicit Interest and Demand Deposit Service Charges**. Journal of Money, Credit, and Banking 17, 528–533.
- Mirrlees, J. (1974) **Notes on Welfare Economics, Information and Uncertainty**, in Balch, M., McFadden, D. and Wu, S. (eds.): Essays on Economic Behavior under Uncertainty. North-Holland, Amsterdam.
- Mitchell, D. (1979) **Explicit and Implicit Demand Deposit Interest: Substitutes or Complements from the Bank's Point of View?** Journal of Money, Credit, and Banking 11, 182–191.
- Mitchell, D. (1988) **Explicit Interest and Demand Deposit Service Charges**. A Comment. Journal of Money, Credit, and Banking 20, 270–274.
- Monti, M. (1972) **Deposit, Credit, and Interest Rate Determination under Alternative Bank Objective Functions**, in Shell, K. and Szegö, G. (eds.): Mathematical Methods in Investment and Finance. North-Holland, Amsterdam.
- Newbery, D. – Stiglitz, I. (1981) **The Theory of Commodity Price Stabilization**. Oxford University Press, Oxford.
- Obst, G. (1923) **Das Bankgeschäft**. Band I. C.E. Poeschel Verlag, Stuttgart.

- Orgler, Y. – Taggart, R. (1983) **Implications of Corporate Capital Structure Theory for Banking Institutions.** *Journal of Money, Credit, and Banking* Vol. 15. No. 2.
- Philips, L. (1983) **The Economics of Price Discrimination.** Cambridge University Press, Cambridge.
- Rae, G. (1910) **The Country Banker. His Clients, Cares, and Work.** John Murray, London (1st edition 1885).
- Roberts, K. (1979) **Welfare Considerations of Nonlinear Pricing.** *Economic Journal* 89, 66–83.
- Rogowski, R. (1984) **Pricing the Money Market Deposit and Super-NOW accounts in 1983.** *Journal of Bank Research* (Summer 1984).
- Rose, P. – Kolari, J. – Riener, K. (1985) **A National Survey Study of Bank Services and Prices Arrayed by Price and Structure.** *Journal of Bank Research* (Summer 1985).
- Santomero, A. (1979) **The Role of Transactions Costs and Rates of Return in the Demand Deposit Decision.** *Journal of Monetary Economics*, Vol. 5, 343–364.
- Saving, T. (1979) **Money Supply Theory with Competitively Determined Deposit Rates and Activity Charges.** *Journal of Money, Credit, and Banking* Vol. 11, No. 1.
- SÅS (1923) **Statistisk årsbok för Sverige.** Stockholm, 1923.
- Schmalensee (1976) **A Model of Promotional Competition in Oligopoly.** *Review of Economic Studies* 43, 493–507.
- Shaffer, S. (1984) **Cross-Subsidization in Checking Accounts. A Note.** *Journal of Money, Credit, and Banking* 16, 100–109.
- Siikala, K. (1992) **Kotitalouksien yleisimmät pankkipalvelumaksut. Lokakuu 1992.** Suomen Pankkiyhdistys, Helsinki.
- Siikala, K. (1993) **Kotitalouksien yleisimmät pankkipalvelumaksut. Maaliskuu 1993.** Suomen Pankkiyhdistys, Helsinki.
- Soltila, H. – Jokinen, H. (1991) **Effects of the Withholding Tax.** *Bank of Finland Bulletin*, Vol. 65, No. 10 (October 1991).
- Startz, R. (1979) **Implicit Interest on Demand Deposits.** *Journal of Monetary Economics*.
- Startz, R. (1983) **Competition and Interest Rate Ceilings in Commercial Banking.** *Quarterly Journal of Economics*, (May 1983), 255–265.
- Stigler, G. (1968) **Price and Non-Price Competition.** *Journal of Political Economy*.
- Suomen Kuluttajaliitto (1992) **Pankkipalvelujen hintavertailuraportti. Maaliskuu 1992.** Suomen Kuluttajaliitto ry, Helsinki.

- Suomen Pankkiyhdistys (1992) **Kotitalouksien yleisimmät pankkipalvelumaksut.** Suomen Pankkiyhdistys, Helsinki.
- Suominen, M. – Tarkka, J. (1991) **Trends in the Demand for and Pricing of Banking Services in Finland.** Bank of Finland Bulletin Vol. 65, No. 10.
- Tarkka, J. (1988) **Tax on Interest and the Competitive Pricing of Deposits and Bank Services.** Bank of Finland Research Department Working Papers No. 3/88.
- Tarkka, J. (1989) **Competitive Deposit Rates and Bank Service Charges.** Bank of Finland Discussion Papers No. 29/89.
- Vittas, D. – Frazer, P. – Metaxas-Vittas, T. (1988) **The Retail Banking Revolution. An International Perspective.** Lafferty Publications, London.
- Walsh, C. (1983) **Taxation of Interest Income, Deregulation and the Banking Industry.** Journal of Finance, Vol. XXVII, No. 5.
- Weninger, J. (1986) **The Responsiveness of Interest Rate Spreads and Deposit Flows to Changes in Market Rates.** Federal Reserve Bank of New York Quarterly Review (Autumn 1986), 1–10.
- Whitesell, W. (1988) **Deposit Pricing, Bank Market Structure, and Welfare with Cost-Minimizing Consumers.** Federal Reserve Board, Finance and Economics Discussion Series No. 48.
- Whitesell, W. (1989) **The Demand for Currency versus Debitable Accounts.** Journal of Money, Credit, and Banking, Vol. 21, No. 2.
- Whitesell, W. (1992) **Deposit Banks and the Market for Payment Media.** Journal of Money, Credit, and Banking, Vol. 24, No. 4.
- Wilson, R. (1993) **Nonlinear Pricing.** Oxford University Press, New York.

Appendix

The deposit account terms offered by a Finnish bank in 1995

The fees and interest rates pertaining to a representative retail transactions bank account in Finland (the MONEX account of Postipankki Oy, as of April, 1995)

1. The interest rates

The bank offers two options, a tax free account, and an account earning higher, but taxable interest.

A. Tax free MONEX account

The deposit interest rate on the tax free account is the maximum tax free rate on transactions deposits as stipulated by the income tax law; since November 15, 1993 this is 2 per cent per annum. The interest is calculated on the lowest balance each month.

B. The taxable MONEX account

The deposit interest rate on the taxable account is graduated according to the minimum monthly balance. The interest rate is linked to the prime rate of the bank, according to the following schedule:

The deposit balance	Interest rate	Presently
FIM 1- 10 000	Prime - 5.0 %	2.5 %
FIM 10 001- 25 000	Prime - 4.0 %	3.5 %
FIM 25 001- 50 000	Prime - 3.0 %	4.5 %
FIM 50 001- 100 000	Prime - 2.0 %	5.5 %
FIM 100 000 or more	Prime - 1.5 %	6.0 %

2. Monthly fees

The account holder may agree on a service package entitling her or him to 30 debit transactions monthly without further service charges. The monthly fee on this service package is as follows:

For the tax free account	FIM 10 per month
For the taxable account	FIM 20 per month

If the minimum monthly balance exceeds FIM 15 000, the monthly fee is reduced by 50 per cent.

If there are less than 5 debit transactions, the monthly fee is reduced by 50 per cent.

If both conditions above apply, the fee is waived entirely.

3. Activity-based service charges

For the account holders who have agreed on the service package, first 30 debit transactions are free of charge.

Transactions in excess of 30, are charged FIM 1.00 per transaction.

For account holders without the service package, debit transactions are charged according to a posted tariff. This includes, inter alia, the following charges:

Paper-based transfers	FIM 2.00
Ditto, with receipt	FIM 4.00
Transfers by ATM	FIM 1.50
Ditto, given with ATM of another bank	FIM 3.00
Transfer orders over the counter	FIM 5.00
Cheques	FIM 2.00

Risk Sharing in the Pricing of Payment Services

Implications for Bank Service Charges and Deposit Interest

Contents

1	Introduction	63
2	Uncertain transaction costs and risk sharing	65
3	The case with moral hazard	71
	3.1 The transactions framework	72
	3.2 The depositor's problem	73
	3.3 The bank's pricing problem	75
4	A case of heterogeneous depositors	79
5	Discussion	81
	References	84
	Appendix	
	Service charges and depositor behaviour	86



1 Introduction

The banking industry has traditionally covered a large part of its operating costs from net interest income derived from the spread between deposit and lending rates. This reflects the common practice of underpricing various services provided to customers, especially depositors. In recent years, banks in many countries have attempted to move towards "direct pricing", meaning more reliance on activity-based service charges. As a result, the pricing of transaction deposits,¹ in particular, has gradually become more sophisticated in many banking markets, but interest rate spreads and underpricing of transaction services have remain ubiquitous (see Vittas et al. (1988), for an international survey). Both the special features of traditional bank service pricing and the recent changes call for a better understanding of the nature of the transaction deposit markets.

The problem has in fact both a positive and a normative dimension. From a positive point of view, one may ask why banks often seem to "cross-subsidize" transaction services provided to depositors (these services include cheque processing, giro transfers, card-activated payments etc.). Put in another way, it can be asked why depositors seem to accept "implicit interest" in the form of free or underpriced services instead of requiring explicit, pecuniary return for their funds. From a normative point of view, in turn, the question is one of what kind of changes, if any, are advisable in the public policies influencing bank pricing, given what the observed pricing patterns may reveal about the operation of the market for bank services.

The standard conjecture in the literature is that, in perfectly competitive, frictionless markets, competition among banks would establish an equilibrium in which the explicit interest on transaction deposits would equal the marginal (opportunity) cost of funds to the bank – perhaps the money market rate. Parallel to this, transactions services provided by banks would be priced according to their marginal factor cost (see Fischer (1983) and Saving (1979), for example). This kind of bank pricing system has been labelled "the Johnson norm", after Harry G. Johnson who advocated it from the efficiency perspective (Johnson, 1968). According to this view, any deviations from the "Johnson norm" type of pricing must be caused by

¹ "Transaction deposits" is a generic term for all deposits which can be used as means of payment. These include cheque accounts as well as various other types of salary accounts with names that vary between countries.

regulation or other market imperfections disturbing the competitive price mechanism.

It is of course obvious that, if the authorities impose ceilings on deposit interest rates, competition for deposits will lead to "implicit interest" (Startz (1979, 1983)). Other types of intervention have also been considered: Walsh (1983) has pointed out that, if explicit interest is taxable, it may be optimal for banks to substitute implicit for explicit interest even in a competitive environment without deposit rate ceilings. The tax argument is generalized further by Tarkka (1992). In the tax-based models, the marginal tax rate of interest income is equated with the marginal efficiency loss of rewarding depositors in kind (with free, non-marketable services).

The purpose of this paper is to present an alternative explanation of the stylized facts of deposit pricing. This explanation does not rely on distortions caused by regulation, taxes, or less than perfect competition. Instead, it is shown that implicit interest may serve as a device which reduces the exposure of depositors to the uncertainty regarding their transaction needs and the costs generated by these needs. If banks are risk-neutral, or at least able to pool away the transactions uncertainty faced by individual depositors, even the competitive equilibrium may well include "below-cost" pricing of transactions services and significant interest rate spreads that compensate for the revenue foregone in the transactions service business.

Demand deposit contracts have previously been analyzed from the risk sharing perspective by Diamond and Dybvig (1983), Smith (1984) and Jacklin (1987). These contributions focus on the "early withdrawal option" characteristic of the deposits, resulting from the banks' capacity to perform maturity transformation. Essentially, the maturity transformation analysis provides a rigorous analysis of the effect of reserve holding costs on deposit interest rates. The resource costs of payment services and the pricing of these services have been almost entirely overlooked in the risk sharing literature, however. In Tarkka (1989), the sharing of transaction cost risk is used to explain the viability of implicit interest in competitive deposit markets. However, that analysis disregards moral hazard. In the present study, this limitation is overcome by applying the principal-agent framework to the deposit pricing problem.

The problem at hand can be seen as a part of a fundamental issue in monetary theory concerning the determination of the rate of return on money in a free competitive equilibrium. The question as to why people voluntarily hold low-yielding, monetary-like assets was presented by Hicks (1935) and it has been a subject of controversy in

recent years as well. According to the widely quoted "legal restrictions hypothesis", equilibria in which money yields a lower rate of return than other assets is possible only because government intervention restricts free competition in the supply of means of payment (see Black (1970) and Wallace (1983)). According to the "transactions cost hypothesis", by contrast, various accounting and administrative costs may be used to explain why assets which are inferior to others in terms of their yield are voluntarily held and used as means of exchange (White (1987)). The present study contributes to the legal restrictions controversy by suggesting that the uncertainty of transaction costs may well be the basic reason why a low-yielding means of payment may be part of a competitive equilibrium.

The structure of the essay is as follows. In Section 2, a simple benchmark model is considered in which competition forces banks to provide transactions services "free of charge" and to cover the costs involved by lowering the rate paid on deposits (indirect pricing). In Section 3, the model is generalized by allowing moral hazard effects in the demand for transaction services, which explains the mixed use of direct and indirect pricing of transaction services. The results on optimal prices are derived by assuming representative agents; the problem of heterogeneous depositors is taken up in Section 4, where some sufficient conditions are given for the results of the previous section to hold also in the case where there is heterogeneity. Section 5 summarizes and discusses the results.

2 Uncertain transaction costs and risk sharing

We consider a competitive market for transaction accounts. These accounts provide deposits with a cheque, giro, or electronic transfer facility so that the account may be used as a store for transactions balances. Such an account can be viewed as vehicle for joint delivery of two distinct bank services, ie depository and transactions services. Obviously, the customer uses the depository service whenever he keeps some funds on the account. Transactions services, on the other hand, are used whenever the funds on the account are used as a means of payment. To execute payments, the bank performs operations such as cheque clearings or payment transfers.

The demand for bank services is derived from a very simple transactions framework, which is a variant of the "cash in advance"

models. We consider a single planning period, so confining the analysis to static equilibria. There is a large number of customers. During the period of analysis, each customer spends a given amount of money. The customers are different from each other with respect to the amount of money they spend, however. The key feature of the model is that the number of transactions in which the money is spent is random. This randomness is idiosyncratic in nature. In order to focus on the effects of uncertain transaction costs, other sources of uncertainty are disregarded. So, for instance, the value of income which is spent during each period is assumed to be known by the customers at the beginning of the period.

More precisely, the representative customer receives an endowment Y at the beginning of each period, holds it in the form of deposits during the period and spends it at the end of the period. Spending takes place in a random number of transactions, each paid separately. The uncertainty regarding the number of transactions is modelled by assuming that the average size of the transactions is a random variable v . Let us denote the demand for deposit balances by D and the number of transactions required by N . The customer's demand for bank services is then given by the equations

$$D = Y \tag{1}$$

$$N = v \cdot Y \tag{2}$$

$$\text{Prob}(v \leq V) = F(v). \tag{3}$$

The size-of-transactions variable v is independently and identically distributed across customers. In this section the probability distribution $F(v)$ is taken to be exogenous. Moral hazard is therefore not present. This assumption will be relaxed in a later section of this essay, however.

The customer chooses between different banks by maximizing the expected value of a convex utility function

$$W = E(U(Y - P)), \tag{4}$$

where E is the expected value operator and P denotes the net payments from the depositor to the bank, ie service charges net of

deposit interest. If service charges are greater than deposit interest payments, P is positive; negative values are also possible, of course. Generally, banks will determine P as function of the depositor's characteristics and conduct. Functions of this kind, which may be nonlinear, are often called tariff functions (see Wilson (1993)).

Here we assume that banks can make the tariff P conditional on the use of bank services only. Thus, no external signals can be used as a basis of bonuses or discounts. This implies that the pricing policies of banks can be characterized by tariff functions of the type

$$P = P(M, N). \quad (5)$$

The partial derivatives of the tariff function define the marginal interest rate on deposits i_m and the marginal service charge s_m on transactions in the following way:

$$i_m(D, N) = -\frac{\partial P(D, N)}{\partial D} \quad (6a)$$

$$s_m(D, N) = \frac{\partial P(D, N)}{\partial N}. \quad (6b)$$

The exclusion of the customers' characteristics from the tariff function could imply serious informational problems. In the present model, however, the consumers' characteristics are completely revealed by their conduct and reflected in quantities which can enter the tariff function. More precisely, the banks can condition their tariffs on an exact indicator D of the customer's characteristic Y , even though the latter is "hidden" in principle. Therefore, in the end the deposit market is not distorted by information asymmetry.

Competition in the deposit market is assumed to operate as follows. In the first stage, each bank posts a tariff function $P(D, N)$, thus fixing the terms it offers to its depositors. In designing this function, the banks have only aggregative information on the depositors. On the basis of the distribution of v , the depositors compare the banks, choosing the bank which promises the highest expected utility W .

In this kind of market, with free and costless entry, competition ensures that the only tariffs which can survive in the market are those which give the depositors the highest possible utility, subject to the

constraint that the expected profit from each deposit constraint must be zero. By symmetry, of course, all banks post the same tariff function in the equilibrium. This function can be derived by maximizing W subject to the break-even constraint.

Now we turn to describe the nature of the break-even constraint. The profit of a representative bank from a representative deposit relationship is

$$\pi = P + r \cdot D - c \cdot N, \quad (7)$$

where c is the unit cost of transaction services and r is the rate of return the bank earns on funds. Both c and r are assumed to be known constants.

Competition ensures that the $E(\pi) = 0$. So, in equilibrium,

$$E(P) = c \cdot E(N) - r \cdot E(D). \quad (8)$$

Note that $E(N) = E(v) \cdot E(Y)$ and $E(D) = E(Y)$. So, in principle,

$$E(P) = [c \cdot E(v) - r] \cdot E(Y). \quad (9)$$

The bank does not observe Y directly and is unable to use it as an argument of its tariff function. However, the present model has the convenient property that $D = Y$ and consequently a customer's Y can be inferred from his deposits. Thus, any equilibrium tariff must satisfy the following property:

$$E(P) = [c \cdot E(v) - r] \cdot D. \quad (10)$$

The actual form of the equilibrium tariff $P(D, N)$ can now be found by maximizing the utility of the representative customer subject to the constraint (10). Formally, this problem can be presented in the form of the following programme:

$$\max_T W = \int f(v)U(Y-P)dv \quad (11)$$

$$\text{s.t. } \int f(v)[cv-r]Y dv = \int f(v)P dv$$

which is equivalent to

$$\max_P W = \int f(v)\{U(Y-P) + \lambda[P - (cv-r)Y]\}dv. \quad (12)$$

Now, the first-order necessary condition for the solution to this programme is simply

$$\frac{\partial U(Y-P)}{\partial P} = -\lambda, \quad (13)$$

where λ is the Lagrange multiplier associated with the zero-profit constraint.

The result has a simple, but crucial interpretation. In this simple case without moral hazard, the optimal deposit contract makes the depositor's marginal utility a constant, ie independent of the transactions cost uncertainty. This is because

$$\frac{\partial U(Y-P)}{\partial Y} = -\frac{\partial U(Y-P)}{\partial P} = \lambda. \quad (14)$$

At the optimum, insurance is thus complete. In particular,

$$P(D,N) = E[P(D,N)], \quad (15)$$

ie the tariff is based only on certain variables. By (10), the equilibrium tariff must be the following:

$$P(D, N) = [c \cdot E(v) - r] \cdot D. \quad (16)$$

Now, due to the linearity of the tariff function just derived, the marginal deposit rate and the marginal service charge are constants. They are defined as

$$i_m = -\frac{\partial P(D, N)}{\partial D} = r - cE(v) \quad (17)$$

$$s_m = \frac{\partial P(D, N)}{\partial N} = 0. \quad (18)$$

The equilibrium deposit rate equals the banks' opportunity cost of funds, less the cost (per unit of deposited funds) of providing the depositor with the required transaction services. The service charge is zero.

These results are actually straightforward implications of the basic theory of competitive insurance markets without moral hazard as developed by Borch and others. The simplicity of the optimal tariff results from the fact that, in this model, the parameters of the probability distribution of transaction costs are completely revealed by the customers' deposit balances. By pricing the transaction account exclusively on the basis of deposit balances, ie through the interest rate spread, the bank is able to provide complete and fairly priced insurance to all deposit customers.

This result is interesting because it constitutes an example in which "free" transaction services are not necessarily due to any frictions or distortions in the price mechanism. Rather, in the simple model just presented, the underpricing of transaction services and the accompanying interest rate spread result from a first-best, competitive insurance arrangement. While obviously not proving anything about actual deposit markets, this demonstrates that, in banking, too, "the user pays principle" and efficiency should not be casually equated.

There are two obvious caveats, however. Both have to do with distortions which may prevent the first-best optimum such as that described above from being realized. First, what happens if the deposit balances do not accurately reveal the probability distribution of transaction costs? Does the problem of adverse selection arise in that case? Second, what if the use of transactions services is endogenous, so that the "free" services characteristic of the complete insurance

solution induces the customer to increase the number of transactions? How is the optimal pricing system changed by the ensuing moral hazard problem? These questions will be addressed in the next section.

3 The case with moral hazard

We now turn to the problem of extending the simple model of the previous section to cover cases where there is a greater role for endogenous behaviour. Essentially, there are two changes to the model. First, the simplistic, mechanical identity between the average deposit balances and the depositor's expenditure is relaxed; second, the expected number of transactions required for a given expenditure flow is made dependent on the depositor's behaviour, and hence on the incentives provided by the bank. With such extensions, the risk sharing model can be used to analyze more complex deposit pricing schemes than the one described above. In particular, it turns out that the levying of a service charge on transaction services can now be explained.

The analysis will be carried out under the restriction that the banks apply a linear tariff. Hence, the representative customer faces a given deposit interest rate and a given service charge (fee) for each transaction. The banks' tariff design problem then simplifies to the task of determining the parameters i (the deposit rate) and s (the service charge) of the tariff function

$$P = s \cdot N - i \cdot D. \quad (19)$$

The linear tariff is not rare in banking, although more complex tariffs have been gaining in popularity (see Vittas (1988), for example). Still, the main reasons for analyzing the linear case here are tractability and robustness. As is well known, only very weak results can be derived from the theory of (nonlinear) optimal risk sharing contracts, except in special cases or unless extremely stylized models are used. Often, almost all that can be said is that an optimal contract must provide partial insurance to the risk averse party. More seriously, even the existence of a well-defined optimum contract requires rather stringent conditions as regards the stochastic specification of the problem (cf.

Mirrlees (1974), Holmström (1979), Rogerson (1985), Jewitt (1988)).² The linear case, by contrast, is applicable to a wider set of problems (see Varian (1980), for an example).

Another simplifying assumption is that the analysis will be carried out using the representative agent approach. Hence, the consequences of customer heterogeneity are mostly disregarded. This issue is, however, discussed briefly in a later section.

3.1 The transactions framework

The simplest way to introduce some flexibility into the transactions framework is to allow the consumer-depositor to choose between two alternative means of payment, eg deposits and currency. We adopt this approach, introducing a choice parameter θ which governs the fraction of expenditure θY which is spent through bank deposits. The remaining part of the expenditure flow $(1-\theta)Y$ is assumed to be spent through currency. Retaining the cash-in-advance framework in other respects, the demand for deposits can now be written as

$$D = \theta \cdot Y. \quad (20)$$

The number of payment transactions is assumed to have an additive random component:

$$N = k \cdot \theta^\rho \cdot Y + e. \quad (21)$$

It is expected that $\rho > 1$, implying that if the depositor increases the share of his expenditure which is paid with deposits, the number of transactions increases relatively faster than the deposit holdings. This phenomenon could result from the fact that the average size of currency transactions is smaller than the average size of deposit

² On the other hand, linear contracts (compensation schemes) can be shown to be optimal in certain cases. Holmström and Milgrom (1987) show that this is the case in a principal-agent model in which the agent's utility is exponential and the compensation is paid at the end of a given time interval, according to the number of realizations of different states during the interval. This result could provide an avenue for extending the present analysis to the multiperiod case.

transactions.³ Hence, channelling more of the expenditure through the bank account will reduce the average size of the transaction.

The stochastic term e in the transactions equation is assumed to have a zero mean and a well-behaved frequency distribution. It should be noted that, in the present formulation, the number of payment transactions is approximated by a continuous variable. This does not, however, appear to have any important consequences in terms of the results obtained below.

In this specification, the use of transaction services is endogenous, and the expected transaction costs cannot be unambiguously estimated from the stock of deposit balances. In these respects, the formulation is more general than the simple benchmark case of Section 2.

3.2 The depositor's problem

The individual utility functions of depositors are written in a fashion which has become common in the principal-agent (optimal contract) literature:

$$W = U(Y - P) + q(\theta). \quad (22)$$

The utility function consists of two additive terms. The first one includes the utility derived from income, net of charges T to the bank. The function $U(\cdot)$ is assumed to be continuous, twice differentiable and to exhibit the following standard properties:

$$U'(Y-P) > 0 \quad (\text{monotonicity})$$

$$U''(Y-P) < 0 \quad (\text{strict concavity})$$

$$U'''(Y-P) \geq 0 \quad (\text{nonincreasing absolute risk aversion})$$

where ' and '' and ''' denote first, second and third derivatives, respectively. The concavity of the utility function with respect to $Y-P$ implies risk aversion, which plays a crucial part in the model. The assumption that absolute risk aversion is nonincreasing is fairly standard in the literature on decision making under uncertainty.

