DISSERTATION

THE EFFECTS OF MERGERS ON BANK EFFICIENCY: STOCHASTIC RISK-RETURN FRONTIER ANALYSIS

Submitted by

Victor Tshisikhawe Munyama

Department of Economics

In partial fulfillment of the requirements

for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2004

COLORADO STATE UNIVERSITY

June 14, 2004

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY VICTOR TSHISIKHAWE MUNYAMA ENTITLED "THE EFFECTS OF MERGERS ON BANK EFFICIENCY: A STOCHASTIC RISK RETURN FRONTIER ANALYSIS" BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Dr. Nancy Jianakoplos

Dr. David Mushinski

Dr. Richard Johnson

Dr. Sherrill Shaffer

Adviser: Dr. Ronnie Phillips

Department Head

ABSTRACT OF DISSERTATION

THE EFFECTS OF MERGERS ON BANK EFFICIENCY: STOCHASTIC RISK-RETURN FRONTIER ANALYSIS

This dissertation is motivated by the need to evaluate the impact of mergers and acquisitions on bank efficiency. Over the past two decades, significant research effort has been devoted into measuring the efficiency of financial institutions, especially commercial banks. Even though many studies found that measured inefficiencies in these financial institutions were high (at least 20% of the total banking industry cost and about half of the industry's potential profits), there is still no consensus about the sources of the differences in measured efficiencies. Differences might be due to various efficiency concept employed, and different measurement methods used to estimate efficiency within each concept. There is also not an agreed theory as to the potential efficiency correlates.

This dissertation assesses bank efficiency under the following considerations: alternative profit efficiency concept, stochastic frontier analysis, flexible translog functional form, generalized normal-truncated normal distribution, and explicit model for inefficiency effects.

Chapter 1 provides background of the study, statement of the problem and hypotheses to be tested. This chapter also outlines the mergers and acquisitions trend that significantly contributed towards the reduction of the number of banks in the U.S.

from a maximum of 14,496 banks during 1984 to 7,887 banks during 2002. These mergers and acquisitions were propelled by the changing market conditions, regulatory changes, and globalization of financial institutions across the U.S. borders.

Chapter 2 put this study into context in terms of the previous literature on bank efficiency. Financial ratio studies concluded that mergers were not associated with significant improvements in cost ratio or any other measure of firm performance. Of the X-efficiency literature, studies that examined merger benefits from the cost or input side also found no benefits on average from mergers. These studies led to several implications that could pave way for future research: first, studies on the implications of bank mergers should move away from simple financial ratios analysis. Second, studies should not only determine factors that predict merger efficiency gains or losses, but should also determine factors that explain differences in measured efficiencies among banks. Third, studies should employ the profit function rather than the cost function in all merger efficiency analysis. Last, following Shaffer (1993) and Berger and Humphrey (1992a) assertion, studies should analyze if merger efficiency effects are time-dependent.

Chapter 3 not only presents the theoretical framework underlying the efficiency studies but also present the analytical foundation of producer theory and efficiency measurement. This chapter explicitly presents a stochastic frontier framework adopted in the study and two distributional assumptions (normal-half normal and normal-truncated normal distributions). The normal-truncated normal distribution, due to its flexibility and its accommodation of a wider range of distributional shapes, which includes ones with nonzero modes, is an adequate representation of the data in this study rather than the normal-half normal distribution adopted by other bank efficiency studies.

Chapter 4 presents the specification of the profit frontier and the inefficiency effects models. This chapter presents the alternative profit efficiency concept and its specification. The alternative profit efficiency concept bridges the gap between the standard cost and profit functions, and is more relevant if the assumptions underlying the cost efficiency and the standard profit efficiency do not hold. This chapter also presents an explicit model for inefficiency effects. Most empirical studies investigated the determinants of technical inefficiency among firms in the banking industry by regressing the predicted inefficiency effects observed from an estimated frontier upon a vector of firm-specific factors in a second-stage analysis.

This study outlines the shortcomings of the two-stage analysis thereby presenting a single-stage maximum likelihood technique that simultaneously estimate the frontier function and the inefficiency effects model. We also model the inefficiency error component, ξi, as independently distributed such that ξ_i is obtained by truncation (at zero) of the normal distribution with mean $Z_i\delta$ and variance σ^2 rather than the normal-half normal distribution adopted by other previous bank efficiency studies.

Chapter 5 presents the empirical results of the study. After estimating the profit function, we found that the inefficiency component should be incorporated into the profit frontier. That is, the auxiliary equation (inefficiency effects model) plays an important role in the estimation of the profit function. This also suggests that we should concentrate on maximum likelihood estimation. We found that the flexible translog is a better representation of the data than the Cobb-Douglas form. In addition, the normal-truncated normal distribution is an adequate representation of the data than the normal-half normal distribution. The measured efficiencies were higher compared to inefficiencies summarized in Berger and Humphrey (1997). We also found that the inefficiency effects variables maintain their explanatory power on average when measuring inefficiency with respect to both risk and return. Also, banks that increased their risk-adjusted performance improved their alternative profit thereby becoming more technically efficient. The diversification hypothesis also holds.

Chapter 6 presents the summary and conclusion of the study. This study indicated that mergers and acquisitions that take advantage of the diversification opportunities after a merger improve their efficiency. However, we take into consideration that this was a cross-sectional study. Therefore, future studies along these lines should employ a panel data to test the robustness of our results and methodology.

Victor Tshisikhawe Munyama
Economics Department
Colorado State University
Fort Collins, CO, 80523
Summer 2004

ACKNOWLEDEMENTS

I approached this dissertation work with a mixture of trepidation and excitement. However, it was a lot of support and help from many people that put me at ease. My sincere thanks and appreciation goes to my money and banking Professor, my advisor, and chairman of my graduate committee Professor Ronnie J. Phillips. He provided me with a lot of support and guidance that not only stimulated my inquiry of the subject, but also gave me the freedom to do so at my own pace.

