Introduction

Competition in banking markets is a major concern these days. Deregulation has fundamentally changed the way banks compete and it has also enabled non-banks to penetrate into markets that were traditionally served by banks. Banks have responded to their eroding market dominance and the liberalised financial scene by entering into mergers and other collaborations, both among themselves and with non-bank financial firms.

There are a number of reasons why competition in banking deserves particular attention. First, the market power of banks affects their profitability and, on a larger scale, financial sector soundness and stability. Second, market imperfections cause allocative inefficiency, which impairs the welfare society derives from banking services. This is of special significance in banking markets, since banks are heavily leveraged intermediaries that have a large part of the nation’s assets on their balance sheets, so that any inefficiency might be expected to have considerable welfare effects. Third, the competitive structure of financial markets has major implications for the effectiveness of certain aspects of monetary policy like the transmission of money market rates into loan and deposit rates.

This thesis thus aims at assessing the degree of competition and cost efficiency among the five dominating Swedish banks, representing some 85% of the total industry assets, during the period 1989-1997. In so doing, it also aims to develop a general oligopoly model with the specific purpose of providing a tool for an improved methodology for the evaluation of market power in oligopolistic markets, and to demonstrate the empirical application of this methodology.

Essay 1

Prices are usually seen as the major instrument of transmitting information about market conditions or product quality. However, it is easy to find examples of markets where quantities, in the form of relative firm sizes or market shares, are of major importance in influencing consumers’ choice. Some examples mentioned by Caminal and Vives (1996) are tourists choosing restaurants (crowded restaurants are preferred to empty ones), and readers choosing books by preferring titles on a bestseller’s list displayed in the shop. Other examples
are in Becker (1991), who relates to sociological reasons to justify the fact that the value of a good for an individual consumer is a function of the number of consumers who choose the same good. Also retail bank markets are clearly subject to the same kind of consumer behaviour, where the mere size of the bank signals confidence to the customers and seems to influence their very preferences.

On the basis of such observations, that consumers rationally believe that a firm with a high market share is likely to produce a high quality good, Caminal and Vives (1996) formulate a model where demand is influenced by the market shares of firms. I take a similar approach to these authors in my first essay, An Asymmetric Oligopoly Model and a Method for Its Empirical Application (a previous version has been published in Oxenstierna (1998)). A general oligopoly model containing parameters reflecting two different types of product differentiation is developed; one parameter vector captures consumer sensitivity to relative price differentials between firms (firm-specific factors $\gamma_i$), and another parameter vector captures consumer sensitivity to relative size differences between firms (firm-specific factors $s_i$). The model furthermore allows firms to be asymmetric in terms of cost levels (firm-specific factors $c_i$). It is solved for four different game-theoretic equilibria, under assumptions of various market regimes. Based on a proposed methodology for empirical applications, the model is suitable for empirical research where the data set contains firm-level data on prices, quantities and costs.

The model was developed with empirical applications in mind. It aims at providing a methodological tool in the empirical characterisation of competition in oligopoly markets. In doing this, it proposes a rather simple method of empirical application, while retaining the full game-theoretic explanatory power. In practice, this means that observed firm behaviour in terms of prices and quantities can be measured to a scale of four different game-theoretic optimal price-quantity levels (the marginal cost pricing, the Nash equilibrium price in competition with respect to prices, the Nash equilibrium price in competition with respect to quantities, and the monopoly price regime). This scale of optimal prices under different market conduct regimes is deduced in the theoretical model, using estimated parameters (firm-specific factors $s_i$, $\gamma_i$ and $c_i$). Thus, there is no estimation of a specific conduct parameter as in, e.g., conjectural variations models. In applied work, this model then makes it possible to gauge empirically observed prices to a range of game-theoretic equilibrium prices that are analytically
deducted, as illustrated in the third essay. It also allows for precise calculations of welfare effects from non-competitive pricing.

