Competition and Market Structure in the South African Banking Industry

Yared Teka Tsegay

A minithesis submitted in partial fulfillment of the requirements for the degree of Magister Commerce in the Department of Economics, University of Western Cape

Supervisor: Prof. P. A. Black

May 2004

Acronyms

CR_4	Concentration ratio of the four largest firms
CR_5	Concentration ratio of the five largest firms
GLS	Generalized least squares
HHI	Herfindahl-Hirschman index
IO	Industrial organization
Ir	Interest revenue
OLS	Ordinary least squares
PPC	Production Possibility Curve
Pl	unit price of labor
P-R	Panzar–Rosse
Pr	unit price of fund
PCM	Perfectly contestable market
NEIO	New empirical industrial organization
SCPP	Structure-conduct-performance paradigm
NPL	Non performing loans

Competition and Market Structure in The South African Banking Industry

Y. T. Tsegay

M.Com. minithesis, Department of Economics, University of Western Cape.

KEYWORDS

Competition
Market structure
Banking industry
Contestability
Conduct
Elasticity
Barriers to entry
Stability
NEIO
Regulation

SACCOL
South African Credit Cooperatives League

22
48
52

Competition and Market Structure in The South African Banking Industry

Y. T.Tsegay

Abstract

The South African banking industry is relatively sound and adequately capitalized high-tech service industry. It provides services to 51% of the economically active population. In the late 2001, 2002 and 2003, a number of small banks failed due to liquidity crises, which subsequently spurred the momentum of consolidation in the industry. During this period, a great deal of mergers and acquisitions has taken place in the industry. The regulatory environment is in line with the Bank of International Settlements (BIS) guidelines. The regulations regarding foreign banks are evidently a barrier to foreign bank entry, which could spur the level of competition in the industry

In this thesis the level of market concentration is calculated using concentration ratios, CR_4 and CR_5, for four product markets. The concentration ratio R_4 ranges from 77% for investment products, 79% for deposit market, 89% for installment sales market to 95% for credit card markets. The Herfindahl-Hirschman Index (HHI) as calculated by reserve bank is 1750 by the end of year 2002, compared with the thresholds set by American Department of Justice; this falls short by only 50 from the range of the highly concentrated markets, which is above 1800 points.

The methodology used to test competition, one of the new empirical industrial organization models, is the Panzar-Rosse reduced form revenue function. The P-R reduced form revenue function estimates the competitive conduct by the extent to which changes in factor price is reflected in revenue. The reduced form revenue function is estimated based on unbalanced panel of 15 banks for the period of 1993 to 2002 using fixed effects panel data econometrics. The estimated H-statistic is 0.516,

which falls between the ranges of zero and one; this clearly indicates monopolistic competition is prevailing in the industry. Monopolistic competition is characterized by product differentiation and excess capacity. This is definitely so in our case, where the banking sector services only 51% of economically active population.

May 2004

Declaration

I declare that *Competition and Market Structure in The South African Banking Industry* is my work, that it has not been submitted for any degree or examination in any other University, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Yared Teka Tsegay May 2004

Signed: .....

Acknowledgment

I thank my family and friends for their love, support and constant encouragement. I would like to express my indebt ness to my supervisor Prof. P. A. Black, for his resourceful comments and guidance. I am grateful to Ms. S. Aklilu for editing the manuscript. I am grateful to Mr. Habtemriam, Mr. Habte and Mr. kesete for their friendship, support and encouragement. Special thanks goes to Ms. R.P. Mosia for her kindness and friendship.

Content

List of Tables, Graphs and Figures **Page**

Figures

Graphs

Chapter I

INTRODUCTION

The analysis of market structure and competition falls within the spectrum of theories of industrial organization. This theory has been evolving from the main stream Structure-conduct-performance hypothesis that holds market structure determines the type of conduct the firms engage in and thereby affect the overall performance. Demsetz's efficiency hypothesis claims that efficiency determines the market structure as well as performance, while Baumol's contestability theory asserts that threat of free entry and exit are the determining factors. The new empirical industrial organization theories (NEIO) emphasizes on the measuring the degree of competition for individual industries irrespective of market structure.

The South African banking industry is characterized by high level of market concentration. The level of concentration and consolidation has dramatically risen since the second half of 2001. The main research problem that this thesis is tying to address will be "is the high level of market concentration and consolidation accompanied by a comparative level of competition in the industry?" For this purpose market concentration ratios are calculated and a model is simulated to measure the degree of competition in the industry.

The market structure of the South African banking industry is measured by concentration ratios (CR_4 and CR_5). These ratios are calculated for four product markets, i.e. the deposit, the credit card, the installment sales and the investment product markets as at June 2003. Dynamics of asset distribution for the period of 1994 to 2003 as well as the Herfindahl-Hirschman Index (HHI) (1995-2002) as calculated by the reserve bank are considered in the analysis of market structure.

The degree of competition is measured by the Panzar-Rosse (1987) reduced form revenue function. This model estimates the level of competition in an industry by measuring the factor input price elasticity, and infers the degree of competition in the industry. The P-R model is simulated for the period of 1993-2002 and for the sample of 15 banks using fixed effects panel data analysis.

Chapter one is an introduction. Chapter two deals with competition and efficiency, subsequent developments in theories of industrial organization and its application to banking industry. Chapter three looks into the structure of the South African banking industry, i.e. the calculation of concentration ratios, the involvement of foreign banks, and regulation, consolidation, and bank failures in the industry. Chapter four deals with the model, methodology and data set. Chapter five deals with econometric simulation. Lastly, the conclusion and implications are dealt in chapter six.

1.1 Overview of the South African Banking Industry

The South African banking industry is the largest and the most sophisticated industry in the continent. The total asset of the South African banking industry is estimated to be R13, 481 billion at the end of December 2003 (Reserve Bank, DI900). It is part of the financial sector, which made 20% contribution to the GDP of the country in year 2000. The number of accounts held by the seven big banks is 27.2million, the big four banks employ 113, 240 people while the foreign banks employ 1,776 people. The big four has 825,000 Internet banking customers (Price Water House Coopers, 2003:10).

The exclusion of a large sector of the South African population from banking services is another feature of the banking industry. A study by ABSA (1999) indicates that only 20% of the country's economically active population of 16.5 million has access to banking services. The availability of banking services is limited to government institution, established business sector and economically upper end of the society. The

very recent Finscope Report (2003) reveals that only 51 % of the economically active population has access to banking services.

The regulatory body, the Registrar of Banks, undertakes its supervisory and regulatory activities through the bank supervision department of the Reserve Bank. The primary objective of the department is to ascertain the safety and soundness of the banking system. The regulatory framework is compatible with the bank of international settlements guidelines. The banks and the registrar of banks are working towards implementation of the new regulatory framework the Basel II Capital Accord.

In December 2002, there were 97 Banks engaged in providing banking services. This figure is composed of 28 locally registered banks, 2 mutual banks, 14 branches of foreign banks and 52 foreign representatives. The number of banks has been on increasing trend from 1994 to 2002. It has reached its peak number of 120 banks in year 2000, since then the number of banks has fallen to 96 as of December 2002, mainly due to failure of small banks in the industry. This phenomenon can be illustrated using the HHI[1]; in 1995 the HHI of the industry was 0.170 (1700), the index has been on decreasing trend since then, and it reached 0.131 by the year 2001. This can be explained by entry of new local as well as foreign firms into the industry. However, the trend was reversed in late 2001, 2002 and 2003 due to wide spread failure of small banks, the exit, merger and acquisitions ensued.

To sum up, the key features of the banking industry are: it provides state-of-the-art banking services, very limited provision of services to the established business sector and to only half of the economically active population. It is highly concentrated industry with the top five firms owning 87% of total banking assets. It has a well-developed regulatory authority which keeps pace with recent international practice.

[1] The HHI is concentration measure, which encompasses all firms in the industry. For detailed discussion refer section 3.1.3.2

1.2 The Research Problem

What is the degree of competition prevailing in the banking industry? What is the level of concentration in the industry? What does the assessment competitive conduct in the industry look like? Does this market structure reduce the competitive conduct of the industry? Is the government intervention in the industry to safeguard competition justified? (E.g. Nedcor and Stanibic).

1.3 Aims of the Research

There is a very limited literature regarding the analysis of market structure and competition in the South African banking industry. The existing literature uses the SCPP methodology, which lacks theoretical vigor. The academic aim of this thesis will be to bring the relatively recent and theoretically robust methodology, namely NEIO, into the discussion of competition and market structure of the South African banking industry. The strategic aim of this research is to reduce the over dependence on the SCPP and the use of concentration ratios on the analysis and appraisal of competition in the banking industry

1.4 Rationale

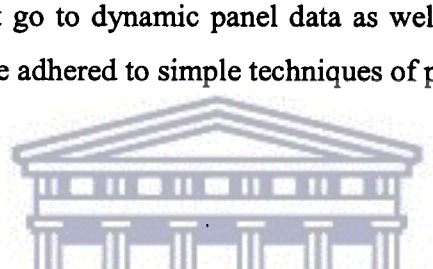
The rationale of this study emanates from the current state of affairs in the industry where some small banks were compelled to cancel their license, some banks failed while others opted for merger and takeovers. Furthermore what makes it more interesting is the macroeconomic effect of competition in this industry. Smith (1998) illustrates some of the macro economic benefits of the more competitive banking system: increased competition in banking tends to increase the level of economic activity and reduce the severity of business cycles.

The choice of the P-R model comes from the understanding, that after the advent of contestability theory and subsequent application of game theory in the industrial economics make the use of structural SCP models theoretically problematic albeit

concentration ratios are still widely used in empirical studies. The South African banking industry is highly concentrated (Competition commission, 2001; Okealaham, 2001). The central theme of this thesis is: does this highly concentrated industry enjoy comparative competitive conduct in the industry?

1.5 Limitations

The basic limitation of this thesis is its data constraint. The size of the sample is only 15 banks. The result of the estimation would be more robust if a larger data set could be accessed. The model simulated used a simple and basic panel data estimation technique. It does not go to dynamic panel data as well as to unit root tests. Thus it must be noted we have adhered to simple techniques of panel data.



Chapter II

THEORETICAL FRAMEWORK AND LITERATURE REVIEW

The analysis of competition and market structure naturally falls within the spectrum of theories of industrial organization. Thus we need to review the subsequent developments in the theories of industrial organization. The relationship between competition and market structure is vague at best. There are some empirical findings, which claim there is a strong correlation between profitability and market structure; however Chicago–UCLA School, has contested the claim of causality. The relatively recent New Empirical Industrial Organization theory, which came into existence as a response to dissatisfactions with the preceding methodologies, has developed a tool to measure competition independently from market structure.

2.1 Competition and Economic Efficiency

Does competition prop up economic efficiency? To give an appropriate answer to this question, we need to look into different types of efficiency. Economic efficiency is multifaceted (Scherer, 2000:211). It comprises allocative efficiency, X-efficiency (technical efficiency) and long run technological efficiency. The traditional "static efficiency" analysis examines efficiency using the static allocative and x-efficiencies, however this is not sufficient. Dynamic efficiency gives an important insight to the analysis. In 1942, Joseph A. Schumpeter challenged the traditional "static" models of analysis. He argued "what really mattered for economic well-being in the long run was not how efficient resources were allocated and utilized at any moment in time, but how well entrepreneurs seized opportunities for reducing cost through technological innovation" (Scherer, 2000:213).

2.1.1 Static Efficiency Analysis

Economic efficiency is conventionally defined in terms of both allocative and technical efficiency. The former estimates an efficient allocation of productive resources amongst alternative uses giving rise to production of optimal mix of commodities while the later refers to a situation in which existing resources are utilized in the most efficient manner producing the maximum possible output from a given set of resources.

Allocative efficiency induces higher level of social welfare. Monopoly creates a loss of social welfare in the economy. It is argued this loss of welfare is mitigated by high level of technical efficiency. Market concentration ratios such as CR_4, CR_5, and HHI are extensively used in static efficiency analysis. Theses measures give a glimpse of the market at a point in time.

There are diverse empirical evidences on whether competition improves x-efficiency. Competition can induce x-efficiency by forcing firms to opt for efficient technical methods and cut corporate fat. Firms will be forced to adapt ways and methods to utilize the existing resource to the optimal level in order to survive the cutthroat competition posed by other firms. They have to tighten their belts and cut organizational fat. The 1990's wave of restructuring of big Multinational Corporation and outsourcing some of their activities can be explained by high level of competition created by globalization. In other extreme, it is argued such arrangement hampers firms from realizing their economics of scale. (Scherer and Weiss 1973, 1970, cited in Scherer, 2000:215) revealed that the incidence of plants operating at size too small to realize all know economics of scale was greater the less concentrated an industry was.

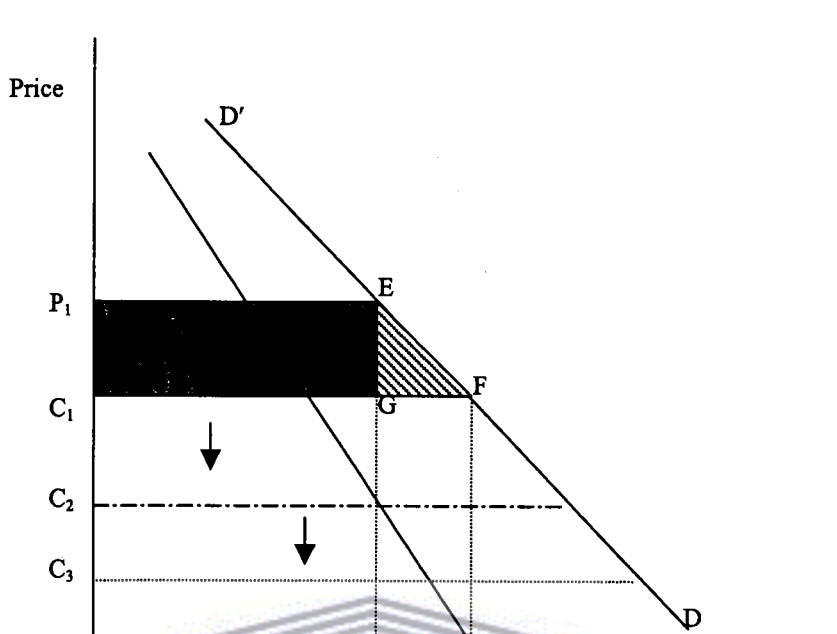


Figure 2.1: Efficiency Analysis

The above figure illustrates the efficiency analysis. DD' is the demand curve. The production and distribution cost is assumed to be $0C_1$ dollar per unit sold. Because a monopolist controls the industry, the price is set at $0P_1$ above the marginal cost, producing $0X_1$ quantity output. The area P_1EGC_1, the producer's surplus, is the wealth redistribution from consumers to producers. It is the profit monopolists enjoy by setting price above the cost. The triangle EGF is a "deadweight loss", as the price goes up to $0P_1$ from the cost $0C_1$ the quantity produced reduced by X_2X_1. In perfectly competitive benchmark, the society would be consuming X_2 quantity of output at price C_1. This "deadweight loss" is attributable to the monopoly.

The monopolistic profit, P_1EGC_1, might not always be realized. In absence of competitive pressure, x-inefficiencies associated with monopolists might let the cost $0C_1$ to go up to $0P_1$ turning the rectangle P_1EGC_1 wealth redistribution into pure x-inefficiency. The monopolists do not have the motive to cut costs to stick to the minimum cost curve $0C_1$, so there is a tendency for the cost to go above the $0C_1$. Conversely, Schumpeter argues monopolists have the motive and the capability to

engage in research and development to bring about reduction in costs through technological inventions and innovations. This can be seen in the above figure, where the cost curve shifts from $0C_1$ to $0C_2$. This is what is called dynamic technological efficiency, Schumpeter advises policy makers should be more concerned with materializing the dynamic efficiency rather than with allocative and x-efficiencies.