I also would like to express a special thanks to Dr. Sherrill Shaffer of the University of Wyoming and a member of my graduate committee. He did not deprive me of his invaluable banking knowledge and experience, an area he is well read. His advice transcended the technical and the formalities of this dissertation work. I am forever grateful for all the time and conversations we had. My graduate committee members, Dr. Nancy Jianakoplos (Dr. J.), Dr. David Mushinski, and Dr. Richard Johnson not only provided positive and constructive comments and suggestions to improve this work, but also showed a lot of interest in the overall progress of my graduate studies. I also would like to thank Dr. Robert Kling who not only made it easy and welcoming for me when I arrived in Fort Collins, but whose support

and interest went beyond the classroom setting. I also would like to thank Dr. Robert DeYoung of the Federal Reserve Bank of Chicago for answering questions on the Call Report data. It would have been impossible to be in Fort Collins, Colorado without the understanding and prayers of my parents, Johannes Ndishavhelafhi and Susan Nnyadzeni. They, together with my siblings, not only provided me with a lot of encouragement, but also believed that I can. My thanks also go to David and Debbie Williams for making my family a significant part of their family and our stay in Fort Collins a pleasant one.

Last, but not least, I would like to thank my wife Vusiwana Rose, my son Nkateko Tshisikhawe, and my daughter Vhonani Tintswalo for the invaluable sacrifices and understanding. Their love, support and patience have always been the inspiration every day.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

1.1: Background

The recent global changes in the structure of financial services industries have motivated observers to determine the cost and revenue efficiency of the evolving financial institutions (especially commercial banks). Researchers have tried to link the improved performance in financial industry to such factors as advancement in financial technologies, new financial engineering tools, advanced statistical techniques, etc. Within the banking industry, studies are still trying to find out how these factors might affect the performance of commercial banks.

The improved or worsened efficiency of these financial institutions have a significant impact on such issues as profitability, taxpayer financial assistance in case of industry bailouts due to substantial losses, prices and quality of services delivered to consumers, and the safety and soundness of the financial system. Because of changing market conditions, changing regulatory environment, and the globalization of the financial institutions, commercial banks have strived to improve their competitiveness by adopting new technologies and services, improving the quality of their services and the delivery process, rapid transmission of information, and better evaluation of risk.

Substantial research effort into measuring the efficiency of financial institutions (particularly commercial banks) found large inefficiencies, about 20% to 25% of total banking costs, and about half of the industry's potential profits. Research also found that less than 5% of costs were attributed to scale and product mix inefficiencies (Berger, Hancock, Humphrey 1993). Berger and Humphrey (1997) surveyed and contrasted results of 130 financial institutions efficiency studies that presented different average efficiency levels both across measurement techniques and across countries. They found that the mean inefficiencies for nonparametric techniques were about 39% (35% median), while the mean inefficiencies for the parametric techniques were about 19% (18% median). The nonparametric studies suggested a greater dispersion in estimated ratios.

The past decade was also characterized by a significant wave of mergers and acquisitions in the U.S. banking industry. Cross-border mergers following the European Community Integration also thrust this merger wave. This shifted the focus to the efficiency implications of bank mergers, especially mergers between large institutions. There is a substantial research effort that is trying to address the question of whether mergers and acquisitions (M&As) in the banking industry lead to efficiency improvement. This also follows the Department of Justice's acceptance of efficiency defense in merger cases during 1984 (U.S. DOJ 1984).

The growing interest in M&As concerns the consequences for consumers and the industry at large. From a public policy standpoint, interests are in the tradeoff between possible gains in operating efficiency versus possible social efficiency losses from a greater exercise of market power. That is, will a particular merger have the potential to fundamentally restructure an industry such that there are fewer firms and the industry output is increasingly concentrated in smaller number of large firms?

Studies using the 1980s data found little or no cost efficiency gains on average associated with M&As. However, banks continued to merge at a dramatic pace further making observers curious. Were these unsatisfactory conclusion a result of employing a cost function framework (that assume that the bank is trying to minimize costs and that managerial mistakes are made in the input usage) rather than the profit function framework? Afterall, recent studies indicated that the analysis of profit efficiency (not cost efficiency) is more appropriate for evaluating M&As. The period under study is also a question under investigation. Structural changes in banking (e.g. changes in the goals of M&As participants, lifting of regulatory restrictions, differences in market conditions, global consolidation, etc.) suggest another closer look at the performance issues especially employing the 1990s data rather than the 1980s data.

This study extends the past efficiency analysis of M&As by applying the stochastic risk-return frontier based on a translog profit function. We also examined bank mergers of all sizes during 1997. We not only measure whether mergers improve profit efficiency on average, but also try to determine some factors that might explain differences in measured efficiencies between banks. We modify the basic frontier estimation approach to incorporate the effects of conditions that may be associated with inefficiency. We, therefore define a stochastic frontier alternative profit function for cross-sectional data on merged U.S. commercial banks, in which the non-negative technical inefficiency effects are assumed to be a function of firm-specific variables. We also assume that the inefficiency effects are independently distributed as truncation of the normal distribution with constant variance, but means that are a linear function of observable firm-specific variables (rather than the mostly applied normal-half normal distribution). We, therefore, test if the simpler normal-half normal distribution is an adequate representation of the data, given the specification of the generalized normal-truncated normal distribution. We apply the test only to the merged banks during 1997.