³ Whitesell (1989, 1992) has analyzed models of bank customer behaviour which have these properties.

The second term in (23) captures the direct utility effects of using the transaction services of a bank. The larger is the part of one's expenditure that is paid from the account, the less one's own effort is required for the execution of the transactions. The $g(\cdot)$ function is assumed to be continuous, twice differentiable and to exhibit the following properties:

$$g'(\theta) > 0 \quad \text{and}$$

$$g''(\theta) < 0.$$

Under the above assumptions on the form of the tariff and on the specification of the transactions technology, we have

$$P = s(k \cdot \theta^p \cdot Y + e) - i \cdot \theta \cdot Y. \quad (23)$$

Observing this, the problem of the representative depositor becomes:

$$\max_{\theta} E(W) = E[U(Y - s \cdot e - s \cdot k \theta^p \cdot Y + i \cdot \theta \cdot Y)] + g(\theta). \quad (24)$$

From this, the following first-order necessary condition for maximum can be derived:

$$E(W') = -E[U'(Y - s \cdot e - s \cdot k \theta^p \cdot Y + i \cdot \theta \cdot Y)](s \cdot k p \theta^{p-1} - i) \cdot Y + g'(\theta) = 0, \quad (25)$$

or, for short,

$$E(U') \cdot (s \cdot k p \theta^{p-1} - i) \cdot Y = g'(\theta). \quad (26)$$

The first-order condition implicitly defines θ as a function of s , i , and other parameters of the depositor's problem, provided that the problem is well-behaved; the latter may be checked by inspecting the concavity of the objective function with respect to θ . The depositor's second-order condition, which ensures concavity, may be expressed as the following inequality:

$$\frac{\partial E(W')}{\partial \theta} = A - B + q''(\theta) < 0, \quad (27)$$

where $A = E(U'') \cdot (sk\rho\theta^{\rho-1} - i)^2 \cdot Y^2$

$$B = E(U') \cdot sYk\rho(\rho-1)\theta^{\rho-2}.$$

In the inequality, the term A is known to be negative by the concavity of the utility function, except in the special case where $i = sk\rho\theta^{\rho-1}$ and $A = 0$ will result. The term B is unambiguously positive, if s is positive; and finally, the term $g''(\theta)$ is negative by assumption. All in all, to ensure the validity of the second-order condition, it suffices to show that s is not negative. Below, it will be established that in a competitive deposit market, it will always be the case that $s \geq 0$ and thus the second-order condition holds.

3.3 The bank's pricing problem

With a linear tariff, the bank's profit from the representative deposit relationship can be written as

$$\pi = (s - c) \cdot N + (r - i) \cdot D. \quad (28)$$

Taking into account the transactions framework specified above, this can be written as

$$\pi = (s - c) \cdot (k\theta^\rho Y + e) + (r - i) \cdot \theta Y. \quad (29)$$

As in Section 2 above, the assumptions of perfect competition and free entry with the same information available to all banks exclude the possibility that the expected profit from any deposit relationship could be different from zero. Therefore, the pricing parameters in the tariff function must be such that $E(\pi) = 0$, which implies

$$i = r + (s - c) \cdot k \cdot \theta^{\rho-1}. \quad (30)$$

In the competitive deposit market, the bank must offer depositors the tariff which maximizes each depositor's *ex ante* utility subject to the break-even constraint, as given by expression (30).

Formally, the bank's problem is then

$$\max_P E(W) = E[U(Y - s \cdot e - s \cdot k \theta^{\rho} \cdot Y + i \cdot \theta \cdot Y)] + g(\theta), \quad (31)$$

$$\text{so that } i = r - (c - s) \cdot k \cdot \theta^{\rho-1} \quad (\text{break-even constraint})$$

$$\text{and } \theta = \theta(s) \quad (\text{incentive constraint}).$$

The function $\theta = \theta(s)$ is implicitly defined by the depositor's first order optimality condition, taking into account the break-even constraint to eliminate i .⁴

Substituting the break-even constraint and the incentive constraint into the maximand (ie into the expression for $E(W)$ in (31)), the bank's tariff problem may be converted to an unconstrained maximization problem:

$$\max_s E(W) = E[U(Y - s e - Y \cdot c \cdot (\theta(s))^{\rho} + Y \cdot r \cdot \theta(s))] + g(\theta(s)). \quad (32)$$

The first-order necessary condition for maximum now reads

$$E[U'(Y - P)(r \cdot Y \cdot \theta' - c \cdot Y \rho \theta^{\rho-1} \theta' - e)] + g'(\theta) \cdot \theta' = 0. \quad (33)$$

Here the symbol θ' denotes the derivative of θ with respect to s when i is allowed to change in a way specified by the bank's break-even constraint. It is possible to show that $\theta' < 0$, meaning that higher service charges must lead to a reduction both in deposit holdings and in the expected number of transactions. This is proven in Appendix 1.

The bank's first-order condition may be simplified further to yield

⁴ To be accurate, the function $\theta(s)$ also includes Y and ρ among its arguments. These depositor characteristics are, however, dropped here to simplify the notation. This should not cause any loss of clarity in the present section where the representative agent assumption (identical depositors) is used.

$$E(U' \cdot e) - E(U') \cdot (r - c\rho\theta^{p-1}) \cdot Y \cdot \theta' = g' \cdot \theta'. \quad (34)$$

The depositor's first-order condition (26) may be multiplied by θ' to yield the equation

$$E(U') \cdot (sk\rho\theta^{p-1} - i) \cdot Y \cdot \theta' = g' \cdot \theta'. \quad (35)$$

Next, this equation (35) may be combined with the bank's first-order condition (34), enabling us to eliminate $g' \cdot \theta'$:

$$E(U' \cdot e) - E(U') \cdot (r - c\rho k\theta^{p-1}) \cdot Y \cdot \theta' = E(U') \cdot (sk\rho\theta^{p-1} - i) \cdot Y \cdot \theta'. \quad (36)$$

This yields, upon simplification,

$$E(U' \cdot e) = E(U') \cdot [r - i - (c - s) \cdot k\rho\theta^{p-1}] \cdot Y \cdot \theta'. \quad (37)$$

Finally, substituting the break-even constraint for $(c-s)$, we obtain

$$E(U' \cdot e) = E(U') \cdot (r - i)(1 - \rho) \cdot Y \cdot \theta'. \quad (38)$$

This formula defines the optimal interest rate spread as predicted by the model. If the model works properly, the spread should, of course, be positive. To find out whether this is the case, it is useful to note that $E(U' \cdot e) = \text{Cov}(U', e)$, which is due to the assumption that $E(e) = 0$. Now, $\partial(U')/\partial e = -s \cdot U''$, which has the same sign as s and is zero if $s = 0$. On the basis of this result, it is easy to check that the optimal interest rate spread is indeed generally positive. This may be done by showing that a zero spread $r-i = 0$ and a negative spread $r-i < 0$ would both lead to a contradiction.

Consider the possibility of a zero spread first. Under the assumption that the utility function $U(Y-P)$ and the incentive function $\theta(s)$ are continuous and differentiable throughout, a zero spread would imply that the right-hand side of equation (38) would be equal to zero. By the break-even constraint, a zero spread also implies $s = c$. However, the left-hand side of the equation is zero only if $s = 0$, which leads to a contradiction.

A negative spread would make the right-hand side of the equation negative. With a negative spread, $s = c - (r-i)/\theta^{p-1} > c$. However, $s > c$

is not possible, because the left-hand side is negative only if $s < 0$. Thus the negative spread also leads to a contradiction. It can be concluded that $(r-i) > 0$.

The result that the equilibrium interest rate spread is positive implies that the equilibrium service charge s is lower than the marginal cost of transaction services, c . This is easily seen from the break-even constraint, eg in the form $s-c = -(r-i)/\theta^{\rho-1}$. The service charge will generally be positive, however. Negative service charges would make the covariance term on the left-hand side of (38) negative, leading to a contradiction with the positive right-hand side. Finally, zero service charges would imply $E(U' \cdot e) = 0$, which is also in contradiction with the positive right-hand side of (38). It can thus be concluded that $c > s > 0$.

Collecting the above results, we have the following expressions for the interest rate spread and the price-cost margin on bank services, respectively:

$$\begin{aligned} (r-i) &= \frac{E(U' \cdot e)}{E(U')(1-\rho) \cdot Y \cdot \theta'} \\ (s-c) &= \frac{-E(U' \cdot e) \cdot \theta^{1-\rho}}{E(U')(1-\rho) \cdot Y \cdot \theta'} \end{aligned} \tag{39}$$

Some insights into the effects of various obstacles for risk sharing in deposit markets can be derived from the formula for the optimal interest rate spread. One pertains to the role of moral hazard. This is captured by the derivative θ' , which measures the effects of the bank's pricing policy on depositor behaviour. The greater is the derivative in absolute terms, the bigger the obstacle moral hazard should constitute for the insurance function of deposit banks. This intuition is confirmed by inspection of formula (38). If θ' is increased in absolute terms, then the interest rate spread must diminish. As the interest rate spread is reduced, the service charge s will rise towards the marginal cost c of transaction services. This increases the covariance term on the left-hand side of the equation, thus reinforcing the effect of θ' on $(r-i)$.

Another insight can be developed on the effects of the properties of the transaction framework. Call the ratio N/D the "velocity" of the deposit account. Our assumptions on the transaction framework imply that we can write $E(N/D) = \theta^{\rho-1}$. We see that the parameter ρ governs the extent to which depositors' behaviour influences the (expected) velocity of deposits. We assumed above that $\rho > 1$. The larger ρ , the more influence depositors have on velocity; if ρ decreases,

approaching 1, the velocity will become totally exogenous at the limit. If other factors in equation (38) are held constant, changes in ρ will have an effect on the interest rate spread and the service charge. The more behaviour influences velocity, the smaller is the interest rate spread (and the higher is the service charge). The more exogenous the velocity becomes, the larger is the interest rate spread and the fuller is the insurance given against liquidity cost uncertainty.

4 A case of heterogeneous depositors

The analysis presented in the previous section was conducted within the representative agent framework. This approach is obviously limited in the sense that the results may not be valid in situations where agents are in fact heterogeneous with respect to some relevant characteristic or characteristics. The interesting kind of heterogeneity is with respect to a variable which cannot be used as a basis for first-degree price discrimination, either because the customer-specific characteristics are unobservable or because price discrimination is not allowed on the basis of the characteristics in question.

In the context of the deposit market, differences in income levels probably constitute an important source of heterogeneity, and it would be desirable to be able to analyze the effects of this heterogeneity. The task is nontrivial, however, since the analysis of optimal contracts is extremely difficult in situations combining moral hazard (hidden action) and adverse selection (hidden information) problems. It is therefore necessary to focus on some tractable special cases and try to derive some general insights from them.

Here, one simple case is considered where the heterogeneity of depositors with respect to income can be allowed without disturbing the results derived in the previous section. The case amounts to making a specific assumption on the type of the utility function and the nature of transactions uncertainty.

First, assume that the depositors' utility functions are logarithmic with respect to income (this case obviously satisfies the assumptions used in Section 3 on the shape of the utility function):

$$W = \text{Log}(Y - P) + g(\theta). \quad (40)$$

Second, assume that the transactions uncertainty is proportional to income. This means that the customer-specific random variable e can be presented as $e = w \cdot Y$, where w is a random variable which has a finite support and is i.i.d across all depositors.

Under these assumptions, we have

$$U'(Y - P) = \frac{1}{(Y - P)} \quad (41)$$

and

$$P = (s \cdot k\theta^p + s \cdot w - i \cdot \theta) \cdot Y \quad (42)$$

so that

$$U' = \frac{1}{[Y \cdot (1 - s \cdot k\theta^p - s \cdot w + i \cdot \theta)]} \quad (43)$$

Note that for the marginal utility to be well defined for all w , the variable w must never obtain too large positive values. This puts some limits on the support of that variable.

It is easy to check that the depositor's first-order condition (26) now becomes

$$(s \cdot k p \theta^{p-1} - i) E \left[\frac{1}{(1 - s \cdot k \theta^p + s \cdot w - i \cdot \theta)} \right] = g'(\theta). \quad (44)$$

The choice or "effort" parameter θ is seen to be determined independently of Y . In equilibrium, all depositors therefore choose the same value for θ . Of course, its derivative θ' is now independent of Y , too. This is important for the determination of the optimal tariff.

The first-order condition for the optimal tariff (38) may now be written as follows:

$$E \left[\frac{w}{(1-s \cdot k\theta^p - s \cdot w + i \cdot \theta)} \right] = \quad (45)$$

$$(r-i)(1-\rho) \cdot \theta' \cdot E \left(\frac{1}{(1-s \cdot k\theta^p - s \cdot w + i \cdot \theta)} \right)$$

This expression, together with the break-even constraint, defines the optimal tariff in the present special case. The optimal tariff will of course have the same general properties which were derived in the more general case of Section 3. There is, however, an interesting additional feature which stems from the fact that Y is no longer present in these optimality conditions. Therefore, in this special case, the equilibrium tariff is the same for all depositors, regardless of their income level.

5 Discussion

The transaction needs of deposit account holders are not entirely predictable. Therefore, deposit pricing according to the "Johnson norm", which amounts to full marginal cost-based activity charges on transactions, leaves the depositor as the "residual claimant" of the transactions cost uncertainty. This would be justified if the depositors were risk neutral. On the other hand, risk averse depositors would prefer contracts which reduce the costs of surprise transactions. In a competitive banking industry, banks would be forced to take this into account and develop deposit pricing mechanisms which satisfy depositor needs better than the full cost-based pricing of services.

The analysis presented in this essay has demonstrated how banks can use underpricing of transaction (payment) services to offer their customers insurance against uncertainty related to transactions. In the simple case where uncertainty regarding transaction needs is the only source of risk, no moral hazard is present and the expected transaction costs are proportional to the deposit balances, full insurance is possible. A competitive banking industry will then provide payment services to depositors without explicit charges. The costs of producing these services will be covered by setting a broad enough spread between deposit and lending rates.

This simple pricing system may no longer be viable if expected transaction costs incurred by the bank from servicing a deposit account

are endogenous and not perfectly revealed by the deposit balance. By "endogenous" it is meant that transaction needs are influenced by the depositor's behaviour and ultimately by the price incentives offered to him. This feature introduces the element of "moral hazard" into the problem. In Section 3 above, it was shown that when moral hazard is present in a tractable linear tariff framework, only incomplete insurance will be provided by a competitive banking industry. More precisely, some service charges on transaction services will have to be instituted, but there will remain a "cross subsidy" on transactions in any case, in the sense that the service charges will be below marginal costs of producing transactions. This "subsidy" is financed from the interest margin.

These results are, in fact, quite contrary to the standard view of the nature and causes of the "implicit interest phenomenon", which is the term commonly used to describe the provision of underpriced transaction services to depositors. Usually, economists have considered the "user pays" system with no cross subsidy on transactions to be the one which would prevail in competitive, undistorted markets. By contrast, in the models analyzed in this essay, the "first-best" solution would be to apply no service charges; but various frictions may alter the situation so that some service charges will be applied. Here, market imperfections are needed to explain the presence of service charges, whereas usually imperfections or regulation have been needed to explain the absence (or lowness) of direct service charges.

In risk sharing problems, moral hazard (hidden action) is not the only one potentially relevant source of frictions. In many instances, adverse selection (hidden information) is considered to be of primary importance. This has been considered in deposit pricing by Shaffer (1984), for instance. However, there are reasons why adverse selection (of depositors) may not be as great a problem in transaction deposit markets as in some other markets. These reasons follow from the fact that the deposit relationship involves a continuous service relationship between the bank and the customer, and it is possible to enter into a new contract as new information on the other party is uncovered. Further, banks are in a good position to obtain information on the different characteristics of their deposit customers. This means that if depositor heterogeneity is important for the determination of service charges and depositor interest rates, the banks would probably be able to avoid the adverse selection problem by price differentiation.

The analysis presented in Section 4 of this paper suggests, however, that the need for price differentiation is not necessarily implied by customer heterogeneity, not even when the customers are different with respect to income, which implies heterogeneity in

transaction needs, deposit holdings and risk aversion. In that section an example was presented in which the same equilibrium tariff was valid for all depositors, regardless of their income. This special case, requiring a particular type of utility function, could perhaps serve as a starting point for future extensions of the models presented in this essay that seek to cover the case of heterogeneous depositors in a more general, yet tractable way.

One broad conclusion emerging from the analysis is that "low" or "inferior" rates of return on assets which help to shield their owner from uncertainty with respect to transaction costs may be compatible with free, competitive equilibrium. Applied to banking problems, this insight warns against casually equating interest rate spreads and "underpriced" services with inefficiencies and lack of price competition. In the models presented in this essay, low-yielding assets emerge as a part of an insurance arrangement between banks and their customers. The novelty here is in the nature of the uncertainty: previous research has focused on idiosyncratic risks in income, or time preference, but has not considered transaction cost uncertainty, although this could prove to be the crucial element in understanding the nature of monetary services provided by banks.

The idea that banks may actually compete in providing transaction services below their actual cost gives new content to the concept of "liquidity creation" and may even suggest an avenue for further research attempting to provide better foundations for the "transaction cost" theory of money and liquidity. In particular, this line of research could help to integrate monetary theory with the theory of banking, which have developed more and more independently of each other since the Diamond and Dybvig (1983) model and the legal restrictions hypothesis have begun to dominate their respective fields.

References

- Black, F. (1970) **Banking and Interest Rates in a World Without Money: The Effects of Uncontrolled Banking.** *Journal of Bank Research* 1, 9–20.
- Diamond, D.W. – Dybvig, P.H. (1983) **Bank Runs, Deposit Insurance, and Liquidity.** *Journal of Political Economy* 91, 401–419.
- Fischer, S. (1983) **A Framework for Monetary and Banking Analysis.** *Economic Journal* 93 (Conference Papers Supplement 1983), 1–16.
- Hicks, J. (1935) **A Suggestion for Simplifying the Theory of Money.** *Economica (N.S.)* 2, 1–19.
- Holmstrom, B. (1979) **Moral Hazard and Observability.** *Bell Journal of Economics* 10, 74–91.
- Holmstrom, B. – Milgrom, P. (1987) **Aggregation and Linearity in the Provision of Intertemporal Incentives.** *Econometrica* 55, 303–328.
- Jacklin, C. (1987) **Demand Deposits, Trading Restrictions, and Risk Sharing.** In Prescott, E. – Wallace, N. (eds.) *Contractual Arrangements for Intertemporal Trade.* University of Minnesota Press, Minneapolis.
- Jewitt, I. (1988) **Justifying the First-Order Approach to Principal-Agent Problems.** *Econometrica* 56, 1177–1190.
- Johnson, H.G. (1968) **Problems of Efficiency in Monetary Management.** *Journal of Political Economy* 76, 971–990.
- Mirrlees, J. (1974) **Notes on Welfare Economics, Information, and Uncertainty.** in Balch, M. and McFadden, D. and Wu, S–Y (eds.): *Essays on Economic Behaviour Under Uncertainty.* North-Holland, Amsterdam.
- Rogerson, W. (1985) **The First Order Approach to Principal-Agent Problems.** *Econometrica* 53, 1357–1368.
- Saving, T. (1979) **Money Supply Theory with Competitively Determined Deposit Rates and Activity Charges.** *Journal of Money, Credit, and Banking* 11, 22–31.
- Shaffer, S. (1984) **Cross-Subsidization in Checking Accounts.** *Journal of Money, Credit, and Banking* 16, 100–109.
- Smith, B. (1984) **Private Information, Deposit Interest Rates, and the 'Stability' of the Banking System.** *Journal of Monetary Economics* 14, 293–317.
- Startz, R. (1979) **Implicit Interest on Demand Deposits.** *Journal of Monetary Economics* 5 (October 1979), 515–534.
- Startz, R. (1983) **Competition and Interest Rate Ceilings in Commercial Banking.** *Quarterly Journal of Economics* 98 (May 1983), 255–265.

- Tarkka, J. (1989) **Competitive Deposit Rates and Bank Service Charges**. Bank of Finland Discussion Paper 26/89.
- Tarkka, J. (1992) **Tax on Interest and the Pricing of Personal Demand Deposits**. Bank of Finland Discussion Papers 11/92.
- Varian, H. (1980) **Redistributive Taxation as Social Insurance**. *Journal of Public Economics* 14, 49–68.
- Vittas, D. – Frazer, P. – Metaxas-Vittas, T. (1988) **The Retail Banking Revolution. An International Perspective**. Lafferty Publications, London.
- Wallace, N. (1983) **A Legal Restrictions Theory of Money and the Role of Monetary Policy**. *Federal Reserve Bank of Minneapolis Quarterly Review* 7, 1–7.
- Walsh, C. (1983) **Taxation of Interest Income, Deregulation and the Banking Industry**. *Journal of Finance* 28 (December 1983) 1529–1542.
- White, L. (1987) **Accounting for a Non-Interest-Bearing Currency: A Critique of the Legal Restrictions Theory of Money**. *Journal of Money, Credit, and Banking* 19, 448–456.
- Whitesell, W. (1989) **The Demand for Currency versus Debitable Accounts**. *Journal of Money, Credit, and Banking* 21, 246–251.
- Whitesell, W. (1992) **Deposit Banks and the Market for Payment Media**. *Journal of Money, Credit, and Banking* 24, 483–498.
- Wilson, R. (1993) **Nonlinear Pricing**. Oxford University Press, Oxford.

Appendix

Service charges and depositor behaviour

In this appendix, the response of depositors to the pricing policy of banks is analyzed. Consider an increase in the service charge s which is accompanied by an adjustment in the deposit rate i such as to keep the expected profits of the bank at the previous (zero) level. This implies that when s is increased, the deposit rate i must increase as well, as specified by the constraint $i = r + (s-c)\theta^{\rho-1}$. It is proved below that such an increase in s will lead to a decrease in the choice parameter θ , and ultimately a reduction in the deposit balances D and the expected number of payments $E(N)$.

By the implicit function theorem,

$$\theta' = - \frac{[\partial E(W')/\partial s]}{[\partial E(W')/\partial \theta]}, \quad (A1)$$

where $E(W')$ denotes the partial derivative of the depositor's objective function with respect to θ . Here, in taking the derivative with respect to s , i must be allowed to vary to maintain $E(\pi) = 0$. Now, to ascertain that $\theta' < 0$ it is sufficient to establish that $\partial E(W')/\partial s < 0$, for the denominator in (A1) is known to be negative. (This property is equivalent to the concavity of the objective function in θ , which was demonstrated in the main body of the text for nonnegative values of s .)

It now remains to evaluate $\partial E(W')/\partial s$ at the depositor's optimum. The starting point is the depositor's first-order condition (31). Substituting the break-even constraint for i yields

$$E(W') = -E[U'(Y-P)] \cdot G \cdot Y + g'(\theta) = 0, \quad (A2)$$

where $G = [s(\rho-1) + c] \cdot k \theta^{\rho-1} - r$.

Note that the factor G must be positive for the first-order condition to hold. After the application of the break-even constraint the tariff P is now

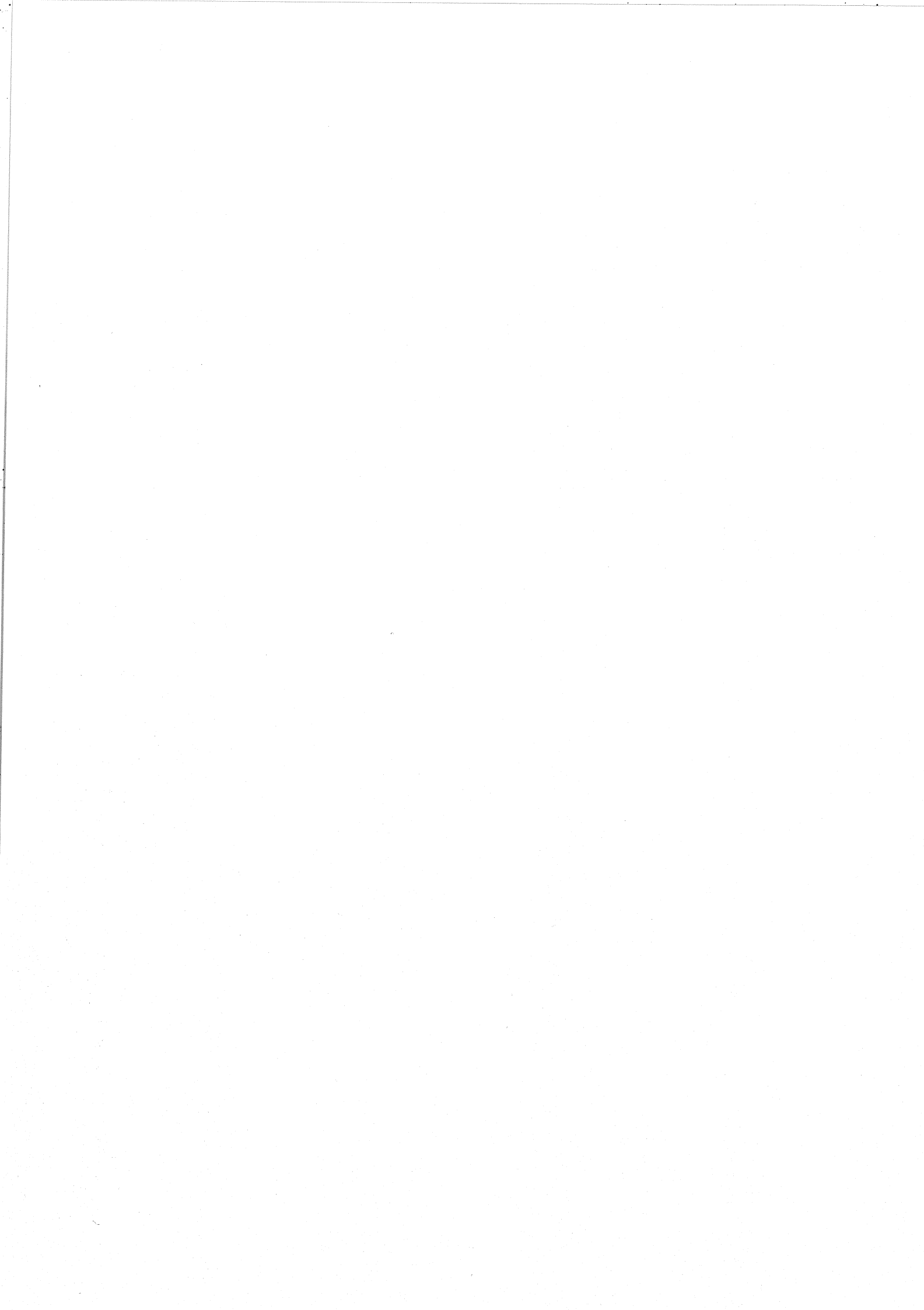
$$P = s \cdot e + c \cdot k \theta^\rho \cdot Y - r \cdot \theta \cdot Y. \quad (\text{A3})$$

Differentiating (A1) with respect to s gives

$$\frac{\partial E(W')}{\partial s} = E[U''(Y-P) \cdot e] G \cdot Y - E[U'(Y-P)] \cdot (\rho-1) k \theta^{\rho-1} < 0. \quad (\text{A4})$$

The negativity of this expression can be demonstrated in the following way. Because of the obvious negativity of the second term, it suffices to show that the first term is nonpositive. Now, clearly $E[U''(Y-P) \cdot e] = \text{Cov}[U'', e]$ and the nonpositivity of the covariance results from the fact that $\partial(U'')/\partial e = -(U''')s$, which is nonpositive for all $s \geq 0$ by the assumption made on the shape of the utility function (nonincreasing absolute risk aversion).

