A Study of Efficiency and Competitive Behaviour of Commercial Banks in Malaysia

Mahadzir Ismail

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy of University of Wales

Cardiff Business School
University of Wales, Cardiff

2005

ABSTRACT

This thesis is mainly concerned with the efficiency and competitive behaviour of the commercial banks in Malaysia for the period 1994 – 2000. Firstly, efficiency, either technical or cost, is defined as a relative concept. Technical efficiency is derived from a production frontier function whilst cost efficiency is obtained from a cost frontier function.

The measurement of technical efficiency uses two basic DEA models that make two assumptions: constant returns to scale and variable returns to scale. Having obtained the efficiency scores, further investigation is made into identifying the sources of productivity growth. This is done by constructing the Malmquist productivity index. In contrast, the estimation of cost efficiency under the parametric approach is performed by using two methods: the stochastic frontier approach and the fixed effects method. For the purpose of measuring efficiency, this study uses three inputs and three outputs selected based intermediation approach.

The competitive behaviour of the banks is determined by employing the Panzar-Rosse methodology. This methodology requires two tests: equilibrium and a competition test. The H-statistic is then obtained that will determine the level of contestability in the market.

This study finds that, in general, the average efficiency measures are subject to the estimation technique used. It shows that the DEA technique generates higher efficiency scores than the parametric methods. It is also shown that the scores are positively related to each other. The study also finds that bank ownership plays important role. It suggests that the foreign banks in general are more efficient than the domestic banks. With regard to sources of productivity growth, the main contributor is improvement in technology rather than efficiency improvement. Finally, the study shows that the commercial banks are best described as operating under monopolistic competition.

ACKNOWLEDGEMENTS

In the Name of Allah, the Most Gracious, the Most Merciful.

All praise to Allah, the Lord of the Universe, for his Blessing and Mercy without which this study might not be realised. Peace be upon the beloved Prophet Muhammad for his teaching that always enlightens my life throughout the challenging period of my stay in Cardiff.

Personally, I am indebted to many individuals and several parties who have kindly provided all the required support and assistance for the completion of this study. It is my great pleasure and honour to express my gratitude to them for their endless contribution.

My sincere thanks to my principal supervisor, Professor Kent Matthews who kindly agreed to accept me as his student and who has truly guided me from the very beginning of my study. His continuous support and expert guidance are priceless. I also thank my other supervisors, Professor Gerry Makepeace and Professor Roger Clarke for their constructive comments and easy accessibility.

I would also like to thank my main sponsor, MARA University of Technology, for giving me the golden opportunity to undertake this study and for providing financial support and other assistance throughout my study at Cardiff Business School. Similar gratitude is due to my fellow Malaysian students and friends in Cardiff for their support and encouragement. Also, my special thanks to the committed staff at the PhD Office: Sara, Elsie and Laine for their continuous support and assistance throughout my study.

My heartfelt thanks go to my beloved mother, Kholijah Othman, my father, Ismail Ariffin, and my parents-in-law who have sacrificed a lot in letting me and my family go far away from themselves and who always pray for my success and the safety of my family.

Last but not least, my utmost gratitude to my beloved wife, Siti Rahani Chin and my children, Siti Qurratu'aini, Siti Nabihah, Muaz, Siti Naimah, Muhammad Naseem and little Marwan Hadid for their patience and support. Their presence with me in Cardiff has truly provided a strong inspiration and motivation to accomplish this study.

List of Abbreviations and Acronyms

ABM	Association of Bankers in Malaysia
APF	Alternative Profit Function
ATM	Automatic Teller Machine
BAFIA	Banking and Financial Act
BCR	Banker, Charnes and Rhodes
BLR	Base Lending Rate
BNM	Bank Negara Malaysia
CCR	Charnes, Cooper and Rhodes
CRS	Constant returns to scale
DEA	Data Envelopment Analysis
DFA	Distribution-Free Approach
DMU	Decision-making units
EPF	Employees Provident Fund
ES	Efficient-Structure
FEM	Fixed Effects Model
FFF	Fourier-flexible Functional Form
FIR	Financial Intermediation ratio
FSMP	Financial Sector Master Plan
GATS	General Agreement on Trade in Services
GDP	Gross Domestic Product
GNP	Gross National Product
IOFC	International Off-Shore Financial Center
KLSE	Kuala Lumpur Stock Exchange
LR	Likelihood ratio

LSDV	Least squares dummy variable
MEPS	Malaysian Electronic Payment Services
MESDAQ	Malaysian Exchange of Securities Dealing and Automated Quotation
MGS	Malaysian Government Securities
MLM	Maximum likelihood method
NEP	New Economic Policy
OLS	Ordinary least square
OTE	Overall technical efficiency
PCSE	Panel-corrected standard errors
PDS	Private Debt Securities
PTE	Pure technical efficiency
REM	Random Effects Model
RM	Ringgit Malaysia
RMP	Relative Market Power
ROA	Rate of returns on total assets
RWCR	Risk weighted capital ratio
SCP	Structure-Conduct-Performance
SE	Scale efficiency
SFA	Stochastic Frontier Approach
SMEs	Small and Medium Enterprises
SPF	Standard Profit Function
TFA	Thick Frontier Approach
TTRS	Two Tier Regulatory System
VRS	Variable returns to scale

WTO	World Trade Organisation

CONTENTS

CHAPTER THREE: Banking Development in Malaysia

CHAPTER FOUR: Analysis of Technical Efficiency

– A Non-parametric Approach

LIST OF CHART AND FIGURES

LIST OF TABLES

Chapter One

CHAPTER ONE

Introduction

1.0 Introduction

Efficiency and competition can be considered key ingredients to ensure the continuous development of the banking industry. Efficiency in simple terms can be related to how best a bank can make use of all available resources by transforming them into the desired outputs. At an efficient bank, the maximum amount of output will be produced given a certain amount of input. Based on this definition, efficiency can be viewed as a dimension of bank performance. For example, a bank is doing well if it is efficient in its operation. Efficiency can be further observed in terms of real quantity of output produced or in terms of cost incurred during the production of a given amount of output. This leads to two types of efficiency: technical efficiency and cost efficiency. Efficiency can be nurtured by a conducive environment that permits the banks to make the best decision regarding the best combination of inputs to be used or the appropriate output mix. In turn, the banking environment is very influenced by the measures introduced and taken by the authority, for example, financial liberalisation.

On the other hand, competition refers to how the banks behave in the market, when there is more than one player and there is no collusive action amongst them. The banking industry is competitive when the banks are able to adjust their pricing strategies given changes in the input prices or input costs. However, the extent of banks' responsiveness is influenced by the market structure. In turn, the market structure itself is influenced by the policies pursued by the authority such as merger programmes and deregulation. Bank mergers will affect the structure of the banking industry and that, subsequently, has an impact on the market power. Meanwhile, deregulation enables the banks to respond appropriately to any changes in the market, especially input and output prices.

This chapter is organised as follows. The immediate section, Section 1.1, introduces the background of this study. Section 1.2 is about the main objectives that this study aims to achieve. Section 1.3 discusses the scope and limitations of the research. Finally, the organisation of this study is presented.

1.1 Background to the study

In developing and Third World countries, the financial sector is not fully developed. Access to financial services might be limited since the financial infrastructure is not well established. Financial institutions might be limited to commercial banks, financing institutions or savings institutions. In addition, financial products are confined to the traditional forms like demand, savings and fixed deposit plus negotiable certificate. In the case of Malaysia, the government takes the initiative and the lead in developing the financial sector. Once it became a sovereign country in 1957, Malaysia wanted to reduce its heavy reliance upon foreign-owned financial institutions especially the banks. The first step was to increase the number of domestic banks and limit the presence of foreign banks. The second has been to pursue financial liberalisation since the early 1970s. This has been demonstrated by relaxing government interventions on the interest rates from time to time. In 1978, the banking institutions were allowed to decide their own interest rates when the government decided to abolish the fixed interest rate. Further liberalisation measures were carried out. This includes the diversification of financial products such as the introduction of a negotiable certificate of deposits and wider scope of investment opportunities (after the introduction of new banking act in 1989).

By the end of the 1970s, the Central Bank believed that Malaysia had too many banks compared to its real size. The creation of new banks was not allowed and the existing banks were encouraged to consolidate. However, the call for bank consolidation throughout the 80s was not received well by the existing banks in the market. Only a few consolidations took place after the economic decline in 1985-86. The Asian financial crisis in 1997-98 had exposed financial difficulties in several banks and this gave the much needed push for the banking institutions to consolidate. The government and the Central Bank announced a major consolidation in 1999 that would reduce the number of domestic banking institutions to ten banking groups by 2000. This consolidation exercise was finally completed in 2001.

1.2 Objectives of the research

Financial liberalisation and bank consolidation serve two main objectives: to enhance the capacity of domestic banks via greater efficiency, and to open the market to greater competition. Financial liberalisation measures like relaxing the control over market interest rates and product diversification give more room for the banking institutions to enjoy greater efficiency, which eventually leads to a competitive edge. Bank consolidation made the domestic banks relatively larger than before. The larger scale of production in theory should enable the banks to achieve greater efficiency and enable them to compete with well-established foreign counterparts.

This thesis is basically concerned with the efficiency and the competitive behaviour of the commercial banks in Malaysia for the period 1994-2000. Given the various measures of financial liberalisation since the late 1970s, this study seeks to determine the level of efficiency of the commercial banks using both a parametric and a non-parametric approach. Having obtained the measures of efficiency, this study will seek to determine whether there is some justification for bank consolidation. Finally, the aim is to identify the competitive behaviour of commercial banks using a non-structural approach (Panzar-Rosse model).

The objectives of this research are as follows:

i. To measure the technical efficiency of the commercial banks using a non-parametric approach

ii. To identify the sources of productivity growth of the commercial banks using a non-parametric approach (given the availability of yearly measure of technical efficiency)

iii. To measure the cost efficiency of the commercial banks using both a parametric and a non-parametric approach

iv. To determine the competitive behaviour of the commercial banks using a non-structural approach.

Objectives (i) and (ii) will be covered in Chapter Four. Objectives (iii) and (iv) will be covered in Chapters Five and Six respectively.

1.3 Scope and limitation of the research

This study applies only to the commercial banks in Malaysia, both domestic and foreign. Other banking institutions, like finance companies and merchant banks, were excluded due to differences in what constitutes input and output. The most significant difference is the provision of current accounts, which can only be offered by the commercial banks. Another distinction refers to the composition of outputs. The finance companies concentrate more on hire and purchase while the commercial banks have a greater range of loans and advances.

Efficiency is confined to a relative concept following Farrel's definition (1957). The relative efficiency is calculated by comparing the performance of a bank with the performance of other banks facing the same technology, regulations and environment. The technical efficiency measure is estimated using the production function while the cost efficiency is based on the cost function.

This study covers the period 1994 to 2000. Although financial liberalisation could be traced back to the end of the 1970s, 1994 is chosen because of the local incorporation of the foreign banks. Before 1994, the foreign banks were operating as foreign branches. They were not required to publish their income statement and this created a chronic data problem. After 1994, they had to publish their annual data in a way similar to their local conterparts. This study could not go beyond the year 2000 because of bank consolidation. This consolidation was finally completed in 2001. After 2000, all domestic banks can be regarded as new entities. Few mergers took place in 1997 and 1999. To maintain the consistency, newly created banks were excluded from this study.

This study relies upon the data produced by the accounting standard. The relevant data are obtained from the income statements of the commercial banks, published by Bankscope.

1.4 Organisation of the thesis

This thesis has seven chapters. Chapter Two deals with the theoretical foundation of efficiency and competitive behaviour of the banks and a review of the previous studies. This chapter will look into the various techniques employed to estimate the efficiency of the banks. In Chapter Three, this study explores the banking development in Malaysia. Several important aspects of banking development are highlighted like the process of financial liberalisation, financial widening and financial deepening. The recent financial crisis (1997-98) and its impact on the banking industry are also discussed.

Chapter Four is intended to serve the first two objectives of this research. Firstly, it presents the measurement of technical efficiency using data envelopment analysis. Having obtained the scores, the banking performance is compared across all banks based on ownership. The profiles of efficient and inefficient banks are also established. Secondly, it presents the construction of the Malmquist productivity index and the decomposition of the sources of productivity growth.

Chapter Five deals with the third objective of the research. It presents the estimation of cost function and cost efficiency using the parametric approach in the context of panel data models. For the purpose of comparison, the cost efficiency is also estimated using data envelopment analysis.

Chapter Six evaluates the competitive structure of the banks and then determines the present level of market contestability created by the process of deregulation. This is to serve the final objective of the research. To achieve this objective, the Panzar-Rosse methodology is used. To determine the type of market that the banks are operating, the H-statistics will be produced.

The final chapter is about summary, policy recommendations and future research. The findings will be discussed in relation to the pre-determined objectives of the study. Based on this discussion, this study presents what can be learnt and proposed for the improvement of banking industry in Malaysia. Further research areas are also considered.

Chapter

Two

CHAPTER TWO

Literature Review:

Efficiency and Competitive Behaviour of Financial Institutions

2.0 Introduction

This chapter reviews previous works related to the efficiency and competitive behaviour of financial institutions, in particular commercial banks. The efficiency issue was sparked by the seminal work of Farrel in 1957. Farrel's idea was further developed by the works of Aigner et al. (1977) and Meeusen and van den Broeck (1977) using a parametric approach, and by the work of Charnes et al. (1978) using a non-parametric approach. Since then, thousands of studies have been undertaken to evaluate the efficiency of financial institutions all over the world. Tavares (2002), for example, compiled more than 3,000 studies that used non-parametric approaches.

The structure of this chapter is as follows. In the next section, we discuss the functions and importance of the commercial banks. A better understanding of the functions of these banks is a pre-requisite in deciding which subsequent approach needs to be taken in analysing banks' efficiency. This is followed by a discussion of the major issues surrounding the financial institutions, such as the definition of what constitutes inputs and outputs and efficiency related issues such as the concept and measurement of efficiency. In fact, the measurement of efficiency is of great interest to researchers, as the outcomes enable the benchmarking of financial institutions. However, in this section, the concept of efficiency is confined to cost and technical aspects only. Section 2.4 is about the non-parametric approach used to estimate efficiency. Productivity growth is also discussed. Section 2.5 summarises the results of the previous studies. Section 2.6 presents the case of profit efficiency as an alternative concept to cost or technical efficiency. This is followed by the discussion on efficiency correlates. Section 2.8 deals with competitive behaviour of the banks. Section 9 concludes.

2.1 Functions and Importance of Commercial Banks

In any economic system, the financial sector fulfills several very important functions. Some commentators consider it the heart of an economy. Fry (1995), for example, stated that there are at least two functions of financial institutions. The first is to create money and facilitate and administer the payments process. Although the central bank is the sole issuer of currency, the financial intermediaries carry out the administration by providing and allocating currency notes of desired denominations when and where they are demanded. They also transfer deposits as instructed by the depositors. The second is to reconcile and fulfil the financial needs of savers and investors, lenders and borrowers. Here, the financial institutions act as risk managers on behalf of savers. The savings are managed in such a way that the savers get reasonable rates of return. The funds are made available and accessible to borrowers at a reasonable cost. Allen and Santomero (1998 and 2001) argue that the traditional functions of providing products and services have been increasingly less relevant in developed countries. According to them, the financial institutions are now the facilitators of risk transfer and deal with the more complex world of financial products and markets. Thus, risk management has become a vital aspect of business activity.

The prime task of the financial institutions, from a macroeconomics point of view, is to raise sufficient funds to meet the required level of investment in the economy. Looking at the Keynesian circular flow of income, financial institutions act as intermediary units, i.e. as mediators of funds between the surplus units and the deficit units. Failure to do so may lead to an investment gap that necessitates the inflow of foreign capital. This task is performed by providing various forms of financial instruments ranging from the so-called traditional forms such as deposits to the most sophisticated ones such as bonds, private securities etc. However, in raising such funds, the financial institutions incur costs, such as interest payments to interest-bearing instruments and other transaction costs. If we want to formulate a cost function for the financial institutions, such costs should be included. In addition, financial institutions can also be seen as *production units.* In this case, the role of the bank is to provide sufficient financial services to its customers, i.e. exactly the same as what other producers are doing in the product market.

The performance of the financial institutions is crucial, particularly in terms of efficient mobilisation of funds. This has attracted the interest of many researchers. At present, most of the studies undertaken focus on developed economies, such as those of Cavallo and Rossi (2002), Elyasiani and Rezvanian (2002), Maudos et al. (2002), Altunbas and Molyneux (1996) and Molyneux and Forbes (1993). The developing and Far East countries too have not been left untouched, such as in Rezvanian and Mehdian (2002), Hardy and di Patti (2001), Abdul_Karim (2001), Edwards (1999), Khatib and Matthews (1999), Chu and Lim (1998), and Bhattacharyya et al. (1997). However, compared to the number of studies which have focused on developed countries, relataively few studies have focused on developing countries and the Far East. Berger and Humphrey (1997) surveyed 130 studies that have employed frontier analysis in 21 countries. Out of these studies, only eight were done in developing and Asian countries (including two in Japan). Studies focusing on US financial institutions were the highest, accounting for 66 out of 116 single country studies. Another recent survey was done by Drake (2004). He, however, focused on three important banking markets, the US, the UK and Japan. In his survey, he highlighted some of the methodologies used like nonparametric frontier approaches (data envelopment analysis and free disposal hull) and parametric approaches (stochastic frontier models, distribution free approach and thick frontier approach).

In fact, there is still some room for exploration. For example, using a non-parametric approach to analyse the performance of commercial banks in

Malaysia, Khatib and Matthews (1999) excluded foreign banks in their study due to the unavailability of data. As a consequence, the relative position of the local banks in terms of efficiency and performance, as compared with the foreign banks, is unknown. Are the local banks at par, better off or worse off than their foreign counterparts? In addition, we cannot determine whether such findings can still stand if we employ alternative techniques such as the parametric approach. Abdul_Karim (2001) and Edwards (1999), on the other hand, focused on the cost efficiency of the banks and the existence of economies of scale, but did not consider the comparison between the domestic and foreign banks. Berger and Mester (1997, pp. 900) suggested that profit efficiency is superior to the cost efficiency concept for evaluating the overall performance of the firm. The reason is simple: profit efficiency is based on the well-known economic goal of profit maximisation. However, they too failed to make a comparison between local and foreign banks. As explained by Elyasiani and Rezvanian (2002), such distinctions should not be ignored, particularly when the presence of foreign owned banks is significant.

In principle, there are several reasons for the performance of banks to be studied, particularly in developing countries. Firstly, since neither capital nor debt markets are well developed and efficient, a substantial part of funds is mobilised through the banking system. This places the survival and the risks of the whole of economy, which needs the funds to finance the required level of investment, in the hands of the bankers. Jomo (2001) argued that one of the reasons for the Malaysian financial crisis in 1997-98, which eventually affected the real sector, was the heavy reliance on the banking sector as the main source of funds in the economy. The performance of the banks, such as their cost efficiency, is therefore the main interest of the researchers and policy makers. Secondly, the domestic banking sector is facing greater and stiffer competition due to globalisation of the financial system. For example, as a member of the World Trade Organisation (WTO), Malaysia has to liberalise its domestic banking sector by removing certain barriers like equity ownership and by allowing a greater number of foreign-owned banks to operate and directly compete with existing domestic-owned banks. This leads to the need to understand the current position of the domestic banks in terms of their performance and efficiency. Thirdly, policy makers such as the monetary authority, on the other hand, are interested to know about the appropriate number and size of the banks in this relatively small but open economy. Should the banks merge if there are too many of them? Can the relatively small domestic banks face competition from the large, well-established foreign banks in the future? Are all banks, regardless of their size, enjoying economies of scale?

In addition, Zhu (2003) stated that there are several advantages of performance evaluation. The first is that it exposes the strengths and weaknesses of the banks or any business operations and activities. It can also assist in meeting customers' needs and identifying potential business opportunities to improve daily activities.

2.2 Related issues: bank input and output and efficiency

Previous studies show that there is considerable variability in the performance of financial institutions. This has been thoroughly discussed by Berger and Humphrey (1997). For example, they showed that the efficiency measures obtained under the nonparametric approach are similar to those from the parametric approach. However, nonparametric methods, on average, produce relatively lower mean efficiency measures and are more likely to have greater dispersion than the results of parametric studies. Sharma et al. (1997) found that the technical efficiency estimates under the parametric approach are higher than those obtained under the non-parametric approach. They studied the swine industry in Hawaii using a cross sectional data set. The technical efficiency estimates were obtained based on a stochastic production function.

The main concern is when the so-called inefficiencies (cost or profit) are quite substantial. According to Berger and Mester (1997), there are at least three main reasons for the differences in the efficiencies of financial institutions. The first is the different concepts of efficiency and data set used. Examples include the studies done by Chu and Lim (1998) and Rezvanian and Mehdian (2002a). Both studies focused on Singaporean banks. Chu and Lim (1998) found that the average cost-efficiency (X-efficiency) of the Singaporean banks was above 95%, a remarkable score. In contrast, Rezvanian and Mehdian (2002a) showed that the banks in their sample were on average 57% overall cost efficient. The second reason is the differences in the measurement methods used to estimate efficiency. Basically, there are two broad approaches that are commonly used by reseachers, i.e. parametric and nonparametric. Under the parametric approach, econometric models are normally used to estimate cost or production functions. In contrast, the nonparametric approach is based on linear programming. Since both use different techniques of estimation, the outcomes are likely to differ. The final reason is the potential correlates of efficiency like bank size, market characteristics (concentrated or saturated) and banking regulation. In other words, efficiency is dependent upon other factors.

Cavallo and Rossi (2002) highlighted two debatable issues concerning the modelling of the cost functions of the banks. Firstly, there is an issue related to the definition of inputs and outputs of the financial institutions. This arises because of different perceptions towards financial institutions, regarding whether they are viewed simply as financial intermediation units or as production units. Empirical studies have shown that many researchers are more interested in using the intermediation approach. It is also found that even if studies use the same approach, the choice of inputs and outputs is still different, due to data availability and topics of interest.

The second issue is whether inputs and outputs should be measured in physical or monetary units. Beck et al. (1999), for example, used nominal and monetary units in constructing the measures of efficiency (net interest margin and overhead cost). They did not deflate the numerator and denominator of these measures for several reasons:

i. there is no obvious deflator for individual banks' assets and income flow

ii. the flows and stocks of individual banks are directly related

iii. financial assets and flows are not the product of quantity times price, as in the national income statistics. However, they calculated the numbers deflated by the Consumer Price Index (CPI) and found that the correlation between real and nominal is high (above 90%). In this case where inflation is constant, using either one of them makes no difference.

Bhattacharyya et al. (1997), on the other hand, used real units when they analysed the productive efficiency of Indian commercial banks. The same approach was taken by Berger and Mester (1997).

2.2.1 Inputs and outputs of commercial banks

The choice of bank inputs and outputs remains debatable. This is due to different perceptions of the ideal function of the bank in an economy. The other reasons are perhaps differences in the focus of study and the types of data available. Siems and Barr (1998) outlined key considerations in choosing appropriate inputs and outputs of the bank. Both must reflect their importance and contribution in attracting deposits and making loans and advances. In their study, they used salary expense, fixed asset expense, interest expense, non-interest expense and the purchased fund as bank inputs. Bank outputs are earning assets, interest income and non-interest income. Chu and Lim (1998) used shareholders' funds, interest expenses and operating expenses (non-interest expense plus provision) as bank inputs. Bank outputs are annual increases in average total assets and income (from interest and non-interest activities).[1]

In the intermediation approach, the choice is based on the bank's assets and liabilities. Bank assets represent outputs and liabilities as inputs. For Berger and Mester (1997), bank inputs are purchased funds, core deposits and labour. Outputs are consumer loans, business loans and securities.[2] Rezvanian and Mehdian (2002a) applied the same method. Inputs are borrowed funds (time deposits and other borrowed funds) and other inputs (labour and capital). Outputs are total loans, securities and other earning assets. Cavallo and Rossi (2002) also treated labour, capital and deposits as bank inputs.

In contrast, the production approach considers bank inputs to be only physical entities such as labour and capital. In relation to the role of deposits, its proponents argue that all deposits should be considered as output, since they

[1] The variable annual increase in average total asset has never before been employed in the literature. Its inclusion is required since the authors want to determine the effect of bank efficiency on the share price. The variable is a proxy for future income and future profit. This shows how the focus of study determines the selection of input and output.

[2] In comparing banks' efficiencies, Berger and Mester (1997) argue that such comparison is between banks producing the same output quality. However, it is most likely that differences in quality cannot be measured. In other words, heterogeneity in bank output is unlikely to be captured.

are associated with liquidity and safekeeping, and are involved in generating value added.[3]

Since differences in the choice of inputs and outputs may eventually affect the scores of cost efficiency, some studies treat deposits firstly as input and secondly as output in the cost function. They then compare the results. Favero and Papi (1995), for example, found that results are not significantly affected by the different definitions used.

2.2.2 Efficiency Issues

Efficiency issues arise when we evaluate the performance of the firms in transforming multiple inputs into multiple outputs. As already known, this transformation is called a production process. To economists, a production function specifies the maximum possible output given a set input. This means that the maximum possible output establishes itself as a boundary or frontier. A best practice firm or bank is one whose observed output lies on the frontier. If the production of an observed output lies below or deviates from the frontier, it reflects production or technical inefficiency. Such deviations can be seen in two ways:

i. in terms of output, i.e. the maximum possible output (Y_{max}) minus the observed output (Y), ($Y_{max} - Y \geq 0$). A positive outcome indicates deviation or inefficiency.

[3] The intermediation approach instead stresses that deposits should be taken as inputs, since they become the raw material that will be transformed into loans and investments.

ii. in term of cost, i.e. the observed cost (C) minus the minimum cost (C_{min}), ($C - C_{min} \geq 0$). A positive outcome shows deviation or inefficiency.

In general, efficiency falls into two categories, technical and allocative efficiency. Technical efficiency (also known as productive efficiency) can be defined as follows;

i. the ability to produce the maximum level of output, given a level of input, or
ii. the ability to employ the minimum level of input, given a level of output.

This requires a specification of a production function. On the other hand, allocative efficiency refers to the ability of the bank or firm to combine optimally various inputs needed for producing output, given the input prices. Optimal combination occurs when the marginal rate of substitution between inputs matches the corresponding input price ratio.

Having discussed the categories of efficiency, another question that needs to be addressed is which concept of efficiency should be used? Is it economic or technical efficiency? Is it productive or allocative efficiency? Nevertheless, is it a sufficient condition to base our discussion on the performance of the bank by only referring to cost or production? In relation to this issue, Tobin (1984) proposed at least four separate concepts of efficiency that can be applied. These are are information arbitrage, fundamental valuation, full insurance, and functional efficiency. Clearly, it is the last category that deals with the basic functions of the financial institutions, i.e intermediating between savers and investors and facilitating and administering the payments process. This further involves risk pooling, allocation of resources etc. Berger and Mester (1997) considered the three most important economic efficiency concepts – cost, standard profit and alternative profit efficiencies. The choice of these three is based on economic optimisation following changes in the market.

Another important concept is cost efficiency. If technical efficiency is estimated based on a stochastic production function, cost efficiency is derived from a stochastic cost function. It provides a measure of how close a bank's cost is to the cost of a so-called best practice bank producing the same type of output under the same market conditions. As known, a bank is a multi-product firm, producing a vector of output from a vector of input. Based on duality theory, its cost function can be simply written as:

$$C = f\,(Y, W)$$

where C is total cost, Y is a vector of output and W is a vector of input price.

Berger and Mester (1997) stated that in the bank cost function, its variable costs depend on the prices of variable inputs, the quantities of variable outputs and any fixed inputs or outputs, environmental factors etc.

2.2.3 Efficiency Measurement Issues

Frontier analysis is about benchmarking the relative performance of decision-making units (hereinafter DMUs). It provides an overall numerical efficiency value and ranking of firms (normally called X-efficiency in economic analysis). X-efficiency is a gap between actual and minimum attainable supply cost. To be efficient, firms must operate on their minimum attainable cost curves. With X-inefficiency, the actual cost lies above the minimum attainable cost. This concept is shown in Figure 2.1.

Figure 2.1: X-inefficiency

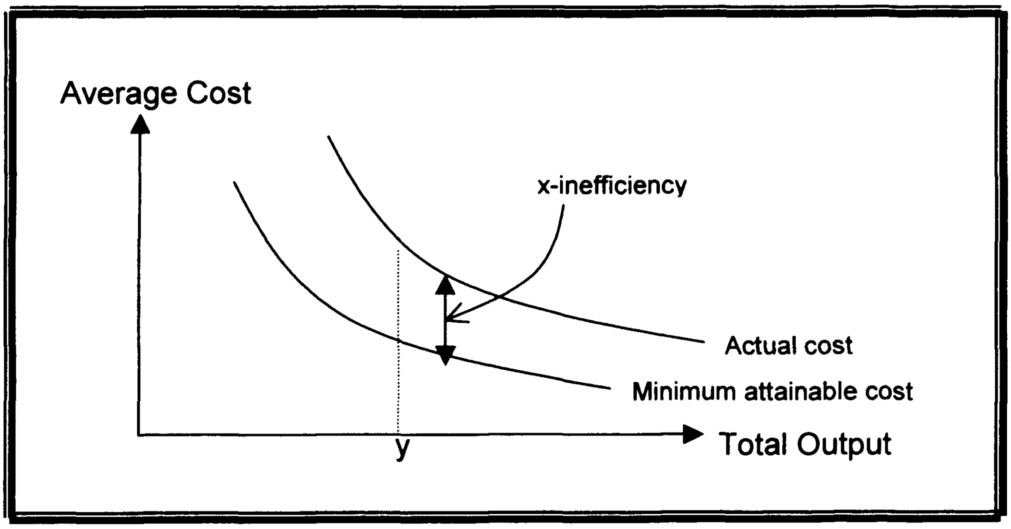


Figure 2.1 shows that the minimum attainable cost curve experiences economies of scale. Since a firm's actual cost curve lies above the minimum level, the vertical distance between the two represents the X-inefficiency or cost inefficiency. This implies the firm's inability to produce a given level of output at the minimum level of cost.

Studies on the performance of the DMUs can be broken down into two broad categories, parametric and non-parametric.

2.3 Parametric approach.

This approach is based on the econometric estimation of a stochastic cost or production function. It can be further divided into three techniques, which are different to each other with respect to distributional assumptions made under each technique.

2.3.1 Stochastic frontier approach

The stochastic frontier approach (hereinafter SFA) was introduced by Aigner et al. (1977) and Meeusen and van den Broeck (1977). A frontier production function is defined as the maximum possible output being a function of a certain input. This frontier production function can be expressed as:

$$y_i = f(x_i, \beta) \qquad \text{Equation 2.1}$$

where y_i is the maximum possible output produced, x_i is a vector of non-stochastic inputs and β is the unknown parameter to be estimated. i runs from 1 to N.

Aigner et al. (1977) further stated two problems that need to be explored. The first is related to variability in the production of output by the firms facing identical input vectors. The second is concerned with the position of the firms below the frontier ($y_i < f(x_i, \beta)$). According to them, the above problem is due to the presence of disturbances. This, therefore, requires the implicit assumption of a disturbance term or error term (ε_i) in the frontier function. The structure of the error term is written as:

$$\varepsilon_i = v_i + u_i$$

This means that the frontier function now becomes stochastic.

$$y_i = f(x_i, \beta) + \varepsilon_i \qquad \text{Equation 2.2}$$

They argued that the production process is subjected to random disturbances with different characteristics. u_i is assumed to be a non-positive disturbance ($u_i \leq 0$). The inclusion of this term is justified on the grounds that each firm is either on or below the frontier, [$f(x_i, \beta) + v_i$]. If the observed output lies below its frontier [$y_i < [f(x_i, \beta) + v_i]$], a deviation from the best practice level has taken place. Such deviation may be due to the presence of internal disturbances or controllable factors such as technical and economic inefficiency, etc. Thus, u_i represents the inefficiency term.

On the other hand, v_i is regarded as random disturbance. It can be either positive or negative ($v_i \geq 0$ or $v_i \leq 0$). It originates from the fact that external events or uncontrollable factors like luck and climate can cause the frontier to vary randomly across the firms, or vary over time for the same firm. It may also come from errors of observation and measurement of the observed output.

By considering a linear model, a stochastic production function, given a set of cross sectional data, can be written as:

$Y_i = X_i\beta + \varepsilon_i$. Since $\varepsilon_i = v_i + u_i$ then,

$$Y_i = X_i\beta + v_i + u_i. \qquad \text{Equation 2.3}$$

In the case of cost efficiency, following the definition set by Berger and Mester (1997), the frontier cost function can be written as:

$$C = C(w,y,z,u_c, v_c) \qquad \text{Equation 2.4}$$

where C represents variable cost, w is the vector of prices of variable inputs, y is the vector of quantities of variable outputs, z is a set of environmental or market variables that may influence the performance of DMUs, u_c denotes an inefficiency factor that may increase costs above the best practice level or the frontier, and v_c denotes random error that takes into account measurement error, which may result in higher or lower bank costs.

To simplify matters, the inefficiency and random terms ($u_c + v_c$) are assumed to be separable from the rest of the function. The function is then re-expressed in natural logarithms as:

$$\ln C = f(w,y,z,) + \ln u_c + \ln v_c \qquad \text{Equation 2.5}$$

Again, the term, $\ln u_c + \ln v_c$ is considered as a composite error term. Various attempts to estimate u_c and v_c have led to various measurement techniques. The main challenge in the parametric approaches is the ability of a relevant technique to disaggregate the composite error. Furthermore, the problem faced in distinguishing the two leads to complications in measuring cost efficiency (Allen and Rai, 1996).

Based on the above frontier function, the cost efficiency (CE) is then estimated. For a typical bank, let us say Bank X , CE_x can be written as:

$$CE_x = \frac{\hat{C}_{min}}{\hat{C}_x}$$

The numerator and denominator represent a best practice bank's and Bank X's estimated cost respectively. The cost efficiency scores range between 0 and 1. For a best practice bank, the score is 1. For example, if CE_x is 0.8, then Bank X is 80% efficient or similarly wastes 20% of its cost relative to a best-practice bank facing the same market environment.

Following Jondrow et al. (1982), the inefficiency measure for a cross sectional data set is derived from the parameter estimates of a stochastic cost function under the assumption that the error terms are half-normally distributed. The inefficiency measure is given by:

$$E(u_i / \varepsilon_i) = \left[\frac{\sigma\lambda}{1+\lambda^2}\right]\left[\frac{\varphi\left(\frac{\varepsilon i \lambda}{\sigma}\right)}{\psi\left(\frac{\varepsilon_i \lambda}{\sigma}\right)} + \frac{\varepsilon_i \lambda}{\sigma}\right]$$ Equation 2.6

where $\lambda = \sigma_u / \sigma_v$ and $\sigma = [\sigma_u^2 + \sigma_v^2]$.

Another advantage of using the stochastic approach is that throughout the analysis, it generates several by-products. These include the estimates of elasticities of substitution, economies of scale and scope and technological change. However, the estimates of ε (error term) are of particular interest (Greene, 1993). Economies of scale, for example, can be calculated by taking the coefficients of the output variable in the cost function. Technological change can be shown by including a time dummy in the function.

Studies of stochastic cost efficiency can be found in Girardone et al. (2004), Cavallo and Rossi (2002), Rezvanian and Mehdian (2002b), Abdul_Karim (2001), Christopolous and Tsionas (2001), Lothgren (2000), Edwards (1999), Maudos (1998), Berger and DeYoung (1997), Altunbas and Molyneux (1996), and Karafolas and Mantakas (1996) .

2.3.2 Distribution-free approach

The distribution-free approach was introduced by Berger (1993) following his criticism of the stochastic frontier approach. The central argument is about the distributional assumptions of the error term. The term “distribution-free” denotes that this approach allows the coefficients to vary over time besides using panel data. It also does not make any strong assumptions regarding the specific distribution of the inefficiencies or random errors. In comparing the stochastic frontier approach and the distribution free approach, Allen and Rai (1996; pp. 660) wrote that:

> "the stochastic cost frontier approach utilises a half-normal distribution under the assumption that X-inefficiency is an asymmetric and increasing but never decreasing cost component. In contrast, the distribution-free model notes that random errors should average out over time leaving a systematic component which is assumed to be time-independent".

The stochastic cost function for firm *i* is written as:

$$C_{it} = c_t(Y_{it}, W_{it}) u_t v_t \qquad \text{Equation 2.7}$$

where;

C_{it} = total cost of firm *i* in year t

Y_{it} = bank outputs (loans and investment assets)

W_{it} = input prices (price of labour, price of capital and interest cost)

u = one-sided inefficiency measure (half-normal distribution)

v = two-sided random noise (normal distribution)

In the above equation, the error component is a multiplicative X-inefficiency factor. By taking the logs, it is written as:

$$\ln C_{it} = \ln c_t (Y_{it}, W_{it}) + \ln u_t + \ln v_t \,. \qquad \text{Equation 2.8}$$

u_t or the inefficiency term remains constant and persistent over time, while other coefficients and variables vary. An average of the residuals, denoted by the ln $u_{it,}$ is then estimated. Since random errors (v_t) are said to phase out in the long term, $\ln u_{it}$ is an estimate of $\ln u_t$. With regards to this, DeYoung (1997a) suggested that six years' worth of data are adequate to reasonably ensure that estimated X-efficiency contains only a small amount of random errors, and that using eight or more years' worth of data may violate the central DFA assumption that cost inefficiency remains unchanged over time.

The inefficiency measure under the distribution free approach can be derived as follows:

$$INEFF_{DFA} = \exp(\min(\ln u_t) - \ln u_{it})$$

$$= \min [u_t] / u_{it} \qquad \text{Equation 2.9}$$

where min ($\ln u_t$) is the minimum $\ln u_{it}$ for the period of estimation t.

According to Allen and Rai (1996), the inefficiency estimate can also be measured as follows:

$$INEFF_{DFA} = E(u_i/\varepsilon_i)$$

They found that the distribution-free approach overestimates the magnitude of the X-inefficiency relative to the stochastic frontier approach. However, DeYoung (1997b) argued that this finding was based on a faulty interpretation of the X-inefficiency estimates from both approaches (SFA and DFA). According to him, Allen and Rai (1996) had mistakenly interpreted the result as being a measure of relative cost inefficiency. Hence, the DFA should not be rejected in favour of the SFA.

Studies using this approach include the work of Hardy and di Patti (2000), Berger and Mester (1997), Bauer et al. (1998) and Pastor et al. (1997).

2.3.3 Thick frontier approach

The thick frontier approach (hereinafter TFA) was introduced by Berger and Humphrey (1991). In the TFA, firms or banks are categorised or grouped by size. Then a frontier is formed based on the lowest quartile of the average costs in each of several size categories. Caudill (2002) argues that the TFA contains several elements of arbitrariness, including the appropriate number of size categories to use and the theoretical justifications for using the lowest quartile instead of some other quartile. This method is less popular amongst researchers. Examples of the studies that have employed this technique are Lozano-Vivas (1997) and Mahajan et al. (1996).

2.3.4 Panel data models

Panel data refers to a set of data in which each observation in the sample is repeated over a period of time. It is a rich dataset, since it can explain variations across the units and over time periods. Under the static model, there are two models that can be used to estimate efficiency measure, namely the fixed effect model and the random effect model. The main difference between the two lies in their assumptions about the effect. Battese and Coelli

(1995) developed a stochastic frontier production function for panel data. This frontier function is written as:

$$\ln Y_{it} = \beta_0 + \beta_{1t} \ln X_{1t} + \beta_{2t} \ln X_{2t} + \beta_{3t} \ln X_{3t} + \beta_{4t} \ln X_{4t} + \beta_{5t} \ln X_{5t} + \beta_{6t} T + v_{it} - u_{it} \qquad \text{Equation 2.10}$$

where

Y_{it} = total output produced in period t

X_{it} = ith input in period t

T = year dummy variable

The estimation procedure of efficiency for panel data was given by Kumbhakar (1987) and Battese and Coelli (1988).

2.3.5 Functional Forms for Parametric Approach

When a parametric model is constructed to estimate efficiency, the next task is to choose an appropriate functional form. In the literature, the most popular functional form is the translog[4]. It was Christensen et al. (1973) who introduced the translog production function. They wrote the frontier production function as:

$$\ln F = \alpha_0 + \alpha_1 \ln Y_1 + \alpha_2 \ln Y_2 + \alpha_3 \ln X_1 + \alpha_4 \ln X_2 + \alpha_5 \ln A$$
$$+ \ln Y_1 (0.5\, \beta_{11} \ln Y_1 + \beta_{12} \ln Y_2 + \beta_{13} \ln X_1 + \beta_{14} \ln X_2 + \beta_{15} \ln A)$$
$$+ \ln Y_2 (0.5\, \beta_{22} \ln Y_2 + \beta_{23} \ln X_1 + \beta_{24} \ln X_2 + \beta_{25} \ln A)$$

[4] Edwards (1999, pp.17-18) summarised the functional forms applied to the cost function starting from Cobb-Douglas, constant elasticity of substitution, translog, generalised translog, generalised Leontif and quadratic.

$+ \ln X_1 (0.5\ \beta_{33} \ln X_1 + \beta_{34} \ln X_2 + \beta_{35} \ln A)$

$+ \ln X_2 (0.5\ \beta_{44} \ln X_2 + \beta_{45} \ln A)$

$+ \ln A (0.5\ \beta_{55} \ln A)$ Equation 2.14

where

F = the production possibility frontier

Y_1= consumption (regarded as output)

Y_2 = investment (regarded as output)

X_1 = capital (regarded as input)

X_2 = labour (regarded as input)

A = an index of technology

Using duality theory, the multiproduct cost function can be written as:

$$C = f(Y,W)$$

where C is the total cost, Y is a vector of outputs and W is a vector of input prices.

Cavallo and Rossi (2002) used 3 outputs and 3 inputs. The translog cost function for *firm s* is written as:

$$\ln C_s = [\ \alpha_0 + \sum_{i=1}^{3} \alpha_i \ln y_{is} + \sum_{k=1}^{3} \beta_k \ln w_{ks} + 0.5 \sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_{ij} \ln y_{is} \ln y_{js}$$

$$+ 0.5 \sum_{k=1}^{3}\sum_{h=1}^{3} \beta_{ks} \ln w_{hs} + \sum_{i=1}^{3}\sum_{k=1}^{3} \delta_{ik} \ln y_{is} \ln w_{ks}\ ] + v_{st} + u_{st}$$

Equation 2.15

where:

C_s = total cost (operating costs + interest expenses) for *firm s*

y_i = the *i*th output (loans, deposits and financial investments)

w_i = the price of the *k*th input (labour, capital and deposits)

v_{st} = random variables assumed to be iid $N(0,\sigma_v^2)$

u_{st} = non-negative random variables assumed to represent cost inefficiency.

Berger and Mester (1997) argued that the translog cost function is not necessarily a good fit for data that are far from the mean in terms of output size or mix. They believed that some of the differences in results on efficiency estimates may be due to the poor fit of the translog function. A more flexible form is needed, like the Fourier-flexible functional form (hereinafter the FFF). The FFF was introduced by Galant (1982). It simply extends the translog functional form by including Fourier trigonometric terms. This means that the FFF specification contains Sines and Cosines of the log of output. This functional form is said to represent a semi-nonparametric approach to the problem of using data so as to deduce relationships among variables when the true functional form of the relationship is unknown (Mitchell and Onvural, 1996). The form is also more flexible than the translog and is a global approximation to any cost or profit function. It is capable of providing a better fit to data. The FFF specification as an extension of the translog functional form can be seen below as employed by Altunbas and Chakravarty (2001).

$$\ln C = \alpha_0 + \sum_{i=1}^{3} \alpha_i \ln Y_i + \sum_{l=1}^{3} \beta_l \ln W_l + 0.5\left[\sum_{i=1}^{3}\sum_{j=1}^{3} \delta_{ij} \ln Y_i \ln Y_j + \sum_{l=1}^{3}\sum_{m=1}^{3} \gamma_{lm} \ln W_l \ln W_m \right]$$

$$+ \sum_{l=1}^{3}\sum_{i=1}^{3} \rho_{li} \ln W_l \ln Y_i + \sum_{i=1}^{3} \left[a_i \cos(Z_i) + b_i \sin(Z_i) \right] + \sum_{i=1}^{3}\sum_{j=1}^{3} [a_{ij} \cos(Z_i + Z_j)$$

$$+ b_{ij}\sin(Z_i + Z_j)] + \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{k\geq 1 k\neq 1}^{3}[a_{ijk}\cos(Z_i + Z_j + Z_k) + b_{ijk}\sin(Z_i + Z_j + Z_k)] + \varepsilon$$

Equation 2.16

where Z_i = the adjusted values of the log output ln Y_i. α_l, β_l, δ_{ij}, γ_{lm}, ρ_{il}, a_i, b_i, a_{ij}, and b_{ij} are coefficients to be estimated.

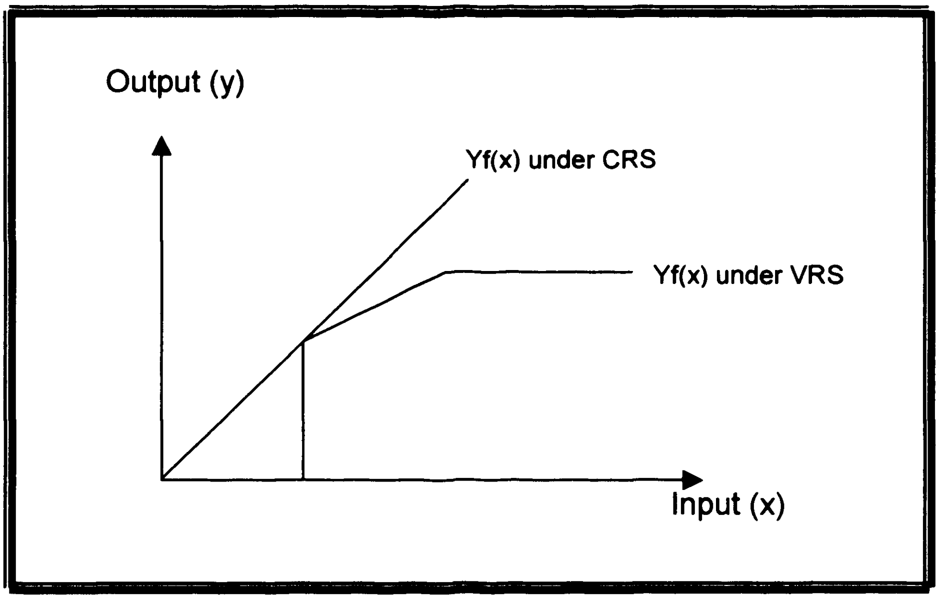
Berger and DeYoung (1997) found that the measured inefficiencies were about twice as large when using the translog in place of the Fourier-flexible form. However, some researchers are cautious about the FFF. Edwards (1999) believed that one main challenge that one needs to face is the large of number of coefficients in the function. Altunbas and Chakravarty (2001) are sceptical about the goodness of fit associated with the FFF. They argued that goodness of fit is not always the most reliable criterion for the selection of a model. Furthermore, it is not necessarily an indicator of goodness in the prediction. Nevertheless, their underlying critique rests on the proposition that mechanical approaches such as the inclusion of Fourier trigonometric terms are unlikely to generate any economic insights into the banking system.

2.4 Non-parametric approach.

In attempting to measure productive efficiency, Charnes et al. (1978) introduced a mathematical programming approach, which aimed to construct a set of benchmark or best-practice frontiers. This approach later became known as data envelopment analysis (hereinafter DEA). Unlike the parametric approach, this linear programming is non-stochastic i.e. total deviations from the best-practice frontier constitute total inefficiency. It is in this way that it differs from the stochastic approach. Under the latter, total deviation or the error term is composed of the inefficiency term and random error. The DEA method transforms multiple inputs and multiple outputs into a scalar measure of relative productive efficiency. It measures the efficiency of a DMU (decision-making unit) relative to other similar DMUs with simple restrictions.[5] Its goal is to 'envelop' the data as a linear combination that connects the set of best-practised observations yielding a convex production possibilities set. Charnes et al. (1978) introduced the first DEA model under the assumption of constant returns to scale (CRS). Under this assumption, the production frontier is a straight line. Their work was later improved upon by Banker et al. (1984) to allow variable returns to scale (VRS). Under this model, the frontier is concave. In general, inefficient firms are the ones whose observed production points lie below the frontier. Figure 2.2 shows the production frontiers made under both assumptions.

[5] Denizer et al. (2000) put forward two restrictions. Firstly, all DMUs must lie either on or below the efficiency frontier. Secondly, the efficiency score for all DMUs must be less than or equal to one. The most efficient DMU has a score of one.

Figure 2.2: The production frontier



Since the production frontier under the CRS assumption is a straight line, the number of efficient firms will be less and the efficiency scores are relatively low. On the other hand, the production frontier under the VRS assumption will produce higher efficiency scores and a higher number of efficient firms, according to Avkiran (1999a) and Bowlin (2002).

Under DEA, the relative productive efficiency of a DMU is defined as a ratio of its total weighted output to its total weighted input. It is written as:

$$EFF_k = \frac{\sum u_{rk} \ OUTPUT_{rk}}{\sum v_{ik} \quad INPUT_{ik}} \qquad \text{Equation 2.11}$$

where *u* and *v* are weights assigned to each output and input respectively (*u* and *v* >0).

A DMU, let us say *k,* is said to maximise its productive efficiency. Thus, the objective function can be expressed as :

$$\text{Max. EFF}_k = \frac{\sum_{r=1}^{s} u_r y_{rk}}{\sum_{i=1}^{m} v_i x_{ik}} \text{ subject to: } \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} x_i x_{ij}} \leq 1 \quad \text{for } j = 1,2,\ldots,n.$$

Equation 2.12

The above objective function is the basic model of DEA. Charnes et al. (1978) showed how to transform this fractional function into a linear function.

Once the best-practice frontiers are derived, several efficiency indices can be calculated for each DMU or bank in the sample. The first efficiency index is overall technical efficiency (OTE), which will be decomposed into pure technical efficiency (PTE) and scale efficiency (SE).[6] Through this decomposition, it is argued that we are able to pinpoint the sources of inefficiency of each DMU or bank in the sample. The overall efficiency can be expressed or summarised as follows:

OTE = PTE*SE Equation 2.13

Another advantage of the DEA is that it does not impose any functional form (as needed under the parametric approach) and error structure on data. Thus,

[6] See Rezvanian and Mehdian (2002) on how each of the indices is measured. Also see Favero and Papi (1995) for a graphical illustration of the measurement of efficiency.

there is no requirement for a strong distributional assumption. It also avoids the possibility of multi-collinearity amongst the variables included in the model.

It should be noted that there are various methods and models currently used within DEA since the original works of Charnes et al. (1978) and Banker et al. (1984). For example, Adler et al. (2002) identified six main methods found in the DEA literature. The first method relates to the evaluation of cross efficiency amongst the DMUs and their peers. The second is a super-efficiency method that ranks an extreme efficient unit to an efficiency score greater than one by removing the unit's constraint in the primal function. The third is based on benchmarking, i.e. using efficient DMUs as a benchmark for inefficient DMUs. The fourth is related to the utilisation of a multivariate statistical technique in order to achieve a complete ranking. The fifth method is to rank inefficient DMUs through proportional measures of inefficiency. The last one is a multi-criteria decision-making method. This method requires the collection of additional, preferential information from relevant decision-makers plus other criteria. However, the suitability of each model depends on the specific objectives of the study. Further discussion of these models can be found in Zhu (2003).

This method is quite popular amongst researchers . As surveyed by Berger and Humphrey (1997), 69 of the 130 studies are found to employ this technique. Another 60 studies use the stochastic frontier approach. Studies using a non parametric approach include Yildirim (2002), Rezvanian and Mehdian (2002), Denizer et al. (2000), Katib and Matthews (1999), Laevan

(1999), Chu and Lim (1998), Resti (1998), Siems and Barr (1998), Bhattacharyya et al. (1997), Resti (1997), Sharma et al. (1997) and Favero and Papi (1995). However, Cornwell and Schmidt (1996) and Greene (1993) did not view DEA as an attractive method for technical efficiency estimation. They all argued that since DEA is non-statistical, the properties of its inefficiency estimates cannot be determined. Secondly, since it is non-stochastic, it is not possible to disentangle inefficiency from random noise. The efficiency measures may be biased if the production process is largely influenced by stochastic elements.

2.4.1 Malmquist productivity growth

The Malmquist productivity index is a number that enables a firm's productivity to be compared over two different time periods. This idea was introduced by Malmquist (1953) who demonstrated the calculation of quantity indices as a ratio of distance function. Grosskopf (2003) recalled the emergence of the Malmquist productivity index by referring to the early work by Caves et al. (1982). They stated that the Malmquist productivity index can be constructed from a ratio of the value of input distance functions before and after a specific event or in different time periods where the earlier time period is taken as the base reference point. In addition, as productivity index is defined based on distance functions, activity analysis can be used to estimate the index directly (Grosskopf, 2003). It was Färe et al. (1994) who initially developed the

Malmquist productivity index using DEA approach based on constant returns to scale.

Given a panel dataset, the annual efficiency estimate can be obtained by using the DEA technique. However, because of the time dimension (present and future), each year's efficiency can be compared to each other. One cannot conclude that efficiency has improved simply because the efficiency estimate in period t+1 is higher than those in period t. The problem is that between period t and t+1, technology might have improved too. With the introduction of the Malmquist productivity index, one can identify the sources of productivity growth, namely efficiency change and technical change. Efficiency change refers to improvement in efficiency, given the same technology or the same production frontier. Technical change refers to improvement in technology, given the same efficiency level. This is indicated by the outward shift of the production frontier.

Färe et al. (1994) analysed productivity growth in 17 OECD countries for the period 1979-1988. The decomposition of productivity growth was done under the assumption of constant returns to scale. They further stated that this decomposition provides an alternative way of testing for convergence of productivity growth, as well as allowing identification of innovation (p.72). They found that the average change in the Malmquist productivity index was less than one percent per year for the whole sample. On average, the growth was due to technical change (innovation) rather than improvement in efficiency.

Canhoto and Dermine (2003) studied the magnitude of efficency gains of Portugese banks for the period of 1991-1995. They first estimated efficiency measures using DEA and then calculated the Malmquist productivity index under constant returns to scale assumption. The results show that productivity had improved by 59% and this was contributed to by improvement in technology rather than improvement in efficiency.

Another study by Casu et al. (2003) compared parametric and non-parametric estimates of productivity change in European banking in five countries for the period 1994-2000. To our knowledge, this is the first comparison between parametric and non-parametric estimates of productivity change. Parametric estimates of productivity growth were initiated by Berger and Mester (1999, 2001). Using the estimates of both cost and profit function, Berger and Mester (1999) decomposed productivity growth into a portion due to changes in business conditions and a portion due to changes in productivity. Later, they decomposed productivity change into the change in best practice and change in inefficiency components. Casu et al. (2003) found that there was clear productivity growth in Italy and Spain, but mixed results for France and Germany. They also observed a favourable trend of productivity growth towards the end of the 1990s. They concluded that the competing methodologies do not yield markedly different results in terms of identifying the main components of productivity growth.