1.2: Statement of the Problem

The past several years have seen substantial research effort directed into measuring efficiency of financial institutions (particularly commercial banks). There is, however, not a generally accepted theory as to the sources of differences in measured efficiency. The question we should ask is *whether these differences are due to the type of efficiency concept used or variation in efficiency measurement methods.*

Studies on bank efficiency have also been motivated by a significant wave of mergers and acquisitions since the mid-1970s. The most frequently cited motivation for bank mergers is that they improve performance by cutting costs and/or capturing economies of scale. As such, a significant percentage of research has surveyed the bank efficiency with respect to costs and/or cost efficiency improvement from mergers. However, there is little evidence to suggest that mergers systematically yield operating efficiencies for the banking firms. This is irrespective of the evidence that merging banks experiences gains in profits largely by expanding their revenue streams.

These efficiency effects of mergers constitute a core policy question since merger applicants often cite prospective efficiency benefits as a justification for merger approval. Research that only used cost data could not find evidence of average efficiency benefits from mergers compared to research that included the revenue effects. Therefore, it is important to continue to consider an application of the profit function in merger efficiency analysis. That is, *we need to employ the profit efficiency concept*, which include the cost efficiency effects of mergers and acquisitions and also incorporate the revenue effects of the changes on output that occur after a merger.

Finally, apart from evidence of no efficiency benefits, on average, from bank mergers, studies also found that some mergers were successful and others were not. It is not only important to look at conditions that might not work in predicting efficiency benefits, but also ask a question: *what are the determinants of bank efficiency?* That is, we need to explore some potential efficiency correlates that not only help predict merger success but also help explain differences in measured efficiency across banks.

1.3: Objectives of the study

In investigating and measuring efficiency, this study has the following objectives:

1. To measure the technical efficiency of the merged banks during 1997. In addition, we shall conduct a generalized likelihood-ratio test to determine whether these banks are perfectly efficient or not.

2. To evaluate the performance of these merged banks using a stochastic frontier approach. We shall also calculate the overall alternative profit function for the period under study.

There are also various methodological matters to consider in estimating technical efficiency. These include, but not limited to, the choice of a functional form, distributional assumptions of the model, validity of efficiency results that rely on cross-sectional data, etc. Apart from these methodological matters, this study has another specific objective in relation to methodology used:

- We shall model the potential inefficiency effects on a one-stage analysis rather than the two-stage analysis normally used in banking efficiency studies.

3. Upon investigating or analyzing the sources of profit efficiency, we shall also draw policy implications based on these efficiency estimates that can indicate both the distribution of bank efficiency and market environment.

Previous studies analyzed the cost efficiency of banks (bank mergers). However, most studies indicated that we should be analyzing bank efficiencies based on the profit function (especially when analyzing bank mergers - Berger, Hunter, and Timme 1993). In employing the alternative efficiency concept, this study uses the stochastic frontier approach based on the flexible translog functional form in analyzing the effects of mergers on bank efficiency. Previous studies also employed a normal-half normal distribution when modeling the inefficiency term. However, this study employs a more general normal-truncated normal distribution, which proved to be an adequate representation of the data.

This study explicitly present a model for inefficiency effects rather than follow a two-stage analysis adopted in other bank efficiency studies. We also adopted a maximum likelihood technique to, simultaneously, estimate the profit frontier and the inefficiency effects model. Previous studies only incorporated a measure of risk as an efficiency correlate in the second-stage regression. However, this study incorporates a measure of risk in both the profit frontier and the inefficiency effects model.

1.4: Hypotheses

The main purpose of this dissertation is to investigate the effects of mergers on U.S. commercial banks efficiencies. We shall focus our attention to the following questions:

1. Does the merged bank show any evidence of inefficiencies? We shall conduct a generalized likelihood-ratio test to test the null hypothesis that these merged banks are completely efficient against the alternative hypothesis that they are not efficient.
2. Does the average degree of measured efficiency improve (or otherwise change) when we correct for variations in bank risk? Previous studies found that banks on average are very inefficient with respect to profit. This should be considered surprising, given the large number of banks and other evidence of strong competition facing any one bank. We shall assess the measured efficiency that incorporates some measure of risk in each bank and compare with estimated efficiencies in other previous studies on bank efficiencies.

3. The Diversification Hypothesis: As the banking organizations increase in size (through mergers), do their risk-return tradeoff improve, because of better diversification of portfolio risk? Merging firms might be able to hold a higher loan/asset ratio (ceteris paribus) due to an improved diversification of loan risks. That is, increased lending should lead to higher output thereby improving efficiency. We, therefore, expect to find an improved risk-return tradeoff in merged banking organization due to better diversification of portfolio risks.
4. Do other variables retain their explanatory power when we measure efficiency with respect to both risk and return? Previous studies found several variables that help explain profit efficiency (e.g. ROA, bank risk, Total Costs/Total Assets, Operating Cost/Total Assets, etc.). We shall inspect and model variables that are hypothesized to have potential effect on measured efficiencies or differences thereof.

We estimated the profit frontier and tested the hypothesis that the inefficiency component is not present in the model against the alternative that it should be incorporated into the profit frontier. We rejected the null hypothesis in favor of the alternative that the inefficiency component should be incorporated into the profit frontier model. Not only should the inefficiency component be incorporated into the profit frontier model but this auxiliary equation or the technical inefficiency equation can explain the difference between each firm's profit and the profit on the profit function.

In line with the "value added approach" (Berger and Humphrey 1992b) we adopted to define and measure bank outputs and inputs, we found that demand deposits and consumer loans were positive and statistically significant in determining profit efficiency. That is, an increase in the portfolio of both demand deposits and consumer loans add value into the firm. Contrary to the value added approach, an increase in the business loans and the real estate loans lead to a decline in alternative profit or depletes the value of the firm. In this study, we corrected for variation in bank risk by incorporating a measure of risk (Altman Z-score that measures the probability of bankruptcy). We found that banks that improve their risk-adjusted performance improve their alternative profit.

We reestimated the profit frontier simultaneously with the inefficiency effects model. We tested the hypothesis that inefficiencies do not exists in these merged banks against the alternative that inefficiencies exist. We found that inefficiencies exist. Also, the measured efficiencies were higher than measured efficiencies in the previous studies on bank efficiencies. We also found that the generalized normal-truncated normal distribution is an adequate representation of the data than the simpler normal-half normal distribution.