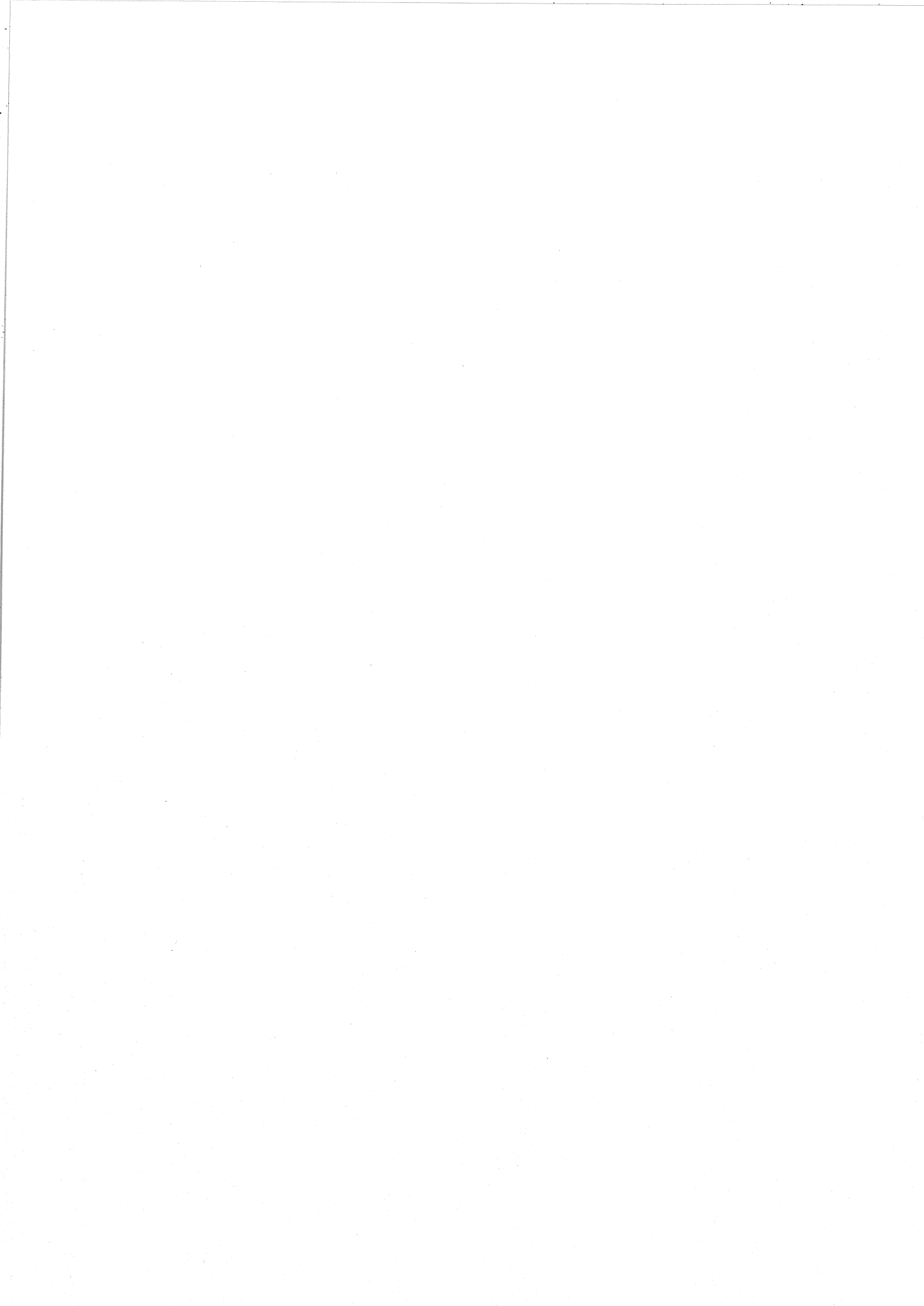
We have thus shown that $\theta' < 0$. Since $D = \theta Y$ and $E(N) = k \theta^\rho Y$ are increasing in Y , we have $\partial D/\partial s < 0$ and $\partial E(N)/\partial s < 0$.



Tax on Interest and the Pricing of Personal Demand Deposits

Contents

1	Introduction	91
2	Some stylized facts of deposit pricing	92
3	The model	94
4	The market equilibria	99
	4.1 The case with dominating taxes	99
	4.2 A neutral system	102
5	The regulated tax-free deposit rate	103
6	Discussion	106
	References	108
	Appendix	
	Some special cases	110



1 Introduction

The fact that most of the services provided by banks are covered indirectly through the interest margin on deposits and not by direct service charges must be regarded as a kind of anomaly from the point of view of standard microeconomic theory. The anomaly is, of course, actually visible only when banks can apply market-based pricing of deposits. Earlier, when interest rate ceilings on deposits existed in many countries, one could often appeal to such legal restrictions when seeking to explain the phenomenon of "implicit interest", as the cross-subsidization of transactions services with profits from deposit-taking activities is often called. After the recent advances in deposit rate deregulation and financial innovation, the legal restrictions argument has now lost force, however, and genuinely economic explanations must be found.

Some authors have noted that a tax asymmetry found in many fiscal systems may provide at least one explanation for the phenomenon (see Walsh (1983), and also Mitchell (1988) and Goodhart (1989, p. 227). Basically, the tax argument goes as follows. If the explicit interest income of depositors is taxable, but the benefits from subsidized transaction services are not, banks may be induced to reward depositors with free services instead of explicit interest even though individuals generally prefer money payments to grants in kind. These authors have not, however, provided a complete analysis of how the costs of using transaction services and the demand for deposits are determined when deposit pricing is distorted by asymmetric taxation. Walsh (1983) comes closest to that goal but he makes the unnecessary and restrictive assumptions that a) interest rates on deposits are uniform (linear pricing) and b) the proportion of free services to deposit balances is given from the point of view of each customer once he has chosen a particular bank.

In this essay, I use a "production model" of banking to study the effects of the aforesaid tax asymmetry. This makes it possible to apply standard microeconomic tools to the problem. The analysis shows that both the demand deposit rate and the service charge may be zero in a competitive equilibrium; and if this is the case, the remission of service charges is conditional on a minimum balance requirement, which dictates the velocity of deposits. The emergence of the cross-subsidizing equilibrium depends on the tax rate on interest income. Formulas are developed for the "effective marginal price" the depositors pay for bank services in the competitive equilibrium. These formulas indicate that the tax-free status of "implicit interest" on

deposits creates a subsidy on the provision of transaction services by banks.

In Finland, interest on personal transactions deposits (accessed mainly by giro and debit cards) has traditionally been tax exempt, if the interest rate is below a ceiling stipulated by tax laws. The ceiling has been low in relation to money market rates, however. Analysis of the Finnish tax system (in Section 5) reveals that it involves an even more extensive tax subsidy to banking than the standard system of deposit interest taxation which was outlined above.

In the literature there exist other explanations, in addition to the tax effect, for the cross-subsidization of deposit-taking and service production. These include price discrimination, monopolistic competition with the currency-based payment system and risk sharing (see Baxter, Cootner and Scott (1977), Whitesell (1988) and Tarkka (1989), as well as the introductory chapter of this study).

Since the purpose of this essay is merely to pinpoint the implications of the tax asymmetry for deposit pricing, the model presented below does not contain other features which might explain the existence of excess interest margins on demand deposits and the underpricing of transaction services. Such complications as imperfect competition, customer access costs and imperfect information are therefore deliberately excluded from the analysis.

2 Some stylized facts of deposit pricing

The following comments are an attempt to establish some "stylized facts" of bank pricing behaviour, especially in deregulated markets. Unfortunately, comparable information on the terms of transactions accounts applied in different countries is very scarce, especially if one tries to advance beyond casual observation. This is actually quite understandable, given the multiplicity of parameters which may be used in pricing demand deposits: apart from straight interest rates and service charges, the banks may use nonlinear pricing by making interest rates or fees conditional on the amount of deposited funds, the number of payments processed, etc. These issues are lucidly surveyed in the U.S. context by Davis and Korobow (1987). Reviews on other countries include Baltensberger and Dermine (1987), Vittas et al. (1988) and SOU (1989).

The freedom of price competition in the deposit market varies considerably across countries. In many countries, the interest paid on chequing accounts or other similar transactions deposits was, right up till the late 1980s, restricted by law or cartel (sometimes sponsored by the authorities; see Bingham (1985)). There are, however, also several major countries where no obvious barriers to price competition exist and the existing structure of deposit interest rates and service charges should therefore reflect "market forces". Examples of countries currently in this group are Germany, Sweden, the United Kingdom and the United States.

In Finland, market-determined deposit terms are a relatively recent development. In the Finnish tax system, personal interest income from bank deposits is tax free if the deposit terms satisfy certain conditions stipulated by tax laws. The rules include interest rate ceilings for tax-free accounts. At present, for example, interest on transactions deposits is exempt from tax if the interest rate does not exceed 2 per cent.

Until recently, the exemption effectively prevented the banks from offering deposit accounts with a higher, and consequently taxable, rate of interest. This is because interest income, if taxable, was taxed according to the marginal income tax rate, which is generally quite high. In the 1990s, however, the general rules of interest income taxation were changed. Interest income was no longer included in the income tax base, and a final tax at source was instituted instead. The evolution of the Finnish deposit tax system in the 1990s is described in Table 1, which is taken from Rantama and Solttila (1994). After this easing of interest taxation, taxable "high-yield" transactions accounts were introduced by banks. These new products have not, however, proved particularly successful in attracting business away from the old style tax-free transactions accounts.

As regards service charges on depositors, banks have attempted to move towards at least approximately cost-based pricing. The general practice seems to be, however, that service charges are waived if the depositor keeps a prescribed minimum balance on his account.¹ The experience suggests that, even with the current low tax rates on interest income, the combination of tax-free interest and no service charges for "good" depositors is quite viable in the competition for deposits.

¹ Cf. The Finnish Bankers' Association (1991) and Halonen (1993).

Table 1. Taxation of deposits in Finland

Effective date	Maximum tax-free interest rate		With- holding tax, %
	Transaction accounts	Fixed-term deposit accounts	
		24-month	
1 Jan 1989	Base rate less 5 %-pts	Base rate less 2 %-pts	
1 Dec 1989		Base rate less 1 %-pt	
1 Jan 1991			10 %
1 Jan 1992			15 %
1 May 1992	4.5 %	Base rate less 2 %-pts	
22 Oct 1992			
1 Jan 1993			20 %
6 May 1993	2.5 %		
1 Nov 1993	2 %		
1 Jan 1994			25 %

3 The model

In the following, the equilibrium deposit price structures are analyzed with the help of a simple model. The model consists of a representative consumer-depositor and a representative bank interacting in a competitive setting. The model is static and perfect information is assumed.

The bank produces "transfer services" which yield utility for the consumer-depositor. These services could consist of, for example, effecting payments through a funds transfer system such as cheque clearing, giro or EFTPOS. The consumer-depositor may pay for these services through direct service charges, or by keeping deposit balances at the bank at a low rate of interest, or both. The customer relationship of the consumer-depositor with the bank may thus consist of purchases of transfer services and the holding of deposits in the bank.

The bank places the deposited funds in "bonds" in a perfect capital market where an interest rate r prevails. However, a fraction q of deposits must be kept in the form of required reserves. For simplicity, these reserves are assumed to earn no interest.

The costs of producing transfer services are assumed to be separable with respect to different customers. We can therefore speak meaningfully of the costs and profits arising from a given customer relationship of the bank. The average (unit) cost of producing a quantity N of transfer services for the representative customer is assumed to be

$$c = c(N). \quad (1)$$

For simplicity, and analogous with the classic Baumol-Tobin model, we treat N as a continuous variable. The model and the results derived below are not very sensitive to the shape of the average cost function $c(N)$. To improve tractability, it is assumed to be continuously differentiable. Also, the marginal costs implied by $c(N)$ must be positive, requiring $c_N > -c(N)/N$ for all N . As long as these conditions are satisfied, the average cost function may be increasing, constant, U-shaped, or decreasing. The last-mentioned possibility is particularly interesting because it seems plausible that the maintenance of bank accounts involves significant, customer-specific fixed costs which are independent of the use of the account. In most of the mathematics that follows, the unit cost is simply denoted by c , even though it is not a constant but a function of N .

The profit π of the bank from a representative deposit relationship can now be defined as follows:

$$\pi = [r(1-q) - i]D + (s - c)N, \quad (2)$$

where D is the deposit balance, i is the deposit rate and s is the service charge for one unit of transfer services.

The only tax considered in the model is the tax on interest income, which is assumed to be asymmetric in the typical way: pecuniary interest income is taxable, but implicit interest on deposits in the form of underpriced services is tax free. Depositors are not able to deduct service charges paid to banks from their taxable income. For simplicity, the tax rate on interest income is assumed to be flat. Usually, in bank pricing models, it is convenient to consolidate the deposit interest payments and service charges into a net concept, called the tariff, which is then specified as a general function of deposit balances and service use. This device is not useful in the present context, for the tax asymmetry makes it necessary to distinguish the gross flows of interest payments and service charges.

Let us now turn to the budget constraint of the depositor. For the present purposes, there is no need to consider income other than interest income. So, the consumer-depositor is assumed to have a given wealth W which he may invest in "bonds" earning the rate of interest r , or in deposits earning the deposit rate i . The tax rate on interest income is t . The after-tax income of the consumer-depositor may thus be defined as $(1 - t)[r(W - D) + iD]$.

The expenditures of the consumer-depositor consist of purchases of a consumption good and bank service charges. Normalizing the price of the consumption good to unity, the budget constraint of the consumer-depositor may now be written as follows:

$$G + sN \leq (1 - t)[rW - (r - i)D], \quad (3)$$

where G is consumption of the consumption good.

The utility of the consumer-depositor is assumed to be a function of the consumption of the consumption good and the transfer services:

$$U = U(G, N). \quad (4)$$

The utility function is assumed to be concave and increasing in its arguments. It is further assumed that $\lim_{G \rightarrow 0} U_G = \infty$ and $\lim_{N \rightarrow 0} U_N = \infty$. These properties are used below to ensure positive demands for both N and G .

Under perfect competition, separability of costs and revenues by customer, and perfect information, the bank cannot earn positive profits from any customer relationship. Further, the perfect information assumption implies that the bank knows the customer's characteristics and thus "prices" (ie the service charge s and the deposit rate i) can be made conditional on the behaviour of the customer. Consequently, the consumer-depositor does not in fact optimize subject to given prices, but is able to obtain any terms for the customer relationship which are allowed by the "feasibility constraint" of nonnegative profits to the bank. Using the profit function (2), the assumption of zero profits in equilibrium implies

$$(r - i)D = (c - s)N + r_0D, \quad (5)$$

so that any interest lost by the depositor is used by the bank for subsidizing his use of transfer services and to cover costs caused by

the reserve requirement. Substitution of the zero-profit constraint (5) into the budget constraint of the representative consumer-depositor yields

$$(1-t)rW - G - sN - (1-t)(c-s)N - (1-t)r_qD \geq 0. \quad (6)$$

The consumer-depositor's decision problem is obviously to maximize utility (4) subject to the "augmented" budget constraint (6). The consumer-depositor does not only maximize with respect to the demands G , N and D ; he may also select the service charge s and deposit rate i so as to maximize his attainable utility level. Whatever the prices chosen, the solution is financially feasible from the bank's point of view as long as the constraint (6) is satisfied.

Finally, we require that the solution must also satisfy the nonnegativity constraints $D \geq 0$, $i \geq 0$ and $s \geq 0$. These constraints are due to the discontinuities which occur in typical tax systems when prices become negative. For example, even though positive service charges are typically not deductible in personal taxation, negative service charges (which would imply paying a bonus to depositors for transactions on the account) would be taxable. Similarly, negative interest rates are assumed not to be deductible in income taxation. Finally, the nonlinearity with respect to D arises from the fact that negative deposits do not allow the bank to have negative interest-free reserves at the central bank.

Formally, the problem of the customer-depositor may be solved by maximizing the following Lagrangean:

$$\begin{aligned} \max \mathcal{L} = & U(N, G) + \\ & k[(1-t)rW - G - sN - (1-t)(c(N) - s)N - (1-t)r_qD] + \\ & h[r(1-q)D + (s - c(N))N] \end{aligned} \quad (7)$$

so that $D \geq 0$, $s \geq 0$.

Besides the utility function, the Lagrangean consists of two constraints. The first, multiplied by the Lagrange multiplier k , is the budget constraint (6). The second, multiplied by another Lagrange multiplier h , is needed to ensure the nonnegativity of the deposit rate i . The constraint is derived from the zero-profit condition (5), which may be written as $iD = r(1-q)D + (s-c)N$. Since $iD \geq 0$ is implied by the nonnegativity constraints discussed above, one may use the condition

$r(1-q)D + (s-c)N \geq 0$ together with $s \geq 0$ and $D \geq 0$ to ensure that the original set of nonnegativity constraints holds.

The Kuhn-Tucker conditions for a maximum (see Intriligator, (1971) for the rules of setting up these conditions) are as follows:

$$\mathcal{L}_G = U_G - k \leq 0 \quad (8)$$

$$G\mathcal{L}_G = G(U_G - k) = 0 \quad (9)$$

$$\mathcal{L}_N = U_N - k[(1-t)(c + Nc_N) + ts] + h(s - c - Nc_N) \leq 0 \quad (10)$$

$$N\mathcal{L}_N = N\{U_N - k[(1-t)(c + Nc_N) + ts] + h(s - c - Nc_N)\} = 0 \quad (11)$$

$$\mathcal{L}_D = -k(1-t)rq + hr(1-q) \leq 0 \quad (12)$$

$$D\mathcal{L}_D = D[k(1-t)rq - hr(1-q)] = 0 \quad (13)$$

$$\mathcal{L}_s = -kNt + hN \leq 0 \quad (14)$$

$$s\mathcal{L}_s = s[kNt - hN] = 0 \quad (15)$$

$$\mathcal{L}_h = (1-q)rD + (s-c)N \geq 0 \quad (16)$$

$$h\mathcal{L}_h = h[(1-q)rD + (s-c)N] = 0 \quad (17)$$

$$\mathcal{L}_k = (1-t)rW - G - sN - (1-t)(c-s)N - (1-t)rqD \geq 0 \quad (18)$$

$$k\mathcal{L}_k = k[(1-t)rW - G - sN - (1-t)(c-s)N - (1-t)rqD] = 0. \quad (19)$$

In addition, h , k , D and s must be nonnegative. Different kinds of equilibria can emerge in the model, depending on the tax rate t and the reserve requirement q . The different possibilities are discussed below (however, only cases with $1 > t \geq 0$ and $1 > q \geq 0$ are scrutinized).

4 The market equilibria

Before starting the analysis of equilibria generated by the model, it is useful to note that the consumption of transfer services N and the consumption good G must be strictly positive in equilibrium because of the assumptions concerning the shape of the utility function as $G \rightarrow 0$ or $N \rightarrow 0$. Similarly, the assumption that the utility function is strictly increasing for all N and G means that the Lagrange multiplier k for the consumer-depositor's budget constraint must always be positive.

4.1 The case with dominating taxes

Let us consider a system in which the tax rate on interest income is greater than the reserve requirement. More precisely, assume that $1 > t > q > 0$. In this case, which is presumably the most common or "realistic" one, the equilibrium amount of deposits is strictly positive. This may be demonstrated by means of a contradiction as follows.

Assume for a moment that $D = 0$. From condition (17) we know that s must be positive whenever $D = 0$ (since N and c are assumed to be positive). With positive s and N we can then divide condition (15) through by Ns to obtain $h = kt$. Substituting $h = kt$ into (12) gives $kt(1-q) - krq(1-t) \leq 0$. Dividing this by the positive factor kr yields $t(1-q) - q(1-t) \leq 0$, or equivalently, $t \leq q$. This contradicts the assumption of $t > q$. Therefore, $D = 0$ is not optimal whenever $t > q$.

Knowing that D must be positive in this case, we can derive from condition (13) the result

$$h = kq(1-t)/(1-q). \quad (20)$$

Substituting this into (15) yields after some manipulation

$$Nsk[q(1-t) - t(1-q)] = 0. \quad (21)$$

With positive N and k , this can hold only if $s = 0$. This is another important result. It means that when $t > q$, transfer services are provided free of charge in equilibrium.

Now it is obvious that, if services are not paid directly, the bank must have a positive interest margin to cover its costs. With positive deposits and zero service charges, the zero-profit constraint allows us to write

$$i = r(1 - q) - cN/D. \quad (22)$$

But, with $h > 0$, $N > 0$ and $s = 0$, we can derive from the condition (17) the result

$$r(1 - q) - cN/D = 0, \quad (23)$$

which, taken together with (22), implies $i = 0$. The optimal deposit rate is thus shown to be zero in this case.

We can now substitute the results for s and h into the original set of Kuhn-Tucker conditions, which under the assumption of an interior maximum for G and N simplifies to

$$U_G = k \quad (24)$$

$$U_N = k(c + Nc_N)(1 - t)/(1 - q) \quad (25)$$

$$(1 - t)rW = G + cN(1 - t)/(1 - q) \quad (26)$$

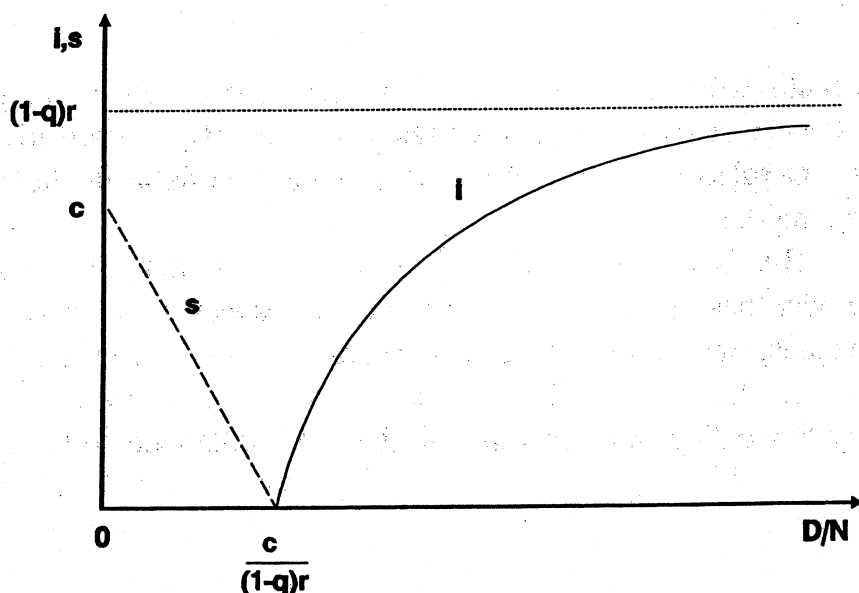
$$D = cN/r(1 - q). \quad (27)$$

From the second first-order condition (25) we see that the effective marginal price paid by consumers for the transfer services is in this case $(c + Nc_N)(1 - t)/(1 - q)$. Note that the expression $(c + Nc_N)$ defines the marginal cost of the production of transfer services. Note also that, in the present case, the effective marginal price paid by consumers is lower than the marginal cost of producing these services. This distortion, which is due to the tax asymmetry, is partly compensated by the reserve requirements. With $t > q$, the "reserve requirement tax" is not, however, sufficient to restore complete neutrality and an effective subsidy on N is inherent in the asymmetric tax system.