**Essay 2**

The current structural changes in the European and Swedish banking industries raise questions concerning the competitive viability of individual banks. Some mergers and acquisitions seem to be driven by either explicit or implicit arguments referring to the cost structure of banking. One example can be the recent rush to merge banks in Sweden, where one often heard argument is that a larger scale of the banking operations should improve cost efficiency.

Swedish banking has been operating with an increased spread between deposit and loan interest rates, see Figure 1 in Essay 2 and Figure 1 in Essay 3. One relevant question to be asked is whether this increased spread can be explained by increased costs or if it has some other explanation. One important factor influencing both cost structures and revenue structures, is the de-regulation of the banking industry that took place in 1985, and which has been followed by a sequence of re-regulations. In order to investigate the effects of de- and re-regulations it is necessary to know the cost structure of the industry.

In the second essay, *Do Swedish Banks Enjoy Economies of Scale or Economies of Scope?*, the extent of production economies on the cost side in the five big Swedish banks are investigated. This essay complements the empirical study of Swedish banking by, *inter alia*, estimating a cost function that includes not only operating costs, but also opportunity costs of equity capital, where the opportunity costs are calculated with the CAPM model. Opportunity costs for equity capital are important to include in the cost function not only because they are an integral part of the economic theory on optimal resource allocation within firms. They are important also because of their inherent feature of reflecting varying risks between banks, and over time. Specifically, they take away a possible source of bias as far as asset portfolios, mainly loans, are concerned, since heterogeneities in loan quality between banks will be reflected in the stock market valuations of bank shares. Banks can easily lower their production costs by more risk-taking behaviour, such as quickly expanding their loan portfolio without increasing monitoring, etc. What appears to be improved production cost efficiency should then be offset by an increase in risk that increases economic costs. These considerations have not only
theoretical relevance, but are moreover of specific relevance to this study, which also covers years during which a profound credit-risk related banking crisis took place.

In order to properly evaluate economies of scope, the study employs the composite functional form for the cost function proposed by Pulley and Braunstein (1992). Results imply small dis-economies of scope between deposits and loans for the banks, as well as slightly negative economies of scale. The result on the lack of economies of scope is essential in formulating the banking model in the third essay, since it means that the analysis of the two aggregated retail banking markets (deposits and loans) can be separated to a large degree.

Essay 3

The traditional and dominating methodology in studies of competition in banking has been conjectural variations models. These models have some theoretical as well as methodological drawbacks which are discussed in both the first and the third essays. Recent methodological developments in the empirical study of oligopoly markets address the issue of product differentiation in detail. Berry (1994), develops a static oligopoly model where there is a range of observable product characteristics, as well as a range of unobservables.

In the study of service markets, like banking, and most product markets, rich data sets on product characteristics and on demand data (prices and quantities) related to these are rarely available. One of the few observable characteristics that distinguishes Swedish banks in terms of product differentiation is their relative sizes. These are highly correlated to the size of the branch networks, since all the five banks have nationwide coverage. They also offer very similar product sets to consumers. Furthermore, payments services that are offered to customers, such as giro systems and automatic teller machines, are not a distinguishing factor between Swedish banks, since they are provided by jointly owned organisations. For these reasons, the models and methods developed by Berry and others are not very appropriate in a banking context, where the typical data set contains only prices, quantities and cost variables.

In the third essay, Testing for Market Power in the Swedish Banking Oligopoly, the model in the first essay is developed into a banking context, allowing for the inherent multi-product (deposits and loans) characteristic in that industry. The method of empirical application is not worked out in full detail in the first essay, only the steps involved
are proposed. In the third essay, these steps are followed in principle, yielding a full demonstration of the methodology involved in a time-series setting. From an econometrical point of view, the two aspects of product differentiation are treated differently. The relative firm sizes are treated as exogenous to demand in the short term, whereas the relative consumer sensitivities to price differentials are endogenously determined by estimating own-price demand equations. Straightforward instrumental variables techniques are employed. All variables are tested for stationarity, and co-integrating vectors are established. From these the long-run market demand equation parameters are determined, which are then used to determine the demand structure. The cost functions of the five banks are estimated separately. All parameters are fed into the optimal price equations given in the theoretical model. The empirical conduct is then evaluated for each quarter and for each bank in both of the two markets.