Drake (2001) studied the efficiency and productivity of the UK banks. His sample consisted of nine UK banks and covered the period 1984-1995. He used two models that differ with regard to the role of deposit. The two models reflected different approaches taken in defining bank input and output (production and intermediation approach). Deposit was used either as input or output. Due to a small sample size, the reference frontier was constructed on a three-yearly basis instead of one. He found that the model based intermediation approach produces lower productivity growth if compared with the one generated by the model based production approach. He also observed that the productivity growth was the net result of a mixture of positive technical change and deterioration in efficiency for most of the banks.

Other studies of productivity growth of the banks include Sturm and Williams (2004), Krishnasamy et al. (2003), Alam (2001), Murkherjee et al. (2001), Stiroh (2000), Worthington (1999), Wheelock and Wilson (1999) and Grifell-Tatje and Lovell (1997). Krishnasamy et al. (2003) studied the Malaysian banks' productivity for the period 2000-2001 (after the major bank consolidation). Using a sample of ten domestic banks, they found that the banks' productivity had improved by about 5%. The growth in productivity was contributed by technological change rather than technical efficiency change. However, it must be noted that their study is small scale and ignores the long presence of foreign banks.

2.5 Results of previous studies

In their survey, Berger and Humphrey (1997, pp.181-184 (Table 1)), found that average cost efficiencies of the commercial banks range from 70 to 90% in the US, about 90% in UK and about 80% in Spain. Table 2.1 shows the results of other studies.

2.6 Profit efficiency

Another concept of efficiency is analysed in terms of profit. Profit efficiency is about performance evaluation in terms of the extent to which firms are able to arrive at the maximum profit. Similar to cost efficiency, this is a relative concept by comparing the profit level of a firm with respect to the level of profit obtained by other firms. Studies on the profit frontier are relatively new. To the best of our knowledge, the works of Berger (1993, 1995) can be considered as the pioneering studies. Theoretically, profit frontier refers to the maximum possible profit that a firm can obtain given a level of input prices and output prices or a level of output and a sample of firms facing the same environment. If the firm is able to arrive at this frontier, then it is regarded as the most efficient firm or the best practice firm. However, if the firm is unable to reach that frontier, then a deviation has taken place. In this context, that deviation is called profit inefficiency. According to DeYoung and Nolle (1996), profit inefficiency can be expressed in two ways. The first is as a ratio of inefficiency to total assets. The second is as a ratio of inefficiency to potential variable profit. Variable profit is that which each firm would have generated had it been free of any inefficiency.

Profit efficiency can be negative, since there is a possibility that a firm can throw away more than 100% of its potential profit or it can simply suffer from losses. Under a parametric approach, the deviation from the frontier is represented by a disturbance term in a profit equation. Furthermore, since the profit frontier takes into account both input and output, it reflects the joint impact of revenue and cost.

Table 2.1: Summaries of the findings on the cost and technical efficiency

Authors	Methods[a] used	Types of institutions	Country	Mean efficiency estimates[b]
Abdul-Karim (2001)	SFA	Commercial banks	Malaysia	0.95
Alam (2001)	DEA	Large banks	US	0.92
Avkiran (1999b)	DEA	Trading banks	Australia	0.79 - 0.91 (Model A) 0.37 - 0.79 (Model B)
Beccalli (2004)	SFA	Investment firms	UK and Italy	0.71 and 0.59 (UK) 0.58 and 0.44 (Italy)
Berger et al. (1997)	SFA	Branches of a large bank	US	0.87(intermediation approach) 0.66 (production approach)
Bhattacharyya et al. (1997)	DEA	Commercial banks	India	0.80
Bonin et al. (2005)	SFA	Commercial banks	Europe	0.73
Canhoto and Dermine (2003)	DEA	Commercial banks	Portugal	0.8
Chen (2001)	DEA	Banks	Taiwan	0.78, 0.76
Chu and Lim (1998)	DEA	Commercial banks	Singapore	0.95
Cuesta and Orea (2002)	SFA	Savings banks	Spain	0.91(non-merged) 0.86 (merged)
Denizer et al. (2000)	DEA	Commercial banks	Turkey	0.59 – 0.87
Drake (2001)	DEA	Commercial banks	UK	0.92
Drake and Hall (2003)	DEA	banks	Japan	0.78
Edwards (1999)	SFA	Commercial banks	Malaysia	0.9
English et al. (1993)	SFA	Small banks	US	0.75
Favero and Papi (1995)	DEA	Commercial banks	Italy	0.91
Fiordelisi and Molyneux (2004)	DEA	Factoring industry	Italy	0.90 – 0.95
Garden and Ralston (1999)	DEA	Credit unions	Australia	0.65 – 0.70
Girardone et al. (2004)	SFA	Commercial, Savings, Popular banks and Credit Co-operatives	Italy	0.80 – 0.90

Authors	Methods[a] used	Types of institutions	Country	Mean efficiency estimates[b]
Greene and Segal (2004)	Random effects	Insurance companies	US	0.80
Isik and Hassan (2002)	DEA, SFA	Commercial banks	Turkey	0.92 (DEA) 0.76 (SFA)
Katib & Matthews (1999)	DEA	Commercial banks	Malaysia	0.68 - 0.82
Laevan (1999)	DEA	Commercial banks	Malaysia Indonesia Korea Philippines Thailand	0.7 0.61 0.89 0.68 0.78
Maudos and Pastor (2003)	DEA	Commercial and savings banks	Spain	0.91 (commercial) 0.80 (savings) 0.87 (total)
McKillop et al. (2002)	DEA	Credit unions	UK	0.74 – 0.82
Miller and Noulas (1996)	DEA	Large commercial banks	US	0.96
Murkherjee et al. (2001)	DEA	Large commercial banks	US	Above 0.9
Pastor et al. (1997)	DEA	Commercial banks	Europe and US	0.54 - 0.95 (Europe) 0.81 (US)
Resti (1998)	DEA	banks	Italy	0.85 – 0.97
Rezvanian & Mehdian (2002)	SFA,DEA	Commercial banks	Singapore	0.57 (SFA), 0.86 (DEA)
Sathye (2001)	DEA	banks	Australia	0.67
Shanmugam and Das (2004)	Fixed effects	Commercial banks	India	0.44
Sharma et al. (1997)	SFA, DEA	Swine industry	Hawaii	0.74 (SFA) 0.72 (DEA)
Sturm and Williams (2004)	DEA,SFA	banks	Australia	0.73 -0.94 (DEA) 0.83 (SFA)
Worthington (1998b)	SFA	Building societies	Australia	0.76 – 0.88
Yildirim (2002)	DEA		Turkey	0.96

Notes: a. DEA is data envelopment analysis. SFA is stochastic frontier approach.
b. The DEA score refers to average pure technical efficiency obtained under variable returns to scale assumption.

2.6.1 Profit function

Following the study undertaken by Berger and Mester (1997), there are two main forms of profit function: standard profit function (SPF) and alternative profit function (APF). The two differ with respect to the presence of some market power in the input or output market. The SPF model assumes that the markets for output and input are perfectly competitive. This implies that the firms or banks are price takers, accepting what is already determined by the market, particularly the market interest rate. Based on the definition provided by Berger and Mester (1997, pp.899), the SPF measures how close a bank is to producing the maximum possible profit given a level of input prices and output prices and some other variables.

The APF model is slightly different from the SPF model. The former emerges due to at least three reasons, as noted by Berger and Mester (1997) and subsequent researchers like Lozano-Vivas (1997), Roger (1998) and DeYoung and Hassan (1998). The reasons are as follows:

- some banks, particularly smaller banks, cannot fully control the scale of output to reach any desired level as can larger banks
- some banks, especially larger or well established banks, have some market power to influence current interest rates
- output prices are not accurately measured. For certain bank outputs like fee-based product or transaction services, such prices are not directly available. This will result in a poor measure of inefficiency.

The APF model assumes that output and input markets are not perfectly competitive. This implies that the decision-making units or banks have some market power to influence market interest rates, such as the ability to set their own interest rates. For example, the established banks, through their experience and expertise, seem able to enjoy some market power by offering quality products and services. As far as the large banks are concerned, they can act as price leaders. Their interest rates will be watched and followed by other banks, particularly smaller banks.

Under the APF model, profit efficiency refers to how close a bank comes to earning maximum profit, given its output levels rather than its output prices. The APF employs the same dependant variable as the SPF and the same independent variables as the cost function.

2.6.2 Profit functional form

There are two widely accepted functional forms in the current literature, the translog functional form and the Fourier flexible functional form. As mentioned above, the translog form was introduced by Christensen et al. (1973). This functional form has been applied by Maudos et al. (2002), Van der Vennet (2002) and Roger (1998). Another functional form is the Fourier flexible form. This has been extensively applied to profit frontier analysis by Berger et al. (1997), Berger and Mester (1997) and DeYoung and Hassan (1998). Altunbas et al. (2001) stressed that it has been widely accepted that the global property is important in banking, where scale product mix and other inefficiencies are often heterogeneous. Therefore, local approximations such as those produced by the translog function may provide a poor approximation to the underlying true frontier. The main difference between these two functions is the introduction of the sine and cosine in the Fourier flexible form. It is argued that a linear combination of the sine and cosine can approximate any well-behaved multivariate function.

van der Vennet (2002) acknowledged, in principle, the superiority of the Fourier flexible form over the translog functional form, particularly its ability to move the function closer to the true path of the data. However, he reminds us that while the formal tests indicate that the Fourier terms are jointly significant, the statistical fit and both the average levels of measured efficiency and their dispersion are very similar for both functional forms. His remark is in line with Berger and Mester's (1997) conclusion that the empirical findings in terms of either average industry efficiency or ranking of individual banks are similar across the methods.

2.6.3 Estimation technique

Similar to cost efficiency, profit efficiency can be estimated using three broad approaches: stochastic frontier approach, distribution-free approach, thick frontier approach and panel data models.

2.6.3.1 Stochastic frontier approach

The stochastic frontier approach (SFA) states that the observed profit of a bank may fall below the profit frontier due to the presence of random disturbance (v) or an inefficiency term (u). To disentangle these errors, the inefficiency term is assumed to be half-normally distributed and random error to be normally distributed. Altunbas et al. (2001) and van der Vennet (2002) use this method.

2.6.3.2 Distribution-free approach

The distribution-free approach (DFA) assumes that the inefficiency term is consistent over time and is normally distributed. Given a course of time, random error will phase out, leaving the inefficiency term alone. DeYoung (1997a) suggests that a period of six years is adequate to estimate efficiency that contains only a small amount of random error. The assumption that the inefficiency term is consistent will be violated if using eight or more years' worth of data. This approach is relatively popular amongst researchers and has been employed by Berger and Mester (1997), Maudos et al. (2002), Roger (1998) and Altunbas et al. (2001)

2.6.3.3 Thick frontier approach

The thick frontier approach (TFA) proposes that the differences in the predicted profit within the quartile of banks with the lowest average costs for a given size are due to random errors, while the differences in predicted profit between the quartiles with the lowest and highest profits are due to inefficiency (See Lozano-Vivas, 1997).

2.6.3.4 Fixed and random effect model approach

This approach relies on the availability of panel data. These two models, fixed and random effects, can be used to estimate cost and profit efficiency without having to make specific assumptions about the distribution of the inefficiency term. In the case of the fixed effect model (FEM), the inefficiency term (u) is considered as a constant specific to each firm. An assumption is made that a bank that has the lowest fixed effect is the most efficient one in the sample.

On the other hand, the random effect model (REM) explicitly takes into account the stochastic nature of inefficiency. The inefficiency term is regarded as part of the random term. According to Maudos et al. (2002), REM is similar to DFA, with the difference that in DFA, the coefficients are allowed to vary over time. This is necessary in order to reflect changes in the level of technology and in the environment.

2.6.4 Results of previous studies

Relatively few studies have been carried out on profit efficiency, as compared to the number of studies on cost efficiency. Based on the survey done by Berger and Mester (1997), out of 130 studies, only nine involved profit efficiency. Secondly, it seems that Berger (1991) initiated the first study on profit efficiency. Based on the limited but growing number of studies, in general, profit efficiency ranges between 50% and 70%, relatively lower than cost efficiency (70% to 90%). (See also Berger and Humphrey, 1997.)

Table 2.2 shows the average profit efficiencies as found in the literature. Generally, the average scores are lower than the average cost efficiencies.

Table 2.2: Summaries of the findings on profit efficiency.

Authors	Countries	Technique of Estimation[1]	Profit Efficiency Scores[2]
Berger & Mester (1997)	US	DFA	0.50 (SPE) 0.46 (APE)
Bonin et al. (2005)	Europe (transition countries)	SFA	0.58
DeYoung & Hassan (1998)	US		0.51 (APE)
Lozano-Vivas (1997)	Spain	TFA	0.72 (APE) 0.57 (SPE)
Maudos et al. (2002)	European countries	DFA FEM REM	0.45 (APE) 0.22 (APE) 0.52(APE)
Rogers (1998)	US	DFA	0.70 (APE) 0.65 – 0.68 (APE)

Notes. 1. DFA = distribution –free approach. SFA = stochastic frontier approach. TFA = thick frontier approach. FEM = fixed effects model. REM = random effects model.

2. APE = profit efficiency measure obtained using alternative profit function. SPE = profit efficiency measure obtained using standard profit function.

2.7 Efficiency correlates

As surveyed by Berger and Humphrey (1997), the estimates of cost and profit efficiency varied across banks. Differences in the efficiency estimates do occur, regardless of several efforts made at the methodological stage, such as controlling the concepts of efficiency and methods of measurement and the use of a similar set of data. So, how can these differences be explained? According to Berger and Mester (1997), the actual reasons behind these differences are of great interest to the policy makers and the management of the banks.

Several authors, including Mester (1996), Berger and Mester (1997) and Maudos et al. (2002), propose the exploration of the effects of potential correlates of efficiency, which may theoretically help to explain the observed differences in efficiency across the banks. This exploration will involve a two-step procedure, according to Berger and Mester (1997). The first step is to obtain estimates of the efficiency score using one of the accepted techniques found in the literature. Secondly, the efficiency score is then regressed on a set of variables (potential correlates). A test for correlation between the efficiency score and potential correlates can also be carried out.

With regard to the use of the econometric approach to test the potential correlates of efficiency, Berger and Mester (1997) raise two critical issues. Firstly, the dependent variable (efficiency score) is just an estimate. Secondly, the variables used in the regression are not completely exogenous. This can bias the estimates of the coefficients on all the regressors. In their study, Berger and Mester (1997) employed both multiple regression and single variable regression. In relation to efficiency correlates, they identified six broad categories – bank size, organisational form and corporate governance, other bank characteristics, market characteristics, state geographical restrictions on competition and primary federal regulators. As a comparison, Maudos et al. (2002) used only four broad categories – a size variable, a specialisation variable, other bank characteristics and market characteristics.

Under each category, at least two variables are used. For example, under bank size, Berger and Mester (1997) used four variables: small bank, medium bank, large bank and huge bank. In contrast, Maudos et al. (2002) used only three variables: medium bank, large bank and huge bank.

Berger and Mester (1997) found that their results were not conclusive. Some of the results had the predicted sign and statistical significance. The others had either unexpected signs or mixed signs. Thus, they concluded that differences in efficiency remain unexplained. They suspected that there are unmeasured factors like differences in managerial ability and measurement error in the efficiency dependent variable. Maudos et al. (2002), on the other hand, found that:

- medium size banks reach the highest level of profit efficiency
- banking specialisation is not significant in explaining the differences in efficiency
- banks with higher loans/assets ratios are more efficient
- market concentration is positively related to profit efficiency
- higher risk, as proxied by the standard deviation of returns on assets, is positively related to profit efficiency

2.8 Competitive behaviour of the banks

Banks, like other production units, are worthy of examination, particularly when their role is of great importance to the whole economy. Realising their great contribution, competitive behaviour by banks is of great interest. Do banks behave as monopolists or as oligopolists? Regardless of their behaviour, the economic welfare is at stake. Monopolist behaviour, for example, might lead to greater welfare loss compared with other types of behaviour. However, such an in-depth investigation requires a particular framework within which the conduct of the banks is observed and analysed. There are two frameworks that can used to study the relationship between competition and market structure: the structural model approach and the non-structural model approach.

2.8.1 Structural model approach

The structural model approach in the first place is based on the monopoly power hypothesis. It hypothesises that the monopoly power indicated by market concentration enables the banks to earn monopolistic profit. Market concentration exists due to collusive actions amongst the banks. It can be measured using *k* bank concentration ratio based on assets or deposits. K can be one, two, three, five or ten, but the most commonly found in the literature is a three and five bank concentration ratio. Another measure of concentration is the Herfindahl index.

The monopoly power hypothesis is also well-known as Structure-Conduct-Performance hypothesis (SCP). The SCP that market concentration is positively and moderately related to profitability (Bourke, 1989). Higher market concentration indicates that a substantial part of the market share (normally in terms of total asset or total deposit, in the case of banking) is controlled by relatively few players in the market. This results in a lower degree of competition, whereby the key players are likely to possess monopoly power, hence charging higher prices and enjoying greater profit.

The studies by Pilloff and Rhoades (2002), Lloyds-Williams and Molyneux (1994), Molyneux and Forbes (1993) Molyneux and Thornton (1992), support the SCP hypothesis that there is a significant and positive relation between market concentration and profitability. However, other studies, including Maudos (1998), Goldberg and Rai (1996), Berger (1995), and Bourke (1989) fail to find support for this hypothesis.

The SCP hypothesis has been challenged by many. Gilbert (1984), for example, argued that the positive correlation between concentration and profit might reflect the superior efficiency among firms rather than the use of market power. In other words, firm efficiency leads to larger size and market share, and finally to the generation of greater profit. The key factor in determining the relation between market structure and firm performance is firm efficiency.

The efficient structure (ES) hypothesis proposes that if a firm is able to enjoy technical efficiency, such as a lower cost structure, it can maximise potential profit by bringing down the price and expanding firm size, hence gaining market share. In empirical studies, market share is used as a proxy variable to firm efficiency (see Bourke 1989, Molyneux and Thornton 1992 and Lloyds-

Williams and Molyneux (1994). However, since firm cost efficiency is indirectly measured, the relationship between firm performance and market structure has been subjected to various interpretations. According to Berger (1995) and Maudos (1998), the critical issue here is that the market share variable may capture the effect of other variables rather than efficiency.

The ES hypothesis gains support from Bourke (1989), while Molyneux (1993) finds little country-specific evidence for this hypothesis. The main weakness of the ES hypothesis is that efficiency is represented by the market share variable. This was found to be problematic, as it has been subjected to various interpretations. To overcome this problem, Berger (1991, 1995) proposed a direct measure of efficiency and its direct inclusion into the profit model. In this new approach, a firm's cost efficiency is directly measured using a stochastic frontier function. There are two types of efficiency, X-efficiency and scale efficiency. The main advantage here is that the relationship between performance and concentration has a clear-cut interpretation.

Instead of just two hypotheses (SCP and ES), Berger (1991, 1995) employed four testable hypotheses using direct measures of both market structure and efficiency. The four hypotheses are:

- structure conduct performance hypothesis (SCP hypothesis), which states that higher profits are due to anti-competitive price settings in a concentrated market. The main indicators are concentration ratio and the Hefindahl index.

- relative market power hypothesis (RMP hypothesis). It states that firms with large market shares are able to exercise market power to earn higher profits. This hypothesis is normally represented by the firms' market size in term of assets and deposits.
- efficiency structure hypothesis (ES hypothesis), which proposes that firms with superior management or production processes can operate at a lower cost and consequently reap higher profits. This hypothesis is represented by any efficiency measures like technical or cost efficiency measure.
- scale-efficiency hypothesis (SE hypothesis). It states that firms that operate at an optimal scale will have lower costs, and this will result in higher profits.

Berger found that the support for the four hypotheses is limited. His findings support the ES and RMP hypotheses, and do not support the ES or SCP hypotheses. Similar results were found by Gorldberg and Rai (1996). Their study failed to support the ES hypothesis, since the results are not robust or particularly sensitive to the measure of performance used. In addition, Maudos (1998) accepted the modified version of the efficient structure hypothesis, since both the efficiency and the market power that are reflected in market share are found to have influences on profit. In the case of Malaysian banks, Tahir (1999) finds support for both the relative market power and the efficiency structure hypotheses.

2.8.2 Non-structural model approach

In contrary to the structural approach, this approach investigates the behaviour of the banks, like the degree of competition, without utilising explicit information about the structure of the market. For example, under this approach, information about the structure of the market such as market power and efficiency are ignored. Bikker and Haaf (2002) identify three models that fall under this approach. These models are the Iwata model (Iwata, 1974), the Bresnahan model (Bresnahan, 1982) and the Panzar-Rosse model (Panzar and Rosse, 1987). The Iwata model and the Bresnahan model were also identified as conjectural-variation models.

The use of a conjectural-variation model in the literature is limited. Uchida and Tsutsui (2005) use this model to investigate whether the degree of competition in the Japanese banking industry has improved or not for the period 1974-2000. The results are mixed. Competition only improved in the 1970s and the early 1980s. Toolseme (2002), meanwhile, looks at the Dutch consumer credit. She finds that there is no evidence of market power.

The Panzar-Rosse model (Panzar-Rosse, 1987) is based on the premise that the banks will employ different pricing strategies given any changes in input costs. However, the ability of the banks to respond to any changes in input costs is influenced by the market structure. Since the output prices for the banks are not directly observed, revenue equations are used. They define a measure of competition, the H-statistic, as the sum of the elasticities of the reduced-form revenues with respect to all input prices. The H-statistic ranges from zero to one. If H = 0, then the market is characterised as a monopoly. H=1 implies that the market is perfectly competitive. The H-statistic that lies between zero and one implies that the market is under monopolistic competition.

Claessens and Laevan (2004) undertook a large scale study covering 50 countries' banking systems for the period 1994-2001. The final sample consisted of 35,834 bank-year observations. Two measures of revenue were used: gross interest revenue and total revenue. The reduced-form revenue equations were estimated using pooled OLS and fixed effect with generalised least squares. The results show that the H-statistic on average varies between 0.6 and 0.8, suggesting that monopolistic competition is the best description of the degree of competition. The work of Claessens and Laevan (2004) is acknowledged by Shaffer (2004) for two main contributions. They extend a proven empirical method (the Panzar-Rosse methodology) to an unprecedentedly large and varied cross-country sample. Secondly, they offer additional insights with regard to factors associated with the variations in measured conduct.

Yildirim (2003) studied the effects of financial liberalisation and deregulation on competitive conditions in the banking industries of fourteen Central and Eastern European transition economies for the period 1993-2000. Similar to Claessens and Laeven (2004,2003), he used two measures of revenue: gross interest revenue and total revenue. However, Yildirim used three estimation techniques: pooled OLS, fixed effects and random effects. The results of the competition analysis suggest that the banking markets in these transition economies operate as under conditions of monopolistic competition. He further observes that large banks operate in a relatively more competitive environment compared to small banks.

In addition, Bikker and Haaf (2002) studied the competitive conditions in the banking industry of 23 countries (European and non-European countries). Again, they used the same measure of revenues. They found that for all-banks samples of all 23 countries, both H=0 (perfect cartel or monopoly) and H=1 (perfect competition) are rejected convincingly. This would imply monopolistic competition in all countries.

Other studies that use the Panzar-Rosse methodoly are Coccorese (2004), Chun and Kim (2004), Shaffer (2002), Hondroyiannis et al. (1999), Molyneux et al. (1996), Molyneux et al. (1994)

2.9 Concluding remarks

This final section has 2 parts: the summaries of the literature reviews and expected contributions by this study. The summaries cover the role of financial institutions, efficiency issues like its concept (technical, cost and profit efficiency), its measurement methods (parametric and non-parametric), efficiency correlates and the competitive behaviour of the banks (structural and non-structural model). The second part looks at the current gap in the literature and presents how this study will contribute towards the existing literature.

This chapter reviews the previous studies of the efficiency and competitive behaviour of the financial institutions. Financial institutions are widely seen and accepted as financial intermediaries rather than simply as production units in the economy. This recognition consequently affects the selection of input and output. It was also found that regardless of the technique of estimation used, findings are highly sensitive to the choice of input and output. In addition, Denizer et al. (2000) mentioned the possibility of a third approach, namely, the modern or risk approach. This approach has several innovative components, such as the introduction of a measure of the banks' asset quality and the probability of the banks' failure.

Efficiency is a relative concept and can be broadly analysed in terms of technical, cost and profit. Technical efficiency requires a production frontier function while cost and profit efficiency require cost and profit frontier function respectively.

It appears that there are two dominant approaches in estimating efficiency namely the stochastic frontier approach and the data envelopment approach. As each technique has its own advantages, this has encouraged researchers to use either or both of them and this trend continues in the current literature.

Under the parametric approach, the popular methods used in the literature are the stochastic frontier approach, the distribution-free approach and the thick frontier approach. The main difference between them is the strong assumption made with regard to the distributional form of each component of error term. In terms of functional form, the translog function emerges as the most popular choice among researchers. However, it has been shown that the variability in the efficiency scores is due to different techniques, sampling and data quality.

There is a wide difference in profit efficiency estimates across countries and banks. As highlighted by Berger and Humphrey (1997), there are still some aspects of this area of study that are left unexplored or that remain to be addressed properly. These include management quality, product quality and measurement error. The profit efficiency follows the trend found in the cost efficiency studies. This includes functional forms (translog and Fourier flexible form) and estimation techniques (SFA, DFA and TFA). Except for the dependent variable, the exogenous variables are the same for both frontiers.

The review discloses the lack of relevant studies concerning the efficiency of financial institutions in the developing countries. The survey of 130 studies done by Berger and Humphrey (1997) clearly showed that only 6 involve developing countries. In our own survey of 37 studies since 1997, we find that only 7 studies involve Asian countries. With regard to the competitive behaviour of commercial banks, studies in the developing countries are left behind except the recent study undertaken by Claessens and Laeven (2004).

Thus, the lack of relevant studies in developing countries drives the motivation of this study.

There are two main approaches in examining the relationship between bank profit and market structure. The first is the profit frontier approach. Profit efficiency is then estimated and scores are tested with potential correlates, which include, among other things, a market structure variable. The second approach uses normal profit function. In this technique, profit function is developed by taking into account its main determinants, which include market structure variables (such as concentration ratio and market power).

The competitive behaviour of the banks can be analysed in two approaches: the structural model approach and the non-structural approach. Under the structural model approach, the competitive behaviour of the banks is evaluated based on the monopoly power hypothesis. This hypothesis states that the existence of market power indicated by the concentration ratio enables the financial institutions to enjoy monopolistic profit. This market power can be realised via collusive actions by the financial institutions. The market power hypothesis was later challenged by the efficient structure hypothesis. Under the non-structural model approach, explicit information about the structure of the market is ignored. Two broad methods were found: the conjectural variation model and the Panzar-Rosse methodology.

We find that there are at least four studies that estimate the efficiency of commercial banks in Malaysia: Katib and Matthews (1999), Abd_Karim (2001), Edward (1999) and Laeven (1999). Both Katib and Matthews (1999) and Laeven (1999) used non-parametric approach while Abd_Karim (2001) and Edward (1999) used parametric approach. Based on the non-parametric approach, the average technical efficiency estimates of commercial banks in Malaysia are between 0.7 and 0.8 while the parametric approach yields higher estimates well above 0.9. The different estimates of efficiency of the above studies trigger the use of both approaches applied to the same data set. Will those results be upheld when utilising both approaches? We believe that the use of both approaches can contribute towards the existing literature by demonstrating that different measurement methods should be a matter of concern.

One important aspect that the previous studies had ignored is the presence of foreign banks in Malaysia. It is true that there are studies in the developed countries that investigated the performance of the banks based on ownership. The existing finding so far is that the foreign banks on average have higher efficiency estimates than the local banks. However, such generalisation from one country to another might be misleading. One needs to acknowledge the differences in the market structure and legal framework and the role of government in the banking sector. As far as the Malaysian banking sector is concerned, it is still relatively closed and well protected from foreign competition. Government controls via legal requirement remain intact but the wind for change is getting stronger. This study aims to consider the presence of foreign banks in Malaysia. An issue that worth to look at is how fast the local banks move closer to the foreign banks and surely this issue will vary

Chapter Three

from one country to another. Thus, the performance comparison between local and foreign banks can be taken as another gap in the existing literature.

Another important aspect that has been ignored is the sources of productivity growth of the banks in Malaysia. None of the previous studies took this into account. This study not only looks at the annual estimate of technical efficiency obtained under non-parametric approach, it also investigates the sources of productivity growth on annual basis and over a short period of time. This has no doubt extended the previous studies done by Katib and Matthews (1999) and Laevan (1999).

The second aspect that lacks investigation is the competitive behaviour of the banks in developing countries using a non-structural model. The work done by Claessens and Laevan (2004) is a large scale study that covers 50 countries. Realising the strength of their study, our work aims to be complement and enrich the existing literature. However, differences with their study can be found in terms of the definition of input prices, types of banks and sample period.

CHAPTER THREE

Banking Development in Malaysia

3.0 Introduction

A clear understanding of the background and the environment of an industry are important before undertaking any study. Such an understanding is useful in designing an appropriate framework. This chapter aims to provide the background to the development of the banking sector in Malaysia. The banking system is said to have begun in 1957 with the establishment of Bank Negara Malaysia (BNM), the Central Bank, soon after Malaysia gained its independence from Great Britain. However, it must be noted that the commercial banks were present in the country long before the independence. Their presence was mainly to facilitate trade and provide short-term financing. The roles of the commercial banks become apparent after independence, in particular when the Malaysian economy underwent a tremendous structural change. The financial and banking system itself had been restructured, reorganized and reshaped to meet the rising demand of the growing economy. The development process included the establishment of a state-owned bank and six other banks during the 1960s. Another five banks were created in the 1970s. After 1970, with the introduction of the New Economic Policy (NEP; 1971-1990), the banks were required to help the country pursue the national agenda[7]. The NEP aimed to achieve two short-term objectives by 1990, the reduction and eventual elimination of poverty, and the restructuring of society in terms of employment and capital ownership. The long-term goal of the NEP was to restore the national unity that had been badly affected by the May 13th incidence. As a result of the NEP and subsequent government policies, the banks were required to participate directly in restructuring the society and overcoming the economic problems; this, amongst other things, included the provision and allocation of funds to certain ethnic groups and certain economic sectors. Thus, the banking sector, as a major mobiliser of the funds, played a crucial role in the transformation and restructuring of the economy.

This chapter is organised as follows. Firstly, we discuss the background of the Malaysian economy from the colonial period until the present day. The focus is the major economic activities and related economic policies. In Section 3.2, we present the framework and structure of commercial banks in Malaysia. The dimensions of the banking development that emphasise the financial widening and financial deepening will follow. Next, we look into the role of commercial banks in the economy, and how they perform this role in Malaysia. In Section 3.5, we consider the financial reforms that have been taking place since late 1970s

[7] The NEP was introduced after the racial riot in 1969 that resulted in the deaths of many Malaysians. This racial clash involved two main ethnic groups, the Malays and the Chinese. One of the factors behind the riot was the economic imbalance between the Malays and the Chinese. The economic wealth was not in favour of the Malays. The triggering factor of the riot was political. In that year, the results of the general election indicated that the Malays had lost their political dominance in certain states.

and become more significant in the aftermath of Asian financial crisis (1997-1998). The development of the Islamic banking system is presented in Section 3.6. Finally, we summarise and make important remarks.

3.1 Malaysian Economy: A background perspective

Malaysia (formerly Malaya) is a relatively small country located in South East Asia. It covers a land area of about 330,000 square kilometers surrounded by close neighbours like Thailand (in the north) and Singapore and Indonesia (in the south). Malaysia has two distinct parts, West Malaysia (Peninsular Malaysia) and East Malaysia (Sabah and Sarawak) located in the island of Borneo. Both parts are separated by the South China Sea. Malaysia gained its independence from Great Britain in 1957 after being colonised for more than 400 years. Foreign colonialisation started with the fall of Malacca to the Portuguese in 1511. Then the British came with the opening of Penang in the 19th century. Because of colonialisation and foreign interests, Malaysia has a multiracial population[8]. In 2003, Malaysia's population is estimated to be around 25.32 million. The Malay or the so-called Bumiputra (the sons of the soil) are the majority (more than 60%), followed by the Chinese (28%), Indians (8%) and others.

[8] Under British rule, the Indians were brought in to work mainly in the rubber estates. Meanwhile, the Chinese were encouraged to join in tin mining. Both rubber plantation and tin mining were new economic activities with which the Malays were not familiar.

From an economic perspective, agriculture was the major sector in the country. It contributed more than 30% to the Gross Domestic Product (GDP) and more than 50% to export earnings in the 1960s. Rubber and tin were the major commodities in the colonial era from which the British had benefited the most. However, the dominance of the primary sector started to decline in 1970s as the country began to focus on the secondary sector. The transformation of the economy was planned based on the NEP and the five-year plans. By the middle of the 1980s, the manufacturing sector took the lead as the main sector in the economy alongside the service sector. A few factors could explain this transformation, such as uncertainties in agriculture (unstable and low income), the lack of effective government policy, effective industrial policy and favourable external conditions[9]. However, Ariff and Khalid (2000, pp. 84-85) argue that the pursuit of the twin policy of rural improvement and increasing the incomes of indigenous people needed to be reviewed in the 1970s due to rising public debt and the loss of competitiveness in commodity production. The review of the existing policy was also triggered by the successful experiments of newly rising countries with a different policy mix. Those countries, such as Singapore, Taiwan, Korea and Hong Kong, had already adopted new policies to restructure their economies by focusing on sustainable growth driven by export-led and labour intensive economic activities. It was further argued that after the oil crisis in the early 1970s, Malaysia chose industrialisation as a route to development.

[9] In the 1960s, the poverty rate stood around 50% and the Malays formed the majority of those living below the poverty line income. Their main activities are farming and fishing. This makes the agriculture unattractive especially to the younger generation.

In 2003, the manufacturing sector contributed 31% to the GDP and up to 80% to export earnings. This successful transformation was taken as a determination in Malaysia that it would become a developed country by the year 2020. This was known as Vision 2020 launched in 1990. The structural transformation from agro-base to manufacturing has witnessed a huge increase in per capita income. Per capita income was only RM720.00 in 1960 and after 40 years, it stood at RM14, 629.00 in 2000. In 2003, per capita income is RM15, 569.00. In US dollars, this is equivalent to $4097.00[10]. Because of her promising performance, Malaysia is seen as a rising star in the East along with Thailand, the Philippines and Indonesia.

3.2 Structure and framework of the financial system

The Malaysian financial system is similar to those of other countries. According to Chin and Jomo (2003), the Malaysian financial system exhibits many features of the Anglo-American model. The main feature of Anglo-American banking model is that the banking activities are restricted to accepting deposits, granting loans and other specified activities[11]. Although there is an Islamic banking system, its share in terms of total banking assets is less than 10%. The total assets of commercial banks are RM629.6 billions in 2003 while the total assets of Islamic banks are only RM29.1

---

[10] Since September 1998, the Malaysian Ringgit has been pegged to US dollar at US$1:RM3.80. This was the drastic measure taken by Malaysia to counter the severe impact of the currency crisis.

[11] Ariff and Khalid (2000) mention three models of banking organisation, namely the American model, the English model and the mono banking of the Communist. (See pp. 50).

billions. Although the Islamic system has been there since 1983, the public acceptance remains low. Since our study ignores the Islamic banking system, the methodology is similar to other previous studies.

The financial system has three main parts, namely, the banking system, the non-bank financial institutions and the financial markets. The banking system comprises the monetary institutions and non-monetary institutions. (See Chart 3.1 for the structure of financial institutions in Malaysia.) The monetary institutions refer to those institutions whose liabilities are generally recognised as money. These institutions include the Central Bank, being the sole issuer of the currency, and the commercial banks, the only institutions allowed to offer current accounts. On the other hand, the non-monetary institutions are closely linked to the monetary institutions whose liabilities are generally accepted as quasi money. They include finance companies, merchant banks and discount houses. The finance companies and the merchant banks are normally subsidiaries of the commercial banks. (See Table 3.1 for the list of banking institutions as at the end of 2004.) For example, the AmBank Berhad has its own finance company and merchant bank, AmFinance Berhad, and AmMerchant Bank Berhad respectively. The same trend applies to Affin Bank Berhad, Bumiputra-Commerce Bank Berhad, RHB Bank Berhad and Southern Bank Berhad.

Jomo and Hamilton-Hart (2001) argue that there is no consensus with regard to the nature of the financial sector. Some argue that the financial sector in

Malaysia is liberal while others say that it is still a controlled system. Although reforms are taking place, many researchers argue that these are not completed yet. Further financial liberalisation is required. In relation to its financial liberalisation, Ariff and Khalid (2000) consider that these reforms ended when the Malaysian currency was pegged to the US dollar in 1998 and other measures of capital controls were introduced.

The banking institutions are currently governed by the Banking and Financial Institutions Act 1989 (BAFIA 1989) with the exception of Islamic banks. The BAFIA 1989 replaced two old Acts that had governed both commercial banks and finance companies, the Banking Act, 1973 and Finance Companies Act, 1969. This Act was introduced due to the failure of the deposit-taking cooperatives, which eventually threatened the health and integrity of the banking system, BNM (1999). The need for a new legal framework became increasingly critical as the growing competition in the banking system had made it hard to draw the lines of business amongst banking institutions, namely, commercial banks, finance companies and merchant banks. The BAFIA (1989) aims to provide an integrated supervision of the banking institutions and to improve the current laws relating to banking operations. With the introduction of the BAFIA 1989, all banking institutions are now under the direct supervision of the Central Bank. The Act also brings other institutions such as discount houses and money and foreign exchange brokers under a single supervisory and regulatory regime. Another important impact of the BAFIA is that all banks, including the foreign banks, had to exchange their licenses for new ones. Unlike the domestic banks, the foreign banks were given a period of up to five years to exchange their licenses. However, in order to obtain the new license, they are now required to be incorporated locally, but are allowed to retain 100% ownership of the new entity. Prior to 1989, the foreign banks operated on a bank branch basis. At least one foreign bank ceased to operate due to the new requirement imposed by the BAFIA 1989.

The BAFIA 1989 also has vast implications for governance of the financial system. For example, it broadens the scope of permissible investment by the commercial banks. The banks are now allowed to invest in blue-chipped shares, other approved shares, manufacturing companies and property trusts. The Act also empowers the central bank to investigate and take corrective measures, including assumption of control of scheduled institutions, in order to promote monetary stability and sound financial structure.

Besides the BAFIA 1989, the commercial banks are required to perform social obligations as stated in well-known national policies such as the New Economic Policy (1971-1990) and National Development Plan (1991-2000). One of the duties of the banks under the NEP is to grant loans to special groups or sectors like the Bumiputra community, small and medium enterprises (SMEs), residential housing and non-resident controlled companies (BNM, 1994). Credit to the Bumiputra community refers to the credit extended to Bumiputra individuals and

Bumiputra-controlled companies[12]. This policy aims to increase the Bumiputra share in the country's capital ownership. In general, the domestic banks provide most of the loans extended to the Bumiputra community as compared to the foreign banks. For example, at the end of 1993, the local banks provided more than 80% of the total credit extended to this community. To assist the banks in meeting these obligations, the Central Bank from time to time issues lending guidelines. (See Table 3.2)

[12] After 1990, the government introduced the creation of the Bumiputra commercial and industrial community. It is the continuity of the restructuring of the society pursued by the previous national policy (New Economic Policy, 1971-1990).

Chart 3.1: The Financial System Structure
As at December 31, 2003

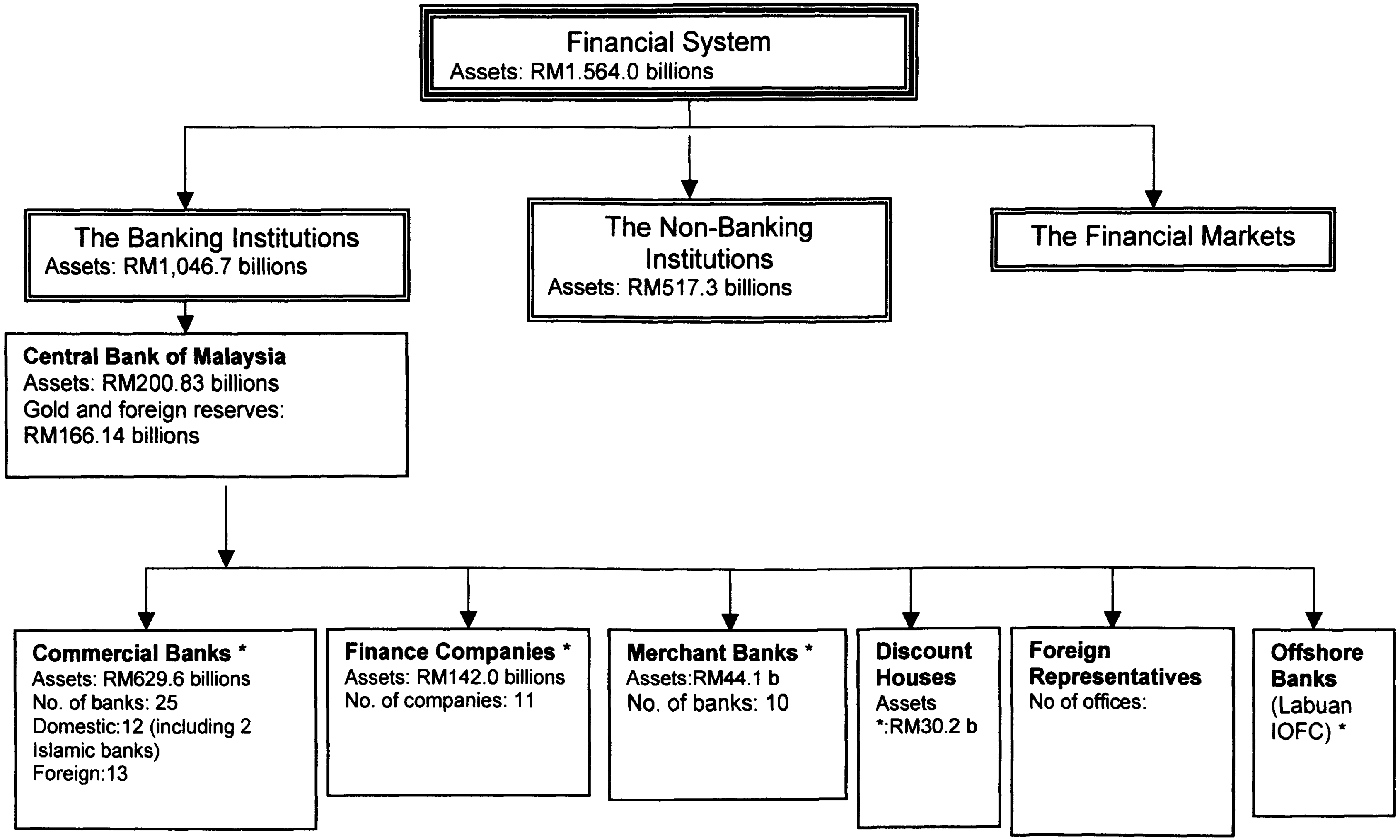


Chart 3.1 (continued)

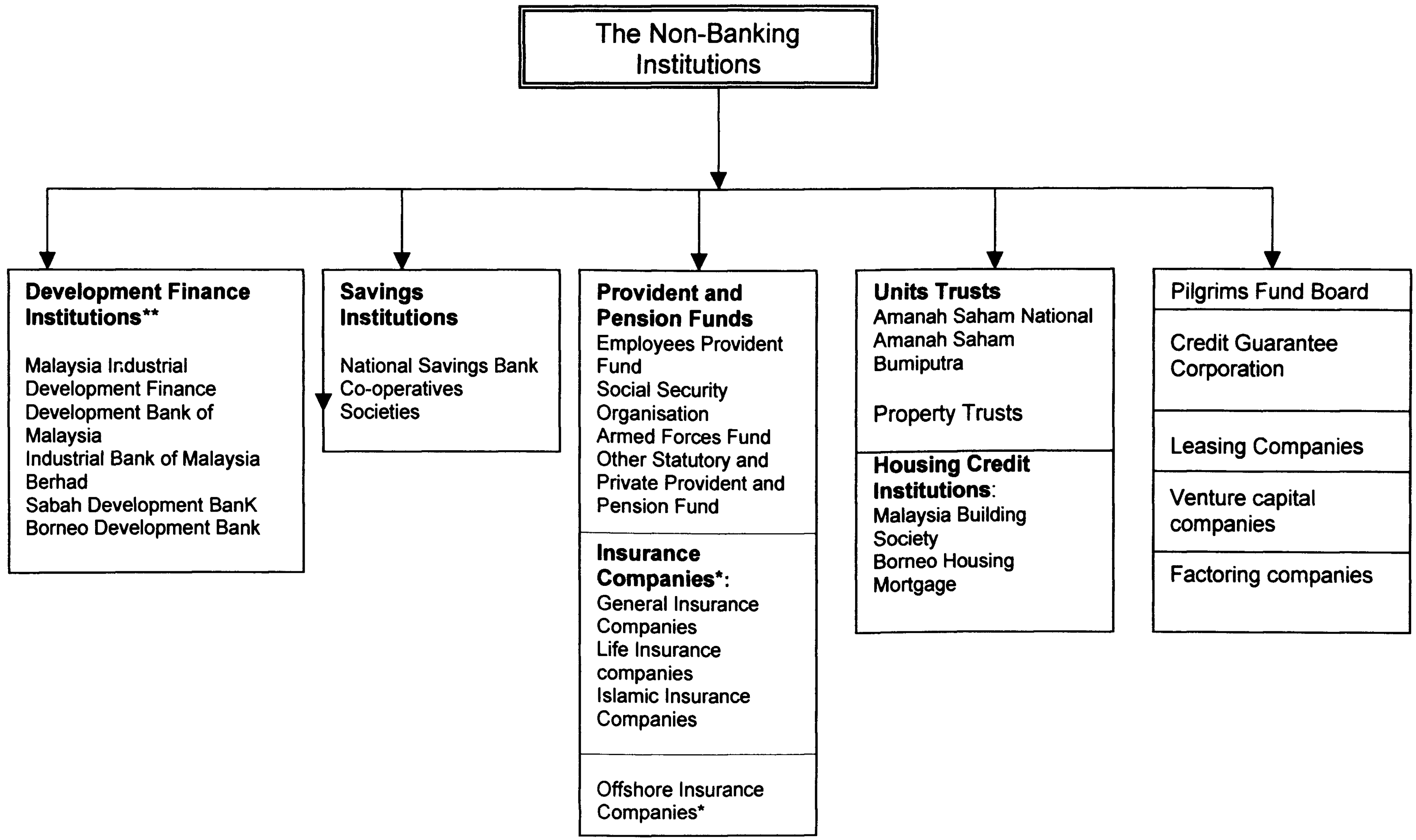
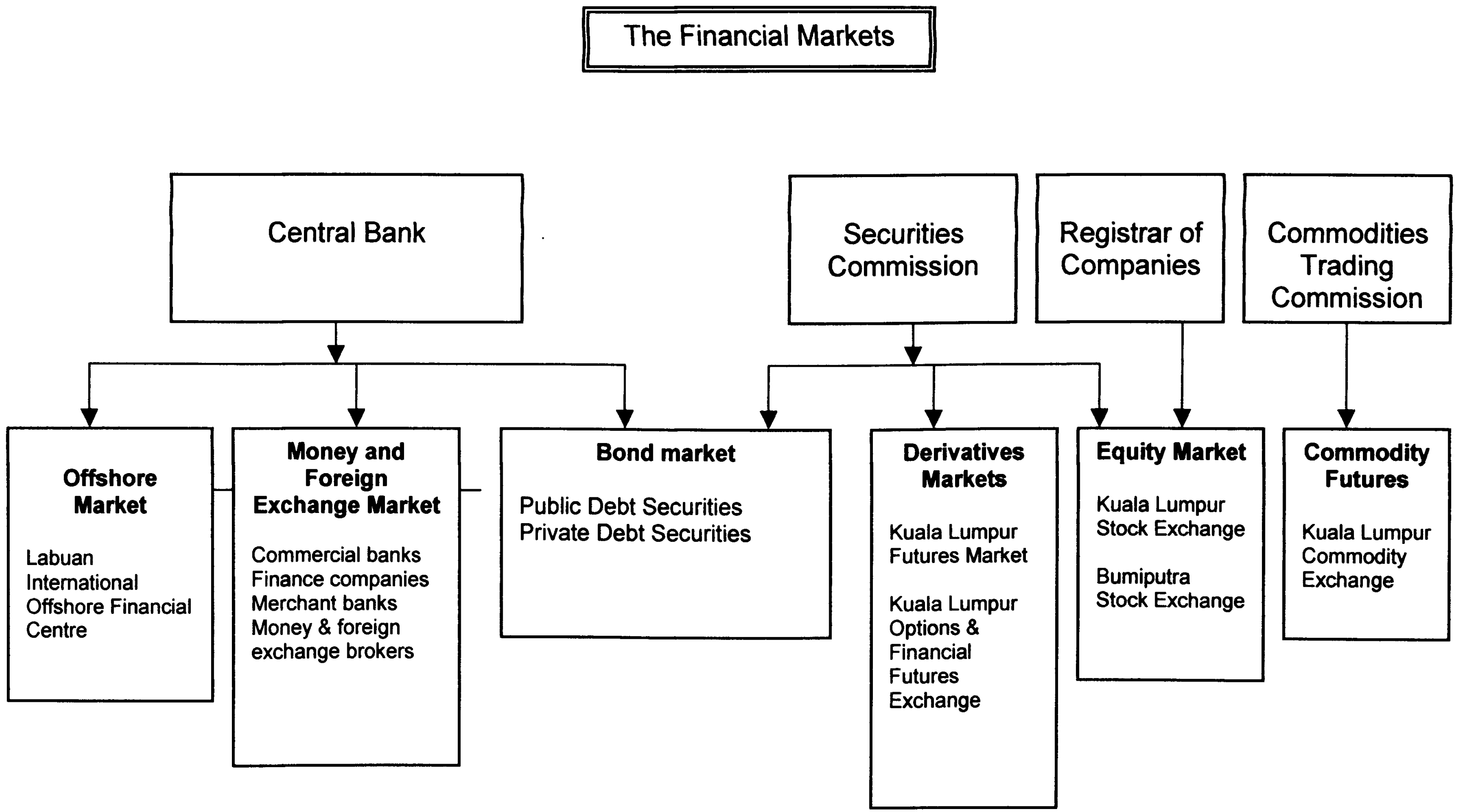


Chart 3.1 (continued)



Notes: * supervised by the Central Bank
** registered with the Central Bank

Table 3.1: List of banking institutions (as at the end of 2004)

Number of institutions	Commercial Banks	Finance Companies	Merchant banks
1	ABN Amro Bank Berhad	AmFinance Berhad	Affin Merchant Bank Berhad
2	Affin Bank Berhad	Affin_ACF Finance Berhad	Alliance Merchant Bank Berhad
3	Alliance Bank Berhad	Bumiputra-Commerce Finance Berhad	AmMerchant Bank Berhad
4	AmBank Berhad	RHB-Delta Finance Berhad	Aseanbankers Malaysia Berhad
5	Bangkok Bank Berhad	Kewangan Bersatu Berhad	Commerce International Merchant Bankers Berhad
6	Bumiputra-Commerce Bank Berhad	Southern Finance Berhad	Malaysian International Merchant Bankers Berhad
7	Bank of America Malaysia Berhad		Southern Investment Bank Berhad
8	Bank of China (Malaysia) Berhad		Public Merchant Bank Berhad
9	Bank of Tokyo-Mitsubishi (Malaysia) Berhad		RHB Sakura Merchant Bankers Berhad
10	Citibank Berhad		Utama Merchant Bank Berhad
11	Deutsche Bank (Malaysia) Berhad		
12	EON Bank Berhad		
13	Hong Long Bank Berhad		
14	HSBC Bank Malaysia Berhad		
15	J.P. Morgan Chase Bank Berhad		
16	Malayan Banking Berhad		
17	OCBC Bank (Malaysia) Berhad		
18	Public Bank Berhad		
19	RHB Bank Berhad		
20	Southern Bank Berhad		
21	Standard Chartered Bank Malaysia Berhad		
22	The Bank of Nova Scotia Berhad		
23	United Overseas Bank (Malaysia) Berhad		
24	Bank Islam Malaysia Berhad		
25	Bank Muamalat Malaysia Berhad		

Source: BNM (2004)

Table 3.2: Lending guidelines to the priority sectors

by the commercial banks

	1998/1999		2000/2001		2003/2004	
	Target	Achieved	Target	Achieved	Target	Achieved
Credits to Bumiputra community Total outstanding loans (billions) Total outstanding loans (%)	71.7 30	78.4 32.8	88.0 30	87.3 29.7		
Credits to SMEs (loans of RM5 millions and below: Total loans approved (billions) Of which: for Bumiputra SMEs Total loans approved (billions)	1.0 0.5	7.9 1.2			20.6	901
Credits for residential housing Total number of houses (units)	105,658	148,502	109,910	116,750	78,114	32,825

Source: BNM (2001a, 2003).

In terms of total loans, at least 30% must be extended to the Bumiputra community either to individuals or to Bumiputra-controlled companies. For example, for the period 1998/99, the total loans granted exceed the target. The target was almost achieved in the subsequent period. Credits to SMEs and residential housing were well achieved in both periods.

The second part of the financial system is the non-bank financial intermediaries. This part comprises five groups of institutions, namely the provident and pension funds, insurance companies (including Islamic insurance companies or Takaful), the development finance institutions, the savings institutions and other non-bank institutions like Pilgrim Fund Board, unit trusts etc. The total assets of the non-bank financial institutions were RM517.3 billions as at the end of 2003. This was equivalent to 33% of the total assets of the financial system. The largest non-bank institution is the Employees Provident Fund (EPF) that has a total asset of RM220.2 billions in 2003 (equivalent to 42% of the total assets of the non-bank financial institutions). As the largest non-bank institution, the EPF has a huge advantage and an important role to perform[13].

The third part of the financial system is the financial markets. The financial markets can be divided into four: the money and foreign exchange markets, the capital markets, the derivatives markets and the offshore markets. The money market refers to the transaction of short-term funds, typically less than 12-months. This market involves the placement of deposits and the purchase and sale of short-term securities like bankers acceptance, negotiable instrument of deposit, treasury bills etc. The interbank players in this market include the commercial banks, the merchant banks, discount houses and eligible finance companies. In 2003, the total money market transactions stood at RM1,538.4 billions. Interbank deposits contributed a total of RM1,084.7 billions or 70% of the total transactions in the money market. Meanwhile, the volume of money market papers stood at RM453.7 billions. The major market papers in 2003 are Malaysian Government Securities (MGS, RM231.4 billions), followed by Bank

[13] One of the main critics against the government is the use of public funds like EPF to bail out certain state-owned enterprises or politically connected companies. This was highlighted by Jomo (2001).

Negara Bills (RM71.5 billions), negotiable instrument of deposits (RM43.1 billions) and bankers acceptance (RM37.3 billions).