Note that equations (24), (25) and (26) constitute a set of first-order conditions and a budget constraint which are sufficient to determine consumption G of the ordinary consumption good as well as the amount N of transfer services. Given N , condition (26) reveals how the stock of deposits held by the representative consumer-depositor is determined simply from the bank's zero-profit condition. With zero interest and zero service charges, the customer is required to keep minimum balances at the bank such that the generated interest margin is sufficient to cover the cost of producing the transfer services used by the customer.

The logic of the pricing system which emerges when taxes dominate reserve requirements may be made clearer by considering the following diagram, describing the tax-minimizing deposit terms which the bank may offer under the break-even constraint (Figure 1). The average service charge s and the average deposit rate i consistent with zero profits are shown as functions of the deposit-to-transactions ratio D/N . If this ratio is less than $c/((1-q)r)$, the deposit rate is zero and the service charge is determined along the straight line s ; when the deposits-to-transaction ratio is higher than $c/((1-q)r)$, the service charge is zero and the deposit rate is determined along the curve i , asymptotically converging to $(1-q)r$. Now, what the above analysis has demonstrated, is that when $1 > t > q > 0$, the depositors will always choose the point $D/N = c/((1-q)r)$ where both service charge and deposit rate are zero. This is evident from the formula (27) above.

Figure 1. Possible tariffs under the banks' zero-profit constraint



In sum, the case $1 > t > q > 0$ leads to the classic deposit pricing system: zero interest on demand deposits, "free" service and minimum average balance requirements. Since the minimum balance requirement is proportional to the volume of transfer services, the deposit terms define an implicit price for these services. Since the minimum balances required are inversely proportional to the interest rate on bonds, the model incidentally provides a novel explanation for the negative interest elasticity of demand deposits.

4.2 A neutral system

In the special case where the reserve requirement rate and the tax rate on interest income are both positive but equal, or $1 > q = t > 0$, the condition (13) may be simplified to

$$D(h - kt) = 0. \quad (28)$$

Thus, if D is not zero in equilibrium, then $h = kt$. It is easy to check that $h = kt$ also holds if D is zero. For when $D = 0$, then $s > 0$ by condition (16); and with a positive s , $h = kt$ is implied by condition (15).

Using the information that h is positive, we can derive from (17) the result

$$(1 - q)rD + (s - c)N = 0, \quad (29)$$

which, when substituted into the zero-profit constraint, gives $D_i = 0$. This means that the deposit rate i is zero, if deposits are positive; this is not ruled out either if $D = 0$, for then the deposit rate has no role in the model.

But is the stock of deposits positive or not? It turns out that the model has no unique solution with respect to D and s . This is so because the Kuhn-Tucker conditions (12) to (15) vanish when $1 > q = t > 0$. The variables D and s are thus left indeterminate, restricted only by the zero-profit constraint, which now takes the form

$$s = c - r(1 - q)D/N$$

(30)

and the nonnegativity constraints $D \geq 0$ and $s \geq 0$. The interpretation is that the consumer-depositor is indifferent about the combinations of D and s allowed by the bank's zero profit condition. The result that the optimal deposit rate is zero thus remains the only precise implication of the model on deposit pricing in this case.

5 The regulated tax-free deposit rate

The model is now extended for application to the Finnish deposit market. In particular, we wish to analyze the case where the banks are allowed to take tax-free deposits at some regulated interest rate. In the following, this legally determined deposit rate is denoted by ρ . This rate is assumed to be set below the after-tax capital market rate, i.e. $\rho < (1-t)r$. Further, the analysis is restricted to the most realistic case of $t > q$, which also produces the most interesting results.

These assumptions simplify the analysis of the competitive equilibrium considerably. First, we can now take for granted that the interest margin $r - \rho$ is positive. This means that the zero-profit condition of competitive equilibrium (analogous to equation 5 above) can be used to solve for the amount of deposits the bank requires to break even at given transaction volumes and service charges:

$$D = [(c - s)/(r(1 - q) - \rho)]N.$$

(31)

Another modification is that the budget constraint of the consumer-depositor must be rewritten to take into account the tax exemption of deposit interest:

$$G + sN \leq (1 - t)rW - [r(1 - t) - \rho]D.$$

(32)

As above, the budget constraint may be developed to take into account the zero-profit constraint. Substituting (31) for D into (32) yields

$$(1-t)rW - G - \{(c-s)[r(1-t) - \rho]/[r(1-q) - \rho] + s\}N \geq 0. \quad (33)$$

The optimization problem of the depositor can now be characterized with the following Lagrangean, which is somewhat simpler than (7) above:

$$\max \mathcal{L} = U(N, G) + k \left\{ (1-t)rW - G - \left\{ (c-s) \frac{[r(1-t) - \rho]}{[r(1-q) - \rho]} + s \right\} N \right\} \quad (34)$$

so that $s \geq 0$.

The conditions for a restricted maximum are now developed with respect to the volumes G and N , as well as the service charge s . Recall that, since the zero-profit condition was taken into account in deriving the above Lagrangean, the depositor is able to select the service charge which maximizes (34). The amount of deposits required by the bank will vary according to the selected charge, however, as indicated in (31). The Kuhn-Tucker conditions are now as follows (taking it for granted that an interior solution for G and N is obtained):

$$\mathcal{L}_G = U_G - k = 0 \quad (35)$$

$$\mathcal{L}_N = U_N - k \{(c + Nc_N - s)[r(1-t) - \rho]/[r(1-q) - \rho] + s\} = 0 \quad (36)$$

$$\mathcal{L}_s = k \{ [r(1-t) - \rho]/[r(1-q) - \rho] - 1 \} \leq 0 \quad (37)$$

$$s\mathcal{L}_s = sk \{ [r(1-t) - \rho]/[r(1-q) - \rho] - 1 \} = 0 \quad (38)$$

$$\mathcal{L}_k = \{(1-t)rW - G - \{(c-s)[r(1-t) - \rho]/[r(1-q) - \rho] + s\}N\} \geq 0 \quad (39)$$

$$k\mathcal{L}_k = k \{(1-t)rW - G - \{(c-s)[r(1-t) - \rho]/[r(1-q) - \rho] + s\}N\} = 0. \quad (40)$$

Under the assumption that the tax rate t is higher than the reserve requirement rate q , condition (37) holds as an inequality. This is because the term $[r(1-t) - \rho]/[r(1-q) - \rho]$ in the expression is smaller than one. By (38), this implies that the optimal service charge must be

zero. As above, the intuition is that, when the reserve requirement tax is lower than the marginal tax rate, taxes are minimized when services are paid for by keeping deposits at the bank.

Substituting this result ($s = 0$) into condition (36) yields the following expression for the optimal use of transactions services N :

$$\mathcal{L}_N = U_N - k(c + Nc_N)[r(1-t) - \rho]/[r(1-q) - \rho] = 0. \quad (41)$$

Now inspection of this expression reveals that the effective marginal price paid by the depositor-customer of transaction services is given by $(c + Nc_N)[r(1-t) - \rho]/[r(1-q) - \rho]$. This is a generalization of the previous result (25), which described the situation prevailing when the maximum tax-free interest was zero. As in that case, the effective marginal price is again lower than the marginal cost $c + Nc_N$.

Further, it can be shown that, under the assumption $t > q$, the effective marginal price paid by the depositor is the lower, the higher is the tax-free deposit rate. An extreme case is reached as ρ approaches $r(1-t)$, the after-tax capital or money market rate. Then the effective price of transactions services approaches zero. Of course, the same results apply for changes in the tax rate on personal interest income. If the tax rate is raised, while keeping the tax-free deposit rate constant, the effective price of transactions services is lowered – and the implicit tax subsidy on the banks' real activities is increased.

A peculiar feature of the Finnish type of deposit tax system is that the effective marginal price of transactions services is a function of the market rate of interest. Assuming that the tax-free deposit rate is not adjusted by the authorities according to the changes in the market rate, the effective price on bank services will decline when the market rate of interest declines and vice versa. When the maximum permissible tax-free rate is strictly positive, the changes in the effective price are more than proportional to the changes in the market rate of interest. The analysis of the tax-free regulated deposit rate regime may be concluded by noting that (31) may be simplified to give the following result for the minimum balance requirement:

$$D = cN/[r(1-q) - \rho]. \quad (42)$$

As before, the minimum balances required by the bank at the customer's optimum can again be calculated simply by dividing the total cost of servicing deposits cN by the interest margin. The minimum balance requirement should increase as the regulated tax-free

deposit rate is increased – or when the net interest rate on bank assets decreases.

6 Discussion

We have analyzed the effects of taxation on the pricing of personal transactions deposits. More specifically, the analysis focused on the competitive price structure which emerges if interest on deposits is taxable while "implicit interest" in the form of underpriced bank services for depositors is not. It was shown that, if the tax rate on interest income is high enough, this asymmetry may lead to zero deposit rates and zero service charges. Minimum balance requirements also emerge. The results differ from Walsh (1983), who concludes that banks would generally pay both implicit and explicit interest in the competitive equilibrium. The difference in the present results arises because we allow more general (ie nonlinear) pricing schemes to be used.

The tax-based model of deposit pricing has a number of attractive features. First of all, it explains the classical system of demand deposit pricing. Secondly, the model gives an explanation for the demand for money which does not rely on including (deposit) money directly in the utility function of the consumers, nor on imposing an *ad hoc* cash-in-advance constraint on transactions. In this model, deposits are simply a device for paying for transaction services in a way which avoids some taxes. Moreover, the model predicts the velocity of deposits, defined as N/D , to be an increasing function of the rate of interest.

Is the tax explanation a serious candidate as "the" explanation for the observed peculiarities of deposit pricing? This is not clear despite the above-mentioned merits of the model. A serious problem with the tax argument is that the required asymmetry does not exist in corporate taxation, since firms can deduct service charges as expenses in taxation (unlike households in most countries). Therefore, the tax-based model can probably be applied only in the case of household deposits.

An important empirically refutable implication of the tax-based model of deposit pricing is that explicit service charges should never coexist with explicit, taxable interest, at least not within a single deposit relationship. This has the intuitive explanation that under perfect information it is financially irrational to pay taxable interest to depositors if the depositors, in turn, pay service charges to the bank

which are not deductible in taxation. If service charges were widely paid by holders of taxable interest-bearing demand deposits, this would probably indicate the presence of market imperfections which would in some way prevent complete tax arbitrage. That would also call for the development of more complicated models of deposit pricing than the simple tax-avoidance model.

Anyhow, the above analysis leads to several important conclusions regarding the distortionary effects of existing tax systems on banking. From the policy point of view, the most important result of the analysis was that the model clearly demonstrates that the tax-free status of "implicit interest" is a subsidy for the production of banking services. Reserve requirements can be used to reduce this distortion, however. The present model is interesting, and perhaps even unique, in that the reserve requirement "tax" presents itself as neutrality improving, rather than distortionary as is usually argued. The analysis suggests that, for personal transaction deposits at least, there is a strong case for imposing reserve requirements. Moreover, the incentives to evade the burden of the reserve requirement on these deposits would not be strong because of the implicit tax advantage.

In the Finnish case, the distortionary effects of not taxing the benefit of free services are amplified by the tax exemption of deposit interest on low-yielding accounts. The implicit subsidy to banks' real activities approaches 100 per cent of costs when the after-tax market rate approaches (from above) the regulated tax-free deposit rate. Thus, the impact of the tax system on the banking industry is variable, and may occasionally be very great, depending of course on the price elasticity of demand for liquidity services provided by banks.

References

- Baltensberger, E. – Dermine, J. (1987) **Banking Deregulation in Europe**. *Economic Policy* (April 1987) 64–95.
- Baxter, W. – Cootner, P. – Scott, K.E. (1977) **Retail Banking in the Electronic Age**. Allanheld, Osmun & Co., Montclair, N.J.
- Bingham, G. (1985) **Banking and Monetary Policy**. OECD, Paris.
- Black, F. (1970) **Banking and Interest Rates in a World Without Money**. *Journal of Bank Research* 1, No. 3, 8–28.
- Davis, R. – Korobow, L. (1987) **The Pricing of Consumer Deposit Products – The Non-rate Dimensions**. *Federal Reserve Bank of New York Quarterly Review*, (Winter 1987), 14–18.
- Fama, E. (1980) **Banking in the Theory of Finance**. *Journal of Monetary Economics*, (January 1980), 39–57.
- Finnish Bankers' Association (1991) **Kotitalouksien yleisimmät pankkipalvelumaksut**. Helsinki.
- Fischer, S. (1983) **A Framework for Monetary and Banking Analysis**. *Economic Journal* (Conference Papers Supplement 1983), 1–16.
- Goodhart, C.A.E. (1989) **Money, Information and Uncertainty**. Second Edition. (Macmillan, London).
- Halonen, U. (1993) **Yksityishenkilöille tarkoitettut tilit**. Lokakuu 1993. Suomen Pankkiyhdistys, Helsinki.
- Intriligator, M. (1971) **Mathematical Optimization and Economic Theory**. Prentice-Hall, Englewood Cliffs, N.J.
- Mitchell, D.W. (1988) **Explicit Interest and Demand Deposit Service Charges**. *Journal of Money, Credit, and Banking* 20, No. 2, 270–274.
- Rantama, J. – Soltila, H. (1994) **Bank Deposits and Their Taxation**. *Bank of Finland Bulletin* 68, No. 4, 8–13.
- Saving, T.R. (1979) **Money Supply Theory with Competitively Determined Deposit Rates and Activity Charges**. *Journal of Money, Credit, and Banking* 11, No. 1, 22–31.
- SOU (1989) **Kostnadsutveckling och konkurrens i banksektorn**. Sveriges Offentliga Utredningar 1989:16.
- Tarkka, J. (1989) **Competitive Deposit Rates and Bank Service Charges**. *Bank of Finland Discussion Papers* No. 29/89.

Vittas, D. – Frazer, P. – Metaxas-Vittas, T. (1988) **The Retail Banking Revolution: An International Perspective.** Lafferty Publications, London.

Walsh, C.E. (1983) **Taxation of interest Income, Deregulation and the Banking Industry.** *Journal of Finance* 28, No. 5, 1529–1542.

Whitesell, W. (1988) **Deposit Pricing, Bank Market Structure, and Welfare with Cost-Minimizing Consumers.** Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series No. 48.

Appendix

Some special cases

The case with dominating reserve costs

Consider the case $1 > q > t \geq 0$. The reserve requirement tax is now higher than the tax on interest income and the equilibrium amount of deposits is zero in this model. This can be demonstrated by means of a contradiction. From (13) we see that in an equilibrium with a positive amount of deposits, $h = kq(1-t)/(1-q)$. When $1 > q > t \geq 0$, this contradicts condition (14), which requires $h \leq kt$.

The principles of deposit pricing are very simple in this case. First, deposit interest rates have no meaning or significance as deposits do not exist. Further, with $D = 0$, the zero-profit condition requires $s = c$. This equilibrium thus involves full-cost pricing of transaction services. Since we have allowed for non-linear price schedules in the model by inserting the zero-profit condition directly into the consumer-depositor's budget constraint, the marginal price of services equals the marginal cost of service production in this case.

Obviously, positive deposits could also be explained in this case when tax incentives for deposit-taking do not exist, by assuming "economies of scope" in banking so that keeping deposits would lower the cost of providing transfer services. This possibility is not, however, explored here.

Zero reserve costs

When $q = 0$ or $q = t$ or both hold, the model is not capable of producing a fully determined solution. Some of the variables D , s , and i are left indeterminate in these border cases, meaning that the consumer-depositor is indifferent with respect to the value given to the indeterminate variables. These cases are briefly commented on below.

Consider the case where $1 > t > q = 0$. When the burden caused by the reserve requirement is zero, but the tax on interest is positive, the model is able to precisely determine only one of the terms of the customer relationship. This is the service charge, which turns out to be zero in this case.

That $s = 0$ may be demonstrated as follows. When $q = 0$, condition (12) simplifies to $hr \leq 0$. With a positive r , this implies $h = 0$. Using $h = 0$, condition (15) may be written in the form $-sktN = 0$, which implies $s = 0$ (k , t , and N are all positive). With $s = 0$ and $q = 0$, condition (16) implies

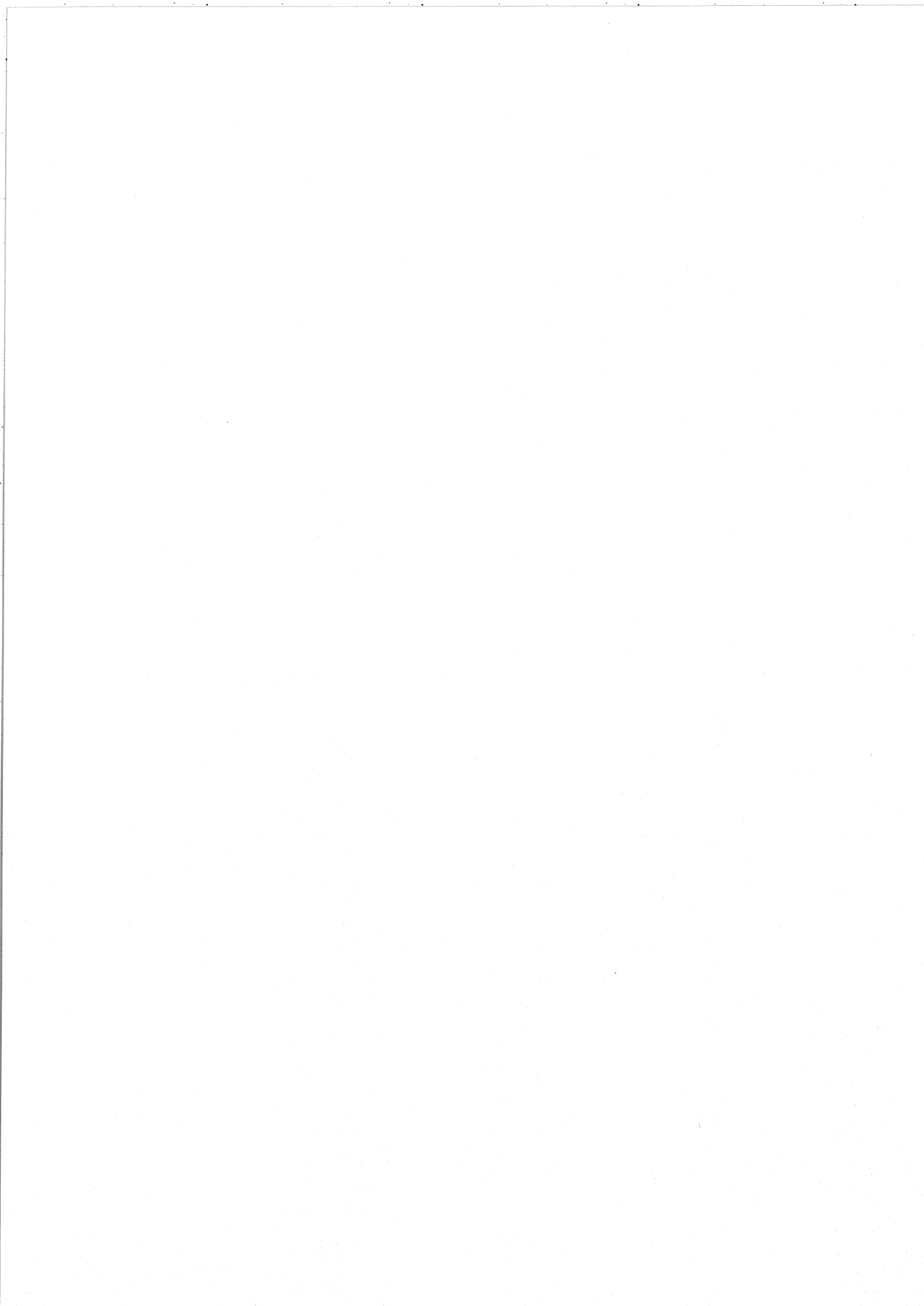
$$D \geq c(N)N/r, \tag{A1}$$

so that a minimum balance requirement is clearly present in this equilibrium. However, D is not uniquely determined, since all other conditions containing D vanish when $q = s = h = 0$. The interpretation becomes clear when we combine (A1) with the zero-profit constraint of the bank, which in this case takes the following form:

$$i = r - c(N)N/D. \tag{A2}$$

The interpretation is that here the consumer-depositor is indifferent with respect to the different combinations of D and i offered by the bank, as defined by (A2); the choice is, however, restricted by the requirement that the resulting $i \geq 0$ is positive, which implies the lower bound for the deposit balance defined by (A1).

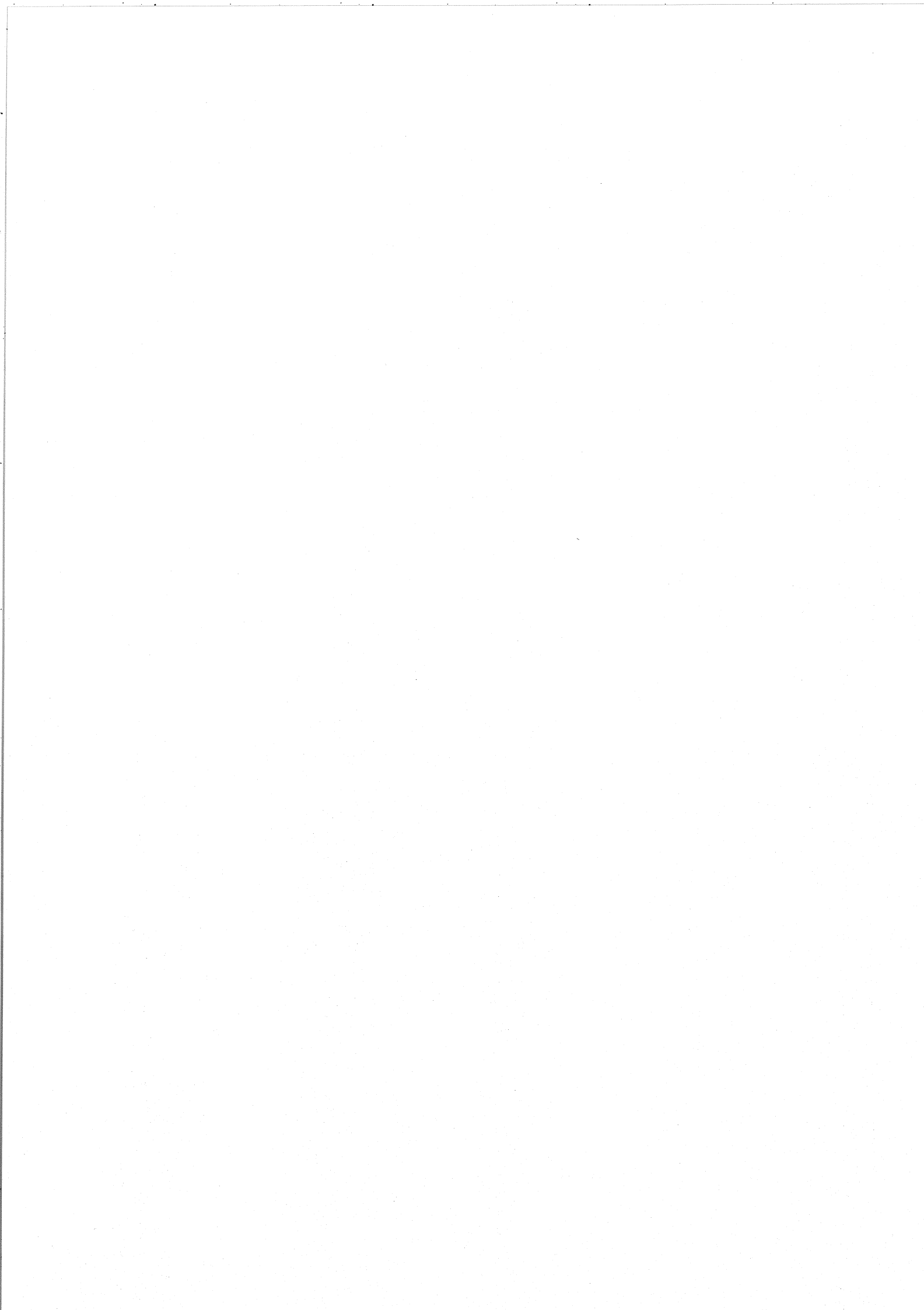
Finally, when neither income tax nor the reserve requirement are present, it is easy to show that the model is totally indeterminate with respect to i , D and s . The reason is of course that all methods of paying for the transfer services are financially equivalent in this case.



Implicit Interest as Price Discrimination in the Bank Deposit Market

Contents

1	Introduction	115
2	The demand for bank services	118
3	Deposit pricing	121
4	The existence of cross-subsidizing optima	126
5	An inventory-theoretic cost function	132
6	Discussion	138
	References	141



1 Introduction

The purpose of this essay is to explore the possibilities of applying the theory of multiproduct nonlinear tariff design to the pricing of cheque accounts. This endeavour is motivated by the need to improve our understanding of some features of deposit pricing by banks, especially the relatively low interest rates on liquid deposits and the apparent underpricing (cross-subsidization) of payment services. Neither of these features is uncommon in the deposit banking industry.