2.1.2 Dynamic Technological Efficiency

The log run dynamics is more important than the efficiency of static resource allocation and utilization. This refers to invention and innovation firms engage on to spur their profitability. Theses efforts bring about technological progress, which enable the society to bolster its production capacity. Scherer (2000:220) identifies the three main Schumpeterian postulates concerning the links between monopoly and competition and the pace of technological change:

I. Large businesses have the capacity to undertake a major research projects, and can afford the risk to undertake them. Some research projects need a large financial and institutional capacity to undertake them.
II. These large businesses can finance the industrial research, development and innovation, using their high monopolist's profit, and
III. The monopoly position ensures a sufficient utilization of the product of the technological and innovative investments. The technological gains made rewards the monopolists, such that the investment made on the research and development appear to be worthwhile.

This was followed by a number of theoretical and empirical findings regarding innovation and firm size as well as innovation and market structure. These studies "compel the abandonment of the romantic but naïve Schumpeterian belief that giant firms organized into highly concentrated oligopolies are essential to maintain the most vigorous pace of technological progress" (Scherer, 2000:223).

Let's sum up by introducing of Production Possibility Curve (PPC) into the discussion. The technical efficiency (x-efficiency) indicates technically efficient production of goods and services, and it ensures that the society is on the PPC. On figure 2.2, point C_o and M_o lays on the PPC M_oN_o, this represents technical efficiency. In absence of competitive pressure, the monopolist is not inclined to make an efficient use of existing resources, so in the figure below the monopolist produces at point M_1 that lays below the PPC M_oN_o. This represents an x-inefficiency, that is, some times termed as organizational slack.

Under perfectly competitive situation, society prefers to produce and consume at point C_0, 0-m_1 quantity good Y and 0-n_2 of good X. Thus point C_0 is allocatively efficient point in the PPC. However the existence of monopoly in sector Y forces the society to produce and consume at the point 0-m_2 quantity of good Y and 0-n_1 of good X. Thus the equilibrium moves from point C_o to M_o, which means the society is worse off.

The Static efficiency is concerned with allocative as well as the x-efficiencies; this means the production and distribution on existing PPC-M_oN_o. The dynamic efficiency is concerned with the leap from the existing PPC-M_oN_o to a higher-level PPC, in our case M_1N_1, through technological invention and innovations. As we have already seen Schumpeter argues in the highly concentrated market, for that matter monopolists, have the capacity and willingness to undertake research and development projects, which could enable the society to be on higher level of PPC. The society would be better off to be at higher PPC at point M_2 albeit it is x-inefficient and allocatively inefficient associated with monopoly.

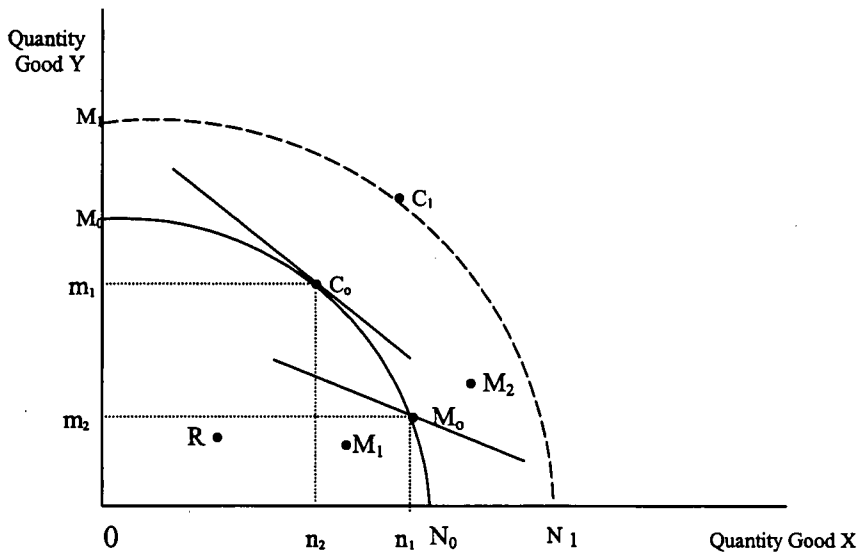


Figure 2.2: The PPC Analysis

To sum up, the inefficiencies involved in the imperfect competition are the raison d'être for competition policy. Policies striving to improve competition in the industries are not aimed to have competition as such but it is with the belief that competition brings about economic efficiency. Economic efficiency is the end result of the competition process in an industry; it is the optimal level production and distribution of goods and services in the economy.

2.2 Theories of Industrial Organization

The evolution of theories of industrial organization emerged with the seminal work of Manson in the 1930's. He brought to attention that the degree of concentration in an industry as an indicator of market power, hence of undesirable effects on economic welfare (Smit, 1999:6). It was followed by major contributions of Bain (1959) which shaped the concept of SCP, Demsetz's (1974) efficiency hypothesis, the recent new industrial organization theory pioneered by Baumol *et al's* (1982) contestability theory; and new empirical industrial organization theories put forward by Panzar-Rosse (1987) and Bresnahan (1982,1989). Recently, of course, game theory has revolutionized the theories of industrial organization.

The SCP approach analyses the interactions among market structure, business conduct, and the social and economic performance of an industry. The essence of the approach is to establish causal links between industrial market structures, how variation in one of the dimensions will lead to variations in the others. It argues a causal link runs from market structure to conduct and from conduct to performance. The structure of the market usually influences the behavior of the firm as to how to compete or collude, and the behavior that affects performance in the market.

Demsetz's (1974) efficiency hypothesis refutes the SCP paradigm. He argues that the micro economic model of profit maximization in perfectly competitive markets in the long-run equilibrium is sufficient to explain real world phenomenon (Smit, 1999:12). Firm's conduct determines the market structure. Competition is viewed as a process that can result in a variety of market structures each yielding efficient industry performance. Thus, if efficient production within a particular industry requires large firm size, then few firms will prevail and the industry will be highly concentrated (and efficient).

The SCP has substantially modified with the advent of game theory to theories of industrial organization. The uni-directional causation proposed by the original SCP paradigm gave its place to bi-directional causation with performance and conduct having a feed back effect on the market structure. Since 1970's the literature have recognized that conduct can influence structure and have abandoned the unidirectional flow of causation fundamental to SCP framework. The application of game theory to industrial organization and an understanding of endogenous forces affecting market concentration indicated that structure is as much dependent on conduct, as conduct on structure. In many instances these attributes were simultaneously determined by the strategic action of firms (Audretsch *et al*, 2001:616).

Contestability theory argues that the threat of entry alone can lead to competitive conduct independent of number of firms actually acting in the market given free entry and exit (Hempell, 2002:3). This theory assumes that there are no barriers to entry and

no significant sunk costs involved, and firms will not experience a disadvantage in terms of production technique or perceived product quality. Firms would enter to industries, which they found profitable and able exit the industry without incurring costs. (Baumol, 1982:5) notes this theory departs from the older theories which implicitly took industry structure to be determined exogenously as the industry structure is determined endogenously along the prices and outputs.

Baumol *et al* (1986:339) explain, "the perfectly contestable markets (PCM) phenomenon, on idealized limiting case, where potential entry imposes the strongest possible symmetric constraint upon incumbents, entry without disadvantage and is perfectly reversible". They propose perfectly contestable markets (PCM), in place of perfect competition, as a general standard of comparison for more-complex and more realistic models of industrial organization. Yet this alteration, in general standards, does not change the implications; since perfect competitive behavior is a necessary but not sufficient condition for PCM. However, we must note that this theory holds in the ideal limiting case; one reason is that it is not robust to small change in some of the assumption. In particular, if entry lag exceeds the price adjustment lags, apparently most industries have considerable sunk costs. But still this theory has been instrumental in causing antitrust analysis to reduce their emphasis on concentration and take proper account of potential competition.

The advent of game theory has revolutionized the theories of industrial organization. It has become the standard language of industrial organization. The contribution of game theory to IO comes in two related forms: as a language as well as (imposition of) a discipline on IO. Game theory as a language makes models to be described in an accurate and economical way using standard familiar formats; furthermore it provides flexibility in the sense that it allows considerations of the situations involving dynamic interactions and imperfect information regarding behavior and the environment.

The second contribution, Game theory imposes a discipline on modeling. Firstly, a non-cooperative game model requires the analysts to specify precisely the actions

available to players, time of actions and information hold by players this forces the modelers to face their assumptions and hence question them. Secondly, with game theoretic framework, results have to satisfy the requirement known solution concepts, Nash equilibrium and its refinements. This forces arguments to be complete in the way dictated by the solution.

Game theory has provided a language and discipline to IO. However, does game theory offer a new insight and better understanding to IO? It does offer a better understanding aspect of actual market behavior. The contribution of Game theory in the areas of oligopoly, barriers to entry, predatory pricing etc is evident. It gave rise to a new policy perspective, Audretsch *et al* (2001:616) explain

> The use of game theory, as a tool in the literature of economics of industrial organization, also illustrates that small adjustments in economic models can generate large changes in the competitiveness of firm's behavior. For competition policy this implies that small variations in the basic circumstances of an industry can often be sufficient to make the difference between warranted and unnecessary regulation. And the argument even raised the possibility that a unique applied economic analysis may be necessary for the formulation of competition policy.

The new developments in industrial organization theory that is the game theory and contestability theory made the interest in concentration measurements to decline. Waterson (1993:112) identifies the common factor in both contestability theory and most game theoretic approaches, which makes linking concentration to theoretical analysis problematic. Both theories do not concern themselves with distribution of firm sizes. But it is fundamental that concentration is a function not only of firm number but also firm sizes. Contestability theory, with its emphasis on technology that determines structure, has no real scope for distribution of firm size within an industry, Game theoretic approach with their emphasis in demonstration, have no need for firms to be equally sized.

2.2.1 New Empirical Industrial Organization Theory

> The NEIO research identifies and estimates the degree of market power, specifies and estimates the behavioral equations that drive price and quantity, and often infers marginal cost or measures market power without it. (Fischer and Kamerschen, 2003:3)

This approach emphasizes on empirical work in specific industries. It makes use of time series data for single industries or cross sectional studies for closely related markets. It differs from the preceding studies on the following fundamental points: What can be observed and on how economic quantities are to be measured. This approach emerged as a response to theoretical and empirical deficiencies of the preceding models. Bresnahan (1989) identifies three hypotheses made by SCP paradigm that NEIO claims as flawed are:

- Economic Price-cost margins (performance) could be directly observed in accounting data.
- Cross section variation in the industry structure could be captured by small number of observable measure.
- Empirical work should be aimed at estimating the reduced form relationship between structure and performance.

The NEIO has developed techniques to assess the conduct and market power of specific industries depending on the parameters estimated based on tailored models developed for specific industries. The widely used models of NEIO are Bresnahan and Panzar-Rosse, which are applied to a number of industries in a number of countries. For example, P-R model is extensively used in banking sectors.

Bresnahan (1989:1012) summarized the central ideas of the new approach:

- The firm's price-cost margins cannot be observed. The economic marginal cost cannot be directly observed therefore it can only be inferred from the behavior. It scrutinizes the difference between closely related markets to trace the effects of change in marginal cost (MC) or comes to a quantification of market power without measuring the cost at all

- It is assumed that individual industries have peculiar features. The idiosyncrasies affect the operation of the firms as well as their conduct. Thus any comparative static of variation across industries are absurd, unless they are closely related.
- Firms and industry conduct are viewed as unknown parameters to be estimated. The behavioral equations by which firms set price and quantity will be estimated, and parameters of those equations can be directly linked to analytical notions of firm and industry conduct.

2.2.1.1 Bresnahan (1982, 1989)

This model aims at estimating parameters, which gives us the measurement of the degree of competition. It assumes the oligopoly interaction on single product industry and assesses the firms and industries conduct. The model has three sets of unknown parameters: costs, demand and firm conduct. The observable variables that are endogenous to industry equilibrium include industry price and each firm's quantity. The observable include variables that shift the cost and demand functions, while the price-cost margins are not directly observable.

Bresnahan (1989) explains that oligopoly theory is used to specify the equations of the model to be estimated. The use of the theory to specify the model will be emphasized over inference. Inferences about market power will be identified only through refutable implications of the theory contained in the comparative static or comparative dynamics of oligopoly equilibrium

Demand function is estimated in this stylized model. For the sake of convenience, it is set as a function of price

$$P_t = D(Q_t, Y_t, \delta, \epsilon_{dt}) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Where *i* will index the firms & *t* will index the time. Since we are dealing with single (homogenous) product case $Q_t = \sum_i Q_{it}$, i.e. demand for the Industrial output is composed of individual firms demand.

P_t market price,

Q_t individual firms' quantity

Y_t all variables shifting the demand function

δ Unknown parameters

ϵ_{dt} Error terms

The marginal revenue will be

$$P_t = D_1 (Q_t, Y_t, \delta, \epsilon_{dt}) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The cost function for i^{th} firm is treated in a similar way

$$C_{it} = c (Q_{it}, W_{it}, Z_{it}, \Gamma, \epsilon_{cit}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Where

W_{it} vector of factor prices for firm i at observation t

Z_{it} variables that shift the cost function for firm i at observation t

Γ unknown parameters, and

ϵ_{cit} error terms

Then the marginal cost function should

$$C_{it} = C_1 (Q_{it}, W_{it}, Z_{it}, \Gamma, \epsilon_{cit}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

At equilibrium, of course, the marginal revenue is equal to marginal cost; however we need to pay attention to the type of competition prevailing in the market. The marginal revenue can be defined in terms of price and elasticity i.e. $MR = P + Q/\alpha$ where α price elasticity of demand. Bresnahan (1989:1016) puts it "outside the perfectly competitive model firms do not have a supply curves (P = MC), instead, price-quantity conduct follows more general supply relation". Thus the monopoly marginal revenue is $P\text{-}D_1Q$, Therefore the supply relation

$$P_t = C_1 (Q_{it}, W_{it}, Z_{it}, \Gamma, \epsilon_{cit}) - D_1 (Q_t, Y_t, \delta, \epsilon_{dt}) Q_{it}. \theta_{it}. \quad . \quad . \quad (5)$$

This has the interpretation of MC = "perceived" MR for oligopoly models.

The θ_{it} is the measure of market power; the parameter θ_{it} is a positive unknown term and indicates the competitiveness of oligopoly conduct. As the θ_{it} moves towards zero

the conduct of the firm moves towards perfectly competitive conduct and as the θ moves further from the zero then the market conduct moves away from the perfectly competitive interaction. After estimating the equation, the θ will be used to infer the type of market conduct prevailing in the market. For example, in case Bertrand oligopolist arrangement the parameter θ_{it} is zero while for cournot oligopolist the parameter θ_{it} is one.

To sum up, the Bresnahan (1989) provides a tool to measure the degree of market power by estimating the parameter θ_{it}. The econometric application of this model has been used in banking industry. Bikker & Haaf (2000) has used this model to assess the market power in European banking industry after the introduction of EURO along the P-R reduced form revenue function.

2.2.1.2 Panzar –Rosse[2]

It estimates the competitive behavior of firms by the extent to which changes in factor prices are reflected in revenues. It uses the reduced form revenue equations derived from marginal revenue and marginal cost functions.

This approach assumes that:

- Firms cost functions are linearly homogenous in the factor prices or that production functions are homothetic,
- Firms operate in their long run equilibrium,
- Factor prices are exogenous to the individual firm, and
- Performance of theses firms needs to be influenced by actions of other market participants

Then banks maximize profits when Marginal revenue equals Marginal cost

$$R'(x_i, z_i) - C'(x_i, w_i, t_i)=0 \quad i, \ldots . n, \text{ where } n \text{ number of firms}$$

R' = marginal revenue for firm i

[2] See page 43 for detailed discussion on the P-R reduced form revenue function.

C' = marginal cost for firm *i*

x_i = the output of firm *i*

m_i = vector of *k* factor of inputs prices of firms

z_i = vector of exogenous variables that shifts the firms revenue function, and

t_i = vector of exogenous variables that shifts the firms cost function.