In testing whether the Cobb-Douglas form was a better representation of the data given the flexible translog functional form, we found that the flexible translog was a better representation of the data. We also rejected the hypothesis that the coefficients of the explanatory variables in the model for the technical inefficiency are equal to zero or have the same truncated normal distribution (that is, the inefficiency effects are not a linear function of the δ_is). This result implies that the joint effects of these variables are significant although the individual effects of one or more variables may not be statistically significant.

When testing the diversification hypothesis, we found that the results were consistent with the diversification hypothesis (i.e. as banking organizations increase is size – through mergers – their risk-return tradeoff should improve because of better diversification of portfolio risk). We also tested if the inefficiency effects variables maintain their explanatory power when compared with what other previous studies on bank efficiency hypothesize. We obtain mixed results, which might be due to different methodologies adopted in this study and other previous studies. That is, this study adopted a single stage maximum likelihood technique in estimating the profit frontier and the inefficiency effects model (a technique not adopted by other previous studies on bank efficiencies).

1.5: Mergers and Acquisitions Trend

For the past two decades, the U.S. banking industry experienced major structural changes. At the core of these structural changes is the record number of banks that consolidated (in terms of both frequency and the size of the merging institutions). Figure 1.1 illustrates changes in the number of commercial banks during the period 1934-2002. The graph shows a remarkable level of stability in the banking industry from 1934 to mid-1980s. However, from the mid-1980s the number of insured commercial banks began to fall dramatically. The number of insured commercial banks peaked at 14,496 during 1984. However, that number declined to 9,143 during 1997 and 7,887 by the end of 2002.