The market for liquid bank deposits such as cheque accounts actually consists of two submarkets: the market for deposited funds, and the market for payment services performed by the bank when deposits are used as a means of payment. Price formation in these two markets is an interesting subject of study, both because the nature of competition in the retail deposit market shapes the whole banking industry in a powerful way and also because the observed pricing practices are not trivially explained by the standard competitive price theory.

A salient feature of (retail) cheque account pricing is the prevalence of cross subsidies between the deposited funds market and the payment services market. This is apparent in the fact that the rate of interest on cheque account deposits is usually significantly lower than the rate of interest on CDs, for example. At the same time, service charge revenues from payment services are small compared with the high costs of these services, including branch networks, large computer systems etc. (this is easily verified from OECD (1993), for example; see also Vittas et al. (1988)). This observed cross-subsidization immediately raises the issues of efficiency and distributional effects of the banks' conduct. The purpose of this essay is to suggest a framework which may be used in analyzing these problems.

During the decades of deposit rate regulation, the cross subsidies on depositors' payment services were easy to explain as "implicit interest" although some disagreement prevailed as to how this "implicit interest" was determined and what kind of effect it had on the demand for chequable deposits (cf. Barro and Santomero (1972), B. Klein (1974) and Becker (1975)). Many economists believed that the cross subsidies would disappear following the deregulation of deposit rates. (cf. Saving (1979) and Fisher (1983)). After the international wave of deregulation in the 1980s, significant interest margins nevertheless still exist, and, while banks have tried to raise service charges, the share of fees in banks' net income has not become dominant.

One explanation for the continued viability of "implicit interest", even in today's unregulated markets, relies on a tax-minimization argument. In many fiscal systems, interest income is taxable, but the benefit of free or underpriced banking services is not taxed (nor are service charges related to payment services deductible in taxation). This kind of system encourages banks to compete for deposits with tax-free implicit interest instead of taxable explicit interest (Walsh (1983)). It should be noted, however, that the tax distortion is not present in corporate taxation. Also, we sometimes observe banks paying interest to customers who are simultaneously paying service charges to banks. This is in contradiction with complete tax arbitrage (see Tarkka (1992)). So, other explanations besides the tax distortion argument should probably be explored as well.

In this essay, imperfect competition is considered as an alternative or potentially complementary explanation. There are already a number of studies which have applied models of imperfect competition to explain deposit pricing, starting from M. Klein (1971). Those which have taken the pricing of payment services explicitly into account include the spatial competition model of Baxter, Cootner, and Scott (1977), Mitchell's (1988) monopoly model of service charge determination and Whitesell's (1992) monopoly model of both demand deposit interest and service charges. Unfortunately, all of these studies share an important shortcoming: they arbitrarily restrict the analysis to the case of uniform pricing, although the uniformity assumption is in contradiction with observed pricing practice and may be seriously questioned on theoretical grounds.

In practice, banks operating in deposit markets where pricing is not regulated typically apply complex nonlinear pricing schemes to cheque accounts. The nonlinearities come in many forms. There may be fixed, monthly or yearly account maintenance fees, for instance. As far as deposited funds are concerned, customers with higher balances on their accounts are often favoured with better interest rates or lower service charges or both. As for payment services, it is not unusual for charges to depend on the number of cheques written etc. These practices are reviewed by Hörngren (1988), Vittas et al. (1988) and Davis and Korobow (1987), for example.

Turning to theory, uniform (non-discriminatory) pricing is generally suboptimal from a monopolist's point of view. Only if the products can be costlessly resold on a "second-hand" market, or if customers can freely form coalitions to purchase a service jointly, can we safely assume that each unit of a monopolist's output must be sold at the same price. In retail banking, where each household and firm typically has its own bank account, the conditions for price

discrimination are clearly present. In assessing the previous theoretical literature, it should be recalled that the possibility of price discrimination invalidates the analytical results obtained under the uniform price assumption even when all customers are identical.

This essay approaches the "implicit interest" issue and other aspects of the cheque account pricing problem from a price discrimination perspective. It presents a model of deposit pricing by a monopoly bank which practices second-degree price discrimination among heterogenous depositors. The pricing problem of a discriminating multiproduct monopoly selling to customers differentiated by an unobservable characteristic was first solved by Roberts (1979) and Mirman and Sibley (1980).¹ In this essay, the model is applied to the deposit pricing problem and the previously neglected possibility of marginal cross subsidies is considered.

It turns out that this kind of model can explain several commonly observed features of cheque account pricing. First, the price discrimination model easily generates a large spread between the security market rate (net of intermediation costs) and the deposit rate of interest. Second, cases where payment services are supplied at prices below marginal cost are shown to be possible. This is perhaps particularly interesting, since sufficient conditions for profit-maximizing nonlinear tariffs to include marginal subsidies are not well established in the literature. Moreover, a model of the demand for money is presented which suggests that conditions for cross-subsidization may actually be inherent in the transaction deposit market. Third, nonlinearities in deposit pricing are given a clear microeconomic interpretation in the model, and may even be used to construct empirical tests of the importance of price discrimination in banking.

The essay is organized as follows. The microeconomic assumptions concerning the demand for deposit services are presented in Section 2. The profit-maximization problem of a monopoly bank is solved in Section 3, yielding a first-order characterization of the optimal nonlinear tariff. The validity of the insights provided by the first-order approach, especially as regards the cross-subsidization issue, is analyzed in Section 4. An inventory-theoretic analysis of the demand for banking services is presented in Section 5, suggesting that incentives to marginal subsidization may be inherent in banking. The results of the essay are summarized and evaluated in Section 6.

¹ Wilson (1993) presents a thorough survey of the relevant techniques.

2 The demand for bank services

We model the demand for deposit services using a liquidity cost approach, as defined in Feenstra (1986), for example. This approach explains the demand for money by assuming that agents hold money in order to reduce liquidity costs (or transactions costs) as implied by the available transactions technology. In this analytical tradition, money balances enter the cost function as an input, reducing transactions costs incurred from earning and spending a given flow of income. In this section, this approach is generalized to the case of two (monetary) banking services: deposit-taking and payment services.

It is assumed that there is a continuum of depositors, identical in all respects except their income level Y . The income level can thus serve as a "type parameter" in the language of price discrimination models. The distribution of Y in the population is according to the continuous cumulative distribution function $F(Y)$, where $Y \in [Y^{\min}, Y^{\max}]$. We require finite densities, so that $F(0) = 0$ and $f(Y) = dF(Y)/dY$ for all Y .

Depositors incur internal transactions costs which depend upon their income level and the amount of deposit services they use. A generic form of the transactions cost function may be written as $G(Y, D, N)$, where D denotes the average deposit balance and N denotes the volume of transactions made using the account (number of cheques and transfers per unit of time, for example). Below, variables D and N will be called simply "deposits" and "payment services", respectively.

The shape of the cost function is obviously crucial for the demands for the two services. We assume that this function is twice continuously differentiable, strictly increasing in the income level and strictly decreasing in both banking services, whenever the income level is positive. Denoting partial derivatives by subscripts, we may summarize these properties by writing:

- (i) $G_Y(Y, D, N) > 0$
- (ii) $G_D(Y, D, N) \leq 0$ and $\{G_D(Y, D, N) < 0 \text{ for all } Y > 0\}$
- (iii) $G_N(Y, D, N) \leq 0$ and $\{G_N(Y, D, N) < 0 \text{ for all } Y > 0\}$.

We also assume that transactions costs are convex in both banking services:

- (iv) $G_{DD}(Y, D, N) > 0$ for all $Y > 0$
- (v) $G_{NN}(Y, D, N) > 0$ for all $Y > 0$
- (vi) $G_{DD}(Y, D, N) G_{NN}(Y, D, N) - [G_{ND}(Y, N, D)]^2 > 0$.

Basically, the convexity assumptions (iv–vi) ensure that well-behaved demand functions exist in the standard neoclassical sense at least if both services (D and N) are available at given marginal prices (ie under a linear price system).

The depositors are assumed to choose the amount of deposits to hold (as an average stock) and the volume of payments they make with their deposits (as the average number of transactions per unit of time) by minimizing total liquidity costs, which are defined as the sum of internal costs $G(Y, D, N)$ and the financial costs of holding deposits D and using payment services N . The financial part of total liquidity costs depends on the interest rate paid on deposits, service charges on payment services and the alternative rate of return on the depositor's funds.

Let us assume that the alternative rate of return, determining the depositor's opportunity cost of funds, is the security market rate r . Collecting the net amount of deposit interest less service charges in a general tariff function $P(D, N)$, the decision problem of a depositor of type Y can then be written as

$$\min_{D, N} [G(Y, D, N) + P(N, D) + rD]. \quad (1)$$

Restricting the tariff function to be independent of the customer's type Y means that we rule out first-degree price discrimination. By allowing the tariff function to be nonlinear we are able to take second-degree price discrimination into account.

In the case of uniform pricing of depositors' services, the deposit rate and the service charge would be constants. In that case, the tariff function would be linear, ie $P(D, N) = sN - iD$, where s is the average service charge and i the interest rate on the average deposit balance. It is, however, well established that linear tariffs are not optimal for a monopolist except under restrictive assumptions (cf. Philips (1983)).

In the case of nonlinear tariffs, it may not be possible to separate the tariff into additive components which could be unambiguously interpreted as price schedules for individual products. So, as concern cheque account pricing, it is perfectly possible that intramarginal interest payments and service charges cannot be defined in a

meaningful sense. Generally, this will be the case whenever the tariff function is not additively separable.² Marginal prices, by contrast, can be defined whenever the tariff function is differentiable. So, we define the following concepts:

1. The marginal interest rate on deposited funds is the negative of the partial derivative of the tariff function with respect to deposits: $-P_D(D,N)$.
2. The marginal service charge is the partial derivative of the tariff function with respect to the use of payment services $P_N(D, N)$.

Let us now turn to analyze the solution of the representative depositor's cost-minimization problem. The first-order conditions for depositor optimum are as follows:

$$G_D(Y, \hat{D}, \hat{N}) + P_D(\hat{D}, \hat{N}) + r = 0 \quad (2)$$

$$G_N(Y, \hat{D}, \hat{N}) + P_N(\hat{D}, \hat{N}) = 0. \quad (3)$$

Here, hats over the variables denote optimal values of D and N . It should be noted that the optimal values are not constants but functions of Y . If the tariff function is linear, the second-order sufficient conditions for a minimum are assured by the convexity assumptions (ii)–(vi) above. In the general nonlinear case, the local second-order conditions require that the sum $G(\cdot) + P(\cdot)$ is locally convex in D and N (see Wilson (1993), pp. 318–319). We will return to this issue in more detail below, after the profit-maximizing tariff function has been characterized.

Finally, for conditions (2) and (3) to represent a true cost minimum, so-called participation conditions must also be fulfilled. These relate to the question of whether it is optimal to hold deposits and use payment services at all. The answer is positive if the following condition holds:

² In practice, tax-minimization reasons could cause one classification of the intramarginal tariff to be preferred to another. See the previous chapter for an analysis of tax arbitrage in the deposit market.

$$G(Y, \hat{D}, \hat{N}) + P(\hat{D}, \hat{N}) + r\hat{D} - G(Y, 0, 0) < 0. \quad (4)$$

The participation conditions are crucial for the analysis of price discrimination, for they determine the upper limit to the revenue which can be extracted from any single individual. From the tariff design point of view, participation conditions are therefore usually called participation constraints. In the present context, the economics behind participation constraints relates to alternative means of payment. That deposits are not held and payment services (produced by banks) are not used given a positive income level and transactions volume is obviously possible only because there exists the alternative of using currency as the only exchange medium. In constructing the transactions cost function $G(Y, D, N)$, the role of currency is not explicitly taken into account, however. This simplification, which is made for the sake of tractability, does not affect our results in a material way.³

3 Deposit pricing

We consider a case where deposits and payment services are supplied by a monopoly bank serving a given population of depositors. The bank invests the deposited funds in securities which yield interest at an exogenously given rate r . In the process, it incurs proportional intermediation costs of δ units per unit of intermediated funds, so that the net rate of return on investment is $r - \delta$. For simplicity, the marginal production costs of payment services are assumed to be constant, denoted by c . The interest paid to depositors and the service charges levied on them are described by the nonlinear tariff function $P(D, N)$, as described above. Under these assumptions, the profit of the bank is given by

³ If currency holdings were explicitly taken into account, the cost-minimization problem of a representative individual could be written as: $\min C(Y, M, D, N) + P(D, N) + rD + rM$, where minimization would now happen with respect to deposits D , payment services N and currency holdings M . By optimizing with respect to M , we obtain $M^* = M(Y, D, N, r)$, so that if we set $G(Y, D, N) \equiv C(Y, M(Y, D, N, r), D, N)$, the problem is of the same type as that in the main body of the text, with the exception that the interest rate r enters the cost function as a parameter.

$$\Pi = \int_{Y^{\text{Min}}}^{Y^{\text{Max}}} f(Y)[P(D,N) + (r-\delta)D - cN]dY. \quad (5)$$

The bank maximizes this expression subject to the constraint that the consumers choose D and N as defined by formulas (2) and (3) above, provided that the sufficient conditions for depositor optima are satisfied.

Mirrlees (1971) demonstrated how this kind of nonlinear tariff problem can be solved by transforming it into a standard control problem. The idea is to treat the type parameter (Y) as the independent variable; the maximand or minimand of the individuals (here the total liquidity cost Γ) as the state variable, and the quantities (here D and N) as the control variables. Once the profit-maximizing assignments $D(Y)$ and $N(Y)$ are found, it may be possible to use the first-order conditions for individual optima to solve for the marginal tariffs.

In the first instance, the Mirrlees approach requires eliminating the tariff altogether from the profit function. This can be done by using the fact that, at the optimum of the representative depositor, total liquidity cost $\Gamma(Y)$ can be expressed as follows:

$$\Gamma(Y) = G[Y, \hat{D}(Y), \hat{N}(Y)] + P[\hat{D}(Y), \hat{N}(Y)] + r\hat{D}(Y), \quad (6)$$

from which

$$P[\hat{D}(Y), \hat{N}(Y)] = \Gamma(Y) - G[Y, \hat{D}(Y), \hat{N}(Y)] - r\hat{D}(Y). \quad (7)$$

To construct the Hamiltonian function for the problem at hand, the "equation of motion" for $\Gamma(Y)$ is needed. Totally differentiating the definition of total liquidity cost (1) and using the first-order conditions for depositor optimum (2) and (3), the equation of motion is obtained as the following envelope condition:

$$\Gamma_Y = G_Y(Y, D, N) \quad (8)$$

The Hamiltonian is now

$$H = f(Y)\{\Gamma(Y) - G[Y, D(Y), N(Y)] - cN - \delta D\} + \lambda(Y)G_Y(Y, D, N). \quad (9)$$

First-order conditions for the optimal quantity assignments $D(Y)$ and $N(Y)$ include the following equations:

$$f(Y)[G_D(Y, D, N) + \delta] = \lambda(Y)G_{DY}(Y, D, N) \quad (10)$$

$$f(Y)[G_N(Y, D, N) + c] = \lambda(Y)G_{NY}(Y, D, N) \quad (11)$$

$$\Gamma_Y(Y) = G_Y(Y, D, N) \quad (12)$$

$$\lambda_Y = -f(Y). \quad (13)$$

We must also consider the transversality conditions of the problem. The starting point for the state variable can be fixed by referring to the participation constraint as follows. Observe that, at the monopolist's optimum, the participation constraint must be binding for some depositors – otherwise some freely available profits would be lost. Moreover, if the net participation benefit from being a customer of a bank, ie the difference $G(Y, 0, 0) - G(Y, D(Y), N(Y)) + P(D(Y), N(Y)) + rD$, is increasing in Y , the participation constraint can only be binding at the lowest point of the income distribution. Here, this monotonicity property is assumed; that it is actually possible will be demonstrated by a quadratic example in the next section. By (13), the costate variable λ is decreasing in Y ; it will be zero at Y^{\max} . Together, the transversality conditions are:

$$\Gamma(Y^{\min}) = G(Y^{\min}, 0, 0); \quad \lambda(Y^{\max}) = 0. \quad (14)$$

The terminal point constraint enables us to integrate the equation of motion for λ to obtain

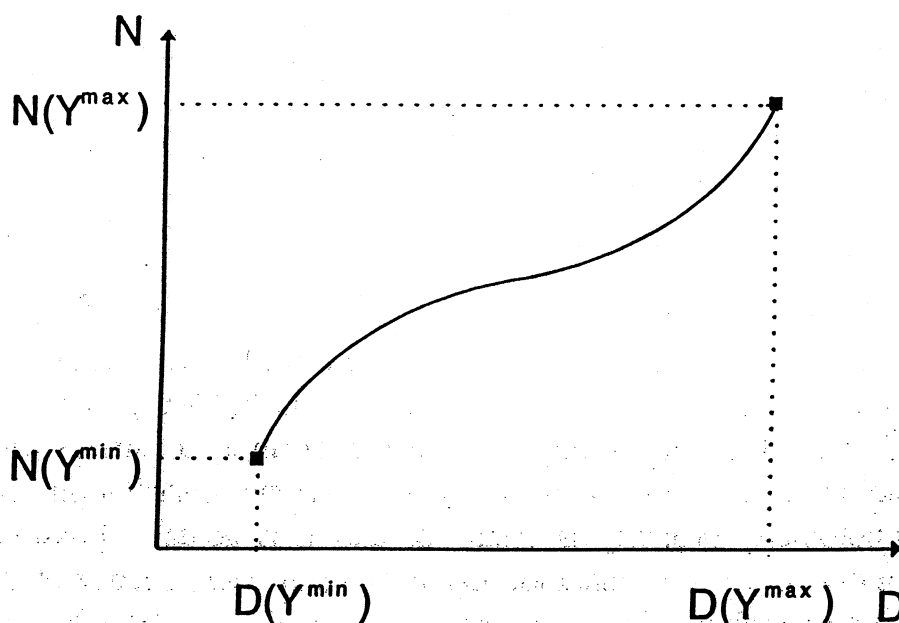
$$\lambda(Y) = 1 - F(Y). \quad (15)$$

Equations (10), (11) and (15) implicitly define the assignments of deposits and payment services to depositors at different points of the

Y-distribution as functions of Y. Recall that these assignments are voluntarily chosen by the depositors on the basis of a nonlinear tariff which is implicit in the solution but will be characterized below.

In the well-behaved case where both of the functions $D = D(Y)$ and $N = N(Y)$ defined by (10), (11) and (15) are monotonic (both strictly increasing, for example), all customers will be induced to choose bundles of D and N such that they lie on a single locus on the D, N surface (Figure 1). By their position on this locus, the customers reveal their type Y. This property of the solution for nonlinear tariff design problems is typical of problems where customers are heterogenous with respect to a single type parameter (Mirman and Sibley (1980); Wilson (1993), pp. 327–331).

Figure 1. An equilibrium locus of (D,N) bundles



What kind of tariff leads to the voluntary choice of the quantity assignments derived above? By applying the depositors' first-order optimality conditions to the results (10) and (11) and using (15), we are able to characterize the optimal tariff with the following equations:

$$-P_D(D, N) = (r - \delta) + \frac{[1 - F(Y)]}{f(Y)} G_{DY}(Y, D, N) \quad (16)$$

$$P_N(D, N) = c - \frac{[1 - F(Y)]}{f(Y)} G_{NY}(Y, D, N). \quad (17)$$

Formula (16) gives the marginal interest rate on deposited funds; (17) gives the marginal service charge. These conditions are analogous to the marginal price functions derived by Mirman and Sibley. These conditions are useful, of course, only if the depositors' first-order conditions suffice to define global optima. Provided that this is the case, the following properties of the optimal tariff emerge:

i) The depositors at the top end of the income distribution (for whom $Y = Y^{\max}$) are charged marginal prices which correspond to marginal costs. This follows from the fact that, under the assumptions which we made on the distribution $F(Y)$, we obtain $[1 - F(Y^{\max})]/f(Y^{\max}) = 0$. In the market for deposits, this implies the result $-P_D(D(Y^{\max}), N(Y^{\max})) = r - \delta$, meaning that, for the largest deposits, the marginal deposit rate equals the bank's marginal net rate of return on investment. As regards pricing in the payment services market, we obtain the result $P_N(D(Y^{\max}), N(Y^{\max})) = c$, implying that (again for the top-bracket customers) the marginal service charge equals the marginal cost of payment services.

ii) The depositors who are located at the interior points or at the low end of the income distribution ($Y^{\min} \leq Y < Y^{\max}$) are charged marginal prices which generally deviate from the marginal cost. Because the hazard rate $f(Y)/[1 - F(Y)]$ is always positive for these depositors, the sign of the price-cost margin on service i ($i = D, N$) is necessarily opposite to that of the cross derivative G_{Yi} . So, if $G_{YD} > 0$, we obtain marginal deposit rates below $r - \delta$; this is the case which seems to prevail in actual markets. In the payment services market, correspondingly, the pricing is "above cost" ($P_N > c$) if $G_{YN} < 0$. However, the optimality conditions suggest below-cost pricing of payment services ($P_N < c$) if $G_{YN} > 0$.

Note that while equations (16) and (17) may be used to characterize the optimal tariff, they alone do not suffice to determine the precise form of the tariff. This is apparent from the fact that the right-hand sides of these equations contain the type parameter Y , whereas the tariff by definition has to be independent of Y . However, once explicit assumptions are made on the functional form of the cost function $G(\cdot)$ and the form of the distribution $F(Y)$, the actual tariff implied by these

may be derived. Instructive examples of how this is done are given by Spulber (1981) and Wilson (1993), for instance.

In accordance with intuition, the above results suggest that the monopoly bank exercising second-degree price discrimination will generally not set the interest paid on deposits equal to its marginal rate of return on investment, nor will it generally set the service charge on payment services equal to the marginal cost of producing these services. Different cases can arise depending on the signs of the cross derivatives of the transactions cost function. The most interesting one of these, cross-subsidization of payment services, occurs if $G_{YD} < 0$ and $G_{YN} > 0$.

All of these predictions of profit-maximizing pricing behaviour are, of course, conditional on the compatibility of the first-order optimality conditions with the sufficient conditions for customers' cost-minimization. This matter is especially interesting in the case of predicted cross-subsidization, which has usually been excluded by certain regularity assumptions which have been adopted for the purpose of ensuring the validity of the second-order conditions for customers' local optima. For this reason, the next section of this paper is dedicated to the analysis of the validity of the first-order conditions.

4 The existence of cross-subsidizing optima

The control-theoretic approach to tariff design relies heavily on the first-order necessary conditions for profit and utility-maximization. However, the first-order conditions are not necessarily relevant for all shapes of the cost function $G(Y, D, G)$ or the type distribution $F(Y)$. This raises the question as to whether the first-order conditions which were derived above satisfy the sufficient conditions for global optima for the customers and the monopoly bank. In other words, is there a genuine, incentive-compatible, separating equilibrium? Can class or classes of cost functions and income distributions be indicated for which the first-order conditions satisfy the relevant sufficient conditions as well?

The conditions for validity of the first-order approach in problems of economic design have been discussed mostly in the principal-agent literature (Rogerson (1985); Jewitt (1988)), where it has been established that sufficient conditions are not trivially fulfilled. In nonlinear pricing literature, the validity of the first-order approach has

usually been guaranteed by very strong regularity assumptions. The most important of such assumptions is the sorting condition. In single-product problems, this condition requires that the size of customers' optimal purchases and their marginal valuation of the good are both increasing functions of their type (see Tirole (1988), pp. 154–157). When multiproduct pricing has been analyzed, an additional assumption of "symmetry" has been utilized, stipulating that the marginal benefit from each good is strictly increasing in the customer's type (see eg Roberts (1979), Mirman and Sibley (1980), Wilson (1993), p. 318).

In the present context, the symmetry assumption would require $G_{YD} < 0$ and $G_{YN} < 0$. A direct and obvious consequence of the symmetry assumption would be the impossibility of marginal subsidies. However, while the symmetry assumption is convenient for tractability, it is not necessary. The sufficient conditions for optimality may well be fulfilled even when symmetry is violated. Below, it will be demonstrated that a separating equilibrium characterized by first-order conditions such as those derived above can be incentive-compatible even when it involves marginal cross subsidies. The argument proceeds in two stages. First, the second-order necessary condition for a separating equilibrium (for which the first-order approach is valid) is developed; second, a quadratic example is considered under which the second-order condition is easy to check and also guarantees global optimality.

The analysis of the incentive compatibility of the separating equilibrium suggested by the first-order conditions is generally quite complicated in multiproduct pricing problems. However, the task is greatly simplified in models such as the one developed in this essay, where customer heterogeneity is described by a single type parameter. In this simple case, the incentive compatibility problem may be analyzed as a unidimensional problem. This is based on representing the customers' choices by the type they "reveal" through their demand behaviour. From this perspective, it is sufficient for incentive compatibility if it is optimal for each individual to reveal his own type through his demand behaviour. In particular, if the customers' preference functions can be shown to be globally concave with respect to their revealed type, the first-order conditions do define a global optimum within the set of points belonging to the optimal assignment locus. The area outside the locus need not be similarly investigated, for there the tariff can always be designed so that all customers stay on the locus (Mirman and Sibley (1980)).