The foreign exchange market is the market for trading the foreign currencies against the ringgit or against other foreign currencies. Any transactions in the foreign exchange market can be undertaken in the spot market besides the forward and swap market. Under the current exchange control regime (introduced in 1998), all forward and swap transactions undertaken by non-bank residents have to be supported by firm underlying trade transactions (BNM, 1999)[14]. All transactions in foreign currencies have to be conducted through authorised dealers like commercial banks and designated merchant banks. By composition, the foreign exchange market continued to be dominated by transactions in the US dollar against the ringgit. In 2003, such transactions amounted to RM323.6 billions, the equivalent of 81% of the total transactions. The dominance of US dollar transactions against the ringgit reflects the importance of the US dollar in the foreign exchange market as well as the high usage of the US dollar in the settlement of trade (BNM, 2003). An interesting development lately is the rising share of the euro and yen transactions against the US dollar. This phenomenon can be attributed mainly to the higher transaction and hedging needs of market participants leading to the significant volatility among the currencies.

---

[14] Since 1998, the ringgit is pegged to US dollar at US$1:RM3.80. In the first quarter of 2005, there has been speculation that the pegging would be reviewed or abandoned completely.

The capital markets are the markets for raising long-term funds and consist of the equity and bond markets. The equity market is the avenue for public-listed companies to raise external funds by issuing stocks and shares. The Kuala Lumpur Stock Exchange (KLSE) and the Malaysian Exchange of Securities Dealing and Automated Quotation (MESDAQ) handle the market trading in stocks and shares. The MESDAQ was established in 1997 as an avenue for small, high growth potential and high technology companies. The KLSE experienced significant expansion with the market capitalisation reaching a peak of RM891 billions in February 1997 before falling to a low of RM181 billions in September 1998 in the aftermath of the Asian financial crisis (BNM, 1999). At the end of 2003, the market capitalisation of the KLSE stood at RM640.5 billions. This is equivalent to 163.4% of the Gross National Product. At the end of 2004, the total number of listed companies is 963 (622 companies listed on the main board and 278 companies on the second board). The increase in the number of listed companies occurred mainly during the 1990s. In April 2004, the KLSE changed its name to Bursa Malaysia.

The bond market is where both the private and public sectors can raise funds by issuing private debt securities (PDS) and Malaysian Government Securities (MGS) respectively. Secondary market trading in unlisted bonds is done through the over-the-counter market and for the listed bonds on the KLSE. Prior to 1990, the funds raised in this market were generally dominated by the public sector, which needed considerable funds to finance its development expenditure. Since then, and in line with the privatisation policy and downsizing government operations, the amount of funds raised by the private sector had outpaced the amount raised by the public sector. At the end 0f 2003, the outstanding PDS stood at RM170 billions. The outstanding MGS were RM149 billions.

The derivatives markets are for trading instruments that provide contingent claims on underlying assets. The values of this instrument depend on the price of the underlying assets or securities. In principle, the use of derivatives aims to guard against volatility in the price of the underlying assets. However, it is possible to use them to speculate for capital gains. Examples of derivatives are forwards, futures, options and swaps. The derivatives markets started in 1980 with the establishment of the Kuala Lumpur Commodity Exchange. It trades in commodity futures like crude palm oil. In 1995, market-traded financial derivatives were introduced in the form of financial futures.

The offshore market is a new addition to the financial structure in Malaysia. It began with the establishment of the Labuan International Offshore Financial Centre (Labuan IOFC) in 1990. The Labuan IOFC aims to enhance the attractiveness of Malaysia as a regional financial centre. It provides a wide range of offshore products like banking, insurance and insurance-related activities, trust business, fund management, investment holding etc. At the end of 1999, there were 63 offshore banks, 50 insurance and insurance-related companies, 20 trust companies, 7 offshore leasing companies and 6 fund managers operating in Labuan IOFC.

3.3 Dimensions of financial development

Financial development can be defined as changes in the financial sector. To understand this process, we need to obtain relevant information regarding any changes in the structure of the sector. These changes might have taken place in the short term as well as in the long term. A country that claims to have passed a primitive stage in the economic development should have several forms of financial instruments besides various financial institutions like commercial banks, finance companies, and saving institutions. This implies that the presence and relative size of financial instruments and financial institutions are the main features of the financial structure of a country. Thus, financial development is best seen from two dimensions, financial widening and financial deepening.

3.3.1 Financial widening of the banking sector

Financial widening takes place when the number of financial institutions and the size of the financial sector increase. The banks normally expand by increasing the number of their branches or its facilities like automated teller machines. At present, the banking expansion is further enhanced via the cyber world by means such as Internet banking. The real growth of the banking sector will enhance people's access to financial services and will further facilitate the mobilisation of the funds in the economy. A relatively cheaper transaction cost is made possible by the growth of the banking institutions. For example, with the availability of nearby banks, people may pay less for transportation cost. They may also spend less time to obtain the required financial services. Financial widening can be demonstrated by the number of financial institutions and the corresponding branches available and the ratio of population to bank branches (refer to Table 3. 3 below.)

Table 3.3: Commercial Banks in Malaysia

Years	Total		Domestic		Foreign		People Served Per Branch
	Banks	Branches	Banks	Branches	Banks	Branches	
1959	26	111	8	12	18	99	71600
1960	28	131	10	29	18	102	62800
1970	38	336	16	177	22	159	32400
1980	38	546	21	398	17	148	25100
1990	38	998	22	852	16	146	18136
1995[a]	37	1433	23	1289	14	144	14438
2000[ab]	26	1755	13	1613	13	142	13384
2001	25	1659	11	1518	14	141	14472
2002[ad]	23	1632	10	1493	13	139	15031
2003[a]	23		10		13		

Notes: a = as at December 31st excluding 2 Islamic banks.
b = domestic banks start to merger.
c = final number of domestic banks is 10.
..........d = 2 foreign banks merged in 2002.
Sources: BNM (1994, and 2001b), ABM (1994, 1996, 1998 and 2000).

In 1959, two years after getting independence from the Great Britain, there were already 26 commercial banks in Malaysia with 111 branches. Although the number of banks could be regarded as reasonably high, they were not well spread across the country. The banks and their branches were confined to urban areas leaving behind rural areas. This was partly contributed to by the concentration of major economic activities in the urban areas. The number of people served per branch stood at 71,600 and this could be regarded as too high. This shows that public access to banking services was very limited. This problem had been given immediate priority by the Bank Negara Malaysia when it was set up in 1959. One of the priorities was to expand the banking sector (a greater number of domestic banks and bank branches) in order to mobilise savings and to reduce national dependency on foreign banks. According to Jomo and Hamilton-Hart (2001), the spread of local bank branches was handled directly by the government. The government-controlled banks expanded their operations by setting up new branches in rural sectors and suburban areas. The monetary authority also used exhortation and incentives to persuade other banks to expand their operation physically. After three decades, the number of bank branches had grown tremendously, given that the number of banks remained the same[15]. With 1,755 bank branches in 2000, the number of people served per branch had been cut to 14,432. After 2000, the number of banks and branches fell further as a result of major consolidation amongst domestic banks. The number of domestic banks in 2003 was only 10 (excluding 2 Islamic commercial

[15] Between 1959 and 2000, the number of banks did increase. For example, in 1990, there were 38 commercial banks. However, due to the merger of several domestic banks and the closure of a few foreign banks, the number of commercial banks remains the same.

banks). As part of the rationalisation process, some bank branches needed to be closed. This has caused the number of people served per branch to increase slightly for example to 15,678 people in 2001. The impacts of financial widening also witness the reduction in the banking concentration in major urban area, the spread of the banking habit and the increased monetisation of the rural sector.

The banking sector has also undergone a modernisation process. Electronic transactions began in the late 1970s with the introduction of automatic teller machines (ATM). This machine allows any withdrawal of money without using counter services. However, the banks had their own payment networks, which prevented the bank customers from using the ATMs of other banks. Following the integration of all ATM networks in 1996, the domestic banks had a fully integrated ATM network of more than 3,900 ATMs by the end of 2001. The switching, clearing and settlement services for the shared ATM network are provided by the payment consortium, Malaysian Electronic Payment System Sdn. Bhd. (MEPS). This facility enables bank customers to use all ATMs within the MEPS network and reduces the need for further expansion of each individual bank's ATM network.

Another important development is the launching of a debit card system by the MEPS in 2002. Under this system, the MEPS would operate the switching network for the domestic debit cards, so that ATM cardholders of participating banks would be able to make payments using their ATM cards at participating shops regardless of the issuing bank. As a pilot case, three domestic banks were chosen to offer this service.

With the growing use of the Internet as a new means of making business transactions, the Central Bank issued new measures and guidelines for Internet banking services. Starting from 1 June 2000, domestic banks were allowed to offer a full range of banking products and services over the Internet. By the end of 2003, 12 banks offered Internet banking services. However, the foreign banks were only allowed to set up communicative websites with effect from 1 January 2001 and transactional websites from 1 January 2002. The delay is consistent with the existing rule that imposes the no-branching restriction on foreign banks. The definition of branches legally includes electronic terminals. Up to 2004, the use of Internet banking was limited and surrounded by controversy. For example, the leading bank, the Malayan Banking Berhad, decided to impose a surcharge on those customers making an Internet transaction.

3.3.2 Financial deepening of the banking sector

Financial deepening refers to the use of financial instruments in relation to real output as development and modernisation is taking place. It is a slow process that takes time. Financial deepening occurs as long as the accumulation of financial assets is faster than the accumulation of non-financial wealth. It can also be seen as a monetisation process whereby money as a medium of exchange and store of value is being widely used in daily transaction. When a country pursues financial liberalisation vigorously, the results should be greater financial depth. Greater financial depth is good for supporting economic growth. Two indicators can be used to explain financial deepening:

a. Financial intermediation ratio. It is the ratio of assets in the financial system as a proportion of national income. Compared to others, this ratio is regarded as the most suitable indicator and is widely used in the literature. Another ratio that can be used is the ratio of money supply to national output. In relation to this, Chin and Jomo (2001, 2003) use the ratio of money supply M1 and M3 to Gross National Product (GNP) in explaining the financial development in Malaysia.[16]
b. The maturity structure of the securities, loans and deposits. The financial sector is said to go deeper when financial institutions are prepared to offer financial products on a long-term basis. For example, deposits may have a maturity period of more than 12 months and loans can be granted with a longer repayment period.

The last three columns in Table 3.4 represent the level of financial deepening in Malaysia. By 1980, the total assets of financial system exceeded national income (TA/GNP = 1.44). By the end of 2003, the size of the financial system was four

[16] Financial intermediation ratio is also used to indicate whether an economy is more liberalised (developed) or less liberalised (developed). A higher ratio implies more liberalised/developed economy. A lower ratio indicates less liberalised/developed economy.

times bigger than the real sector. This implies that the financial system is no longer underdeveloped. Increased monetisation in the economy is demonstrated by M3/GNP. By 1990, the money supply M3 was greater than the national income. As M3/GNP was greater than one, Malaysia showed a good performance in mobilising the savings. By the end of 2003, the money supply M3 was 150% bigger than the real sector. The rising ratio from time to time implies that financial liberalisation is an on-going process.

Table 3.4: Indicators of Financial Deepening in Malaysia

(1960 - 2003)

Year	TA[a] (billions)	M1[a] (billions)	M3[a] (billions)	GNP[a] (billions)	TA/GNP[b]	M1/GNP[b]	M3/GNP[b]
1960	4.00	0.48		5.84	0.68	0.08	n.a
1970	11.60	1.03	4.66	11.64	1.00	0.08	0.40
1980	74.15	9.76	32.69	51.39	1.44	0.19	0.63
1990	329.31	24.24	115.44	110.76	2.97	0.22	1.04
1995	747.10	51.92	271.94	212.10	3.50	0.25	1.28
2000	1243.00	78.22	456.50	312.15	3.98	0.25	1.46
2001	1295.20	80.72	469.52	308.68	4.20	0.26	1.52
2002	1,395.12	89.07	501.13	335.60	4.16	0.27	1.49
2003	1,564.06	99.70	548.43	369.40	4.23	0.28	1.53

Notes: a = TA is total assets of the financial system. M1 and M3 are measures of money supply. GNP is Gross national product measured at market price.

b. TA/GNP = the ratio of total financial assets to GNP. M1/GNP = the ratio of money supply M1 to GNP. M3/GNP = the ratio of money supply M3 to GNP.

Sources: Bank Negara Malaysia (1994, 1999, 2001b and 2003).

In 1979, the Central Bank introduced negotiable certificates of deposits as an alternative avenue for banking institutions to source funds. Initially, the maximum tenor was restricted to three years. However, in response to developments in the local debt market, the maximum was lengthened to five years. Further to the need to deepen the capital market, efforts were taken by the Central Bank to expand the range and tenure of debt instruments. In the early of 2001, the maximum tenure for the issuance of negotiable certificates of deposits was lengthened from five years to ten years. In addition, the variety and diversity of the longer dated papers in the debt market are also encouraged to cater for the specific needs of various groups of investors and to allow the banking institutions to find longer term funds for hedging purposes (Bank Negara Malaysia, 2001a).

3.4 Types and functions of commercial banks

According to Parkin et al. (2000), the commercial banks can be divided into two categories: retail and wholesale banks. A wholesale bank accepts a minimum deposit of GBP250, 000 and has a few branches that operate mainly in London. A retail bank accepts deposits as small as £1 and has extensive branch networks. In Malaysia; with the exception of few foreign banks, most of the banks that have extensive branch networks all over the country are retail banks. However, an interesting feature in the Malaysian banking sector is that the banks belong to two main groups: conventional banks and Islamic banks. The two differ with respect to the charging of interest in doing banking business. Most of the banks are conventional. The conventional banks use interest in accepting the deposits and in making loans and advances. In contrast, the Islamic banks are interest-free banks. Their operation avoids the use of interest but instead relies on Islamic principles in accepting the deposits and granting the loans and advances. For example, the deposits are accepted based on profit sharing. The depositors are considered as investors and the banks act as entrepreneurs. The rate of return is not pre-determined, but depends on the performance of the projects undertaken. Under Islamic law, the pre-determined rate of return constitutes a rate of interest and is totally prohibited. Further discussion of the Islamic banking system is presented in Section 3.6.

As the most important financial intermediaries, the commercial banks perform various functions. Freixas and Rochet (1997), for example, state that contemporary commercial banks have four functions: offering access to a payment system (facilitating transaction processes), transforming assets, managing risks and processing information and monitoring borrowers. In addition, Parkin et al. (2000) argue that the banks create liquidity, minimise the cost of funds, minimise the cost of monitoring borrowers and pool the risks. Allen and Santomoro (2001) offer new insights into the new role of financial intermediaries. They agree that the basic duty of the financial intermediaries is to reallocate the resources from the surplus units to the deficit units. However, they argue that this role may vary across time and from one country to another. They also observe that other forms of intermediaries such as pension funds and mutual funds have grown significantly. If this is so, then does the role of banks decline relative to other intermediaries? One way to determine whether the financial system has changed across time is by observing the distribution of financial assets based on the main types of financial intermediaries. By referring to the US financial system, Allen and Santomoro (2001) conclude that the share of assets held by banks had declined and the proportion of assets held in the form of pension funds, trusts and investment companies had grown.

Table 3.5 and Figure 3.1 show the distribution of financial assets by the major financial intermediaries in Malaysia. Between 1980 and 2003, the financial system underwent few changes. The banking system remained the largest component of the financial world led by the commercial banks. The share of the commercial banks was still approximately 40% of the total financial assets. A gradual increment can be seen in the provident and pension funds and insurance funds. This implies that in the long run, people started to save in long maturity products. However, it should be noted that the employees' contribution to the Employees Provident Funds are contractual savings. Therefore, it should not be concluded that the public have switched to long-term saving products. Part of the insurance funds like life insurance can be used to indicate the change in the pattern of holding the assets. As a kind of voluntary savings scheme, the share of insurance funds is rising. However, its contribution is still relatively low, only 5.2% of the total assets of the financial system.

Table 3.5: Distribution of financial assets by the major financial institutions

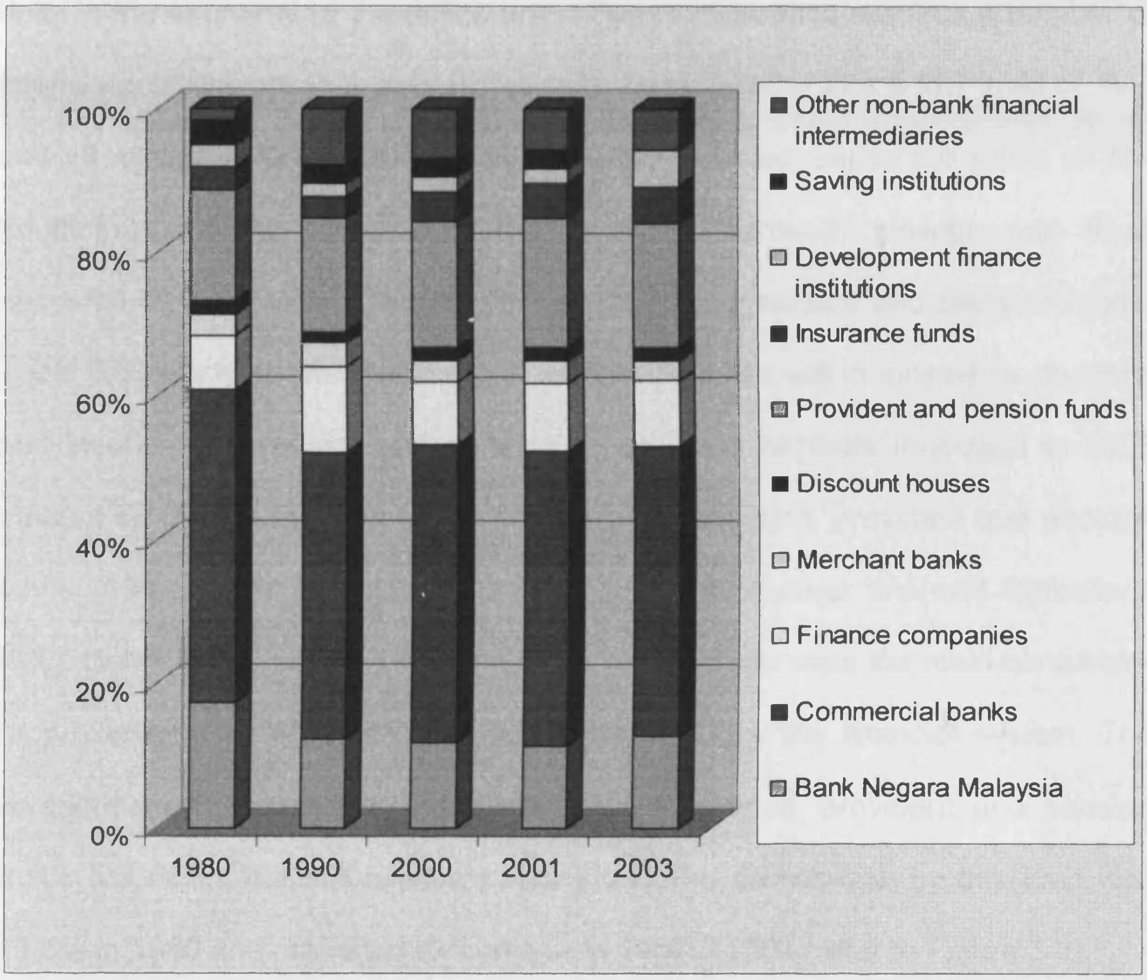
(1980-2003)

Financial intermediaries	1980	1990	2000	2001	2003
	(billions)	(billions)	(billions)	(billions)	(billions)
Banking system	54.3	226.9	829.9	865.8	1,046.7
Bank Negara Malaysia	13.0	40.9	148.9	149.7	200.8
Commercial banks	32.3	130.6	513.6	529.5	629.6
Finance companies	5.6	39.4	109.4	121.7	142.0
Merchant banks	2.2	11.1	36.9	41.1	44.1
Discount houses	1.2	4.9	21.1	23.8	30.2
Non-bank financial intermediaries	19.8	102.2	413.1	429.2	517.3
Provident and pension funds	11.3	52.2	217.6	231.4	266.4
Insurance funds	2.3	9.4	52.2	62.9	81.3
Development finance institutions	2.2	6.2	25.1	27.9	79.1
Savings institutions*	2.4	8.3	32.3	30.7	
Other non-bank financial intermediaries	1.3	26.1	85.9	76.3	90.5
TOTAL	74.1	329.1	1,243.0	1,295.0	1,564.0

Note:* Total assets of saving institutions are included in the development of finance institutions in 2003.

Sources: BNM (1994, 2001a,2003)

Figure 3.1: Distribution of financial assets by major financial intermediaries (1980-2003)



Note: The share of saving institutions in 2003 is already counted in development finance institutions.

3.4.1 Sources and uses of funds of the financial system

The main role of the financial institutions is to mobilise the funds from the surplus units in the economy to the deficit units. Such mobilisation requires a number of financial instruments that possess certain characteristics like a low level of risk, and an attractive rate of returns. In Malaysia, this mobilisation still relies on the traditional financial instruments like deposits (demand, savings and fixed deposits) as well as contractual savings such as provident and pension funds. Table 3.6 clearly shows that the bulk of the funds are still mobilised by deposits and insurance, provident and pension funds. Total deposits mobilised in 2003 amount to 45% of the total funds, followed by insurance, provident and pension funds (19%), other liabilities (14%) and funds from other financial institutions (6%). Between the period 1980 and 2003, the deposits were the main contributor (approximately 47%) to the total funds mobilised by the financial system. The contributions by alternative instruments like insurance, provident and pension funds are rising, but still relatively very slow. The contribution by this fund was 16.5% in 1980 and managed to increase to 19% in 2003 (refer to Figure 3.2).

The mobilised funds are classified into five main uses: loans and advances, securities, deposits with other institutions, gold and foreign reserve, and other assets. The data in 2003 shows that loans and advances are the main users of the mobilised funds, consuming up to 38% of the total funds, followed by securities (25%), deposits with other financial institutions (14%), gold and foreign reserve (11%) and other assets (10%). By referring to Figure 3.3, it can be seen that the allocation of funds between 1980 and 2003 has not significantly changed. Loans and advances remain the most popular use of the funds. In terms of securities, its composition has changed since 1990. The MGS were the largest component in 1990 (70%) while the corporate securities constituted only 30%. By the year 2000, the position changed when the corporate securities became the largest component (66%) while the MGS represented only 31%. This trend continued in 2003. We next move to the commercial banks to investigate whether the trend in the sources and uses of funds is similar or not. It is worth observing the behaviour of the banks since they are the backbone of the financial system in terms of the funds mobilised in the economy.

Table 3.6: Sources and uses of funds of the financial system

	1980 (million)	1990 (million)	1995 (million)	2000 (million)	2003 (million)
Sources of funds:					
Capital, reserves & profit	3,894.0	30,676.0	78,696.0	120,990.7	158,808.6
Currency	5,104.0	11,224.0	18,913.0	26,708.9	29,445.4
Demand deposits	5,326.0	20,539.0	46,156.0	83,205.3	92,094.8
Other deposits[1]	34,374.0	127,277.0	290,225.0	503,079.1	617,353.0
Borrowings	1,668.0	7,556.0	6,322.0	34,820.5	48,035.6
Funds from other financial institutions	3,915	19,642.0	78,122.0	67,603.3	88,877.7
Insurance, Provident & pension funds	12,218.0	60,882.0	127,870.0	236,640.1	297,068.8
Other liabilities	7,654.0	51,517.0	100,796.0	190,138.4	232,378.1
Total liabilities	74,153.0	329,313.0	747,101.0	1,263,186.3	1,564,062.0
Uses of funds:					
Currency	420.0	1,478.0	1,929.0	8,834.4	6,264.4
Deposits with other financial institutions	6,910.0	50,503.0	139,217.0	183,470.2	224,560.8
Bills	3,790.0	6,668.0	16,392.0	16,572.4	16,751.2
Loans & advances	29,892.0	141,560.0	305,751.0	512,428.5	600,844.8
Securities	17,991.0	80,219.0	160,281.0	281,460.8	393,321.3
Gold & foreign reserves	9,709.0	25,886.0	61,682.0	109,835.5	166,139.3
Other assets	5,441.0	22,,999.0	61,850.0	150,584.5	156,180.2
Total assets	74,153.0	329,313.0	747,101.0	1,263,186.3	1,564,062.0

Note: 1.Equals savings, fixed and other deposits plus negotiable instrument of deposits and repos.

Sources: Bank Negara Malaysia (1994, 1999 and 2003)

Figure 3.2: Financial system's sources of funds by types

(1980 – 2003)

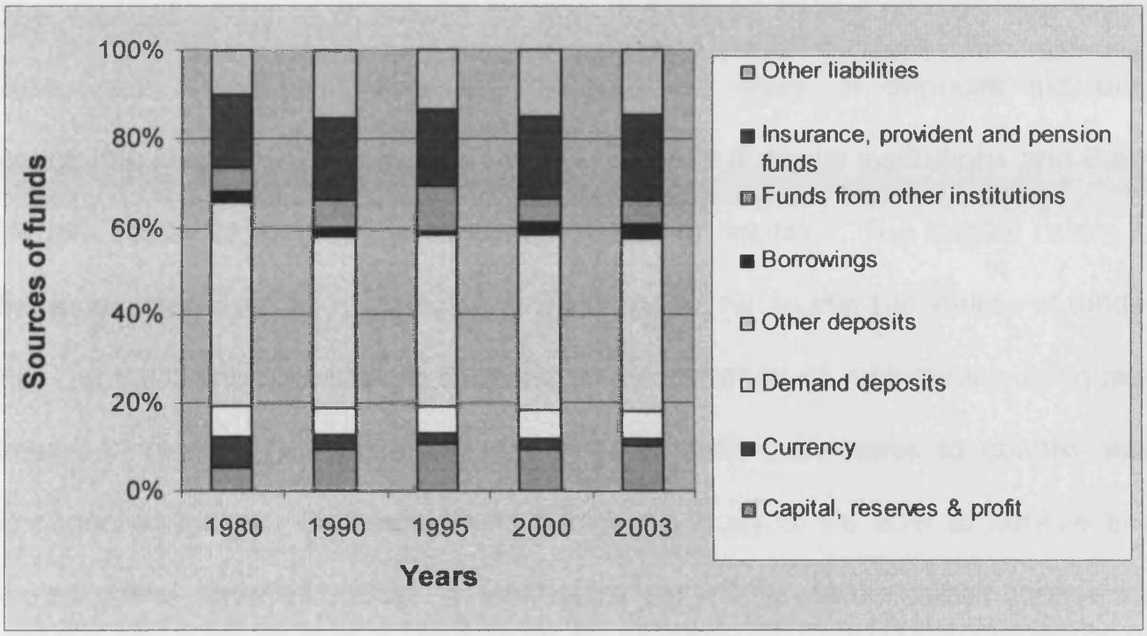
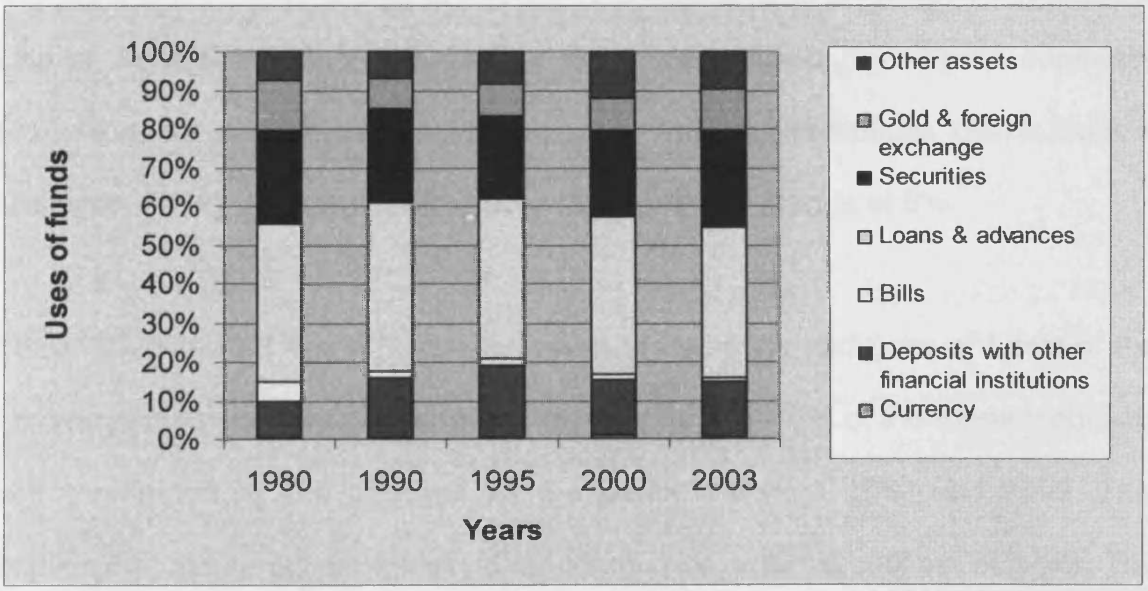


Figure 3.3: Financial system's uses of funds by types

(1980 – 2003)



3.4.2 Sources and uses of funds of commercial banks

The sources of funds mobilised by the commercial banks fall into five broad categories: capital and reserves, deposits (all types of deposits including negotiable instrument of deposits), amounts due to financial institutions and Bank Negara Malaysia, bankers acceptance and other liabilities. The capital refers to the funds raised by the shareholders of the banks. As an internal source of funds, the Central Bank has always stressed the importance of maintaining adequate capital in relation to the volume and risks of their businesses to counter any unexpected losses. Well-capitalised banks are likely to be able to survive any hard times or external shocks. In addition, there will be greater public confidence in the banking industry given a strong capital. Prior to 1989, there was no standard capital adequacy framework for all banking institutions. Commercial banks were required to maintain a minimum capital adequacy of 4% for domestic banks and 6% for foreign bank branches. Following the introduction of the Basle Capital Accord in 1988, the Central Bank streamlined the capital adequacy framework for commercial banks and other banking institutions (Bank Negara Malaysia, 1999). The capital adequacy ratio currently stands at 8%.

Table 3.7, Figure 3.4 and Figure 3.5 show the sources and uses of funds of the commercial banks. The data shows that, on average, 70% of the funds mobilised are contributed by the deposits for the period between 1980 and 2003. This implies that the banks are heavily dependent upon external sources of funds. The ratio of capital and reserves to total liabilities is about 8%. As the capital ratio is relatively low, the banks are vulnerable in terms of solvency, in particular when its deposit maturity does not match its loan maturity in the short term. The data show that more than 60% of the loans granted by the banks are for a term of more than five years while more than 60% of the fixed deposits have a maturity of less than one year. This phenomenon may place the banks into a liquidity problem with regard to the management of assets and liabilities.

Amongst the three main types of deposit, the proportion of fixed deposits has been the biggest. On average, the fixed deposits represent up to 50% of the total deposits received by the commercial banks. In 2004, the shares of each deposit (demand deposit, savings deposit and fixed deposit) are 20%, 14% and 66% respectively (Bank Negara Malaysia, 2004). In terms of maturity, 39% of the fixed deposits are held up to between 9 and 12 months and 34% up to one month in 2004. A higher percentage in the one month fixed deposit indicates that the depositors are sensitive toward any changes in the deposit rate. According to Hashim (2001), when the deposit rates are not stable or are uncertain, the fixed deposits would be transferred to other forms of deposits or other forms of investments either locally or abroad.

Table 3.7: Sources and uses of funds of commercial banks

(1980 – 2003)

	1980 (million)	1990 (million)	1995 (million)	2000 (million)	2003 (million)
Sources of funds:					
Capital & reserves	965.2	6,133.1	20,937.4	43,516.7	53,766.6
Deposits	24,257.4	82412.8	196,850.6	362,991.2	432,985.3
Amounts due to financial institutions and Bank Negara Malaysia	3,561.1	18,193.6	31,994.5	35,848.6	48,175.1
Bankers acceptance	546.3	2,706.7	15,781.1	15,778.4	18,503.9
Other liabilities	2,856.1	19,838.6	26,648.0	54,579.7	76,459.5
Total liabilities	32,186.1	129,284.8	292,211.5	512,714.7	629,890.5
Uses of funds					
Cash & reserves with Bank Negara Malaysia	2,422.8	6,204.7	20,108.5	15,388.0	17,118.9
Amounts due from financial institutions and Bank Negara Malaysia	2,693.1	15,897.0	45,257.3	84,542.8	111,148.0
Negotiable instruments of deposits	155.1	4,187.2	12,528.2	4,692.4	6,744.7
Marketable securities	3,948.4	12,639.35	24,434.4	56,908.4	81,842.4
Loans & advances	21,031.1	80,758.0	175,007.4	303,366.6	355,610.1
Others	1,935.6	9,598.6	14,875.7	47,816.6	57,426.4
Total assets	32,186.1	129,284.0	292,211.5	512,714.7	629,890.5

Source: Bank Negara Malaysia (1994, 1999 and 2003).

Figure 3.4: Commercial banks' sources of funds

(1980 – 2003)

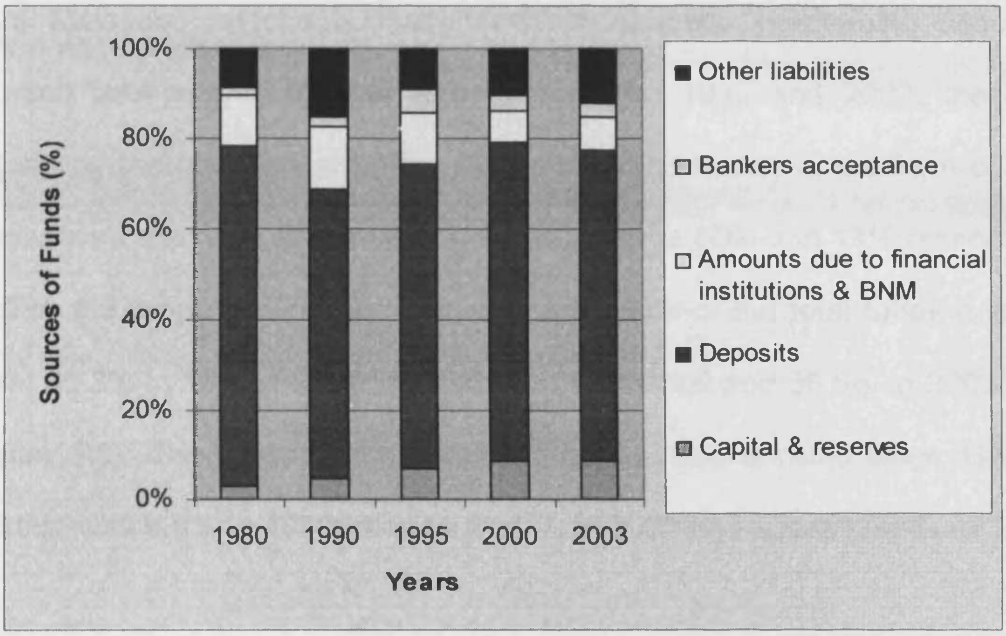
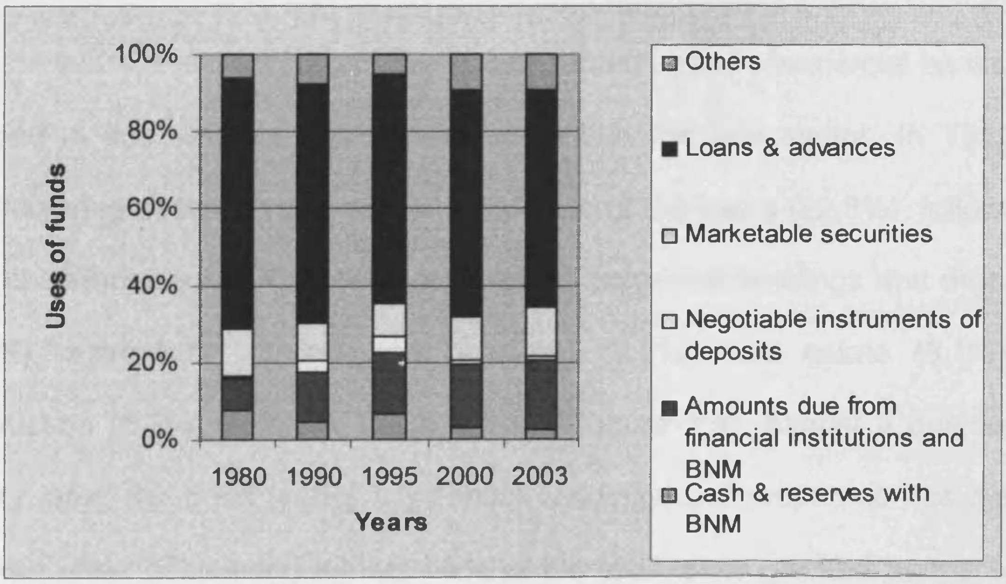


Figure 3.5: Commercial banks' uses of funds

(1980-2003)



The uses of funds of the commercial banks are made up of cash and reserves with Bank Negara Malaysia, amounts due from financial institutions and Bank Negara Malaysia, negotiable instruments of deposits, marketable securities, loans and advances and other assets. Between 1980 and 2003, the funds mobilised by the commercial banks were channeled mainly in the form of loans and advances and marketable securities, on average 60% and 11% respectively. However, the proportion of the loans and advances of the total funds is on the decline; 65.3% in 1980, 62.5% in 1990, 59.2% in 2000 and 56.5% in 2003 (refer to Figure 3.5). The proportion of marketable securities is rising since 1990. Its proportion was 9.8% in 1990, rose to 11.1 % in 2000 and subsequently to 13% in 2003.

3.4.3 Directions of loans of the commercial banks

Over the years, the directions of the loans granted by the commercial banks have changed in line with the rapid development in the real sector. In 1980, the manufacturing sector topped as the main user of the loans (22.3%), followed by general commerce (22.1%), the purchase of residential buildings and properties (10.6%), agriculture, mining and quarrying (8.8%), real estate (8.1%) and construction (6.7%) (refer to Table 3.8 and Figure 3.6). Almost a quarter of a century later, the purchase of residential buildings and properties has become the main user, consuming almost 35% of the total loans granted by the banks.

Manufacturing is second with 13.4%, followed by general commerce (9.2%), finance, insurance and business services (6.6%), consumption credit (6.5%) and construction (5.8%). The main changes in the directions of loans are from the so-called productive sectors (manufacturing and general commerce) to unproductive sectors (residential buildings and properties). This implies that the banks are moving towards less risky business activities. This is true since the loans granted to the purchase of residential building and properties are secured on these properties. The loans granted to manufacturing and general commerce in contrast are based on the viability of the projects. This is what Chin and Jomo (2003) state as a source of financial fragility or, to be precise, banking fragility. According to them, financial fragility is caused mainly by high loan exposure to asset price inflation, a build-up of private external debt and a surge in portfolio investment inflows. Ariff and Khalid (2000, pp. 49), on the other hand, argue that this phenomenon implies that more loans are granted on the basis of the connectedness of the borrowers to the government, big businesses and cross-border investments. They add that this idea of lending was wrongly advocated as the Asian way of doing business. In fact, this networked business disregards all the basic rules of lending. They conclude that the provision of an implicit guarantee by the presence of the government in the transactions was at the root of the problem.

Table 3.8: Commercial banks' classification of loans and advances by sectors

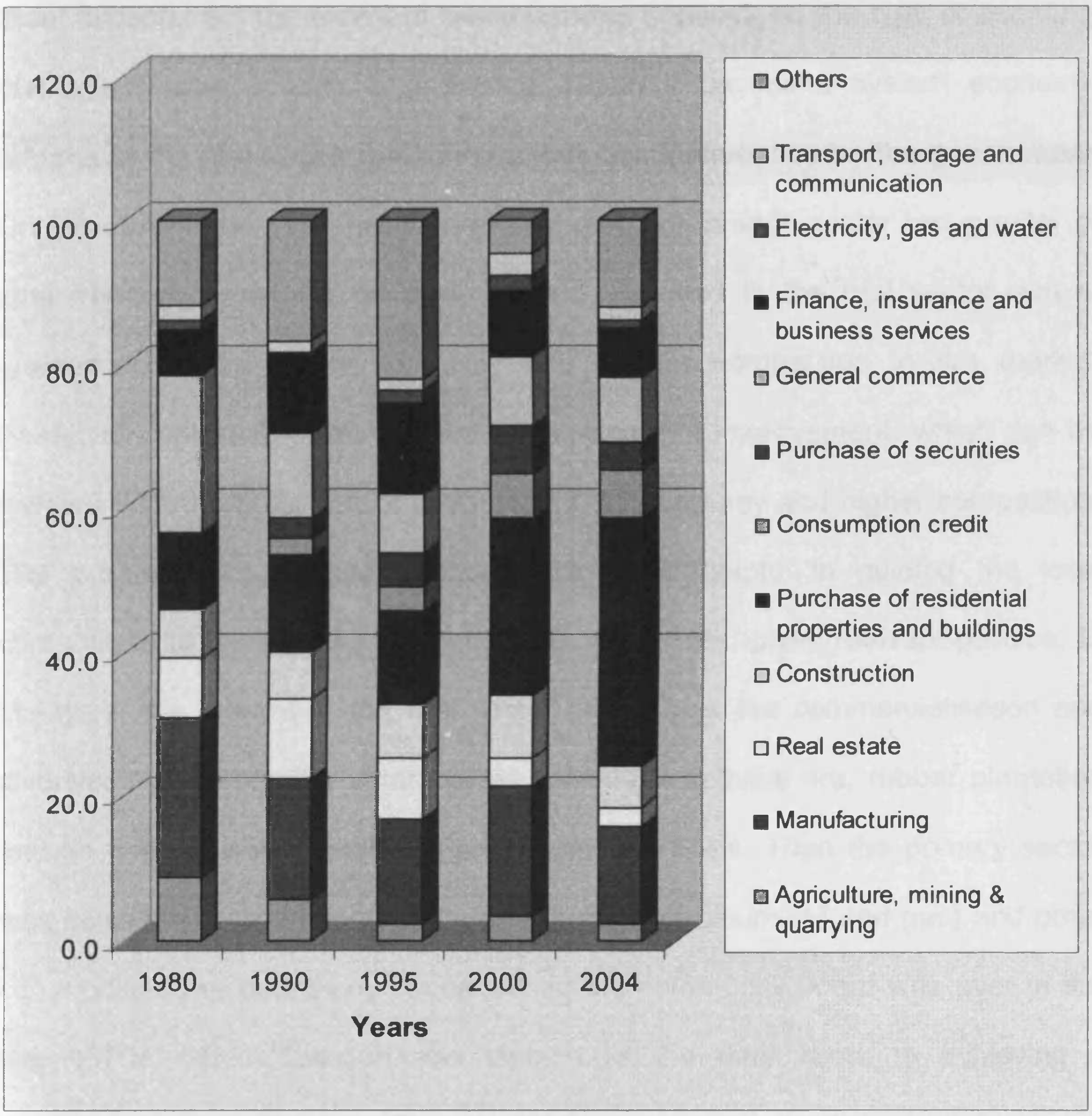
(1980 – 2004)

Sectors	1980	1990	1995	2000	2004
Agriculture, mining & quarrying	1,858.9	4,809.1	4,173.8	11,395.7	10,833.5
Manufacturing	4,693.8	13,697.4	2,6391.3	62,822.3	59,820.9
Real estate	1,710.2	9,080.1	15,816.7	13,672.6	12,299.1
Construction	1,406.6	5,417.7	13,386.6	30,361.9	26,082.2
Purchase of residential properties and buildings	2,232.4	11,054.2	23,473.0	84,963.3	15,4438.0
Consumption credit	n.a	1,906.1	6,354.1	20,861.0	29,205.0
Purchase of securities	n.a	2,280.6	8,177.2	23,178.4	17,195.8
General commerce	4,644.2	9,853.3	15,340.0	33,654.3	41,157.4
Finance, insurance and business services	1,297.3	8,834.3	22,798.9	30,647.3	29,686.3
Electricity, gas and water	278.9	201.4	3,573.8	7,005.2	4,756.6
Transport, storage and communication	400.1	1,317.2	2,918.7	10,750.7	8,349.2
Others	2,508.7	13,772.3	39,675.0	16,145.4	53,956.6
TOTAL	21,031.1	82,223.6	182,078.7	345,458.0	447,780.6

Sources: Bank Negara Malaysia (1994, 1999, 2001a, 2004)

Figure 3.6: Commercial banks' classifications of loans by sectors

(1980-2004)



3.5 Financial reforms and the banking crisis

An economy can be reformed in four broad aspects: real, financial, external and fiscal aspects, but the extent of these reforms depends on the type of economy that a particular country is practising. Given a command system economy, reforms in the real sector may bring about less intervention by the government. On the other hand, in a free-market economy, the private sector has greater or total freedom in making its own decision. Reforms in the real sector aim at greater openness of the economy and greater competition in the market. Production efficiency becomes the main aspect of improvement, which can be realised with the presence of openness of the economy and higher competition. The presence of foreign participation may be helpful in guiding the local competitors to achieve the pre-determined objectives, given relevant policies. In Malaysia, the reforms in the real sector bring about the commercialisation and diversification of primary commodities. Since the colonial era, rubber plantation and tin mining were carried out on commercial basis. Then the primary sector was diversified in the late 1960s with palm oil, petroleum (oil and gas) and other commodities like copra and cocoa. When the commodity boom was over in the late 1970s, industrialisation was chosen as the main route to achieving a developed country status by the year 2020.

Financial reforms will follow the reforms taken in the real sector. They focus on the easy entry into the market and openness of the financial sector. By removing the entry barriers, one may expect the inflow of foreign players into this sector. As highlighted by Yildirim (2003), the collapse of command system economies in the Central and Eastern Europe witnessed the inflow of foreign banks. In certain countries, these banks control up to 70% of the market. This reform in Malaysia will be discussed in details later.

Fiscal reforms are related to tax reforms and fiscal policies. Under tax reforms, tax rates will be reduced without jeopardising government revenue. The tax base may be diversified depending, on not only the direct tax, but also indirect tax. Tax collection needs to be improved. This, perhaps, requires a new technique in collecting taxes especially when this involves the private sector. In the case of Malaysia, the Inland Revenue Department has been corporatised with the hope that such a move will improve tax collection and yield greater revenue for the government. Fiscal reforms may also require the introduction of suitable fiscal policies to attract foreign investments. For example, the new fiscal policies may offer new tax incentives like pioneer status and tax exemption for re-investment purposes.

Reforms in the external sector involve two aspects: current account and capital account. Under current account, the reforms will remove any restrictions imposed on foreign currency trade. The foreign exchange is freely determined by market forces while intervention by the government via the central bank is kept to a minimum. Following the Asian financial crisis, the Malaysian ringgit has been pegged to the US dollar at RM3.80 for US dollar. Although it was argued that the move was to stabilise the ringgit in the short term, the pegging remains intact after almost eight years. Under capital account, the reforms will grant the freedom of flow for capital into and out of the country. Other aspects of reform in this sector include reductions in the trade barriers like quota, import duties and import licenses. In addition, foreign firms might find it easier to penetrate into the domestic market.

Financial reforms in Malaysia can be said to be selective. While one may expect that there should be freely determined market interest rates (deposit and lending rates) and greater foreign participation in the financial sector, the financial institutions still face lending guidelines from the Central Bank. As mentioned in the earlier section, the government influence over the banks takes the form of a requirement to lend to itself (the need to purchase government securities), to grant loans to certain priority sectors and certain categories of borrowers (in line with the national policies pursued by the government), setting up the interest rates (up to 1978) and controlling the entry of foreign banks including the license needed to set up new branches. After the BAFIA 1989, the foreign banks must be locally incorporated and should have more stringent capital requirements than the local banks (Jomo and Hamilton-Hart, 2001).

The main idea behind financial liberalisation is to improve the operating efficiency of the financial institutions and to promote competition amongst the players in the market[17]. One of the major reforms was the relaxation of government control over interest rates. Under section 37(1) (b) of the Central Bank of Malaysia Ordinance 1958, the Bank Negara Malaysia is empowered to determine the rates of interest payable to or by banks. However, since the early 1970s, interest rate determination had been gradually liberalised. The idea was to establish a responsive market-oriented banking system that would adjust each interest rate due to changes in the market conditions. After 1978, the deposit rates were totally free from the control of the Central Bank. All banks may determine their own rates for the deposits, subject to certain rules (Bank Negara Malaysia, 1994). With regard to lending rates, the commercial banks began to introduce their own base-lending rate (BLR) in 1981. The BLR was a rate based on their cost of funds. In 1983, all banks officially accepted the BLR system when the Rules of the Association of Banks in Malaysia (ABM) incorporated a rule requiring commercial banks to tie their lending rates (except loans to special groups or priority sectors) to the BLR.

The move to free the banking institutions' BLR from the Central Bank's administrative control was accompanied by the requirement for them to compute and quote their BLR based on a standardised formula to ensure objectivity and comparability. When the BLR was formally adopted in 1983, the cost components were not defined. Thus, a standardised format would ensure that the BLR would move in symmetry with changing market conditions and not be stuck downwards.

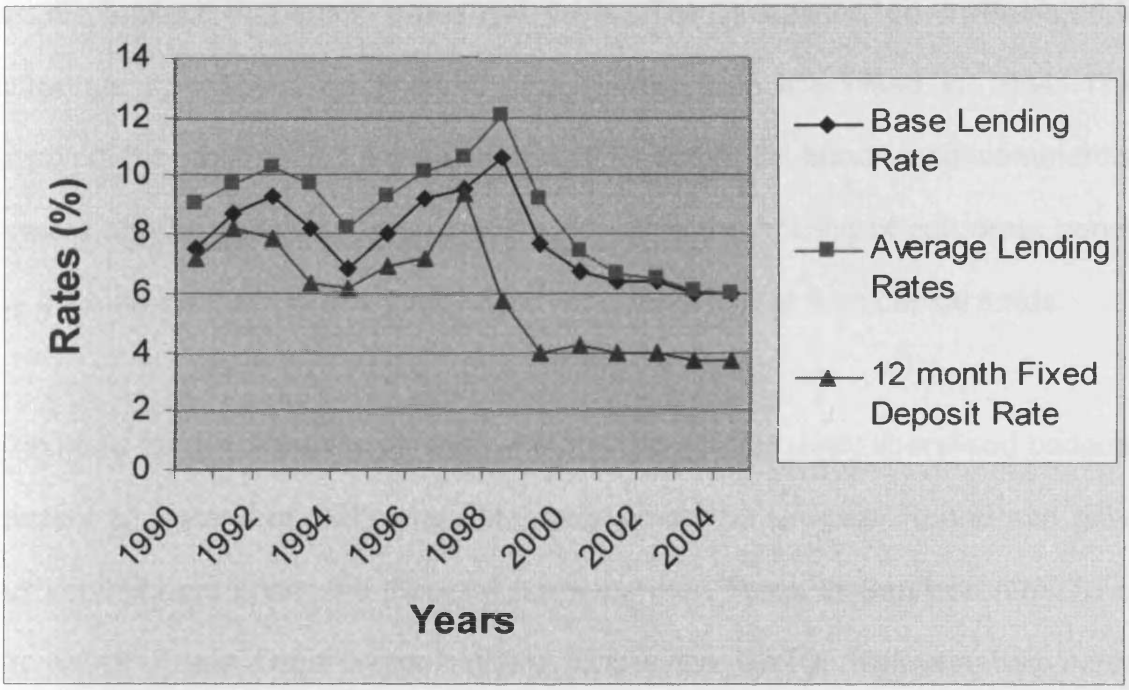
[17] Hashim (2001) states that the aims of deregulation comprise of opening local market, promoting transparency and accountability.

The Central Bank in 1991 introduced the standardised BLR formula that incorporated factors like funding cost, administrative cost and profit margin. Nevertheless, over the years, a number of weaknesses surfaced. The averaging of the banking institution's cost of funds has introduced a lagging effect and delayed the impact of the transmission of the monetary policy into the credit market. For example, the changes in direction of the monetary policy would conflict with the market lending rates. This would create confusion in the market (Bank Negara Malaysia, 1999). The BLR framework was revamped in 1995 in order to promote greater efficiency and, more importantly, to ensure the speedy transmission of changes in the direction of the monetary policy into the credit market. Under the revised framework, the banking institutions were free to quote their BLR below the ceiling BLR computed for the industry. This was considered a major step towards fostering greater competition between banking institutions and permitting greater flexibility in their lending strategies. The BLR framework was further refined in 1998 by substituting the weighted average interbank rate with the intervention rate.

Figure 3.7 shows the commercial banks' interest rates for the period 1990 to 2004. The BLR gradually increased from 1990 until 1992 and then declined in consecutive years. It rose to 8% in 1995 and reached the highest level of 10.6% in 1998 when the financial crisis was taking place. The same trend was shown by the average lending rate. After 1998, both rates dramatically declined. This was in response to an expansionary monetary policy that aims to speed up economic recovery. In 2004, both rates were equal. The deposit rate represented by the 12 month fixed deposit rate reached the highest point in 1997 during which there was a massive withdrawal of deposits from the domestic banks.[18]

Figure 3.7: Commercial banks' interest rates

(1990 – 2004)



Source: Bank Negara Malaysia (1999, 2004).

Besides relaxing the control over interest rates, progressive deregulation was also made with regard to the investment guidelines by the banking institutions. As

[18] There were reports that by early 1998 a total of RM54 billions (equivalent to 20% of GDP) was deposited outside of Malaysia. Only a fraction had returned whilst the rest remained in hard currencies (Ariff and Khalid, 2000).

stated in the previous section, with the introduction of BAFIA 1989, the commercial banks and other banking institutions were allowed to invest in shares of selected companies subject to prudential limits stipulated by the Central Bank. In 1991, greater autonomy for all banking institutions was granted to invest or have interest in trustee and selected non-trustee shares without prior approval from the Central Bank. Three years later, the scope of investment in non-trustee shares was widened to include any corporation listed on the Main Board of the KLSE, subject to certain prudential limits. The guidelines on investment in corporate bonds and commercial papers were also liberalised in 1994. This enabled the commercial banks to invest in corporate bonds and commercial papers approved by the Central Bank. However, the holding of corporate bonds by banking institutions was not allowed to exceed 10% of their capital funds.