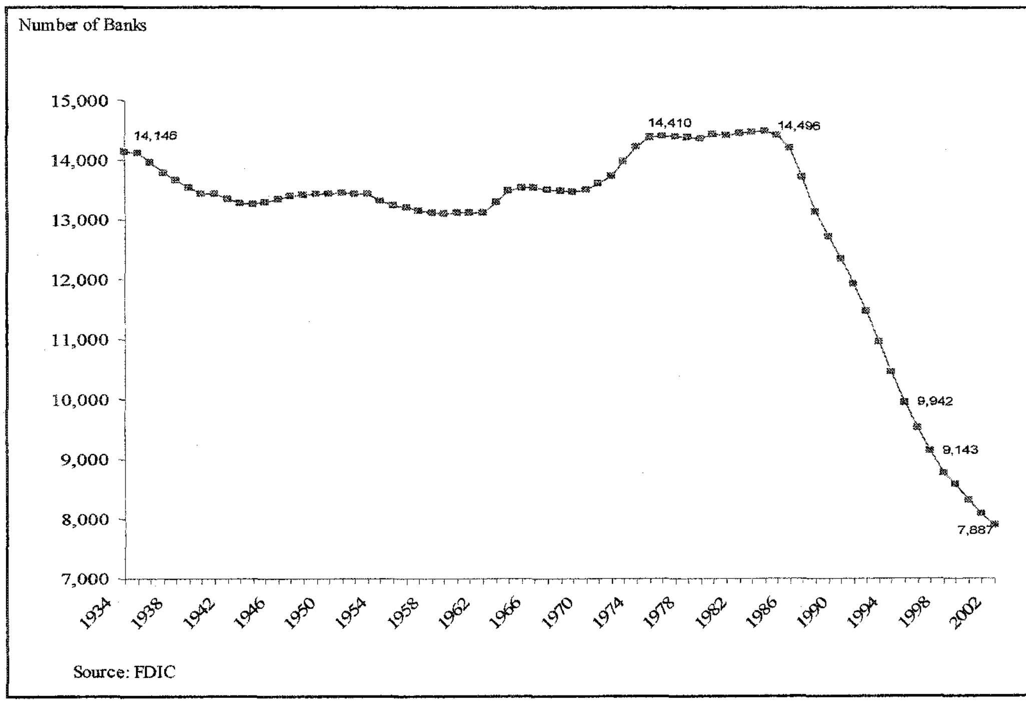


Figure 1.1: Changes in the Number of FDIC-Insured Commercial Banks 1934-2002

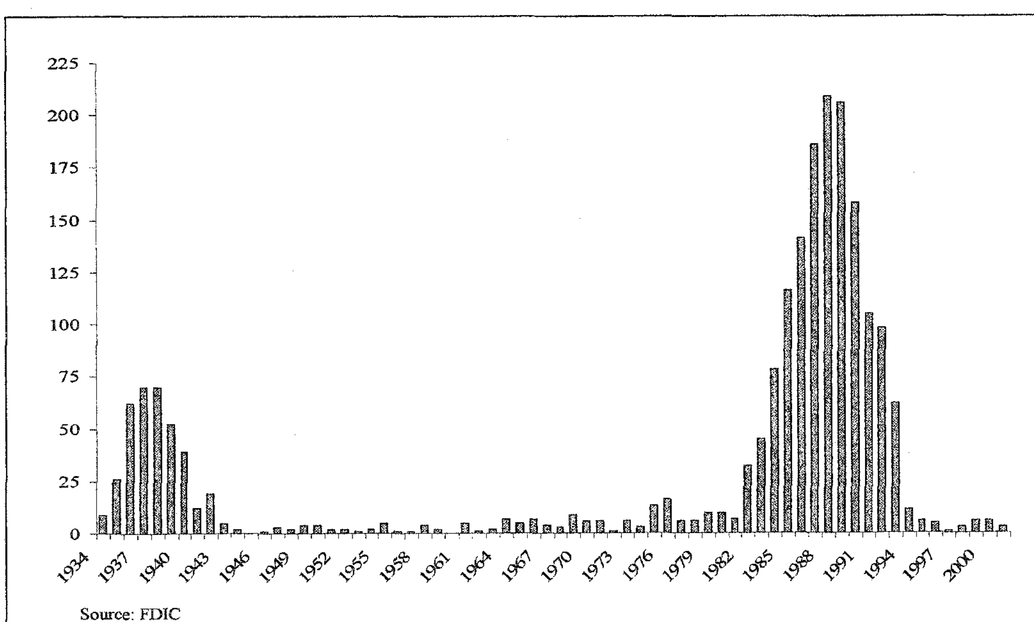


Figure 1.2: Number of Bank Failures 1934 - 2001

The banking industry also experienced major problems during the 1980s and early 1990s. This is evident from the number of bank failures during that time. Bank failures were also common during the period before 1934, a period were depositors suffered tremendous losses. However, between 1943 and 1981 bank failures only averaged about 15 a year for commercial banks and less than 5 a year for savings and loans. Figure 1.2 above shows that from early 1980s, the stable environment in the banking industry changed as bank failures started increasing dramatically.

It should be noted that bank failures was not the sole reason for the disappearance of banks in the U.S. Between 1980 and 2000, banks also disappeared through mergers and acquisitions (M&As) at a rate of about 1.75 per business day. During the period 1985-1992, the number of commercial banks declined by about 3000, which was more than double the number of bank failures thereby attesting to the increased rate of M&As that increased dramatically throughout the 1980s and 1990s. A further look at the pattern of these M&As shows their significant contribution to the disappearance of commercial banks in the U.S. According to the FDIC, there were 8,386 unassisted mergers and approximately 1,100 assisted mergers between 1980 and 1999. Figure 1.3 shows a significant increase in unassisted mergers from the early 1980s throughout the 1990s.

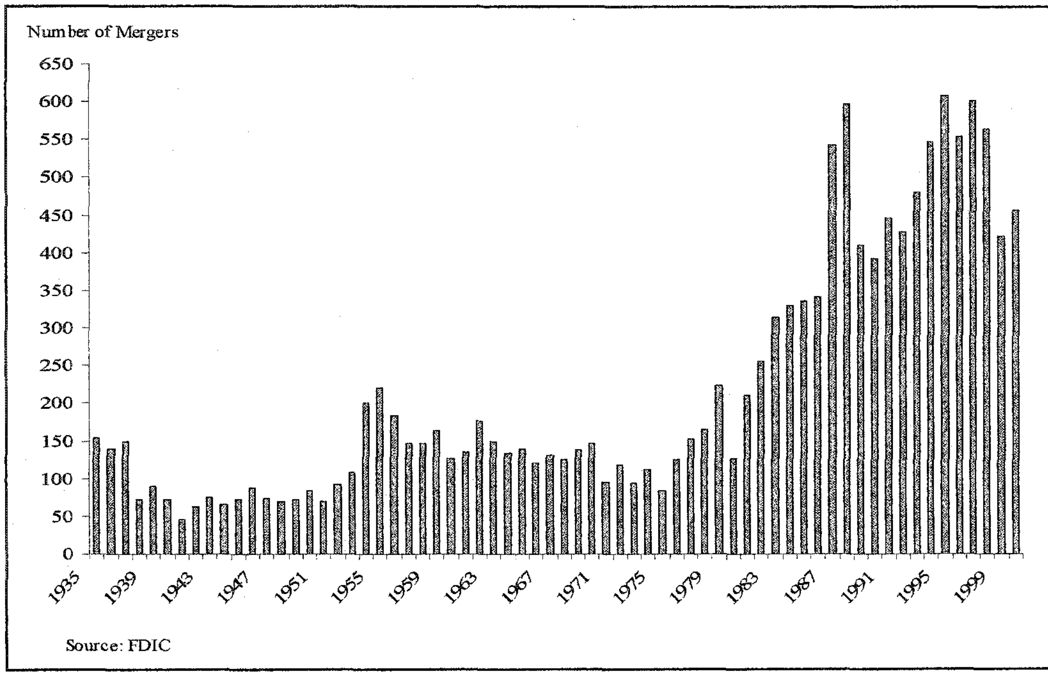


Figure 1.3: Unassisted Mergers 1935-2001

Of significance in this merger wave is not only the number of banks merging but also the size of the mergers and the total assets involved. Before 1980, almost no merger occurred where the acquiring and acquired banks had assets in excess of $1 billion. However, the period 1980s and 1990s saw some of the largest mergers in banking history especially during the 1990s. More banks with assets greater than $1 billion were acquired during the 1990s than during the 1980s. The mergers during the period 1980-1999 involved about $2.5 trillion in assets, which was more than half the U.S. total banking asset. About 50% of this $2.5 trillion in acquired assets during this period was acquired during 1995-1999 period. Figure 1.4 below show estimates of the U.S. bank mergers and bank assets acquired during the period 1980-1998.