Specifically, in the present context, the first-order condition defines a global minimum at least if the total liquidity cost given by

$G(Y, D(Z), N(Z)) + P(D(Z), N(Z)) + r \cdot D(Z)$ is strictly globally convex with respect to Z . Here Z is the "revealed Y " indicating the customer's choice of position on the $\{D(Y), N(Y)\}$ locus. The checking of this condition is facilitated by the following result (Roberts (1979), p. 82).⁴

Define $K(Y, Z) \equiv G(Y, D(Z), N(Z)) + P(D(Z), N(Z)) + r \cdot D(Z)$. This is the total liquidity cost incurred by the customer of type Y when he behaves like a customer of type Z . Note that the first-order conditions now read $K_Z(Y, Y) = 0$. Totally differentiating this yields the result $K_{ZZ}(Y, Y) = -K_{YZ}(Y, Y)$. In terms of the present model, this result implies that $K_{ZZ}(Y, Y) > 0$ and the problem is thus locally convex if $K_{YZ}(Y, Y) < 0$, a condition which can be written out as

$$G_{DY}(Y, D(Y), N(Y)) \cdot \frac{\partial D(Y)}{\partial Y} + G_{NY}(Y, D(Y), N(Y)) \cdot \frac{\partial N(Y)}{\partial Y} < 0. \quad (18)$$

Global convexity requires $K_{ZZ}(Y, Z) > 0$ for all Z , not just $Z = Y$; it can be shown that this is satisfied if $K_{YZ}(Y, Z) < 0$ (for a proof, see Tirole (1988), p. 156n). In the present model, this condition amounts to

$$G_{DY}(Y, D(Z), N(Z)) \cdot \frac{\partial D(Z)}{\partial Z} + G_{NY}(Y, D(Z), N(Z)) \cdot \frac{\partial N(Z)}{\partial Z} < 0. \quad (19)$$

According to condition (19), the first-order conditions characterize a global optimum at least if a) assignments grow with the type, in the sense that $\partial D/\partial Z$ and $\partial N/\partial Z$ are both positive, and b) cross derivatives G_{DY} and G_{NY} are both negative for all Y . The (b) part of the condition is the symmetry assumption mentioned above.

However, convexity under normality with respect to type is also possible when cross derivatives are of opposite signs. This is due to the fact that, if the goods D and N are complementary, the assignments may be increasing in type even though the direct effects

⁴ In some (more complex) cases, so-called "bunching" equilibria may emerge, in which many customers with a range of different types purchase the same quantities. Features which may give rise to bunching include income effects on the part of the consumer, and fixed costs on the part of the producer (Wilson (1993), pp. 166, 203). In the well behaved case where second-order conditions suffice for global incentive compatibility, the equilibrium is a separating one and there is no bunching. Note that income effects are excluded in the model presented in this chapter by assuming that the tariff has no impact on Y .

of the customer type on their marginal valuation is asymmetric. In other words, one of the terms in (19) can be positive without violating the inequality. In this case, however, convexity depends on the relative magnitudes of the relevant derivatives.

The easiest way to demonstrate that convexity of the consumer's problem is compatible with asymmetric valuation effects is to use a specific example. Consider the case where the internal cost function $G(\cdot)$ is quadratic, with the properties (i)–(vi) above, and the distribution $F(Y)$ has an increasing hazard rate. In the quadratic case, the derivatives of the assignment functions are the following:

$$\begin{aligned}\frac{\partial D}{\partial Z} &= (1 + h(Z)/H(Z)^2) \frac{G_{DN}G_{NY} - G_{DY}G_{NN}}{\Delta} \\ \frac{\partial N}{\partial Z} &= (1 + h(Z)/H(Z)^2) \frac{G_{DN}G_{DY} - G_{NY}G_{DD}}{\Delta}\end{aligned}\quad (20)$$

where $H(Z) = f(Z)/[1 - F(Z)]$

$$h(Z) = \partial H(Z)/\partial Z$$

and $\Delta = G_{DD}G_{NN} - G_{DN}G_{DN}$.

These results can now be substituted into the inequality (19). Under the assumptions on the shape of the $G(\cdot)$ function (assumption (vi) above) we have $\Delta > 0$. Further, the assumption of an increasing hazard rate means that $(1 + h(Z)/H(Z)^2) > 0$. The global sufficiency condition (19) can then be reduced to the following form:

$$(G_{DY})^2 G_{NN} + (G_{NY})^2 G_{DD} - 2G_{DY}G_{NY}G_{DN} > 0. \quad (21)$$

This can be broken down into two cases, depending on whether the cross derivatives G_{DY} and G_{NY} are of different sign or not:

$$G_{DN} > \frac{(G_{DY})^2 G_{NN} + G_{DD} (G_{NY})^2}{2G_{DY}G_{NY}}, \quad \text{if } G_{DY}G_{NY} < 0 \quad (22)$$

$$G_{DN} < \frac{(G_{DY})^2 G_{NN} + G_{DD} (G_{NY})^2}{2G_{DY} G_{NY}}, \text{ if } G_{DY} G_{NY} > 0. \quad (23)$$

Thus, in the quadratic case with an increasing hazard rate in the depositors' income distribution, the convexity of the problem relates to the complementarity of the products as captured by the derivative G_{DN} . If the cross derivatives G_{DY} and G_{NY} are of opposite signs, G_{DN} must not be too negative (condition 22); in the other case, it must not be too positive (condition 23).

Let us now focus our attention on a particularly interesting case which can be called "normality with respect to type". In economic terms, this implies that the customers' cost or preference functions are such that, if goods are available at fixed marginal prices, the demands for both of the goods are increasing functions of the customer's type.⁵

It is easy to show that, given the assumptions (i) to (vi) above, normality with respect to type requires that the following inequalities hold:⁶

$$G_{DY} G_{DN} - G_{DD} G_{NY} > 0 \quad (24)$$

$$G_{NY} G_{DN} - G_{NN} G_{DY} > 0. \quad (25)$$

Clearly, these inequalities must hold if G_{DN} and G_{NY} and G_{DY} are all negative. In that case, increasing the income level Y makes both deposits D and payment services N more efficient cost-saving factors at the margin. However, normality with respect to type is also possible in the asymmetric cases when the cross derivatives G_{DY} and G_{NY} have different sign. When $G_{NY} > 0$ and $G_{DY} < 0$, for example, normality with respect to type requires $(G_{DY}/G_{NY}) \cdot G_{NN} < G_{DN} < (G_{NY}/G_{DY}) \cdot G_{DD} < 0$.

Consider again the case of quadratic transactions costs and an income distribution with an increasing hazard rate. Assume also that the necessary condition for cross-subsidization holds, ie $G_{DY} G_{NY} < 0$.

⁵ The properties of the $G(\cdot)$ -function which imply normality with respect to type are mathematically equivalent to the properties of production functions which give rise to positively sloped expansion paths.

⁶ This is done by applying the implicit function theorem to the customer's first-order conditions (2) and (3) and setting $P_{DD} = P_{NN} = P_{DN} = 0$. The conditions are for $dD/dY > 0$ and $dN/dY > 0$.

Then for both incentive compatibility and normality with respect to type conditions to be fulfilled, the value of the cross derivative G_{DN} must be a negative number in the range

$$0 > \frac{G_{NY}G_{DD}}{G_{DY}} > G_{DN} > \frac{G_{NN}(G_{DY})^2 + G_{DD}(G_{NY})^2}{2G_{DY}G_{NY}}. \quad (26)$$

Such a range exists, and cross-subsidization is thus compatible with normality with respect to type and incentive compatibility, if

$$\frac{(G_{NY})^2}{G_{NN}} < \frac{(G_{DY})^2}{G_{DD}}. \quad (27)$$

The interpretation of conditions (26) and (27) is that cases where profit-maximizing nonlinear tariffs display cross subsidies between goods which are normal with respect to type may occur only if the goods are complementary enough and the asymmetry in the effects of the type on the marginal valuations of the goods is not too extreme.

To conclude the analysis of the quadratic case, we must check the validity of the assumption that the participation constraint holds only at the point $Y = Y^{\min}$. This assumption was used in solving the maximum principle conditions (10)–(13). We will see that this assumption actually amounts to a requirement concerning the shape of the internal cost function G . This can be demonstrated as follows.

It was noted in the previous section that the assumption that the participation constraint is binding only at the low end of the type distribution is certainly valid if the net participation benefit $B = G(Y, 0, 0) - G(Y, D(Y), N(Y)) + P(D(Y), N(Y)) + rD$ is increasing in Y . Now, differentiating B with respect to Y and applying the envelope theorem yields $B_Y = G_Y(Y, 0, 0) - G_Y(Y, D, N)$. When G is quadratic, this can be developed into $B_Y = -(G_{DY}D + G_{NY}N)$, and the condition $B_Y > 0$ becomes equivalent to $G_{DY}D + G_{NY}N < 0$. This is clearly ensured for all $D = D(Y)$ and $N = N(Y)$ if it holds for $D^{\min} = D(Y^{\min})$ and $N^{\min} = N(Y^{\min})$, and the convexity condition (18) holds, too. On this basis, it is immediately obvious that the assumption concerning the participation constraint is valid if the marginal valuation effects of the type are symmetric ($G_{DY} < 0$ and $G_{NY} < 0$) and the minimum assignments D^{\min} and N^{\min} are positive. However, the assumption can clearly be valid in the asymmetric case as well. Consider for example the case where $G_{DY} < 0$ and $G_{NY} > 0$, leading to marginal subsidies on payment services. In that case, the net benefit of

participation is increasing at Y^{\min} if the cost function is such that $-G_{DY}/G_{NY} > N^{\min}/D^{\min}$. Coefficients of the linear terms of the quadratic $G(.)$ function provide sufficient degrees of freedom for this condition to be satisfied without violating the other regularity conditions referred to above. If the convexity conditions developed above also hold, net participation benefits are increasing for all Y and the required assumption concerning the participation constraint is valid.

5 An inventory-theoretic cost function

In this section we look at a specific example. We derive a transactions cost function in a model of multiple means of payment and show that it naturally displays the properties which in the above analysis led to cross-subsidization of payment services. This is seen by examining its cross derivatives with respect to the relevant variables.

The cost function is derived in the Baumol-Tobin tradition, with two extensions. First, we assume positive transaction costs associated with the spending of money, unlike in the original Baumol-Tobin framework.⁷ Second, we assume two means of payments, which are substitutable with each other: currency and bank deposits.

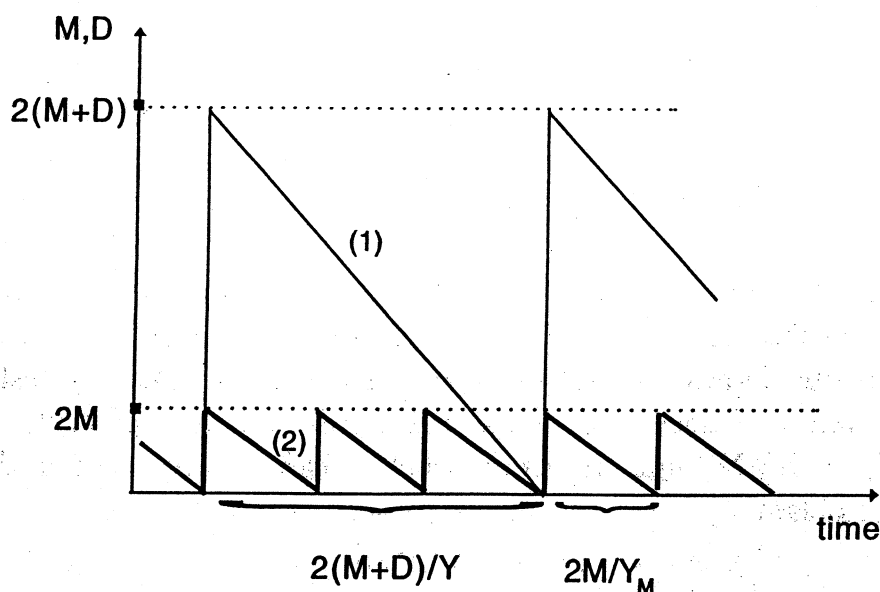
We consider the behaviour of a representative individual over some fixed time period (say a year). The individual holds currency and deposits. During the period, the individual uses a given amount of money Y in the form of payments, which he can make either with currency or with a transfer of deposit money. The total frequency of payments is X and the frequency of those paid with deposits is N . These are volume magnitudes. In value terms, we denote the part of expenditure paid with cash by Y_M and the part paid with deposit money by Y_D . By definition, $Y = Y_M + Y_D$. For simplicity, the individual's disposable income, which is also Y , is assumed to be received in the form of an investment asset ("bonds").

The payment patterns arising from the use of the income flow are assumed to be as depicted in Figure 2, where the payment flows from monies to goods are drawn as continuous for simplicity. Periodically,

⁷ The case of discrete payments in "goods transactions", ie in the ultimate use of money, has been analyzed by Santomero (1979), for example. In his analysis, the payment pattern is entirely endogenous. We do not go that far, but assume an exogenous distribution of payments instead.

the individual sells some bonds in order to increase his money holdings ($M + D$). As in the Baumol-Tobin framework, the frequency of these transactions is $Y/[2(M + D)]$, where M denotes average currency holdings and D denotes average deposit holdings. Both monies are used for making payments. It is assumed that the initial currency balance $2M$ is depleted to zero much quicker than the initial deposit balance $2D$, meaning that $M/Y_M < D/Y_D$. Each time the currency balance reaches zero it is restored to the level $2M$ by a withdrawal from the deposit balance until the deposit balance too has been depleted and the payment cycle starts anew (here, we follow the tradition of neglecting cumbersome integer constraints, see eg Barro (1976)).

Figure 2. The time pattern of money holdings in the two payment media model



1 = currency + deposits
2 = currency

The costs arising from this payment pattern are assumed to derive from four different sources:

1. Holding costs of the currency (loss, inconvenience of storage and carrying), which is proportional to the amount of currency held. These costs are measured by the parameter h .
2. Transactions costs of paying with deposit money, with the exception of outright service charges to the bank (including the

value of time lost in making payments, for example). The cost per deposit payment is measured by the parameter q .

3. Transactions costs of currency withdrawals from the bank. These are assumed to be fixed and are measured by the parameter k_M .
4. Transactions costs incurred when "bonds" are sold for money. These are also fixed and are measured by the parameter k_D .

Adding these cost items, the transactions costs C may be written as follows:

$$C = hM + k_M \frac{Y_M}{2M} + qN + k_D \frac{Y}{2(D+M)}. \quad (28)$$

There is an exogenous size distribution of payments over some interval $[p^{\min}, p^{\max}]$. This distribution is defined as a function $H(Y, x)$, indicating the value of payments due to the x largest transactions. For tractability, the function is assumed to be continuous and twice differentiable. From the definition it follows that $\lim_{x \rightarrow \infty} H(Y, x) = Y$ as $x \rightarrow \infty$ and $H(Y, 0) = 0$ for all Y . By definition, the function must be increasing and concave $\partial H(Y, x) / \partial x > 0$, $1 > \partial H(Y, x) / \partial Y > 0$, and $\partial^2 H(Y, x) / \partial^2 x < 0$. Now, given the cost structure described above, and given that both Y_M and Y_D are positive, it will always be rational to pay the largest transactions with deposits and the smallest one with currency. Hence, the total value of deposit payments is simply $Y_D = H(Y, N)$ and, further, $Y_M = Y - H(Y, N)$. Substituting this for Y_M we obtain

$$C = hM + k_M \frac{Y - H(Y, N)}{2M} + qN + k_D \frac{Y}{2(D+M)}. \quad (29)$$

In order to differentiate the transactions cost function in a way comparable to the derivatives of $G(\cdot)$ used in the previous sections of this essay, we must allow for optimal adjustments in the currency holdings M . The demand for currency as a function of Y , D , N and r may be derived by minimizing the transactions cost plus opportunity cost of currency holdings $r \cdot M$ with respect to M . The first-order condition for optimum demand for currency is

$$\frac{\partial(C+rM)}{\partial M} = h+r - \frac{k_D Y}{2(D+M)^2} - \frac{k_M[Y-H(Y,N)]}{2M^2} = 0. \quad (30)$$

From this we obtain the following results on the adjustment of M when D , N , or Y change:

$$M_Y = \frac{\frac{k_M \left(1 - \frac{\partial H}{\partial Y}\right)}{2M^2} + \frac{k_D}{2(D+M)^2}}{\Omega} > 0$$

$$M_N = \frac{-k_M \left(\frac{\partial H}{\partial N}\right)}{2\Omega M^2} < 0 \quad (31)$$

$$M_D = \frac{-k_D Y / (D+M)^3}{\Omega} < 0, \text{ and } -1 < M_D$$

where $\Omega = \frac{k_D Y}{(D+M)^3} + \frac{k_M [Y-H]}{M^3} > 0.$

Here the subscripts in M_Y , M_N and M_D indicate partial derivatives of currency demand with respect to Y , N and D , respectively. The reactions of the demand for currency to changes in income, bank-intermediated payments and the stock of deposits are intuitively plausible. Increases in income will induce greater demand for currency. By contrast, if more payments are made with bank deposits, currency demand will decrease. Deposit holdings have a negative effect on currency, too. The result that deposits do not fully crowd out currency ($M_D > -1$) is important for the cross derivatives which are evaluated below. Now we can find the derivatives of the cost function $G(Y,D,N) \equiv C(Y, M(Y, D, N), D, N)$ with respect to its arguments:

$$\frac{\partial G}{\partial Y} = \frac{k_M \left(1 - \frac{\partial H}{\partial Y}\right)}{2M} + \frac{k_D}{2(M+D)} > 0$$

$$\frac{\partial G}{\partial D} = \frac{-k_D Y}{2(D+M)^2} < 0 \quad (32)$$

$$\frac{\partial G}{\partial N} = q - \frac{k_M \left(\frac{\partial H}{\partial N}\right)}{2M}$$

These are as envisaged in assumptions (i)–(iii) provided that the transaction cost q of paying with deposit money is not too high, so that the derivative $\partial G/\partial N$ can obtain its assumed negative value. If the demand price of payment services is to be positive, these parameter values are necessary.

The second derivatives, which are crucial for both incentive compatibility and cross-subsidization, are presented in (33a) and (33b):

$$\frac{\partial^2 G}{\partial D^2} = \frac{k_D(1+M_D)Y}{(D+M)^3} > 0$$

$$\frac{\partial^2 G}{\partial N^2} = k_M \left\{ \frac{M_N \left(\frac{\partial H}{\partial N}\right)}{2M^2} - \frac{\partial^2 H}{\partial^2 N} \right\} \quad (33a)$$

$$\frac{\partial^2 G}{\partial D \partial N} = \frac{k_D Y M_N}{(D+M)^3} < 0.$$

$$\frac{\partial^2 G}{\partial D \partial Y} = \frac{-k_D(1+M_D)}{2(D+M)^2} - \frac{k_M \left(1 - \frac{\partial H}{\partial Y}\right) M_D}{2M^2} < 0 \quad (33b)$$

$$\frac{\partial^2 G}{\partial N \partial Y} = \frac{k_M \left(M_Y \frac{\partial H}{\partial N} - M \frac{\partial^2 H}{\partial N \partial Y} \right)}{2M^2}$$

The first three of the second derivatives in (33a) determine how well the cash management problem at hand behaves. The signs are as required, if the size distribution of payments fulfils certain requirements. In particular, the (strict) convexity of the cost function requires the cumulative size distribution of payments to be sufficiently curved. More specifically, it is required that $M_N/M > (\partial^2 H / \partial^2 N) / (\partial H / \partial N)$; otherwise the second derivative with respect to N is not positive and the determinant condition for convexity of G does not hold. The interpretation of this condition is that when more payments are made with deposits (ie N increases), the size of the smallest deposit payment must decrease faster (in relative terms) than the demand for currency. Suppose, for example, that the size distribution is exponential so that $Y_D = Y(1 - \text{Exp}(-aN))$. In this case, it is straightforward to show that the $\partial^2 G / \partial N^2 > 0$, and the determinant condition for convexity also holds.

The last two of the second derivatives, $\partial^2 G / \partial D \partial Y$ and $\partial^2 G / \partial N \partial Y$, in (33b) determine how changes in income affect the marginal valuation of bank services. This, in turn, determines how a (monopoly) bank serving customers with this type of liquidity costs will construct its tariff. If these derivatives are both negative, the bank will collect profits on both of the products D and N. However, if $\partial^2 G / \partial N \partial Y$ is positive, there will be marginal subsidization of payment services. Now, the sign of this derivative depends on the term $\partial^2 H(Y,N) / \partial Y \partial N$. This describes the effect of income on the size of the smallest deposit transaction when the number of deposit transactions is held constant. If this term is not positive and too large, then $\partial^2 G / \partial N \partial Y > 0$ and marginal subsidies emerge. The precise condition is

$$\frac{\partial^2 H(Y,N)}{\partial Y \partial N} < \frac{M_Y}{M} \cdot \frac{\partial H(Y,N)}{\partial N} \quad (34)$$

The right-hand side of the condition (34) is positive.

To illuminate this, consider the special case where the size of the N :th largest payment made is independent of the income of the individual. Then the term $\partial^2 H / \partial Y \partial N$ is zero and the condition (34) is satisfied. This is an example of a situation where the bank would cross-subsidize payment services in the profit-maximizing equilibrium.

An intuitive explanation of the asymmetry which arises in the model if the direct effect on income or the size of the smallest deposit transaction is weak can perhaps be developed along the following lines. The marginal benefit of holding deposits is proportional to the cost of bond liquidations, and with given deposits, bond liquidation costs increase when income grows. On the other hand, the marginal benefit of payment services is determined at the currency/bank payment margin. There, an increase in income, if the use of payment services is held constant, means that more payments are made with currency. That means more cash withdrawals and a lower marginal benefit from payment services.

6 Discussion

The analysis presented in this essay is based on the observation that, given a monopolistic market structure, the market for cheque accounts clearly satisfies the conditions for price discrimination to be possible. From this starting point, we study the problem of optimal pricing of cheque accounts, emphasizing that the cheque account market consists of two submarkets: the market for deposited funds and the market for payment services. These cannot be considered a single good, because depositors are free to vary the ratio of (average) deposited funds to the volume of payment services they use. However, the two submarkets are obviously interdependent. In the present analysis, the interaction is modelled through cross effects in the depositors' preference function (liquidity cost function) and the important feature that the depositors' participation decisions are made jointly for both markets.

Nonlinear multiproduct pricing often becomes quite complicated mathematically. To retain some tractability, we analyzed price discrimination in a model where depositors were heterogenous with respect to a single type parameter ("income") only. The analysis suggests that cross subsidies are quite possible as an outcome of profit-maximizing tariff design by a monopolist. That the cross-subsidizing solution is not only possible but even likely in a banking context was suggested by an inventory model of the parallel use of two payment media in Section 5.

Heuristically, cross-subsidization may occur when there are effects in the population (such as the variation of the type parameter) which increase the marginal valuation of one product and decrease the marginal valuation of another. We pointed out that, under complementarity, this asymmetry does not rule out normality with respect to type, ie the property that the realized demands of the two goods are positively related to each other in the population.

There is a well-known analogy between nonlinear pricing and bundling. For example, in the single-product case, quantity discounts can be analyzed as instances of bundling. The analogy can be extended to multiproduct bundles as well. Adams and Yellen (1976) and Schmalensee (1984) have analyzed examples of multiproduct pricing in which independent variation of pairs of reservation prices in the customer population causes bundling to be profitable. The incentives to bundle are especially strong when reservation prices are negatively correlated – something closely related to asymmetric valuation effects of the customer's type. These bundling models are very restrictive, however, in the sense that customers' demand behaviour is assumed to be of the (0,1) type, and thus the bundles are "fixed" by assumption. These assumptions are of course unsuitable for the analysis of banking, for instance. The analysis in the present essay seems to generalize Adams and Yellen's and Schmalensee's insights to the marginalist world of continuous demand functions and smoothly variable bundles.

A general property of optimal nonlinear tariffs is that marginal prices should equal marginal costs for those customers who purchase the largest quantities (or, to be more precise, for those customers whose net benefits from participation are the greatest). This property may perhaps be used to derive tests of whether price discrimination is indeed the reason why deposit pricing is often nonlinear and seems to involve cross subsidies. On the other hand, if one takes the price discrimination model as given, then marginal prices charged to the largest customers may be used to measure the marginal costs of bank services, which are not usually directly observable.

The full implications of the above findings have not yet been explored systematically. The welfare aspects of bank deposit pricing constitute an important area which must be left to later research. It is clear, however, that the results suggest that even under imperfect competition, and without interest rate regulation or distorting taxes, payment services may be supplied in excess of the "first-best" allocation. However, this does not necessarily imply that nonlinear pricing of cheque accounts is socially undesirable. Wilson (1993) points out that optimal nonlinear tariffs are a case of Ramsey pricing,

and Ramsey pricing is a second-best way of collecting sufficient revenue for natural monopolies to be viable without direct subsidies. Furthermore, as pointed out by Whitesell (1992), there may be negative externalities in the use of currency, and some subsidization of paying with deposit money may thus be warranted. The negative externalities stem from costs incurred by the central bank, for example, in the production and distribution of banknotes. These costs are, of course, not directly charged to agents using currency in payments.