At equilibrium the zero profit constraint holds

$R^*(x_i, z_i) - C^*(x_i, m_i, t_i) = 0$

The market power is measured by the extent to which a change in factor input prices (∂m_{n_i}) is reflected in equilibrium revenue (∂R_i^*) earned by firm *i*. Panzar-Rosse define measure of competition, the H- statistic as the sum of elasticity's of the reduced from revenue function with respect to factor prices.

$$H = \sum_{n=1}^{k} \frac{\partial R_i^*}{\partial m_{n_i}} \frac{m_{n_i}}{R_i^*}$$

As De Bandt and Davis (2000:162) explained if perfect competition prevails in the market, banks operate at their long run equilibrium. Then proportional increase in factor prices induces an equi-proportional change in gross revenues; output does not change in volume terms, while the output prices rise by the same extent as the input price. On the other hand under monopolistic competition, revenues will increase less than proportionally to changes in factor prices, in limiting case of monopoly there may be no response or even a negative response of gross revenue to change in output cost.

Thus, the value H-statistic, according to the P-R, is zero or negative ($H \leq 0$) when the competitive structure is monopoly, perfectly colluding oligopoly, or conjectural variations of short run oligopoly. While the value of H-statistic is between zero and one ($0<H< 1$), when there is a monopolistic competition. The H value is zero ($H=0$) in perfectly competitive market.

2.2.1.3 Application to Banking Industry

Gelos and Roldós (2002) have applied Panzar–Rosse (1987) methodology to sample of eight European and Latin American countries. They have examined the evolution of market structure in emerging market banking systems during the last decade. Their findings indicate that though a significant amount of bank consolidations took place in these countries, it is not associated with high level of concentration when estimated by the standard indices. The resulting econometric estimation indicates that the market have not become less competitive. They concluded that lowering barriers to entry have been effective, specially the entry of foreign banks, in preventing a decline of competitive pressure due to consolidation.

Hempell (2002) conducted an empirical study to test the degree of competition in German banking industry by making use of P-R model. The P-R reduced form revenue function was estimated based on the micro data of banks' balance sheets, profit and loss accounts for the period 1993-1998. The study concludes despite the decreasing number of banks in Germany and a slight increase in concentration during the study period, these are no clear indication of different competitive behavior in the second half of the time period under investigation.

Bikker & Haaf (2000) undertook a study on European banking industry based on non-structural models of competition the P-R and the Bresnahan models. The P-R was estimated for 23 countries overtime span of 10 years. The resulting H-statistic provided strong evidence that the banking markets in the industrial world are characterized by monopolistic competition.

Okealaham (2001), which tried to evaluate competition and concentration in the South African commercial banking, used the SCP paradigm on analysis of the banking industry. A model was developed based on Reserve Bank's DI900 RETURNS data. Since commercial banks in South Africa offer a wide array of retail deposits, average of six types of deposit accounts were taken. The model was specified, interest rate as

dependent variable and two concentration measures (CR_3 and HHI) and other exogenous variables as explanatory variables. The finding depicts that concentration ratios are significant and main determinants of interest rate payments. From a policy perspective, the model clearly indicates that consumers were paying for profit of banks in the high concentration level and this contributed to welfare costs (Okealaham, 2001:14).

Conclusion

The structure-conduct-performance, the efficiency hypothesis, the contestability theory, and the new empirical industrial organization theory are milestones in the evolution of theories of industrial organization. The introduction of game theory to the theories of industrial organization has had a significant impact on the contemporary approach. For purposes of measuring degree of competition in specific industries, the new empirical industrial organization theory gives the appropriate tool that is independent of the market structure. The reason for the emphasis on competition comes from the understanding competition brings about economic efficiency thereby improving the welfare of the society.

Chapter III

THE STRUCTURE OF THE SOUTH AFRICAN BANKING INDUSTRY

For purposes of analyzing the market structure and competition in banking industry, it is important to discuss the nature of banking industry and its product markets and types of banking services. There are two theories in the banking industry, namely the "production approach" and the "intermediation approach". Intermediation approach views banks as firms producing intermediation services, value of loan and investments are output; while labor, capital and deposits are inputs. Production approach treats banks as firms that use capital and labor to produce different types of loan and deposit accounts; output can be identified with the number of accounts.

To undertake competition analysis, the industry is categorized into product markets and geographical markets. This is in line with the European competition authorities practice. Thus the industry can be categorized in to three broad categories.

- Retail banking is composed of deposit taking, lending and payments handling.
- Corporate banking comprises payments handling, deposit operations, lending, documentary business and foreign trade financing.
- Investment banking consists of corporate finance, stock market listing and share and bonds

3.1 Banking Structure

According to the Banking Act of 1990, only public companies[3] are allowed to engage in business of banking in South Africa. In year 2001 there were 58 registered banking firms engaged in the industry; however, in 2002 the number of firms has decreased to 45, eight banks has cancelled their licenses while one bank is in curator ship, two

[3] It must have at least 7 share holders and if any single private entity ownership exceeds 25% threshold it must notify to the regulatory bodies (Feasibility, 2001:9)

banks are in receivership and one bank is in liquidation. Ned Bank has taken over BOE, after BOE experienced initial difficulties.

Table 3.1: Number of banks in the industry

Bank type	As at 31 December					
	1994	1996	1998	2000	2001	2002
Registered banks	39	39	39	41	39	28
Mutual banks	2	3	4	3	2	2
Local branches	N/a	6	12	15	15	14
Foreign representatives	44	58	58	61	56	52
Total	N/a	106	113	120	112	96
Registration cancellation		4	7	3	4	9
Banks under curator ship	1	1			1	1
Bank in final liquidation	1			1	1	1
Banks under Receivership						2

Sources: South African reserve bank, Bank supervision department annual reports 1994,1996, 1998, 2000, 2001 and 2002.

The drastic decrease in the number of banks in 2002 can be attributed to the recent wave of bank runs in small banks, which was followed by subsequent consolidation as well as exit by small banks from the industry. The number of banks reached its peak in year 2000; 120 banks including branches and representative offices were operating in the industry. This has ascended from 85 banks in operation in 1994 (with exception of branches). The number of registered banks reached its peak in 2000 to 41 which by the end of year 2002 fallen to 28, the trend has continued to the first quarter 2003. The number of representative offices has decreased from the peak 61 to 52.

3.1.1 Foreign Banks[4]: Entry & Liberalization

The internalization of banking sector has become a recent global phenomenon spurred by liberalization of financial markets in cross-section of countries. In South Africa, the foreign bank entry has increased after the 1994 democratic election. According to Competition commission (2001) the foreign banks hold 4.3% of total bank assets or 31 billion of assets as at 2001. In the year of 2002, they made 14 branches and 54 representative offices as well foreign controlled locally registered banks.

The entry of foreign banks can come as de novo foreign entry or entry through acquisition of domestic entities (Claessens *et al,* 2001:3). In our case, the operation of foreign banks[5] is limited to niche markets, which include government institutions, corporate and wealthy individuals. Price Water House Coopers, (2003) survey indicates that in the future, foreign banks will take Treasury, Foreign Exchange, Capital Markets, Commercial Lending and Structured Products as of high importance product markets.

According to Feasibility (2001), in the South African banking industry there is no policy restriction regarding new entry of foreign firms. However, they are subjected to certain regulations and criteria. The South African Banks Act 1990 and its subsequent amendments (1994,2003) sets the requirement for foreign banks' operation in South Africa, foreign institutions can operate in the republic of South Africa as representative offices, branch, subsidiaries and can be incorporated in South Africa as South African entity.

According to the amendment to Banks Act 2003, a foreign[6] institution, which intends to open a branch bank in the republic, should at least hold USD1billion of net assets at the beginning as well as during the operation of the branch bank. The branch should

[4] See Claessens *et al*, 2001 for detailed account on the competition effect as well as other advantages of foreign entry into domestic banks.

[5] CITI Bank (American bank) had shown an interest to take over BOE, after the later shown a liquidity strain in 2002.
[6] See http://www.alifferdekker.co.za/litreature/regulation.forigne.html. For further information

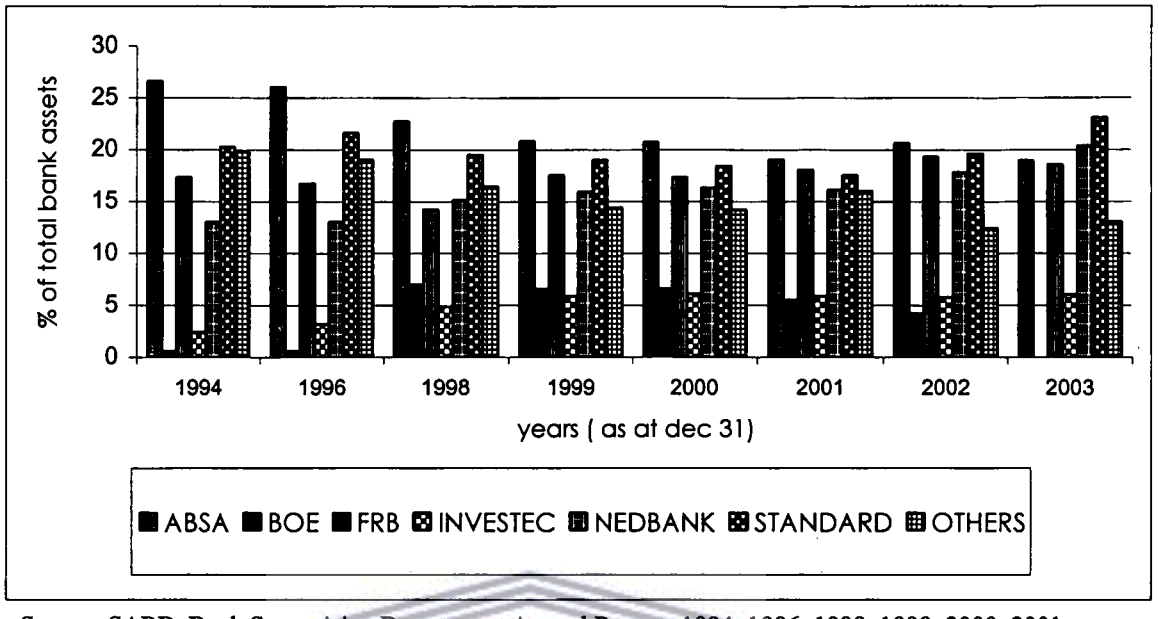
set its endowment capital a minimum of R250 million or 8% percent of assets and other risk capital. The banking survey conducted by Price Water House Cooper (2003) indicates that the majority of the respondents of the survey participants consider capital requirements, compliance and regulations as a primary cause of the exit of foreign banks as well as cutting of foreign banks' operation in South Africa. According Feasibility (2001) Branch banks are considered in South Africa as entities to be regulated on their own right. The independent capitalization, the net open position of foreign branch banks, double regulation (the parent company is regulated at home country), and the foreign exchange controls have detrimental effect on the entry of foreign banks into South African market.

Feasibility (2001) states that branch offices and subsidiaries can operate fully as banks without any restrictions while representative offices may only play a facilitating and marketing role, and may not accept deposits. There is no foreign subsidiary while there are 14 branches banks South Africa. This is due to the mere fact that international rating agencies rate branch banks as parent banks while subsidiaries are rated as South African entities, which happen to be rated lower than some of foreign banks.

3.1.2 Asset Distribution

The total asset of the South African banking industry as of June 2003 was Rand 1,348,113 million. As the figure below indicates, the top five firms that is Standard Bank, Ned bank limited, ABSA, FRB and Investec account for the 86% of the total assets. While back in December of year 2000 the five top banks held only 79% of the total assets.

Graph 3.1: Asset distribution of the South African banking firms [7](1994-2003)



Source: SARB, Bank Supervision Department Annual Reports 1994, 1996, 1998, 1999, 2000, 2001, 2002, and DI 900 Returns.

The above asset distribution shows dynamics in assets distribution. For instance, in 1994 ABSA had the largest asset in the industry. Its share of assets has been declining as shares of assets for Ned Bank has been rising steadily. As of 2003 Standard Bank has the largest assets base followed by Ned Bank.

3.1.3 Market Structure of The South African Banking Industry

In order to understand the current state of the South African banking, it is necessary to look at the current market structure as well as the change of the market structure. Market structure is appraised by market concentration. Market concentration is measured by the concentration ratios, such as the CR_4 & CR_5, and Herfindahl-Hirschman Index (HHI). A concentration ratio is defined as CR_N the combined share of the N largest firms in the relevant market. The basic problem with concentration ratio is that it describes only one point on the entire distribution of sellers. It does not give comprehensive information on the level of competition within the industry. While

[7] All the data are as of 31 December (end of the year). For the year 2003, the data is taken from DI 900 returns -line 224.

the concept and computing of HHI is quite different, it takes all the firms in the industry into account not just the largest firm.

These diverse market concentration measures have been used to assess the market structure of a particular industry. The robustness and comprehensiveness of these measures vary along the continua. However, there exists a precondition for the quality of measurement, robust measurement needs a substantial amount of data. For instance, Okealaham (2001) claims that it would have been of a better quality measurement if HHI could be estimated, but due to data limitation; he had to resort to other types of measurement.

It is always difficult to settle to one choice of measurement of concentration, assuming that it is plausible to estimate it. In making choice of concentration ratios, the main criteria is the suitability to the issue at hand, it is important to note that each measure highlights different dimension of the phenomenon and shades light on different aspects or issues. Ideally the choice of a measure should depend on the question that is being analyzed and should not be decided in isolation or on a priori grounds (Faurie, 1996:15).

It is important to bear in mind that concentration index is exclusively concerned with actual competition and ignores potential competition. For this reason a concentration index cannot fully assess the competitiveness of a particular industry. Some other information will have to be provided so as to take into account of the degree of potential competition. The concentration measurements remains popular albeit their static nature.

3.1.3.1 Concentration Ratios in the South African Banking Industry

The concentration ratio is a measure of industry output (or other relevant measure of economic activity) accounted for by a given number of large firms (Competition

Commission, 2002:12). The analysis of market share and concentration ratio for the purposes of competition analysis involves the delimitation of the markets. Usually such market delimitation is done based on product market and Geographic markets. This study follows the suit of European competition authorities, which categorizes the market in three broad service categories.

- Retail banking
- Corporate banking
- Merchant and investment banking

The next step will be to identify different products, which fall in each broad service categories. However the incompatibility of this kind of categorization with the available data compels us to resort to broad product markets. This is in line with competition commission (2001) methodology. We have selected a sample of five Broad product markets

- Deposit market
- Credit card market
- Investment market
- Installments sales market

Deposit Market

It is a single important product market in the banking industry. The main business of banking involves deposit-taking operation. According to Competition Commission (2000) this product market includes a number of specialized facilities such as call accounts, investments accounts and fixed deposits. The relative market share of the top player in this market is as fellows

Table 3.2: Market shares in the deposit market

RANK	BANK	IN THOUSANDS	MARKET SHARE IN %
1	ABSA	188,217,102	22.69
2	NED bank	169,080,361	20.39
3	Standard bank	154,460,846	18.62
4	FRB	150,589,513	18.16
5	Investec	51,989,072	6.2
	Others	114.823353	13.84
	Total	829,160,247	100
	CR_4	662,347,822	79.88
	CR_5	714,336,894	86.15

Source: SARB, DI900 June2003

As of June 2003 the CR_4 in this product market is 79.88 % of the total deposit in the industry. In other words, this means the largest four firms hold 79.88 % of the total deposit in the industry. The CR_5 is estimated to account 86.15% of the total deposits in the industry.

Installment sales market[8]

This category includes all installment sales; this includes vehicles, equipment and machinery.

Table 3.3: Market share in the installment sales market

RANK	BANK	IN THOUSANDS	MARKET SHARE IN %
1	FRB	24,007,826	28.84
2	ABSA	21,338,258	25.63
3	Standard bank	17,808,263	21.39
4	NED bank	11,382,781	13.67
5	Investec	1,410,326	1.6
	Others	7,288,138	8.7
	Total	83,235,592	100
	CR_4	74,537,128	89.54
	CR_5	51,939,628	91.2

Source: SARB, DI900 DATA SET June2003

[8] DI900, this item is reported in line 114 under bank liabilities

FRB has the largest market share in this product market followed by ABSA. The big four makes 89.5 percent of the whole market. The CR_5 is 91.2%.

Credit Card Market

They are credit instrument with a limit of specific Rand value depending on the type of the credit card.