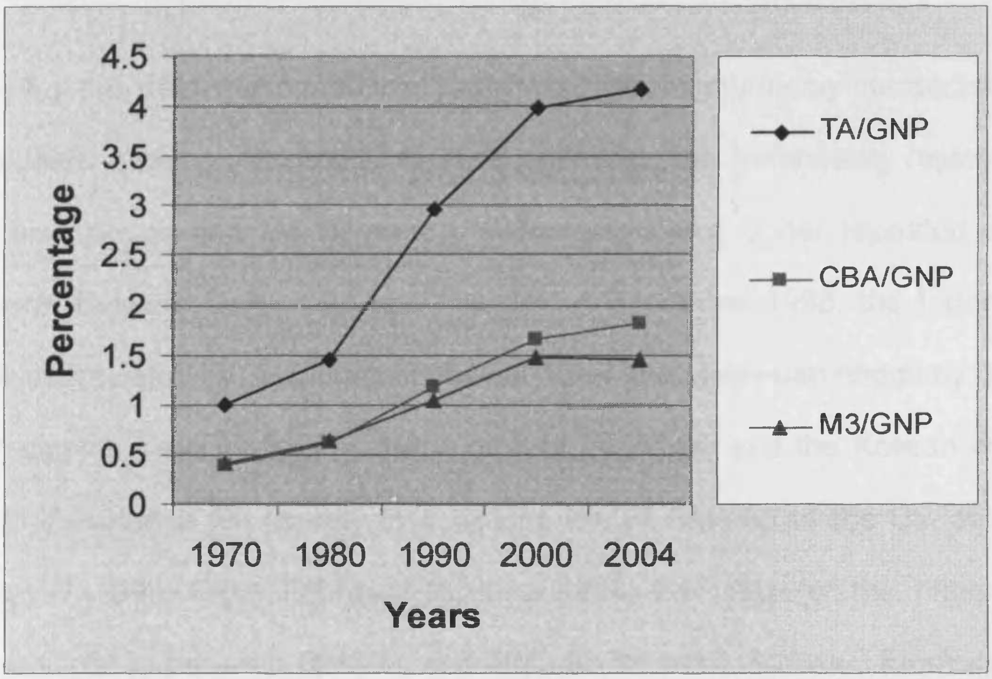
The need for the domestic banking sector to be progressively liberalised became evident at the end of 1993 when Malaysia joined the Uruguay Round and gave full commitment under the General Agreement on Trade in Services (GATS) of the World Trade Organisation (WTO). Under the GATS, Malaysia had made substantial commitments like an increase in the aggregate foreign equity limit in the insurance companies to 51%, liberalisation of offshore investment banking, offshore insurance and offshore financial leasing (Bank Negara Malaysia, 1999). The subsequent strategy was to develop a more competitive banking sector that would be able to respond to the changing needs of the economy for new financial products and services. A two-tier regulatory system (TTRS) was introduced in

1994 with the aim of providing the impetus for the emergence of a core of strong and well-developed commercial banks. Under the TTRS, the commercial banks, which were adequately capitalised and well managed, would be classified as Tier I institutions. To qualify for Tier-I status, commercial banks were required to meet minimum shareholders' funds of RM500 million by the end of 1995. The funds would then be increased to RM1 billions by the end of 1998 and eventually reach a target paid-up capital of RM1 billions by the end of 2000. Enjoying Tier-I status, the banks would be allowed to conduct certain aspects of their operations under a more liberal regulatory environment, as determined by the Central Bank from time to time. By the end of 1996, 11 banks qualified for Tier-I status. The TTRS was then extended to finance companies and merchant banks in 1996. It seems that the TTRS managed to segregate the weaker banks from the sound ones. It had a significant impact on the balance sheet of commercial banks.

One important effect of financial liberalisation is the deepening of the financial sector. One of the indicators of financial deepening is financial intermediation ratio (FIR). Figure 3.8 below shows the FIRs for the period 1970 until 2004. The ratio of financial assets to GNP (TA/GNP) was already 1 in 1970. It rose to 1.44 and 2.97 in 1980 and 1990 respectively. By 2004, the financial assets were more than four times bigger than GNP. The same trend is shown by CBA/GNP (ratio of commercial bank assets to GNP). The total assets of the commercial banks exceed national income in 1990. By 2004, CBA/GNP is 1.8 (180 percent bigger than national income). The FIR reflects the public's willingness to hold assets in the financial institutions like commercial banks and others. In addition, it also reflects the development of the banking institutions that could facilitate the holding of assets in such institutions. According to Ariff and Khalid (2000), this level of intermediation is very close to the values of developed economies. It could also reflect a greater commercialisation of economic activities over the years and the liberal policies that promote competition.

Figure 3.8: Indicators of Financial Deepening

(1970 – 2004)



3.5.1 Impact of financial crisis on the banking system

The East Asian Financial Crisis occurred in 1998 following the currency crisis in Thailand in the middle of 1997. While the actual causes of the crisis remain debatable, there are at least two main views with regard to the issue. The first is the fundamentalist view, which blames the structural weaknesses prevalent in the domestic financial institutions together with unsound macroeconomic policies and the issue of moral hazard. The alternative view is financial panic, which stresses the role of expectation, overreaction and over adjustment in handling the crisis (Bank Negara Malaysia, 1999).

Following the depreciation of the Thai baht, the neighbouring currencies also depreciated starting with the Philippine peso and the Indonesian rupiah. The Malaysian ringgit and the Singapore dollar also came under repeated selling pressure. Between July 1997 and the end of December 1998, the Indonesian rupiah depreciated by 70% against the US dollar, the Malaysian ringgit by 33.6%, the Philippine peso by 32.8%, the Thai baht by 29.4% and the Korean won by 25.5%. The ringgit fell sharply to a historic low of RM4.88 to the US dollar on January 7 1998. From February to June 1998, the value of the ringgit was relatively stable between RM3.84 and RM3.98 for the US dollar. From June to August, the ringgit faced further downward pressure following the depreciation of the Japanese yen, the decline in domestic economic activities and increased speculative activity (Bank Negara Malaysia, 1999). To stabilise the exchange rate, the ringgit was fixed at RM3.80 for the US dollar and selective exchange controls were imposed on September 2 1998.[19]

The immediate impact of the fall of the ringgit was on the stock market. The Bursa Malaysia, formerly the Kuala Lumpur Stock Exchange, declined by about 44% in the second half of 1997 and its composite index reached a low of 286 points on September 1 1998. Since then, the index has never returned to the pre-crisis level of well above 1,200 points. The market capitalisation was RM374.52 billions as at the end of 1998, compared with RM806.77 billions at the beginning of 1997.

The sharp decline of the ringgit together with the fall of the share prices eventually had an adverse effect on earnings and overall performance of the banking sector. Given the contraction in the real sector, bank borrowers found it difficult to service their debts. This resulted in deterioration in the assets quality of the portfolio of the banking institutions, causing the banks to adopt a pessimistic outlook of the economy that led to their being overly cautious in granting new loans. The reluctance of the banking institutions to grant new loans coupled with higher interest rates exacerbated the situation. Individuals and businesses found it very hard to obtain financing, and viable projects in productive activities were put on hold. As a result, the loan growth started to slow down from a high of

---

[19] This drastic change of policy followed the sacking of the Deputy Prime Minister who was also the Finance Minister on September 2 1998. As the real sector started to feel the severity of the currency crisis, it emerged that there was a rising conflict between the Prime Minister and his deputy with respect to appropriate measures to overcome the economic problem.

26.5% at the end of 1997 to 1.3% at the end of 1998 (Bank Negara Malaysia, 1999). At the same time, the rising level of non-performing loans also eroded the capital base of the banking institutions. While the Risk Weighted Capital Ratio (RWCR) for the banking system as a whole remained well above the minimum 8% requirement, some banking institutions required recapitalisation. In addition, merger and take-over are seen as business solutions. For example, Sime Bank was taken over by the RHB Bank while Bank Bumiputra merged with Bank of Commerce in 1999. However, if compared with Thailand, Korea and Indonesia, Malaysia did not need to seek emergency assistance from the International Monetary Fund. Thus, Malaysia avoided the forced closure of banks and financial institutions that took place in those countries. Such closure would have exacerbated the panic and undermined confidence in the domestic financial system. Realising the severity of the crisis and eager to avoid a deeper recession, the government and the Central Bank adopted a four-pronged approach to strengthen the resilience of the banking sector. This included a merger programme, the setting up of an assets management company, a special purpose vehicle to recapitalise the banking institutions and the Corporate Debt Restructuring Committee.

3.5.2 Banking consolidation

The commercial banking industry had gone through some degree of rationalisation. While new domestic banks were continuously established in the 1960s and 1970s, several rationalisation programs took place in the 1980s. Following the economic decline in 1985-86, three ailing banks needed capital injection. One of them, United Asian Bank merged with Bank of Commerce. In 1997, the need to increase banking institutions' capital prompted two more mergers in the industry. The first was between DCB Bank and Kwong Yik Bank, which produced RHB Bank. The second merger was between Chung Khiaw Bank and United Overseas Bank. It appears that the speed of mergers in the commercial banking industry was slow. However, the financial crisis in 1997-98 had exposed financial vulnerabilities in a few banks and given the much-needed push to the industry to consolidate. For example, the Bank Bumiputra, a state-owned bank, merged with another smaller bank (Bank of Commerce) in 1999 following serious financial problem faced by the former. Meanwhile, RHB Bank absorbed Sime Bank in the same year.

Although the Malaysian banking sector remained resilient and was able to perform its role efficiently and effectively, the Central Bank believes its ability must be strengthened in order to operate in a more competitive and increasingly difficult environment. In 1999, the government announced a major bank consolidation amongst domestic banks, which would eventually produce ten banking groups. (See Table 3.9 for merger programme of domestic banking institutions). Each banking group would comprise of commercial banks, finance companies and merchant banks. In fact, this decision was in line with the existing policy that the Malaysian banking sector has too many banks. The bank consolidation was expected to be completed by the end of 2000, but was extended until 2001 when two last banks (RHB Bank and Bank Utama) finally merged[20]. The merger had succeeded in consolidating the fragmented domestic banking sector without causing disruptions to the provision of banking services.

The merger programme undertaken by the banking sector has re-shaped the financial sector. In turn, the financial development should be aligned with the envisaged development of the rest of the economy. This is an important element behind the Financial Sector Master Plan (FSMP), which was launched by the Bank Negara Malaysia in 2001. The FSMP is the landscape for the financial sector until 2010. It outlines the medium and long term agenda and has three broad objectives:

- to enhance domestic capacity by building the capabilities of domestic banking institutions and increased deregulation in certain areas to increase competition
- to promote financial stability through strong, risk-adjusted prudential regulations and supervision

[20] The main disagreement between these two banks was which one should be the anchor bank. By size, it should be the RHB Bank but due to political reason, the Bank Utama was finally chosen as the anchor bank.

- to meet the socio-economic objectives of Malaysia, which includes increasing the level of consumer activism (Bank Negara Malaysia, 2001, p 112).

Following the consolidation programme for the domestic banking institutions, the newly created banking groups would need to assess their position from time to time. This includes benchmarking their position relative to that of their rivals in the market (both domestic banks and foreign banks). Through benchmarking, the domestic banks would be on the right track to gain the competitive edge and would be able to set realistic targets for potential improvement. After consulting with the banking industry, the Bank Negara Malaysia has produced a set of general benchmarks that can be used to measure the relative position of banking institutions. These benchmarks include indicators on operating and financing ratios, broad indicators on customer service and ratios to determine operational efficiency of the banking institutions.

Table 3.9: Merger programme for domestic banking institutions

	Original Anchor Banking Group	Merged with	Resultant Entity After Merger
1	**Affin Bank Berhad Group** Perwira Affin Bank Asia Commercial Finance Perwira Affin Merchant Bank	BSN Commercial Bank BSN Finance BSN Merchant Bankers	Affin Bank Affin ACF Finance Affin Merchant Bank
2	**Alliance Bank Berhad Group** Multi-Purpose Bank	International Bank Malaysia Sabah Bank Sabah Finance Bolton Finance Amanah Merchant Bank Bumiputa Merchant Bankers	Alliance Bank Alliance Finance Alliance Merchant Bank
3	**Arab-Malaysian Bank Berhad Group** Arab-Malaysian Bank Arab-Malaysian Finance Arab-Malaysian Merchant Bank	MBf Finance	Arab-Malaysian Bank Arab-Malaysian Finance Arab-Malaysian Merchant Bank
4	**Bumiputra Commerce Bank Berhad Group** Bumiputra Commerce Bank Bumiputra Commerce Finance Commerce International Merchant Bankers		Bumiputra Commerce Bank Bumiputra Commerce Finance Commerce International Merchant Bankers
5	**EON Bank Berhad Group** EON Bank EON Finance	Oriental Bank City Finance Perkasa Finance Malaysian International Merchant Bankers	EON Bank EON Finance Malaysian International Merchant Bankers
6	**Hong Leong Bank Berhad Group** Hong Leong Bank Hong Leong Finance	Wah Tat Bank Credit Corporation	Hong Leong Bank Hong Leong Finance
7	**Malayan Banking Berhad Group** Malayan Banking Mayban Finance Aseambankers Malaysia	The Pacific Bank PhilleoAllied Bank Sime Finance Kewangan Bersatu	Malayan Banking Mayban Finance Aseambankers Malaysia
8	**Public Bank Berhad Group** Public Bank Public Finance	Hock Hua Bank Advance Finance Sime Merchant Bankers	Public Bank Public Finance Public Merchant Bank
9	**RHB Bank Berhad Group** RHB Bank RHB Sakura Merchant Bankers	Delta Finance Interfinance	RHB Bank RHB Delta Finance RHB Sakura Merchant Bankers
10	**Southern Bank Berhad Group** Southern Bank	Ban Hin Lee Bank United Merchant Finance Perdana Finance Cempaka Finance Perdana Finance	Southern Bank Southern Finance Southern Investment Bank

Notes:

Source: Bank Negara Malaysia (2001a, p. 111)

3.6 Islamic banking system

The Islamic financial system encompasses the Islamic banking system, Islamic money market, Islamic insurance or Takaful, Islamic capital market and the specialised financial institutions that provide alternative sources of financing. The Islamic banking system is the backbone of the Islamic financial system. It begins with the establishment of the first Islamic bank (Bank Islam Malaysia Berhad) in 1983. The second Islamic bank (Bank Muamalat) was established in 1999[21]. Islamic banking system is governed by the Islamic Banking Act (1983). Since its launch in 1983, the Islamic banking system has been well accepted by the mostly Muslim community and non-Muslim customers. To strengthen the Islamic banking system, conventional banks were allowed to open Islamic counters by offering an interest-free banking scheme in March 1993. The three largest domestic banks were selected to offer this scheme on a pilot basis. Six months later, another six banks joined the scheme. By 2001, four foreign banks had joined this scheme by offering interest-free financial products. They were Citibank, HSBC Bank, OCBC Bank and Standard and Chartered Bank.

The Financial Sector Master Plan, which was launched in 2001, provided a clear strategic focus to develop and promote the expansion of the Islamic banking system. One of the strategies was to allow a banking group to set up its own Islamic subsidiary instead of just an Islamic window. Thus, on March 1 2005, a

[21] The creation of the second Islamic bank (Bank Muamalat) is actually a by-product of the merger between two banks: Bank Bumiputra and Bank of Commerce.

new era of Islamic banking emerged when an Islamic subsidiary of a banking group was established. The RHB Banking Group launched its own Islamic bank, called RHB Islamic Bank Berhad. The move was in line with the Central Bank's plan of strengthening the institutional structure for the Islamic banking operation. One of them was by transforming an existing Islamic banking window to an Islamic subsidiary. The Islamic subsidiary would have its own chief executive and a board of directors solely in charge of that bank. The Islamic banking products and services would continue to be available at the existing branches of the conventional commercial banks. Similar to the existing Islamic banks, the Islamic subsidiaries would be operating under the Islamic Banking Act 1983.

By the end of 2003, the Islamic banking system comprises of two Islamic banks and 31 conventional banking institutions (nine domestic banks, four foreign banks, seven finance companies, four merchant banks and seven discount houses). These conventional banking institutions join the Islamic banking scheme by setting up the so-called Islamic counters. Besides their conventional products, these Islamic counters offer Islamic financial products, which are free from interest. At present, the Islamic banking system offers a wide range of financial products and services ranging from savings, current and investment deposit products to financing products such as property financing, working capital financing, project financing, plant and machinery financing, hire purchase, education financing and others. The ability of the Islamic banking institutions to design and offer attractive products at competitive prices has appealed to both

Muslim and non-Muslim customers. In addition, the extensive distribution network of Islamic banking institutions is supported by an efficient, secure and effective payment system. In 2003, the Islamic banking system accounts for 10% of the total assets in the banking system. Based on the FSMP, it is intended that its share will increase to 20% by the year 2010.

3.7 Concluding remarks

This chapter provides the background and environment of the Malaysian banking industry. The banking system is one of the three components of the financial system besides the non-banking institutions and the financial market. Since its early stage, the banking industry has grown in terms of size, facilities and financial products. This has been discussed based on financial widening and deepening.

The sources of funds of financial system come from the deposits mobilised by the banking and non-banking institutions. In term of the uses, loans and advances are the main uses. This trend continues as we analyse the sources and uses of funds of commercial banks. The deposits contribute about 70% of the funds mobilised. The directions of loans granted by the banks have changed from the manufacturing sector and general commerce to the purchase of residential buildings and properties.

Financial reforms did take place in Malaysia, but they remain selective in nature. The government abolished the fixed interest rate in 1978 and since then the banks independently set their interest rates (deposit and lending rates). However, lending guidelines issued by the BNM will remain as long as the relevant policies, like NEP, are still pursued by the government. The financial liberalisation also helps to facilitate the process of financial deepening and widening. The financial products have a greater choice as new products were introduced into the market.

The health of the banking industry was finally affected by the Asian financial crisis. This led to the merger of a few banks, but sparked greater bank consolidation in 1999. Although the number of banks has declined, the domestic banks are now better capitalised and, on average, are bigger in size. As outlined in the FSMP, the banking consolidation aims for greater efficiency and capacity that subsequently will enable the domestic banks to enjoy a competitive edge against the foreign counterparts.

The Islamic banking system remains as a major issue for Malaysian financial development. As stated in the FSMP, this sector is set to control at least 20% of the total assets of the banking system by 2010. Future development will see the growth of Islamic banks as subsidiaries of the existing banking groups. However, an issue that remains a major challenge to the Islamic banking system is its ability to offer competitive products at least on a par with the products offered by the conventional banking system.

Chapter

Four

Important aspects of the Malaysian banking system that can be related to our study are financial liberalisation and bank consolidation. Financial liberalisation aims for higher efficiency and a more competitive environment. Although the financial reforms are still selective, the key development was that the banks are now free to set their own deposit and lending rates. Both rates represent the price of deposit and the price of outputs. Thus, this freedom together with financial deepening can bring about significant influence over banking efficiency and competitive environment.

Another critical development is the recent merger of domestic banks. The basic argument behind this policy is the pursuit of higher efficiency. It was argued that by being bigger in size and by having the ability to operate on a larger scale; the domestic banks might enhance their efficiency. It is hoped that by being bigger and more efficient, the domestic banks can face stiffer competition from foreign counterparts. The support for the merger policy can be demonstrated if we manage to establish a positive relationship between bank efficiency and bank size.

CHAPTER FOUR

Analysis of technical efficiency: A DEA approach

4.0 Introduction

This chapter has two main objectives. The first is to measure the efficiency of the commercial banks using a non-parametric approach, popularly known as data envelopment analysis (DEA). The DEA method then divides the efficiency into three sub-categories: overall technical efficiency, pure technical efficiency, and scale efficiency. This division is needed in order to identify and determine the sources of inefficiency in the banking sector. These sources of inefficiency are of great interest to bankers. They can offer golden opportunities and new directions for further improvement in the banks' operations, and enhance banks' competitiveness.

Banking performance is then compared across all commercial banks located in Malaysia. In this case, the comparison is based on the grouping of the banks by their ownership: domestic and foreign banks. The main question is whether ownership can play an effective role in explaining the variability in banking performance. Elyasiani and Rezvanian (2002), for example, raised the issue of whether the production technology of the foreign banks is different in principle from that of the local banks. If that occurs, then any reduction in costs can be materialised through foreign ownership. Having obtained the technical efficiency scores, the profiles of efficient and inefficient banks are then established. Such profiles are useful in preparing practical benchmarks for the banking industry, since inefficient banks would surely follow their efficient counterparts.

The second objective is to construct the Malmquist productivity index, which then helps to identify the sources of productivity growth of the banks. Given the availability of panel data set, this study will construct the Malmquist productivity index using DEA-like programming and then identify the sources of the growth i.e. technical change or shifting-up effect, and efficiency improvement or catching-up effect.

It should be noted that any analysis of banking efficiency is relevant in an attempt to find some justification for supporting or rejecting the recent bank consolidation after the last financial crisis. The banking consolidation was finally completed in 2001 and resulted in a reduction in the number of local banks from twenty to ten banking groups. Two foreign banks also merged in 2002. Although the number of banks has decreased, the average size of the domestic banks is bigger than before. Before 2000, the average size of the domestic banks was twice the average size of the foreign banks. However, after 2000, its average size was three times bigger than the average size of the foreign banks.

This chapter has six sections and is organised as follows. Firstly, we introduce the concept of efficiency and how it is measured. The non-parametric approach is presented and its case is discussed. Next, we discuss the measurement of bank input and output and how the methodology will be carried out. This methodology is confined to the measurement of efficiency using the non-parametric approach. In Section 4.3, we introduce the DEA model and how it works. Two main models are put forward: the CCR model, based on the work of Charnes et al. (1978), and the BCR model, based on the work of Banker et al. (1984). Next, we present the analysis of the efficiency measure. Three aspects are taken into account. They are the components of efficiency, the comparison of efficiency scores across all banks grouped by ownership, and the profiles of efficient and inefficient banks. Several key indicators are used, such as size of the bank, market power, rate of return and asset quality. Section 4.5 is about the construction of Malmquist productivity index using DEA-like programming. The study will then identify the sources of productivity change over the sample period. Finally, the last section provides some concluding remarks.

4.1 Concept and measurement of efficiency

From an economic point of view, efficiency is said to be all about the relationship between scarce factor inputs and the output of goods and services. This relationship can be seen and evaluated in terms of either physical output or cost.

If we wish to identify and determine the best possible (optimal) combination of inputs to produce a given level of output in physical terms, then we are concerned with technological or technical efficiency. This implies that we are looking for the minimum level of inputs needed to produce a certain amount of output. It also implies that technical inefficiency is caused by the failure to achieve the best possible output levels and/or usage of an excessive amount of inputs. On the other hand, if we want to determine the optimal combination of inputs that will minimise the cost of producing a given level of output, then we are talking about economic efficiency or cost efficiency. This kind of efficiency requires the availability of input prices such as the price of labour and capital.

As is already known, any business entity has a set of inputs or resources that is ready to be employed and has output to be produced i.e. products or services. It is in the context of this relationship that the performance of the firm needs to be addressed and assessed, hence the efficiency analysis. Farrel (1957) tried to measure the efficiency of a firm in the single input-output case. It involved the measurement of technical and allocative efficiency and the derivation of the efficient production function. In the case of the efficient production function, he suggested the use of either a non-parametric piecewise linear convex frontier or a parametric function such as the Cobb-Douglas form. However, his attempt was considered incomplete since he did not provide a way to summarise all the various inputs and outputs into a single virtual input and single output.

Farrel's idea was later picked up and extended by Charnes et al. (1978). They proposed a model that can generalise the single input-output ratio of efficiency of a single decision making unit in a multiple input-output setting. Technical efficiency is measured as a ratio of virtual output produced to virtual input used. Their work was later popularly known as the CCR model (after their names), and latterly generated the Data Envelopment Analysis (DEA) approach.

4.2 Bank inputs and outputs

Commercial banks perform various functions. In Malaysia, a commercial bank offers various financial products, including conventional deposits (demand and time deposit), negotiable certificates of deposits, and other financial services (foreign exchange, business consultancy etc.) Based on its nature of business, the bank is best seen as a multi-product firm.

This study acknowledges the fact that a bank can be broadly seen as either a production unit or an intermediate unit. The difference between these two approaches has been discussed in the literature review. One of the debatable issues concerns the role of a bank deposit. Does it constitute a bank input or a bank output? In this study, we use bank deposit as either input or output. The result shows that, regardless of its position, it does not have a significant influence over the efficiency measure. It turns out that the difference is only 2%.

This result is in line with the findings of Favero and Papi (1995). Therefore, we use bank deposit as an input in line with our treatment of the bank as an intermediation unit. The same action was taken by Maudos and Pastor (2003), Cavallo and Rossi (2002), Rezvanian and Mehdian (2002) and Miller and Noulas (1996).

In this study, we identify and select other bank inputs and outputs using the intermediate approach. Three inputs are used; the number of employees (LAB), fixed assets (FA) and total deposits (TD). TD is made up of demand deposits, saving deposits, and fixed deposits. In contrast, Katib and Matthews (1999) use total deposit as bank output in their study of banking performance in Malaysia. The three outputs are total loans (LOANS), other earning assets (OEA), and other operating income (OOY). The OOY variable is selected to reflect the growing contribution of non-interest income to a bank's total income. In 1999, the average OOY of the banks in the sample stood at 11.38%. The inclusion of this variable is in line with the works of Maudos and Pastor (2003), Yildirim (2002), Siems and Barr (1998). The inclusion of this output was initially suggested by Roger (1998).

4.2.1 Descriptive statistics of the data

This study aims to include all commercial banks but not Islamic banks. The sample period runs from 1994 to 2000. The study starts from 1994 because from this period the data of foreign banks are available. Because of the BAFIA 1989, the foreign banks are now locally incorporated and are required to publish their own financial statement. However, not all foreign banks managed to be locally incorporated by 1994. Only seven did so but the rest were finally incorporated a year later. Most of the foreign banks had been operating Malaysia for a long time, but remained as foreign branches. They were not required by the law to publish their accounts. There were 22 local banks in 1994. However, because of the incomplete data provided by BANKSCOPE, only 13 local banks were included in the sample. There were 20 banks in the sample in 1994.

In 1995 and 1996, there were 37 commercial banks in Malaysia but only 32 banks were included in the sample. Out of 32 banks; 21 were local banks and 11 foreign banks. Until 2000, the number of foreign banks in the sample remains at 11. The number of local banks starts to fall in 1997 due to merger activity. The number was 31 in 1997, 30 in 1998, 28 in 1999 and only 9 in 2000. In 2000, data of some domestic banks were not available since the banks were about to merge.

Table 4.1 shows the descriptive statistics of the bank inputs and outputs used. Except for LAB, other variables are in real value (1994 = 100). The total number of bank-year observations for the period 1994 – 2000 is 193. From one year to another, the number of banks is not balanced, due to the unavailability of data and merger activities.

After 1997, some data were not included due to merger activities amongst the banks. As stated in Chapter 3, merger activities have been taking place, particularly since the financial crisis. For example, the bank that is currently the second largest (Bumiputra Commerce Bank Berhad), which was created in 1999 as a result of a merger between two banks (Bank Bumiputera Malaysia Berhad and Bank of Commerce), is excluded from our sample. This also means that data for Bank Bumiputera and Bank of Commerce cease to exist in 1999 and 2000 respectively. The same happens to RHB Bank, which merged with Kwong Yik Bank in 1997 and with Sime Bank in 1999. Its data for the period between 1997 and 2000 were excluded. Because of the unavailability of data and merger activities, the number of banks in 1994 and 2000 is only 20. The average number of banks for other years (1995 to 1999) is 30.

It must be noted that the Malaysian financial system was badly affected by the Asian Financial Crisis (1997-98). The crisis had exposed the banking sector into liquidity problem with higher interest rates and higher non-performing loans. The turbulence finally called for speeding up the bank consolidation. It started in 1999 when the second largest bank (Bank Bumiputra) merged with Bank of Commerce. Then, the Central Bank announced a major bank consolidation that would reduce the number of domestic banks from 21 banks in 1999 to 10 banks by 2001. Consequently, the merger not only reduced the sample size but also made it unbalanced. However, given that the population size of the banks is already small, we chose unbalanced panel so that the sample size could be maximised. The newly merged banks were excluded from this study. This is to be consistent with the current trend in the literature.

The financial crisis has no doubt brought about high volatility in banking activities. There was evidence as stated by Ariff and Khalid (2000) that substantial deposits were transferred to foreign accounts and that non-performing loans were on the rise due to higher lending rates. As a consequence, the banks suffered losses and some of them required recapitalisation with the help of a special purpose vehicle set up by the government.

It should be restated that, unlike the parametric approach, the DEA technique includes measurement error when measuring the deviations from the best-practice frontier. Given the high volatility in banking activities, annual estimates of technical efficiency might reflect this volatility. Thus, the DEA estimates of efficiency might be better seen over medium term rather than short term. Annual estimates of technical efficiency should be interpreted cautiously.

The main source of the data is BANKSCOPE, which compiles the balance sheets of commercial banks all over the world for up to eight years for each bank. Each balance sheet and income statement is transformed into a standard format. We also use the Banker's Directory published by the Association of Banks in Malaysia every two years. This report contains data on labour, number of bank branches and bank shareholders.

4.3 Methodology

As outlined by McKillop et al. (2002), the usual DEA procedure for measuring efficiency begins with collecting data on input and output quantities. This data is used to construct a non-parametric frontier or best practice amongst the decision-making units (DMUs). Finally, an efficiency score for each DMU is measured in relation to this frontier. Under DEA, there are two basic measurements, which are based on the assumption of constant returns to scale (CRS) and variable returns to scale (VRS). The CRS model is based on the pioneering work of Charnes et al. (1978) and the VRS model is the work of Banker et al. (1984).

The set of three inputs and three outputs was run under both assumptions; CRS and VRS. If the efficiency score of each DMU produced by these models varies, then the banks are said to experience variable returns to scale (Avkiran, 1999). In addition, under the VRS, a model can be orientated either by using input minimisation or output maximisation. This orientation is important in order to seek any potential areas of improvement. By input minimisation, what we mean is how to seek potential improvement expressed in terms of how input or resource levels could decrease while maintaining the current production level. It can be used to address a question such as by how much input quantities can be proportionally reduced without changing the output quantities. This is important when a firm is facing a slump in business due to declining demand, for example. On the other hand, output maximisation means that we seek potential improvement in terms of how production could increase, given the current input level. This is crucial when a firm is facing the question of by how much output quantities can be proportionally expanded without altering the input quantities employed. This orientation is applicable in the case of rapid business expansion due to a continuous increase in demand.

Table 4.1: Descriptive Statistics of Bank Inputs and Outputs

1994 (n = 20)	**LAB**	**TD**	**FA**	**LOANS**	**OEA**	**OOY**
Mean	1740.2	4630.2	76.5	3797.8	3125.1	54.6
Std Deviation	1830.6	4774.9	110.4	3574.2	4624.2	57.6
Maximum	7633.0	17768.8	463.7	13378.1	17818.6	210.0
Minimum	116.0	283.0	2.0	271.0	76.0	2.0
1995 (n = 32)						
Mean	1730.8	5187.5	87.0	4679.9	3003.2	69.0
Std Deviation	2347.6	7354.2	142.3	6331.8	4794.8	96.3
Maximum	10824.0	36282.4	622.2	32543.0	23276.8	450.3
Minimum	70.0	131.4	2.1	0324.3	100.3	0.7
1996 (n = 32)						
Mean	1877.9	6426.4	101.7	5895.0	3415.1	79.1
Std Deviation	2452.8	8067.2	147.3	7213.4	5045.8	98.4
Maximum	11200.0	39300.2	620.7	36374.0	25404.8	452.4
Minimum	71.0	173.8	2.2	299.0	166.4	4.0
1997 (n = 31)						
Mean	1900.5	8246.7	113.0	7167.9	4543.3	97.6
Std Deviation	2504.2	9851.7	148.8	8539.3	6559.8	123.3
Maximum	11521.0	46918.4	627.2	45297.7	31326.1	586.3
Minimum	72.0	181.0	1.7	334.5	132.9	4.7
1998 (n = 30)						
Mean	2083.8	8535.0	127.3	8108.3	4467.5	117.4
Std Deviation	2555.8	10684.5	160.1	10533.0	5678.4	159.7
Maximum	12000.0	51894.4	693.8	56277.2	24470.1	725.4
Minimum	73.0	350.0	2.1	263.5	161.3	-4.6
1999 (n = 28)						
Mean	1890.3	8611.7	122.1	7449.5	4713.9	107.8
Std Deviation	2343.1	11214.8	165.2	10516.2	6032.2	152.5
Maximum	12100.0	57581.1	725.4	57489.4	26646.6	772.0
Minimum	74.0	320.7	1.9	212.6	251.8	3.3
2000 (n = 20)						
Mean	2135.0	10781.5	150.0	9466.7	6453.7	150.2
Std Deviation	2729.9	13348.4	200.8	12957.9	7913.3	186.6
Maximum	12200.0	60260.4	792.2	61003.9	32091.0	800.7
Minimum	75.0	508.4	1.7	146.3	360.7	4.6

Notes: a. LAB is the number of bank employees. TD is total deposits. FA is total fixed assets. LOANS is total loans issued by the banks (overdraft, term loans and others). OEA is other earning asset and OOY is other operating income. n is the number of commercial banks.

b. LAB figures for 1995, 1997 and 1999 are replacement value. The method used is mean substitution for each of the bank involved as suggested by Hair et al. (1998).

c. Figures are in thousands of ringgit Malaysia (RM) except for the number of bank employees.

Sources: ABM (1994, 1996, 1998 and 2000) and BANKSCOPE

According to Färe and Lovell, (1978, quoted in Coelli, 1996), the output and input-oriented measures will provide equivalent measures of technical efficiency when constant returns to scale exist. This efficiency will no longer be equal when variable returns to scale are present. Given this choice of orientation, the question of which one is appropriate is not very crucial. Many previous studies are more inclined to select input-oriented models because many firms or DMUs have particular demands to fill. Moreover, in some industries, the DMUs may be given a fixed amount of resources and asked to produce a certain amount of output. Therefore, input quantities appear to be of greater concern than output quantities. However, this argument may not be applicable in all industries. In principle, one should select an orientation based on which quantities (inputs or outputs) the DMUs have most control over.

4.3.1 Strengths and weaknesses of DEA

Hababou (2002) and Avkiran (1999a) provide a relatively thorough discussion of the merits and limits of the DEA. Amongst the strengths of the DEA is that there is no need for a preconceived structure or specific functional form to be imposed on the data in identifying and determining the efficient DMUs (Hababou, 2002; Favero and Papi, 1995 and Banker et al. 1984). Hababou (2002) adds that it is better to adopt the DEA technique when it has been shown that a commonly agreed functional form relating inputs to outputs is difficult to prove or find. It is truly difficult to show such specific functional form for financial services entities. Avkiran (1999a) acknowledges the limits of the DEA by stating that this technique allows researchers to choose any kind of input and output of managerial interest, regardless of different measurement units. There is no need for standardisation.

In addition, Charnes et al. (1994, quoted in Hababou 2002) put forward three useful features of the DEA. The first is that each DMU is assigned a single efficiency score, thus allowing ranking amongst the DMUs in the sample. Secondly, it highlights the areas of improvement for each single DMU. For example, since a DMU is compared to a set of efficient DMUs with similar input-output configurations, the DMU in question is able to identify whether it has used input excessively or its output has been under-produced. Finally, there is the possibility of making inferences about the DMU's general profile. We should be aware that the technique used here is a comparison between the production performance of each DMU and a set of efficient DMUs. The set of efficient DMUs is called the reference set. The owners of the DMUs may be interested to know which DMU frequently appears in this set. A DMU that appears more than others in this set is called the global leader. Clearly, this information gives huge benefits to bank owners, especially in positioning their entities in the market.

The first weakness of the DEA is that it assumes that data are free from measurement error. It is well known that such an error can cause a significant problem (Hababou, 2002; Avkiran 1999a and Mester, 1996). Secondly, since efficiency is measured in a relative way, its analysis is confined to the sample used. This means that an efficient DMU found in the analysis cannot be compared with other DMUs outside the sample. In addition, the DEA fails to handle absolute efficiency. Another weakness is that statistical testing is difficult to perform, because the DEA is a non-statistical approach. It is also apparent that the DEA is not sensitive to intangible and categorical variables such as service quality and service delivery.

4.3.2 The DEA model

Technical efficiency demonstrates how a firm can best transform various inputs into single or multiple products. However, Sealey and Lindley (1977) made a distinction between technical production and economic production in the case of financial firms. According to them, the former refers to a process of transforming input into output to the extent that input loses its identity. In contrast, economic production means the creation of a product that is more highly valued than the original input element. To begin with, we should start with production function. Production simply refers to the transformation of input into output. Like the consumers under the utility theory who purchase goods that yield satisfaction, the producers acquire input, for example labour and capital, with which they produce commodities. They are acting as rational economic agents seeking to maximise the profit they get from the production and sale of their commodities. However, the production theory differs from the utility theory with regard to its objective function. Its objective is clear and measurable. Maximum profit, for example, is expressed clearly in monetary value. The output produced, meanwhile, can be measured in terms of physical units or monetary units. In contrast, consumers' satisfaction or utility is very subjective and is hard to measure. The choice of measurement available is either cardinal or ordinal.

Let us consider a simple production process whereby a producer employs two inputs (x_1 and x_2) in order to produce a single output (y). The production function states that the quantity of output (y) produced depends on the quantity of input (x_1 and x_2). This simple function can be written as:

$$y = f(x_1 , x_2) \quad \text{Equation 4.1}$$

Gravelle and Rees (1998) state some assumptions with regard to the above function as follows:

i. input (x_1 and x_2) and output (y) are divisible.

ii. maximum output (y_{max}) is feasible for every input combination of x_1 and x_2

$$Y_{max} = f(x_1, x_2)$$

iii. $0 \leq y \leq y_{max}$ are feasible. This implies that wasteful production is possible. Output efficiency occurs when $y = y_{max}$. In contrast, $y > y_{max}$ is technically impossible, given the current state of technology. Other factors like prior contractual commitments or legal restriction may also play their parts.

iv. when $Y > 0$, then either $x_1 > 0$ or $x_2 > 0$.

Equation 4.1 is the explicit form of a production function from which a firm produces only one product. Now, by considering the various roles of commercial banks, it is clear that the banks produce more than a single product (refer to Chapter 3 where we discuss the role and function of the commercial bank in the economy.) Since the firm or bank produces various outputs, it is better to employ a multi-product production function. In this case, the production function is stated in its implicit form (Gravelle and Rees, 1998; Henderson and Quandt, 1980).

Let us say that a production process generates *s* outputs using *r* inputs. Then following Equation 4.1 and conditions for output efficiency ($y - f(x_1, x_2) = 0$), we have an implicit production function as follows:

$$g(y_1, y_2, \ldots, y_s, x_1, x_2, \ldots\ldots, x_r) = 0 \qquad \text{Equation 4.2}$$

Equation 4.2 is assumed to possess continuous first and second-order partial derivatives, which are different from zero for all its nontrivial solutions (Henderson and Quandt, 1980).

As absolute efficiency remains unknown due to the absence of known technology, we need to rely on the relative efficiency. This is where the DEA is put forward to facilitate efficiency analysis. As explained in the literature review

(Chapter Two), the DEA uses linear programming in order to construct the so-called best-practice frontier. The relative efficiency of each DMU in the sample refers to the position of that DMU in relation to the frontier. If the DMU lies on the frontier, then it is an efficient entity. If it lies below, then it is regarded as inefficient.

By assuming that there are *n* number of DMUs (banks), each producing *s* different outputs using *r* different inputs, the efficiency ratio for a DMU is measured as:

$$Eo = \frac{\sum_{i=1}^{s} u_i y_{io}}{\sum_{j=1}^{r} v_j x_{jo}} \qquad \text{Equation 4. 3}$$

where

E_o= relative efficiency of the DMU *o*

s = number of outputs produced by the DMU *o*

r = number of inputs employed by the DMU *o*

y_i = the i th output produced by the DMU *o*

x_j = the j th input employed by the DMU *o*

u_i = s x 1 vector of output weights and

v_j = r x 1 vector of input weights..

i runs from 1 to *s* and *j* runs from 1 to *r*.

The above model is designed to evaluate the relative performance of some DMU, designated as DMU *o*, based on the observed performance of m = 1, ..., n DMUs. The numerator in Equation 4.3 indicates a set of desired outputs and the denominator represents a number of resources utilised to produce these outputs. This ratio generates a scalar value which can be interpreted as an efficiency rating, since it satisfies $0 \leq E_o \leq 1$. The objective function E_o can only assume a maximum value that is equal to 1 or becomes less than 1 but remains positive. $E_o = 1$ means full efficiency and $E_o < 1$ shows the presence of inefficiency for each DMU being evaluated. Also, $1 - E_o$ proves the existence of inefficiency. This equation also assumes that there are constant returns to scale and all inputs are controllable. In addition, both inputs and outputs can be entered without any kind of standardisation. However, the determination of weights for input and outputs can bring in a major problem, since the DMUs might have different valuations concerning the two. Charnes et al. (1978) address this difficulty. They permit a DMU to adopt a set of weights that will maximize its relative efficiency ratio without the same ratio for other DMUs exceeding one. According to Siems and Barr (1998), the DEA model will place higher weights on those inputs that the DMU uses least and those outputs that the DMU produces most. In practice, the determination of the weights is settled by the DEA computer software.

Equation 4.3 is re-written in the form of fractional programming as:

$$\max Eo = \frac{\sum_{i=1}^{s} u_i y_{io}}{\sum_{j=1}^{r} v_j x_{jo}} \qquad \text{Equation 4. 4}$$

subject to

$$\frac{\sum_{i=1}^{s} u_i y_{im}}{\sum_{j=1}^{r} v_j x_{jm}} \leq 1$$ for each DMU in the sample, where m=1,....,n (number of DMUs).

To enable the measurement of the relative efficiency, Equation 4.4 is transformed into a linear programming problem, as carried out by Charnes et al. (1978). This transformation is performed by algebraically replacing all efficiency variables with an optimisation problem expressed purely in terms of weights. Next, an additional constraint is introduced whereby the denominator of the objective function is set to equal one.

$$\max E_o = \sum_{i=1}^{s} u_i y_{io} \qquad \text{Equation 4.5}$$

subject to

$$\sum_{j=1}^{r} v_j x_{jo} = 1$$

$$\sum_{i=1}^{s} u_i y_{im} - \sum_{j=1}^{r} v_j x_{jm} \leq 0,\ m = 1,......,n.$$

u_i, v_j $\geq$ 0.

u and v are small but positive in quantity. The first constraint ($\sum v_j x_{jo} = 1$) guarantees that it is possible to move from (4.5) to (4.4) as well as from (4.4) to (4.5) (Bowlin 2002).

The dual formulation of the objective function can be expressed as:

$$\text{Minimise } \alpha_o = \sum_{j=1}^{r} \alpha_o x_{jo} \qquad \text{Equation 4.6}$$

subject to;

$$\sum_{m=1}^{N} \varphi_m y_{im} \geq y_{io}, i = 1, \ldots, s.$$

$$\alpha_o x_{jo} - \sum_{m=1}^{N} \varphi_m x_{im} \geq 0,\ j = 1, \ldots, r; \varphi_m \geq 0; 0 < \alpha_o < 1.$$

The variable α_o, is a scalar and ϕ is nx1 vector of constants. According to Miller and Noulas (1996) and Yildirim (2002), α_o represents the overall technical efficiency (OTE) of the *o*-th DMU. Its value must be between zero and one, where 1 means that the DMU is fully and technically efficient. This is in line with Farell's (1957) definition. It should be noted further that the linear programming problem must be solved N times, i.e. once for each DMU in the sample. A value of α is then obtained for each DMU.

Bowlin (2002) states that the above dual formulation is the basis upon which the name DEA is derived. Any choice of ϕ_m produces an upper limit for the outputs and a lower limit for the inputs of DMUs. It follows that α is linked to ϕ_m so that optimising choices are associated with minimising $\alpha = \alpha^*$. By collecting and plotting such solutions, an upper bound is produced which envelops all of the observations, creating the name DEA.

However, the CCR model under the assumption of CRS is only appropriate when all DMUs are running at an optimal scale, and this requires the DMUs to operate at the flat portion of the long run average cost (LRAC) curve. In practice, some factors may prevent a DMU from operating at an optimal scale, such as financial and legal constraints, imperfect information etc. Coelli (1996) highlights that the use of the CRS specification when some of the DMUs not running at optimal scale will result in measures of technical efficiency which are mixed up with scale efficiency. Banker et al. (1984) suggest their model (known as the BCR model) to handle this problem. It improves the CCR model by introducing a variable that represents the returns to scale. The BCR model will allow the calculation of technical efficiency that is free from scale efficiency effects.

In the BCR model, the primal formulation is written as:

$$\text{Maximise}\, E_o = \sum_{i=1}^{s} u_i y_{io} - c_o \qquad \text{Equation 4.7}$$

subject to;

$$\sum_{j=1}^{r} v_j x_{jo} = 1$$

$$\sum_{i=1}^{s} u_i y_{im} - \sum_{j=1}^{r} v_j x_{jm} - co < 0,\, m=1,.........,N.$$

u_i, v_j > 0; (small but positive). The parameter c_o is unconstrained in sign. It indicates the various possibilities of returns to scale. $C_o > 0$ indicates increasing returns to scale and $c_o = 0$ implies constant returns to scale. Finally, $c_o < 0$ implies decreasing returns to scale. This model forms a convex hull of intersecting planes, which envelop the data points more tightly than the CRS model. Therefore, it enables technical efficiency scores to be greater than or equal to those obtained under the CRS model.

The dual formulation is expressed as:

Minimise α_o Equation 4. 8

subject to;

$$\sum_{m=1}^{N} \varphi_m y_{im} \geq y_{io},\ i = 1,.....,s.$$

$$\alpha_o x_{jo} - \sum_{m=1}^{N} \varphi_m x_{im} \geq 0,\ j = 1,.....,r;\ \varphi_m \geq 0;\ 0 < \alpha_l < 1.$$

$$\sum \varphi m = 1;\ m = 1,.......,N.$$

The main difference between the CCR model and the BCR model is that the ϕ_m are now restricted to summing to one. $\sum \phi_m = 1$ is taken as a convexity constraint. This implies that the constraint in the CCR model, whereby the DMUs must be scale efficient (under the CRS assumption), is now relaxed. By allowing this, the BCR model permits variable returns to scale and measures only technical efficiency for each DMU. As a result, under the CCR model, a DMU is said to be efficient if it is both scale and technically efficient. Unlike the BCR model, it is enough to be technically efficient in order to be an efficient DMU.

4.3.3 Productivity growth

Given the availability of panel data set, it is possible to compare the performance of the banks over time. However, the present technique of measuring the technical efficiency does not allow a direct comparison of banks' efficiency from one period to another. The efficiency scores obtained from a particular sample are confined to that particular sample and cannot be compared with another sample in different period. Therefore, this limitation does not allow the measurement of productivity growth over a period of time. To those responsible for making the decisions, information about improvement in efficiency is of great interest. For example, by holding the same level of inputs, can the firms increase the future level of outputs? Put another way, holding the same level of outputs, can the firm reduce the inputs in another time period? This perhaps can happen via improvement in efficiency. Besides that, technical improvement will also take place over the time, hence contributing to productivity growth.

The idea of comparing the input of a decision making unit over two periods of time (period t and period $t+1$) by which the input in period t could be decreased holding the same level of output in period $t+1$ is said to come from Malmquist (1953), quoted in Zhu (2003)[22]. Malmquist proposed the calculation of quantity indices as ratio of distance functions. His idea eventually led to the Malmquist input index. This was further developed into a Malmquist productivity index by

[22] A brief history of the Malmquist productivity index can be found in Grosskopf (2003). He also discusses some theoretical and empirical issues related to the index.

Caves et al. (1982). Finally, Färe et al. (1994) developed Malmquist productivity measures using DEA approach based on constant returns to scale[23]. Studies that use Malmquist productivity index are those such as Casu et al. (2004), Canhoto and Dermine (2003), Tortosa-Ausina et al. (2003), Färe and Grosskopf (2001), Drake (2001), Mukherjee et al. (2001), Wheelock and Wilson (1999), Worthington (1999) and Grifell-Tatjé and Lovell (1997).

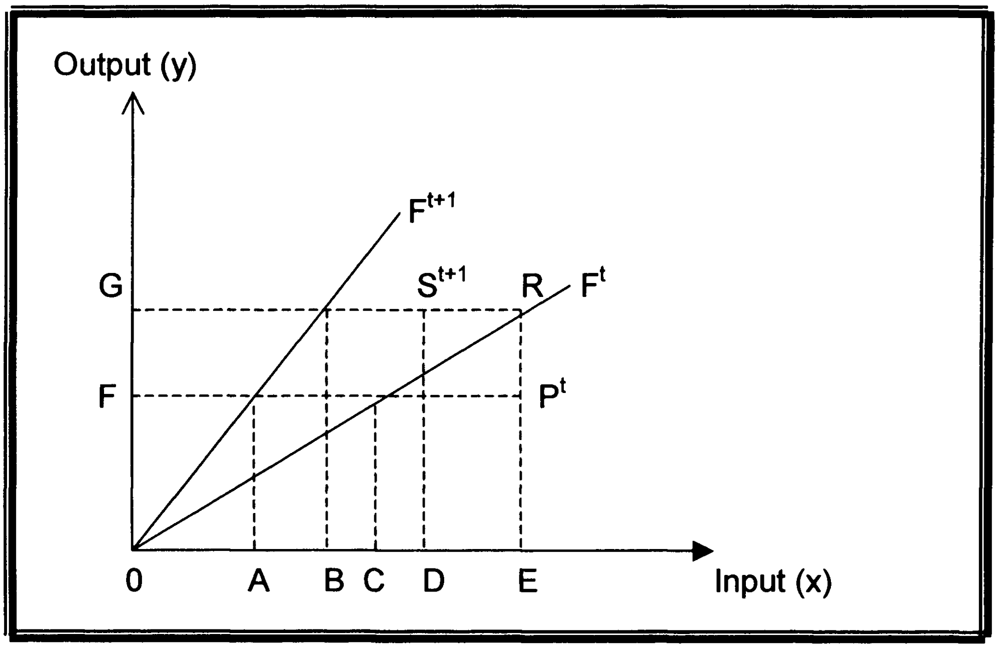
The Malmquist productivity index can be defined as an index number that enables a productivity comparison of the same firm over two different periods. In computing this index, two issues are being addressed. The first is about productivity change over a period of time. The second is the decomposition of the changes in productivity into catching-up effect (efficiency change) and frontier – shift effect (technical change). Improvement in efficiency means that the production is getting closer to the frontier and deterioration in efficiency implies the production is getting farther from the frontier. On the other hand, Färe et al. (1994) consider any improvement in technical progress as evidence of innovation.

The framework used in constructing a Malmquist productivity index is shown in Figure 4.1. Under constant returns to scale assumption, production frontiers are represented by F^t in period t and F^{t+1} in period $t+1$. Both represent efficient level

[23] Following the work of Färe et al. (1994), Ray and Desli (1997) proposed the decomposition of the same Malmquist index using a variable returns to scale frontier as the benchmark. This might lead to different conclusions concerning the sources of productivity growth.

of output (y) that can be produced from a given level of input (x). It is also assumed that the frontier can shift over time. Let us say that given a financial institution in period *t*, the production point is represented by P^t. Note that P^t lies below frontier F^t. An input-based efficiency measure can be deduced by the horizontal distance ratio OC/OE. The value of the distance ratio is less than one implying that the observed point is relatively inefficient. To make production technically efficient in period *t*, inputs must be reduced i.e. shown by the movement onto frontier F^t. Next, consider point S^{t+1} that lies above frontier F^t. The value of distance function evaluating S^{t+1} relative to technology in period *t* is given by OE/OD, which is greater than one. Based on technology in period *t*, point S^{t+1} is infeasible. However, it is attainable due to technological improvement or innovation over time. This is shown by the shift of the frontier from F^t to F^{t+1}. Thus, when inefficiency is present, the magnitude of the movement of any financial institution over time is influenced by its position relative to the corresponding frontier (technical efficiency) and the position of the frontier itself (technical change or progress). By ignoring inefficiency, productivity growth fails to distinguish between improvement in efficiency (catching-up effect to its own frontier) and improvement in technology (frontier shift effect).

Figure 4.1: The Malmquist input-based index and productivity changes over time



The Malmquist productivity index can be constructed based two broad orientations; partially oriented (input-oriented and output-oriented) and simultaneously output and input-oriented (Lovell, 2003). Two assumptions are normally made, either constant returns to scale or variable returns to scale. The DEA-based Malmquist productivity index developed by Färe et al. (1994) was constructed under the constant returns to scale.

Before constructing a Malmquist productivity index, four sets of linear program problems need to be solved. The main differences lie in term of observed production and reference technology. Recall that there are *n* firms (m = *1, 2,…,*

n) producing a vector of s outputs by using a vector of r inputs. The four sets of linear program problems are as follow:

$$D_o^t(x_o^t, y_o^t) = \min \theta_o \qquad \text{Equation 4.9}$$

subject to

$$\sum_{m=1}^{n} \lambda_m x_m^t \leq \theta_o x_o^t$$

$$\sum_{m=1}^{n} \lambda_m y_m^t \geq y_o^t$$

$$\lambda_m \geq 0, m = 1,2,\ldots,n.$$

Equation 4.9 is to compare x_o^t to the frontier in period t.

$$D_o^{t+1}(x_o^{t+1}, y_o^{t+1}) = \min \theta_o \qquad \text{Equation 4.10}$$

subject to

$$\sum_{m=1}^{n} \lambda_m x_m^{t+1} \leq \theta_o x_o^{t+1}$$

$$\sum_{m=1}^{n} \lambda_m y_m^{t+1} \geq y_o^{t+1}$$

$$\lambda_m \geq 0, m = 1,2,\ldots,n.$$

Equation 4.10 is to compare x_o^{t+1} to the frontier in period $t+1$.

$$D_o^{t+1}(x_o^t, y_o^t) = \min \theta_o \qquad \text{Equation 4.11}$$

subject to

$$\sum_{m=1}^{n} \lambda_m x_m^{t+1} \leq \theta_o x_o^t$$

$$\sum_{m=1}^{n} \lambda_m y_m^{t+1} \geq y_o^t$$

$$\lambda_m \geq 0, m=1,2,...,n.$$

Equation 4.11 is to compare x_o^t to the frontier in period *t+1*.

$$D_o^t(x_o^{t+1}, y_o^{t+1}) = \min \theta_o \qquad \text{Equation 4.12}$$

subject to

$$\sum_{m=1}^{n} \lambda_m x_m^t \leq \theta_o x_o^{t+1}$$

$$\sum_{m=1}^{n} \lambda_m y_m^t \geq y_o^{t+1}$$

$$\lambda_m \geq 0, m=1,2,...,n.$$

Equation 4.12 is to compare x_o^{t+1} to the frontier in period *t*.

In this study, the construction of the Malmquist productivity index is to follow the works of Caves et al. (1982) and Zhu (2003). These studies employ the technology in period *t+1* as the reference technology (see Equation 4.13 below). Alternatively, the technology in period *t* (base period) can also be used as reference technology. This approach is taken by Casu et al. (2004), Canhoto and Dermine (2003), Wheelock and Wilson (1999) and Färe et al. (1994). The difference in the reference technology used affects the magnitude in interpreting the index. When the reference technology is based on period *t+1*, then $M_o > 1$ implies deterioration in productivity over the period under study. On the other hand, when reference technology is based on period *t*, then $M_o > 1$ implies an improvement in productivity.

Our empirical Malmquist productivity index (M_o) is written as:

$$M_o = \left[\frac{D_o^t(x_o^t, y_o^t)}{D_o^t(x_o^{t+1}, y_o^{t+1})} \frac{D_o^{t+1}(x_o^t, y_o^t)}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \right]^{1/2} \qquad \text{Equation 4.13}$$

An equivalent way of writing the index for a specific DMU *o* is:

$$M_o = \frac{D_o^t(x_o^t, y_o^o)}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})} \left[\frac{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{D_o^t(x_o^{t+1}, y_o^{t+1})} * \frac{D_o^{t+1}(x_o^t, y_o^t)}{D_o^t(x_o^t, y_o^t)} \right]^{1/2} \qquad \text{Equation 4.14}$$

or

$M = E * T$

where

$$E = \frac{D_o^t(x_o^t, y_o^o)}{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}$$

$$T = \left[\frac{D_o^{t+1}(x_o^{t+1}, y_o^{t+1})}{D_o^t(x_o^{t+1}, y_o^{t+1})} * \frac{D_o^{t+1}(x_o^t, y_o^t)}{D_o^t(x_o^t, y_o^t)} \right]^{1/2}$$

where

M = the Malmquist productivity index

E = a change in efficiency over the period t and t+1 (the term outside the square bracket)

T = a measure of technical progress measured by shifts in the frontier over the period t and t=1 (the two ratios in the square bracket).