Results show that there is significant market power mainly in the deposit market, although with a strongly time-varying pattern. The loan market is found to be more competitive, but also displays strong time variations. In order to further analyse the short-term volatility in conduct that is found, an error correction mechanism is deducted and estimated. This reveals different short-term behaviour in the stocks of deposits and loans, respectively, a result which sheds some light on the monetary transmission mechanisms in the Swedish banking system. Welfare losses to the society from non-competitive pricing of loans and deposits are calculated to be ca 1.1% of GDP as a yearly average during the sampling period. There is a persistent trend from the end of the banking crisis in 1993 until mid-1997 towards a more competitive conduct. Furthermore, the combined results of the empirical analyses in the second and third essays, indicate that there are no cost-side arguments for the many mergers and acquisitions that have taken place in the industry proper, and that small, specialised banking institutions might be competitively viable.

At the bottom line, the worth of different methods in analyzing oligopoly markets must be established in empirical applications. It is hoped that the contribution made in the first essay might advance such methodologies and that the application to Swedish bank markets in the third essay will demonstrate the viability of the proposed new methodology. If so, a further development of the oligopoly model in the first essay into a genuine multi-product one might be worth wile.
Finally, it should be noted that the analysis of competition and cost efficiency in Essay 2 and Essay 3 has been restricted to only on-balance sheet products. Off-balance sheet products are excluded from the outputs since there is a profound lack of data on them; they are only reported as a net revenue item in annual and quarterly reports. On the other hand, they comprise a substantial and increasingly important activity in Swedish banks, with revenues comparable in size with the net interest revenue from on-balance sheet products in later years.

A Multi-product Banking Model

Much of the analysis in the second and third essays is built on a specific banking model, that recognises the multi-product character of banks’ activities. The model specification is related to a large debate among bank researchers on how to define banking inputs and outputs, see, e.g., the papers and discussions in Griliches (1992). The main issue of this debate is the nature of deposits, which have sometimes been defined only as an input (the “asset approach”), sometimes only as an output (the “production” approach, see Berger and Humphrey (1992) from which these classifications are taken), sometimes as either input or output (the “user cost” approach, see Fixler and Zieschang (1992), Hancock (1985 and 1992)), and sometimes as both (the “value-added” approach, see Arndt (1984), Wykoff (1992), Triplett (1992), Vesala (1995, Ch. 4)).

Banks aim to establish a profit maximising spread between loan and deposit interest rates. From a simplified loans-supply-side point of view, deposits are then an input in the production of revenue generating securities, such as loans. More generally, however, a bank is a service company, selling its financial products in a market. From a deposits-demand-side point of view, the bank provides a multitude of services bundled with the deposit service, such as payments services, liquidity services, insurance of the value of the customer’s assets, etc, for which the customer is ready to pay, either in the form of explicit fees or in the form of implicit interest foregone on the deposit holdings. In this sense, deposits are simultaneously both an input and an output in the bank’s activities. To illustrate the point: How profitable

---

1 This has often been called the “intermediation approach” and has been implemented by, e.g., Sealey and Lindley (1977); Berger, Hanweck, and Humphrey (1987); Noulas, Ray, and Miller (1990); and Schaffer (1993).
to the bank is a customer that is only depositing money? Or: How profitable is a branch in the bank, which has large deposits but little loans? From an accounting point of view, the deposits would be regarded as a cost. But clearly, banks would not accept a customer that generates only costs and no revenues. The deposited money has a value to the bank and is part of the funding placed in revenue-generating securities. Furthermore, all banks are assumed to expose depositors to equal risks. If that was not the case, differentials in deposit rates could reflect perceived risk differentials. This assumption is natural in a system where there is deposit insurance\(^2\) and bank failures are practically non-existent, with the national bank serving as a lender of last resort.