Table 3.4: Market share in the credit card market

RANK	BANKS	IN THOUSANDS	MARKET SHARE IN %
1	Standard bank	4,260,249	26.88
2	ABSA	3,939,256	24.8
3	FRB	3,573,577	22.55
4	Ned Bank	3,411,975	21.5
5	Investec	552,158	3.4
6	Others	108,176	0.7
	Total	15,845,391	100
	CR_4	15,185,057	95
	CR_5	15,737,215	99.3

Source: SARB, DI900 DATA SET June2003

In this product market the largest four firms account for 95% of the whole product market, and the top five firms account for 99% of the whole credit card product market. While the remaining firms market share is limited to 5% and 1% respectively. This product market is the most concentrated product market.

Investment

We have followed the suit of the Competition Commission (2001) in identifying the products, which fall into this product market. This product market consists of

- Securities issued by the public enterprises/corporations

- Securities of local authorities and regional services council
- Securities of other public sector bodies
- Other subsidiary companies and associates
- Other companies, and
- Other investments (derivative instrument and banks and other companies).

Table 3.5: Market share in the investment market

RANK	BANK	IN THOUSANDS	MARKET SHARE (%)
1	Standard	73835757	32.9
2	Ned bank	40077043	17.8
3	ABSA	32714871	14.5
4	FRB	28271598	12.5
5	Investec	18853886	8
	Others	30635087	15
	Total	224388242	100
	CR_4	174899269	77.7
	CR_5	193753155	85

Source: SARB, DI900.DATA SET

In this product market the market share of the four largest companies is 77.7% of the total market for investment products. And the CR_5 in this market is 85%. The rest of the firms in the product market will be covering the remaining 15% of the market. Unlike retail banking, this product market does not require extensive establishments; however this product market is still very concentrated, albeit it is relatively lower than in other product market.

3.1.3.2 Herfindahl-Hirschman Index (HHI)

HHI is a more comprehensive measurement. It takes into account the market share of all relevant firms in the industry. It is basically the sum of the squared market shares of all firms in the industry, that is

$$HHI = \sum_{i=1}^{n} Ms_i^2 \quad - \quad - \quad - \quad - \quad - \quad (1)$$

Where n is a number of firms and Ms_i is the market share of the i^{th} ranked firms

As shown above, it is a summation of squared markets shares of individual firms. HHI ranges Between 1 for a monopoly and $1/n$ where n is the number of firms. It can be expressed in percentage-points or decimal points. If the market share is expressed in percentage points then the index will range from close to zero (when many firms in the industry) to 10,000 for monopoly and the index will range from close to 0 to 1 in decimal points. According to competition commission (2002:12)

> . . .HHI differs from concentration ratios in that it takes all the firms in the industry into account not just the largest firms. Unlike the concentration ratios the HHI reflects both the distribution of market shares of the top firms and the composition of the market outside the largest firms. In addition, the HHI has a strong point, which means it particularly useful in competition analysis, it reflects market dominance appropriately with a very high HHI values.

Benchmark

The benchmark used in utilizing concentration ratios and HHI, for competition analysis purposes by competition authorities, can differ between countries. But we can take the American benchmark used by the Department of justice as a tool in scrutinizing the merger cases.

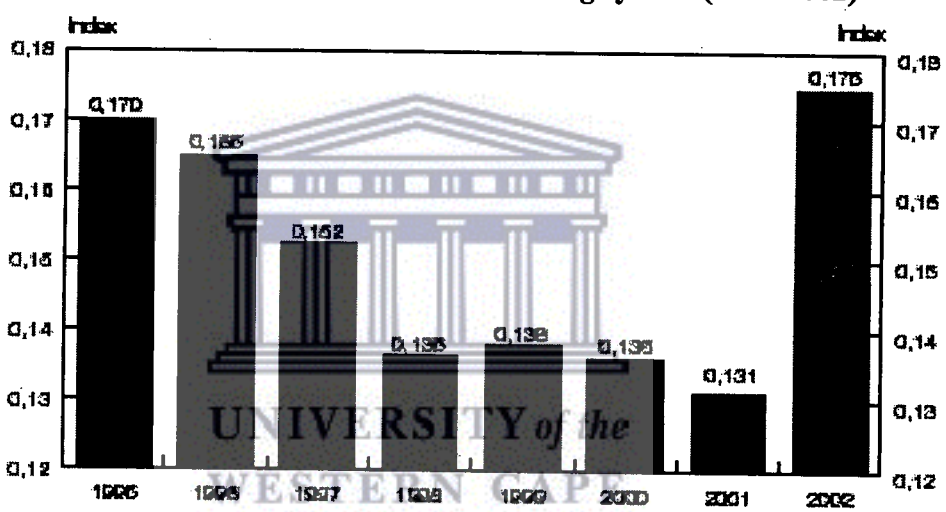
- ✓ Below 1000 points / 0.1 /
  If the HHI is below this point, then it considered as un-concentrated market
- ✓ Between 1000 and 1800 points /0.1 to 0.18/

If the HHI falls within the range, then the market is considered as moderately concentrated. Any merger that occurs within this range and brings more than 100 points (0.01) increase in the HHI is met by a serious competition concern

✓ Above 1000 points / >0.18/

This range is considered to be highly concentrated (this is approximately equivalent to 5 firms of equal size). Any merger that happens in this category is met by serious competition concern unless it brings an increase of less than 50 points (0.005).

Graph: 3.2 HHI for the South African banking system (1995-2002)



Source: South African Reserve Bank, Banks supervision department annual report 2002

As discussed above, the measurements should be compared against some sort of benchmark to give us an insight of the competition in the industry. Competition authorities usually set a HHI threshold, for purposes merger and acquisition cases, against which the post merger HHI is compared. The South African competition commission used the HHI in the impact assessment of the proposed merger of Nedcor and Stanibc in year 2000. The Commission compared post merger HHI with the thresholds.

The HHI in the South African banking industry has been decreasing steadily from 1995 to 2001. This fact could be due to the entry of a number of local and foreign firms into industry in post the 1994 democratic election. The recent hike of the HHI in the South African banking industry is induced by the failure of small banks, which culminated with the exit of a number of banks from the industry and the merger and acquisition followed, i.e. Ned bank took over BOE.

An international comparison of the concentration index, the South African Banking industry has higher HHI as at 2001 than the average G-8 countries. Furthermore the recent bank runs has resulted in a very high HHI. South Africa is among the countries that have the most concentrated banking industry.

3.2. The Bank Regulation

The primary reasons for bank regulation are to ensure safety and soundness of the banking system. This involves prudential regulations, which sets conditions and requirements to be met by individual banks. The authority to regulate banks is vested on the registrar of banks that conducts its operation through banking supervision department of reserve bank. The registrar of banks issues licenses to banking firms based on conditions set by Banking Act, 1990 and its subsequent amendments. Thus it requires the banks to pay license fees, submit a detailed business plan and comply with the regulation and prudent requirements set by Banks Act.

The registrar of banks has been revising the Bank's Act, 1990 to ensure its compatibility with international and local developments. The registrar has been conducting the amendment of the regulatory framework, the supervisory process and procedures, as well as the development of complementary tools; in order to cater for the requirements of the Basel II capital accord (South African Reserve Bank 2003:17). The registrar undertakes monitoring and surveillance operations of individual banks to ensure the soundness of the industry.

The 1997 Asian financial crisis is a good example as to how the crisis of a banking system could disrupt the over whole economic activity of a country. Thus the primary goal of banking regulation is to prevent a large-scale bank failure that would adversely affect the level of economic activity (Fraser *et al,* 2001:30). The depositors, which are creditors to the banks, are always of a large number; and they need an apparatus to monitor the activities of the banks in order to protect their interest. Hence in absence of such an apparatus the government assumes the responsibility to protect the interest of the depositors.

3.2.1 Prudential Regulation

Activities of the bank regulating authority differ from one country to another. However, basically it involves chartering, regulating and supervising banking firms. The authorities set prudent requirements and assess the safety and soundness of the bank by examining a number of indicators. The major risks which banks face are credit risk, liquidity risk, market risk, interest rate risk and currency risk.

Credit risk[9] is the single and most significant factor for bank failure. According to Fraser *et al.* (2001), credit risk occurs when a debtor fails to meet the terms of any contract with the bank. Bad loans and NPL's deplete the bank's capital and reserves. The capital adequacy requirement is a safety net that is used to replenish the reduction in assets due to the default. In December 2002 Net overdue as a percentage of net qualifying capital and reserves reached 13,2 % that has improved from 17,3 % in December 2001. This figure is well below the international benchmark of 25 percent. According Feasibility (2001) the banks supervision department has set the minimum capital adequacy requirement of 250 million or 8% of their risk weighted assets. This is consistent with the Basel Accord of 1988. The adequacy requirement is likely to be increased in line with BIS revision in 2002. The South African banks are well capitalized the capital adequacy ratio of 12.6% for the year 2002.

[9] Credit risk is measured as the ratio of net overdue (that is, gross overdue less specific provisions) to net qualifying capital and reserves is used internationally to benchmark the extent of amounts overdue in a banking sector.

Liquidity risk: It lays at the heart of confidence in the banking system. It represents the bank ability to meet its financial obligations. It is the availability of liquid assets that might be used to extend loans and deposit withdrawals. The importance of liquidity goes beyond the individual institution; it has spillover effect on other banks as well .The bank supervision department has set a statutory liquid asset requirement. This requirement is calculated on total liabilities less capital and reserves (referred to as adjusted liabilities). South African banks were operating well above the requirement; however during the second half of 2001 the banks had fallen short of the requirement.

Market risk: It is a situation when a bank experiences a loss in on-or-off balance sheet positions arising from unfavorable movements in market prices. It results from changes in prices of equity instruments, commodities, money and currencies. Its major components are therefore equity position risk, commodities risk, as well as interest rate risk. The increasing exposure of banks to market risk is due to the trend of business diversification from the traditional intermediation function towards trading and investment in financial products that provide better potential for capital gain which exposes banks to significantly higher returns (Greuning and Bratanovic, 2000:21).

Interest rate risk: It is risk of interest rate volatility. Banks are porn to interest rate fluctuations. The market value of bank assets and liability depends on the interest rate. For example, the market value of bank assets will increase with the increase in interest rates. Furthermore earnings from assets, fees and the cost of borrowed funds are affected by changes in interest rate.

Currency risk: Currency risk results from changes in exchange rates between banks domestic currency and other currencies. The supervisory body has set net open position limit on the banks, which is an aggregate limit of bank's currency risk

exposure. The net foreign exchange open position (NOP) of commercial bank is the difference between its total foreign exchange assets and total foreign exchange liabilities at the end of each day, after adjusting for all foreign exchange transactions done during the day. It is used as one of the prudential requirements for purposes of bank supervision. The supervisory body set a limit on NOP because a large net open position may subject banks to excessive risk due to exchange rate volatility. In our case, Feasibility (2001) notes the prudential limit in NOP is set not to exceed 10% of its capital and reserves (It is recently tightened 15% in 2001).

3.2.2 The South African Bank Failures

There are four major reasons for bank failure; that is, credit risk, market risk, bank runs, and fraud (Fraser *et al,* 2001:33). We have already discussed these risks above. However, in relation to the recent South African bank crises, which was caused mainly by a liquidity strain. Thus it would be interesting to describe the bank runs phenomena.

Bank runs (Liquidity Strain) A run on a bank occurs when a large number of depositors, fearing that their bank will be unable to repay their deposits in full and on time, try to withdraw their funds immediately. This creates a problem because banks keep only a small fraction of deposits on hand in cash; they lend out the majority of deposits to borrowers or use the funds to purchase other interest-bearing assets like government securities. Bank run always reflects herd behavior of depositors to obtain a limited amount of cash on hand. Sometimes a "silent run" might occur when large creditors, such as banks and investment companies, withdraw their funds due to one to many reasons. Usually a bank run might induce a contagion effect whereby other banks might fall into the same liquidity strain.

Against the above theoretical background the bank failures unfolded in the last two years are mostly bank runs. As bank supervision department noted within the period from the beginning of first quarter 2002 until the end of the first quarter 2003, 22

banking firms exited the industry. Ned Bank took over BOE after the later experienced initial liquidity strain. The onset of bank failures began with imposition of curator ship over the FBC Fidelity Bank limited (FBC) in the last quarter of 1999. This ensued a gradual loss of depositor's confidence in small to medium sized banks (SARB, 2002:). The crisis gained momentum, fuelled by the contagion effect, in the year of 2002 where 22 banks exited from the industry.

Table 3.6: The number of banks that exited from the South African banking industry

	As at 31 December					
	1994	1996	1998	2000	2001	2002
Registration cancellation		4	7	3	4	9
Banks under curator ship	1			-	1	1
Bank in final liquidation	1			1	1	1
Banks under Receivership				-	-	2

Sources: South African reserve bank, Bank supervision department annual reports 1994, 1996, 1998, 2000, 2001 and 2002.

ABSA (2001) identifies predicaments which small banks faced before the crisis unfolded

- Small banks had to relay on expensive wholesale deposits.
- Small banks were often more prepared to take risks and venture into new markets, which bigger banks tend to avoid. Thus these banks took micro-finance as one of their main product lines. The micro lending is highly risky line of business.
- Poor domestic saving ratios, that is, 0.3% of the personal disposable income. This entails small banks were forced to pay higher interest rate to attract retail deposits.
- Concentration of market power in South African money market. The big institutional portfolio investors such as Old Mutual and Sanlam are either stockholders or involved in some arrangement with major banks. The

possibility for these institutions to deposit on small banks is unlikely. They face competitive disadvantage versus big banks.

The Failure of Saambou and BOE: scenarios

Saambou bank was the eightieth largest bank in South Africa with deposits around Rand 16 billion. It had 52 branches throughout the country. The failure of this bank can be explained as cumulative effect of the above theoretical background. However, the principal factor is loss of confidence in the bank. The scenarios preceding the crises are:

- ✓ After the failure of FBC, there was a loss of confidence in small banks
- ✓ The Fed Sure limited (FED SUR)[10] that holds the 47% of Saambous' shares was also the largest shareholder of FBC. This had an indirect pressure on Saambou.
- ✓ Speculation of under performance was triggered by the sale of Saambou's shares by its executive directors.
- ✓ The subsequent profit warning issued by the company took the pressure to a higher level. The bank under performed due to the loss incurred in its Micro finance business and internet venture
- ✓ The announcement by the Uni bank limited (Uni bank) that disclosed the bank has depleted its capital base due to loss sustained in its micro finance business.
- ✓ The Rating agency Fitch placed Saambou on a negative rate watch, while Auditors ascertained the solvency of the bank.

The public confidence on the bank was eroded and depositors began withdrawing their deposit in order to secure them. When the wholesale depositors began to withdraw their deposits the bank run phenomenon set in, in two days time, depositors withdrew 1 billion Rand from the bank. The bank ran into liquidity strain.

BOE:

The contagion set-in after the Saambou was put to curator ship. A number of small banks excited from the industry. BOE, the sixth largest bank in South Africa with assets of R67 billion and 1,4 million depositors; started to experience a bank run. 14 billion Rand deposits were withdrawn by the march 2002, mainly by wholesale depositors. It began to experience a liquidity strains. This time, however, authorities extended their assistance to the ailing bank. Initially Reserve Bank extended liquidity assistance but this could not stop the bank run. Then the national treasury issued a guarantee to all depositors. This had stopped wave of withdrawals. NEDBANK took over BOE.

3.2.3 Consolidation of The South African Banks

Consolidation in banking industry is a global phenomenon; recently we have witnessed a wave of mergers and acquisitions in banking industries across countries spurred by development in information technology, globalization, and financial and capital liberalization. The wave of consolidation in the South African banking industry began when ABSA, Allied Bank, Trust Bank and Bankrop merged in early1990's. The momentum reached its peak on 2002, spurred by the failure of small banks, which were subsequently swallowed by the big four.

In the year 2000, the aggressive bid by Nedcor to takeover Stanbic was stopped by the ministry of finance, on recommendation from Competition Commission. The commission, on its report (2000), made an extensive assessment of competition effect of proposed takeover. It looked at the high level of concentration in the three broad product markets with a serious concern and concluded the proposed merger could give Nedcor a market power, which could lessen the competition in the industry. This could ultimately affect pricing, employment, and accessibility of banking services to the society.