4.4 Analysis of results

The process of estimating the individual efficiency of the commercial bank is carried out by using Excel Solver Software developed by Zhu (2003).

4.4.1. Efficiency scores and their components

Overall technical efficiency (OTE) and pure technical efficiency (PTE) are calculated directly by the CCR (CRS) and BCR (VRS) models respectively. Scale efficiency (SE), on the other hand, is given by OTE/PTE.

Under the assumption of VRS, the average pure technical efficiency between 1994 and 2000 is 83.21%. This means that the commercial banks could have produced, on average, the same amount of outputs with approximately 16.79% fewer resources than they actually employed (see Table 4.2 below). This finding is similar to what has been found in the literature. For example, in another study using the Malaysian banking industry, Katib and Matthews (1999) found that the average score for the period between 1989 and 1995 was 86%. During the sample period, the average efficiency score had declined from 90% in 1989 to

82% in 1995. In another similar study, Laeven (1999) found that the average efficiency measure was 70%. A similar result was obtained by Canhoto and Dermine (2003). They found the average technical efficiency of Portuguese banks was 80% for the period between 1990 and 1995. However, in other studies, the average efficiency measures are higher. For example, Yildirim (2002) found that the average score for Turkish banks was 96.06%. Favero and Papi (1995), meanwhile, found that the average efficiency score for Italian banks was 91%. For Miller and Noulas (1996), the average efficiency score for the large banks in the United States was 96%. In another similar study concerning the large banks in the United States, Mukherjee et al. (2001) found the average efficiency score to be above 90% and they, in fact, included in their sample the same banks used by Miller and Noulas (1996). The main difference lies in the choice of inputs and outputs.

Although the average efficiency scores from this study are similar to what has been found by Katib and Matthews (1999), several differences should be highlighted. Firstly, the type and number of inputs and outputs used is different. In this study, we use three inputs and three outputs, while Katib and Matthews (1999) used only two inputs and three outputs. Secondly, our sample includes both local and foreign banks, but they limited their sample only to local banks. Our total year-bank observation is 193 compared to theirs of 140 bank-years. Thirdly, the sample period also differs, but shares the same number of years. Our sample period is from 1994 until 2000 whilst their sample period ran from 1989 until 1995.

It should be re-emphasised that the DEA technique is very sensitive towards the number of inputs and outputs used and sample size; different combinations of inputs and outputs will influence the estimates of technical efficiency. It can also affect the ranking of the banks in the sample. Secondly, Katib and Matthews (1999) excluded foreign banks from their study. By including the foreign banks in this study, we are able to compare the performance of domestic and foreign banks. We can also do the bank ranking, which is of great interest to the policy maker and the bankers.

Under the CRS assumption, as expected, the average efficiency score is around 56.57%. Graphically, under this assumption, the production frontier is a straight line. In contrast, the frontier under the VRS assumption is concave. Thus, the latter can accommodate more efficient DMUs than the former. For scale efficiency, the average score is 67.44%. This implies that the actual scale of production has diverged from the most productive scale size by about 32.56%.

If PTE is greater than SE, as shown by the table below, then inefficiency is caused by scale inefficiency. In other words, if there is a large difference between the efficiency score obtained under the CRS and VRS assumptions, then this is evidence of scale inefficiency. The results show that, on average, the inefficiency that the banks have experienced is due to scale inefficiency. This implies that the banks have difficulty in finding optimal combinations of various inputs to produce the desired output. The result is in accordance with what has been found by Sturm and Williams (2004), Yildirim (2002) and Katib and Matthews (1999). However, Miller and Noulas (1996) found that pure technical inefficiency was twice as great as scale inefficiency.

Table 4.2: Summary of efficiency score by types (1994 – 2000)

Types of efficiency	1994 (%)	1995 (%)	1996 (%)	1997 (%)	1998 (%)	1999 (%)	2000 (%)	1994-2000 (%)
Overall Technical Efficiency								
Mean	89.41	53.80	48.14	50.99	60.45	43.79	62.37	56.57
Standard Deviation	12.27	25.49	26.36	25.35	23.29	23.01	20.47	26.28
Minimum	67.56	24.03	22.13	20.55	36.40	23.65	35.50	20.55
Maximum	100	100	100	100	100	100	100	100
Pure Technical Efficiency								
Mean	93.56	82.58	81.48	83.95	78.88	78.22	88.95	83.21
Standard Deviation	10.18	22.69	22.48	17.83	20.62	20.29	13.71	19.68
Minimum	69.16	35.78	31.38	34.03	46.18	34.76	64.24	31.38
Maximum	100	100	100	100	100	100	100	100
Scale Efficiency								
Mean	95.67	65.17	59.50	60.15	76.91	55.24	69.72	67.44
Standard Deviation	8.59	21.65	25.03	23.48	19.33	19.30	17.22	23.40
Minimum	73.43	29.68	28.73	36.84	48.83	34.30	35.50	28.73
Maximum	100	100	100	100	100	100	100	100

Note: a. Inputs are the number of labour (LAB), fixed asset (FA) and total deposits (TD).
b. Outputs are totals loans (TERM), other earning asset (OEA), and other operating income (OOY).

The average efficiency score in 1994 is always the highest, regardless of the types of efficiency. Based on Table 4.2, the overall technical efficiency (OTE) is 89.41%, the average pure technical efficiency (PTE) is 93.56%, and the average scale efficiency is 95.67%. The standard deviation in this year is also relatively low compared to those in other years. We have tried to seek an explanation for this phenomenon. Factors such as the number of observations and the different types of banks were considered. In 1994, the total number of banks observed was just 20, compared to 32 banks in 1995 and 1996, 31 banks in 1997, 30 banks in 1998 and 28 in 1999. So, can the number of banks explain this? The answer is no, simply because in 2000, the number of banks is also 20. Therefore, in our opinion, the number of banks cannot be the factor. Perhaps, one plausible reason is the different number of local and foreign banks. In 1994, the number of foreign banks is 7 compared to 13 local banks. From 1995 onwards, the number of foreign banks has increased to 11. Although the number of local banks has increased too, the impact of the greater presence of foreign banks reduces the average efficiency score drastically after 1994. In fact, the average efficiency score of local banks (in the case of OTE and SE) in 1994 is greater than that of foreign banks.

Another interesting result is that in all three types of efficiency measure with the same set of inputs and outputs, the lowest score occurs in 1999. (Refer to Table 4.2). The average score for OTE, PTE, and SE are 43.79%, 78.22%, and 55.24 % respectively. Why was the performance in 1999 relatively poor? Based on Table 4.1, we find that the volume of bank outputs had fallen in 1999. Two outputs showed the declining trend in volume. They are total loans (LOANS) and other operating income (OOY). Although the declining trend occurred in 1998, it was not as bad as in 1999. One possible factor for the declining volume is caused the exclusion of two banks from the sample. These two banks merged in 1999 and one of them used to be the second largest bank in Malaysia. Therefore, the declining volume of bank output might explain the low score in 1999.

It is well known that the period between 1997 and 1999 witnessed economic decline in several Asian countries due to the financial crisis. Malaysia too could not escape from this crisis. One of the sectors that was most affected was the banking industry. During the crisis, the commercial banks faced liquidity problems, lower total deposit growth, and higher non-performing loans. (Refer to Chapter 3, which discusses the development of banking in Malaysia.) The results show that the average efficiency score (PTE) declined during the 1998-99 period. However, the decline is relatively small and cannot be accounted for by the crisis alone. Moreover, the crisis affected both sides of the banks, liabilities, and assets. The score rises slightly in the year 2000.

It is well established that efficiency scores are affected by various types of input and output, and various combinations of input and output. Besides using three inputs and three outputs, we tried various sets of input and output based on what other researchers had previously done. We found that regardless of the number and types of the input and output employed the most efficient banks in the ranking are almost the same. For example, the top ten of the best-practice banks are the same. The same trend happens at the bottom of the ranking; the rest, which are in the middle of the ranking, are also almost the same. Although the scores for each individual bank may vary, the banks' rankings in general are not affected. Hence, the results from the DEA analysis can be said to be robust.

4.4.2. Efficiency score grouped by bank ownership

Further analysis of efficiency is now focused on bank ownership. Two forms of ownership are identified, namely domestic banks and foreign banks. Out of 32 banks observed between 1994 and 2000, 21 are domestic banks and the rest are owned by foreigners. Table 4.3 shows that the foreign banks have the highest average score in all types of efficiency except in 1994 and 1998 for scale efficiency. The average efficiency score of the foreign banks for OTE, PTE, and SE are 71.74%, 97.14%, and 75.01%. Meanwhile, the average efficiency score of the domestic banks for OTE, PTE, and SE 50.93%, 78.94% and 64.86% respectively. Our finding is different from that of Sathye (2001). He, in contrast, finds that domestic banks are more efficient than foreign banks. The average pure technical efficiency was 90% for the domestic banks and 71% for the foreign banks. He uses a cross-sectional data set from Australian commercial banks in 1996. In another study, Yildirim (2002) finds the efficiency score as follows: state-owned banks (98.5%), foreign banks (96.63%) and private banks (96.08%). The study was done in Turkey, and involved commercial banks for the period between 1988 until 1999. Compared with his findings, it is obvious that the performance of domestic banks in Turkey is better than their performance in Malaysia.

The number of efficient banks varies from one year to another. Throughout the period of study, only eight banks remained fully efficient, i.e. having a perfect score of 100 in every year. In the group of the eight most efficient banks with respect to the pure technical efficiency measure, the foreign banks emerge as the highest contributors with 62.5%. For the inefficient groups (where the efficiency score is less than 100), local banks make up 75% compared with only 25% by the foreign banks. It also turns out that the most inefficient banks are always privately owned. In fact, there are at least ten domestic banks that never reach the best-practice frontier as compared to only one foreign bank.

Banker (1993, quoted in Bowlin, 2002) and Yildirim (2002) proposed that a student t-test could be used to determine whether the difference in efficiency scores for two groups of banks is statistically significant. The t-test procedure compares the means of the two groups of banks and assumes that each group comes from a population that is normally distributed, particularly with respect to skewness. This can be shown by implementing the one-sample Kolmogorov-Smirnov test (Siegel and Castellan, 1988). (The t-test result is shown in Table 4.4 below.)

Table 4.3: Efficiency measures and bank ownership

Efficiency scores	1994	1995	1996	1997	1998	1999	2000
Average OTE all banks	0.8941	0.5380	0.4814	0.5099	0.6045	0.4379	0.6237
Std. Deviation	0.1227	0.2549	0.2636	0.2535	0.2329	0.2301	0.2047
Minimum	0.6756	0.2403	0.2213	0.2055	0.3640	0.2365	0.3550
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average OTE domestic banks	0.8940	0.4234	0.3657	0.4304	0.5308	0.3381	0.5093
Std. Deviation	0.1202	0.1747	0.1590	0.2130	0.1822	0.1017	0.0997
Minimum	0.6756	0.2403	0.2213	0.2055	0.3640	0.2365	0.3993
Maximum	1.0000	0.8958	0.9650	1.0000	0.9869	0.6487	0.7209
Average OTE foreign banks	0.8944	0.7567	0.7025	0.6545	0.7319	0.5921	0.7174
Std. Deviation	0.1371	0.2452	0.2881	0.2663	0.2633	0.2886	0.2241
Minimum	0.7068	0.4051	0.4000	0.3596	0.3962	0.2963	0.3550
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average PTE all banks	0.9356	0.8258	0.8148	0.8395	0.7888	0.7822	0.8895
Std. Deviation	0.1018	0.2269	0.2248	0.1783	0.2062	0.2029	0.1371
Minimum	0.6916	0.3578	0.3138	0.3403	0.4618	0.3476	0.6424
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average PTE domestic banks	0.9322	0.7475	0.7486	0.7785	0.6979	0.6897	0.7894
Std. Deviation	0.1088	0.2444	0.2370	0.1819	0.1903	0.1962	0.1434
Minimum	0.6916	0.3578	0.3138	0.3403	0.4618	0.3476	0.6424
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average PTE foreign banks	0.9421	0.9753	0.9412	0.9505	0.9460	0.9251	0.9714
Std. Deviation	0.0954	0.0521	0.1325	0.1076	0.1235	0.1120	0.0534
Minimum	0.7487	0.8412	0.6152	0.6819	0.6394	0.6921	0.8430
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average SE all banks	0.9567	0.6517	0.5950	0.6015	0.7691	0.5524	0.6972
Std. Deviation	0.0859	0.2165	0.2503	0.2348	0.1933	0.1930	0.1722
Minimum	0.7343	0.2968	0.2873	0.3684	0.4883	0.3430	0.3550
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Average SE domestic banks	0.9610	0.5891	0.5199	0.5572	0.7701	0.5050	0.6486
Std. Deviation	0.0829	0.1846	0.2158	0.2231	0.1721	0.1191	0.0801
Minimum	0.7606	0.2968	0.2873	0.3809	0.5126	0.3430	0.5232
Maximum	1.0000	0.9372	0.9769	1.0000	0.9869	0.8355	0.7468
Average SE foreign banks	0.9487	0.7712	0.7385	0.6820	0.7675	0.6146	0.7501
Std. Deviation	0.0975	0.2308	0.2579	0.2445	0.2346	0.2414	0.1791
Minimum	0.7343	0.4051	0.4155	0.3684	0.4883	0.3638	0.4717
Maximum	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

The results show that the differences in all types of efficiency scores for the two groups of banks are statistically significant. The results imply that foreign banks are relatively more efficient than local banks.

Table 4.4: Efficiency scores and bank ownership
(1994 – 2000)

Efficiency	Domestic banks[a]	Foreign banks	t-statistics	Significance level
Mean OTE	0.5000	0.6737	-4.5309*	0.0000
Standard deviation	0.0554	0.0735		
No. of observations	120	73		
Mean PTE	0.7816	0.9151	-5.2087*	0.0000
Standard deviation	0.0422	0.0223		
No. of observations	120	73		
Mean SE	0.6444	0.7238	-2.2866**	0.0236
Standard deviation	0.0516	0.0567		
No. of observations	120	73		

Note: [a] Local banks consist of both private banks and state-owned banks.
* significant at 1% level.
** significant at 5% level.

The relative superiority of foreign banks over domestic banks should be viewed cautiously. The efficiency measure discussed here refers to the technical aspect of production. We are looking at the bank's performance in producing the maximum possible output from a given set of inputs. In our analysis, we use three inputs (labour, fixed asset and total deposits) and three outputs (total loans, other earning asset and other operating income). Thus, the relatively superior performance of the so-called best-practice banks relates to how much better they are in terms of producing greater quantities of output given the same level of input. It may be the case that economic efficiency, which seeks to identify possible combinations of input that minimise the cost of producing a given level of output, can offer an alternative explanation. However, this requires the availability of input prices such as the price of labour and capital[24].

4.4.3. Profiles of efficient and inefficient banks

We have so far identified the so-called efficient and inefficient banks. The difference is identified in terms of their efficiency scores. Efficient banks are those that have a score of 100, and inefficient banks have scores of less than 100. Bauer et al. (1998) propose that efficiency measures should meet a set of consistency conditions like rank-order correlation and efficiency correlation with standard non-frontier performance measures. In a similar study, Yildirim (2002) uses several non-frontier performance measures in his search for what constitutes a best-practice bank. He uses indicators such as profitability, asset quality (loan loss allowance and bad loans) and size.

The main idea of this section is to set up a profile of efficient and inefficient banks, not the determinants of efficiency. At this stage, we choose a bivariate analysis. A multivariate analysis is not required since we do not plan to identify the determinants of efficiency. To set up the profile, we use four variables. Two variables are the determinants of efficiency: market power and size. The other

[24] The following chapter focuses on estimating and measuring the cost efficiency using banks' cost function. Two approaches are used: stochastic frontier approach and non-parametric approach.

two are effects of efficiency: return on asset and asset quality. What should be done here is to apply test statistics in order to determine whether the difference in these criteria for the two groups of banks (efficient and inefficient banks) is statistically significant. In contrast, Katib and Matthews (1999) use regression techniques to identify the determinants of efficiency in the Malaysian banking sector. Amongst the factors that they have identified are bank size (represented by the number of bank branches), labour cost, and market structure (shown by the ratio of bank deposit to the total deposit).

With respect to the pure technical efficiency measure, we divide the banks into two groups: efficient and inefficient. Then we employ test statistics for group return on asset (ROA), market power (D/TD), bank size (A/TA), and asset quality (LLP/LN). In fact, there is another indicator for asset quality, i.e. non-performing loans (NPL). However, at this stage, data for NPL for each bank is not yet available. What is published by the BNM is an aggregate figure (sum of all individual NPLs).

Table 4.5: Profile of efficient and inefficient banks

	Efficient banks	Inefficient banks	t-statistics	Significance level
Return on asset (ROA)				
Mean	0.0198	0.0091	3.9083*	0.0001
Variance	0.0003	0.0004		
Market power (D/TD)				
Mean	0.0471	0.0283	2.6172**	0.0102
Variance	0.0037	0.0007		
Size (A/TA)				
Mean	0.0491	0.0268	2.9401**	0.0041
Variance	0.0042	0.0006		
Asset quality (LLP/LN)				
Mean	0.0128	0.0181	-1.4751	0.1420
Variance	0.0003	0.0011		
Number of observations	82	111		

Notes: a. ROA = percentage of pre-tax profit to total asset.
b. D/TD = percentage of bank deposit to total deposits in the banking sector.
c. A/TA = percentage of bank size to total banking assets.
d. LLP/LN = percentage of loan loss provision to total loans issued by the banks.
e. * = significant at 1% level. ** = significant at 5% level.

Table 4.5 shows that efficiency is related to the selected indicators of financial performance. For example, the efficient banks have a higher rate of return (ROA) than the inefficient banks. The null hypothesis of equality of pure technical efficiency among group of banks with different ROA is rejected ($p \leq 0.0001$ in this case). The same finding was obtained by Yildirim (2002). This suggests that the efficient banks enjoy higher rates of profit compared to the inefficient banks.

It also appears that efficient banks are characterised by their size ($p \leq 0.0102$). Since the mean size of the efficient banks (0.0491) is larger than that of the inefficient banks (0.0268), this suggests that efficient banks are relatively larger than inefficient banks. However, this finding should be treated cautiously. The standard deviation tells us that there is substantial deviation in the asset size of the efficient banks, i.e. the existence of both extremely large and small banks. Yildirim (2002) also reports the same finding.

The two banking groups also differ significantly with respect to their market power. The market power represented by the percentage of bank deposit to total deposit refers to banks' ability to influence the market price (in this case, market interest rate). Our finding shows that the efficient banks have stronger market power (the mean is 0.0471) compared to the inefficient banks (whose mean is 0.0283). Nevertheless, again, the standard deviation of the market power in the efficient banks is larger than that in the inefficient banks. Cautious interpretation is required here.

The last indicator is asset quality, shown by the percentage of loan loss provision to bank loans. We find that these two groups do not differ with respect to this indicator.

4.4.4 Malmquist productivity index and productivity growth

Using the same inputs and outputs, the Malmquist productivity index is constructed with the aid of Excel Solver software, developed by Zhu (2003). The results are shown in Tables 4.6 and 4.7 below. Table 4.6 shows the Malmquist productivity index and the sources of productivity growth: efficiency change and technical change or catching-up effect and shifting-up effect. The table reports the changes in productivity during two consecutive years (taking the second year to construct the benchmark technology or reference technology) as well as changes between 1994 and 2000. Recall that the Malmquist productivity index (M_o) measures the change in productivity between two periods. Since technology in the second period is used as reference technology, then if M_o is less than 1, there is productivity growth. If M_o greater than 1, productivity deteriorates and if equal to one, productivity remains unchanged.

On average, productivity has increased over the 1994-2000 period for the banks in our sample. The average change in the Malmquist productivity index is about 21.8% per year for the sample as a whole. The index is 0.782. The main source of productivity growth comes from technical change or innovation. The index for technical change for this period is 0.498. This implies that the frontier has shifted outward by 50.2%. On the other hand, technical efficiency has deteriorated by 57.1%. However, it must be noted that the Asian Financial Crisis occurred during the sample period. We think that the crisis had badly affected the performance of the banks as a result of high interest rates and high non-performing loans. To determine whether the crisis had any influence or not, we divide the sample into pre-crisis period (1994-1998) and post crisis period (1998-2000). During 1994-1998 where the crisis took place, technical efficiency deteriorated by 52% compared with only 5% after the crisis. This shows that the banks have difficulty improving their production level or bringing their operation closer to the frontier. In terms of technical change, it improves in both periods.

Turning to the period-by-period results, we have mixed trends. During the first two periods (1994/95 and 1995/96), productivity declines, 1.7% and 8% respectively. It appears that this is mainly caused by the decline in technical efficiency (16.4% during 1994/95 and 17.1% during 1995/96). However, the frontier-shift effect shows positive movement. Technical progress is about 12.6% during 1994/95 and 7.8% during 1995/96. During 1996/97, productivity increases by 9.3%. Both efficiency change and technical change show positive development. Technical efficiency increases by 3.9% and technical progress by 5.6%. The next two periods witness the declining trend in the productivity growth. Productivity declines during 1997/98 by 17.6% and this is mainly caused by the fall in technical innovation by 41.5%. It is well-known that during 1997-98 the financial crisis was taking place. This implies that during this period the frontier had shifted inward, given the severity of the problem that the banks were facing. The impact of financial crisis on the banking sector has been discussed in Chapter 3 (Section 3.5). Productivity continues to decline during 1998/99 by 1% before rising back during 1999/2000 periods.

Besides the turbulence of the Malaysian financial system during the sample period, it should be realised that the DEA technique includes measurement error in estimating the deviations from the frontier. If there were huge deviations from the frontier due to measurement error or high volatility in banking activities, this would have a significant impact on productivity measures. It is recommended that efficiency and productivity measures should be observed on a medium term basis rather than a short-term basis.

The mixed trend in the productivity growth has been found by other studies. Färe et al. (1994) studied the productivity growth in 17 OECD countries over the period 1979-1988. They found that over this period, the average change in the productivity index was less than 1% and the main contribution to this growth came from innovation or technical progress. They also found that five countries experienced a decline in the average change of productivity. The results also showed that during two periods (1979/1980 and 1981/1982), the frontier had shifted backward indicating a deciine in technical change.

In another study, Canhoto and Dermine (2003) look at the productivity growth of Portuguese banks during 1991-1995. They find that during this period, productivity increased on average by 59%. The main factor was an increase in technical progress on average by 74%. However, the catching-up effect shows a decline on average by 9%. In terms of a productivity component, they find that the frontier shift has always been on the positive trend whilst the catching up effect is always on the opposite trend. This implies that while the banks manage to enjoy technical progress, their production points are getting farther from the frontier. Further analysis is done by dividing the banks into three groups. They find that, on average, the Malmquist productivity index for the new banks is larger than that for the old banks.

Casu et al. (2004) study five European countries: France, Germany, Italy, Spain, and the United Kingdom, over the period 1994-2000. Their results seem to indicate that there is significant growth in productivity, particularly in Spanish and Italian banks (9.5% and 8.9% respectively). In the rest of the countries, productivity has been modest ranging from 1% to 2%. Although their study exhibits productivity growth, its components, technical change and efficiency change show a different trend. The result suggests that there has been a considerable frontier shift of the best practice or technological innovation but the technical efficiency shows little change.

Table 4.6: Malmquist productivity index and sources of productivity growth

Periods		Malmquist Productivity Index (M_o)[a, b]	Efficiency Change	Technical Change
1994/95	Mean	1.017	1.164	0.874
1995/96	Mean	1.080	1.171	0.922
1996/97	Mean	0.907	0.961	0.944
1997/98	Mean	1.176	0.831	1.415
1998/99	Mean	1.010	1.476	0.684
1999/2000	Mean	0.961	0.732	1.313
1994/2000	Mean	0.782	1.571	0.498

Note: a. The calculation of productivity index is done based on the assumption of constant returns to scale and under input orientation.
b. $M > 1$ means deterioration in productivity, $M=1$ means no change in productivity and $M < 1$ means improvement in productivity.

We now turn to the comparison of productivity growth over two groups of banks. This is to follow the approach taken by Grifell-Tatjé and Lovell (1997), and Canhoto and Dermine (2003). Grifell-Tatjé and Lovell (1997) divided the Spanish banks into savings and commercial banks while Canhoto and Dermine (2003) divide the Portuguese banks into old, new, and savings banks. The banks in our sample are divided into domestic banks and foreign banks. Table 4.7 shows the Malmquist productivity index and the sources of productivity growth for these two groups.

For the period 1994-2000, both domestic and foreign banks experienced an increase in productivity (20% and 26% respectively). Table 4.7 shows that the main source of the productivity growth is technical change or frontier shift effect (increase by 50% for both groups). However, the technical efficiency on average has declined (60% for domestic banks and 48% for foreign banks).

Table 4.7: Malmquist productivity index and sources of productivity growth by types of bank

Banks	1994-1995	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	1994-2000
Malmquist productivity index [a, b]							
Domestic banks	1.102	1.105	0.897	1.195	1.023	0.926	0.806
Foreign banks	0.869	1.028	0.900	1.130	0.975	0.978	0.740
Efficiency Change @ catching-up effect							
Domestic banks	1.216	1.187	0.890	0.798	1.584	0.651	1.600
Foreign banks	1.068	1.140	1.089	0.888	1.310	0.798	1.515
Technical change or frontier shift effect							
Domestic banks	0.906	0.931	1.008	1.498	0.646	1.422	0.503
Foreign banks	0.814	0.902	0.826	1.272	0.744	1.225	0.499

Note: a. The calculation of productivity index is done based on the assumption of constant returns to scale and under input orientation.
b. M > 1 means deterioration in productivity, M=1 means no change in productivity and M < 1 means improvement in productivity.

Period by period results show that the temporal pattern for both groups of bank has mixed behaviour of the Malmquist productivity index. With regard to the domestic banks, the index exceeds one in 1994/95 and 1995/96 (1.102 and 1.105), and is then less than one in 1996/97 (0.897). The index is greater than one in 1997/98 and 1998/99 (1.195 and 1.023). It finally becomes less than one in the final period (0.926). The productivity of domestic banks deteriorates in the first two periods due to a considerable decline in technical efficiency. Efficiency declines by 21.6% in 1994/95 and further by 18.7% in 1995/96. On the other hand, improvement in technical progress by 9.4% and 6.9% in these periods fails to trigger productivity growth. In 1996/97, the domestic banks' productivity grows by 10.3% due to an improvement in technical efficiency whilst technical change is almost unchanged. In the following period, 1997/98, productivity declines by 19.5% due to a sharp decrease in technical change (49.8%). The technical change of the foreign banks also plunges by 27.2%. It is well known that this period witnessed the Asian financial crisis. In contrast, technical efficiency improves substantially for both groups, 20.2% by the domestic banks and by 11.2% by the foreign banks. Interesting observations can be seen in 1998/99 and 1999/2000. In fact, this trend can be traced back to 1997/98. In 1998/99, the domestic banks experienced a modest decline in productivity by 2.3% while the foreign banks enjoy a modest growth of 2.5%. However, major changes take place in both efficiency change and technical change. Whilst efficiency falls sharply (41.6% and 31% for domestic and foreign banks), technical change improves substantially (35.5% and 25.6% for domestic and foreign banks). In 1999/2000, it is technical efficiency that improves significantly (34.9% and 20.2% for domestic and foreign banks) whilst technical change declines (42.2% and 22.5% for domestic and foreign banks). The results suggest that since the turbulence started in 1997, the catching up effects and frontier shifts effect move in opposite directions. While some banks try to improve their relative efficiency substantially, the remaining banks fail to keep up their trend in best practice. These opposite changes cancel out each other and, at the end, there are moderate changes in productivity.

Over the 1994-2000 periods, on average, productivity increased for both domestic and foreign banks in this sample. The average change in the Malmquist productivity index was about 20-25%. The main source of productivity growth came from technical change or innovation. This implies that the frontier had shifted outward by 50.2%. On the other hand, technical efficiency had deteriorated. The Malmquist productivity index was also constructed for the periods 1994 – 1998 (pre-crisis period) and 1998 – 2000 (post crisis period). During the first period (1994/98), productivity improved, with foreign banks showing a much faster pace of productivity growth (26%) compared with only 9% by the domestic banks. Again, the source of growth comes from technical improvement. Productivity declined during 1998/2000 for domestic banks (2.1%) caused mainly by the deterioration in efficiency.

4.6 Concluding remarks

In this chapter, we sought to achieve two main objectives. The first was the measurement and decomposition of bank efficiency. To arrive at this objective, we used a non-parametric approach, namely the Data Envelopment Analysis, for measuring the efficiency scores. Since our objective was to measure the relative efficiency (not absolute efficiency) of each commercial bank in Malaysia, there are two main models available in the DEA: the CCR and the BCR model. Both models are set up by assuming constant returns to scale and variable returns to scale respectively. For bank inputs, we used three variables – the number of bank employees, fixed assets and total deposits. Outputs of the bank are total loans, earning assets, and operating income. The sample period was from 1994 until 2000 with a total of 193 observations.

We found that the average overall technical efficiency was 56.57%, pure technical efficiency was 83.21%, and scale efficiency was 67.44%. These scores are within the range of what others have found. The results indicate that the main source of inefficiency in the Malaysian banking system is caused by scale inefficiency (failure to find the optimal combination of inputs to produce the desired level of output). Based on pure technical efficiency, the performance of the Malaysian banks is relatively stable, with the score always remaining above 80% except in 1998 and 1999. The same pattern applies to overall technical efficiency and scale efficiency, with the exception that their scores are relatively lower.

The performance of the foreign banks is relatively better than that of the domestic banks. The test statistics show that the difference in the efficiency score between these two groups is significant. However, it should be noted that the analysis is so far related to technical efficiency. It looks at how best a bank uses its resources to produce a given amount of output. Therefore, it should be made clear that the superior performance of the foreign banks over the local banks (both private and state-owned banks) does not necessarily mean that the former are also economically efficient. To be economically efficient, input prices such as the price of labour and capital need to be used in the analysis.

The analysis of the profile of the efficient and inefficient shows that there are two characteristics of the so-called efficient banks – higher rate of return (ROA), market power, and the size of the asset. The efficient banks seem to enjoy greater profits as compared to less efficient banks. On the third indicator, we believe that the connection between the efficient banks and the size of the banks needs to be treated cautiously. However, the last indicator of financial performance is less satisfactory. Asset quality is not statistically related to the efficient banks. Although the sign of this relationship is as expected, it is not significant.

The last part of this chapter dealt with the construction of the Malmquist productivity index and the identification of the sources of productivity growth. The study follows the approach taken by Cave et al. (1982) and Zhu (2003). The results show that over the period 1994/2000, on average the Malmquist productivity index changed about 20-25% and such growth had been contributed by an improvement in technical change rather than an improvement in technical efficiency. The temporal pattern in period-by-period observations is mixed.

Chapter

Five

CHAPTER FIVE

Analysis of cost efficiency: Parametric and non-parametric approach

5.0 Introduction

Firms' relative efficiency can be measured by adopting either non-parametric or parametric approaches. The non-parametric approach, popularly known as Data Envelopment Analysis, has been applied and discussed in the preceding chapter. Although the DEA is a non-statistical technique, the way we analyse its efficiency measure is statistical. In this chapter, we turn to estimating cost efficiency by using statistical approaches such as the stochastic frontier approach as well as non-parametric approach. Cost efficiency is related to one of the basic behavioural assumptions concerning firms. It is normally assumed that firms seek to minimise the cost of producing a given amount of output. This behavioural assumption, whilst it is given paramount consideration by economists, is largely ignored by the non-parametric approach. It is widely accepted that minimisation of the cost of production is a necessary condition in order to maximise profit. In addition, one of the advantages of the parametric or statistical approach is that the researcher can characterise the degree of uncertainty associated with the estimation of firm efficiency (Cornwell and Schmidt, 1996; Greene, 1993).

This chapter focuses on estimating cost functions and cost efficiency using the parametric approach in the context of panel data models. This study uses both the fixed effects model and stochastic frontier model. The estimates of the cost functions are obtained by using the least squares dummy variable (LSDV) for the fixed effects model and the maximum likelihood method (MLM) for the stochastic frontier model. Based on these estimates, we obtain the measures of cost efficiency, which are time-invariant, and then rank the banks based on their efficiency scores. We proceed further by measuring the efficiency scores via the Data Envelopment Analysis (DEA) technique with the same input prices and outputs. This enables us to compare the efficiency measures obtained by two different techniques. This comparison is very interesting since different techniques might lead to different scores and rankings.

The structure of this chapter is as follows. In the next section, we discuss the concept of cost efficiency and how to estimate it. In Section 5.2, we put forward the superiority of the panel data and its models. This is followed by model building. Section 5.4 is about panel data models. In this section, the study presents its case for using a fixed effects model rather than a random effects model. The Hausman's test that needs to be performed in deciding the appropriate model is presented. Section 5.5 is about estimation technique. Two techniques are discussed; the least squares dummy variable and the maximum likelihood method. In Section 5.6, we present the empirical models. Besides the stochastic models, we also present the DEA model. Again, the aim is to determine the extent that these models might have affected the cost efficiency measures. The likelihood ratio test is also explained. It is a test to decide which functional specification should be used. In this study, the choice is between log linear cost function and translog cost function. In Section 5.7, the descriptive statistics of bank input prices and bank outputs are presented. Our findings are presented and discussed in Section 5.8. The discussion is focused on the ranking of banks and the comparison of efficiency measures across the technique employed. Sections 5.9 and 5.10 look at the relationship between efficiency and bank size and other standard measures of performance. In the last section, we draw conclusions from the data presented in this chapter.

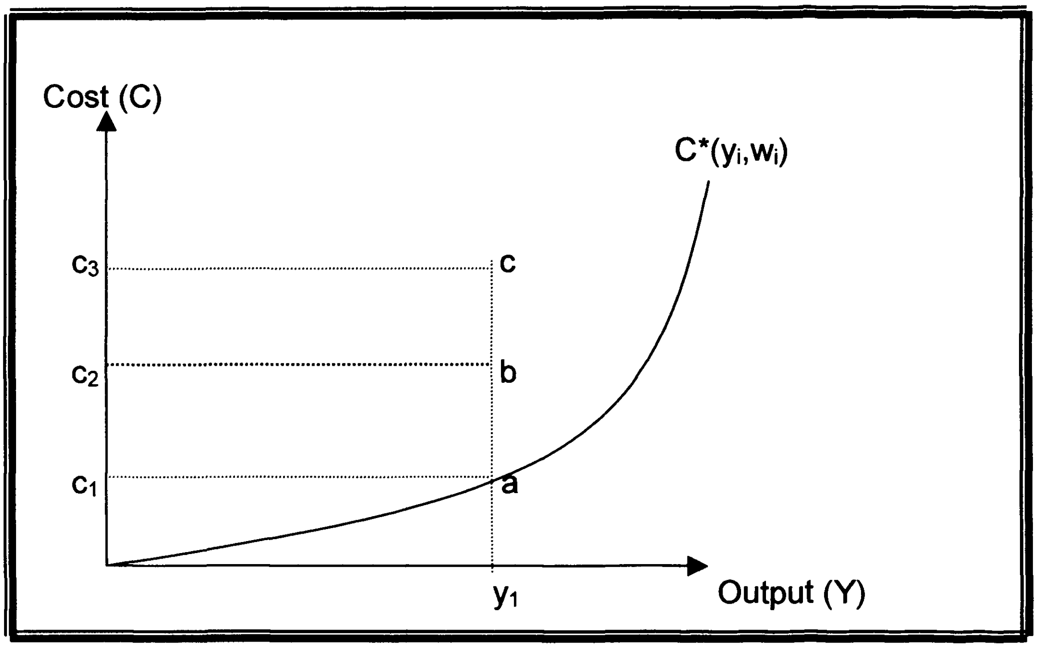
5.1 Concept and measurement of cost efficiency

In the theory of firms, economists are normally interested in studying the behavioural objectives of a firm, such as the pursuit of cost minimisation. Firms are assumed to employ any combination of input factors that minimise the cost of production, given a set amount of output. This is what the economists define as 'economic efficiency' or 'cost efficiency'. The term is normally equated with the effectiveness of resource allocation in the economy. According to Kumbhakar and Lovell (2000, pp.51), if cost minimisation is an appropriate behavioural objective, then the cost frontier and its associated system of cost minimising input demand equations can be used to measure cost efficiency. However, the actual cost frontier is hard to establish empirically and instead we work with the 'best-practice' cost frontier. We therefore consider a relative measure of cost efficiency, not an absolute one.

5.1.1 Concept of cost efficiency

The cost frontier refers to the minimum expenditure required to produce any amount of output (y), given input prices (w). Let $C^*(y_i,w_i;\beta)$ represent the cost frontier. The observed expenditure of a firm, let us say firm *i* (C_i), must be on or above the frontier. The measure of cost efficiency is given by the ratio of minimum cost to observed cost. In this case, firm *i* is cost efficient if $C_i = C^*(y_i,w_i;\beta)$ and is cost inefficient if $C_i > C^*(y_i,w_i;\beta)$. This relationship is shown in Figure 5.1 below. To be cost efficient, the observed cost of the firm must lie on the cost frontier. Any point on the left of the frontier indicates that the firm is not cost efficient. Let us say that the observed cost for firm *i* is at point c, given output level is at y_1. The cost efficiency of this firm is given by c_1/c_3.

Figure 5.1: A cost frontier



In frontier analysis, when an observed cost does not lie on the frontier, the difference is captured by the error term, given a cost function. This is equivalent to the vertical distance between the observed cost and the frontier.

As stated above, the cost frontier is expressed as $C^*(y_i,w_i;\beta)$. y_i and w_i are the vectors of output and input prices respectively. I is the number of firms. β is a vector of unknown parameters to be estimated. This cost frontier is common to all firms and thus represents the minimum cost of production. The observed cost of firm *i* is represented by C_i which is firm *i* 's expenditure. C_i is given by $\sum_k w_{ki}\, x_{ki}$. x is firm *i's* inputs and k indexes the inputs, k = 1, 2, ..., K.

The cost efficiency (CE) of firm *i* is given by

$$CE_i = \frac{C^*(y_i, w_i; \beta)}{C_i}$$ Equation 5.1

This equation shows that the cost efficiency of firm *i* is the ratio of minimum feasible cost to the observed expenditure. Since C_i is either larger than or equal to $C^*(y_i, p_i; \beta)$, it follows that;

i. $CE_i = 1$. This means that the observed cost reaches its lowest feasible value as represented by the cost frontier. Firm *i* is said to be cost efficient.
ii. $CE_i < 1$.This means that the observed cost exceeds the lowest feasible value.

The excess is then attributed to cost inefficiency. However, there are random shocks that are beyond the control of the producers, such as pure luck and measurement error. Thus, the excess is attributed to cost inefficiency and random shocks (See Aigner et al., 1977; Meeusen and van den Broeck, 1977; Greene, 1993; Kumbhakar and Lovell, 2000).

With the presence of random shocks, the stochastic cost frontier can be rewritten as $C^*(y_i, w_i; \beta).\exp\{v_i\}$ and has two parts. The first component is a deterministic part that is common to all producers. The second is a firm's specific random part, and is meant to capture the presence of random shocks on each firm. Given a stochastic cost frontier, the cost efficiency measure of firm *i* now becomes:

$$CE_i = \frac{C^*(y_i, w_i; \beta)\exp\{v_i\}}{C_i}$$ Equation 5.2

As argued by Parkin et al. (2000) and Kumbhakar and Lovell (2000), technical efficiency is necessary but not sufficient for the achievement of cost efficiency. This implies that a firm may be technically efficient (able to minimise the use of inputs in producing a given amount of output), but not necessarily cost efficient (able to minimise the cost of producing a given amount of output). There is no guarantee that the two can co-exist.

Thus, cost efficiency can be decomposed into two elements: technical efficiency and allocative efficiency.[25] While technical efficiency has been measured and discussed in the previous chapter using a non-parametric approach, allocative efficiency refers to firms' ability to use the inputs in optimal proportion, given their respective prices.

5.1.2 Measurement of cost efficiency

The main challenge under the stochastic frontier approach is the disaggregation of the error components into an inefficiency term and random error[26]. The reason is simple. The inefficiency term needs to be isolated from the other factors that are beyond the control of the firms, such as measurement error and luck. To do so, several strong assumptions need to be made about each error component

[25] The decomposition of cost efficiency is beyond the scope of this study and, therefore, further discussion is not presented here. However, Kumbhakar and Lovell (2000) provide an interesting explanation about the composition of cost efficiency.
[26] In contrast, DEA considers total deviations from the cost frontier are caused by firms' inefficiency.

and these assumptions define or differentiate each respective method[27]. The three methods are the stochastic frontier approach, distribution free approach, and thick frontier approach. All of them were discussed in Chapter II under the methods of estimating efficiency. In this study, the stochastic cost frontier approach with panel data model (fixed effects model) is used.

5.2 Panel data

Panel data or longitudinal data refers to repeated observations of the units over a period of time. The potential advantages of panel data over cross-sectional and time series data have been discussed, for example by Hsiao (2003), Baltagi (2001), Verbeek (2000), Greene (2000) and Kumbhakar and Lovell (2000). This discussion can be summarised below.

5.2.1 Strengths of panel data

Panel data gives both a cross-sectional and a time-series dimension. This probably provides a larger number of data points, which can be more informative, reduce co linearity amongst the explanatory variables and increase the degrees of freedom. The dual dimensions can also produce more accurate and efficient

---

[27] Strong assumptions like functional form and distributional form of the error components are not required under non-parametric approach like DEA. To certain extent, the absence of these assumptions makes researchers favour DEA.

estimates, as compared to other sources of data (Verbeek, 2000). Secondly, panel data can handle individual heterogeneity. This is to suggest that the decision-making units are heterogeneous regardless of whether they are individuals, firms or countries. On the other hand, failure to account for this heterogeneity may result in biased results. This is likely to take place in time series or cross section studies. Let say that the production cost of a firm is modeled as a function of output produced and prices of inputs used. These variables (output and input prices) vary with state and time. However, there are other factors that can make each firm heterogeneous to each other like management style, product and service quality and efficiency level. As stated by Hsiao (2003) and Baltagi (2001), some of these variables are qualitative in nature and are, therefore, difficult to measure. In addition, they are also hard to obtain. Variable like efficiency is not observed at all. If these variables are omitted in the cost function, the cost estimates may be biased. Panel data are able to control for these variables. Its analysis will show that the unobserved and omitted variables are captured by a firm's specific effects. In turn, this specific effect can be taken as either fixed or random.

Let say that the cost function is written as $C_{it} = \alpha_i + x_{it}\,\beta + \varepsilon_{it}$. β measures the partial effects of x_{it} in period *t* for firm *i*. x_{it} is a *K*-dimensional vector of explanatory variables (output and input prices), excluding a constant. This implies that the effects of a change in x are the same for all firms and all periods. However, the average cost level for firm *i* may be different from that for other firms. Thus, the α_i captures the effects of those variables that are peculiar to the i-th firm. If we take the α_i as fixed unknown parameters, this cost function is referred to as a fixed effects model. On the other hand, if the α_i are treated as drawings from a distribution with mean μ and variance σ^2_α, this leads to a random effects model (Verbeek, 2000). The essential assumption under the random effects model is that the drawings of α_i are independent of the explanatory variables (output and input prices). The error term (ε_{it}) in the cost model consists of two components: a time invariant component and a random error that is not correlated over time (v_{it}). The cost function can be rewritten as $C_{it} = \theta + x_{it}\,\beta + \alpha_i + v_{it}$. θ is the intercept term. α_i now represents firm's specific effects.

According to Kumbhakar and Lovell (2000), panel data can avoid the strong distributional assumptions normally used with cross-sectional studies or lead to estimates of cost efficiency with more desirable statistical properties. There are at least three difficulties identified with cross-sectional studies. They include the need to make strong distributional assumptions about the error components (inefficiency term and random error), independence assumptions relating to the relationship between the error components and other variables in the function and the inconsistency of technical efficiency estimates produced by the early estimates such as JLMS technique as the size of cross-section increases.[28] It is argued that the repeated observations on a sample of firms can overcome these

[28] The JLMS technique refers to the prediction of firm's specific inefficiency term (u_i) based on the work of Jondrow et al. (1982). The inconsistency of this technique is shown by Kumbhakar and Lovell (2000).

limitations or avoid strong distributional assumptions. Baltagi (2001, p 7) further argues that panel data models enable us to construct and test more complicated behavioural models than purely cross section or time series data. Technical and cost efficiency is better studied and modelled with panels.

In favour of panel data, Baltagi (2001) argues that this form of data set offers better opportunities to study the dynamics of adjustment by revealing the multitude of changes that occur and the speed of those adjustments. In the case of economic problems such as poverty and unemployment, panel data are able to show how the rates change over time or perhaps remain the same; only panel data can estimate what proportion of those who are poor or unemployed in one period remain poor (unemployed) in another period. Given relevant government policies concerning these problems, they can also show the reactions or responses of the poor and the unemployed.

In addition, Greene (2000, pp. 559-560), arguing in favour of the use of panel data, states that:

> "the fundamental advantage of a panel data set over a cross-section is that it will allow the researcher far greater flexibility in modelling differences in behaviour across individuals".

5.3 Model building

The main role of firms is to transform inputs into outputs. By assuming that firms always seek to produce the maximum feasible amount of output (y), given an amount of input (x), a production function defining this relationship can be written as:

$$y = f(x) \qquad \text{Equation 5.3}$$

The quantity $y = f(x)$ represents the maximum output level that can be produced, given a level of input x. A firm is said to be technically efficient if it can produce that maximum level of output. However, there is no guarantee that a technically efficient firm will be able to maximise profit. Given that profit is the excess of revenue over the cost, and that the firm operates to maximise profit, it must, after deciding an output level, select inputs that minimise the cost. Therefore, the assumption of cost minimisation is consistent with that of profit maximisation.

Production cost depends directly on the cost of input factors. If input prices are fixed, let say w_1, w_2, ……, w_k for the *k* input factors, the cost for input level x_1, x_2, ……, x_k is $w_1x_1 + w_1x_1 + ... + w_kx_k$. Therefore, the cost function, defined for all input prices and all output level, is the minimum-value function.

$$c(y,w) = \min w.x \qquad \text{s.t.} \quad f(x) \geq y. \qquad \text{Equation 5.4}$$

By assuming that the firms make perfectly rational and efficient decisions, the cost function is expressed as

$$c = (y,w) \qquad \text{Equation 5.5}$$

However, under the stochastic cost approach, the presence of disturbances may prevent the firms from being cost efficient, i.e. minimising their production costs, or positioning the actual cost on the cost frontier. Factors such as errors, time lag between the choice of the plan and its implementation, and human factors can push the firms' actual cost away from the frontier. Taking these disturbances into account, Aigner et al. (1977) and Meeusen and van den Broeck (1977) wrote a stochastic cost function for firm *i* as

$$C_i = f(y_i,w_i; \beta) + \varepsilon_i \qquad \text{Equation 5.6}$$

where ε_i is the error term.

The structure of the error terms can be separated into two elements: u and v. This means that $\varepsilon_i = u_i + v_i$. u represents internal disturbances or any factors under the control of the firm, such as technical or cost inefficiency. It is always negative. On the other hand, v is regarded as random disturbance and it can be either positive or negative. The stochastic cost function can be rewritten as:

$$C_i = f(y_i,w_i; \beta) + v_i + u_i \qquad \text{Equation 5.7}$$

The main challenge now is how to separate u and v. As discussed in Chapter Two, the various distributional assumptions of u and v generate various estimation methods like SFA, DFA and TFA. Given a panel dataset, there are panel data models.

5.4 Panel data models

Let us say that we have repeated observations on I firms, indexed by i = 1, 2, ..., I, through T time periods, indexed by t = 1, 2, ..., T. We simply assume that cost efficiency is time invariant i.e. remains constant over a short period of time or sample period of time. Following the works of Shanmugam and Das (2004), Christopoulus and Tsionas (2001) and Battese and Coelli (1995), a Cobb-Douglas cost function in natural logarithm is written as:

$$\ln c_{it} = \sum_{k} \delta_k \ln w_{kit} + \sum_{m} \gamma_m \ln y_{mit} + u_i + v_{it} \qquad \text{Equation 5.8}$$

where

c_{it} = the operating cost of firm *i* in period *t* in log value

y_{mit} = the *m*-th output of firm *i* in period *t* in log value

w_{kit} = the *k*-th input price of firm *i* in period *t* in log value

u_i = inefficiency term for firm *i*

v_{it} = random disturbance for firm *i* in period *t*

k = 1, ..., K for the number of input prices paid

m = 1, ..., M for the number of outputs produced.

In Equation 5.8, the total number of regressors depends on *K* and *M*, excluding the constant term (β_i). All w_{kit} and y_{mit} are independent of the error components. v_{it} is assumed to be independently, identically distributed with zero mean and variance. The firm effect is u_i, which is taken to be fixed over a period of time and specific to the individual unit *i*. According to Greene (2000, 2001, 2002), there are two basic methods by which to generalise this regression model. The first is the fixed effect approach, where u_i is taken to be a group-specific, constant term. It is argued that this constant term represents the firm's inefficiency. The second is the random effect approach. Under this approach, u_i is taken as random drawing from a distribution and treated as a component of a composite error term with ε_{it}.

5.4.1 Fixed effects model

The fixed effects model can be defined as a linear regression model in which intercept terms vary across the individual units but remain constant over the sample period. The model assumes that the differences across units can be captured in differences in these constant terms. By including a dummy variable for each unit *i* we get:

$$\ln c_{it} = \sum_j \varphi_j d_{ij} + \sum_k \delta_k \ln w_{kit} + \sum_m \gamma_m \ln y_{mit} + u_i + v_{it} \qquad \text{Equation 5.9}$$

where $d_{ij} = 1$ if $i = j$ and 0 elsewhere.

The parameters can be estimated by ordinary least squares (OLS). In fact, Equation 5.9 is the Least squares Dummy Variable estimator (Greene, 2000 and Verbeek, 2000). It can also be said that this estimator is equivalent to the 'within' estimator. The 'within' estimator looks at the variation within the groups or units over time. For example, as w_{1t} increases from one year to another, C_i will increase by certain units determined by the size of the coefficient (β_1). However, this may be cumbersome if the sample contains a large number of firms (Cornwell and Schmidt, 1996). The 'within' estimator can be written as:

$$(C_{it}-\overline{C_i})=(W_{1t}-\overline{W_1})\beta_1+(W_{2t}-\overline{W_2})\beta_2+(W_{3t}-\overline{W_3})\beta_3+(Y_{1t}-\overline{Y_1})\beta_4$$

$$+(Y_{2t}-\overline{Y_2})\beta_5+(Y_{3t}-\overline{Y_3})\beta_6+(v_{it}-\overline{v_i}) \qquad \text{Equation 5.10}$$

where $\overline{C_i} = \Sigma C_{it} / T_i$., $\overline{W_1} = \Sigma W_{1t} / T_i$, $\overline{W_2} = \Sigma W_{2t} / T_i$, $\overline{W_3} = \Sigma W_{3t} / T_i$, $\overline{Y_1} = \Sigma Y_{1t} / T_i$, $\overline{Y_2} = \Sigma Y_{2t} / T_i$, $\overline{Y_3} = \Sigma Y_{3t} / T_i$, $\overline{v_i} = \Sigma v_{it} / T_i$.

In contrast to the 'within' estimator, the 'between' estimator looks at the variation across the units observed. However, this estimator requires additional assumptions, such as that v_i and $\overline{Wi}$ are uncorrelated. It is argued that the 'between' estimator is less efficient in the sense that it relies heavily on the sample means and ignores the over time information[29]. This estimator can be written as:

$$\ln\overline{C_i} = \beta_0 + \overline{W_1}\,\beta_1 + \overline{W_2}\,\beta_2 + \overline{W_3}\,\beta_3 + \overline{Y_1}\,\beta_4 + \overline{Y_2}\,\beta_5 + \overline{Y_3}\,\beta_6 + D_i\,\beta_8$$

[29] For academic purposes, we obtained the 'between' estimator estimates and compared them with the estimates of the 'within' estimator. We found that some of the coefficients are similar in both cases and the rest vary. It is noteworthy to mention that we are looking at different magnitudes of variations, cross-sectional variation against time-series variation.

$+ u_i + \bar{v}_i$ Equation 5.11

5.4.2 Random effects model

The fixed effects model may involve many parameters. This is likely to happen when the number of cross-section units is large and each unit is given a dummy. This in turn leads to the loss of degrees of freedom. According to Hsiao (2003) and Baltagi (2001), the random effects model is more appropriate when the sample of the units is randomly drawn from a large population. Maddala (2001, p.576) raises another argument that u_i can be taken as a total of several factors specific to the cross-section units. Thus, u_i represent "specific ignorance" and can be treated as random variables. This is the same argument put forward for v_{it} that represent "general ignorance" and are treated as random variables. The u_i are assumed to be independent of the error v_{it} and also mutually independent.

5.4.3 Fixed effects model versus random effects model

The choice between the fixed effects model and the random effects model has been critically examined by Hsiao (2003) and Baltagi (2001). According to Hsiao (2003), the inferences that the researchers are looking for can be used in deciding which model should be used. If the inferences are to be based on the effects that are in the sample, then the fixed effects model is the answer. However, if the researchers want to make unconditional inferences with regard to the population of all effects, then one should choose the random effects model. This argument can be taken to establish a general guideline in assisting the researchers to make a decision. The appropriate model depends on the context of the data, how the data were collected, and on the environment relating to them. In this study, if the banks were randomly selected from a common population, then the effects can be assumed random. However, if each bank was not randomly sampled from a population of all banks and the researchers wanted to assess the differences between those individual banks, then the fixed effects model would be more appropriate. In this study, the banks in the sample are all that are available in Malaysia and, therefore, constitute the population rather than a random sample. The fact that several banks were excluded from the sample was, in fact, caused by the lack of appropriate data and bank consolidation.

The selection between the fixed effects model and the random effects model can be determined by using the specification error test (Hsiao, 2003; Baltagi, 2001 and Maddala, 2001). Hausman's (1978) test or *m*-statistic can be used to test hypotheses in terms of bias or inconsistency of an estimator. According to Baltagi (2001), a critical assumption that is always made in the regression model is that the error components (u_i and v_i) are independent of the regressors (x_{it}). This implies that $E(u_i, v_i / x_{it}) = 0$. The unobserved disturbances contain individual specific effects (u_i) which may be correlated with the x_{it}. For example, in a cost function, the u_i may represent unobserved management superiority of individual firm and this may be correlated with the output variables included in the function. If this happens then $E(u_i,v_i /x_{it}) \neq 0$. If we consider the regression model as $y_{it} = \beta x_{it} + u_i + v_{it}$, then if u_i is correlated with x_{it}, it is called fixed effects model. When u_i is independent of x_{it}, it is called random effects model. (See argument put forward by Hsiao, 2003 pp.49.)

The way to decide whether to use a fixed effects model or a random effects model is to test for misspecification. Let H_o denote the null hypothesis that there is no misspecification and let H_a denote the alternative hypothesis that there is misspecification. As shown by Maddala (2001, pp. 494 - 496), the null and alternative hypotheses are;

H_o : x and u are independent

H_a : x and u are not independent

To perform the Hausman's test, two estimators are constructed, β_1 and β_2. These estimators must have the following properties;

- β_1 is consistent and efficient under H_o but is not consistent under H_a.
- β_2 is consistent under both H_o and H_a but is not efficient under *Ho*.