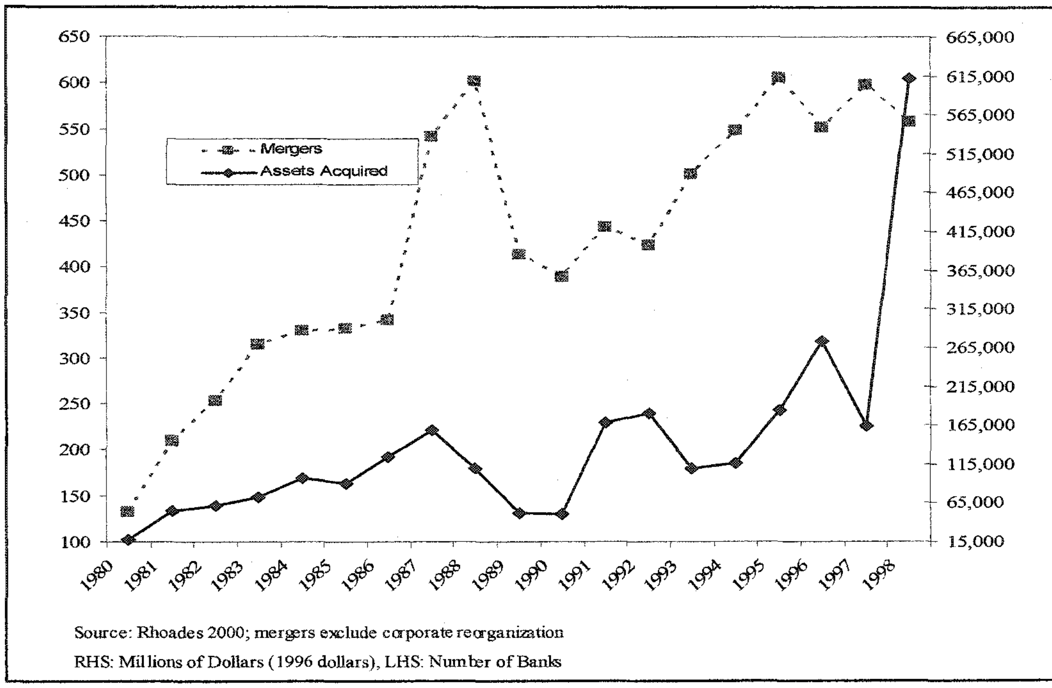


Figure 1.4: U.S. Commercial Bank Mergers and Assets Acquired 1980 - 1998

There has also been an increase in the number of megabanks in the U.S. During 1992, there were 51 banks with assets greater than $10 billion, and their combined assets constituted about 40% of the U.S. total banking assets. By the end of 2000, that number has increased to 83 banks controlling about 70% of the total U.S. banking assets. Figure 1.5 shows the increased concentration of banking assets in the largest banks between 1980 and 1998. Moore and Siems (1998) indicated that, based on market values, nine of the ten largest M&As in the U.S. history in any industry occurred during 1998, and four of these (Citicorp-Travelers, BankAmerica-NationsBank, Bank One-First Chicago, Norwest-Well Fargo) occurred in the banking industry.

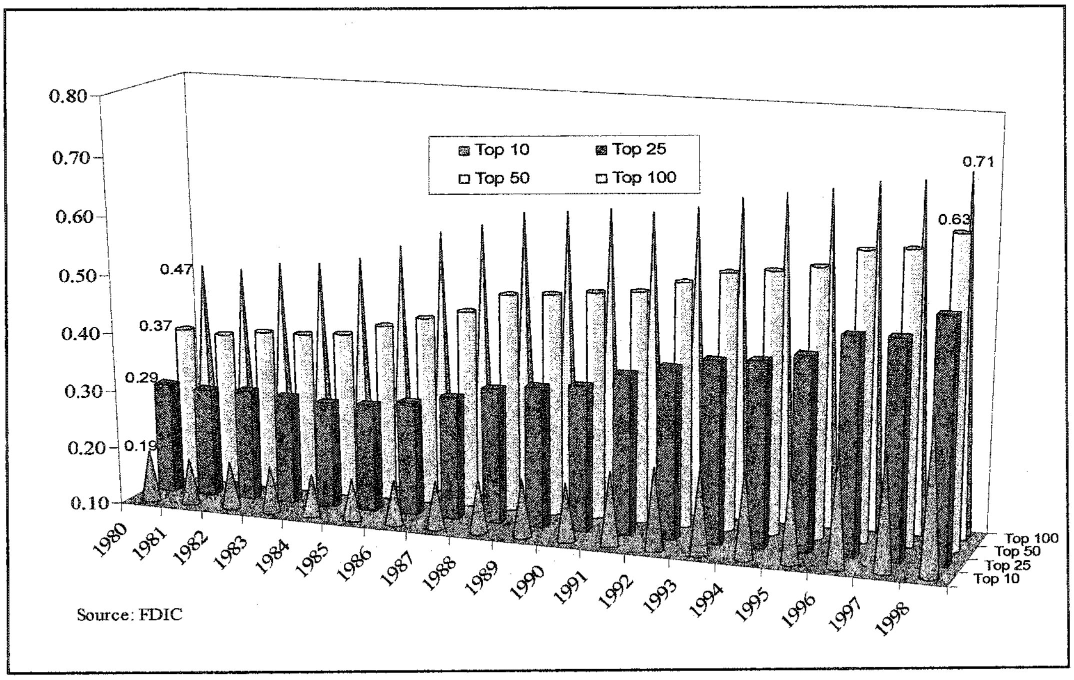


Figure 1.5: Concentration of Banking Assets in the Largest Banks in the U.S. 1980 - 1998