[10] Fed sure was taken over by Investec in 2001 that made Investec a principal shareholder in Saambou

According Ernest & Young (2002), the year 2002 witnessed a large number of mergers and acquisition in history of the South African banking industry. The major merger and acquisitions, which took place in 2002, were: BOE after experiencing initial liquidity strain was taken over by Nedcor for R 8.1billion of mix of cash and shares. This friendly bid, which is seen as a rescue operation, was given a green light by the Ministry of Finance to go ahead. First Rand bought certain assets From Saambou. Initially First Rand and Nedcor went to bid for Saambou, but Nedcor withdrew from the bid ones BOE became available. Thus First Rand bought the major chunk of Saambou's assets. R2.8 billion Saambous micro loans book was sold to African Bank Investments. While BOE's NBS home loans was sold to First Rand. ABSA acquired 61% of PSG investment Bank (PSGI) after its initial attempt to acquire the group. The PSGI remained as a subsidiary of ABSA.

Conclusion

The South African banking industry is a large and high-tech service industry, which makes a significant contribution to the South African economy. Its contribution is multifaceted. It has a significant employment potential, and its contribution in the gross domestic product is substantial, however its main contribution comes from the efficient intermediation of financial resources in the economy. Yet it is characterized by high level of market concentration, which potentially can hamper the level of competition in the industry. The regulatory environment in the industry has been in line with international practice, and the authorities as well as the banks are working towards the adoption of the Basel II Capital Accord. The recent failure of small banks, which was about to trigger a contagion effect, had a serious ramification in the economy.

Chapter IV

THE MODEL, METHODOLOGY AND DATA SET

The model used in this research is the P-R reduced from revenue function. To apply this model to the South African banking industry, there are two methodological options to choose from cross sectional data analysis and panel data analysis. The considerable merit of the panel data analysis and data limitations (smaller number of Cross sections) makes the use panel data analysis preferable. The data set has to be unbalanced panel, so that we may be able to benefit from larger data set.

4.1 The Model

The model used to estimate the competitive conduct of the South African banking industry is P-R reduced form revenue function. It is a new empirical industrial organization theory (NEIO), which measures the level of competition in the industry irrespective of the market structure. P-R function estimates the factor price elasticity. It gauges the degree with which the change in factor prices is reflected in revenues. Based on the estimated elasticities, it concludes the level of competition prevailing in the industry.

4.1.1 The Panzar–Rosse approach (1987)

The P-R measures the competitive behavior of firms on bases of comparative static properties of the reduced form revenue function, it estimates the factor input price elasticities, and infers the competitive conduct based on it. This approach assumes:

- The firm produces a single output.
- Firms cost functions are linearly homogenous in the factor prices or the production function is homothetic.
- Firms operate in their long run equilibrium
- Factor prices are exogenous to the individual firm, and
- Performance of these firms needs to be influenced by actions of other market participants.

Revenue function: revenue is a function of a vector of decision variables (x_i) [11] which affect revenue and a vector of exogenous variables (z_i) that shifts the Bank's revenue function, the revenue function for bank *i*

I.e. $R = R(x_i, z_i)$ . . . . . (1)

Cost function: it is a function of a vector of decision variables (x_i), and m_t is a vector of *k* exogenous factor input prices for bank *i,* and t_i is vector of exogenous variables that shift the bank's cost function.

I.e. $C = C(x_i, m_i, t_i)$ . . . . . (2)

The vectors *t* and *z*, which represents the variables, which shift the cost and the revenue functions respectively may or may not have components in common.

Thus the profit function can be written

$\Pi = R - C = \Pi(x_i, z_i, m_i, t_i)$ . . . . . (3)

Profit is maximized when

Max $\Pi = r(x_i, z_i, m_i, t_i) = 0$. . . . . (4)

The Panzar-Rosse (P-R) put the reduced revenue function as

$R^*(z, t, m)$ . . . . . (5)

Conversely, Profit is maximized when marginal cost is equal to marginal revenue

$MR(x_i, z_i) = MC(x_i, m_i, t_i)$ . . . . . (6)

At equilibrium the zero profit constraint holds at market level

$R^*(x^*, z) - C^*(x^*, m, t) = 0$ .. . . . . (7)

The decision variable (x_i) does not only represent the out put levels but also it includes prices, advertising expenditure and quality; in other words it represents any variable which structurally enters the firm's revenue function.

The (*) represents the equilibrium value

The market power is measured by the extent to which a change in factor input prices (∂m_{n_i}) is reflected in equilibrium revenue (∂R_i^*) earned by bank *i*, P-R define measure of competition, the "H-statistic", as the sum of elasticity's of the reduced from revenue function with respect to factor prices.

$$H = \sum_{n=1}^{k} \frac{\partial R_i^*}{\partial m_{n_i}} \frac{m_{n_i}}{R_i^*} \quad \ldots \ldots \ldots \ldots \ldots \quad (8)$$

Bresnahan (1989) has made two critical observations for the methodology to give convincing results. Firstly, all factors that shift cost function must be correctly specified and entered to the equation. Secondly, the shifts in the cost function are unlikely to fully identify the degree of market power in any industry. This is because the market power coefficient is a part of supply relationship and its identification requires more information from the demand side of the market.

4.1.2 Interpretation

The interpretation of the H-statistic[12] according to P-R is summarized in the table below.

Table 4.1 Interpretation of Panzar-Rosse "H-statistic"

VALUES OF H	COMPETITIVE ENVIRONMENT
H< 0	Monopoly or perfectly collusive oligopoly
0< H< 1	Monopolistic competition
H=1	Perfect competition, natural monopoly in a perfectly contestable market, or sales maximizing firm subject to break even constraint

In the extreme monopoly case, the sum of elasticities of firms reduced form revenue with respect of factor prices is less than zero. The increase in input prices will shift all

[12] See Panzar and Rosse (1987), for detailed account and proof on the estimation and interpretation of H-statistic,

cost curves. Therefore it will increase equilibrium output price and reduce equilibrium quantity output, since the monopolist's faces an elastic demand curve, the total revenue decreases.

In the other extreme case of perfect competition, the sum of factor price elasticities is equal to unity. This means when banks operate at their long run equilibrium, a proportional increase in factor prices induces an equi-proportional change in gross revenues; output does not change in volume terms, while the output prices rises by the same extent as the input prices (De Bandt and Davis, 2000:1048). The intuition behind this result is that a one-percent increase in all factor prices will result; because average cost is homogeneous, in a shift of the average cost curve up by one percent for all output levels. Consequently, the minimum point remains unchanged. In the long run, the competitive equilibrium the firm operates at minimum average cost and the competitive output remains unchanged. However, in equilibrium, the competitive price must be equal to minimum average cost, which has increased by one percent. Therefore, price must have increased by one percent, driving up total revenues by the same percentage.

In between these two extreme conditions lays the monopolistic competition. This is more realistic environment in the analysis of banks. It is characterized by product differentiation and excess capacity. Bikker & Haaf (2000:2) put it, "Monopolistic competition models are *a priori* most plausible for characterizing the interaction between banks. The monopolistic competition model recognizes the existence of product differentiation and is consistent with the observation that banks tend to differ with respect to various product quality variables and advertising, although their core business is fairly homogeneous". In the symmetric chamberlain equilibrium, the sum of elasticities of the firms reduced form revenues with respect to factor prices is less or equal to unity. The intuition, behind this preposition, assumes the elasticity of perceived demand facing the individual firm is a non-decreasing function of the number of rivals. As De Bandt and Davis assert "revenues will increase less than proportionally to changes in factor prices, as demand for banking products facing

individual banks are inelastic". Thus Panzar and Rosse (1987:449) has summarized the process as

> . . . immediate effect of an increase in input prices the upward shift of the firms average and marginal cost curves and reduce its output. However, this would result in loss for the representative firm, inducing exit by some firms, which would, in turn, result in an upward shift in the demand curve facing the representative firm until the tangency equilibrium was re established.

Thus the H-statistic, which is the summation of the factor price elasticity's for the factors of production, is between zero and one.

4.1.3 Limitation

Hempell (2002:9) has recognized several limitations in applying Panzar-Rosse model to the banking industry to assess bank's market conduct and interpreting the results. The limitations are regarding the scope of models captured by static approach, the assumption made, and the resulting biases when applying to the "real world" data.

a. Panzar-Rosse H-statistic was developed on the basis of static (oligopoly) model; it ignores the dynamic process of entry, exit, and innovation. However this model assumes long run equilibrium. Dynamic models question the existence of market equilibrium in the market overtime span. Thus this imposes further limitation in the model.

b. The other limitation is the assumption of perfect competition on input markets (banks as price takers on input markets). Generally, the assumption of perfect competition in input markets is problematic. It is less plausible to have perfectly competitive market for deposits. The second assumptions on input prices, that is the use of differing bank specific input prices might indicate the existence on imperfect factor markets; however, they may also be interpreted as the result of local (competitive) factor markets.

c. A downward bias in the estimated elasticity results from the maturity structure of banks' asset portfolios. As longer maturities in fixed rate contracts prevent

banks from direct price adjustments, even perfectly competitive markets delayed changes in pricing imply lower elasticties estimated.

4.2. Empirical Models[13]: a comparative analysis of applications to banking industry

In adapting an empirical model, let's briefly review the empirical models developed based on the initial Panzar-Rosse reduced form revenue function to assess the level of competitive conduct. Among the vast application of this model in different industries, we will be particularly interested in contributions made on banking industry. This model has been extensively used in banking industry to test competition in a number of countries. Let's take Molyneux *et al* (1994, 1995) as benchmark, and then we will review the main features of the model. .

In using the P-R in banking sector for purposes of measuring competition requires the assumptions made above to hold. The endogenous variable is the total revenue or total interest income the bank drives from its operation. The choice of the endogenous variable depends on the banking approaches. The bank makes use of a vector of factors of production in the process, i.e. labour, capital and deposits, and more interestingly are the exogenous and bank specific variables to consider. As we have mentioned above, for P-R model to give us unbiased estimates, we need to account for elements, which shift the cost and revenue functions.

Molyneux *et al* (1994,1995)

The reduced form revenue function for European banking industry was estimated as

The model:

$$\ln TRESS_{it} = a + b \ln PL_{it} + c \ln PK_{it} + d \ln PF_{it} + e \ln ASS_{it} + g \ln CAPASS_{it} + h \ln IBDEP_{it} + I \ln LNASS_{it}$$

For $t = 1, \ldots, T$, where T is the number of periods observed and $i = 1, \ldots, I$, where I is the total number of banks ; where

ln = natural logarithm

Dependent variable

$TRASS_{it}$ = Total interest revenue per dollar of assets.

Independent variables

PL_{it} = Personal expenses per dollar of assets (proxy for unit price of labour)

PK_{it} = Capital expenses per dollar of fixed assets (proxy for unit price of capital)

PF_{it} = Annual interest expenses to total funds (proxy for unit prices of funds)

Bank specific

ASS_{it} = Banks assets

$CAPASS_{it}$ = Total risk capital to assets ratio

$IBTDEP_{it}$ = Inter bank deposits to total deposits, and

$LNASS_{it}$ = Loans to assets ratio

The dependent variable in this application is interest revenue per dollar assets; the choice of the dependent variable is based on the banking approaches adopted. That is the intermediation approach versus production (traditional) approach of banking functions[14]. While the production approach uses total income as dependent variable instead of gross interest revenue. Molyneux *et al* (1994,1995), Bikker & Haaf (2000,2001), Gelos & Roldós (2002) opt for the intermediation approach thus they specified interest revenue as the dependent variable; while Nathan & Neave (1989),

[13] See appendix III for descriptive analysis

Hempell (2002) specified the total revenue as their dependent variable. The other difference in the specification of the endogenous variable is the algebraic representation –Molyneux (1994), Bikker & Haaf (2000,2001) specified the dependent variable as a ratio of interest revenue to total asset balance sheet while Nathan & Neave (1989) took the logarithm endogenous variable. (Vesala, 1995; cited in De Bandt & Davis, 2002:1049) observes – a ratio of interest revenues to assets provides a price equation, the log specification may also reduce possible simultaneity bias.

The independent variables are mainly a vector of input factor prices; this is composed of unit price of labour, unit price of capital and unit price of fund. Various studies on European banking industry have comparable definition on unit factor prices, the distinction lays on the computation of the factor prices, for example Molyneux (1994), Hempell (2002), and Bikker & Haaf (2000) compute the labour cost as personal expenses to total assets; while Nathan & Neave compute the unit labour cost total wage expenses to number of employees.

The other variables included are the bank-specific and exogenous factors, that may shift the cost and revenue functions; and of course, the scale measure. Diverse variables are considered depending on the availability of data and structure of the banking system under investigation. Molyneux *et al* (1994, 1995) includes the firm specific explanatory variables to take into account "other risks", cost and size characteristics. Since the dependent variable is not risk adjusted, the total risk capital to asset ratio taken to proxy firm specific risks; loans to asset ratio and the inter bank deposits to total deposits ratio to account for difference in deposit mix; total assets is included to account the scale factor. De Bandt and Davis (2000) use equity and fixed assets to as a measure of scale capacity and loans asset ratio, and deposits as a proportion of deposits plus money market liabilities to account for business mix (factors that may shift the revenue and cost schedule). Gelos and Roldós (2002) have included in their specification capacity indicators such as fixed assets, business mix

[14] See page 19

and the size of non-performing loans to account for factors affecting the interest revenues.

Bikker & Haaf (2001) argue bank-specific factors (*BSF*) should stem from the marginal revenue and cost functions underlying the empirical P-R equation. The BSF encompasses the risk, structure, deposit (business mix) and economics of scale. The risk component can be proxied by the ratio of risk capital or equity to total assets, the ratio of loans to total assets and the ratio of non-performing loans to total loans (*NPL*). More than one variable for risk is considered, as there are cases where one of these variables may be unavailable for a particular bank. The ratio of inter-bank deposits to total customers and short-term funding and the ratio of demand deposits from customers to total customer and short-term funding are used to capture differences in the deposit mix. Correspondent bank activities are taken into consideration when the ratio of cash and due from depository institutions (or banks) to total deposits is included. Total assets are used as a scaling factor.

Another feature of the applications is the econometric methodology utilized. The methodologies used to estimate the P-R model are the cross-sectional and panel data econometric analysis. The recent P-R models developed for European banking industry has used the Panel data econometric tools. Bikker & Haaf has used the panel data to assess the effect of Euro (European monetary policy union) on competition in European banking industry. Gelos & Roldós have used the panel data to assess the effect of bank consolidation on competition in emerging market economies. Recently, Panel data has been increasingly utilized in research of industrial organization.

4.3 The Specification of the Panzar-Rosse Reduced Form Revenue Function For the South African banking industry.

In extending the above model to the South African banking, we need to look at contending specifications and assumptions, the features of the South African banking

and the nature of the data being used to estimate the model. The variables used to estimate this model are the endogenous variables, exogenous variables and the bank specific variables. The specification of the model in the preceding studies took the logarithmic expression. This is inherent in Panzar-Rosse Model, because it aims to estimates the elasticity in the model.