References

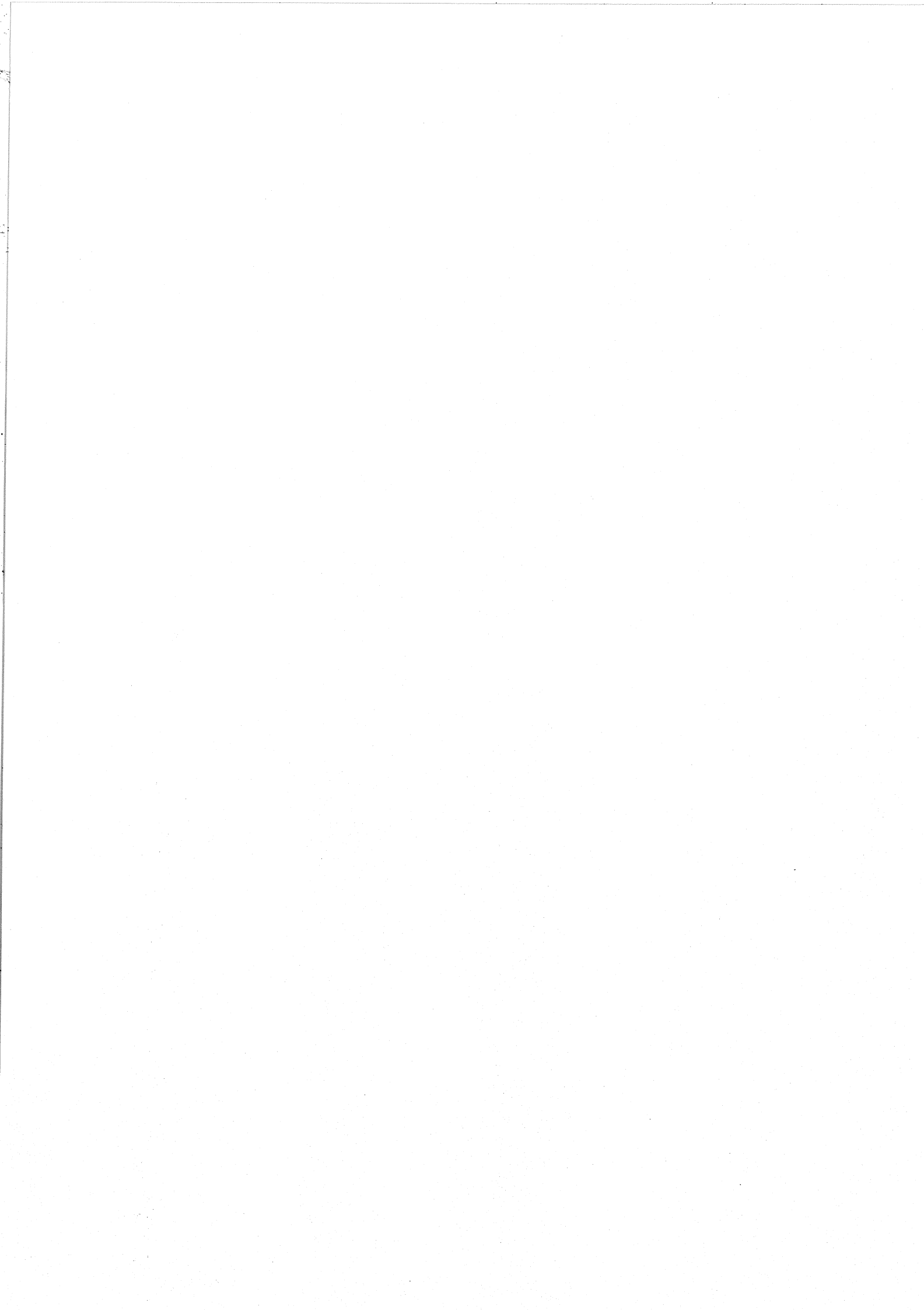
- Adams, W.J. – Yellen, J.L. (1976) **Commodity Bundling and the Burden of Monopoly.** *Quarterly Journal of Economics* 90, 475–498.
- Barro, R. (1976) **Integral Constraints and Aggregation in an Inventory Model of Money Demand.** *Journal of Finance* 31, 77–87.
- Barro, R. – Santomero, A. (1972) **Household Money Holdings and the Demand Deposit Rate.** *Journal of Money, Credit, and Banking*, May.
- Baxter, W. – Cootner, P. – Scott, K.E. (1977) **Retail Banking in the Electronic Age.** Allanheld, Osmun & Co., Montclair, N.J.
- Becker, W.E. (1975) **Determinants of the U.S. Currency – Demand Deposit Ratio.** *Journal of Finance* 30, 47–57.
- Davis, R. – Korobow, L. (1987) **The Pricing of Consumer Deposit Products – The Non-Rate Dimensions.** *Federal Reserve Bank of New York Quarterly Review* (winter 1987), 14–18.
- Feenstra, R. (1986) **Functional Equivalence between Liquidity Costs and Utility of Money.** *Journal of Monetary Economics* 17, 271–291.
- Fischer, S. (1983) **A Framework for Monetary and Banking Analysis.** *Economic Journal* (Conference Papers supplement), 1–16.
- Hörngren, L. (1988) **Kostnadsutveckling och konkurrens i banksektorn.** *Sveriges Offentliga Utredningar* 1989:16.
- Jewitt, I. (1988) **Justifying the First-Order Approach to Principal-Agent Problems.** *Econometrica* 56, 1177–1190.
- Klein, B. (1974) **Competitive Interest Payments on Bank Deposits and the Long Run Demand for Money.** *American Economic Review* 64, 931–945.
- Klein, M. (1971) **A Theory of the Banking Firm.** *Journal of Money, Credit, and Banking* 3 (May 1971), 205–218.
- Mirman, L. – Sibley, D. (1980) **Optimal Nonlinear Prices for Multiproduct Monopolies.** *Bell Journal of Economics* 11, 659–670.
- Mirrlees, J. (1971) **An Exploration in the Theory of Optimal Taxation.** *Review of Economic Studies* 38, 175–208.
- Mitchell, D. (1988) **Explicit Interest and Demand Deposit Service Charges.** *Journal of Money, Credit, and Banking* 20 (May 1988), 270–274.
- OECD (1993) **Bank Profitability. Financial Statements of Banks 1982–1991.** Statistical Supplement. OECD, Paris.

- Phlips, L. (1983) **The Economics of Price Discrimination**. Cambridge University Press, Cambridge (U.K.).
- Roberts, K. (1979) **Welfare Implications of Nonlinear Pricing**. *Economic Journal* 89, 66–83.
- Rogerson, W.P. (1985) **The First-Order Approach to Principal-Agent Problems**. *Econometrica* 53, 1357–1367.
- Santomero, A. (1979) **The Role of Transactions Costs and Rates of Return on the Demand Deposit Decision**. *Journal of Monetary Economics* 5, (May 1979), 343–364.
- Saving, T. (1979) **Money Supply Theory with Competitively Determined Deposit Rates and Activity Charges**. *Journal of Money, Credit, and Banking* 11, No. 1, 22–31.
- Schmalensee, R. (1984) **Gaussian Demand and Commodity Bundling**. *Journal of Business* 57, S212–S230.
- Spulber, D. (1981) **Spatial Nonlinear Pricing**. *American Economic Review* 71, 923–933.
- Tirole, J. (1988) **The Theory of Industrial Organization**. MIT Press, Cambridge, Mass.
- Vittas, D. – Frazer, P. – Metaxas-Vittas, T. (1988) **The Retail Banking Revolution: An International Perspective**. Lafferty Publications, London.
- Walsh, C. (1983) **Taxation of Interest Income, Deregulation and the Banking Industry**. *Journal of Finance* 28, No. 5, 1529–1542.
- Whitesell, W. (1992) **Deposit Banks and the Market for Payment Media**. *Journal of Money, Credit, and Banking* 24 (November 1992), 483–498.
- Wilson, R.B. (1993) **Nonlinear Pricing**. Oxford University Press, New York.

Switching Costs and Price Discrimination in the Bank Deposit Market

Contents

1	Introduction	145
2	The demand for bank services	146
3	Switching costs and participation	148
4	Tariff design	151
4.1	The first-order conditions	151
4.2	The participation constraint	155
4.3	The switching constraint	156
5	Incentive compatibility	158
6	Some properties of the tariff	161
7	Conclusions	164
	References	166



1 Introduction

Banks often apply complex nonlinear tariff structures in pricing the payment services used by depositors, at least when pricing is not restricted by regulation (see Vittas et al. (1988) and the introduction of the present study for details). Judging from the revenue and cost structures of banks, deposit pricing strategies seem to imply cross-subsidization of transaction (payment) services by deposit-taking activity. This tendency for complicated pricing strategies suggests that price discrimination may be important in the bank deposit market.

In the previous chapter, I applied a two-product model of second-degree price discrimination to the pricing of chequable bank deposits, or so-called transaction accounts. In that context, the depositor's banking relationship was assumed to consist of purchasing two products, ie the deposit facility and payment services. Some kind of imperfect competition must be assumed in models explaining price discrimination. The pricing problem of the bank was specified under the assumption of monopoly.

The purpose of the present chapter is to generalize the model of two-product nonlinear pricing developed in the previous chapter to a case where there is some competition. To make room for price discrimination, competition is not assumed to be perfect, however. Instead, it will be assumed that deposit markets are characterized by switching costs, as introduced by von Weizsäcker (1984) and Klemperer (1987). Switching costs are defined as once-and-for-all sunk costs incurred each time the consumer establishes a customer relationship with a new seller. As noted by Klemperer, the market for chequing accounts is one obvious area of application of the theory of switching costs, since "there are high transaction costs in closing an account with one bank and opening another with a competitor" (Klemperer (1987)).

As in the previous chapter, the analysis is presented in the deposit banking context. In the specific example used, consumers' demand behaviour is derived from minimizing a quadratic liquidity cost function (see Feenstra (1986) for a discussion of liquidity costs and the demand for monetary services). The heterogeneity of customers is modelled by assuming that the distribution of the (single) type parameter, or "income", is uniform over a given range. The income parameter determines the demands for bank services. There is a constant switching cost, which is equal for each depositor.

It is demonstrated that, within this formulation, parameter values exist which give rise to a cross-subsidizing pricing scheme. This is

interesting because in the standard one-good formulation of the price discrimination problem marginal subsidies (in the sense that marginal prices are below marginal costs) are not part of the optimal nonlinear tariff (see Wilson (1993)). In this essay, emphasis is put on the scrutiny of second-order conditions for customer choice, which determine the incentive compatibility (self-selection) constraints for tariff design. It also turns out that the switching cost model of price discrimination is less restrictive than the monopoly model, in the sense that less can be said about the properties of the optimal tariff. For example, the biggest customers need not pay marginal-cost-based marginal prices, as in the monopoly model.

2 The demand for bank services

The cheque account market is modelled as consisting of two submarkets, the market for deposited funds and the market for payment services. These two markets are connected by complementarity in customers' preferences and the restriction that there is a common participation constraint. Customers' preferences are modelled by assuming that they minimize a liquidity cost function describing how the use of banking services reduces the costs of earning and spending a given income flow (Feestra (1986)). Consumers are heterogeneous with respect to the "income" parameter of the liquidity cost function.

The previous chapter presented an analysis of general cost functions and type distributions, and the results were thus – inevitably – mainly qualitative. In the present essay, however, explicit functional forms are used and explicit solutions are obtained for quantities purchased and marginal prices charged. In this section, we present an analysis of the demands for bank services, conditional on the existence of a customer relationship with a particular bank. The questions of switching and market participation are taken up in the next section.

Assume that a representative customer behaves so as to minimize the following expression:

$$\Gamma = G(D, N, Y) + P(D, N) + r \cdot D. \quad (1)$$

The first term is an internal cost function, which has the "goods" D (the deposit balance), N (the volume of payment services used) and the type parameter Y ("income") as its arguments. The second term is a

general tariff function which gives the financial (pecuniary) costs of purchasing the goods D and N. The tariff function captures the net effect of all service charges and deposit interest, and is defined as a net outlay for the customer. Since the function $P(D, N)$ is not specified as a function of the customer type, we are ruling out first-degree price discrimination; since it is allowed to be nonlinear and is not constrained to pass through the origin, we do allow for second-degree price discrimination. The third term is the opportunity cost of funds for the depositor, which is determined by the exogenous "security market rate" r .

Suppose now that the customer has chosen the optimal D and N. We define the minimum attainable cost (if the customer stays with the bank) as

$$\hat{\Gamma}(Y) \equiv G(\hat{D}, \hat{N}, Y) + P(\hat{D}, \hat{N}) + r \cdot \hat{D}. \quad (2)$$

Here $\hat{D} = \hat{D}(Y)$ and $\hat{N} = \hat{N}(Y)$ denote values of D and N, respectively, which minimize Γ for the given $P(\cdot)$ function.

At this point we depart from the model of the previous chapter by giving the internal cost function an explicit form. It is assumed to be of quadratic form in D, N and Y:

$$G = a_1 D + a_2 N + a_3 Y + (\frac{1}{2})b_{11} D^2 + b_{12} DN + (\frac{1}{2})b_{22} N^2 + b_{13} DY + b_{23} NY + (\frac{1}{2})b_{33} Y^2. \quad (3)$$

This quadratic form is assumed to be convex in D and N so that we have $b_{11}b_{22} > 0$ and $b_{11}b_{22} - b_{12}b_{12} > 0$. It is also required¹ that $b_{13}^2 b_{22} - 2b_{12}b_{13}b_{23} + b_{11}b_{23}^2 > 0$. The first-order necessary conditions of the representative, type Y customer for optimal N and D are

$$-P_D(\hat{N}, \hat{D}) - r = a_1 + b_{11}\hat{D} + b_{12}\hat{N} + b_{13}Y \quad (4a)$$

¹ This parameter restriction is needed to ensure that the second-order conditions of the consumer's problem can hold. See Section 4 of the previous chapter and Section 5 below.

$$-P_N(\hat{N}, \hat{D}) = a_2 + b_{12}\hat{D} + b_{22}\hat{N} + b_{23}Y. \quad (4b)$$

These equations define the demand behaviour of the customer given a tariff structure and assuming that sufficient conditions for a true cost minimum hold at the point in the (D, N) space defined by (4a, b). Note that the marginal interest rate on deposits is $i = -P_D(D, N)$ and the marginal service charge is $P_N(D, N)$.

3 Switching costs and participation

Switching costs are a theoretical construct intended to capture some aspects of imperfect competition. Methodologically, they serve a similar function as adjustment costs in dynamic models, or transportation costs in spatial models. Switching costs can be defined as once-and-for-all sunk costs incurred when a new customer relationship is formed. The analytical significance of switching costs is that they allow imperfect competition even in the case where firms and their products are "functionally identical". Typically, the seller has some limited monopoly power over the consumer who is subject to switching costs.²

Klemperer (1992) mentions several categories of switching costs, including psychological costs and artificial costs (which are created by discount coupon plans etc.). There are also reasons for genuinely economic switching costs in many instances.

In the context of the bank deposit market, the transaction cost category is probably the most relevant motivation for the use of the switching cost model. This is so because the time and trouble which a customer must spend if he wants to move his bank account and all business related to it from one bank to another may be considerable.

We analyze the effects of switching costs on the bank deposit market under very simple assumptions. The model is static, which implies that there are no "customer flows" in the market in the form of entering or exiting customers.³ All customers are assumed to be part

² This conclusion depends on the ability of the sellers to discriminate between their present and potential customers; see Klemperer (1987).

³ The impact of the presence of new, unattached consumers is analyzed by Klemperer (1987b); see also Klemperer (1992).

of some bank's clientele, and the pricing problem is solved so that this situation is just barely preserved.

The competition restricting the banks' pricing conduct is modelled by assuming that the markets are contestable⁴ in the following way. There is free entry without any firm-specific sunk costs. From this it follows that there is a potential competitive fringe in the deposit market in which any customer is able to obtain the products D and N (deposits and payment services) at the break-even, marginal cost prices. In other words, if the representative customer quits his customer relationship with the representative bank and goes to the "competitive fringe" instead, it will be possible for him to obtain banking services at a tariff reflecting the marginal costs of supply. The competitive fringe may be thought to consist of potential entrants to banking, ready to respond to the demands of any switching customers. Clearly, however, equivalent results would be obtained if the competitive fringe consisted of incumbent firms, if these are able either to discriminate between old and new customers, or to give their new customers an outright switching bonus.

It is assumed that the marginal cost of producing payment services is constant, denoted by c , and that the banks have a given opportunity cost of funds $r-d$, where r is the security market interest rate and d is proportional intermediation costs, possibly due to reserve requirements. Under these assumptions, the tariff available in the competitive fringe of the market is the following:

$$P(D,N) = c \cdot N - D \cdot (r - d). \quad (5)$$

The presence of switching costs implies that the banks can offer their depositors somewhat less favourable terms than available in the fringe. For simplicity, the individual switching costs S are assumed to be a constant, common to all customers regardless of their income level.

Now, a necessary condition for the representative consumer to remain in the representative bank's clientele is that the present value of the gain from switching is not greater than the switching cost. To see what this condition entails, we must first compute the (reduced) cost attainable after switching. This is

⁴ For an exposition of the concept of contestability, see Baumol, Panzar and Willig (1988).

$$\begin{aligned} \bar{\Gamma} = & a_1 \bar{D} + a_2 \bar{N} + a_3 Y + (1/2)b_{11} \bar{D}^2 + b_{12} \bar{D}\bar{N} + (1/2)b_{22} \bar{N}^2 + b_{13} \bar{D}Y \\ & + b_{23} \bar{N}Y + (1/2)b_{33} Y^2 + c \cdot \bar{N} + d \cdot \bar{D}. \end{aligned} \quad (6)$$

Here the bar mark denotes the optimal values of the variables obtained when the marginal-cost-based "fringe" tariff prevails. The values \bar{D} and \bar{N} can be solved from the first-order conditions

$$-d = a_1 + b_{11} \bar{D} + b_{12} \bar{N} + b_{13} Y \quad (7a)$$

$$-c = a_2 + b_{12} \bar{D} + b_{22} \bar{N} + b_{23} Y. \quad (7b)$$

The possibility of switching implies that for each customer remaining in the bank's clientele, the following must hold.

$$S - \frac{[\hat{\Gamma}(Y) - \bar{\Gamma}(Y)]}{r} \geq 0. \quad (8)$$

The presence of the interest rate in the above condition makes the once-and-for-all switching cost commensurable with the liquidity costs, a flow concept. Essentially, condition (8) is about the present values of the expected utility gain from switching from one bank to another, compared with the immediate loss caused by the switching cost. In the present, static formulation, the condition obtains the very convenient form above.

Provided that condition (8) is satisfied, the competition from the fringe is not a strong enough incentive for the customer to quit. The condition serves as a restriction on the bank's pricing behaviour. We may call it the switching constraint.

It should be noted that switching cost models may involve a time inconsistency problem. The type of switching constraint developed here is built on the assumption that the potential suppliers constituting the competitive fringe are able to commit themselves credibly to the lower prices which attract the potentially switching customers.

In addition to switching costs, there is a participation constraint. This says that for a depositor to stay in the market for bank services, his utility must not be less than that obtainable by keeping $D = 0$ and $N = 0$. This constraint restricts the pricing policy of banks. As will be shown below, it may set limits to fixed charges that can be imposed

on customers regardless of their use of services, or, equivalently, for those customers with the smallest demands for services.

It is a property of second-order price discrimination situations that the participation constraint will be binding for some of the customers only. There is thus some surplus left for most of the customers. This is because the so-called incentive compatibility (self-selection) constraint restricts the shape of the tariff function to a form which is generally not compatible with extracting all consumer surplus from all customers.⁵ This will be shown below for the linear/quadratic case at hand.

4 Tariff design

We now turn to the problem of determining the optimal tariff structure. It is convenient to do this in two stages. First, the profit-maximizing quantity assignments of the products for each customer type are solved. The tariff required to induce customers to purchase just the optimal assignments is obtained in the second stage by using the customers' first-order conditions.⁶

4.1 The first-order conditions

The profit of the bank from business with a customer of type Y is given by

$$\pi(Y) = P[\hat{D}(Y), \hat{N}(Y)] + (r - d) \cdot \hat{D}(Y) - c\hat{N}(Y). \quad (9)$$

The first step in tariff design using the two-stage method is to eliminate the tariff function from the profit-maximization problem. For this purpose, the tariff faced by a representative customer located at point Y of the type distribution may be written as the minimum cost attainable less the internal cost:

⁵ In the present context, "consumer surplus" must be interpreted as the minimum of a) the difference of costs from the level which would induce the customer to switch and b) the difference of costs from the level which would induce the customer to withdraw from the market.

⁶ This is the Mirrlees method. Another, direct method is presented by Roberts (1979).

$$P[\hat{D}(Y), \hat{N}(Y)] = \hat{\Gamma}(Y) - G[\hat{D}(Y), \hat{N}(Y), Y] - r \cdot \hat{D}(Y). \quad (10)$$

This is, of course, subject to the reservation that the second-order sufficient conditions for the customer's cost minimum do hold. These conditions will be discussed below in some detail. Assuming for now that the first-order conditions do indeed characterize a customer optimum, the profit of the bank from doing business with this customer can be expressed as a function of the customer's minimum attainable cost and the customer's chosen quantities:

$$\pi(Y) = \hat{\Gamma}(Y) - G[\hat{D}(Y), \hat{N}(Y), Y] - d \cdot \hat{D}(Y) - c \hat{N}(Y). \quad (11)$$

Assume that the density function of the type parameter Y is $f(Y)$. The profit of the bank from the whole of the customer population is then

$$\Pi = \int_{Y_{\min}}^{Y_{\max}} f(Y) \pi(Y) dY. \quad (12)$$

Henceforth, we assume, for simplicity, that the income distribution of the customer population is uniform at the interval $Y \in [0, Y_{\max}]$ with a constant density $f(Y) = 1/Y_{\max}$.

The problem of maximizing the profit integral can be approached as a control problem, with Y as the independent variable, Γ as the state variable and variables D and N as controls. These are treated as functions of Y , even though the function notation is omitted below for brevity. The equation of motion for Γ which is needed for the control-theoretic approach is derived by differentiating the representative customer's total (internal plus financial) cost function (3) and applying the first-order conditions (4a, 4b) for customer optimum. This gives

$$\Gamma_Y = a_3 + b_{13}D + b_{23}N + b_{33}Y. \quad (13)$$

This constraint incorporates part of the implications of customer optimization for the bank's profit-maximization problem; in particular, it captures the differential implications of the customer's first-order

conditions (4a, 4b). Therefore, constraints analogous to (13) are often referred to in the literature as incentive-compatibility constraints.⁷

Using (13), we are able to write the Hamiltonian relevant to the bank's profit-maximization problem, subject to the customers' incentive constraints. The expression to be maximized is

$$H = \left(\frac{1}{Y_{\max}} \right) \cdot [\Gamma - a_1 D - a_2 N - a_3 Y - (\frac{1}{2})b_{11} D^2 - b_{12} DN - (\frac{1}{2})b_{22} N^2 - b_{13} DY - b_{23} NY - (\frac{1}{2})b_{33} Y^2 - d \cdot D - cN] + \lambda(a_3 + b_{13} D + b_{23} N + b_{33} Y). \quad (14)$$

This gives rise to the following maximum principle conditions for an extremum:

$$-\frac{\partial H}{\partial D} = a_1 + b_{11} D + b_{12} N + b_{13} Y + d - Y_{\max} \cdot \lambda \cdot b_{13} = 0 \quad (15a)$$

$$-\frac{\partial H}{\partial N} = a_2 + b_{12} D + b_{22} N + b_{23} Y + c - Y_{\max} \cdot \lambda \cdot b_{23} = 0 \quad (15b)$$

$$\frac{\partial H}{\partial \lambda} = a_3 + b_{13} D + b_{23} N + b_{33} Y = \Gamma_Y \quad (15c)$$

$$-\frac{\partial H}{\partial \Gamma} = -\frac{1}{Y_{\max}} = \lambda_Y \quad (15d)$$

Integrating the equation of motion for λ (15d) yields a very simple expression for the costate variable: $\lambda(Y) = C - Y/Y_{\max}$. Using this to eliminate λ from equations (15a, b), the profit-maximizing quantity assignments to different customer types can be characterized by the following equations:

⁷ This is, however, somewhat inaccurate since incentive compatibility is not necessarily guaranteed even when (9) is satisfied. One must therefore also examine the second-order (sufficient) conditions for customer optimization, as will be done later in the text.

$$a_1 + b_{11}D + b_{12}N + b_{13}Y + d = (C \cdot Y_{\max} - Y) \cdot b_{13} \quad (16a)$$

$$a_2 + b_{12}D + b_{22}N + b_{23}Y + c = (C \cdot Y_{\max} - Y) \cdot b_{23} \quad (16b)$$

From these equations, the profit-maximizing quantity assignments to each customer type can be solved:

$$D = \frac{[a_2 b_{12} - a_1 b_{22} - b_{22}d + b_{12}c + (b_{13}b_{22} - b_{23}b_{12})(C \cdot Y_{\max} - 2Y)]}{\Delta} \quad (17a)$$

$$N = \frac{[a_1 b_{12} - a_2 b_{11} + b_{12}d - b_{11}c + (b_{23}b_{11} - b_{13}b_{12})(C \cdot Y_{\max} - 2Y)]}{\Delta} \quad (17b)$$

where $\Delta = b_{11}b_{22} - b_{12}b_{21} > 0$, by the assumed convexity of the internal cost function.

Given the constant C , these assignment equations define an "optimal assignment locus" which in the present model is actually a straight line in the (D, N, Y) space. This result can be interpreted to mean that, although the customers can choose any nonnegative values for D and N , the profit-maximizing tariff induces them to select points along the straight line defined by (17a, b).