The gross margin a bank earns has to be allocated in parts, each representing the opportunity cost of the asset or liability in question. One way of doing this is to peg the bank as a financial intermediary around an (internal) transfer interest rate, normally equal to the short-term money market interest rate or the interbank interest rate. This market rate then represents a market valuation of the opportunity cost of liquidity operating in the bank. The banking model employed in the second and third essays can be illustrated by defining a bank objective function built on the argumentation above. First, assume that each bank \(i\) operates subject to the following simplified balance sheet restriction:

\[
L_i + B_i + R_i + F_i = D_i + K_i,
\]

where \(L_i\) is the stock of all kinds of loans; \(B_i\) is securities, e.g. T-bills and bonds; \(R_i\) is the value of compulsory liquidity reserves (cash + central bank account); \(F_i\) is the book value of fixed assets; \(D_i\) is the stock of all kinds of deposit liabilities; and \(K_i\) is the book value of equity capital.

Banks are regulated in several ways. They might be required to hold reserves for their deposit liabilities with \(\bar{\delta}\) as the legally required reserve ratio, and \(\delta_i\) as the actual rate held, \(0 < \bar{\delta} \leq \delta_i < 1\). Reserves are kept as assets in the form of cash, or in a zero-interest account in the

\(^2\) In Sweden there was a system of implicit deposit insurance as evidenced during the banking crisis in the beginning of the 1990’s. This implicit guarantee has recently been replaced by an explicit deposit insurance system.
central bank. Thus the real interest cost for deposits is higher than the deposit interest rate paid to customers: $(\tilde{p}_i + \delta_i p_M)D_i$, where:

\[ p_M \] is the (internal) transfer price of funds in the bank, normally equal to the exogenous short-term money market interest rate.

\[ \tilde{p}_i \] is the average deposit interest rate paid by bank \( i \)

Thus, \( \delta_i p_M D_i \) is not a direct cost, but merely reflects the opportunity cost of revenues forsaken. Banks are furthermore required to keep a certain minimum ratio between assets and equity capital, the capital adequacy ratio, with \( \bar{\rho} \) as the legally required minimum rate, and \( \rho_i \) as the actual rate held, \( 0 < \bar{\rho}_i \leq \rho_i < 1 \). The amount of assets included in the calculation of the capital adequacy ratio is evaluated according to risk weights assigned to the different classes of assets, with e.g. loans and fixed assets carrying a 100% weight and government securities 0%. The balance sheet restriction can therefore be written as:

\[
L_i + B_i + \delta D_i + F_i = D_i + \rho_i (L_i + F_i), \text{ or } (1 - \rho) (L_i + F_i) - (1 - \delta)D_i + B_i = 0.
\]

The level of equity is thus determined as \( \rho_i (L_i + F_i) \). Disregarding equity costs on fixed assets, the objective function for bank \( i \) can then be written as:

\[ \max \rho_i (L_i + F_i) \]

---

3 Risk weights are agreed internationally through the Bank of International Settlements and implemented by central banks or financial supervisory authorities domestically. For simplicity I here assume that there are only two weight classes.

4 To illustrate the argument, the bank is assumed to have a given ratio between own capital and total assets. Any net increase in deposits that is intermediated into assets in the form of loans must therefore be compensated for by a proportional increase in the amount of capital. This in turn means that the “real” interest obtained on loans is lower than the nominal interest rate. E.g., if the opportunity cost of capital equals the risk-free money market rate + a risk premium for bank stocks, \( p_M + r = 10\% + 2\% \) and the capital adequacy ratio is 8%, the real marginal interest on a new loan yielding 15% will be \( 0.15 - 0.12\times0.08 = 14.04\% \). Another way to express the same thing is to say that the cost of funds is higher than the deposit interest rate paid, in this case \( \tilde{p}_i + 0.96\% \). This example also serves to illustrate the arbitrariness of designating equity costs only to the bank lending, which seems to be the common practice of the banking industry, see Matten (1996, ch 3).