The reduced form P-R model is specified as

$$\ln IR_{it} = C + \beta \ln pf_{it} + \alpha \ln pl_{it} + \delta \ln pk_{it} + \eta \ln ass_{it} + \theta \ln mix_{it} + \mu_{it} + \varepsilon_i$$

For $t = 1, \ldots, T$, where T is the number of periods observed and $i = 1, \ldots, I$, where I is the total number of banks; where

ln = natural logarithm

IR_{it} = ratio of total interest revenue to total assets for bank i at time t

C = constant

PF_{it} = unit price of fund: interest expenses to total fund for bank i at time t

PL_{it} = unit price of labor: staff wage to total assets for bank i at time t

PK_{it} = unit price of capital: depreciation and other capital expenditures to fixed assets for bank i at time t

Bank specific

Ass_{it} = total assets for bank i at time t

Mix_{it} = business mix: total deposits (fund) to assets for bank i at time t

Endogenous Variables / Dependent Variable

There are two contending approaches[15] over the preference of the dependent variable in stating the Panzar-Rosse model in the banking industry. The intermediation

[15] The "production approach" held banks as a firm using capital and labour to produce different products and services including loans, investments and other services. While the "intermediation approach' treats banks as financial intermediary.

approach considers banks primarily as financial intermediaries. Therefore, this approach prefers to use gross interest revenues as the dependent variable. Molyneux *et al* (1994) have chosen the dependent variable to be the ratio of total interest revenue to the total balance sheet. The choice for taking only interest rate part of the total revenue of banks is consistent with the underlying notion inherent on P-R model, that financial intermediation is the core business of the most banks (Bikker & Haaf, 2002:2196). The dependent variable for our model is the interest income to total assets, though it contributes only 58% to total revenue

Table 4.2: Composition of Gross Revenue for the South African Banking as at December 2001

COMPOSITION OF REVENUES	R MILLION	%
Interest income	7627	58 %
Transaction-based fee income	1559	11.4 %
Income from management of funds	2889	21.24 %
Investment and trading income	1253	9.2 %
Knowledge-based fee income	273	2 %
Total	13601	100%

Source: South African Reserve Bank, Banks Supervision Department annual report 2002

Nathan & Neave (1989) on Canada, Hempell (2002) on Germany embarked on total revenue as dependent variable. The rational behind their choice lays on production approach on the assumption though the principal task of a bank is financial intermediation; it is supplemented by fee income. De Bandt and Davis (2000:1049) argue the very fact that for banks in the competitive struggle for survival, the distinction between interest and non-interest revenue becomes less relevant, competition being equally vigorous for both. In this model the dependent variable is specified ratio interest revenue to total assets as in Molyneux *et al* (1994).

Exogenous variables

Factor inputs

It consists of three principal inputs used in production of banking services; these are capital (physical), labor, and funds (deposits).

- Personal expenses

  It covers all staff expenses that consist of wages, salary and benefits, incurred in providing banking service. There is not much difference in the definition of personal expenses, however there are apparent differences in the scaling factor. Shaffer (1982), and Nathan & Neave (1989) scale the total labor cost to total employees by the end of the budget year. Molyneux *et al* (1995), Bikker & Haaf (2002) scale personal expenses to dollar of assets. The availability of data is a major factor in deciding which way to go, the financial statements of banks usually does not report the number of employees. Bikker & Haaf (2002:2197) claim that empirical exercise using personal expenses to the number of employees closely approximates " ratio of personal expenses to balance sheet". In this model the variable will be scaled to total assets. In our case the staff expenses to assets is taken to proxy to personal expenses.

- Interest expenses

  The annual interest expenses incurred in production of banking services. The annual payment made on deposits, it is the main input in the banking industry. All pervious empirical studies have similar definitions for funding, the ratio of interest expenses to sum of deposits and other liabilities. The only difference is on the scaling factor, whether the dominator includes other entries or not. According to Hempell (2002) "other liabilities" may constitute deposits by banks, customer accounts, and certified liabilities; or it may include inter-bank time and demand deposits, long term borrowing subordinated debt, participating debt and hybrid capital. In our model interest paid to deposits is taken as a unit price for fund.

- Capital expenses

  It is the ratio of the physical capital and other expenses to fixed assets. It includes the cost of premises, equipment and information technology. It refers to the depreciation of fixed assets, and it is scaled to fixed asset.

Banks specific factors

These additional variables included in bank specific factors (BSF) are supposed to take into account the other risk, cost and size characteristics, and structure of banks.

- The ratio of non-interest income to total income. This variable will be included in order to take into account the shift of banks from interest revenue to non-interest fees. This variable represents the business mix of the bank. We have taken deposits total assets as a proxy to business mix.
- The capacity indicator is included in the estimation to reflect the economics of scale in the industry. Molyneux *et al* (1994), used banks assets to reflect the economies of scale. We took total assets to account for economics of scale.
- Since the dependent variable is not risk adjusted, it is important to include a variable, which takes into account the risk involved in the industry. This includes ratio of loans to total assets, and a ratio of non-performing loans to total assets.

4.4 Econometric Methodology

Industrial economists have used at least three methodologies to study what happens in the markets (Geroski and Mata, 2001:1000). It has evolved from the case study method of particular sectors to cross sectional methodology. This methodology goes away from the sector specific studies to the assessment of broad features, which many industries share. However this methodology is not without a drawback. It can face a

measurement difficulty of very significant variables. Thus, in some cases, it resorts to omitting very significant variables.

The third methodology is a panel data analysis, among of its advantages to enable to control some important unobservable variables by using fixed effects panel data analysis, the another advantage is the time series dimensions of such data enables to give insight to market dynamics. The fourth and relatively recent methodology is longitudinal case studies. This methodology makes use of almost the entire history of the particular market. The virtues of making use of a long time series data enables to make a solid inferences about market dynamics unlike panel data analysis which may reflect the features for a particular four or five year period of time.

Shaffer (1982), Nathan & Neave (1989), Molyneux *et al* (1994,1995) have used the cross sectional data analysis to estimate the Panzar-Rosse reduced revenue function. While De Bandt & Davis (2000), Bikker & Haaf (2002), and Gelos & Roldós (2002) have used panel data econometrics to estimate the reduced form revenue function for European banking industry.

4.4.1 Econometric Analysis of Panel Data

Panel data is a pooling of observations on cross section of households, countries, firms etc. over several periods of time (Baltagi, 2002:1). This data set, contains a repeated observations of the same unites collected over a number of periods, allows us to specify and estimate more complicated and more realistic models than a single cross section or a single time series would do.

A panel data regression differs from a regular time-series or cross section regression in that it has a double subscript in its variables, i.e.

$$Y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \qquad i = 1, \ldots., N; t = 1, \ldots., T$$

With i denoting firms, and t denoting time. The i subscript, therefore, denotes the cross section dimension whereas t denotes the time series dimension. Most panel data applications utilize a one-way error component model for the disturbances,

$$\varepsilon_{it} = \alpha_i + \varepsilon_{it}$$

α_i denotes the unobservable individual specific effect and ε_{it} denotes the remained disturbance. For example, in the P-R reduced form revenue function Y_{it} measures interest revenue of banks, whereas X_{it} contains the factor inputs. Note that α_i is time-invariant and accounts for any individual specific effect that is not included in the regression, in this case managerial quality can be taken as unobservable variable. The remained disturbance ε_{it} varies with individuals and time, and can be thought of as the usual disturbance in the regression.

The pros and cons of the panel data econometrics analysis can be summarized as

Pros:

- Dynamic of economic process or decision. The dynamic economic behavior can be comfortably estimated using the panel data. For example, the dynamics of entry and exit of firms in an industry can be estimated by panel data.
- Reducing the estimation bias.
- Efficiency of parameters: Panel data are more likely to be efficient than any other data.
- Identification and discrimination between competing hypotheses. Panel data are not only suitable to model or explain why individual units behave differently at different time periods. Panel data suggest that individuals are heterogeneous and control for individual and time specific heterogeneity.

Cons:

- Attrition: some individual may be dropped due to a number of reason thus it can result to unbalanced panel data series. Unbalanced panel data occurs when the time periods for different observations is not the same.
- Heterogeneity bias

- Selectivity bias: this type of problem emanates from the technique of sampling. If samples are not randomly drawn from the population then there will be such a bias.

4.4.1.1 Fixed Effects

The fixed effects model is simply a regression model in which the intercept term vary over the individual unites, i.e.

$$Y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \quad . \quad . \quad . \quad . \quad . \quad \varepsilon_{it} \sim \text{iid } (0, \delta^2_{\varepsilon})$$

The time-invariant α_i error term is assumed to be fixed parameter to be estimated, and the remained disturbances stochastic with ε_{it} independently and identically distributed.

This can be written in usual regression form including dummy variable for each unit *i* in the model

$$y_{it} = \sum_{j=i}^{N} \alpha_j d_{ij} + x'_{it}\beta + \varepsilon_{it} \qquad \text{Where } d_{ji}=1 \text{ if } i = j, \text{ and it is 0 if } i \neq j$$

We have *n* dummy variables in the model, the parameters $\alpha_1, \alpha_2, \ldots\ldots, \alpha_n$ and the β can be estimated by ordinary list squares (OLS). The β estimator is called **least squares dummy variable estimator**. However, it is numerically unattractive to estimate a number of regressors, fortunately there are simpler way to compute the estimator β. It can be shown that exactly the same estimator β is obtained if the regression is performed in deviation from individual means. This implies we eliminate the individual effects α_i by transforming the data

The simple regression:

$$Y_{it} = \alpha_i + \beta x_{it} + \epsilon_{it} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The averaging over time gives us

$$y^*_{it} = \alpha_i + \beta x^*_{it} + \varepsilon^*_i \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Subtracting (5) from (4)

$$y_{it} - y^*_t = \left(x_{it} - x^*_i\right)\beta + \left(\varepsilon_{it} - \varepsilon^*_i\right) \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This is a regression model in deviation from individual means and does not include the individual effects α. This transformation, which produces observations in deviation from individual means, is called within transformation. The estimator β obtained from the transformed model is called within estimator or fixed effects estimator.

Properties of fixed effects

- Assuming that all x_{it} are independent of all ε_{it}. The fixed effects estimator β is unbiased. We assume the normality ε_{it} then, the fixed effects estimator β is normal.
- For the estimator to be consistent, it is required E $\{(x_{it} - x^*_i)\varepsilon_{it}\}=0$, it is sufficient for this that x_{it} is uncorrelated with ε_{it} and that x_i has no correlation with the error term. This in turn implies that E $\{x_{it}\ \varepsilon_{is}\}= 0$ for all *s, t*, this means x_{it} strictly exogenous.
- Fixed effects estimator of intercept α_i, with the explanatory variables independent of all errors, is consistent provided that T goes to infinity. Otherwise α is inconsistent if the T is fixed, even when the number of individual increases y_i and x_i do not converge to anything.
- Under weak regularity conditions, the fixed effects estimator is asymptotically normal, so that the usual inference procedure can be used (like *t* and wald tests).

Note: when interpreting the results from fixed effects regression, it may be important to realize that the parameters are identified through the within dimension of the data. The fixed effect model concentrates on difference "within" individuals. That is, it is explaining to what extent Y_{it} differs from Y^*_{it} and does not explain why Y^*_{it} from Y^*_j. The parametric assumptions about β on the other hand, impose that change in x has the same (ceteris paribus) effect, whether it is a change from one period to the other or a change from individual to the other.

4.4.1.2 Random Effects

The random effects model incorporates the omitted variables as a random terms α_i which are independently and identically distributed over individuals. The model can be written as

$Y_{it} = \mu_i + \beta x_{it} + \alpha_i + \varepsilon_{it} \quad i = 1, \ldots . N; t = 1, \ldots , T$

And it assumed

$\varepsilon_{it} \sim$ iid $(0, \delta^2_\varepsilon)$, $\alpha_i \sim$ iid $(0, \delta^2_\alpha)$; the error term ε_{it} & α_i are mutually independent and independent of x_{js}(for all j and s). This assumption imply that the OLS estimator for μ & β are unbiased and consistent. However, OLS estimator would be unbiased and consistent as far as $\delta^2_\alpha = O$.

There are two components errors term α_i & ε_{it}, i.e. α_i is the individual specific component; while ε_{it} is the reminder that is assumed to be uncorrelated over time. Since composite error term α_i & ε_{it}, exhibit a particular form of autocorrelation, thus the standard OLS estimator will not be best linear unbiased estimator (BLUE). Hence we make use the more efficient GLS estimator by exploiting the structure of the error covariance matrix.

To derive a GLS estimator, first we estimate the covariance matrix Ω,

$$\text{Var}(\alpha_i T + \varepsilon_{it}) = \Omega = \delta^2_\alpha T'T + \delta^2_\varepsilon I_T$$

Where $T = (1,1, \ldots 1)'$ of dimension T and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots .. \varepsilon_{iT})$; I_T is the T-dimensional identity matrix.

Now we can estimate the GLS first by transforming the data by multiplying the vectors,

$Y_{it} = (y_{i1}, y_{i2}, \ldots y_{iT})'$ by Ω^{-1}

Then we can run the β_{GLS} estimator. This GLS estimator is composed of between estimator and fixed effects estimator, their relation ship could be written as:

$\beta'_{GLS} = \Delta\beta'_B + (I_k - \Delta)\beta_{FE}$

The GLS estimator is a matrix weighted average of the between estimator and the within estimator, where the weight depends upon the relative variances of the two

estimators. The between estimator ignores any information within individuals. The GLS estimator is the optimal combination of the within estimator and the between estimator, and therefore more efficient than either of theses two estimators.

The choice of the fixed vis-à-vis random effects, (Verbeek, 2000:318) explains the fixed effects basically considers the distribution Y_{it} given the α_i, where α_i has to be estimated. This makes sense intuitively if the individuals in the sample are not of "one kind" and cannot be assumed as random draw from the population. The interpretation is more appropriate when, *i*, the observations are companies, industries, countries. The random effects is not interested on individual effect α_i, the distribution is not conditional on α_i thus it integrates them out. Without going further, the choice of the estimation tool is fixed effects estimator because it cannot be assumed each observation of "one kind" for the banking industry. Each individual bank has its own features and idiosyncrasies.

4.5. Data Set

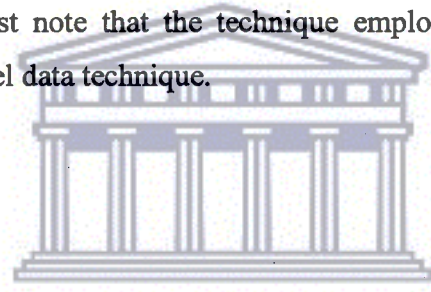
The data set used to estimate the model is acquired from the McGregor's word database, the South African Reserve Bank DI 900 database, and additional data collected directly from the individual firms concerned. The data is extracted from annual financial statements of the individual banks. The sample is drawn from the cross section of South African banks. However due to data compilation problem in small banks, the data could be biased towards large banks.

The model is making use of unbalanced panel data. This is the case because some of the banks entered into the banking industry only recently while the others have been in the industry for over a decade. Furthermore, for some banks some of data is missing for the initial years in the sample period. The choice of unbalanced panel data set entails the advantage of permitting a greater number of observations into the estimation thereby solving the degree of freedom problem. There was no need to make a stationarity test, because the model is estimated using static fixed effects estimator.

The choice of the estimator is made a priori, there was no need to undertake Hausman specification test.

Conclusion

The model being simulated to measure the level of competition in the South African banking industry is, a new empirical industrial organization theory, the Panzar-Rosse reduced form revenue function. The model is tailored to fit the South African context as well as the data availability. The econometric methodology used in the estimation is the fixed effects estimator in order to take into account the dynamism in the industry as well as to control for the unobservable bank specific features. The data set is chosen to be unbalanced panel data, in order to benefit from more comprehensive sample and observations. We must note that the technique employed in the simulation of the model is a simple panel data technique.



Chapter V

REGRESSION RESULTS AND INTERPRETATIONS

In previous section we have seen the specification of the P-R reduced form revenue equation for the South African banking industry. In this section we will simulate the model, a fixed effects estimator is simulated using unbalanced panel data for a sample of 15 banks for the period of 1993 to 2002. The resulting H-statistics is analyzed in relation to the literature reviewed; a comparative analysis is undertaken in relation to emerging market countries' banking systems.

5.1 Estimating the P-R Reduced form revenue function

As we have already underlined in the pervious section, the choice of fixed effects versus Random effects estimator depends on the nature of the model estimated and the population. In panel data econometrics, the error term is composed of the individual specific time invariant component as wells as the remainder time varying error term. The fixed effects estimator estimates the individual specific error term, while the random effects estimator assumes this individual specific error term to be randomly distributed. In heterogeneous population involving banks, countries, and regions; the fixed effects estimator is compatible because the individual specific error term cannot be expected to randomy distributed, thus the individual specific error should be estimated.