Now, the consumer's first-order optimum conditions can be utilized to characterize the profit-maximizing tariff, which leads to the above-derived assignments being voluntarily chosen by the intended consumer types. Substituting (4a, b) into (16a, b) gives for the marginal prices:

$$P_D(N, D) + r - d = (Y - C \cdot Y_{\max}) \cdot b_{13} \quad (18a)$$

$$P_N(N, D) - c = (Y - C \cdot Y_{\max}) \cdot b_{23} \quad (18b)$$

The tariffs cannot be satisfactorily characterized, however, until the value of the constant C (which emerged from integrating equation (15d)) is known. The determination of C requires inspection of the participation and switching constraints. This will be done in the following subsections.

4.2 The participation constraint

Let us consider the participation constraint first. Participation means purchasing the assigned quantities as opposed to $D = 0$ and $N = 0$. Define the net benefits from participation to a customer at Y as $B(Y)$, with

$$B(Y) = a_3 Y + (1/2)b_{33} Y^2 - \Gamma(Y). \quad (19)$$

This benefit is calculated as the hypothetical internal cost with $D = 0$ and $N = 0$ (from (3)) less the minimum attainable cost $\Gamma(Y)$ with the assigned demands D and N .

Now, the participation constraint states that $B(Y)$ must be non-negative for all Y (in the range of customers). It is helpful to note that, according to equation (13) above, $B_Y(Y) = -(b_{13}D + b_{23}N)$. We will restrict the analysis below to such parameter values that this expression is positive, implying that the net benefit from participation must be increasing in the income level.⁸ This being the case, and if the participation constraint is binding for some customers, it follows that it can be binding only for those customers who are located at the lowest point of the income distribution. Since it would be irrational for the bank to leave the participation constraint unbinding,⁹ $B(Y_{\min}) = 0$ must hold in equilibrium. We may conclude that the participation constraint actually implies the following terminal-point condition for our control problem:

$$\Gamma(Y_{\min}) = a_3 Y_{\min} + (1/2)b_{33}(Y_{\min})^2. \quad (20)$$

Under the simplifying assumption made above that $Y_{\min} = 0$, the terminal condition (20) reduces to

$$\Gamma(0) = 0. \quad (20a)$$

⁸ The result that the participation constraint is binding only for the "smallest" customers (smallest in terms of the size of their purchases) is a standard feature of price discrimination models (Phlips (1983), p. 169).

⁹ This would mean a lost profit opportunity, since if the participation constraint is nowhere binding, there is consumer surplus which can be extracted without loss of clientele.

4.3 The switching constraint

Turning now to the switching constraint, the net benefit for the customer from not switching can be written in flow form as

$$Z(Y) = \bar{\Gamma}(Y) + r \cdot S - \Gamma(Y). \quad (21)$$

This is an application of condition (8) above. We know that in equilibrium it must be nonnegative for all customers. The relationship between income and the benefit of not switching can be analyzed by differentiating (21) with respect to Y , which gives

$$Z_Y(Y) = [b_{13}(\bar{D} - D) + b_{23}(\bar{N} - N)]. \quad (22)$$

Using equations (7a) and (7b) for \bar{D} and \bar{N} and the assignment equations (17a) and (17b) for D and N we obtain, after some manipulation,

$$Z_Y(Y) = \frac{[(b_{13}^2 b_{22} - 2b_{23} b_{12} b_{13} + b_{11} b_{23}^2)(Y - C \cdot Y_{\max})]}{\Delta}. \quad (23)$$

The derivative $Z_Y(Y)$ is clearly negative for all $Y < C \cdot Y_{\max}$, as long as the parameter values are as assumed. This is demonstrated in the section below, which is dedicated to the incentive compatibility question. Thus, the benefit of not switching is decreasing in Y , if $Y < C \cdot Y_{\max}$. At $Y = C \cdot Y_{\max}$, the benefit from not switching reaches a minimum and grows thereafter (in the region where $Y > C \cdot Y_{\max}$).

The region where $Z_Y(Y)$ is positive exists only if the tariff is such that $C < 1$, since $Y \leq Y_{\max}$. If $C \geq 1$, then the benefit from not switching is monotonous decreasing in Y . However, it is possible to rule out the case $C > 1$ by the following argument.

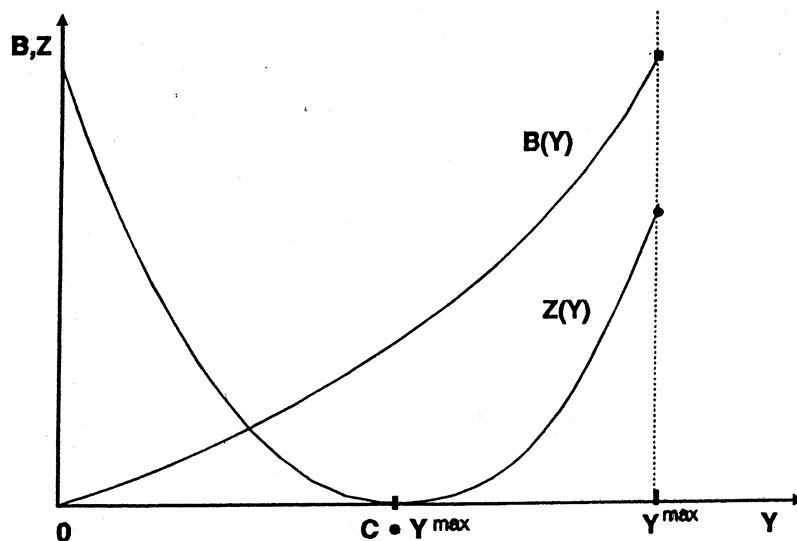
Let us focus on the marginal prices of the largest quantities sold, ie on $P_D(D(Y_{\max}), N(Y_{\max}))$ and $P_N(D(Y_{\max}), N(Y_{\max}))$. Call them marginal prices at the top. Now, if $C > 1$, then at least one of the marginal prices at the top is greater than the corresponding marginal cost ($d-r$ or c , respectively). This is clear from equations (17a, b) and the limits for the parameter values b_{13} and b_{23} obtained from the requirement that the problem is well behaved (see next section). Now, the selling of a product at a marginal price at the top higher than the cost of producing it is clearly suboptimal, for there is always the

possibility of selling an additional unit of the product in question at a price which adds to profits, without disturbing the equilibrium in the market for the units sold previously. This means that such a situation cannot represent a true optimum.¹⁰

Now, the limiting case where $C = 1$ gives the tariff which is exactly the same as that of a monopolist; therefore, that case must arise when the switching constraint does not restrict the pricing conduct of the representative firm in any way.

The conclusion which follows is that when the switching constraint is binding, the constant $C < 1$. This implies that then the switching constraint is binding at an interior point of the interval $(0, Y_{\max})$, and there is a region in which the benefit from not switching is increasing in Y . The behaviour of the consumer surplus $B(Y)$ and the net benefit from not switching $Z(Y)$ are depicted in Figure 1 for a well-behaved case. Consumer surplus is increasing in Y , while the benefit from not switching is a parabola with a minimum at the point $C \cdot Y_{\max}$.

Figure 1. Participation and switching constraints



¹⁰ The argument is also cited in Wilson (1993).

5 Incentive compatibility

Any suggested scheme for second-degree price discrimination must be checked for incentive compatibility. Incentive compatibility means that, for a nonlinear pricing scheme to operate as intended, it must give rise to a separating equilibrium in the sense that customers with different characteristics do indeed choose different bundles of goods with different prices. Obviously, first-order necessary conditions for incentive compatibility are guaranteed by construction when the design of the tariff is based on the family of customers' first-order conditions. There is, however, the further question of whether the points defined by these first-order conditions really represent true global optimum (cost minimum) for each customer type. Thus, to check incentive compatibility, sufficient conditions for the customer optimum need to be examined. It turns out that these are essentially similar to those developed in the previous chapter for the monopoly model.

Since (in the present model) the optimum tariff is unambiguously defined only along the linear assignment locus, it is important to examine whether the (global) sufficient conditions hold along that locus. Now, a customer's chosen point along that locus may be interpreted as defining his revealed type (distinct from his actual type, which is exogenous for each customer). If the optimal revealed type is equal to the actual type, the pricing scheme is incentive-compatible and the resulting equilibrium is a separating equilibrium. Expressed in this way, the question of whether the tariff is incentive-compatible can be rephrased as whether the marginal price schedules induce customers to reveal their actual type through their demand behaviour.

Let us revert for a while to the general nonlinear formulation of the representative customer's problem. The cost-minimization problem in (1) can be rewritten in the following way as a problem of choosing the revealed Y , denoted as \hat{Y} :

$$\min \Gamma(\hat{Y}, Y) \text{ with respect to } \hat{Y},$$

$$\text{where } \Gamma(\hat{Y}, Y) = G(D(\hat{Y}), N(\hat{Y}), Y) + P(D(\hat{Y}), N(\hat{Y})).$$

The first-order condition for a cost minimum is

$$\Gamma_{\hat{Y}}(\hat{Y}, Y) = 0, \quad (24)$$

which holds when $\hat{Y} = Y$. This is actually just a very compact way of writing the solution to the customer's problem, observing, however, the assignment equations as restrictions which define the optimal \hat{D} and \hat{N} as functions of only one variable, the revealed type (\hat{Y}).

The second-order optimality condition requires $\Gamma_{\hat{Y}\hat{Y}}(Y, Y) \geq 0$; with a quadratic cost function, this (local) second-order condition also ensures global convexity. The evaluation of the second-order condition can be simplified using a "trick" described in Tirole (1988, p. 156 n.). Setting $\hat{Y} = Y$ and differentiating with respect to Y yields

$$[\Gamma_{\hat{Y}\hat{Y}}(Y, Y) + \Gamma_{\hat{Y}Y}(Y, Y)]dY = 0, \quad (25)$$

from which

$$\Gamma_{\hat{Y}\hat{Y}}(Y, Y) = -\Gamma_{\hat{Y}Y}(Y, Y). \quad (26)$$

It is seen that the second-order condition for a strict cost minimum can be transformed to the requirement that $\Gamma_{\hat{Y}Y}(Y, Y)$ is negative. In the present quadratic case, we can utilize (17a) and (17b) and write this condition in the following form:

$$\begin{aligned} \Gamma_{\hat{Y}Y}(Y, Y) &= b_{13} \cdot \frac{\partial D}{\partial Y} + b_{23} \cdot \frac{\partial N}{\partial Y} \\ &= - \left[\frac{2b_{13} \cdot (b_{13}b_{22} - b_{23}b_{12})}{\Delta} + \frac{2b_{23} \cdot (b_{23}b_{11} - b_{13}b_{12})}{\Delta} \right] < 0. \end{aligned} \quad (27)$$

This condition can be given an economic interpretation by referring to the analysis of the participation constraint in section 4.2. It is easy to check that in fact $\Gamma_{\hat{Y}Y}(Y, Y) = -B_{YY}(Y)$ and our second-order condition states that the net benefits from participation are a convex function of Y along the assignment locus.

Taking into account that $\Delta > 0$, condition (27) simplifies to

$$b_{13}^2 b_{22} - 2b_{12} b_{13} b_{23} + b_{11} b_{23}^2 > 0. \quad (28)$$

This is ensured by the assumptions we made on the shape of the $G(D,N,Y)$ function.

An important question is whether tariffs involving marginal subsidies (which are suggested by the first-order conditions if b_{13} or b_{23} is positive) can be incentive-compatible. In the present context, this amounts to asking when (28) is consistent with either b_{13} or b_{23} being positive. It turns out that this depends on the other parameters of the problem, and actually it is useful to interpret the incentive compatibility condition (28) as a condition for the cost complementarity parameter b_{12} . This can be broken down into three cases, depending on b_{13} and b_{23} :

$$A) \quad b_{12} < \frac{(b_{13}^2 b_{22} + b_{11} b_{23}^2)}{(2b_{13} b_{23})} \quad \text{if } b_{13} b_{23} > 0 \quad (29a)$$

$$B) \quad b_{12} > \frac{(b_{13}^2 b_{22} + b_{11} b_{23}^2)}{(2b_{13} b_{23})} \quad \text{if } b_{13} b_{23} < 0 \quad (29b)$$

$$C) \quad b_{13} b_{23} = 0. \quad (29c)$$

In case A, both b_{13} and b_{23} are negative by the assumption that the benefit from participation is increasing in Y , and the participation constraint is binding at $Y = 0$. This is the symmetric case which is exclusively treated in the previous literature on nonlinear multiproduct pricing. Case B is the most interesting one, for it has asymmetric valuation effects (b_{13} and b_{23} have opposite signs) which generate cross subsidies, as is seen from the tariff function. The inequality in B) provides a condition for the parameter b_{12} (the complementarity of bank services) in which the cross-subsidizing tariff is compatible with the individual incentives.¹¹ Case C) is uninteresting.

¹¹ Normality with respect to type, if required, imposes a further condition on b_{12} (see Section 4 of the previous chapter).

6 Some properties of the tariff

Summarizing the analysis presented above, the following observations on the shape of the optimal tariff can be made. The remarks are made under the assumption that the sufficient conditions hold so that there is a separating equilibrium and that the assignments of D and N are increasing in Y (normality with respect to type).

First, the marginal prices deviate from marginal costs almost everywhere in the range $0 \leq Y \leq Y_{\max}$. The equality of marginal prices with marginal costs occurs only at one point on the assignment locus, ie at the point where customers with $Y = C \cdot Y_{\max}$ are located (see equations 18a and 18b and note that $C < 1$ if the switching constraint is binding). For those customers, the marginal interest rate on bank deposits equals the marginal return on funds for the bank ($r - d$), and the marginal price of transaction services equals c , the marginal cost of production.

For customers with income less than $C \cdot Y_{\max}$, the marginal prices are above cost (meaning that the marginal deposit rate is low), if both b_{13} and b_{23} are negative. This is the symmetric case. Marginal subsidies for one product occur if one of the parameters is positive, however; for example, if b_{23} is positive, transaction services are priced below marginal cost. For customers with relatively large incomes ($Y > C \cdot Y_{\max}$), the deviations of marginal prices from marginal costs are the opposite from those charged to low-income customers.

The direction of the marginal price distortions emerging in the model can be presented in the form of the following table.

Table 1. **Marginal price distortions in the switching cost model. m_i = the marginal price-cost spread for product i .**

Valuation effects	$Y < C \cdot Y_{\max}$	$Y > C \cdot Y_{\max}$
$b_{13}, b_{23} < 0$	$m_1 > 0$ $m_2 > 0$	$m_1 < 0$ $m_2 < 0$
$b_{13} < 0, b_{23} > 0$	$m_1 > 0$ $m_2 < 0$	$m_1 < 0$ $m_2 > 0$
$b_{23} > 0, b_{13} < 0$	$m_1 < 0$ $m_2 > 0$	$m_1 > 0$ $m_2 < 0$

Although equations (18a, b) characterize the optimal tariff in a very revealing way, these equations alone do not suffice to determine the explicit form of the optimal tariff. This is because, in (18a, b), the marginal prices are presented as functions of the customer type Y , whereas in reality, the tariff must be posted as conditional only on the purchased quantities D and N . This is the restriction inherent in second-degree price discrimination (and is due to the unobservability of types). The tariff can be cast into the required form by eliminating Y using the quantity assignment equations (17a, b). This yields the following equations for marginal prices:

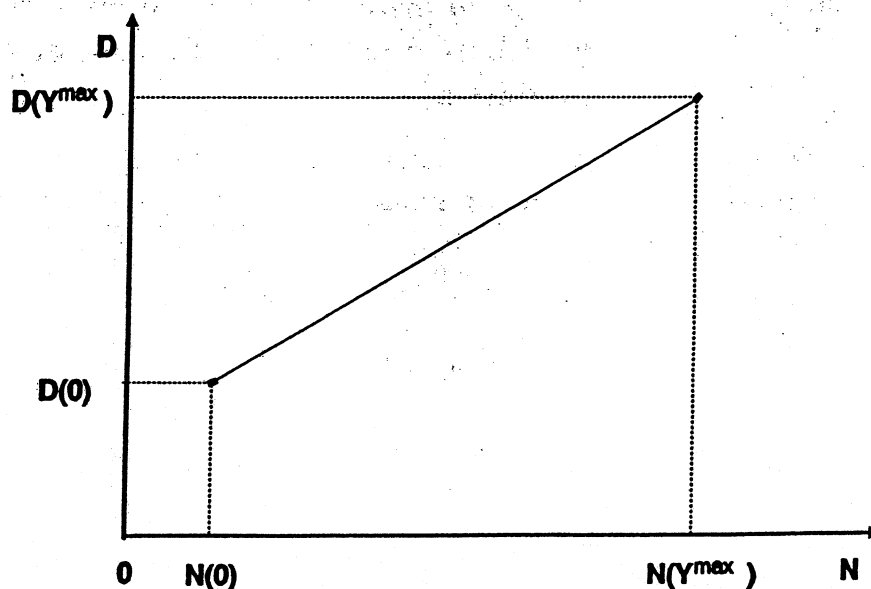
$$P_D(N, D) = -r + (1/2)(d - a_1 - b_{11}D - b_{12}N - b_{13}C \cdot Y_{\max}) \quad (30a)$$

$$P_N(N, D) = (1/2)(c - a_2 - b_{12}D - b_{22}N - b_{23}C \cdot Y_{\max}). \quad (30b)$$

It should be noted that Y can be eliminated from (18a, b) in several ways using either (17a) or (17b) or both, each method resulting in somewhat different marginal price equations. The different solutions are, however, necessarily equal for all values of D and N included in the optimal assignment locus. Outside the locus, there is ambiguity.

The quantity assignments and marginal prices predicted by the model behave as presented in the following figures. These figures are drawn for the case where $b_{23} > 0$ so that payment services are subsidized at the margin for low-income customers.

Figure 2. The assignment locus



The assignment locus is linear with positive assignments for $Y=0$. The marginal price schedules intersect the relevant marginal "costs" $r-d$ and c , respectively at the points $Y=C \cdot Y_{\max}$ where the benefit from net switching is at minimum.

Figure 3. The marginal deposit rate along the assignment locus

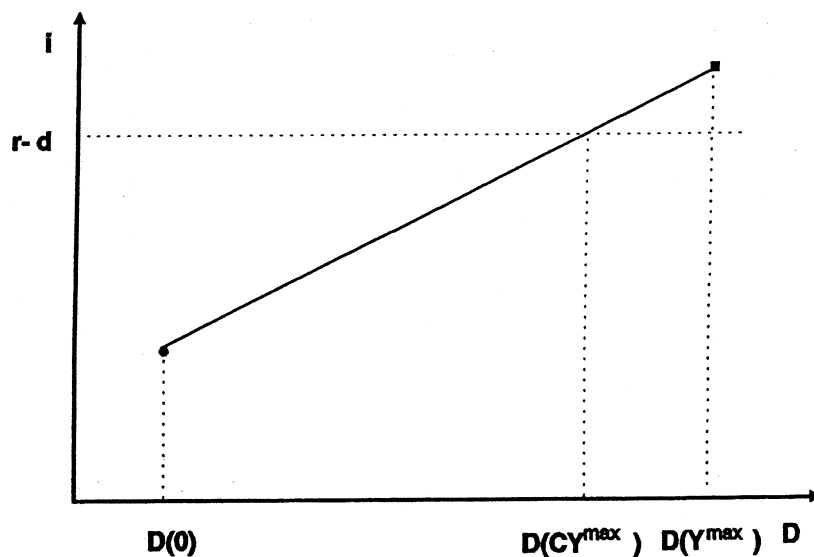
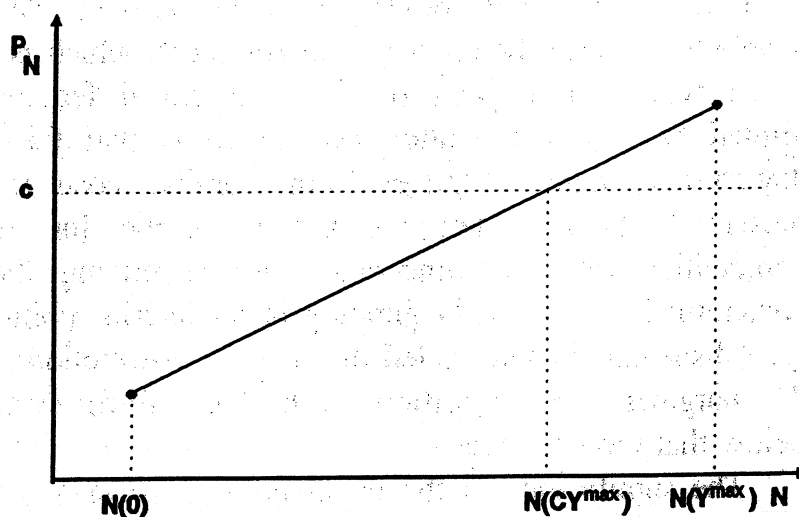


Figure 4. The marginal service charge along the assignment locus



Generally, we would like to characterize the optimal tariff in the entire nonnegative quadrant of the D, N space. It is, however, a general property of single-type parameter pricing problems that the first-order conditions define the optimal tariff only along an optimal assignment locus such as the line $D(Y), N(Y)$ defined by equations (17a, b) above. For other combinations of D, N the marginal prices do not really matter as long as the tariff is expensive enough so that the incentives to stay on the assignment locus do dominate. The restrictions implied by profit-maximization for the optimal tariff outside the assignment locus can therefore be expressed only in the form of inequalities, and belong to the realm of "global sufficient conditions".

7 Conclusions

In the literature, nonlinear pricing problems are usually formulated for a discriminating monopolist. This paper has modified the standard analysis by adding some competition, in the form of a competitive fringe with a switching cost. The analysis is performed in a two-product context, where the two products are payment services and deposit taking services produced by banks. For simplicity, the analysis is carried out for customers with quadratic preferences with a uniform "income" distribution.

Like the monopoly model in the previous chapter the model here is able to produce the cross-subsidizing tariff which is often considered to be typical of deposit pricing. The novel feature caused by the limited type of competition considered is that the direction of the deviation of marginal prices from marginal costs (as defined in the present banking context) is not the same for all customers. If competition with switching costs restricts pricing, then the switching constraint is shown to be binding at an interior point of the customer type distribution (here called the income distribution). The direction of the marginal price "distortion" is the opposite for customers above and below that income level.

The significance of the result is negative in a certain respect: one of the most important results from the analysis of optimal nonlinear pricing for monopolies is that the biggest purchases of products are sold at marginal prices which are equal to marginal cost. In the switching cost model, this precise result is lost. It is replaced by the interesting, but less easily testable, outcome that there are marginal cross subsidies between the largest and smallest purchases: if the

smallest purchases are sold at prices above marginal cost, then the largest purchases are subsidized at the margin. If the tariff happens to be cross-subsidizing between products, the direction of the cross subsidy is reversed when moving from the smallest to the largest purchases.

In the banking context, these results can be illuminated by the following example. If the deposit pricing system has the property that the deposit rate is increasing in the deposit balance and the service charge is increasing in the usage of payment services, then the tariff is cross-subsidizing. For the smallest customers, deposit taking is "expensive" and payment services are "cheap" relative to the marginal costs of the bank. If switching costs of the type analyzed in this paper constitute a binding constraint for pricing, then the largest deposits (on transactions accounts) should be paid marginal rates which exceed the bank's opportunity cost of funds, and the transactions should be priced above marginal cost for the biggest users.

References

- Baumol, W. – Panzar, J. and Willig, R. (1988) **Contestable Markets and the Theory of Industry Structure**. Harcourt Brace Jovanovich, San Diego.
- Feenstra, R. (1986) **Functional Equivalence between Liquidity Costs and the Utility of Money**. *Journal of Monetary Economics* 17, No. 2, 271–291.
- Klemperer, P. (1987a) **Markets with Consumer Switching Costs**. *Quarterly Journal of Economics* 102, No. 2, 375–394.
- Klemperer, P. (1987b) **The Competitiveness of Markets with Switching Costs**. *Rand Journal of Economics* 18, No. 1, 138–150.
- Klemperer, P. (1992) **Competition with Switching Costs**. CEPR Discussion Papers, London.
- Phlips, L. (1983) **The Economics of Price Discrimination**. Cambridge University Press, Cambridge (U.K.).
- Roberts, K. (1979) **Welfare Considerations of Nonlinear Pricing**. *Economic Journal* 89, No. 1, 66–83.
- Vittas, D. – Frazer, P. and Metaxas-Vittas, T. (1988) **The Retail Banking Revolution. An International Perspective**. Lafferty Publications, London.
- von Weizsäcker, C. (1984) **The Costs of Substitution**. *Econometrica* 52, No. 5, 1085–1116.
- Wilson, R. (1993) **Nonlinear Pricing**. Oxford University Press, Oxford.

Publications of the Bank of Finland

Series E (ISSN 1238-1691)

(Series E replaces the Bank of Finland's research publications series B, C and D.)

- E:1** Jukka Vesala **Testing for Competition in Banking: Behavioral Evidence from Finland.** 1995. 206 p. ISBN 951-686-447-3.
- E:2** Juha Tarkka **Approaches to Deposit Pricing: A Study in the Determination of Deposit Interest and Bank Service Charges.** 1995. 166 p. ISBN 951-686-457-0.