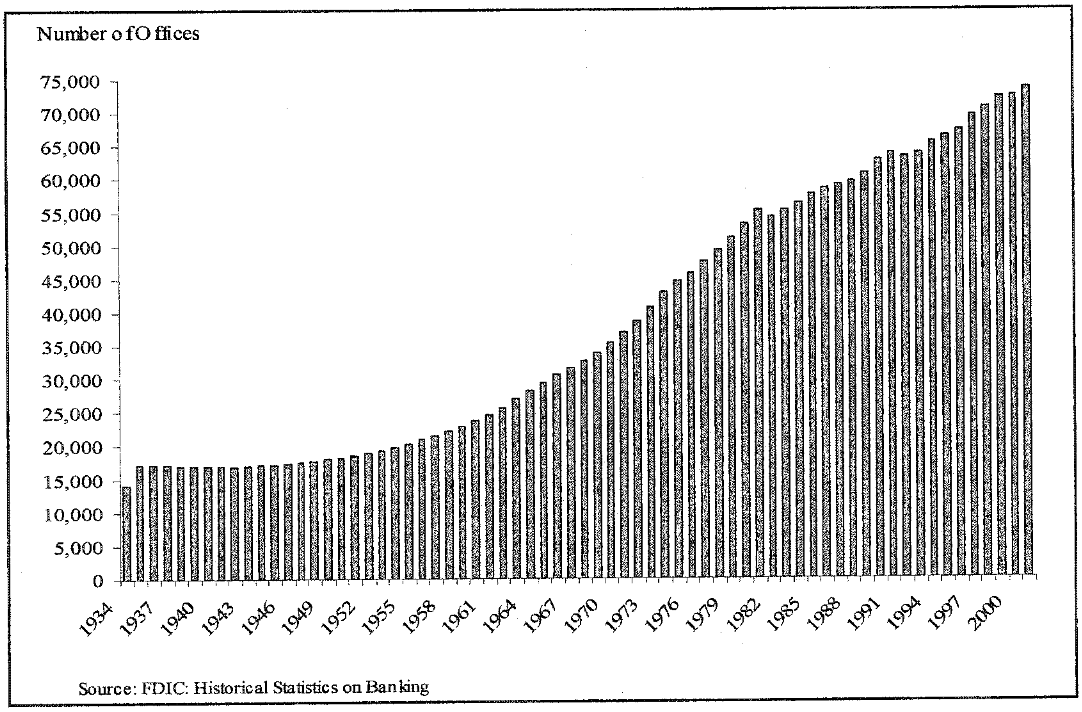


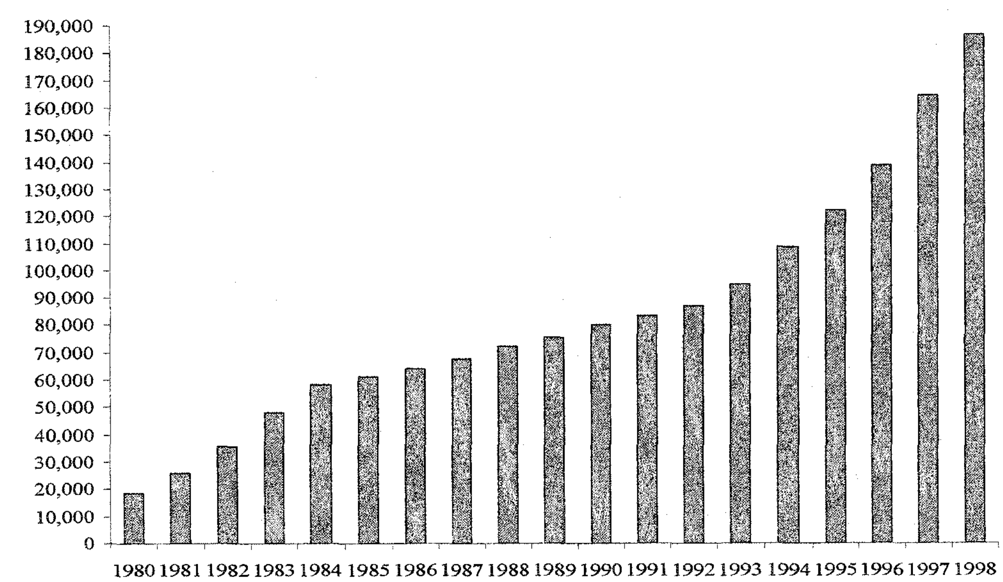
Figure 1.6: Offices of FDIC-Insured Institutions 1934 - 2001

The industry structural changes had a significant impact on such issues as competition, economic performance of the industry, quality of product and services, prices, and efficiency of production. Apart from the decrease in the number of banks across the U.S., and the large increase in the nationwide concentration of banking assets and deposits in the largest banks, there were also a surprisingly increase in the number of banking offices (figure 1.6 above).

These structural changes in the U.S. banking industry also raised several questions about the causes, consequences, and the future implications to the financial services industry. That is, why are banks merging at such a dramatic pace; will these mergers enhance efficiency in banking; what are the consequences to consumers, industry, economy etc.

We can look at three broad explanations for these merger waves: Changing market conditions, regulatory changes, and globalization or consolidation of financial institutions across border. The **changing market conditions** or economic environment tends to not only alter the constraints faced by firms within a particular industry but also accelerate the pace of consolidation. Some of the changing market factors that contributed to the accelerated pace of consolidation in the banking industry include the technological progress, improvements in the financial conditions, excess capacity, etc. The financial services industry experienced a significant technological progress that not only increased scale economies in producing financial services, but also improved the delivery and access to these financial services.

Larger institutions, faced with new financial engineering tools (derivative instruments, off-balance sheet guarantees, and risk management tools) were able to broaden the scope of their products. This is also evident in the explosion of automatic teller machines (ATMs), and on-line banking. From about 2,000 ATMs during 1973, the number increased to 18,000 during 1980, and 187,000 during 1998. Figure 1.7 below shows the increasing number of ATMs during the period 1980-1998.



Source: Rhoades 2000

Figure 1.7: Number of ATMs 1980 - 1998

Berger et al. (1999) indicated that the scale economies in banking significantly increased during the 1990s, which was consistent with technological progress that favored larger institutions. During the 1970s, commercial banks earnings declined significantly as evident in figure 1.8 below. Due to a combination of macroeconomic adversity and regulatory structures, banks could not easily borrow low and lend high, as was the case before. Figure 1.8 also shows a decline in banks' interest margin (i.e.

interest income relative to interest cost) between 1970 and 1981. On the other hand, figure 1.9 shows a significant increase in interest expenses during the early 1980s.