The "fixed effects" estimator

The P-R model is estimated based on unbalanced panel data for the sample of 15 South African banking firms for the period of 1993 to 2002, using fixed effects panel data econometrics. The reduced form revenue function estimated is:

$$\ln IR_{it} = C + \beta \ln pf_{it} + \alpha \ln pl_{it} + \delta \ln pk_{it} + \eta \ln ass_{it} + \theta \ln mix_{it} + \mu_{it} + \varepsilon_I.$$

The regression result[16] is reported in the following table,

Table 5.1: The fixed effects estimator

	Variables		Fixed effects (with in)
Dependent variable	Interest revenue	Log (IR_{it})	Coefficient & t-statistic
Factor prices	Funds	Log (pf_{it})	0.412
			(18.2)
	Labor	Log (pl_{it})	0.147
			(7.9)
	Capital	Log (pk_{it})	-0.043
			(-2.3)
H-statistic			0.516
Bank specific variables	Business mix	Log (mix_{it})	-0.134
			(-1.87)
	Total assets	Log (ass_{it})	-0.0066
			(-8.7)
Diagnostic test: R^2=0.87, Adjusted R^2=0.8, Durbin-Watson statistic = 1.71			

Note: all the variables are statistically significant at 5% level except for business mix, which is only statistically significant at 10% level. The null hypothesis that H=1 and H=0 is rejected. The figures in the parenthesis represent the t-statistic for each coefficient.

The dependent variable is the ratio of interest revenue to total assets (IR_{it}); the independent variables are the interest expenses to total funds (pf_{it}), personnel expenses to total assets (pl_{it}), and the capital expenses to fixed assets (pk_{it}). The bank specific variables included are the ratio of total fund to total assets (mix_{it}) and total assets (ass_{it}) that are incorporated to explain the effects of business mix in the industry and the economics of scale respectively

[16] See appendix I for the fixed effect estimator print out.

The diagnostic test: the adjusted R^2 that measures the goodness of fit is 0.8, this means 80% variation in the dependent variable is explained by a vector of input factor prices and a vector of bank specific variables. The Durbin-Watson statistic (DW) used to a test for autocorrelation. The DW is 1.711. In order to make provision for autocorrelation a SUR (seemingly unrelated regression) is estimated, thus variables are transformed with a form of Cochrane-Orchutt correction to model autocorrelation. White heteroscedasticity consistent variance is chosen to solve the problem of heteroscedasticity.

The coefficient of the funds, capital, and personnel expenses are the factor price elasticities, their summation makes the H-statistic $H= \beta + \alpha + \delta$ .The coefficient (β) of interest expenses is 0.412, the coefficient of personnel expenses (α) is 0.147 and the coefficient of capital expenses (δ) is –0.043. The coefficient of the interest expenses β is more significant and of higher magnitude; the coefficients of personnel expenses are positive and significant but still it is smaller than β. The coefficient of the capital expenses δ is significant but it is negative. The findings of Bikker & Haaf (2002) are comparable to our results, the relative contribution of each coefficient to the H-statistic. The elasticity the interest expense is more significant and positive; the elasticity of labor cost is significant and positive, but usually smaller than β, the coefficient of price of capital expenses δ varies in size, sign and level of significance. Molyneux *et al* (1994) asserts the sign of capital expenses varies across years and countries, and is significant at 5% level, whereas the coefficients of interest expenses are almost always positive and statistically significant.

The coefficient of the bank specific variables, the total asset, which is included in the model to account for the economics of scale, is significant but negative. Molyneux *et al* (1994:454) explains this phenomenon as, "the sign of asset size coefficient is mainly negative indicating that size-induced difference between banks may induce to lower total revenues per dollar asset". The other variable is the loan to asset ratio,

which indicates the business mix has a negative sign and it is insignificant at 5% level, but it is significant at 10% level.

The estimated H-statistic[17] is 0.516. This falls within the range of monopolistic competition between zero and one. Based on results of Vesala (1995, as cited in De Bandt and Davis, 2000:1048), we interpret the h-statistic as a continuous measure of the level of competition, in the sense that high a value of h-statistic within the range indicates stronger competition than lower value. In our case the H-statistic falls approximately in the middle of the continuous range. This indicates the level of competition in the industry unequivocally monopolistic competition.

The estimated h-statistic can be compared with the h-statistic estimated for other countries using the same model and econometric methodology. Let's take a sample of five emerging market economies' banking systems and look at their respective market concentrations and the H-statistic.

Table 5.2: International Comparison[18]

COUNTRY	NUMBER OF BANKING FIRMS	HHI	H-STATISTIC	TYPE OF COMPETITION
Greece	32	2800	0.032	Mc/monopoly
Mexico	23	1360	0.51	Mc
Argentina	113	865.7	0.95	Mc
Brazil	193	1278.6	0.66	Mc
South Africa	41	1750	0.516	Mc

Source: The data is adapted from Gelos and Roldós for Latin American countries. Honderoyiannnis *et al*, for Greek Banking industry, and South African reserve Bank supervision annual report 2003

Greece has the most concentrated banking system with the HHI 2800, followed by South Africa with 1750 points; Argentina has the lowest concentration, 865.7 points. Greeks' banking systems H-statistic is 0.032, which falls at the lower bound of the

[17] Refer back to section 4 for detailed discussion on the H-statistic

range for monopolistic competition i.e 0<h<1. Given the continuous nature of H-statistics, the Greek's banking system conducts itself almost in monopolistic manner, or we can suspect for some sort of tacit collusion of the banking firms. Argentina's banking system h-statistic, on the other hand, is 0.95 that definitely fall in the upper bound of the range of 0 and 1. We can make an assertion that there is more competition in the Argentinean banking system than in the Greek banking system.

The international comparison gives us an insight to competition test vis-à-vis the market concentration. In the case of Greek banking systems the H-statistic, which is highly concentrated, is close to zero. Though it is within the range of monopolistic competition, that is 0<H<1. In case of Argentina that has a relatively less concentrated banking system, the estimated H-statistic is on the upper bound of the range. There is a pattern of high correlation between the market structure and the h-statistic. But still the methodology holds itself from making any conclusion regarding causal link between market structure and H-statistic.

[18] The number of the banks refers only to the locally registered banks for the year of 2000.

Chapter VI

CONCLUSION

In the last decade, we have witnessed a steady growth of the banking industry, the growth rate ranges over 25% in 2001 to 4% by the end of 2002. The main features of the industry can be summarized as high-tech service industry, which is updated with recent information technology developments and substantial Internet banking activities. It is well capitalized with capital adequacy ratio 12.6% and relatively stable, though in 2001 and 2002 the industry had faced liquidity crises as a number of small banks failed including 6th and 8th biggest banks. The provision of banking services is very limited, 49 % of economically active South Africans still are not able to access banking services.

The market structure of the banking industry is highly concentrated. The CR_4[19] in the sample of four-product markets, that is, deposit, credit card, investment and installments market are 79.88 %, 95%, 77.7% and 89 % respectively. The Herfindahl-Hirschman Index (HHI) of the industry, as of December 2002, is 0.175(1750). This figure, according to the guidelines of American Department of Justice, is considered to be highly concentrated.

The methodology used is the Panzar-Rosse reduced form revenue function, which is a new empirical industrial organization theory [NEIO] that tests (measures) competition irrespective of the market structure. Looking back at subsequent developments in theories of industrial organization. That is, SCP paradigm which proposes the causation of market structure to conduct and conduct to performance, the contestability theory claims irrespective of the market structure the threat of free entry and exit of firms can bring about perfectly contestable markets, and recent game theoretic approaches. Therefore, it theoretically impossible to use the market structure

[19] It is calculated as of June 2003

to measure the level of competition in an industry. The natural choice will be to adopt a model, which can measure the competitive conduct irrespective of market structure.

The findings of the estimated Panzer-Rosse reduced form revenue function for a sample of 15 South African banking firms for the period of 1993-2002 indicates monopolistic competition prevails in the South African banking industry with estimated H-statistic 0.516. This result compared with respective H-statistic of four emerging market economies, the South African banking has more competition than of Greek and Mexican banking systems but lower competition than that of Argentinean and Brazilian banking systems. These findings are comparable with Okealaham (2001) findings that conclude South African banks are making comfortable returns, which reduces the incentive to seek markets that are considered difficult.

6.1 Implication

The regression results has clearly indicated a monopolistic competition is prevailing in the industry; it would be much interesting to explore what this means in terms of performance of the industry by revisiting the characteristic of monopolistic competition. A monopolistically competitive firm faces downward sloping demand curve and it is characterized by product differentiation and "excess capacity". Competition in this industry takes the form of price and non-price competition. In the long run equilibrium price is equal to long run average cost but it is greater than the marginal cost, this implies the existence of allocative inefficiency. Moreover, in the long run production takes place before the average cost curve reaches its minimum point, this denotes a sub optimal firm size but at the same time in short run the firms do not reach at the minimum short run average cost curve implying the existence of excess capacity. Monopolistic competition brings in allocative inefficiency as well as technical inefficiency[20] into the industry.

[20] See section 2.3 for detailed discussion on efficiency.

The interest rate[21] charged by banks would be higher than what it would be the case if the industry were perfectly competitive. The quantity of output falls short to satisfy the market demand. This particularly interesting in South African banking industry, which services only 51% of the economically active population. This gets even more interesting when we look further into informal sector and the small-scale enterprises which have a restricted access to financial services, particularly credit access; lack of access to such credit facilities is a detrimental for the growth of the small-scale enterprises. Theses enterprises, unlike big companies, do not have the capacity to raise capital independently of the banking industry.

The macro economic implication of the bank competition arises from the central role of banking industry has to play in the economy. The findings of Smith (1998:793) indicates that

> . . . increased bank competition raises the level of income and reduces the level of severity of business cycles. The quantitative effect on macro economic performance of less competition in banking can be large: for instance, an imperfectly competitive banking system can produce a worse macroeconomic out come than if the economy had no banks.

The intuition behind these findings is that as banks get more competitive then the fall (increase) in agency costs associated with the bank intermediation results in reduction (increase) of interest and non-interest payments. In imperfectly competitive banks, the reduction of agency cost realized by bank would be their profits. Furthermore firms as well as individuals tend to face higher cost of financing. This worsens when there is a recession in the economy because the credit worthiness of the firms is likely to deteriorate; their ability to raise funds directly is diminished. In such a scenario, these firms have to depend on the banking system. But in imperfectly competitive banking system the firms will have a limited finance at higher interest rate. This, therefore, exacerbates the recession and slows down the over all economic activity[22].

The macro & microeconomic ramification of imperfect competition in banking industry is of a great magnitude. Thus the regulatory body as well as competition

[21] Banking firms charge fees, besides the interest rate, for different type of services rendered to the public. Hence the monopolistically competitive banks can charge fees above the marginal cost.
[22] See Smith (1998) on bank competition and macroeconomic performance.

authorities needs to reformulate their policies so as to bolster competition in the industry. These policies can come up as a reduction of the regulatory barriers to entry; entry by local as well as by foreign firms into the industry can boost competition in banking industry.

The effect of foreign entry into domestic bank competition is vigorous. Claessens *et al* (2001) using 7900 bank observations from 80 countries for the 1988-1995 period, found out that " . . .increased presence of foreign banks is associated with a reduction in profitability and margins of domestic banks ". This being the case, when we look at this issue in the light of PriceWaterHouseCooper (2003) survey, which shows foreign banks view towards South African Bank regulations as barriers to entry. The policy option in this case is abolition the double capitalization requirement for foreign banks.

To conclude, the imperfect competition, namely the monopolistic competition will have micro as well as macro economic affects which needs to be mitigated by bolstering competition through a set of policy measures. However, the stability and soundness of banking industry should not be jeopardized.

Bibliography

ABSA, (2002) Is the South African banking sectors still sound? *Economic perspective*, 3rd quarter.

Audretsch, D.B., Baumol, W. J. & Burke, A.E. (2001) Competition policy in dynamic markets. *International journal of industrial organization,* 19: 613-634

Baltagi, H. B. (2003) *Econometric analysis of panel data*. Chichester; New York: John Wiley & Sons.

Baumol, J.W. (1982) Contestable markets an uprising in the theory of industry structure. *The American economic review*, 72(1) 1-15 March

Baumol, J.W., Panzar, J.C. & Willig, R.D. (1986) On the theory of Perfectly Contestable markets. Williamson, E.O. (ed.), *Industrial organization.* Aldershot, Hants: Edward Elgar.

Bhattacharya, S., Boot, A.W. & Thakor, A.V. (1998) The economics of bank regulation. *The journal of money, credit and banking, November*: 745-767

Bikker, J. A. & Haaf. k. (2000) *Competition, concentration and their relation ship: an empirical analysis of the Banking Industry (preliminary draft). Paper for the Financial Structure, Bank Behavior and Monetary Policy in the EMU Conference*, October 5-6, 2000, Groningen

- - - - - - - - - - - - (2002) Competition, concentration and their relation ship: an empirical analysis of the Banking industry. *Journal of banking and finance*, 26(2002): 2191-2214

Bresnahan, T. (1997) Testing and measurement in competition models. Kreps, D.M. & Walls, F.K. (eds.), *Advances in economics and econometrics: theory & application.* Economic society monograph, NO.28. New York: Cambridge University Press

- - - - - - - - - - - -(1989) Empirical studies of industries with market power. Schmalensee, R. & Willig, R. (eds.), *Handbook of industrial organization*, Vol. 11. Amsterdam: North-Holland.

Caminal, R. & Matutes, C. (2002) Market power and banking failures. *International journal of industrial organization,* 20 (2002): 1341-1361

Cavalluzzo, K.S. & Cavelluzzo, L.C. (1998) Market structure and discrimination: the case of small business. *The journal of money, credit and banking, November*: 771-791

Claessens, S., Demirgüc-kunt, A. & Huizinga, H. (2001) How does foreign entry affect domestic banking markets? *Journal of banking and finance,* 5(2001): 891-911

CliffeDekker (2003) South African banking regulation. [Online]. Availablehttp://www.cliffedekker.co.za/litreature/regulation/foreigm.htm

Competition Commission. (2001) *Industry overview and database of the South African banking industry: policy & research.* [Online]. Available http://www.compcom.co.za/policyresearch.esp

- - - - - - - - - - - - - - - - (2000) *Competition commission reports to South African Reserve Bank: the proposed merger between NEDCOR and STANBIC.* [Online]. Available http://www.compcom.co.za/policyreaserch/competition commission/stanbic.pdf

De Bandt, O. & Davis, E.P. (2000) Competition, contestability and market structure in European banking sector on the eve of EMU. *Journal of banking and finance, 24(*2000): 1045-1066

Feasibility (2001) *Liberalization, regulation and provision: the implications of compliance with international norms for the South African financial sector. Tips annual forum 2001.* [Online]. Available http://www.tips.org. za/research papers/list.esp

Finscope (2003) Finscope 2003 survey. Finmark trust. [Online] available http://www.finscope.co.za/brochure/2003/BNLSfin2003.pdf

Fischer, T. & kamerschen, R.D. (2003) Measuring competition in the U.S airline industry using the Rosse-Panzar test & cross–sectional regression analysis. *Journal of applied Economics*, Vol. VI no.1 (may 2003): 73-93

Fourie, F.C.v.N. & Smit, M.R. (1989) Import adjusted concentration trends in South African manufacturing industry. *South African journal of economics*, 57(3).

Fourie, F.C.v.N. (1996) Industrial concentration levels and trends in South Africa: completing the picture. *South African journal of economics,* 64(1)

Fourie, F.C.v.N. & Smith, A. (1993) Concentration, tariff protection and industrial performance in South Africa. *South African journal of eco*nomics, 61 (3)

Fraser, R., Gup, E. & Kolari, W. (2001) *Commercial banking: the management of risk.* Cincinnati: South Western collage publishing.

Gelos, R.G. & Roldós, J. (2002) Consolidation and market Structure in emerging market-banking system. IMF working paper, WP/02/186.

Geroski, A.P.& Mata, J. (2001) The evolution of markets. *International journal of industrial economics,* 19 (2001): 999-1002

Greuning, V.H. & Bratanovic, S.B. (2000) *Analyzing banking risk.* Washington D.C: The World Bank.

Hempell, H.S. (2002) Testing for competition among German banks. Discussion paper 04/02. Economic research center of the Deutsche Bundesbank.