If the alternative hypothesis H_a holds, the fixed effects model should be used. If the null hypothesis H_o holds, the random effects model should be used.

With regard to the assumption that firms' specific inefficiency effects are fixed over a short period, DeYoung (1997a) proposes that a panel set of between six and eight years is adequate to make reasonably sure that this assumption holds and contains only a small number of random errors. A panel set that exceeds eight years, on the other hand, may violate this central assumption.

5.5 Estimation technique

According to Kumbhakar and Lovell (2000), there are at least three ways to estimate the cost function depending upon the data set available. These techniques start with ordinary least squares, generalised least squares, and maximum likelihood methods. For example, Equation 5.9 can be estimated using ordinary least squares. After the estimation, the cost frontier intercept (θ) is represented by the minimum value of the estimated u_i:

$$\hat{\theta}_i = \min\{\hat{u}_i\} \qquad \text{Equation 5.12}$$

E_i are estimated from $\hat{E}_i = \hat{u}_i - \hat{\theta}$. Finally, the firm's specific estimate of cost efficiency (CE) is given by:

$$CE_i = \exp\{-\hat{E}_i\} \qquad \text{Equation 5.13}$$

Under the least squares approach, at least one firm has $CE_i = 1$ and the rest have $CE_i \leq 1$. The estimates of cost efficiency are consistent as the number of firms and the time period increase.

Another estimation technique is the maximum likelihood method. Let us assume that the u_i is not a generalisation to the truncated normal assumption. Recall that we have repeated observations on I firms, indexed by i = 1, 2, ..., I, through T time periods, indexed by t = 1,2,..., T. Based on the work of Kumbhakar and Lovell (2000, pp 105), the log likelihood function is given by

$$\ln L = \text{constant} - \frac{I(T-1)}{2}\ln\sigma_v^2 - \frac{I}{2}\ln(\sigma_v^2 + T\sigma_u^2)$$

$$- I\ln\left[1-\Phi\left(-\left(\frac{\mu}{\sigma_u}\right)\right)\right] + \sum_i \ln\left[1-\Phi\left(-\frac{\tilde{\mu}_i}{\sigma_*}\right)\right]$$

$$-\frac{\sum_i \varepsilon'_i\varepsilon_i}{2\sigma_v^2} - \frac{I}{2}\left(\frac{\mu}{\sigma_u}\right)^2 + \frac{1}{2}\sum_i\left(\frac{\tilde{\mu}_i}{\sigma_*}\right)^2,$$

Equation 5.14

where

$$\tilde{\mu}_i = \frac{\mu\sigma_v^2 - T\bar{\varepsilon}\sigma_u^2}{\sigma_v^2 + T\sigma_u^2}$$

$$\sigma_*^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma_v^2 + T\sigma_u^2}$$

$$\bar{\varepsilon} = \frac{1}{T}\sum_t \varepsilon_{it}$$

This function can be maximised with respect to the parameters in order to produce the required estimates, (Kumbhakar and Lovell, 2000 and Griffiths et al., 1993). It is further argued that the estimates obtained by the MLM will be similar to those obtained under OLS as $I \rightarrow \infty$ and $T \rightarrow \infty$.

The conditional distribution (u | ε) is given by

$$f(u \mid \varepsilon) = \frac{f(u,\varepsilon)}{f(\varepsilon)}$$

$$= \frac{1}{(2\pi)^{1/2}\sigma_{*}\left[1-\Phi(-\tilde{\mu}/\sigma_{*})\right]}.\exp\left\{-\frac{(u-\tilde{\mu})^2}{2\sigma_{*}^2}\right\},$$

Equation 5.15

which is distributed as $N^{+}(\tilde{\mu}, \sigma_{*}^2)$.

The derivation of u_i under panel data was originally shown by Kumbhakar (1987) and Battese and Coelli (1988)[30]. Estimates for u_i are obtained from the mean and the mode of the conditional distribution $f(u \mid \varepsilon)$.

$$E(u_i \mid \varepsilon_{it}) = \tilde{\mu}_i + \sigma_{*}\left[\frac{\varphi(-\tilde{\mu}_i/\sigma_{*})}{1-\Phi(\tilde{\mu}_i/\sigma_{*})}\right]$$ Equation 5.16

---

[30] See also Kumbhakar and Lovell (2000), Giannakas et al. (2003) and McKillop et al. (2004). For cross-sectional data, see Jondrow et al. (1982).

and

$$m(u_i \mid \varepsilon_{it}) = \tilde{\mu}_i \text{ if } \tilde{\mu}_i \geq 0 \text{ or } = 0 \text{ if otherwise.}$$

Finally, the estimates of cost efficiency for firm *i* is given by

$$CE_i = \exp\left(-\hat{u}_i\right)$$

5.6. Empirical models

Besides the input prices and bank outputs, other factors need be considered too. For example, Kwan (2003) argues that time-specific dummies should be included in the cost function. The basic idea is to test whether the production costs have undergone any systematic changes across all banks during the period from 1994 until 2000. It is well known that the banking industry in Malaysia faced severe problems during this period, due to the Asian financial crisis (1997-1998). Besides this crisis, there were also technological advances, which might have helped the banks reduce their operating costs. This study employs two dummy variables: for time and type of bank. For the latter, a value of 1 is given for domestic banks and 0 for foreign banks. However, under the fixed effects model, the bank dummy is dropped since no variation in the group takes place.

5.6.1 Functional form

As highlighted in the literature review, one of the major factors behind the variation of the efficiency estimates is that the relevant studies employ different functional forms. Indeed, an appropriate functional form is a major problem under stochastic approach. In the first place, one must decide the functional form of the cost frontier before the estimation can take place. The frontier functional forms can start with a simple function like a log linear cost function, followed by translog cost function and more complex functions like Fourier flexible functional form. The last two forms are the most commonly used. It is argued that the translog function has been a popular choice because of its ability to fit the data better than a simple function like linear function. It can also accommodate the U-shaped of the cost curve. However, the use of the translog cost function faces a multicollinearity problem since the interaction variables are closely related to each other. This might have contributed to poor basic results. But, in this study, we are more concerned with the inefficiency term, a component of the error term.

In relation to the choice of appropriate functional form, Giannakas et al. (2003) have demonstrated that the efficiency measures are sensitive to the functional specification in the stochastic approach. To identify the most appropriate form, they use six functional specifications of a production frontier model, which amongst others include translog and Cobb-Douglas form. Their study shows that the magnitude of the bias in the estimates of efficiency is very small when the frontier is correctly specified. The smallest bias is found in the translog specification. In contrast to the translog specification, there are studies that use a simpler model. For example, Shanmugam and Das (2004), Sharma et al. (1997), and Battese and Coelli (1995) used just log linear specification.

One-way to decide which functional form is better is to apply the likelihood ratio test (LRT). The LRT is a statistical test to find out the goodness of fit between two competing models. Under this procedure, a relatively more complex model is compared to a simpler model focusing on which one significantly fits better a particular dataset. A more complex model normally has more parameters. As more parameters are added to the model, the likelihood scores will get higher. However, there is a point where additional parameters are no longer justified based on significant improvement in fit of a model to the dataset. According to Pyndick and Rubenfeld (1998), this procedure can be used to test whether some of the coefficients are equal to zero under the null hypothesis.

The LRT begins by comparing the value log- likelihood function of the two models. Let us say that L1 and L2 are the unrestricted model and restricted model respectively. The likelihood ratio test statistic is given by

$$LR = 2*(\ln L1 - \ln L2) \qquad \text{Equation 5.17}$$

The test has a chi-square $\left(\chi_r^2\right)$ distribution if the joint null hypothesis is true (Pindyck and Rubinfeld, 1998 and Griffiths et al., 1993). *r* is the number of restrictions. If the null hypothesis is not supported by the data, then the value of the test statistic becomes larger. The null hypothesis is rejected if LR is greater than or equal to χ_r^2 distribution.

The LRT is only valid if used to compare hierarchically nested models. This means that the more complex model differs from the simpler model with regard to the addition of one or more parameters. In the case of cost function, the new parameters refer to interaction variables based on the existing variables.

Equation 5.9 is re-expressed in translog functional form as:

$$\ln c_{it} = \beta_0 + \beta_1 \ln w_{1t} + \beta_2 \ln w_{2t} + \beta_3 \ln w_{3t} + \beta_4 \ln y_{1t} + \beta_5 \ln y_{2t} + \beta_6 \ln y_{3t} + \beta_7 \ln y_{1t}\, y_{1t} + \beta_8 \ln y_{1t}\, y_{2t} + \beta_9 \ln y_{1t}\, y_{3t} + \beta_{10} \ln y_{1t}\, w_{1t} + \beta_{11} \ln y_{1t}\, w_{2t} + \beta_{12} \ln y_{1t}\, w_{3t} + \beta_{13} \ln y_{2t}\, y_{2t} + \beta_{14} \ln y_{2t}\, y_{3t} + \beta_{15} \ln y_{2t}\, w_{1t} + \beta_{16} \ln y_{2t}\, w_{2t} + \beta_{17} \ln y_{2t}\, w_{3t} + \beta_{18} \ln y_{3t}\, y_{3t} + \beta_{19} \ln y_{3t}\, w_{1t} + \beta_{20} \ln y_{3t}\, w_{2t} + \beta_{21} \ln y_{3t}\, w_{3t} + \beta_{22} \ln w_{1t}\, w_{1t} + \beta_{23} \ln w_{1t}\, w_{2t} + \beta_{24} \ln w_{1t}\, w_{3t} + \beta_{25} \ln w_{2t}\, w_{2t} + \beta_{26} \ln w_{2t}\, w_{3t} + \beta_{27} \ln w_{3t}\, w_{3t} + \beta_{28} D_{95} + \beta_{29} D_{96} + \beta_{30} D_{97} + \beta_{31} D_{98} + \beta_{32} D_{99} + \beta_{33} D_{00} + \beta_{33}\, DDOM + u_i + v_{it}$$

Equation 5.18

where

w_{1t} = price of labour (Wlab) in period t

w_{2t} = price of capital (Wcap) in period t

w_{3t} = price of deposits (Wdep) in period t

y_{1t} = total loans (LOANS) in period t

y_{2t} = other earning assets (OEA) in period t

y_{3t} = other operating income (OOY) in period t

$y_{1t}y_{1t}$ = 0.5*LOANS*LOANS

$y_{1t}\,y_{2t}$ = 0.5*LOANS*OEA

$y_{1t}\,y_{3t}$ = 0.5*LOANS*OOY

$y_{1t}\,w_{1t}$ = LOANS*Wlab

$y_{1t}\,w_{2t}$ = LOANS*Wcap

$y_{1t}\,w_{3t}$ = LOANS*Wdep

$y_{2t}\,y_{2t}$ = 0.5*OEA*OEA

$y_{2t}\,y_{3t}$ = 0.5*OEA*OOY

$y_{2t}\,w_{1t}$ = OEA*Wlab

$y_{2t}\,w_{2t}$ = OEA*Wcap

$y_{2t}\,w_{3t}$ = OEA*Wdep

$y_{3t}\,y_{3t}$ = 0.5*OOY*OOY

$y_{3t}\,w_{1t}$ = OOY*Wlab

$y_{3t}\,w_{2t}$ = OOY*Wcap

$y_{3t}\,w_{3t}$ = OOY*Wdep

$w_{1t}\,w_{1t}$ = 0.5*Wlab*Wlab

$w_{1t}\,w_{2t}$ = 0.5*Wlab*Wcap

$w_{1t}\,w_{3t}$ = 0.5*Wlab*Wdep

$w_{2t}\, w_{2t}$ = 0.5*Wcap*Wcap

$w_{2t}\, w_{3t}$ = 0.5*Wcap*Wdep

$w_{3t}\, w_{3t}$ = 0.5*Wdep*Wdep

D_{95}, D_{96}, D_{97}, D_{98}, D_{99} and D_{00} = year dummies for 1995 until 2000

DDOM = bank dummy; 1 for domestic banks and 0 for foreign banks

u_i = bank's inefficiency term

v_{it} = random error

Several assumptions about the error components (u_i and v_{it}) need to be made.

i. $v_{it} \sim N(0,\sigma_v^2)$

i. $u_i \sim N(\mu, \sigma_u^2)$

iii. v_{it} and u_i are distributed independently of each other and have no relationship with other regressors.

Following the intermediation approach of what constitutes bank inputs and output, we used three inputs: labour, fixed assets and total deposits. Their prices are the price of labour (Wlab), price of fixed asset (Wcap) and price of deposit (Wdep). Three bank outputs are total loans (LOANS), other earning assets (OEA) and other operating income (OOY). The use of OOY was suggested by Rogers (1998) to represent the growing contribution of non-interest based products or fee-based services. This variable was latterly used by Shanmugam and Das (2004). All variables are in real values.

All variables are in log value except the dummy variables. As suggested by Verbeek (2000) and Maddala (2001), the use of log value depends on the motives of the study and acts as a solution to possible econometric problems. For example, Maddala (2001) states that in the presence of heteroskedasticity, two remedies are commonly used to solve the problem. The first is to transform the data into log value and the second is to deflate the variables by some measures of size. In addition, the parameters in the log function have the interpretation of elasticities. For example, the log linear cost function allows interpretation of the coefficients of outputs as cost elasticities. This is given by the estimated coefficients of the output used in the function. Thus, the log function facilitates any analysis relating to the function in question, or it can help the researcher to arrive at the objectives of the study (Gujarati, 1992).

5.6.2 Measurement of cost efficiency using DEA technique

Variations in the efficiency scores can differ when different techniques are employed. For the purpose of comparison, this study also uses DEA technique to obtain the cost efficiency of the commercial banks besides the stochastic approach. As assumed before, there exist I banks (i = 1, ..., I) that produce a vector of m outputs $y_i = (y_1, \ldots, y_m)$ using a vector of k inputs $x_i = (x_1, \ldots, x_k)$ for which they pay prices $w_i = (w_1, \ldots, w_k)$. The cost efficiency for a particular firm o is defined as;

$$CEo = \frac{C_o^*}{C_o} = \frac{\sum_{i=1}^{k} w_{io} x_{io}^*}{\sum_{i=1}^{k} w_{io} x_{io}}$$ Equation 5.19

This definition implies that the cost efficiency of firm *o* is indeed the ratio between the minimum cost (C_o^*) – associated with the use of the inputs vector (x_o*) that minimise costs and the observed cost (C_o).

The cost efficiency for firm *o* can be calculated by solving the following linear programming problem;

$$\min \sum_{k} w_{ko} x_{ko}$$ Equation 5.20

subject to;

$$\sum^{I} \lambda_j y_{kj} \geq y_{ko}$$

$$\sum^{I} \lambda_j x_{kj} \leq x_{ko}$$

$\lambda_j \geq 0$ for constant returns to scale and $\lambda_j = 1$ for variable returns to scale.

The efficiency scores are within the range 0 and 1.

5.7 Descriptive statistics: input prices and output

The data comprise the majority of the commercial banks in Malaysia. We made every attempt to ensure that our sample was close to the total population. However, for technical reasons such as bank consolidation and incomplete data, we had to exclude several banks. For example, we excluded one bank in 1997 and another in 1998 as they merged with another bank. Two more banks were removed as they merged in 1999. In 2000, several domestic banks were excluded since the data were not available or they had merged. Our total number of bank-year observations was 193. The period runs from 1994 until 2000. After 2000, all domestic banks, excluding two Islamic banks, are considered as new entities as a result of merging with other banks. The maximum number of banks in a single year was 32 (in 1995 and 1996) and the lowest was 20 banks (in 1994 and 2000). Hence, our panel is not balanced.

We used the annual balance sheet and income statement of the banks in order to identify and construct the variables for empirical analysis. Three inputs and three outputs were used, which were selected based on the intermediation approach. This approach is in accordance with the functions of commercial banks in the economy and is the approach that has normally been undertaken by other studies in the literature. The inputs are labour, fixed assets and total deposits. The main problem here is that the input prices are not directly observed. To overcome this obstacle, we calculated each of them as follows:

- the price of labour (Wlab) is given by personal expenses divided by the total number of bank staff
- the price of capital (Wcap) is given by real non-interest expenses divided by real fixed assets
- the price of deposit (Wdep) is given by real total interest expenses divided by total deposits received. Total deposits include demand deposits, savings deposits, and fixed deposits.

The three bank outputs are total loans (LOANS), other earning assets (OEA) and fee-based products. The fee-based products are represented by other operating income (OOY). Table 5.1 shows the descriptive statistics of input prices and bank outputs. By looking at the outputs, it is clearly shown that there are substantial gaps amongst the banks. For example in 2000, the minimum LOANS was RM 146.30 millions, while the maximum was RM 61,003.90 millions. This indicates the presence of both very large banks and small banks in the sample. The same comparison applies to other outputs such as other earning assets and other operating income. Figures 5.2 and 5.3 show the pattern of relationship between the total operating cost (OPC) and the three outputs. The OPC and LOANS rose steadily from 1994 until 1998 and declined between 1998 and 1999. However, it should be noted that in 1999, the second largest bank merged with another bank and the new bank was excluded from the sample. This might help to explain the decline in the OPC and bank outputs. The OOY, on average, was relatively stable throughout the period. In term of input prices, the average cost of bank inputs had generally increased, with the exception of the price of deposits, which continued to rise up to 1998 but declined after that. (See Figures 5.4 and 5.5.below.) This was in line with current monetary policy that tried to lower the cost of borrowing in order to stimulate the economy, which was badly affected by the financial crisis in 1997-98.

Table 5.1: Descriptive statistics of banks input prices and output

Years	OPC[2]	Wlab[1]	Wcap[1]	Wdep[1]	LOANS[3]	OEA[3]	OOY[3]
1994							
Average	338.97	0.0303	0.7929	0.0447	3797.80	3125.14	54.56
Std. Dev.	421.65	0.0075	0.4434	0.0198	3574.17	4624.17	57.63
Maximum	1683.70	0.0429	1.6755	0.0898	13378.1	17818.60	210.00
Minimum	9.60	0.0106	0.1971	0.0141	271.00	76.00	2.00
1995							
Average	404.68	0.0401	1.0007	0.0649	4679.94	3003.24	68.97
Std. Dev.	572.38	0.0149	0.6427	0.0489	6331.81	4794.77	96.27
Maximum	2879.30	0.0781	2.7407	0.2470	32543.0	23276.80	450.30
Minimum	23.40	0.0103	0.1663	0.0079	0324.3	100.30	0.70
1996							
Average	540.09	0.0434	0.8552	0.0751	5894.96	3415.14	79.06
Std. Dev.	701.80	0.0181	0.5290	0.0540	7213.47	5045.81	98.42
Maximum	3510.90	0.1079	2.5556	0.3170	36374.00	25404.80	452.40
Minimum	32.30	0.0189	0.2495	0.0415	299.00	166.40	4.00
1997							
Average	697.72	0.0585	0.9497	0.0838	7167.86	4543.30	97.57
Std. Dev.	821.62	0.0497	0.8236	0.0629	8539.29	6559.80	123.31
Maximum	4028.80	0.3048	4.5882	0.3901	45297.70	31326.10	586.30
Minimum	52.80	0.0299	0.3033	0.0474	334.50	132.90	4.70
1998							
Average	991.98	0.0514	0.9420	0.1066	8108.27	4467.53	117.39
Std. Dev.	1295.44	0.0271	0.6311	0.0500	10533.01	5678.35	159.67
Maximum	6774.60	0.1337	3.1905	0.2686	56277.20	24470.10	725.40
Minimum	58.20	0.0288	0.3599	0.0475	263.50	161.30	-4.60
1999							
Average	685.14	0.0562	1.1160	0.0655	7449.49	4713.92	107.76
Std. Dev.	943.92	0.0402	0.8767	0.0404	10516.18	6032.22	152.51
Maximum	5074.10	0.2229	4.1818	0.1890	57489.40	26646.60	772.00
Minimum	32.10	0.0292	0.3354	0.0196	212.60	251.80	3.30
2000							
Average	906.14	0.0659	1.3658	0.0585	9466.69	6453.67	150.25
Std. Dev.	1500.61	0.0314	1.1084	0.0621	12957.92	7913.27	186.56
Maximum	6333.10	0.1438	3.9630	0.3075	61003.90	32091.00	800.70
Minimum	25.10	0.0323	0.3589	0.0241	146.30	360.70	4.60

Note: 1. Input prices are calculated manually. The price of labour (Wlab) = total personal expenses divided by total staff. The price of capital (Wcap) = total non-interest expenses divide by real fixed assets. The price of deposit (Wdep) = total interest expenses divided by total deposits received.
2. OPC is total operating cost.
3. LOANS is total loans issued by the banks. OEA is other earning assets held by the banks. OOY is other operating income to represent banks' fee based product or services.

Sources: ABM (1994, 1996, 1998, and 2000) and BANKSCOPE

Figure 5.2: Average value of operating cost, total loans, and other earning assets (1994 – 2000)

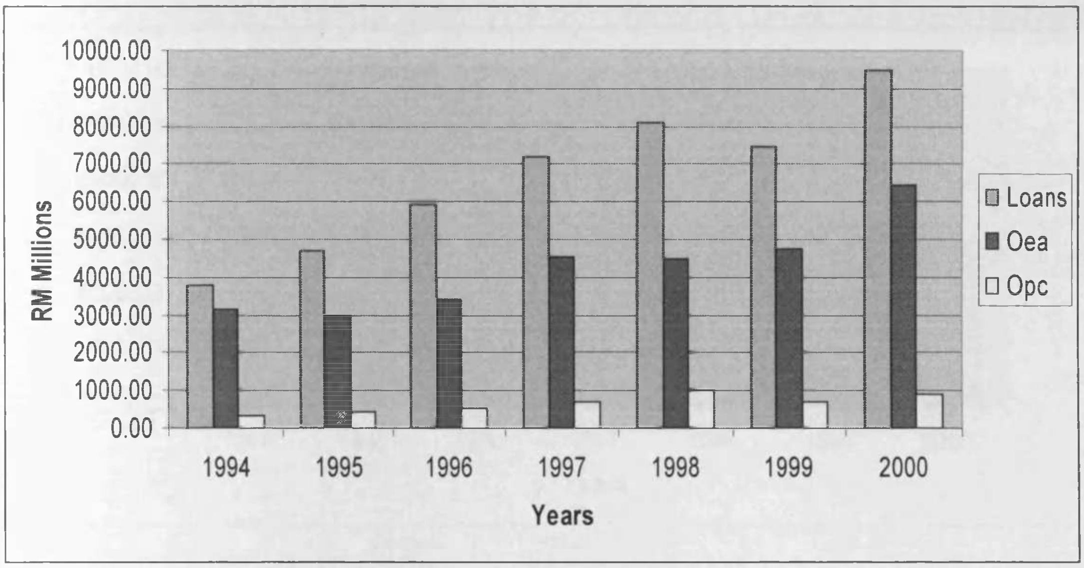


Figure 5.3: Average operating cost and other operating income (1994 –2000)

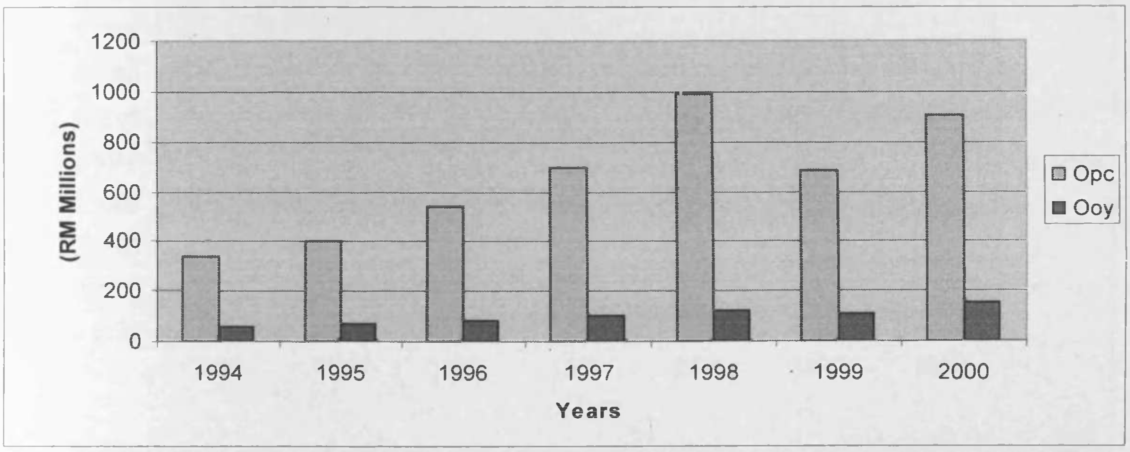


Figure 5.4: Price of labour and price of deposits

(1994 –2000)

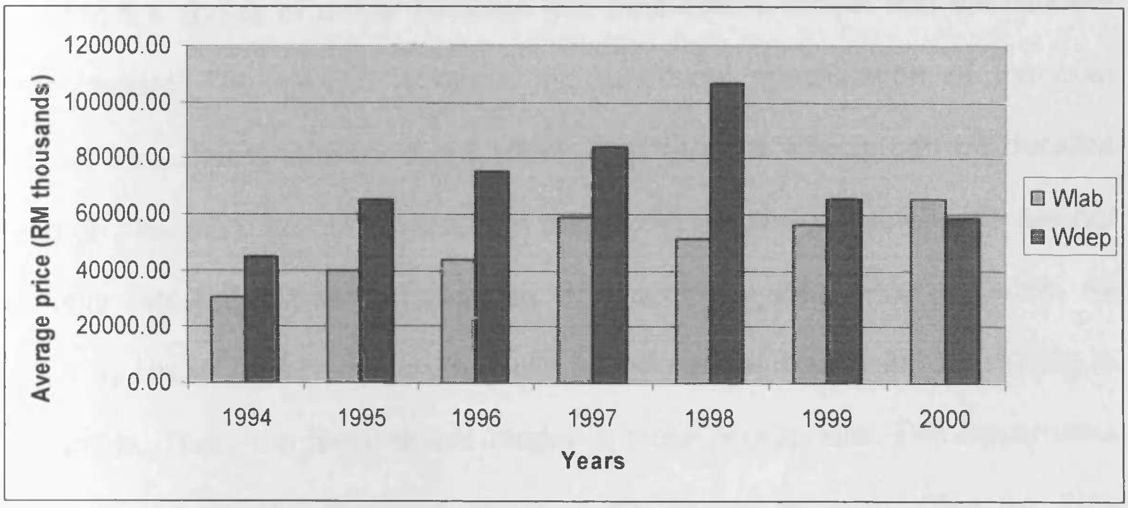
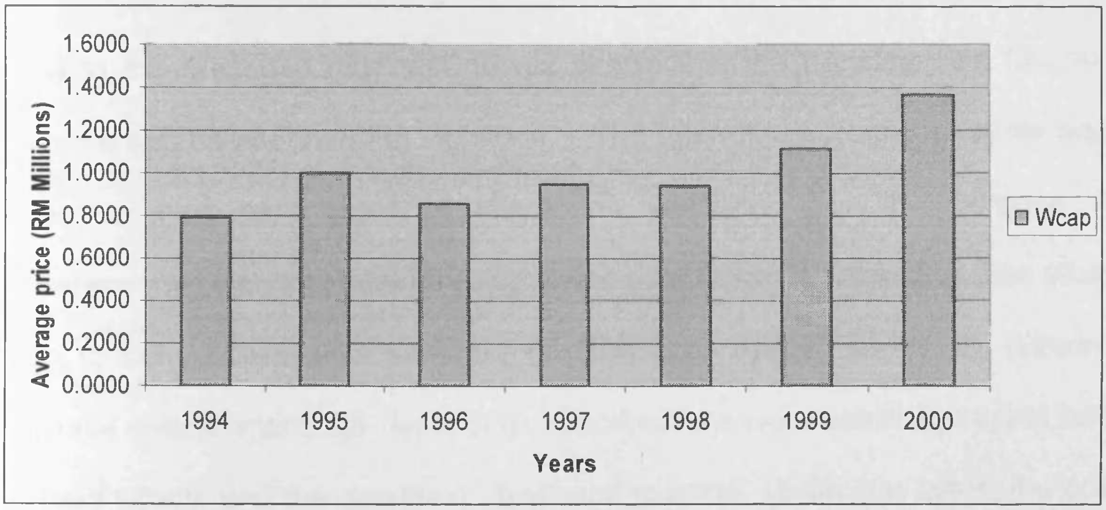


Figure 5.5: Price of capital

(1994 –2000)



5.8 Empirical results

Two main issues need to be tackled before the estimation takes place. The first is related to the choice of model between the fixed effects model and the random effects model. The second is about the functional specification of the cost function. The choice between fixed effects and random effects can be decided based on how the data are collected. In this study, the commercial banks are not randomly selected but are all included whenever the data were available. As argued by Hsiao (2003), this study looks for inferences based on the effects in the sample. Thus, the fixed effects model is more appropriate. The Hausman's test also suggests that the fixed effects model should be used. (See pp. 202-205)[31] To perform the Hausman's test, one need to estimate the random effects model. Because of the Hausman's test results, we did not pursue the random effects model.

Following the likelihood ratio test, it was shown that the translog cost function was more appropriate than the log linear[32]. The later differs from the former with respect to new parameters or interaction variables. For the purpose of comparison, we also estimated the log linear cost function. Moreover, this study wants to demonstrate how sensitive the efficiency measures are as different functional specifications are used. Both specifications were estimated using both the fixed effects and the maximum likelihood method. Under the latter, the cost

[31] Chi square (X^2) = 53.04 and Prob> X^2 = 0.0000

[32] LR chi square (X^2) = 106.08 and prob> X^2 = 0.0000.

efficiency measure is assumed to be time-invariant. Variable DDOM (dummy variable for bank ownership) is dropped under the fixed effects model since there is no variation within the group. The estimation process was done using Stata software version 8.

5.8.1 Results by functional specifications

Besides the translog cost function, we also estimated the log-linear cost function. Tables 5.2 and 5.3 present the regression estimates of Equation 5.18. The dependent variable is the total operating cost. As mentioned before, the two specifications of the cost function differ from one another in terms of interaction variables. It is clearly shown that the log linear specification under both fixed effects and stochastic frontier models produces the expected estimates with correct signs except for w_{2t} (price of capital). In contrast, the estimates produced by the translog cost function are mostly insignificant.

- *Input prices*. Under translog specification in both models, the coefficients of w_{1t}, w_{2t} and w_{3t} have mixed signs. However, none of them are significant. Under log-linear specification (Table 5.3), we found that the coefficient of w_{3t} is greater than the coefficient of w_{1t} and w_{2t}. This shows that the price of deposit has a substantial impact on the operating cost. For example, a 1% increase in the price of deposit will cause about a 0.56% increase in the operating cost. The coefficient of w_{1t} (price of labour) is 0.1739, smaller than the coefficient of w_{3t}. Since the value is small, it implies that the cost of labour forms a relatively minor component of the total operating cost compared to the cost of total deposits.

- *Bank outputs*. We used three outputs in this study. Under translog specification in both models, we found that only one is significantly positive, y_{3t} (other operating income). The other two inputs have negative signs, but are not significant. However, the size of the y_{3t} coefficient is greater than one. Under the fixed effects model (Table 5.2), the estimate of the y_{3t} coefficient is 1.0567. This means that if y_{3t} increases by 1%, then the operating cost will increase by 1.06%. Under the stochastic frontier model, the estimate of y_{3t} is 1.1319. Under the log-linear specification, all outputs are significantly positive with y_{1t} has the largest coefficient, followed by y_{2t} and y_{3t}.
- *Interaction variables*. These variables are created by crossing amongst input prices and output. All output in quadratic terms ($y_{1t}y_{1t}$, $y_{2t}y_{2t}$ and $y_{3t}y_{3t}$) are significantly positive except $y_{2t}y_{2t}$ in fixed effects model. Input prices in quadratic terms have a positive sign but are not significant. This can imply that the translog specification can capture the underlying bank technology. In other cross prices and cross outputs, the signs are uncertain depending on individual interactions.
- *Dummy variables*. In the stochastic frontier model, all time dummies are not significant under translog specification, but significant under the log linear specification. All time dummies except in 1995 are significantly

positive. The coefficients are close to zero between 1995 and 1997 and become the largest in 1998 (0.2294). This suggests that the event in 1998 (financial crisis was taking place) had a substantial impact on the operating cost of the banks. The value of the coefficients gets smaller in 1999 and 2000. Under the fixed effects model, five time dummies are significantly positive, D_{96}, D_{97}, D_{98}, D_{99} and D_{00}. Again, the largest coefficient occurs in 1998. The Bank dummy (DDOM - 1 for domestic banks and 0 for foreign banks) is significantly positive under the stochastic frontier model. This means that domestic ownership has a significant impact on the operating cost. It implies that domestic ownership contributes towards a higher operating cost as compared with foreign ownership. According to Okuda and Hashimoto (2004), high and positive DDOM implies cost inefficiency while lower and negative DDOM implies cost efficiency of the banks.

Since we used panel data, the presence of heteroskedasticity might be a matter of concern. Heteroskedasticity occurs when one of the main assumptions under OLS is violated i.e. the variance of error term is no longer constant. However, under a fixed effects model, this problem can only occur through the v_{it} or the noise component (refer to Equation 5.18.) Since we assume that the v_{it} are iid $(0,\sigma_v^2)$ and that the $u_i \geq 0$ and is time-invariant, an MLM estimator produces unbiased estimates of the intercept and the coefficients as in Equation 5.18. If we decide to neglect the heteroskedasticity in the noise component of the fixed effects model, this is not a serious problem. All cost frontier parameters and producer-specific cost efficiency are consistently estimated (Kumbhakar and Lovell, 2000; pp. 123).

Table 5.2: Estimation results of fixed effects model

Variables	Translog cost function			Log linear cost function		
	Coefficients	Std Error	P>z	Coefficients	Std Error	P>z
constant	4.5337	1.9567	0.022	2.5971	0.3829	0.000*
w_{1t}	0.2149	0.5914	0.717	0.1951	0.0405	0.000*
w_{2t}	-0.2526	0.2910	0.387	-0.0274	0.0334	0.412
w_{3t}	0.2857	0.3637	0.434	0.6038	0.0319	0.000*
y_{1t}	-0.0752	0.4024	0.852	0.3572	0.0390	0.000*
y_{2t}	-0.4316	0.3647	0.239	0.2724	0.0366	0.000*
y_{3t}	1.0567	0.2982	0.001**	0.0820	0.0240	0.001**
$y_{1t}\ y_{1t}$	0.1424	0.0731	0.054***			
$y_{1t}\ y_{2t}$	-0.0929	0.0960	0.335			
$y_{1t}\ y_{3t}$	-0.2650	0.0847	0.002**			
$y_{1t}\ w_{1t}$	-0.0248	0.0666	0.711			
$y_{1t}\ w_{2t}$	0.0021	0.0484	0.965			
$y_{1t}\ w_{3t}$	-0.0520	0.0557	0.353			
$y_{2t}\ y_{2t}$	0.0728	0.0529	0.171			
$y_{2t}\ y_{3t}$	0.0316	0.0688	0.647			
$y_{2t}\ w_{1t}$	-0.0900	0.0854	0.294			
$y_{2t}\ w_{2t}$	0.0147	0.0430	0.733			
$y_{2t}\ w_{3t}$	-0.0383	0.0428	0.372			
$y_{3t}\ y_{3t}$	0.1170	0.0473	0.015***			
$y_{3t}\ w_{1t}$	0.0125	0.0839	0.882			
$y_{3t}\ w_{2t}$	-0.0385	0.0473	0.416			
$y_{3t}\ w_{3t}$	0.1497	0.0547	0.007**			
$w_{1t}\ w_{1t}$	-0.0523	0.1150	0.650			
$w_{1t}\ w_{2t}$	-0.1051	0.1536	0.495			
w_{1t} w3t	-0.4514	0.2130	0.036			
$w_{2t}\ w_{2t}$	0.0774	0.0735	0.294			
$w_{2t}w_{3t}$	-0.0987	0.0823	0.233			
$w_{3t}\ w_{3t}$	0.1168	0.0633	0.067			
D_{95}	0.0162	0.0352	0.647	0.0472	0.0366	0.199
D_{96}	0.0956	0.0427	0.027***	0.1126	0.0403	0.006**
D_{97}	0.1440	0.0530	0.007**	0.1650	0.0466	0.001*
D_{98}	0.1924	0.0589	0.001**	0.2811	0.0472	0.000*
D_{99}	0.1714	0.0508	0.001**	0.2404	0.0461	0.000*
D_{00}	0.1263	0.0564	0.027***	0.2340	0.0519	0.000*
Sigma_u	0.4433			0.4504		
Sigma_e	0.0954			0.1201		
rho	0.9557			0.9337		
R^2 within	0.9702			0.9527		
R^2 between	0.9068			0.9047		
R^2 overall	0.9054			0.9001		

Note: * = significant at 1% level.
** = significant at 5% level
*** = significant at 10% level.

Sigma_u is the standard deviation of the individual effect (u_i) and sigma_e is the standard deviation of the random error (e_i). Rho is fraction of the unexplained variance due to differences among the units. It is given by $Var[u_i\}/(Var[u_i]+Var[e_{it}])$. R^2 within is a measure of the explained variation within units. It is defined as the squared correlation between deviations of y_{it} values from unit means (y_{it} – y_i bar) and deviations of predicted values from unit mean predicted values (y_{it} hat – y_i hat). R^2 between is a measure of the explained variation between units. It is defined as the squared correlation between unit means (y_i bar) and y_i hat values predicted from unit means of the independent variables.

R^2 overall is a measure of the explained overall variation. It is defined as the squared correlation between observed (y_{it}) and predicted (y_{it}hat) values. The fixed effects model performs a good job fitting the observed operating cost overall (R^2 = 0.9054). The model is also able to explain about 90% of the variations amongst the banks (R^2 = 0.9068). With regard to the variations within the groups, this model explains up to 97% of these variations (R^2 = 0.9702).

Table 5.3: Estimation results of stochastic frontier model

using maximum likelihood method

Variables	Translog cost function			Log linear cost function		
	Coefficients	Std Error	P>z	Coefficients	Std Error	P>z
constant	2.0415	1.6348	0.212	0.3632	0.3697	0.326
w_{1t}	-0.2287	0.5433	0.674	0.1739	0.0404	0.000*
w_{2t}	-0.4223	0.2655	0.112	-0.0636	0.0324	0.050**
w_{3t}	0.4058	0.3205	0.206	0.5645	0.0322	0.000*
y_{1t}	-0.3412	0.3515	0.332	0.4380	0.0379	0.000*
y_{2t}	-0.0614	0.3156	0.846	0.3234	0.0321	0.000*
y_{3t}	1.1319	0.2701	0.000*	0.1026	0.0240	0.000*
$y_{1t}\,y_{1t}$	0.1840	0.0664	0.006**			
$y_{1t}\,y_{2t}$	-0.1488	0.0848	0.079***			
$y_{1t}\,y_{3t}$	-0.2700	0.0782	0.001**			
$y_{1t}\,w_{1t}$	-0.0586	0.0605	0.333			
$y_{1t}\,w_{2t}$	-0.0156	0.0440	0.722			
$y_{1t}\,w_{3t}$	-0.1043	0.0518	0.044**			
$y_{2t}\,y_{2t}$	0.0997	0.0465	0.032**			
$y_{2t}\,y_{3t}$	0.0087	0.0624	0.889			
$y_{2t}\,w_{1t}$	0.0294	0.0766	0.701			
$y_{2t}\,w_{2t}$	0.0033	0.0385	0.931			
$y_{2t}\,w_{3t}$	-0.0820	0.0384	0.033**			
$y_{3t}\,y_{3t}$	0.1496	0.0437	0.001**			
$y_{3t}\,w_{1t}$	-0.0303	0.0768	0.693			
$y_{3t}\,w_{2t}$	-0.0020	0.0427	0.963			
$y_{3t}\,w_{3t}$	0.2228	0.0503	0.000*			
$w_{1t}\,w_{1t}$	0.0252	0.1009	0.803			
$w_{1t}\,w_{2t}$	-0.2162	0.1399	0.122			
w_{1t} w3t	-0.5817	0.1876	0.002**			
$w_{2t}\,w_{2t}$	0.0584	0.0670	0.363			
$w_{2t}w_{3t}$	-0.1189	0.0767	0.121			
$w_{3t}\,w_{3t}$	0.0555	0.0582	0.340			
D_{95}	-0.0194	0.0323	0.549	0.0407	0.0366	0.267
D_{96}	0.0146	0.0378	0.699	0.0713	0.0340	0.074***
D_{97}	0.0254	0.0442	0.565	0.0947	0.0452	0.036**
D_{98}	0.0764	0.0495	0.122	0.2294	0.0459	0.000*
D_{99}	0.0622	0.0416	0.135	0.1574	0.0442	0.000*
D_{00}	0.0067	0.0468	0.886	0.1415	0.0496	0.004**
DDOM	0.3289	0.0978	0.001**	0.3290	0.1037	0.002**
mu	0.6706	0.2314	0.004	0.7724	0.2713	0.004
sigma2	0.0588	0.0176		0.0784	0.0226	
gamma	0.8640	0.0465		0.8142	0.0616	
sigma_u2	0.0508	0.0178		0.0639	0.0230	
sigma_v2	0.0080	0.0010		0.0146	0.0018	
Log likelihood	133.0084			81.0828		
Wald χ^2	6080.02			3366.21		
Prob > χ^2	0.0000			0.0000		

Note: * = significant at 1% level.
** = significant at 5% level
*** = significant at 10% level.

mu is the estimate of μ. The estimated μ is positive and is significantly different from zero. sigma2 is the estimate of σ_s^2. $\sigma_s^2=\sigma_v^2+\sigma_u^2$. Gamma is the estimate of γ. $\gamma=\sigma_u^2/\sigma_s^2$. Both σ_s^2 and γ are the variance parameters. The estimate of γ is also significantly from zero which implies the inefficiency effects are significant in determining the level and variability of the production cost of the banks. sigma_u2 is the estimate of σ_u^2 . sigma_v2 is the estimate of σ_v^2 .

5.8.2 Measurement of cost efficiency

The cost efficiency of bank *i* (CE_i) is estimated based on Equation 5.13 (under the fixed effects model) and Equation 5.16 (under the stochastic frontier model). Since two cost specifications were estimated, four estimates of cost efficiency are generated. They are shown in Table 5.4. Under the fixed effects model, the average cost efficiency measure for CE1 and CE2 is 32%. Under the stochastic frontier model, the average cost efficiency measure is 52.65% for CE3. On the other hand, the average cost efficiency measure by log linear function (CE4) is slightly lower. It stands at 47.89%. The mean cost efficiency measure of 53% indicates that, on average, the sample banks waste 47% of their costs relative to a best practice bank facing the same environment. The results of the paired t-test show that the mean efficiency scores for the two functional specifications (CE3 and CE4) are significantly different. The null hypothesis is that the difference between CE3 and CE4 is zero. However, the null hypothesis that the difference between CE1 and CE2 is zero cannot be rejected.

The mean cost efficiency measures generated by the fixed effects model are lower than the recent findings obtained by Shanmugan and Das (2004). They found that the mean technical efficiency value of Indian banks was 44.6%. Their study covered the period between 1992 and 1999 and used a fixed effects model. Instead of a stochastic cost frontier, they used a stochastic production function with Cobb-Douglas functional form. This means that their efficiency measures refer to technical efficiency not cost efficiency.

The findings in this study differ from the findings obtained by Abd_Karim (2001) and Edward (1999). Abd_Karim (2001) studied banking efficiency in four ASEAN countries (Thailand, Malaysia, Philippines, and Indonesia) for the period between 1989 and 1996. Input prices are the price of capital, the price of labour, and the price of funds. Bank outputs are deposits, loans, and investments. Using a translog cost function, he found that the average cost efficiency was around 95% in Malaysia, 98% in Thailand, 82% in Indonesia and 66% in the Philippines. Edwards (1999) studied banking markets in six emerging countries, Malaysia, Philippines, Thailand, India, Indonesia, and Korea. His study also used translog cost specification. Input prices are the price of labour, the price of capital and the price of deposit; only two outputs are used, loans and other earning assets. In the case of Malaysia, he found that the average cost efficiency of the commercial banks was 90% for the period between 1986 and 1995.

We should note that both Abd_Karim (2001) and Edwards (1999) used a single approach (parametric approach) to estimate the cost efficiency of commercial banks in Malaysia. This study not only uses a parametric approach but also a non-parametric approach. The use of both approaches enables a better understanding of the banks' performance and banks' ranking. We also take into account the presence of foreign banks in Malaysia when analysing the performance of the banks. Thus, it can be shown whether the performance and ranking of the banks are affected by the various techniques used.

In another recent study, Girardone et al. (2004) studied Italian banks using a Fourier-flexible stochastic cost frontier for the period between 1993 and 1996. Using cross-sectional data, they found that the average values of X-efficiency exceeded 80% and average inefficiency levels ranged between approximately 13% and 15%. They also found that X-inefficiency levels showed a decreasing trend over time.

Table 5.4: Efficiency scores by methods and functional forms

Banks	Types of banks[a]	CE1[b]	CE2[c]	CE3[d]	CE4[e]
Bank 1	1	0.2152	0.2121	0.4891	0.4324
Bank 2	1	0.2300	0.2203	0.4747	0.4038
Bank 3	2	0.5493	0.7524	0.5623	0.7165
Bank 4	1	0.3330	0.3462	0.6808	0.6063
Bank 5	1	0.2201	0.2292	0.4530	0.4297
Bank 6	2	0.4656	0.5060	0.4709	0.4781
Bank 7	1	0.1580	0.1526	0.4362	0.3693
Bank 8	1	0.2695	0.2591	0.4972	0.4471
Bank 9	1	0.4017	0.4493	0.8048	0.8151
Bank 10	1	0.1916	0.1860	0.4617	0.3954
Bank 11	2	0.8751	0.5963	0.8285	0.5669
Bank 12	2	0.2511	0.2526	0.3955	0.3558
Bank 13	1	0.3058	0.3233	0.5863	0.5571
Bank 14	1	0.2281	0.2197	0.4224	0.3767
Bank 15	1	0.2129	0.2172	0.4641	0.4320
Bank 16	2	0.1794	0.1709	0.3261	0.2732
Bank 17	1	0.3366	0.3672	0.5082	0.4970
Bank 18	1	0.1605	0.1548	0.5175	0.4148
Bank 19	2	1.0000	1.0000	0.9175	0.8895
Bank 20	2	0.2198	0.2089	0.3794	0.3180
Bank 21	1	0.2577	0.2597	0.5322	0.4755
Bank 22	2	0.3362	0.3096	0.4762	0.3888
Bank 23	1	0.2301	0.2322	0.4833	0.4397
Bank 24	1	0.2618	0.2722	0.5870	0.5527
Bank 25	1	0.1636	0.1600	0.4050	0.3747
Bank 26	1	0.2231	0.2088	0.5753	0.4669
Bank 27	1	0.3166	0.3532	0.5976	0.5817
Bank 28	1	0.2341	0.2343	0.4756	0.4332
Bank 29	2	0.2717	0.2762	0.4809	0.4314
Bank 30	2	0.5427	0.4872	0.6388	0.5550
Bank 31	2	0.2758	0.2904	0.4198	0.3989
Bank 32	1	0.3318	0.3485	0.4994	0.4504
Average		0.3203	0.3205	0.5265	0.4789
Standard Deviation		0.1898	0.1830	0.1307	0.1336
Minimum		0.1580	0.1526	0.3261	0.2732
Maximum		1.0000	1.0000	0.9175	0.8895

Notes: a. 1 = domestic banks, 2 = foreign banks.

b. Cost efficiency scores with translog cost function under fixed effects model.

c. Cost efficiency scores with log linear function under fixed effects model.

d. Cost efficiency scores with translog cost function under stochastic frontier model.

e. Cost efficiency scores with log linear function under stochastic frontier model.

5.8.3 Cost efficiency measures by bank ranking

The cost efficiency of the banks (obtained under the stochastic frontier model) is now analysed in terms of their rankings. The banks are divided into two types: domestic banks and foreign banks. The ranking of the banks is presented in Table 5.5.

One significant aspect of the managerial interest of the banks' rankings is their positions at the top and the bottom of the distribution. In the top ten, the domestic banks dominate the rankings. For example, under CE3, six domestic banks dominate the top positions. Meanwhile, if we examine the bottom ten, they are also the main contributors. They contribute six banks under CE3 and five banks under CE4. At the very bottom of the rankings, regardless of the models used, three banks remain as the most inefficient banks in this study. All of them are owned by foreigners.

Table 5.5: Cost efficiency scores by bank ranking

Banks	Type[a]	CE1	Banks	Type[a]	CE2	Banks	Type[a]	CE3[b]	Banks	Type[a]	CE4[c]
19	2	1.0000	19	2	1.0000	19	2	0.9175	19	2	0.8895
11	2	0.8751	3	2	0.7524	11	2	0.8285	9	1	0.8151
3	2	0.5493	11	2	0.5963	9	1	0.8048	3	2	0.7165
30	2	0.5427	6	2	0.5060	4	1	0.6808	4	1	0.6063
6	2	0.4656	30	2	0.4872	30	2	0.6388	27	1	0.5817
9	1	0.4017	9	1	0.4493	27	1	0.5976	11	2	0.5669
17	1	0.3366	17	1	0.3672	24	1	0.5870	13	1	0.5571
22	2	0.3362	27	1	0.3532	13	1	0.5863	30	2	0.5550
4	1	0.3330	32	1	0.3485	26	1	0.5753	24	1	0.5527
32	1	0.3318	4	1	0.3462	3	2	0.5623	17	1	0.4970
27	1	0.3166	13	1	0.3233	21	1	0.5322	6	2	0.4781
13	1	0.3058	22	2	0.3096	18	1	0.5175	21	1	0.4755
31	2	0.2758	31	2	0.2904	17	1	0.5082	26	1	0.4669
29	2	0.2717	29	2	0.2762	32	1	0.4994	32	1	0.4504
8	1	0.2695	24	1	0.2722	8	1	0.4972	8	1	0.4471
24	1	0.2618	21	1	0.2597	1	1	0.4891	23	1	0.4397
21	1	0.2577	8	1	0.2591	23	1	0.4833	28	1	0.4332
12	2	0.2511	12	2	0.2526	29	2	0.4809	1	1	0.4324
28	1	0.2341	28	1	0.2343	22	2	0.4762	15	1	0.4320
23	1	0.2301	23	1	0.2322	28	1	0.4756	29	2	0.4314
2	1	0.2300	5	1	0.2292	2	1	0.4747	5	1	0.4297
14	1	0.2281	2	1	0.2203	6	2	0.4709	18	1	0.4148
26	1	0.2231	14	1	0.2197	15	1	0.4641	2	1	0.4038
5	1	0.2201	15	1	0.2172	10	1	0.4617	31	2	0.3989
20	2	0.2198	1	1	0.2121	5	1	0.4530	10	1	0.3954
1	1	0.2152	20	2	0.2089	7	1	0.4362	22	2	0.3888
15	1	0.2129	26	1	0.2088	14	1	0.4224	14	1	0.3767
10	1	0.1916	10	1	0.1860	31	2	0.4198	25	1	0.3747
16	2	0.1794	16	2	0.1709	25	1	0.4050	7	1	0.3693
25	1	0.1636	25	1	0.1600	12	2	0.3955	12	2	0.3558
18	1	0.1605	18	1	0.1548	20	2	0.3794	20	2	0.3180
7	1	0.1580	7	1	0.1526	16	2	0.3261	16	2	0.2732

Notes: a. 1 = domestic banks, 2 = foreign banks.
b. CE1 = cost efficiency scores with translog cost function under fixed effects model. CE2 = cost efficiency scores with log linear function under fixed effects model. CE3 = cost efficiency scores with translog cost function under stochastic frontier model. CE4 = cost efficiency scores with log linear function under stochastic frontier model.

5.8.4 Efficiency scores and bank ownership

It is interesting to look at the banks' performance from an ownership point of view. The banks are categorised into two types: domestic banks and foreign banks. The performance of each group is shown in Table 5.6. Under the fixed effects model (CE1 and CE2), the average efficiency scores of the foreign banks are always higher than are those of the domestic banks. For CE1, the mean efficiency measures are 45.15% and 25.15% respectively. Similar scores are obtained under CE2. In contrast, under the SFA, the performance of the two groups, on average, is not much different. The mean efficiency measures (CE3) for domestic banks and foreign banks are 52.15% and 53.60% respectively. For CE4, the mean efficiency measures are 47.39% and 48.84% respectively. We implement the one sample t-test to determine whether the difference in efficiency scores is significant. The result shows that the difference in the mean efficiency score between the two groups is statistically significant under the fixed effects model, but not statistically significant under the SFA. In a similar study, Shanmugam and Das (2004) categorised the banks into 4 groups. They found that state-owned banks are more efficient than their counterparts (foreign banks, private domestic banks and nationalised banks). The average technical efficiency values are 59% for state bank group, 52% for foreign banks, 46% for nationalised banks, and finally 32% for private domestic banks.

Table 5.6: Average efficiency scores by types of bank

Types of banks	CE1	CE2	CE3	CE4
Domestic banks				
Average efficiency	0.2515	0.2574	0.5215	0.4739
Std Deviation	0.0653	0.0793	0.0931	0.1033
Minimum	0.1580	0.1526	0.4050	0.3693
Maximum	0.4017	0.4493	0.8048	0.8151
Foreign banks				
Average efficiency	0.4515	0.4410	0.5360	0.4884
Std Deviation	0.2725	0.2587	0.1884	0.1838
Minimum	0.1794	0.1709	0.3261	0.2732
Maximum	1.0000	1.0000	0.9175	0.8895
All banks				
Average efficiency	0.3203	0.3205	0.5265	0.4789
Std Deviation	0.1898	0.1830	0.1307	0.1336
Minimum	0.1580	0.1526	0.3261	0.2732
Maximum	1.0000	1.0000	0.9175	0.8895

Notes: CE1 = cost efficiency scores with translog cost function under fixed effects model. CE2 = cost efficiency scores with log linear function under fixed effects model. CE3 = cost efficiency scores with translog cost function under stochastic frontier model. CE4 = cost efficiency scores with log linear function under stochastic frontier model.

5.8.5 Economies of scale

Economies of scale refer to the ability of firms to lower the average cost of production as they expand their output. One way to explain this is by looking at the relationship between total cost and total output. Since all variables are in log value except the dummy variables, the estimated coefficients can be interpreted as cost elasticities. This shows us how the total cost responds to any changes in output. In other words, the economies of scale in banking are measured by the reciprocal of the elasticity of cost with respect to outputs. For example under the log linear model (Table 5.3), the coefficient of y_{1t} (loans) is 0.4380. This means that a 1% increase in loans will cause the operating cost to increase by 0.4380%, *ceteris paribus.* Since the percentage increase in the operating cost is smaller than the percentage increase in output, the average cost must be falling. This proves the existence of economies of scale. [33]

It is widely accepted that banks produce multiple outputs. So, we need to consider the impact of all outputs produced on the firms' cost of production. Under the multiple outputs cost function, the economies of scale can be defined as the sum of cost elasticity with respect to each output (Girardone et al., 2004; Edward, 1999). This can be written as:

---

[33] If we use a production function, then by adding the coefficients of the inputs used, we obtain an economically important parameter known as the returns to scale parameter. This parameter will tell us the response of output to a proportional change in inputs, Gujarati (1992).

$$\varepsilon_S = \sum_m \frac{d \ln c}{d \ln y}$$

where

ε_S = the economies of scale

$\frac{d \ln c}{d \ln y}$ = the coefficients of firm's *m* output

The economies of scale parameters are calculated and shown in Table 5.7.