5 This formulation of the objective function has some similarity with the well-known model by Klein (1971). However, Klein’s model doesn’t comprise an op-
\[
\max \pi_i = \left[ p_i - p_M - r_i (p_M + r_i) - c_i \right] L_i + \left[ p_M (1 - \delta_i) - \tilde{p}_i - \tilde{c}_i \right] D_i + (p_B - p_M - c_B) B_i - C_{F_i}
\]
s.t. \( \rho_i \geq \tilde{\rho}, \delta_i \geq \tilde{\delta} \) \quad (1)

where:
- \( p_i \) is the average lending interest rate for loans given by bank \( i \)
- \( r_i \) is the risk premium for investors in bank equity, \( i \)
- \( p_B \) is the yield on the securities portfolio
- \( c_i \) is the operational average cost for providing loans
- \( \tilde{c}_i \) is the operational average cost for providing deposit services
- \( c_B \) is the operational average cost for managing the bond portfolio
- \( C_F \) is the total opportunity cost for having fixed assets.

The formulation of the payoff function (1) requires some qualification, mainly regarding the treatment of deposits. It is a well-known fact that large banks manage their balance sheets financially with an (internal) transfer price of funds \( p_M \). \(^6\) Contemporary Asset and Liability Management (ALM) systems comprise management of the bank’s balance sheet as a portfolio of assets and liabilities. The transfer price of funds \( p_M \) is then decided by the bank itself to serve as a tool for allocating costs and revenues, thus facilitating profitability analysis of products and customers as well as individual branches, and a distribution of costs and revenues within the organisation’s departments and branches. The use of the transfer price of funds \( p_M \) can be illustrated with the following stylised figure of a bank’s balance sheet: \(^7\)

\[\begin{array}{c}
\% \\
MR \\
\end{array}\]

\[\begin{array}{c}
A \\
\end{array}\]

\[\text{Operating cost. Furthermore, the level of equity capital is indeterminate, as pointed out by Dermine (1986). Another similar formulation is in Neven and Röller (1996).}\]

\(^6\) See, e.g., Matten (1996), Smullen (1996) and Bergendahl (1989), which also contains an interview survey with financial officers of 11 major banks in Sweden and the UK. The author has also conducted interviews with financial management officers in two major Swedish banks, which verify that they use an average of short-term interest rates to calculate \( p_M \), ranging from overnight rates to 3 month money market rates.

\(^7\) Marginal costs are here assumed to be linear and monotonically increasing. Fixed assets, bonds, and purchased funds are omitted for clarity.
The bank in Figure 1 has one aggregated loan category earning interest along the marginal revenue line starting in \( p_i = MR_i \) and one aggregated deposit category paid interest along the \( MC_{FC} \) line starting in \( \tilde{p}_i = MC_{FC} \), where FC denotes funding cost. The bank is assumed to be operating in an oligopolistic market, so it is an interest rate setter, as is evident from the negative slope of the marginal revenue curve for loans and the positive slope of the \( MC_{FC} \) curve for deposits.

The amount of equity capital is the area A and the stock of compulsory reserves is the area B. Given the interest rates set by the bank, it is optimising the level of \( V \), i.e. its activities both regarding deposit-raising and loan-giving, since the marginal cost equals marginal revenue for both. We should specifically notice the double role of the (internal) transfer price of funds \( p_M = MR \) for deposits and \( p_M = MC_{FC} \) for loans, i.e. the marginal cost for fund-raising. This has important implications for the specification of the statistical cost function, since it needs to contain only such cost items which are not per se separable between the different outputs, see further Essay 3.