Houdroyiannnis, G., Lolos, S. & Papapetrou, E. (1999) Assessing competitive conditions in the Greek banking system. *Journal of international financial markets, institutions and money,* 9 (1999): 377-391

KPMG. (2002) South African banking survey. [Online]. Available http://www.kpmg.co.za/moduls

Leach, D.F. (1992) Absolute vs. relative concentration in manufacturing Industry: 1972-1985. *South African journal of economics.*

Molyneux, P. & Forbes, W. (1995) Market structure and performance in European banking. Applied economics, Vol.27 (2): 155-59

Molyneux, P., Lloyd-Williams, D.M. & Thornton, J. (1994) Competitive conditions on European banking. *Journal of banking and finance,* 18 (1994): 445-459

- - - - - - - - - - - - - - (1996) Competition and market contestability in Japanese commercial banking. *Journal of economics and business,* 1996.48: 33-45

Nathan, A. & Neave, E.H. (1989) Competition and contestability in Canadian financial system: empirical results. *Canadian journal of econo*mics, 22(3): 576-594

Okealaham, Charles (2001) Structure and conduct in the Commercial banking sector of South Africa. Tips annual forum 2001. [Online]. Available http://www.tips .org.za

Panzar, J.C. & Rosse, J.N. (1987) Testing for "Monopoly" equilibrium. *The journal of industrial economics*, XXXV (4): 443-456

Price Water house coopers (2003) Strategic and emerging issues in South African banking.[Online].Available.http://www.pwc.com/ZA/ENG/INSSOL/PUBL/TAX/pwc_SA2003banking.pdf

Scherer, F.M. (2000) Antitrust, efficiency, and progress. Scherer, F.M. (eds.). *Competition policy, domestic and international.* New York: Edward Elgar publishing.

Smit, R. M. (1999) Competition policy: South African literature, researches and data. Tips working paper. [Online]. Available http://www.tips.org.za/researchpapers/list.esp

Smith, T.R. (1998) Bank competition and macro economic performance. *Journal of money, credit and banking, November*: 767-771

South African reserve bank (2000) *Bank supervision department annual report.* Pretoria. [Online]. Available http://www.resbank.co.za

---------- (2001) *Bank supervision department annual report.* Pretoria. [Online]. Available http://www.resbank.co.za

---------- (2002) *Bank supervision department annual report.* Pretoria. [Online]. Available http://www.resbank.co.za

Verbeek, M. (2000) *A guide to modern econometrics.* Chichester; New York: John Wiley and Sons, LTD.

Waterson, M. (1993) Are Industrial economists still interested in concentration? Casson, M. & Creedy J. (eds.), *Industrial concentration and economic inequality.* Aldershot, Hampshire: Edward Elgar: 107-121.

Vol. 70, pp. 59-62

Appendix I

Dependent Variable: LNINTR?
Method: Seemingly Unrelated Regression
Date: 12/02/03 Time: 03:50
Sample: 1993 2002
Included observations: 10
Number of cross-sections used: 14
Total panel (unbalanced) observations: 94

Variable	Coefficient	Std. Error	t-Statistic	Prob.
LNPL?	0.147067	0.018569	7.919946	0.0000
LNPK?	-0.043527	0.018838	-2.310637	0.0236
LNPF?	0.412214	0.022582	18.25375	0.0000
LNASS?	-0.006615	0.000756	-8.752069	0.0000
LNMIX?	-0.222520	0.074299	-2.994928	0.0037
Fixed Effects				
1--C	0.801052			
2--C	0.792096			
3--C	0.796785			
4--C	0.788542			
5--C	0.794420			
6--C	0.775040			
8--C	0.765033			
9--C	0.709143			
10--C	0.742825			
11--C	0.757473			
12--C	0.727035			
13--C	0.776545			
14--C	0.665494			
15--C	0.825461			

Weighted Statistics			
Unweighted Statistics			
R-squared	0.845800	Mean dependent var	0.859940
Adjusted R-squared	0.808792	S.D. dependent var	0.050429
S.E. of regression	0.022051	Sum squared resid	0.036470
Durbin-Watson stat	1.718823		

Appendix II

South African Literature on Competition and Concentration

The interest on identifying market structure and market concentration started with seminal work of Du Plessis (1977,78a, 78b). He estimated market concentration for South African manufacturing industry. To mention some major studies: Fourie and Smit (1989), Leach (1992) using the Rosenbluth index and 50 percent occupancy count, He found a strong and highly significant correlation between concentration and industry profitability for South African manufacturing industry.

Fourier and Smith (1993) extend the simple analysis of the relationship between industrial concentration and profitability to include import protection into account. This gives an important insight by assessing the impact of import competition on effective levels of relative concentration as well as market power. Fourie and Smith (1993:25) come to conclude, "There are sufficient empirical grounds for suspecting as significant interaction between concentration and import protection in determination of industrial profitability".

Leach (1992), challenged Fourie and Smit (1989) and Du Plessis (1977), concluded that in South African manufacturing industries; a discrete measure of concentration on the 80% occupancy rate shows on average no change in concentration over the period 1972-1985. A summary measures of concentration the Rosenbluth index-shows, on average, a decrease in concentration over the same period. In his other work titled *Concentration and profits in South Africa: monopoly or efficiency* emphasized the importance of competition policy, and reviewed Reekie (1984) which made the use of two measurements that is measurement of concentration and measurement of profitability. He found a strong and highly significant correlation between concentration and industry profitability for South African manufacturing industries. He also found the Rosenbluth index to be superior to the occupancy count in the sense of a higher and statistically more significant correlation with industrial profitability.

He goes on to check its consistency with alternatives industrial organization theories that is SCP and efficiency theories. Using Demsetz methodologies He comes to conclude that it is consistent with efficiency theory.

Fourie (1993) on *Economic concentration: a survey and evaluation of policy options* argues that public concern regarding the concentration of economic power in South Africa is justified from both equity and efficiency point of views. In his other study (1991) he tried to assess the relationship between economic concentration and inflation, his findings depicts that prices in more concentrated industries are relatively rigid with regard to spending fluctuations, and specially downswing in spending (Fourie, 1991:32). Fourie (1996), reviewed measurements used till 1996, and he made use of CR_3, CR_4 and CR_{10} concentration ratios, the Herfindhal-Hirshman, and Rosenbluth indices, as well as Gini-indices; all these were used for the South African manufacturing industries for the period 1982-1988. He came to conclude that "concentration in the South Africa manufacturing industry is high and seemingly on increase" (Fourie; 1996:31).

Fourie and Smith (2001) noted the longest running debates in the field of industrial organization could not come to conclusion. However, they believed that cross sectional study could still give important insights. Analyzing the concentration profits relationship using simultaneous equations highlighted some important multi-causality and identification of number of exogenous variables, which can influence the concentration-profits relationship.

Appendix III

4.2. Empirical Models: Comparison

In adapting an empirical model, let's briefly review the empirical models developed based on the initial Panzar-Rosse reduced form revenue function to assess the level of competitive conduct in the banking industry. Among the vast application of this model in different industries, we will be particularly interested in contributions made on banking industry. This model has been extensively used in banking industry to test competition in a number of countries.

4.2.1. Shaffer (1982)

He has used the Panzar-Rosse (1977) reduced form revenue equation to study a sample of unit banks in New York. He postulated that total revenue; the dependent variable can be explained by input prices such as interest rate, wages, and other variables.

Model:

$$\ln TR = a + b(\ln PL) + c(\ln PK) + d(\ln PF) + e(\ln ASST) + g(\ln MKT) + s[(c+d)/dep] + h[(h+I)/loans],$$

Where

Ln	=	natural logarithm
TR	=	total interest revenue
PL	=	annual wages, salary and benefits per full time equivalent employee (unit price of labor)
PK	=	ratio of annual expenses of premises, furniture, fixture and equipment to the balance sheet total of these items. (Unit price of premises)
PF	=	ratio of interest expenses to total depositors (unit price of funds)
AST	=	total bank assets

MKT	=	market deposits with commercial banks.
(C+D) / Dep	=	ratio of cash and due from depositors to total deposits
(C+I) / Loans	=	ratio of commercial and industrial loans to total loans

The variables PL, PK, PF are input prices while (c + d) / dep and (h + I) / loans, are included to account for other variables affecting equilibrium revenues. Total assets are included with the intent to identify the presence or absence of scale economies. Market is included as a proxy for aggregate market demand. The variables (c + d) / dep and (h + I) / loans are used to correct for effects of differing levels of correspondent activity and differing business mix. Then the H-statistic is calculated as a summation of coefficients of input prices, that is, H = b + c + d.

In his later work, Shaffer made use of Panzar–Rosse by relating return on assets and return on equity to input prices.

ln ROA = a +b (ln plk) + c(ln pk) + d(ln pf) + e(ln Asst) + f(ln Mkt) + g[(c+d)/dep] + h[(h+I)/loans],

He relaxed the long run equilibrium assumption and assumed that the dependent variable should not be correlated with input prices. In both cases the test conducted on Canadian banking found H-statistic less than zero which implied disequilibrium while H=0 is consistent with equilibrium.

4.2.2. Nathan & Neave (1989)

Using the Shaffer's method, Nathan & Neave (1989) estimated Panzar –Rosse reduced form revenue function for Canadian banks, trust companies and mortgage companies. Shaffer's model was modified for three reasons:

i. To test the relation ship function using net revenue as dependent variable;
ii. To account for difference between US and Canadian financial business; and
iii. To eliminate some possible incidence of multicollinarity

The Shaffer's regression did not take the effects of loan losses into account, but Nathan & Neave (1989) believes its conceptual importance, thus, they incorporated the "loan losses' in their Panzar–Rosse reduced form revenue estimated for Canadian banks.

Model:

ln TRLL = a + b (lnPF) + c (lnpkb) + d (lnpl) + e (ln) + f (lnbr) + g06

Where

TRLL	=	total revenue less provision for loans losses
PF	=	interest expense to total deposits (unit price of funds)
PKB	=	premises expenditure to number of branches (unit price of capital)
PL	=	wages and salary expenses to number of employees. (Unit price labor)
AST	=	total assets
BR	=	number of branches / total numbers of branches in the system.
BR	=	1, for six largest banks
		0, for the other banks

Loan loss is determined by administrative mechanism; as a result, it may not reflect the workings of market forces exactly in any given year. Nathan & Neave (1989:582) concluded that the loan loss variable does not turn out to have significant impacts on their estimates of H-statistic. Both regression that omit the banks loan loss and the regression that use loan losses as a input price variable were estimated but with no significant changes to the results

4.2.3. Molyneux *et al.* (1994,1995)

The reduced form revenue function for European banking industry was estimated as The model:

L n TRESS = a + b ln PL + c ln PK + d ln PF + e ln ASS + g ln CAPASS + h ln IBDEP

Where

ln	=	natural logarithm
TRASS	=	total interest revenue per dollar of assets
PL	=	Personal expenses per dollar of assets (proxy for unit price of labor)
PK	=	capital expenses per dollar of fixed assets (proxy for unit price of capital)
PF	=	ratio of annual interest expenses to total funds (unit prices of funds)
ASS	=	banks assets
LMASS	=	loans to assets ratio, and
IBTDEP	=	inter bank deposits to total deposits

Molyneux *et al* (1994,1995) include the explanatory variables firm specific variables in the estimation. These control variables are included to take into account "other risks", cost and size characteristics. Since the dependent variable TRASS is not risk adjusted, they employed two variables to account for firm specific risks; that is, the total risk capital to asset ratio (CAPASS), loans to asset ratio (LAMASS) and the inter bank deposits to total deposits (IBTDEP) ratio to account for difference in deposit mix.

4.2.4. Bikker and Haaf (2000)

They adopted the Molyneux *et al* (1994) model and specified the Panzar -Rosse reduced form revenue equation to asses the competitive conduct in European banking industry and the effect of introduction of Euro, the new European currency, on the level of competition in European banking.

The model

$$\ln Intr = \alpha + (\beta \ln AFR + \gamma \ln PPE + \delta \ln PCE) + \zeta \ln BSE + \eta \ln OI + e$$

Where

ln inter = log of the ratio of total interest revenue to total balance sheet,

AFR = the ratio of annual interest expenses to total funds, or the average rate,

PPF = is the ratio of personal expenses to total balance sheet, or the (approximated) price of personal expenses,

PCE = is the ratio of physical capital expenditure and other expenses to fixed assets, or the price of capital expenditure,

BSF = Bank specific exogenous factor (without explicit reform to their origin from cost or revenue function),

OI = the ratio of other income to total balance sheet.

The above specification is modified in order to unify whether the competitive structure has changed over time due to liberalization and deregulation. Thus they applied a pooled data analysis, and they assumed the market structure shifted gradually over time " Ignoring market dynamics may lead to imprecise parameter estimates and biased H-statistic, which could in turn result in wrong inference about the competitive nature of the banking industry" (Bikker & Haaf, 2000:). Therefore, they included the continuous time-curve model to account for the market dynamics. This incorporated by multiplying the estimates of H-statistic by continuous time curve model exp (ϵ *time).

$$\ln Intr = \alpha + (\beta \ln AFR + \gamma \ln PPE + \delta \ln PCE) EXP (\epsilon . time) + \zeta \ln BSE + \eta \ln OI + e$$

4.2.5. De Bandt and Davis (2000)

The extension of the Panzar-Rosse (1987) methodology to banking requires assuming that banks are treated as a single product firm (De Bandt and Davis, 2000). The model is based on intermediation approach to banking where banks are viewed as financial intermediaries. Their basic difference from the traditional approach is that they use total income as dependent variable instead of gross interest revenue. Their justification for this, " . . . given that for banks in the competitive struggle for survival, the distinction between interest and non interest income becomes less relevant, competition being equally vigorous for both" (De Bandt and Davis, 2000:1049).

Among the number of specification in the banking literature, they opted for the Nathan & Neave (1995) on Canada and Vesala (1995) on Finnish banking specification, which use the logarithms of interest revenue, instead of Molyneux *et al* (1994) specification on European banks which use the ratio of interest revenue to total balance sheet as endogenous variable. The fundamental reason behind this is economic and econometric reasons, a ratio of interest revenue to assets provides price equation, and the log specification may also reduce possible simultaneity.

The model

$$\text{Log } R_{it} = \sum \alpha_j \log w^j_{it} + \sum \beta_k \text{logs}^k_{it} + \sum \gamma_n x^n_{it} + \epsilon_{it}$$

For t = 1…T, where T is the number of periods

I = 1…I, where I is the total number of banks

R = is gross interest revenue or total gross revenue

W = vector of factor prices

S = are scale variables measuring the capacity level at which level the bank operates

X is a vector of exogenous bank – specific variables that may shift the cost and revenue schedule (business mix)

4.2.6. Hempell (2002)

The Panzar–Rossee (1987) was applied to test the level of competition in German banks. With this purpose, the sum of the factor price elasticities with respect to banks revenue is estimated.

The model:

$$\ln TIN1tTA_{i,t} = a_1 + b_1 \ln IETF2_{i,t} + b_3 \ln wage1_{i,t} + b_3 \ln aCEtFA_{i,t} + c_1 LtTL_{i,t} + c_2 IDtTD_{i,t} + c_3 imab1t_{i,t} + c_4 cashi_{i,t} + d_1 t + \lambda_t + \mu_i + \mu_{i,t}$$

Where:

TIN1tTA	=	total income to total assets,
IETF2	=	interest expenses to total funds
Wage1	=	Personnel expenses to total assets
CetFA	=	capital expenses to fixed assets

Other bank specific variables

LtTL	=	customer loans to total loans
IdtTD	=	inter bank deposits to total deposits
Imab1t	=	maturity structure of customers loan portfolio
Cashi	=	cash flow to business volume by sector of borrower, weighted with the portfolio of loans to enterprises

The selection of variables and their specification takes into account the work done by her predecessors. It takes total revenue instead of interest revenue as the dependent variable, which is consistent with De Bandt and Davis (2000). The dependent variable is expressed as a ratio to total assets to abstract from size effects (Hempell, 2002:14). De Bandt and Davis as well as Neave and Nathan (1989) criticize this specification by saying that it is just a price equation and it is inappropriate.