Table 5.7: The economies of scale parameters

	Fixed Effects Model		**Stochastic Frontier Model**	
Economies of scale	**Translog function**	**Log linear function**	**Translog function**	**Log linear function**
Total parameters	0.49	0.71	0.75	0.86

Economies of scale occur when $\varepsilon_S < 1$. If $\varepsilon_S > 1$, diseconomies of scale will take place. Our findings show that the commercial banks enjoy economies of scale. Given that $\varepsilon_S = 0.75$, this means that a 10% increase in the production of output will increase the production costs by 7.5%, given the level of input prices. However, our estimates are lower than those found by Hashim (2001). His estimate of economies of scale was 0.9983 (close to unity, implying constant returns to scale). His study covered the period between 1988 and 1998 and used three outputs (investments, loans, and bank branches). Another study by Abd_Karim (2001) found that the banks experience economies of scale (increasing returns to scale). However, the ASEAN banks exhaust scale economies at an asset size of US$3 billion.

5.8.6 Comparison of efficiency scores: SFA and DEA

In the preceding chapter, we calculated the technical efficiency of the banks by using the DEA approach. In that chapter, we assessed the banks in terms of how good they are at using inputs to produce a given amount of output. This was done by using a production frontier model. In this chapter, we estimated the cost efficiency of the banks by using a cost frontier model. Two techniques of estimation were used: the DEA and the parametric approach. The DEA estimates of cost efficiency are done under two main assumptions: constant returns to scale and variable returns to scale. However, both assumptions are not made under the parametric approach. Table 5.8 shows this comparison. However, it must be noted that the DEA provides yearly efficiency scores unlike the fixed effects model and the SFA that computes efficiency scores for the whole period. To enable this comparison, we computed average efficiency scores for the DEA. This means that the CE_{crs} and CE_{vrs} are the averages of yearly efficiency scores for the whole period of seven years.

The average cost efficiency measures (CE_{vrs}) calculated under the DEA input – oriented approach are higher than the average cost efficiency measures estimated under the fixed effects model. The average CE_{vrs} is 75.39%, compared to 52.65% under CE3. However, under CE_{crs}, the DEA produces lower scores, 48.96%.

We should also take note of the fact that the two approaches work differently. Under the DEA, the calculation is based on two assumptions made with regard to returns to scale. If there are many production cost points that constitute the frontier, then many banks will be regarded as efficient. This is likely under variable returns to scale as compared to constant returns to scale. Under regression analysis, the technique is to find an appropriate line that can fit all the observed production costs. Given multiple outputs and input prices, the number of observed points located on or near to the line is fewer than the number of points found on the cost frontier.

Table 5.8: Comparison of efficiency scores by estimation techniques

	CE1	CE2	CE3	CE4	CE_{crs}	CE_{vrs}
Average efficiency	0.3203	0.3205	0.5265	0.4789	0.4896	0.7539
Standard deviation	0.1898	0.1830	0.1307	0.1336	0.2108	0.1912
Minimum	0.1580	0.1526	0.3261	0.2732	0.2329	0.3728
Maximum	1.0000	1.0000	0.9175	0.8895	1.0000	1.0000

Notes: a.CE1 = efficiency measures obtained by FE model using translog function
b.CE2 = efficiency measures obtained by FE model using log linear function
c.CE3 = efficiency measures obtained by SFA using translog function
d.CE4 = efficiency measures obtained by SFA using log linear function.
e.CE_{crs} and CE_{vrs} = efficiency measures obtained by DEA under constant and variable returns to scale assumption.

To determine whether the efficiency scores obtained by the stochastic frontier model and the DEA are related or independent, we performed the Spearman rank correlation test. The null hypothesis is that the efficiency scores obtained by both methods are independent. In general, the results show that there is a positive rank-order correlation between these efficiency estimates (Table 5.9). However, higher rank-order correlation coefficients exist when efficiency estimates are compared within the same framework. The rank-order correlations amongst the parametric efficiency scores (CE1, CE2, CE3 and CE4) are higher ranging from 0.65 to 0.98. The same happens to the correlation between CE_{crs} and CE_{vrs} (0.8061). The correlation coefficients between DEA efficiency scores and stochastic efficiency scores are relatively lower. For example, the correlation coefficients between CE_{vrs} and CE2 and between CE_{vrs} and CE4 are small and close to zero: 0.0291 and 0.0284 respectively. The null hypothesis between CE_{vrs} and all efficiency measures of the stochastic approach cannot be rejected. However, the correlation coefficients between CE_{crs} and all efficiency measures of the stochastic approach are always statistically significant i.e. rejecting the null hypothesis. The correlation coefficients between CE_{crs} and the stochastic efficiency measures range from 0.40 to 0.48.

The results of this study are partly similar to the findings of Isik and Hassan (2002). They find that the correlations between parametric and non-parametric efficiency scores are highly positive and significant in all cases. An interesting comparison is that they too find the correlation between the stochastic efficiency score and CE_{crs} (0.884) is higher than the correlation between the stochastic efficiency score and CE_{vrs} (0.611). However, they only use two approaches, the SFA and the DEA. Weill (2004) and Bauer et al. (1998) also used the Spearman rank-order correlation among the efficiency measures created by various techniques. Weill (2004) finds that the correlations between the stochastic efficiency score and the DEA scores are always negative, but only significant in the case of Germany. The correlation coefficients are near to zero or one. Bauer et al. (1998) found that the DEA and the parametric technique provide very weakly consistent rankings with each other. The average of rank-order correlations was only 0.098. Thus, they argued that there is no guarantee that both techniques can generally rank the banks in the same order. In addition, Cummins and Zi (1998) stated that, whilst the efficiency rankings are established among the parametric approach, the rankings are less consistent between the econometric and non-parametric. This implies that the choice of efficiency estimation is likely to have a significant effect on the findings of an efficiency study.

Table 5.9: Spearman rank correlation matrix of cost efficiency rankings obtained from different methods

	CE1[a]	CE2[b]	CE3[c]	CE4[d]	CE_{crs}[e]	CE_{vrs}[e]
CE1	1.000					
Prob > | t |						
CE2	0.9878	1.000				
Prob > | t |	(0.0000)					
CE3	0.6514	0.6558	1.0000			
Prob > | t |	(0.0000)	(0.0000)				
CE4	0.7503	0.7889	0.9115	1.0000		
Prob > | t |	(0.0000)	(0.0000)	(0.0000)			
CE_{crs}	0.4543	0.4078	0.4870	0.4096	1.0000	
Prob > | t |	(0.0090)	(0.0205)	(0.0047)	(0.0199)		
CE_{vrs}	0.1006	0.0291	0.1953	0.0284	0.8061	1.0000
Prob > | t |	(0.5838)	(0.8743)	(0.2841)	(0.8775)	(0.0000)	

Notes: a.CE1 = efficiency measures obtained by FE model using translog function
b.CE2 = efficiency measures obtained by FE model using log linear function
c.CE3 = efficiency measures obtained by SFA using translog function
d.CE4 = efficiency measures obtained by SFA using log linear function.
e.CE_{crs} and CE_{vrs} = efficiency measures obtained by DEA under constant and variable returns to scale assumption.

5.9 Efficiency and bank size

The relationship between cost efficiency and bank size is of great interest following the major bank consolidation in 2001. One of the main objectives of the merger policy is that it eventually leads to greater efficiency in the banking industry. The correlation results between cost efficiency scores and the bank size are shown in Table 5.10. The results in general indicate a significantly negative relationship between efficiency scores and bank size. The only positive relationship is shown by the CE_{vrs} (significant in both the whole sample and the domestic banks sample). The results provide little support for the bank consolidation policy. In another study that used Malaysian banking industry, Katib and Matthews (1999) also found negative relationship between the DEA efficiency score and bank size. This relationship was determined based on the pooled estimates of regression results. Their study covered the period 1989 to 1995. Similar results were found by Isik and Hassan (2002) who studied the Turkish banks. However, Weill (2004) finds a mixed result of the relationship between bank size and efficiency in European countries. The correlations between the stochastic efficiency scores and bank size are positive but not significant. Using the DEA scores, the correlation is positive and significant in Italy but not significant in France, Germany and Spain. He finds negative and significant correlation in Switzerland. Thus, we may say that the relationship between efficiency and bank size is not conclusive.

Table 5.10: Correlation results of efficiency measures with bank size.

Bank size	CE1	CE2	CE3	CE4	CE_{crs}	CE_{vrs}
	Whole sample					
Asset	-0.8317	-0.8519	-0.4582	-0.6151	-0.0539	0.2600
	(0.0000)	(0.0000)	(0.0083)	(0.0002)	(0.7695)	(0.0918)
	Domestic banks					
Assets	-0.7883	-0.8247	-0.2883	-0.5065	0.1818	0.7048
	(0.0000)	(0.0000)	(0.2050)	(0.0191)	(0.4302)	(0.0004)
	Foreign banks					
Assets	-0.8727	-0.8909	-0.6727	-0.7818	-0.3395	-0.0512
	(0.0005)	(0.0002)	(0.0233)	(0.0045)	(0.3071)	(0.8812)

Note: values in parentheses are P-values.

5.10 Efficiency and standard measures of performance

It is a well-known fact that efficiency must be closely related to other aspects of a firm's performance. If a firm is cost efficient, then it should also be profitable and negatively correlated with the other cost performance like ratio of operating cost to total asset or the ratio of operating cost to income. As stated by Bauer et al. (1998), there should be consistency conditions so that the estimated efficiency scores are valid and reliable. We proceed to analyse the correlations between efficiency scores and four standard measures of performance. These measures are the same as those used by Weill (2004). The four measures of performance are the average cost ratio (OCTA), defined as ratio of total operating cost to total assets; the cost to income ratio (OCINC), defined as the ratio of total operating cost to total income; the rate of returns on total assets (ROA), defined as the ratio of pre-tax profit to total assets; and the rate of returns on total equity (ROE), defined as the ratio of pre-tax profit to total equity. The relationships between cost efficiency scores and measures of cost performance are expected to be negative. There should be positive relationship between cost efficiency scores and indicators of profitability. The correlation results are shown in Table 5.11

Given the signs are as expected in most of the cases, we find that the correlations are not significant. The negative correlation between efficiency scores and cost performance (OCTA and OCINC) is only significant in one case (CE_{vrs}). The positive correlations between efficiency scores and profitability only occur in the case of the DEA efficiency scores. There is a significantly negative correlation between CE4 and ROE. This seems to suggest the existence of the quiet life hypothesis as noted by Weill (2004). The quiet life hypothesis refers to the situation where profitable firms do not bother to pay attention to operating efficiency. It is likely that if firms face less competitive pressure this may result in a lessened effort to maximise efficiency. Thus, there may also be a higher cost per unit of output produced in concentrated markets because of slight management.

Our results are similar to those of Weill (2004) but differ from those of Isik and Hassan (2002). Isik and Hassan (2002) find that the correlations between efficiency scores and OCTA are significantly negative. With regard to profitability, they find that the efficiency scores have positive correlations with ROA and ROE in all cases. Although these correlations are positive, they are not strong, ranging between 0.1 and 0.4.

Table 5.11: Efficiency scores and standard measures of performance

Efficiency Scores	OCTA	OCINC	ROA	ROE
CE1	-0.0475	0.0396	0.2273	-0.1070
Prob > | t |	(0.7963)	(0.8297)	(0.2109)	(0.5598)
CE2	-0.0119	0.1188	0.1826	-0.1613
Prob > | t |	(0.9484)	(0.5174)	(0.3173)	(0.3778)
CE3	-0.1886	-0.0070	-0.0251	-0.2592
Prob > | t |	(0.3014)	(0.9698)	(0.8915)	(0.1520)
CE4	-0.1067	0.1118	-0.0781	-0.4087
Prob > | t |	(0.5609)	(0.5424)	(0.6710)	(0.0202)
CE_{crs}	-0.1530	-0.0832	0.3882	0.0950
Prob > | t |	(0.4031)	(0.6506)	(0.0281)	(0.6051)
CE_{vrs}	-0.3408	-0.3499	0.4426	0.2889
Prob > | t |	(0.0563)	(0.0496)	(0.0112)	(0.1088)

Notes: OCTA = ratio of operating cost to total assets. OCINC = ratio of operating cost to total income. ROA = ratio of pre-tax profits to total assets. ROE = ratio of pre-tax profit to total equity.

5.11 Concluding remarks

This chapter aims to estimate the cost function and cost efficiency of the commercial banks using a stochastic frontier model. Two models were used: a fixed effects model and a stochastic frontier model. The study employs three input prices and three bank outputs. With regard to the cost function, two specifications are available: translog cost function and log linear cost function. The two differ in terms of additional parameters. The estimation techniques used are the fixed effects method and the maximum likelihood method. For the purpose of comparison, this study also uses the DEA technique to obtain the efficiency measures. This is to determine whether or not banking efficiency measures are subject to the technique employed.

Once we had estimated the cost function, we obtained the measures of cost efficiency. Under the stochastic frontier model, the mean efficiency score is 52.65% (using translog function). The mean efficiency score is 47.89% under the log linear specification. The mean difference between these two scores is statistically significant. This shows that efficiency scores are dependent upon the functional specification. However, under the fixed effects model, the average efficiency measures (using translog and log linear function) are lower. The scores are not statistically different from one another. By bank ownership, the performance between the domestic banks and foreign banks are not much different under the SFA, but are significantly different under the fixed effects model and the DEA. The commercial banks in Malaysia enjoy economies of scale. Under both models, the economies of scale parameters are less than 1.

In general, the average efficiency scores obtained by the stochastic approach are found to be lower than the mean scores calculated by the DEA approach. The CE_{vrs} the one calculated by the DEA under variable returns to scale emerges as the highest efficiency score, followed by CE3 (estimated under translog specification), CE_{crs} (calculated by the DEA under constant returns to scale) and CE4 (estimated under log linear specification). The Spearman rank-order correlation coefficients between the DEA efficiency measures and the stochastic efficiency are always positive, but only statistically significant between CE_{crs} and CE1, CE2, CE3 and CE4.

Efficiency is found to be unrelated to bank size. The correlation results show that there is a significantly negative relationship between these two variables except in the case of CE_{vrs}. This finding provides little support for the recent bank consolidation of domestic banks. We also find that the efficiency scores are not closely related to standard measures of performance (cost performance and profitability) except in the case of the DEA estimates.

Chapter

Six

CHAPTER SIX

Competitive behaviour of commercial banks: A Panzar-Rosse methodology

6.0 Introduction

The competitive environment plays a crucial role in shaping banks' behaviour. Any absence of competitive pressure may induce the banks concerned not to utilise their capacity fully. Potential markets might be left untapped and potential business adventures unexplored. Ultimately, banks will tend to remain conservative under such circumstances and continue to rely on the current marketing strategies. However, this type of behaviour might change in the face of competitive pressures such as market-driven interest rates and the presence of foreign ownership, which force the banks to be resilient and efficient. The threat of potential take-overs and mergers may encourage the bank managers to search continually for new strategies, new business potentials and investment opportunities. In addition, the presence of foreign competitors may drive the domestic banks to set higher standards for doing business. As found in the preceding chapters that deal with efficiency, foreign banks show superior performance in terms of both technical and cost efficiency.

Since the late 1970s, a mild and gradual financial liberalisation has been adopted by the Malaysian monetary authority in managing the banking industry. As discussed in Chapter Three, this liberalisation was brought about by relaxing the fixed interest rate in 1978. Since then, the banks have been allowed to set their own deposit and lending rates, but their actions are always monitored by the Central Bank (Bank Negara Malaysia). At the end of the 1980s, regulatory reform occurred with the passing of the Banking and Financial Act (BAFIA 1989). This new Act is crucial to ensure prudential regulation governing most of the financial institutions in Malaysia. Since the early 1990s, the commercial banks have been allowed to invest in certain blue-chip companies as well as in the traditional business sectors. Based on the various forms of financial liberalisation, it is believed that these measures might have influenced the competitive environment of the commercial banks. Have banks become more competitive as a result? Alternatively, have they become less competitive or become collusive?

It must be emphasised that the financial liberalisation in the Malaysian financial system is limited. While some may expect that the banking sector should be widely open to foreign banks, this is not the case in Malaysia. Since the 1960s, the Malaysian authorities have attempted to reduce the presence of foreign banks by encouraging the establishment of local banks. The growth of foreign banks is limited by forbidding the creation of new banks and bank branches. In addition, the pursuit of preferential policies such as the New Economic Policy (1971-1990) and the National Development Plan (1991-2000) places additional constraints on the banks' lending. As discussed in

Chapter Three, the banks are still required to fulfil social obligations by allocating a proportion of their loans to certain priority sectors like the Bumiputra community, small and medium industries, and the housing sectors. The banks are also expected to purchase substantial amounts of government bonds and securities.

This chapter aims to evaluate the competitive structure of the Malaysian banking industry and to measure the present level of market contestability created by the process of financial liberalisation and deregulation. To achieve these two objectives, we use the Panzar-Rosse methodology, named after Panzar and Rosse (1987). This methodology requires, firstly, that banks are in long-run equilibrium, whereby any changes in input prices have zero impact on the rates of return. This can be shown by conducting an equilibrium test. Secondly, the H-statistic will be constructed to indicate the level of competition in the banking industry. To set up this statistic, we need to estimate a reduced form revenue equation. Given the estimated coefficients of the input prices, the H-statistic is obtained. $H = 0$ indicates that the banks behave like a monopoly. If $H = 1$, then the banks are in a perfectly competitive market. If H lies between 0 and 1, then the banks are in monopolistic competition. Given the various measures of financial liberalisation and deregulation, this study expects that the Malaysian banking industry will be operating under monopolistic competition.

This chapter is organised as follows. In the next section, we discuss the theoretical framework by emphasising the various aspects of financial liberalisation and deregulation and how this policy could have affected the competitive environment of the banking sector. In Section 6.2, we put forward two main approaches for the analysis of this competitive behaviour. These methods are structural models, which are comprised of the structure-conduct-performance (SCP) paradigm and the efficient-structure hypothesis, and non-structural models, represented by the conjectural variation model and the Panzar-Rosse model (Panzar and Rosse, 1987). We then proceed by setting up our empirical models, which need to undergo two critical tests. The first is the equilibrium test that determines whether the banks are in long-run equilibrium or not. A critical assumption in the Panzar-Rosse model is that the banks are in long-run equilibrium. The second is the competition test, in which H-statistic will be estimated and tested to determine its value. In Section 6.4, we present the descriptive statistics of the key variables. Empirical results of both the equilibrium and the competition test are presented in the subsequent section. This is followed by the estimation results for both domestic and foreign banks. Section 6.6 closes by making important remarks about the chapter.

6.1 Theoretical framework: market structure, financial liberalisation and competition

The presence of more than one bank is likely to make competition inevitable. Under perfect competition or monopolistic competition, banks in the long run are expected to experience normal profit only, given free entry and exit. The popular objectives of having a larger market share and a higher level of profits might be compromised due to the existence of closed rivalries. If a large number of banks exist in the market, then it is likely that the market share will be smaller and the level of profit will be lower (assuming no collusion amongst the banks). On the other hand, the rivalry and competition may require the banks to be more efficient, innovative and resilient in order to pursue their current objectives. Underperforming banks may encounter the risk of going bankrupt or being taken over by more efficient and well-performing banks. In this way, competition is widely seen and accepted as a positive phenomenon.

Claessens and Laeven (2003) argue that there are several reasons why competition should exist in the financial sector. Firstly, the degree of competition can affect the productive efficiency of financial services. Given the large number of players in the market, financial institutions must make the best of available resources, especially borrowed or purchased funds. Competition also demands a greater quality of financial products, as close substitutes are available. Furthermore, it requires bankers to be more innovative in offering new products. Following the Asian financial crisis in the late 1990s, some theorists have argued that excessive competition amongst the financial institutions in offering loans has been acknowledged as one of the factors behind the crisis (Noy, 2004; Ariff and Khalid, 2000; Vives, 2001 and Jomo, 1998). In addition, the degree of competition can influence the access of potential bank customers to banking facilities. This in turn helps to facilitate the mobilisation of funds and the financing of required investments in the country.

Liberalisation refers to the removal of government interference in financial markets, capital markets and international trade (Stiglitz, 2004). In general, it aims to promote competition. According to Yildirim (2003), there are at least three ways in which financial liberalisation may take place. The first is by liberalising the capital account. This requires free movement of capital into and out of the country and no restrictions on foreign exchange transactions. A country's currency is allowed to be freely determined by the market forces. The second way is through the deregulation of financial institutions. This requires a reduction of government intervention in the financial sector. Government controls may take place in the form of fixing the deposit and lending rates, credit allocation, and relevant legal requirements. Through deregulation, the banks are allowed to determine their own deposit and lending rates. The third method is the internationalisation of financial services. This is implemented by giving fair treatment to both domestic and foreign banks. In other words, the foreign banks are not discriminated against in favour of domestic banks. Internationalisation of financial services also requires the removal of cross border business activities.

As financial liberalisation exposes the banks to greater competition, how do they respond? Firstly, the banks can no longer rely on their traditional market. New markets must be sought and the banks need to go beyond their current boundaries. New sources of revenue must be sought. As competition increases, the interest rates are likely to go down and this puts further pressure on the profit levels, i.e. reducing the profit margins. The high level of dependence on interest-based products must be gradually reduced by exploring fee-based products. If available, excess capacity must be eliminated via mergers or the closure of bank branches.

Fry (1997, pp. 755), in favour of financial liberalisation, puts forward his arguments against financial repression. According to him, interest rate ceilings distort the economy in several ways. It is argued that low interest rates discourage current saving in favour of current consumption. Lower saving rates might place investment (future consumption) at risk. In addition, because of lower lending rates, the banks may prefer relatively low-yielding direct investment. The bank borrowers, on the other hand, may opt for capital-intensive rather than labour-intensive projects. This technique of production is, perhaps, inappropriate when the economy is full of unemployed labourers.

There are some studies that advise caution on the role of financial liberalisation. According to Stiglitz (2004, p 59), the IMF agrees that it has pushed the liberalisation agenda too drastically. Liberalisation in both capital and financial markets was acknowledged as a contributing factor to the global financial crises of the 1990s. It can also wreak havoc on a small emerging country. In addition, Stiglitz (1994, quoted in Fry (1997) believes that financial liberalisation may render the financial markets vulnerable to market failures. Vives (2001) states that given financial liberalisation, competition in banking can be excessive, which can contribute to crises. With regard to banking crises, Noy (2004) asserts that there are two contending hypotheses: 'lax supervision' and monopoly power. Under the 'lax supervison' hypothesis, financial liberalisation brings about changes in the rules of the game, which eventually drive the banks to take excessive risks in order to increase their profit. In addition, the changes in the rules may place the regulators and the supervisors in a new environment surrounded by new rules. This may jeopardise and reduce the effectiveness of the supervision. Therefore, it is argued that financial liberalisation in the absence of efficient supervision and prudential regulation will probably contribute towards a banking crisis. Under the monopoly power hypothesis, financial liberalisation leads to intense competition. Increased competition may shrink the profit margins and place some banks into situations of loss. It is likely that the loss-making banks will go bankrupt or face the risk of being taken over by the efficient ones. In addition, the increased competition may force the banks to take risky investments. Eventually, systemic problems and distress in the banking sector are inevitable.

6.2 Approaches to competitive behaviour

There are two broad approaches in analysing the competitive behaviour of the banks: structural models and non-structural models. The structural models are popularly represented by the structure-conduct-performance (SCP) paradigm and the efficient structure (ES) hypothesis. The non-structural models are shown by the conjectural variation model and the Panzar-Rosse (1987) methodology.

6.2.1 Structural models

Two widely used models are the SCP paradigm and the ES hypothesis. The difference between these two lies in the role of market power indicated by industrial concentration and efficiency (initially represented by the market share, but lately by its own measure).

6.2.1.1 The Structure-Conduct-Performance paradigm

Under the SCP paradigm, the market structure is comprised of organisational characteristics of a market that seem to impose a strategic influence on the ways in which firms behave within the market (Bain, 1968). In relation to this, Howe (1978) added that these organisational characteristics have several dimensions as follows:

- seller concentration (leading to oligopoly and monopoly)
- buyer concentration (leading to oligopsony and monopsony)
- product substitution (either homogeneity or differentiation)
- entry conditions (either easy or blockaded entry) .

This market structure is then set to have a substantial influence on the market conduct, which has its own dimensions. These dimensions are:

- price policies that involve its objectives, bases, tactics and co-ordination

- product and sale policies such as product development, product design etc.

The conduct eventually generates market performance. Bain (1968) stated that by pursuing whatever lines of conduct they choose, there will be a composite of end results at which firms arrive. As a composite, market performance has several indicators, including profitability, such as returns on assets (ROA), returns on capital (ROC) and returns on equity (ROE), efficiency, such as technical efficiency, product quality, design and variety.

Lloyds-Williams and Molyneux (1994) stated that SCP studies fall into two categories, depending on which measure of performance is taken into account. The first is a measure of the prices of certain banking products and services. This choice is very problematic because for certain products, such as fee-based products, price is unlikely to be measured accurately. The second is a measure of profit such as returns on assets, returns on capital, returns on equity, price-cost margin and net interest margin. These measures are more suitable since they are directly available from the annual report published by the banks.

Since a higher market concentration is expected to lead to a lower degree of competition but eventually higher profit, then a profit function under the SCP paradigm can be expressed as follows:

$$\pi_i = \pi\ (CR, z_i) + \varepsilon_i \qquad \text{Equation 6.1.}$$

where π_i is the variable profit of bank *i*. CR is a concentration ratio, say 3-firm concentration ratio or Herfindahl index. Z is a set of market variables that may affect bank *i* performance. ε_i is the normal random error of bank *i*. .

This SCP model postulates that in a concentrated industry, all firms can increase the prices of their products. Given barriers to entry, profits can be protected. Therefore, it is likely that some firms will pursue anti-competitive conduct in this kind of market where only a small number of large firms prevail.

6.2.1.2 Efficient structure hypothesis

As discussed in Chapter Two, the SCP hypothesis has been questioned by many. The opponents of the SCP argue that the positive correlation between concentration and profit might reflect the superior efficiency among firms rather than the use of market power. In other words, it is the efficiency that permits larger size and market share, and subsequently generates greater profit. The key factor between market structure and firm performance is efficiency. The ES hypothesis proposes that given the presence of efficiency, a firm can maximise potential profit by bringing down the price and expanding firm size, hence gaining market share.

A profit function under the ES hypothesis is written as:

$$\pi_i = \pi\ (CR, MS_i, z_i) + \varepsilon_i \qquad \text{Equation 6.2.}$$

where π_i is the variable profit of bank *i*. CR is a concentration ratio, say 3-firm concentration ratio. MS_i is the market share (proxy to bank *i* 's efficiency). Z is a set of market variables that may affect the performance of bank *i*. ε_i is the normal random error of bank *i*.

With a direct estimate of efficiency, the profit function is re-written as:

$$\pi_i = \pi\ (CR, EFF_i, z_i) + \varepsilon_i \qquad \text{Equation 6.3.}$$

where π_i is the variable profit of bank *i*. CR is a concentration ratio, say 3-firm concentration ratio. EFF_i is bank *i's* efficiency measure. Z is a set of market variables that may affect bank *i*'s performance. ε_i is the normal random error of bank *i*.

6.2.2 Non-structural models

An issue that remains debatable is the relationship between market structure and firms' conduct or between market concentration and competition. As an alternative to the structural models, the non-structural models were introduced to offer new insights with regards to this relationship. According to Bikker and Haaf (2002), three non-structural models have been developed, namely the

Iwata model (Iwata,1974), the Bresnahan model (Bresnahan,1982) and the Panzar-Rosse model (Panzar and Rosse, 1987). Toolsema (2002), on the other hand, considers both the Iwata and the Bresnahan model as conjectural-variation models. All these models are considered as New Empirical Industrial Organisation approaches. The non-structural models basically measure the degree of competition. However, while analysing the competitive behaviour of the banks, these models ignore information about the structure of the market such as market power and efficiency.

6.2.2.1 Conjectural-variation model

This method requires a system of equations that consist of inverse market demand and suppy equations. One of the parameters of the supply function represents the conjectural variation, i.e. the anticipated response of the bank's rivals to an output change. Since this method uses a system of equations, an identification problem might arise. However, the work of Bresnahan (1982) and Lau (1982) has solved this problem. Bresnahan (1982) and Lau (1982) used an alternative parameterisation given by the conjectural variation elasticity besides concentrating on aggregate data in their articles. Since this study uses a different non-structural model, their works are not discussed in detail.

The use of the conjectural-variation model is limited. To the best of our knowledge, the most recent works that use this method are Uchida and

Tsutsui (2005) and Toolsema (2002). Uchida and Tsutsui (2005) investigate whether competition in the Japanese banking industry has improved for the period 1974 to 2000. They find that competition improved in the 1970s and the first part of 1980s. They also reveal that the Cournot oligopoly cannot be rejected. Meanwhile, Toolsema (2002) looks at the competitiveness of the Dutch consumer credit market. Using aggregate data, she finds that there is no evidence of market power.

6.2.2.2 Panzar-Rosse (1987) method

The Panzar-Rosse model employs simple models to determine the type of market in question. Whether the market is oligopolistic, competitive, or monopolistic will be determined by a competition test. This method estimates a reduced-form revenue equation by using bank-specific data. After estimating this equation, the so-called H statistics are constructed to serve as a measure of the competitive behaviour of the banks. Details of this method are explored in the next section.

6.3 Empirical models

The focus of this chapter is to evaluate the competitive structure of the commercial banks in Malaysia and to measure the existing level of market contestability generated by the process of financial liberalisation and deregulation. To arrive at these objectives, we use the Panzar-Rosse methodology (PR) as introduced by Panzar and Rosse, 1987. The PR method is a technique that estimates the overall competitive conduct in a given banking system. This study follows the steps taken by previous studies in terms of using relevant functions and variables. In the selection of bank inputs, we choose the intermediation approach so as to be consistent with the choice taken in the preceeding chapters (Chapters Four and Five). As discussed above, this approach of bank modelling is closedly related to the the role of financial institutions. Thus, we use three bank inputs: labour, fixed assets, and total deposits. Their respective prices are the price of labour, the price of capital and the price of deposits.

Under the PR method, the competitive analysis begins by estimating the reduced form revenue equation. While the quantity of output and its price are not observable, total revenue is observable. The premise that the PR method wants to show is that the banks will use different pricing strategies as a response to any changes in input costs. However, the market structure in which the banks operate may influence the extent of these responses.

To arrive at an equilibrium output and an equilibrium number of banks, profits are to be maximised both at industry and firm level. Profit is normally defined as the excess of total revenue over total cost. As stated before, there exist I banks (i = 1, ..., I) that produce a vector of m outputs $y_i = (y_1, \ldots, y_m)$, which they sell at prices $p_i = (p_1, \ldots, p_m)$ using a vector of k inputs $x_i = (x_1, \ldots, x_k)$ for which they pay prices $w_j = (w_1, \ldots, w_k)$. The total profit for bank i is written as:

$$\Pi_i = R_i - C_i. \qquad \text{Equation 6.4}$$

Since $R_i = p_i{*}y_i$ and $C_i = w_i{*}x_i$, then

$$\pi_i = \sum_{i=1}^{m} p_i y_i - \sum_{j=1}^{k} w_j x_j$$

Let $R_i = R(y_i, z_i)$ represent the revenue of bank *i*. y_i is a vector of bank *i* outputs and z_i is a vector of exogenous variables that shift the bank revenue function. The cost function is given by $C_i = C(y_i, w_i, g_i)$. w_i is a vector of k input prices that are exogenous to the bank. g_i is a vector of exogenous variables that shift the cost function. The profit maximisation of bank *i* requires that its marginal revenue (R'_i) is equal to its marginal cost (C'_i). Thus,

$$R'_i (y_i, z_i) - C'_i (y_i, w_i, g_i) = 0. \qquad \text{Equation 6.5}$$

It also implies that in equilibrium, the following condition holds in the industry for all firms.

$$R^*_i (y^*, z^*) - C^*_i (y^*, w^*, g^*) = 0. \qquad \text{Equation 6.6}$$

where * represents equilibrium values.

Panzar and Rosse (1987) define a measure of competition, the H-statistic, as the sum of the elasticities of the reduced–form revenues with respect to all input prices.

$$H_i = \sum_{j=1}^{k} \frac{\delta R_i{*}}{\delta w_{ji}} \frac{w_{ji}}{R_i{*}} \qquad \text{Equation 6.7}$$

The H statistic ranges from zero or less than zero to one. It measures the responsiveness of total revenue to a proportional increase in all input prices. From an economic point of view, this statistic can be interpreted as follows. Let us assume that the banking industry operates under a monopoly. Under this market, total revenue will respond in the opposite direction to a change in input prices. A ten percent increase in input prices will bring about a ten percent increase in both marginal and average costs, assuming that the cost function is homogenous of degree one in input prices. Increases in marginal and average costs lead to reductions in equilibrium output and revenue. Thus, the H statistic is less than or equal to zero. Panzar and Rosse (1987) also show that the H statistic becomes nonpositive when the market structure is a perfectly colluding oligopoly or a conjectural variations short-run monopoly.

If the H statistic is equal to one, then the market is characterised as perfectly competitive in long-run equilibrium. Given a proportional increase in input prices, both marginal and average costs will increase by the same proportion without altering the equilibrium output of the banks. The increased costs may place certain firms in losses and market adjustment takes place. Loss-making firms might be driven out of the market and inefficient firms might be taken over by the efficient ones. Market adjustment will witness a reduction in the number of firms in the industry. The remaining firms will experience increases in demand that will lead to a rise in the price of output and revenue by the same amount as costs.

The H statistic might also lie between zero and one. This indicates that revenues will increase less than proportionally to changes in input prices. In this case, the industry is described as a monopolistic competition whereby the market is free from entry and exit barriers and the products of other banks are not perfect substitutes.

6.3.1 Reduced-form revenue equation.

Two measures of revenue were used: total revenue (TR) and total interest revenue (INT). Both are normally calculated as ratios to total assets. TRA is the ratio of total revenue to total assets and INTA is the ratio of interest revenue to total assets. Besides input prices, several control variables at the individual bank level are also included. This is to account for the effect of bank size, risk factors and asset quality. Claessens and Laeven (2004, 2003), for example, use three control variables, namely the ratio of equity to total assets, the ratio of net loans to total assets, and total assets. Besides the above control variables, Yildirim (2003) also included interbank deposits. Meanwhile, Bikker and Haaf (2002) employ control variables such as the ratio of non-performing loans to total loans and used more than one variable for risk. Chun and Kim (2004) use the number of bank branches to account for bank size.

The three-input empirical model for the competition structural test is written as follows.

$$\ln(TRA_{it}) = \alpha + \beta_1 \ln(W_{1,it}) + \beta_2 \ln(W_{2,it}) + \beta_3 \ln(W_{3,it}) + \theta_1 \ln(NB_{it}) + \theta_2 \ln(EQTA_{it}) + \varepsilon_{it} \qquad \text{Equation 6.8}$$

where TRA is the ratio of gross operating revenue to total assets (proxy for output price), $W_{1,it}$ is the ratio of personnel expenses to total labour (proxy for input price of labour), $W_{2,it}$ is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), $W_{3,it}$ is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB_{it} is the number of bank branches (to control for potential size effect), $EQTA_{it}$ is the ratio of total equity to total assets (to control for bank-specific risk) and ε_{it} is the random error. The subscripts *i* and *t* denote bank *i* and year *t* respectively.

The other measure of revenue is interest revenue (INT). The main difference between total revenue and interest revenue is non-interest revenue. Non-interest revenues are generated by fee-based financial services such as foreign exchange transactions, financial advice etc.

$$\ln(INTA_{it}) = \alpha + \beta_1 \ln(W_{1,it}) + \beta_2 \ln(W_{2,it}) + \beta_3 \ln(W_{3,it}) + \theta_1 \ln(NB_{it}) + \theta_2 \ln(OY_{it}) + \theta_3 \ln(EQTA_{it}) + \varepsilon_{it} \qquad \text{Equation 6.9}$$

where $INTA_{it}$ is the ratio of interest revenue to total assets (proxy for output price). $W_{1,it}$ is the ratio of personnel expenses to total labour (proxy for input price of labour), $W_{2,it}$ is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), $W_{3,it}$ is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB_{it} is the number of bank branches (to control for potential size effects), OY_{it} is other operating income (to control for non-interest based services), $EQTA_{it}$ is the ratio of total equity to total assets (to control for bank-specific risk) and ε_{it} is the random error. The subscripts *i* and *t* denote bank *i* and year *t* respectively.

Under the PR method, the H statistic is equal to the sum of the elasticities of the total revenue (interest revenue) with respect to the three input prices:

$$H = \beta_1 + \beta_2 + \beta_3 \qquad \text{Equation 6.10}$$

The testable hypothesis that the industry is under monopolistic competition is given by

$$0 < H = \beta_1 + \beta_2 + \beta_3 < 1 \qquad \text{Equation 6.11}$$

where $H \leq 0$ is monopoly and $H=1$ is perfect competition. The Wald test is used here to determine the value of the H statistic.

6.3.2 Equilibrium test

The critical aspect of the PR methodology is that the changes in input prices have no impact on the returns on assets. To conduct the equilibrium test, we use the rate of returns on total assets, i.e. ratio of pre tax profit to total assets, as the dependent variable. The right hand side variables remain the same as those used under the competition test (Equations 6.8 and 6.9).

$$\ln(ROA_{it}) = \alpha + \beta_1 \ln(W_{1it}) + \beta_2 \ln(W_{2it}) + \beta_3 \ln(W_{3it}) + \theta_1 \ln(NB_{it}) + \theta_2 \ln(OY_{it}) + \theta_3 \ln(EQTA_{it}) + \varepsilon_{it} \qquad \text{Equation 6.12}$$

where ROA_{it} is the ratio of pre-tax profits on total assets of bank *i* (proxy for rate of returns on total assets). $W_{1,it}$ is the ratio of personnel expenses to total labour (proxy for input price of labour), $W_{2,it}$ is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), $W_{3,it}$ is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB_{it} is the number of bank branches (to control for potential size effects), OY_{it} is other operating income (to control for non-interest based services), $EQTA_{it}$ is the ratio of total equity to total assets (to control for bank-specific risk) and ε_{it} is the random error. The subscripts *i* and *t* denote bank *i* and year *t* respectively.

The null hypothesis that the banks operate in long-run equilibrium is given by:

$$H = \beta_1 + \beta_2 + \beta_2 = 0. \qquad \text{Equation 6.13.}$$

H=0 implies that long-run equilibrium exists. $H \neq 0$ implies disequilibrium. Again, we use the Wald test to decide whether to accept or reject this hypothesis.

6.3.3 Estimation technique

The competition and equilibrium test can be performed using a number of techniques. The most popular techniques are pooled ordinary least squares (OLS), the fixed effects method, and the random effects method. Claessens and Laeven (2004, 2003) used both pooled OLS and the fixed effects method. Meanwhile, Chun and Kim (2004) used both the fixed and random effects methods. On the other hand, Yildirim (2003) used all three techniques; pooled OLS, fixed and random effects. In this study, we used pooled OLS and the fixed effects model.

6.4 Descriptive statistics

This study estimated two equations, a reduced-form revenue equation and a returns on assets equation. For the first equation, we used two dependent variables: total revenue (Equation 6.8) and total interest revenue (Equation 6.9). The independent variables comprised of input prices (price of labour, price of capital and price of deposits) and a set of control variables. The control variables were selected to account for the effect of bank size, risk aspect and asset quality. The control variables are the number of bank branches (NB), the ratio of equity to total assets (EQTA) and non-interest income (OY). For the second equation (Equation 6.12), the dependent variable is the ratio of pre-tax profit to total assets (ROA). The independent variables are the same as in Equations 6.8 and 6.9. The descriptive statistics of all variables are presented in Table 6.1. All figures are in RM millions except the number of bank branches. Figures 6.1 and 6.2 show the relationship between total revenue and input prices and between total revenue and control variables, respectively.

The same considerations as used in the previous chapters were taken into account in determing the final sample of commercial banks. This includes factors like bank consolidation and data unavailability. Total bank-year observation encompasses 193 banks and covers the period from 1994 until 2000. The commercial banks after 2000 were excluded, mainly due to major consolidation that involved domestic banks.

Table 6.1: Descriptive statistics of the data (1994 – 2000)

Variables	No of banks	Mean (RM Millions)	Standard Deviation (RM Millions	Minimum (RM Millions	Maximum (RM Millions
1994					
TR	20	459.16	554.46	14.20	2205.50
INT	20	404.61	510.17	12.20	2070.60
ROA	20	0.01	0.01	0.00	0.03
PL	20	0.0302	0.0074	0.0106	0.0429
PK	20	0.7929	0.4434	0.1971	1.6755
PD	20	0.0447	0.0198	0.0141	0.0898
NB	20	39.35	42.39	1.00	169.00
EQTA	20	0.0811	0.0197	0.0464	0.1127
OY	20	54.56	57.62	2.00	210.00
1996					
TR	32	752.37	960.69	56.60	4908.20
INT	32	673.32	869.90	50.20	4455.80
ROA	32	0.02	0.01	0.01	0.04
PL	32	0.0406	0.0169	0.0177	0.1009
PK	32	0.8552	0.5290	0.2494	2.5556
PD	32	0.0751	0.0540	0.0414	0.3170
NB	32	44.12	58.78	1.00	273.00
EQTA	32	0.0903	0.0252	0.0584	0.1814
OY	32	79.06	98.42	4.00	452.40
1998					
TR	30	1335.15	1794.77	74.00	9497.00
INT	30	1217.76	1648.82	67.40	8771.60
ROA	30	0.01	0.03	-0.07	0.08
PL	30	0.0444	0.0234	0.0249	0.1155
PK	30	0.9420	0.6311	0.3599	3.1904
PD	30	0.1066	0.0500	0.0475	0.2686
NB	30	49.83	64.60	1.00	289.00
EQTA	30	0.0961	0.0440	0.0148	0.2103
OY	30	117.39	159.67	-4.60	725.40
2000					
TR	20	1391.14	1860.79	39.50	6890.90
INT	20	1240.90	1714.37	34.90	6508.60
ROA	20	0.02	0.02	-0.05	0.06
PL	20	0.0546	0.0260	0.0268	0.1191
PK	20	1.3658	1.1084	0.3589	3.9630
PD	20	0.0585	0.0621	0.0241	0.3075
NB	20	49.10	71.37	1.00	294.00
EQTA	20	0.0997	0.0611	-0.0469	0.2204
OY	20	150.24	186.56	4.60	800.70
1994-2000					
TR	193	910.42	1299.59	14.20	9497.00
INT	193	815.10	1183.36	12.20	8771.60
ROA	193	0.01	0.02	-0.09	0.08
PL	193	0.0442	0.0277	0.0099	0.2773
PK	193	0.9923	0.7429	0.1663	4.5882
PD	193	0.0734	0.0531	0.0079	0.3901
NB	193	45.03	59.21	1.00	294.00
EQTA	193	0.0934	0.0396	-0.0469	0.2218
OY	193	95.32	130.68	-4.6	800.70

Notes: TR = total bank revenue. RY = total interest revenue. ROA = rate of returns on total assets. PL = average price of labour (total personnel expenses divided by the number of bank staff). PK = average price of capital (non-interest expenses divided by fixed assets). PD = average price of deposit (total interest expenses divided by total deposits. NB = number of bank branches. EQTA = ratio of total equity to total assets. OY = other income (non-interest income)

Sources: ABM (1994, 1996, 1998, 2000) and BANKSCOPE

Figure 6.1: Relationship between total revenue and input prices (1994-2000)

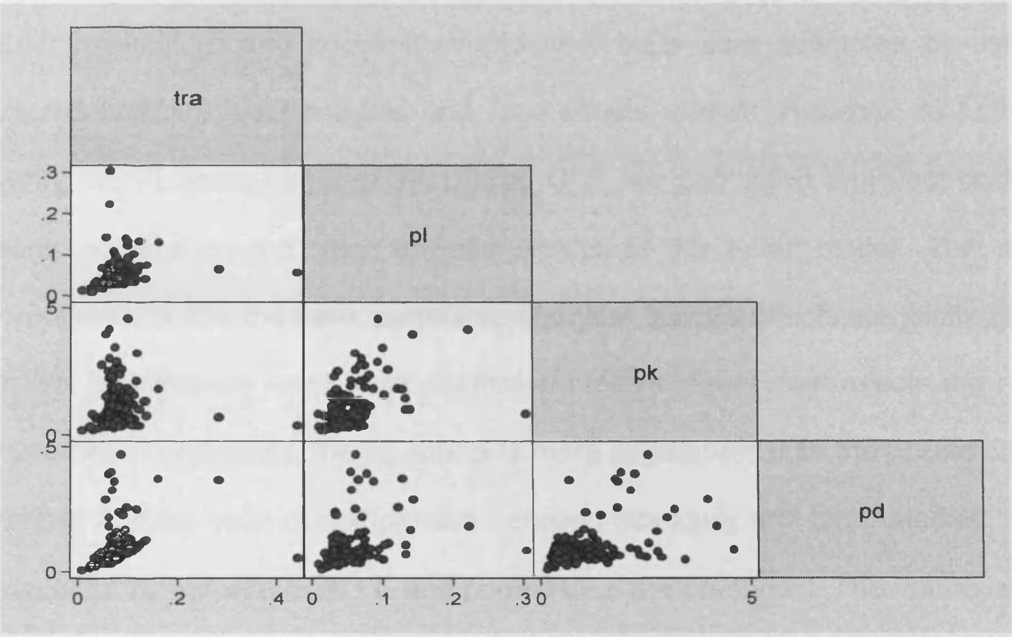
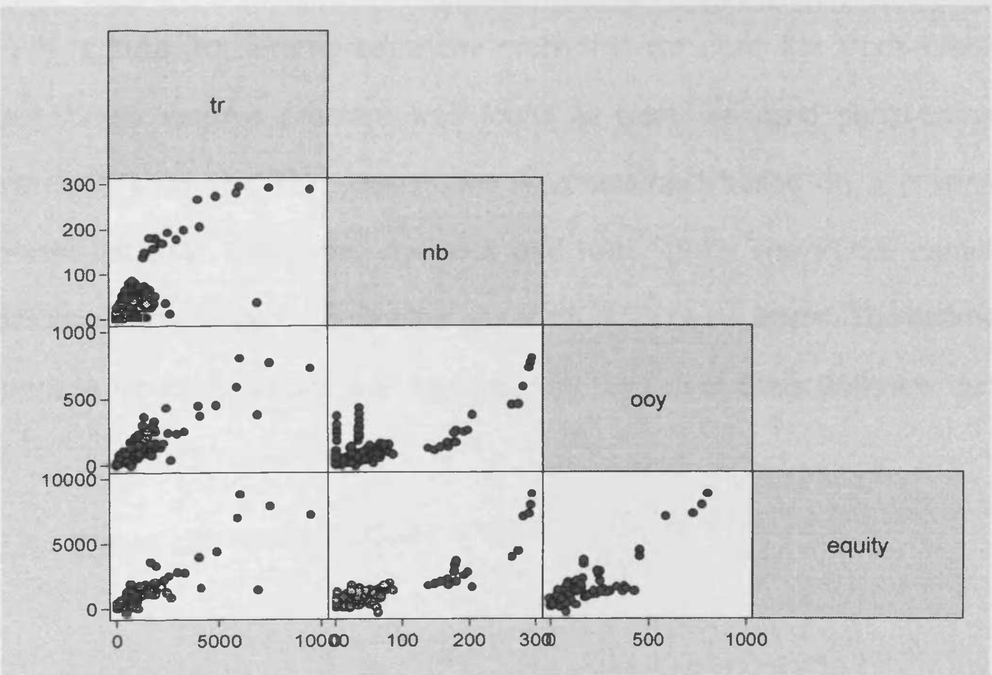


Figure 6.2: Relationship between total revenue and control variables (1994-2000)



6.5 Empirical results

Both equilibrium and competition structural tests were estimated by using pooled ordinary least squares and fixed effects models. However, to further verify the FE model against the pooled OLS, we performed an F-test on the bank ownership and year specific effects in the latter model. The null hypothesis is that the bank ownership and year specific effects are jointly zero at 5% significance level in all estimated models. The F-test rejects the null hypothesis. Therefore, the FE model is more appropriate than the pooled OLS model. For the sake of comparison between this study and other studies, the results obtained with both FE and pooled OLS are presented. The results also show how the H-statistic changes as we move from one specification to another.

With regards to heteroskedasticity problems, we used the Cook-Weisberg test. Since such a problem was found to exist, we used panel-corrected standard error (PCSE). The PCSE is constructed based on a covariance matrix estimate introduced by Beck and Katz, 1995. The PCSE formula is designed to take into account the panel structure of the errors. The estimation process for all equations was facilitated by the use of Stata Software Version 8.

6.5.1 Equilibrium test

This test is to ensure that the banks are in long-run equilibrium, whereby any changes in the input prices have no effect on the rate of returns. The zero effect on the rate of returns is shown by the H-statistic that equal the sum of input price elasticities. Table 6.2 shows the estimation results for Equation 6.12. The dependent variable is the rate of returns on total assets (ROA)[34]. The estimated H-statistics are 0.0252 (under pooled OLS) and 0.2130 (under the fixed effects model). Using the Wald test, the null hypothesis that H=0 cannot be statistically rejected. The results show that the banks are indeed in long-run equilibrium.

[34] Besides returns on total assets, we also used rate of returns on total equity as suggested by Molyneux et al., 1996. Since the results are the same, we report the estimates on ROA only.

Table 6.2: Estimation results for ROA

(all banks, 1994-2000)

Independent variables	Pooled OLS		Fixed Effects Model	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.1270	0.1437	0.2092	0.1362
LPK	0.0466	0.0616	0.0759	0.1111
LPD	-0.1483	0.1131	-0.0722	0.0959
LNB	-0.1876	0.0403*	-1.0270	0.2416*
LOY	0.2163	0.0263*	0.3997	0.0970*
LEQTA	0.5004	0.0950*	0.4414	0.1276**
constant	-3.2613	0.4676*	-1.1930	0.8594
R^2	0.3991		0.2566	
H-statistic	0.0252		0.2130	
F value on Wald test for H=0	0.03		1.52	
P value	0.8654		0.2196	
No of observation	175		175	

Notes: ROA is the ratio of pre-tax profit to total assets, PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effect), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.
*,**,*** represent significance levels at 1%, 5% and 10% respectively.

6.5.2 Competition test

The Competition test was used to determine the degree of competitiveness in the banking sector. This was done by estimating the reduced-form revenue equations (Equations 6.8 and 6.9). Two measures of revenue were used:the ratio of total revenue to total assets (TRA) and the ratio of interest revenue to total assets (INTA). The revenue model with TRA as the dependent variable will be referred to as Model 1 and the model with INTA as the dependent variable will be referred to as Model 2. When Model 1 is estimated using pooled OLS, it will be referred to as Model 1a. Model 1b is referred to when Model 1 is estimated using the fixed effects model. The same applies to Model 2. The estimation results for both measures of revenue are shown in Tables 6.3 and 6.4 respectively.

The estimated coefficients of input prices (β_1, β_2 and β_3) are always positive and significant except for the price of capital (LPK). The coefficient of LPK (β_2) is positive but not significant under Model 2b. This implies that any changes in the price of capital have no significant impact on the banks' revenue. The comparison between LPL and LPD shows interesting results. The coefficient of LPD (β_3) is always greater than the coefficient of LPL (β_1) and the rest of the coefficients in all models. β_3 ranges from 0.3188 to 0.5615 compared to β_1, which ranges between 0.1984 and 0.3029. Taking into account other variables, the results show that LPD has the largest impact on the banks' revenue, followed by LPL. Given β_3 = 0.5397 (Model 1b), a one percent increase in the price of deposits will cause approximately a 0.5% increase in the total revenue (proxy to output prices), ceteris paribus. In all models, β_3 and β_1 are the largest coefficients. This means that the price of deposits provides the largest contribution to the explanation of bank revenue, followed by the price of labour.

The coefficient of determination (R^2) ranges from 0.3892 to 0.6747. Given $R^2 = 0.6747$ (Model 1b), about 67% of the variation in the bank revenues has been explained by the model.

Our main aim is to determine the degree of competitiveness in the banking sector. The H-statistic is the sum of input prices elasticities. $H = \beta_1 + \beta_2 + \beta_3$.. The H-statistic ranges from 0.5495 to 0.8460. By model, the H-statistics are smaller under Model 1a and Model 2a (the model estimated using pooled OLS). On the other hand, Model 1b and Model 2b generate higher H-statistic (the model was estimated using the fixed effect model). The Wald test shows that the hypotheses (H = 0 and H = 1) are rejected. Since both are rejected, then the H values are significantly different from zero and unity. This leads to the rejection of the monopoly hypothesis, the perfect competition hypothesis, and the conjectural SR oligopoly hypothesis. Thus, the banking sector in Malaysia is best described as operating under monopolistic competition.[35] This study confirms the finding made by Claessens and Laeven (2004, 2003). Claessens and Laeven (2004) found that the average of the four H statistics for Malaysia is 0.68. The average of the four H statistics in this study is 0.69. Thus, we can say that our results are robust.

Claessens and Laeven (2004) covered the period 1994 to 2001, while this study covers the period from 1994 to 2000. We excluded commercial banks after 2000, since all domestic banks are new entities following the major

[35] Based on the market power hypothesis, the Hirshman-Herfidahl Index (HHI) based on bank assets for this study is less than 1,000. The HHI implies that the banking industry in Malaysia is competitive (based on USA'a competitive guidelines).

consolidation with the exception of two Islamic banks. This is to take into account the structural change in the Malaysian banking system. Other differences between this study and theirs refer to the different control variables used and the different definitions of price of labour and price of capital. Claessens and Laeven (2004, 2003) defined the price of labour as the ratio of personnel expense to total assets (we use the ratio of personnel expense to the number of labour). In the case of the price of capital, they defined it as the ratio of other operating and administrative expenses to total assets while we define it as the ratio of non-interest expenses to fixed assets. They also reported that there were 41 Malaysian banks used in their sample. However, the official statistics released by BNM (1994, 1999) show that the maximum number of commercial banks in a single year is only 37. Thus, we believe that Claessens and Laeven (2004) included other banking institutions, not only commercial banks.

Table 6.3: Estimation results for TRA

(all banks, 1994-2000)

Independent variables	Model 1a		Model 1b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.3029	0.0683*	0.2759	0.0616*
LPK	-0.0576	0.0444	0.0304	0.0536
LPD	0.3189	0.0774*	0.5396	0.0416*
LNB	0.0636	0.0139*	0.1259	0.0924
LEQTA	0.0157	0.0384	0.0328	0.0390
constant	-0.9792	0.2433*	-0.6246	0.3906
R^2	0.3957		0.6539	
H-statistic	0.5642		0.8460	
F value on Wald test for H=0	42.32		133.53	
P value	0.0000		0.0000	
F value on Wald test for H=1	25.24		4.42	
P value	0.0000		0.0372	
No of observation	192		191	

Notes: TRA is the ratio of gross operating revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effect), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.
*,**,*** represent significance levels at 1%, 5% and 10% respectively.