The argument elaborated above for separating the fund-raising (mainly deposits) and asset-placing (mainly loans) activities of the bank, and pegging them to an internal transfer interest rate \( p_M \), normally equal to the short-term money market interest rate, is firmly

---

\(^8\) This model specification avoids the disequilibrium properties regarding the amount of deposits raised when \( \delta > 0 \) that were demonstrated by Sealey and Lindley (1977, Appendix 1).
based in the practice of modern banks, although it has been more observed in the bank management literature than in academic bank studies. For example, all major Swedish banks have a “flat” decentralised organisation in the sense that branches in the bank are made into rather independent profit centres.\footnote{Swedish examples are Handelsbanken, which has been organised in this way since the early 1970’s as an early example of “business process re-engineering”, see Wallander (1991). Swedbank changed its organisation accordingly in the early 1990’s. All major Swedish banks are nowadays organising their branches as profit centers where branches manage their own balance sheets.} $p_M$ then functions as the peg of the bank not only in its market relations as a financial intermediary, but also in its internal organisation as a financial entity, subject to modern financial management methods. $p_M$ brings independence between the deposit and the loan rate decisions. The reason for this is that all funds in the bank must be traded with the bank’s treasury unit, or “internal bank”.$^{11}$ The (internal) transfer price of funds $p_M$ is used as a managerial tool in two respects: 1. To manage liquidity within the whole bank organisation, and: 2. To manage interest rate risks.

The first aspect of how $p_M$ is used regards the short-term liquidity management. Each branch is free to have imbalances in its day-by-day balance sheet. Liquidity surpluses or deficits are automatically traded with the treasury unit of the bank. Surplus liquidity is sold to the treasury at $p_M$ and deficits are automatically borrowed at $p_M$. The consequence is that in a situation when $p_M$ is exactly in the middle of the spread, then (non-cash) surplus liquidity in one branch will earn equally much nominal interest revenue to that branch as a liquidity deficit in another branch.

A profit-oriented bank branch will also use $p_M$ to decide the value of generated deposits and loans as parts of the branch’s total revenues, $^{11}$

\footnote{In practice this applies to any specialized department in a bank. Here I take branches merely as an example. The remainder of this section serves to support the assumption of separability between the loan and deposit market activities of a modern bank. No attempt is made to model the optimisation of the internal bank organisation; the objective function (1) refers to the bank as a whole.}

\footnote{An optimising bank earns exactly its opportunity cost of funds at the margin on each asset category: interest revenue less operating costs. And it uses each category of funds exactly to the extent that the opportunity cost on the margin, i.e. interest cost plus operating costs, is equal for all categories.}

---

9 Swedish examples are Handelsbanken, which has been organised in this way since the early 1970’s as an early example of “business process re-engineering”, see Wallander (1991). Swedbank changed its organisation accordingly in the early 1990’s. All major Swedish banks are nowadays organising their branches as profit centers where branches manage their own balance sheets.

10 In practice this applies to any specialized department in a bank. Here I take branches merely as an example. The remainder of this section serves to support the assumption of separability between the loan and deposit market activities of a modern bank. No attempt is made to model the optimisation of the internal bank organisation; the objective function (1) refers to the bank as a whole.

11 An optimising bank earns exactly its opportunity cost of funds at the margin on each asset category: interest revenue less operating costs. And it uses each category of funds exactly to the extent that the opportunity cost on the margin, i.e. interest cost plus operating costs, is equal for all categories.
as illustrated in Figure 1. When the top management of the bank changes the level of $p_M$, it accordingly changes the value of deposits and loans since opportunity costs change. This will also influence the interest rate risk exposure for the bank. Deposits have a different maturity structure than loans, so the interest rate risk exposure changes accordingly. However, branches are not allowed to take the interest rate risk that ensues from intermediation. Instead, since all deposits raised in branches are sold internally to the treasury department at $p_M$, and all loans granted by the branches are funded with liquidity bought at $p_M$, the interest rate risk from intermediation is lifted away from the branches and become centralised in the treasury unit.\footnote{One non-branch example is the Swedish telephone bank Sesam, which is owned by S-E-Banken. Even though it operates as an independent entity, it isn’t allowed to take any interest rate risk. It has to transfer all its deposited funds through the S-E-Banken group treasury, thus pegging all interest rates to $p_M$. (Interview by author with Sesam CEO, Nov.1996.)}
References

Arndt, H.W., 1984, Measuring Trade in Financial Services, Banca Nazionale del Lavorno Quarterly Review, 149: 197-213