Table 6.4: Estimation results for INTA

(all banks, 1994-2000)

Independent variables	Model 2a		Model 2b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.2569	0.0797**	0.2226	0.0662**
LPK	-0.0850	0.0452***	0.0174	0.0552
LPD	0.3776	0.0733*	0.5615	0.0433*
LNB	0.0914	0.0214*	0.0059	0.1082
LOY	-0.0189	0.0186	0.0844	0.0399***
LEQTA	0.0007	0.0393	0.0384	0.0395
constant	-1.1530	0.2772*	-0.8054	0.4049
R^2	0.3892		0.6706	
H-statistic	0.5495		0.8054	
F value on Wald test for H=0	26.90		91.07	
P value	0.0000		0.0000	
F value on Wald test for H=1	18.08		5.31	
P value	0.0000		0.0226	
No of observation	191		191	

Notes: INTA is the ratio of interest revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effect), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.

*,**,*** represent significance levels at 1%, 5% and 10% respectively.

6.5.3 Estimation results for domestic and foreign banks

We want to check whether the banks might behave differently if different sets of data are used in the estimation. This decision is in accordance with the works of Coccorese (2004) and Bikker and Haaf (2002). Besides estimating the whole sample of Italian commercial banks, Coccorese (2004) also estimated the H-statistics by macro-regions: North-West, North-East, Centre, South and Islands. On the other hand, Bikker and Haaf (2002) split the sample based on the size of the banks: small banks, medium-sized banks and large banks. So, this study divides the banks into domestic and foreign banks. Out of 32 banks, 21 are domestic banks. The rest belong to foreign banks. Using both samples of domestic and foreign banks, we re-estimated Equations 6.8 and 6.9. The results for domestic banks are shown in Table 6.5 (equilibrium test) and Tables 6.6 and 6.7 (competitive test). For foreign banks, the estimation results are shown in Table 6.8 (equilibrium test) and Tables 6.9 and 6.10 (competitive test).

6.5.3.1 Equilibrium test for domestic banks

The equilibrium test (Table 6.5) shows that the domestic banks operate in long-run equilibrium. The null hypothesis that H = 0 is not rejected under the pooled OLS and fixed effects models. The H-statistics are -0.08 and 0.2881 under the pooled OLS and fixed effects models respectively.

6.5.3.2 Competition test

The estimation results for the bank revenue equation are shown in Tables 6.6 and 6.7. All coefficients of input prices have positive signs but only LPD and LPL remain statistically significant in all models. Again, β_3 is always the largest coefficient. The H-statistics range from 0.7235 to 0.8492. In all models, the hypothesis that H = 0 is rejected at the 1% significance level. This leads to the rejection of monopoly and the conjectural short-run oligopoly hypothesis. The hypothesis that H = 1 shows mixed results. Under Model 2a (dependent varaible is INTA), the hypothesis cannot be rejected. The H value is 0.8492 This implies that the banks operate as though under perfect competition. The mixed results prompt us to conduct another F-test to determine which model specification is more appropriate. The F-test again shows that the FE model is more appropriate than the pooled OLS.

Three models (Model 1a, Model 1b and Model 2b) show that the banks operate under monopolistic competition, as found when using the whole sample. The average of the four H values is 0.76. Bikker and Haaf (2002) also found the same trend. Using a sample of 23 countries, they found that all banks operate under monopolistic competition in all countries, without exception. However, when the markets were split into segments, the results became diversified. They split the sample into small, medium, and large banks. For example, in Australia the market of small banks is characterised by perfect collusion. In other countries like Denmark, Germany, Greece, Switzerland and United Kingdom, the market of large banks is characterised by perfect competition. In these countries, the hypothesis H =1 cannot be rejected.

Another interesting observation is that when using a sample of domestic banks only, the pooled OLS generates a higher H value than the fixed effects model. Model 1a and Model 2a yield H values of 0.7525 and 0.8492, respectively. Meanwhile, the fixed effects model produces H values of 0.7235 and 0.7258, respectively. Recall that when using the whole sample, the fixed effects model produced higher H values than those produced by the pooled OLS.

Table 6.5: Estimation results for ROA

(domestic banks only)

Independent variables	Pooled OLS		Fixed Effects Model	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.0899	0.1704	0.3125	0.1782***
LPK	0.2489	0.0937**	0.3118	0.1936
LPD	-0.4188	0.2638	-0.3362	0.1694***
LNB	0.1554	0.1218	-0.9506	0.3366**
LOY	-0.0068	0.0754	0.2462	0.1340***
LEQTA	0.2293	0.1424	0.3248	0.1594**
constant	-5.2052	0.9566*	-0.7510	1.4428
R^2	0.1324		0.1994	
H-statistic	-0.08		0.2881	
F value on Wald test for H=0	0.11		0.96	
P value	0.7416		0.3308	
No of observation	106		106	

Notes: ROA is the ratio of pre-tax profit to total assets, PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.
*,**,*** represent significance levels at 1%, 5% and 10% respectively.

Table 6.6: Estimation results for TRA

(domestic banks only)

Independent variables	Model 1a		Model 1b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.1721	0.0970***	0.0833	0.0424***
LPK	0.1065	0.0351**	0.0974	0.0428**
LPD	0.4739	0.0733*	0.5428	0.0311*
LNB	0.0290	0.0122**	0.2066	0.0599**
LEQTA	0.0567	0.0311***	0.0089	0.0222
constant	-0.7358	0.3482	-1.6477	0.3331*
R^2	0.5564		0.8517	
H-statistic	0.7525		0.7235	
F value on Wald test for H=0	43.38		135.27	
P value	0.0000		0.0000	
F value on Wald test for H=1	4.69		19.74	
P value	0.0303		0.0000	
No of observation	119		119	

Notes: TRA is the ratio of gross operating revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.
*,**,*** represent significance levels at 1%, 5% and 10% respectively.

Table 6.7: Estimation results for INTA

(domestic banks only)

Independent variables	Model 2a		Model 2b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.2213	0.0991**	0.0861	0.0418**
LPK	0.0965	0.0354**	0.0614	0.0434
LPD	0.5314	0.0649*	0.5783	0.0311*
LNB	0.1237	0.0399**	0.2187	0.0671**
LOY	-0.0816	0.0282**	0.0042	0.0230
LEQTA	0.0433	0.0306	0.0111	0.0215
constant	-0.6019	0.3413***	-1.7048	0.3234*
R^2	0.5928		0.8649	
H-statistic	0.8492		0.7258	
F value on Wald test for H=0	52.58		126.62	
P value	0.0000		0.0000	
F value on Wald test for H=1	1.66		18.05	
P value	0.1980		0.0001	
No of observation	118		118	

Notes: INTA is the ratio of interest revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.

*,**,*** represent significance levels at 1%, 5% and 10% respectively.

6.5.3.3 Equilibrium test for foreign banks

The equilibrium test (Table 6.8) shows that the foreign banks operate in long-run equilibrium. The null hypothesis that H = 0 is not rejected under the pooled OLS and fixed effects models. The H-statistics are -0.0177 and 0.0082 under the pooled OLS and fixed effects models respectively.

6.5.3.4 Competition test for foreign banks

The estimation results for bank revenue (TRA and INTA) are shown in Tables 6.9 and 6.10. The H-statistics range from 0.4733 to 0.9043. The average of the four H values is 0.66. This is close to the average H values for domestic banks only (0.76). In all models, the hypothesis that H=0 is rejected at the 1% significance level. This leads to the rejection of monopoly and the conjectural short-run oligopoly hypothesis. Again, the hypothesis H=1 shows mixed results. This hypothesis is rejected under pooled OLS but is accepted under the fixed effects models. The mixed results prompt us to conduct another F-test to determine which model specification is more appropriate. The F-test again shows that the FE model is more appropriate than the pooled OLS. Since the hypothesis H=1 is accepted, this suggests that the market for foreign banks in Malaysia is best described as perfectly competitive.

Table 6.8: Estimation results for ROA

(foreign banks only)

Independent variables	Pooled OLS		Fixed Effects Model	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	-0.0374	0.1980	-0.1856	0.2305
LPK	0.0079	0.0882	0.0012	0.1204
LPD	0.0118	0.1025	0.1024	0.1035
LNB	-0.1115	0.0584***	-1.1123	0.3386**
LOY	0.2702	0.0475*	0.5867	0.1560*
LEQTA	0.9113	0.1827*	0.9239	0.2351*
constant	-2.6868	0.6292*	-2.5031	1.2469***
R^2	0.5787		0.5268	
H-statistic	-0.0177		-0.0082	
F value on Wald test for H=0	0.01		0.13	
P value	0.9178		0.7240	
No of observation	69		69	

Notes: ROA is the ratio of pre-tax profit to total assets, PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.
*,**,*** represent significance levels at 1%, 5% and 10% respectively.

Table 6.9: Estimation results for TRA

(foreign banks only)

Independent variables	Model 1a		Model 1b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.5175	0.0924*	0.5138	0.1331*
LPK	-0.1972	0.0464*	-0.1014	0.0982
LPD	0.2789	0.0744*	0.4918	0.0802*
LNB	0.1179	0.0276*	-0.2948	0.2501
LEQTA	0.1668	0.1323	0.4629	0.1809**
constant	-0.2073	0.4004	1.6420	0.7409**
R^2	0.4535		0.6387	
H-statistic	0.5992		0.9043	
F value on Wald test for H=0	39.32		51.23	
P value	0.0000		0.0000	
F value on Wald test for H=1	17.57		0.57	
P value	0.0000		0.4518	
No of observation	73		73	

Notes: INTA is the ratio of interest revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.

*,**,*** represent significance levels at 1%, 5% and 10% respectively.

Table 6.10: Estimation results for INTA

(foreign banks only)

Independent variables	Model 2a		Model 2b	
	Coefficients	Standard Error	Coefficients	Standard Error
LPL	0.3405	0.1047**	0.2595	0.1858
LPK	-0.2511	0.0581*	-0.0494	0.1022
LPD	0.3839	0.0756*	0.4920	0.0864*
LNB	0.0713	0.0345**	-0.5603	0.2921***
LOY	0.0594	0.0284**	0.2454	0.1309***
LEQTA	0.0191	0.1269	0.4170	0.1882**
constant	-1.1326	0.4372**	0.1268	1.0533
R^2	0.4425		0.6497	
H-statistic	0.4733		0.7021	
F value on Wald test for H=0	18.29		13.24	
P value	0.0000		0.0006	
F value on Wald test for H=1	22.65		2.38	
P value	0.0000		0.1282	
No of observation	73		73	

Notes: INTA is the ratio of interest revenue to total assets (proxy for output price), PL is the ratio of personnel expenses to total labour (proxy for input price of labour), PK is the ratio of non-interest expenses to fixed assets (proxy for input price of capital), PD is the ratio of interest expenses to total deposits (proxy for input price of deposit), NB is the number of bank branches (to control for potential size effects), OY is other operating income (to control for non-interest based services), EQTA is the ratio of total equity to total assets (to control for bank-specific risk). All variables are in logarithm.

*,**,*** represent significance levels at 1%, 5% and 10% respectively.

6.6 Concluding remarks

Following the gradual financial liberalisation and deregulation of the Malaysian banking industry, it is interesting to know its impact on the competitive environment of the banking sector. Although this policy is limited, it gives greater freedom to the banks in terms of deciding their own deposit and lending rates. This has undeniably opened the door to competition amongst the banks. This study seeks to evaluate the competitive structure of the Malaysian banking industry and to estimate the current level of market contestability from 1994 to 2000. To achieve these objectives, we use the Panzar-Rosse model named after Panzar and Rosse (1987). This model requires us to perform equilibrium and competition tests using the return on assets equation and the reduced-form revenue equation. The equilibrium test is conducted to ensure that the banks are in long-run equilibrium. The competition test is performed to obtain the H statistics, which are then used to indicate the degree of competitiveness in the banking industry. This study employed two estimation techniques, pooled OLS and the fixed effects model. Following the F-test, the FE model is found to be more appropriate than the pooled OLS.

Using the whole sample, the results show that the banking industry is uniformly characterised by monopolistic competition for the period between 1994 and 2000. The $H = 0$ and $H = 1$ hypotheses are rejected convincingly. This implies that given a large number of banks and differentiated products and services, the banks can relatively exert monopoly power to influence the output prices. The average of the four H statistics is 0.69. This means that the market is competitive to a certain degree. Lower H values indicate less competitiveness and higher values indicate a highly competitive market. The main contributor to the H statistics is the price of deposits, followed by the price of labour. The fixed effects model was found to generate higher H values.

Using domestic banks only, the H statistics were found to be higher. The average of the four H values was 0.76. This indicates that the market of domestic banks is competitive. However, in one model estimated under the pooled OLS, the banking industry is characterised by perfect competition. The H = 1 hypothesis cannot be rejected. However, since the FE model is a better specification than the pooled OLS, then the results remain the same i.e. the domestic banks in Malaysia seem to operate under monopolistic competition.

Using foreign banks only, the average H values are similar to the one estimated by using the whole sample (0.66 and 0.69 respectively). This suggests that the banks are best described to be operating under monopolistic competitition. However, under the FE model, perfect competition is more appropriate to describe the market for foreign banks.

Chapter Seven

CHAPTER SEVEN

Summary, policy recommendations and future research

7.0 Introduction

This research has been an attempt to look at the efficiency and competitive behaviour of commercial banks in Malaysia. Understanding of these two aspects of the banks performance is of great interest to the policy makers, the bankers, and the economists. Such understanding can provide some explanation of what has happened to these banks following the implementation of financial liberalisation since the late 1970s. Furthermore, it is also important to find some justification for the recent consolidation of domestic banks.

This study wants to achieve four objectives: the measurement of technical efficiency, the identification of productivity growth of the commercial banks using a non-parametric approach, the measurement of cost efficiency of the commercial banks using both parametric and non-parametric approaches and the determination of the competitive behaviour of the banks using a non-structural model. The work on the first two objectives is presented in Chapter Four. Chapters Five and Six deal exclusively with the third and fourth objectives.

To maintain consistency, this study considers the banks as financial intermediaries rather than production units. This subsequently leads to the selection of total deposit as one of the three inputs. The other inputs were labour and capital. The three outputs were total loans, other earning assets, and other operating income. The study covers the period 1994 – 2000.

The rest of this chapter is as follows. The immediate section summarises the main findings of the study. The discussion is closely linked to the objectives of the study. Section 7.2 recommends policies that can be derived from this study. Finally, Section 7.3 offers potential aspects for further research.

7.1 Summary of findings

The findings of the study are summarised in three main themes so that appropriate conclusions can be reached. The three themes are technical and cost efficiency, productivity growth, and competitive behaviour.

7.1.1 Technical and cost efficiency

Two signicant outcomes of the financial liberalisation in the Malaysian banking industry are that it permits the banks to decide their own interest rates (both deposit and lending rate) and to widen the scope of outputs produced. In turn, deposit price or deposit rate and bank output are key determinants of both production and cost. This study looks at both technical and cost efficiency. The former is derived from a production frontier function by using the DEA technique while the latter is estimated based on a cost frontier function by using both the parametric approach and the DEA technique.

In general, the average efficiency measures obtained under the DEA technique is higher than those generated by the parametric approach (SFA and FE method). For example, the average PTE and CE_{vrs} were 83% and 75% respectively compared with CE1 (32%) and CE3 (52%). This finding suggests that the measurement of efficiency is subject to which type of frontier function is being used (either production or cost), given the same inputs and outputs. The estimation technique is also important as demonstrated by both the parametric and the non-parametric approach. Although the average efficiency scores vary from one method to another, the rank order correlation results show that these scores are related to each other with the exception of correlation between CE_{vrs} and other parametric efficiency scores.

Bank ownership plays an important role in explaining the differences in the efficiency scores. This study splits the commercial banks into two categories: domestic and foreign banks. On average, the foreign banks are significantly more efficient than the domestic banks except in the case of CE3 and CE4 (not significantly different). In the case of OTE, PTE and SE (obtained under the DEA technique), the foreign banks have obvious superiority over local counterparts. This implies that, given an amount of outputs, the foreign banks manage to use fewer inputs than the domestic banks. They are also in a better position than their local counterparts for finding the optimal combinations of various inputs to produce the desired output as indicated by the scale efficieny. The comparison of efficiency scores by bank ownership also discloses that the superior performance of the foreign banks is not influenced by the choice of the estimation technique used. In all techniques, the foreign banks on average have higher efficiency estimates than the domestic banks.

This study finds that the empirical support for the recent banking consolidation is not conclusive. As stated before, one of the arguments for this consolidation is that by increasing the size of the domestic banks, efficiency can be enhanced. It appears that this consolidation programme gets some support from the efficiency measures obtained by using the DEA technique (except CE_{crs}) and no support from the parametric estimates of cost efficiency. This is not surprising as the literature has shown that the relationship between efficiency and size is mixed.

Using ROA as a measure of bank performance, the notion that the most profitable banks are also the most efficient ones is only established when applying the DEA efficiency scores. This notion cannot be proven when using parametric estimates of efficiency. Again, this finding is not unusual as other studies have shown.

7.1.2 Productivity growth

This study constructs the DEA-based Malmquist productivity index and then decomposes the sources of productivity growth of the banks. On average, the productivity of the banks has increased over the sample period by about 22%. The increment in productivity was contributed by technical improvement or innovation rather than efficiency improvement. Comparing the productivity growth of domestic and foreign banks, this study suggests that while there was little difference in the average efficiency between domestic and foreign banks at the beginning of the period, it was the growth in productivity over the period that provided foreign banks with an efficiency advantage. In particular, the Asian financial crisis exposed domestic banks to greater financial stress. Our results show that the differences in productivity are not related to differences in technical efficiency. We can interpret the results as differences caused by risk exposure. Malaysian domestic banks were more exposed to the Asian financial crisis than foreign banks operating in Malaysia.

7.1.3 Competitive behaviour of the banks

This study finds that the banking industry is best described as behaving under monopolistic competition. Under this kind of market structure, the banks have some kind of market power that enables them to adjust the output prices (lending rates) given changes in input costs. This adjustment is made possible through financial liberalisation that grants greater freedom and flexibility to the banks. Equipped with control over their own deposit and lending rates, the banks can now compete effectively in the market.

7.2 Policy recommendations

The magnitude of the relationship between efficiency and size is not conclusive. Thus, there is limited support for the recent banking consolidation. Since the banking consolidation aims to ehance efficiency, greater size cannot be taken as the only means to do so. Other measures must be pursued in the aftermath of the merger programme. Perhaps, greater participation of the foreign ownership can be exploited as a catalyst for a higher efficiency level. This study has so far found that the foreign banks have higher efficiency measures than the domestic banks. It is an interesting idea to see foreign participation in the domestic banks in terms of equity participation. The recent announcement by the BNM might be a practical step to enhance the current efficiency of the domestic banks. The BNM has made a decision to allow foreign participation in the local banks up to 49%. Besides equity participation, there is a clear signal from the BNM to encourage the foreign banks to penetrate further into the banking industry. This suggests that the foreign banks might be allowed to set up more branches all over the country, in particular in rural and suburban areas.

The analysis of the sources of productivity growth seems to suggest that, in general, technical change or innovation is the main contributor besides improvement in efficiency. Improvement in technology will shift the production frontier upwards assuming other things remain the same. Such improvement might require greater freedom for the banks and perhaps further incentives by the BNM. However, greater freedom for the banks should not compromise prudential regulation. Some studies suggest that financial liberalisation in the absence of prudential regulation will only bring the banking industry into a crisis.

Since technical innovation has a positive impact on the banks' productivity, the internet banking and computerised banking activities should be pursued more vigorously. It has been shown that this type of banking transaction is capable of reducing operating costs and creates greater access for the bank customers. This has been proven by the use of electronic networks and the cash machines that connect with the customers. In general, the Malaysian commercial banks with the exception of foreign banks are new to this kind of business transaction and a lot must be done. One important aspect is the internet access and the availability of competent internet service providers. At present, there are only two internet service providers and the internet costs are still high. This has been demonstrated when some domestic banks impose additional charges for using their internet banking. Perhaps, the higher costs are due to the lack of competition in the market. By bringing more players into the market, it is hoped that the internet banking can be improved and the operating cost can be lowered. This will lead to cheaper prices and greater access for the public. Subsequently, the Malaysian banking system can be transformed into a commercially-driven high-tech business.

Since 2001, the domestic banks on average have increased in size and the banking industry itself has undergone a structural change due to a major bank consolidation. Given a reduced number of banks, the Central Bank should, from now on, monitor or investigate the contestability conditions in the market. A smaller number of players is normally linked with collusive strategy and anti-competitive behaviour which may place public welfare at risk. In addition, excessive monopolistic power may result in higher lending rates, unattractive deposit rates and lesser loanable funds. At present, there is no guideline that monitors the competitive conditions of the banking industry in particular after the recent merger. It is the right time to have competitive guidelines similar to those used in the USA and other developed countries.

7.3 Future research

This study seeks to measure efficiency using two broad techniques, namely parametric and non-parametric approaches. The beauty of the non-parametric approach is that it can deal with multiple outputs. The parametric approach cannot deal with multiple outputs. So, it might be a good idea to compare the technical efficiency with the cost efficiency of multiple output firms using the same DEA technique. This might provide an answer to the contention that a technically efficient firm is also cost efficient.

Another concept of efficiency that can be explored is profit efficiency. This requires the setting up of a profit frontier function similar to a production and cost frontier function. What is required is data on bank profit like rates of return on assets or equity. Again, the technique of estimating the profit efficiency is also similar: parametric and non-parametric.

Another aspect that can be further explored is the determinants of productivity growth. As discussed in the previous chapter, the sources of productivity growth are two: technical change and efficiency change. If we were to identify the determinants of productivity growth, then we should encompass all the factors that can influence both technical change and efficiency change.

This study is mainly concerned with efficiency and competitive behaviour. However, both were analysed separately. Perhaps, it is better to see the link between them. It is an interesting idea to determine whether the competitive behaviour of the banks eventually leads to greater efficiency. Theoretically, the presence of intense competition may require firms to be more efficient in terms of using the available but limited resources. We would like to suggest that such study may require a larger sample and a longer period of time. Secondly, instead of time invariant measures, time varying cost efficiency by a parametric approach can be an alternative measure of efficiency.

The inefficiency model is worth exploring. So far, this study has measured technical and cost efficiency using both parametric and non-parametric methods. Having obtained scores, it will be interesting to look at factors that may affect efficiency. This can be done by setting up inefficiency model. This model can offer a greater scope for exploring the determinants of inefficiency.

Besides market structure and ownership variable, one should also look at other variables like balance sheet items, corporate governance, management performance and the risk taking variable. The findings generated by this inefficiency model might offer new insights to assist the local banks so that their performance can get closer to that of the foreign banks.

This study shows that, on average, the foreign banks are more efficient than the local banks. Given the same access to technology, we argue that the differences in efficiency measures may source from managerial aspects. Perhaps, the study suggests, this is because the foreign banks have superiority in management over the local banks. Managerial efficiency might be the result of staff training and flexible movement of labour. This can be explored further using questionnaires or a qualitative survey.

References

Abd_Karim, M.Z. 2001. Comparative bank efficiency across select ASEAN countries. *ASEAN Economic Bulletin* 18, pp. 289-304.

Adler, N., Friedman, L and Sinuany-Stern, Z. 2002. Review of ranking methods in the data envelopment analysis context. *European Journal of Operational Research 140*, pp. 249-265.

Aigner, D., Lovell, C.A.K and Schmidt, P. 1977. Formulation and estimation of stochastic production function models. *Journal of Econometrics* 6, pp. 21-37.

Akhavein, J.L., Berger, A.N and Humphrey, D.B. 1997. The effects of mega mergers on efficiency and prices: Evidence from a bank profit function. *Review of Industrial Organization* 12, pp. 95 – 139.

Akhigbe, A and McNulty, J.E. 2003. The profit efficiency of small US commercial banks. *Journal of Banking and Finance* 27, pp. 307 – 325.

Alam, I.M.S. 2001. A nonparametric approach for assessing productivity dynamics of large U.S. banks. *Journal of Money, Credit and Banking* 33, pp. 121-139.

Allen, F and Santomero, A.M. 1998. The theory of financial intermediation. *Journal of Banking and Finance* 21, pp. 1461-1485.

________ 2001. What do financial intermediaries do? *Journal of Banking and Finance* 25, pp. 271 – 294.

Allen, L and Rai, A. 1996. Operational efficiency in banking: An international comparison. *Journal of Banking and Finance* 20, pp. 655-672.

Altunbas, Y and Chakravarty, S.P. 1998. Efficiency measures and the banking structure in Europe. *Economic Letters* 60, pp. 205-208.

________. 2001. Frontier cost functions and bank efficiency. *Economic Letters* 72, pp. 233-240.

Altunbas, Y and Molyneux, P. 1993. Scale and scope economies in European banking. *Institute of European Finance* RP 93/11.

________. 1996. Cost economies in EU banking system. *Journal of Economics and Business* 48, pp. 217-230.

Altunbas, Y. Liu, M-H., Molyneux, P and Seth, R. 2000. Efficiency and risk in Japanese banking. *Journal of Banking and Finance* 24, pp. 1605-1628.

Altunbas, Y., Evans, L and Molyneux, P. 2001. Bank ownership and efficiency. *Journal of Money, Credit and Banking* 33, pp.926-954.

Ariff, M and Khalid, A.M. 2000. *Liberalization, growth and the Asian financial crisis*. Edward Elgar Publishing, Inc.

Association of Bankers in Malaysia. 1994. Bankers' Directory.

___________, 1996. Bankers' Directory.

___________, 1998. Bankers' Directory.

___________, 2000. Bankers' Directory.

Avkiran, N.K. 1999a. An application reference for data envelopment analysis in branch banking: Helping the novice researcher. *International Journal of Bank Marketing* 17, pp.206-220.

__________. 1999b. The evidence on efficiency gains: The role of mergers and the benefits to the public sector. Journal of Banking and Finance 23, pp.991-1013.

Bain, J.S. 1968. *Industrial Organization*. Second Edition. Wiley, New York.

Baltagi, B.H. 2001. *Econometric Analysis of Panel Data*. Second Edition. John Wileys & Sons, Ltd.

Bank Negara Malaysia. 1994. *Money and banking in Malaysia*. Forth Edition. Bank Negara Malaysia, Kuala Lumpur.

___________. 1999. *The central bank and the financial system in Malaysia*. Bank Negara Malaysia, Kuala Lumpur.

___________. 2001a. *Annual Report 2001*. Bank Negara Malaysia, Kuala Lumpur.

___________. 2001b. Annual Reports 1991 – 2000. Bank Negara Malaysia, Kuala Lumpur.

___________. 2003. *Annual Report 2003*. Bank Negara Malaysia, Kuala Lumpur.

__________. 2004. [WWW] URL:http:/www.bnm.gov.my. [Accessed 15 March 2005]

Banker, R.D. 1993. Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science* 39, pp. 1265-1273.

Banker, R.D., Charnes, A and Cooper, W. 1984. Some models for estimating technical and scale efficiencies in data envelopment analysis. *Management Science* Volume 30, pp. 1078-1092.

Barth, J.R., Caprio, G and Levine, R. 2000. Banking systems around the globe: Do regulation and ownership affect performance and stability? *World Bank*, Policy Research Working Papers 2325.

Battese, G.E and Coelli, T.J. 1988. Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data. *Journal of Econometrics* 38, pp. 387 – 399.

__________. 1995. A model for technical inefficiency effects in a stochastic frontier production function for panel data. Empirical Economics 20, pp. 325 – 332.

Bauer, P.W., Berger, A.N., Ferrier, G.D and Humphrey, D.B . 1998. Consistency conditions for regulatory analysis of financial institutions: A comparison of frontier efficiency methods. *Journal of Economics and Business* 50, pp. 85-114.

Beccalli, E. 2004. Cross-country comparisons of efficiency: evidence from the UK and Italian investment firms. *Journal of Banking and Finance* 28, pp. 1363-1383.

Beck, N and Katz, J.N. 1995. What to do (and not to do) with time series cross-section data. *American Political Science Review* 89, pp. 634-647.

Beck, T., Demirguc-Kunt, A and Levine, R. 1999. A new database on financial development and structure. *World Bank*, Policy Research Working Papers 2146.

Berger, A.N. 1991. *The Profit-concentration relationship in banking.* Federal Reserve Board, Washington D. C. Working Paper 176.

__________. 1993. Distribution-free estimates of efficiency in the U.S. banking industry and tests of the standard distributional assumptions. Journal of Productivity Analysis 4, pp. 261-292.

________ 1995. The profit-structure relationship in banking: Tests of market power and efficient-structure hypotheses. *Journal of Money, Credit and Banking* Volume 27, pp.404-431.

Berger, A.N and DeYoung, R. 1997. Problems loans and cost efficiency in commercial banks. *Journal of Banking and Finance* 21, pp. 849-870.

Berger, A.N and Humphrey, D.B. 1991. The dominance of inefficiencies over scale and product mix economies in banking. *Journal Of Monetary Economics* 28, pp. 117-148.

________. 1997. Efficiency of financial institutions: international survey and directions for future research. *European Journal of Operational Research* 98, pp. 175-212.

Berger, A.N and Mester, L.J. 1999. *What explains the dramatic changes in cost and profit performance of the U.S. banking industry?* Working Paper No. 99-1. Federal Reserve Bank of Philadelphia.

__________. 1997. Inside the black box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance* 21, pp. 895-947.

Berger, A.N., Demsetz, R.S and Strahan, P.E. 1999. The consolidation of the financial services industry: Causes, consequences, and implications for the future. *Journal of Banking and Finance* 23, pp.135-194.

Berger, A.N., Leusner, J.H and Mingo, J.J. 1997. The efficiency of bank branches. *Journal of Monetary Economics* 40, pp.141-162.

Bhattacharyya, A., Lovell, C.A.K and Sahay, P. 1997. The impact of liberalization on the productive efficiency of Indian commercial banks. *European Journal of Operational Research* 98, pp. 332-345.

Bikker, J.A and Haaf, K. 2002. Competition, concentration and their relationship: An empirical analysis of the banking industry. *Journal of Banking and Finance* 26, pp. 2191-2214.

Bonin, J.P., Hasan, I and Wachtel, P. 2005. Bank performance, efficiency and ownership in transition countries. *Journal of Banking and Finance* 29, pp. 31-53.

Bourke, P. 1989. Concentration and other determinants of bank profitability in Europe, North America and Australia. *Journal of Banking and Finance* 13, pp.65-79.

Bowlin, W.F. 2002. Measuring performance: An introduction to data envelopment analysis (DEA). University of Northern Iowa. mimeo

Bresnahan, T.F. 1982. The oligopoly solution concept is identified. *Economic Letters* 10, pp. 87-92.

________ 1989. Empirical studies of industries with market power. In: Schmalensee, R and Willig, R.D. eds. *Handbook of Industrial Organizations*. Volume II, Elsevier Science Publishers B.V. pp. 1011-1057.

Brown, J.H. 2002. Structure-conduct-performance: A comment on Balug's "Is competition such a good thing? Static efficiency versus dynamic efficiency". *Review of Industrial Organization* 21, pp. 103-105.

Canhoto, A and Dermine, J. 2003. A note on banking efficiency in Portugal, New vs. Old banks. *Journal of Banking and Finance* 27, pp. 2087-2098.

Casu, B. 2000. A comparative study of efficiency in European banking. *PhD Thesis*, University of Wales, Bangor.

Casu, B, Girardone, C and Molyneux, P. 2003. Productivity change in European banking: A comparison of parametric and non-parametric approaches. *Journal of Banking and Finance* 28, pp. 2521-2540.

Caudill, S.B. 2002. SFA, TFA and a new thick frontier: Graphical and analytical comparisons. *Applied Financial Economics* 12, pp. 309-317.

Cavallo, L and Rossi, S.P.S. 2002. Do environmental variables affect the performance and technical efficiency of the European banking systems? A parametric analysis using the stochastic frontier approach. *The European Journal of Finance* 8, pp. 123-126.

Caves, D.W., Christensen, L.R and Diewert, W.E. 1982. The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica* 50, pp. 1393-1414.

Cetorelli, N. 2001. Competition among banks: Good or bad? *Economic Perspective* Federal Reserve Bank of Chicago Volume XXV (2), pp. 38-47.

Charnes, A., Cooper, W.W and Rhodes, E. 1978. Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, pp. 429-44.

Charnes, A., Cooper, W.W., Lewin, A.Y and Seiford, L.M. 1994. *Data Envelopment Analysis: Theory, Methodology and Application*. Kluwer Academic Publishers, Boston.

Chen, T.Y. 2001. An estimation of X-inefficiency in Taiwan's banks. *Applied Financial Economics* 11, pp.237-242.

Chin, K. F and Jomo, K.S. 2001. Financial liberalisation and system vulnerability. In: Jomo, K.S. ed. *Malaysian Eclipse: Economic Crisis and Recovery.* Zed Books Ltd, London.

________. 2003. From financial liberalisation to crisis in Malaysia. In: Chung, H.L. ed. *Financial liberalisation and the economic crisis in Asia.* European Institute of Japanese Studies East Asian Economics and Business Studies, RoutledgeCurzon, London.

Christensen, L.R., Jorgenson, D.W and Lau, L.J. 1973. Transcendental logarithmic production frontiers. *The Review of Economics and Statistics* 55, pp. 28-45.

Christopoulus, D.K and Tsionas, E.G. 2001. Banking economic efficiency in the deregulation period: Results from heteroscedastic stochastic frontier models. *The Manchester School* 69, pp. 656-676.

Chu, S.F and Lim, G.H. 1998. Share performance and profit efficiency of banks in an oligopolistic market: evidence from Singapore. *Journal of Multinational Financial Management* 8, pp. 155-168.

Chun, S.E and Kim, B.H. 2004. The effect of financial restructuring on the degree of competition in the Korean banking industry. Mimeo.

Claessens, S and Laeven, L. 2003. What drives bank competition? Some international evidence. World Bank Working Paper 3113.

________. 2004. What drives bank competition? Some international evidence. *Journal of Money, Credit and Banking* 36, pp. 563-583.

Coccorese, P. 2004. Banking competition and macroeconomic conditions: a disaggregate analysis. *Journal of International Financial Markets, Institutions and Money* 14, pp.203-219.

Coelli, T. 1996. "*A guide to DEAP Version 2.1: A Data Envelopment Analysis (Computer) Program*", CEPA Working Paper 96/08, University of New England, Australia.

Cornwell, C and Schmidt, P. 1996. Production frontiers and efficiency measurement. In: Matyas, L and Sevestre, P. *The Econometrics of Panel Data.* Second Edition, Kluwer Academic Publishers.

Cuesta, R.A and Orea, L. 2002. Mergers and technical efficiency in Spanish savings bank: a stochastic distance function approach. *Journal of Banking and Finance* 26, pp. 2231-2247.

Cummins, J.D and Zi, H. 1998. Comparison of frontier efficiency methods: An application to the U.S. life insurance industry. *Journal of Productivity Analysis* 10, pp. 131-152.

Denizer, C.A., Dinc, M and Tarimcilar, M. 2000. Measuring banking efficiency in the pre- and post-liberalization environment: Evidence from the Turkish banking system. *World Bank*, Policy Research Working Papers 2476.

DeYoung, R. 1997a. A diagnostic test for the distribution-free efficiency estimator: An example using U.S. commercial bank data. *European Journal of Operational Research* 98 pp.243-249.

__________. 1997b. Comment on operational efficiency in banking: An international perspective. *Journal of Banking and Finance* 21, pp.1325-1329.

DeYoung, R and Hasan, I. 1998. The performance of de novo commercial banks: A profit efficiency approach. *Journal of Banking and Finance* 22, pp.565-587.

DeYoung, R and Nolle, D.E. 1996. Foreign-owned banks in the United States: Earning market share or buying it? *Journal of Money, Credit and Banking* Volume 28 (4), pp.622-636.

Dietsch, M and Lozano-Vivas, A. 2000. How the environment determines banking efficiency: A comparison between French and Spanish industries. *Journal of Banking and Finance* 24, pp. 985-1004.

Drake, L. 2001. Efficiency and productivity change in UK banking. *Applied Financial Economics* 11, pp.557-571.

________. 2004. Costs and efficiency in banking: a survey of the evidence from the US, the UK and Japan. In: Mullineux A.W and Murinde, V. eds. *Handbook of International Banking*. Cheltenham: Edward Elgar.

Drake, L and Hall, M.J.B. 2003. Efficiency in Japanese banking: An empirical analysis. *Journal of Banking and Finance* 27, pp. 891-917.

Edwards, S.J. 1999. Bank efficiency and performance: A study of developing countries. *PhD Thesis*, University of Wales, Cardiff.

Elyasiani, E and Rezvanian, R. 2002. A comparative multiproduct cost study of foreigned-owned bank and domestic-owned US banks. Applied Financial Economics 12, pp. 271-284.

English, M., Grosskoph, S., Hayes, K and Yaisawarng, S 1993. Output allocative and technical efficiency of banks. *Journal of Banking &Finance* 17, pp. 349-366.

Färe, R and Grosskopf, S. 2001. APEC and the Asian economic crisis: Early signals from productivity trends. *Asian Economic Journal* Volume 15, pp. 325-341.

Färe, R., Grosskopf, S., Norris, M and Zhang, Z. 1994. Productivity growth, technical progress and efficiency change in industrialised countries. *The American Economic Review* 84, pp. 66-83.

Färe, R and Lovell, C.A.K. 1978. Measuring the technical efficiency of production. *Journal of Economic Theory* 19, pp. 150-162.

Farrell, M.J. 1957. The measurement of productive efficiency of production. *Journal of the Royal Statistical* Society, Series A, 120, pp. 253-281.

Favero, C.A and Papi, L. 1995. Technical efficiency and scale efficiency in the Italian banking sector: Non-parametric approach. *Applied Economics* 27, pp. 385-395.

Fiordelisi, F and Molyneux, P. 2004. Efficiency in the factoring industry. *Applied Economics* 36, pp. 947-959.

Fischer, T and Kamerschen, D.R. 2003. Measuring competition in the US airline industry using the Panzar-Rosse test and cross-sectional regression analyses. *Journal of Applied Economics* 6, pp. 73-93.

Freixas, X and Rochet, J.C. 1997. *Microeconomics of Banking.* Cambridge MIT Press.

Fry, M.J. 1995. *Money, interest and banking in economic development.* Baltimore Md, London.

________. 1997. In favour of financial liberalisation. *The Economic Journal* 107, pp. 754-770.

Gallant, A.R. 1982. Unbiased determination of production technologies. *Journal of Econometrics* 20, pp. 285-323.

Garden, K.A and Ralston, D.E. 1999. The X-efficiency and allocative efficiency effects of credit union mergers. *Journal of International Financial Markets, Institutions and Money* 9, pp. 285-301.

Giannakas, K., Tran, K.C and Tzouvelekas, V. 2003. Predicting technical efficiency in stochastic production frontier models in the presence of misspecification: a Monte Carlo analysis. *Applied Economics* 35, pp. 153 – 161.

Gilbert, A. 1984. Studies of bank market structure and competition: A review and evaluation. *Journal of Money, Credit and Banking* 16, pp.617-644.

Girardone, C., Molyneux, P and Gardener, E.P.M. 2004. Analysing the determinants of bank efficiency: the case of Italians banks. *Applied Economics* 36, pp. 215 - 227.

Goldberg, L.G and Rai, A. 1996. The structure-performance relationship for European banking. *Journal of Banking and Finance* 20, pp.745-771.

Gravelle, H and Rees, R. 1981. *Microeconomics*, Longman London and New York.

Greene, W.H. 1993. The econometric approach to efficiency analysis. In: Fried, H. et al eds. The measurement of productive efficiency: Theory and applications, Oxford University Press, New York.

__________. 2000. *Econometric Analysis*, Fourth Edition, Prentice Hall Publications.

__________. 2001. *New developments in the estimation of stochastic frontier models with panel data.* Department of Economics, Stern School of Business, New York University.

________. 2002. Fixed and random effects in stochastic frontier models. Department of Economics, Stern School of Business, New York University. Mimeo.

Greene, W.H and Segal, D. 2004. Profitability and efficiency in the U.S. life insurance industry. *Journal of Productivity Analysis* 21, pp. 229-247.

Grifell-Tatjé, E and Lovell, C.A.K. 1997. The sources of productivity change in Spanish banking. *European Journal of Operational Research* 98, pp 364-380.

Griffiths, W.E., Hill, R.C and Judge, G.G. 1993. *Learning and Practicing Econometrics*. John Wiley and Sons, Inc.

Grosskopf, S. 2003. Some remarks on Productivity and its decompositions. *Journal of Productivity Analysis* 20, pp. 459-474.

Gujarati, D. 1992. *Essentials of Econometrics*. McGraw-Hill International Editions.

Hababou, M. 2002. Tutorial in DEA. [WWW] URL:http://members.tripod.com/moezh/DEAtutorial/ [Accessed 29 August 2002]

Hair, J.F., Anderson, R.E., Tatham, R.L and Black, W.C. 1998. *Multivariate Analysis*. Fifth Edition. Prentice Hall.

Hannan, T.H. 1991. Foundations of the structure-conduct-performance paradigm in banking. *Journal of Money, Credit and Banking* 23, pp. 68-84.

Hardy, D.C and di Patti, B.E. 2001. Bank reform and bank efficiency in Pakistan. *IMF Working Paper*, WP/01/138.

Hashim, M.P. 2001. "Economies of scale and economies of scope of Malaysian banks", *PhD Thesis*, Putra University of Malaysia.

Hausman, J.A. 1978. Specification tests in econometrics. *Econometrica* 46, pp. 1251-1271.

Henderson, J.M and Quant, R.E. 1980. *Microeconomic Theory: A Mathematical Approach*. McGraw-Hill.

Hollingsworth, B. 2004. Non parametric efficiency measurement. *The Economics Journal* 114, pp. 307 – 311.

Hondroyiannis, G., Lolos, S and Papapetrou, E. 1999. Assessing competitive conditions in the Greek banking system. *Journal of International Financial Markets and Money* 9, pp. 377-391.

Howe, W.S. 1978. *Industrial Economics: An applied approach*. London: The Macmillan Press Ltd.

Hsiao, C. 2003. *Analysis of Panel Data*. Second Edition. Cambridge University Press.

Humphrey, D.B and Vale, B. 2004. Scale economies, bank mergers and electronic payments: A spline function approach. *Journal of Banking and Finance* 28, pp. 1671-1696.

Isik, I and Hassan, M.K. 2002. Technical, scale and allocative efficiencies of Turkish banking industry. *Journal of Banking and Finance* 26, pp. 719-766.

Iwata, G. 1974. Measurement of conjectural variations in oligopoly. *Econometrica* 42, pp. 947-966.

Jomo, K.S. 1998. Financial liberalisation, crises, and Malaysian policy responses. *World Development* 26, pp. 1563-1574.

Jomo, K.S. 2001. From currency crisis to recession. In: Jomo, K.S. ed. *Malaysian eclipse: Economic crisis and recovery*. Zed Books Ltd, London and New York.

Jomo, K.S. and Hamilton-Hart, N. 2001. From financial liberalisation to crisis. In: Jomo, K.S. ed. *Malaysian eclipse: Economic crisis and recovery*. Zed Books Ltd, London and New York.

Jondrow, J.C., Lovell, C., Materov, I and Schmidt, P. 1982. On the estimation of technical inefficiency in the stochastic frontier production function model. *Journal of Econometrics* 19, pp. 233-238.

Karafolas, S and Mantakas, G. 1996. A note on cost structure and economies of scale in Greek banking. *Journal of Banking and Finance* 20, pp. 377-387.

Katib, M.N and Matthews, K. 1999. A non-parametric approach to efficiency measurement in the Malaysian banking system. *The Singapore Economic Review* 44, pp. 89-114.

Krishnasamy, G., Ridzwa, A.H and Perumal, V. 2003. Malaysian post merger banks' productivity: application of Malmquist productivity index. *Managerial Finance* 30, pp. 63-74.

Kumbhakar, S.C. 1987. The specification of technical and allocative inefficiency in stochastic production and profit frontiers. *Journal of Econometrics* 34, pp. 335 – 348.

Kumbhakar, S.C and Lovell, C.A.K. 2000. *Stochastic Frontier Analysis*. Cambridge University Press.

Kwan, S. H. 2003. Operating performance of banks among Asian economies: An international and time series comparison. *Journal of Banking and Finance* 27, pp. 471 – 489.

Laeven, L. 1999. Risk and efficiency in East Asian banks. *World Bank*, Policy Research Working Papers 2255.

Lang, G and Welzel, P. 1996. Efficiency and technical progress in banking: Empirical result for a panel of German cooperative banks. *Journal of Banking and Finance* 20, pp.1003-1023.

Lau, L.J. 1982. On identifying the degree of competitiveness from industry price and output data. *Economic Letters* 10, pp. 93-99.

Leibenstein, H. 1966. Allocative efficiency vs. "X-efficiency". *American Economic Review* 56, pp. 392 – 415.

Lin, S.Y. 1977. Malaysia: money and monetary management since independence 1957 1- 1976. *PhD Thesis*, Harvard University, Cambridge, Massachusetts.

Lloyds-Williams, D.M and Molyneux, P. 1994. Market structure and performance in Spanish banking. *Journal of Banking and Finance* 18, pp.433-443.

Lothgren, M. 2000. Specification and estimation of stochastic multiple-output production and technical inefficiency. *Applied Economics* 32, pp. 1533-1540.

Lovell, C.A.K. 2003. The decomposition of Malmquist productivity indexes. *Journal of Productivity Analysis* 20, pp. 437-458.

Lozano-Vivas, A. 1997. Profit efficiency for Spanish savings banks. *European Journal of Operational Research* 98, pp.381-394.

Maddala, G.S. 2001. *Introduction to Econometrics*. Third Edition. John Wiley & Sons, Ltd.

Mahajan, A., Rangan, N and Zardkoohi, A. 1996. Cost structures in multinational and domestic banking. *Journal of Banking and Finance* 20, pp. 283-306.

Malmquist, S. 1953. Index, numbers and indifferent surfaces. *Trabajos de Estadistica* 4, pp. 209-242.

Maudos, J. 1998. Market structure and performance in Spanish banking using a direct measure of efficiency. *Applied Financial Economics* 8, pp. 191-200.

Maudos, J and Pastor, M. 2003. Cost and profit efficiency in the Spanish banking sector (1985-1996): a non-parametric approach. *Applied Financial Economics* 13, pp. 1-12.

Maudos, J., Pastor, J.M., Perez, F and Quesada, J. 2002. Cost and profit efficiency in European banks. *Journal of International Financial Markets, Institutions and Money*, 12, pp. 33-58.

McKillop, D.G., Glass, J.C and Ferguson, C. 2002. Investigating the cost performance of UK credit unions using radial and non-radial efficiency measures. *Journal of Banking &Finance* 26, pp.1563-1591.

McKillop, D.G., Glass, J.C and Ward, A-M. 2004. Cost efficiency, environmental influences and UK credit unions, 1991 to 2001. mimeo

Meeusen, W and van den Broeck, J. 1977. Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18, pp. 435 – 444.

Mendes, V and Rebello, J. 1999. Productive efficiency, technological change and productivity in Portuguese banking. *Applied Financial Economics* 9, pp. 513 – 521.

Mester, L.J. 1996. A study of bank efficiency taking into account risk-preference. *Journal of Banking and Finance* 20, pp.1025-1045.

Miller, S.M and Noulas, A.G. 1996. The technical efficiency of large bank production. *Journal of Banking and Finance* 20, pp. 495-509.

Mitchell, K and Onvural, N.M. 1996. Economies of scale and scope at large commercial banks: Evidence from the Fourier flexible functional form. *Journal of Money, Credit and Banking* 28, pp. 178-199.

Molyneux, P. 1993. Market structure and profitability in European banking. *Research Papers in Banking and Finance* RP 93/9. Institute of European Finance. University College of North Wales, Bangor.

Molyneux, P and Seth, R. 1998. Foreign banks, profits and commercial credit extension in the United States. *Applied Financial Economics* 8, pp.533-539.

Molyneux, P and Forbes, W. 1993. Market structure and performance in European banking. *Research Papers in Banking and Finance* RP 93/5. Institute of European Finance. University College of North Wales, Bangor.

Molyneux, P., Thornton, J and Lloyd-William, D.M. 1996. Competition and market contestability in Japanese commercial banking. *Journal of Economics and Business* 48, pp. 33-45.

Molyneux, P., Lloyd-William, D.M and Thornton, J. 1994. Competitive conditions in European banking. *Journal of Banking and Finance* 18, pp. 445-459.

Mukherjee, K., Ray, S.C and Miller, S.M. 2001. Productivity growth in large US commercial banks: The initial post-deregulation experience. *Journal of Banking and Finance* 25, pp. 913-939.

Noy, I. 2004. Financial liberalization, prudential supervision, and the onset of banking crisis. *Emerging Markets Review* 5, pp. 341-359.

Okuda, H and Hashimoto, H. 2004. Estimating cost functions of Malaysia commercial banks: the differential effects of size, location and ownership. *Asian Economic Journal* 18, pp. 233-259.

Panzar, J.C and Rosse, J.N. 1987. Testing for "monopoly" equilibrium. *The Journal of Industrial Economics* 35, pp. 443-455.

Parkin, M., Powell, M and Matthews, K. 2000. *Economics*. Fourth Edition. Addison-Wesley

Pastor, J.M., Perez, F and Quesada, J. 1997. Efficiency analysis in banking firms: An international comparison. *European Journal of Operational Research* 98, pp. 395-407.

Pilloff, S.J and Rhoades, S.A. 2002. Structure and profitability in banking markets. *Review of Industrial Organization* 20, pp. 81-98.

Pindyck, R.S and Rubinfeld, D.L. 1998. *Econometrics models and economic forecasts.* Fourth Edition, McGraw-Hill International Editions.

Ray, S.C and Desli, E. 1997. Productivity growth, technical progress and efficiency change in industrialised countries: Comment. *The American Economic Review* 87, pp. 1033-1039.

Resti, A. 1997. Evaluating the cost-efficiency of the Italian Banking System: What can be learned from the joint application of parametric and non-parametric techniques. *Journal of Banking and Finance* 21, pp. 221-250.

________. 1998. Regulation can foster mergers, can mergers foster efficiency? The Italian case. *Journal of Economics and Business*, 50, pp. 157-169.

Rezvanian, R and Mehdian, S. 2002a. An examination of cost structure and production performance of conventional banks in Singapore. *Journal of Banking and Finance* 26, pp. 79-98.

__________. 2002b. A comparative multiproduct cost study of foreigned-owned and domestic-owned US banks. *Applied Financial Economics*, 12, pp. 271-284.

Roger, K.E. 1998. Nontraditional activities and the efficiency of US commercial banks. *Journal of Banking and Finance* 22, pp.467-482.

Sathye, M. 2001. X-efficiency in Australian banking: An empirical investigation. *Journal of Banking and Finance* 25, pp.613-630.

Sealey, C and Lindley, J. 1977. Inputs, outputs and a theory of production and cost at depository financial institutions. *Journal of Finance* 32, pp. 1251-1266.

Shaffer, S. 2004. Comment on "what drives bank competition? Some international evidence" by Stijn Claessens and Luc Laeven. *Journal of Money, Credit, and Banking* 36, pp. 586-592.

Shaffer, S. 2002. Conduct in a banking monopoly. *Review of Industrial Organization*, 20, pp. 221-238.

Shanmugam, K.R and Das, A. 2004. Efficiency of Indian commercial banks during the reform period. *Applied Financial Economics* 14, pp. 681 – 686.

Sharma, K. R., Leung, P and Zaleski, H.M. 1997. Productive efficiency of the swine industry in Hawaii: Stochastic frontier vs. data envelopment analysis. *Journal of Productivity Analysis* 8, pp. 447 – 459.

Siegal, S and Castellan, N.J. 1988. *Nonparametric statistics for the behavioral sciences*. Second Edition, McGraw-Hill International Editions.

Siems, T.F and Barr, R.S. 1998. Benchmarking the productive efficiency of U.S. banks. *Financial Industry Studies*, Federal Reserve Bank of Dallas.

Stiglitz, J. 1994. The role of the state in financial markets. In: Bruno, M and Pleskovic, B. eds. *Proceedings of the World Bank Annual Bank Conference on Development Economics 199*".

________. 2002. *Globalization and its discontent*. Penguin Books.

Stiroh, K.J. 2000. How did bank holding companies prosper in the 1990s? *Journal of Banking and Finance* 24, pp. 1703-1745.

Sturm, J-E and Williams, B. 2004. Foreign bank entry, deregulation and bank efficiency: Lessons from the Australian experience. *Journal of Banking and Finance* 28, pp. 1775 – 1799.

Tahir, I.M. 1999. Market structure and efficiency in ASEAN banking. *PhD Thesis*, University of Wales, Bangor.

Tavares, G. 2002. "*A bibliography of data envelopment analysis (1978-2001)*", Rutcor Research Report, Rutgers University, USA.

Tobin, J. 1984. On the efficiency of the financial system. Lloyds Bank Review 153, pp. 1-15.

Toolsema, L. 2002. Competition in the Dutch consumer credit market. *Journal of Banking and Finance* 26, pp. 2215-2229.

Tortosa-Ausina, E., Grifell-Tatje, E., Armero, C and Conesa, D. 2003. *Sensitivity analysis of efficiency and Malmquist productivity indices: An application to Spanish savings banks*. Working Paper 2003/02, Centre For Applied Economic Research, The University of New South Wales.

Uchida, H and Tsutsui, Y. 2005. Has competition in the Japanese banking sector improved? *Journal of Banking and Finance* 29, pp. 419-439.

Verbeek, M. 2000. *A guide to modern econometrics*, John Wiley and Sons, Ltd.

van der Vennet, R. 2002. Cost and profit efficiency of financial conglomerates and universal banks in Europe. *Journal of Money, Credit and Banking* 34, pp.254-282.

Vives, X. 2001. Competition in the changing world of banking. *Oxford Review of Economic Policy* 17, pp. 535-547.

Weill, L. 2004. Measuring cost frontier in European banking: A comparison of frontier techniques. *Journal of Productivity Analysis* 21, pp. 133-152.

Wheelock, D.C and Wilson, P.W. 1999. Technical progress, inefficiency and productivity change in U.S. banking, 1984-1993. *Journal of Money, Credit and Banking* 31, pp. 212-234.

Williams, B. 2003. Domestic and international determinants of bank profits: Foreign banks in Australia. *Journal of Banking and Finance* 27, pp. 1185 – 1210.

Worthington, A.C. 1998a. The determinants of non-financial institution efficiency: A stochastic cost frontier approach. *Applied Financial Economics* 8, pp. 279-287.

________. 1998b. Efficiency in Australian building societies: An econometric cost function approach using panel data. *Applied Financial Economics* 8, pp. 459-467.

________. 1999. Malmquist indices of productivity change in Australian financial services. *Journal of International Financial Markets, Institutions and Money* 9, pp. 303-320.

Yildirim, H.S. 2003. Competition and contestability in Central and Eastern European banking markets. *PhD Thesis*, The University Of Tennessee, Knoxville, Tennessee.

Yildirim, C. 2002. Evolution of banking efficiency within an unstable macroeconomics environment: the case of Turkish commercial banks. *Applied Economics* 34, pp. 2289-2302.

Zhu, J. 2003. *Quantitative models for performance evaluation and benchmarking*. Kluwer Academic Publishers.