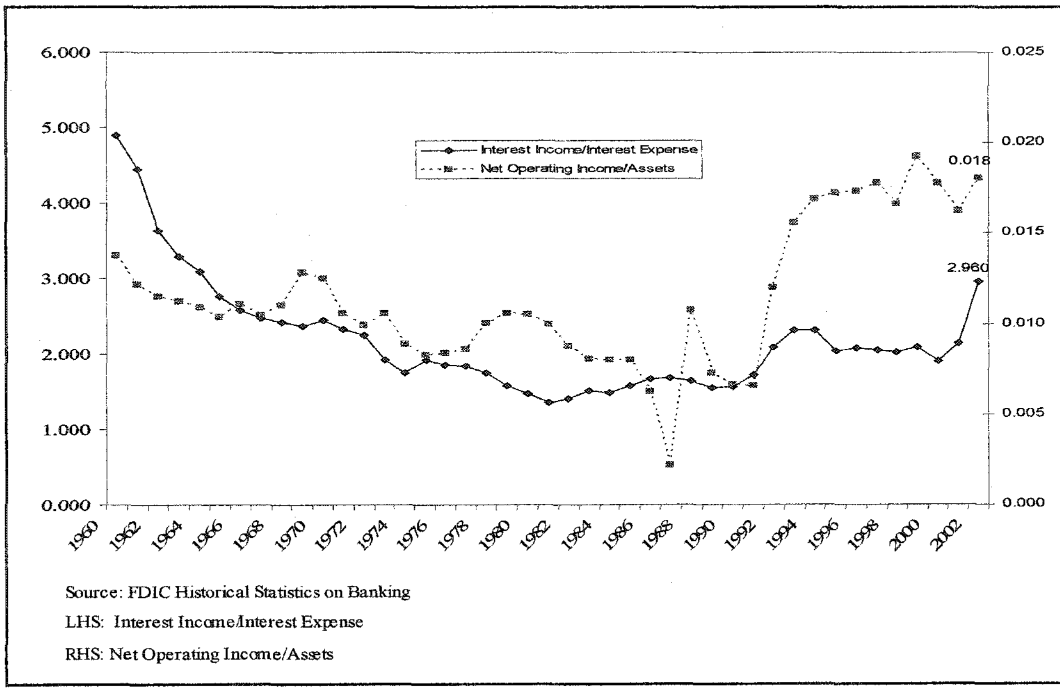


Figure 1.8: Commercial Banks Earnings Trend 1960 - 2002

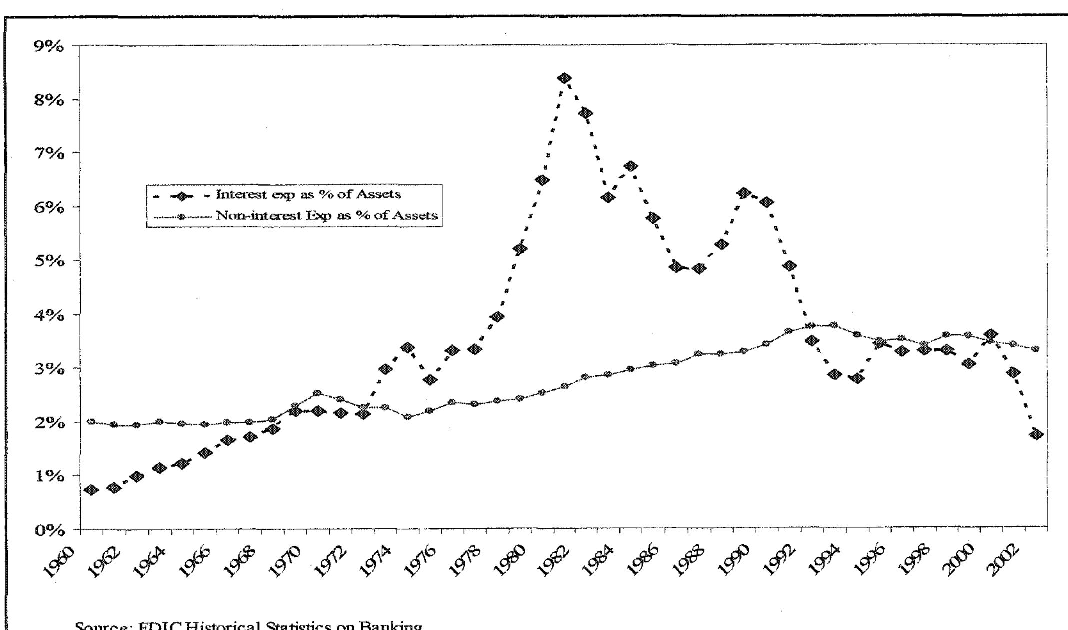


Figure 1.9: Bank Interest and Non-interest Expenses relative to Assets 1960 - 2002

During this distressed period of the 1970s and 1980s, commercial banks also experienced an excess capacity problem because of lost market share to other competing financial institutions on both sides of the balance sheet (Berger et al. 1999). On the asset side, a significant proportion of corporate lending was lost to foreign banks, other intermediaries and the capital markets. The innovation of off-balance sheet activities did not fully compensate for this loss. On the liability side, the boom of such products as mutual funds posed a direct and intense competition to the banking industry. This was also promulgated by the fact that households began holding a greater portion of their wealth outside of bank deposits. However, market conditions changed drastically during the 1990s. Banks experienced a low interest rates and high stock prices environment. As Berger et al. (1999) indicated, M&A activity in banking responded positively to this low interest rates environment during the mid-1990s than any M&A activity in nonfinancial industries.

The analysis above does not imply that changes in these market conditions or economic environment lead to improved or worsened efficiency in banks. It only hypothesizes that these changes might have accelerated the M&A activities within the banking industry. The acceleration of M&A activities was either a regulatory solution (assisted mergers) to assist troubled insolvent institutions or was an individual firm's initiative to address the impact of changing market environment (unassisted mergers).

Another major factor behind the banking merger wave was the changing **regulatory structure** or deregulation of the banking industry. Figure 1.10 below illustrates the timeline of regulatory changes in the U.S. banking industry.

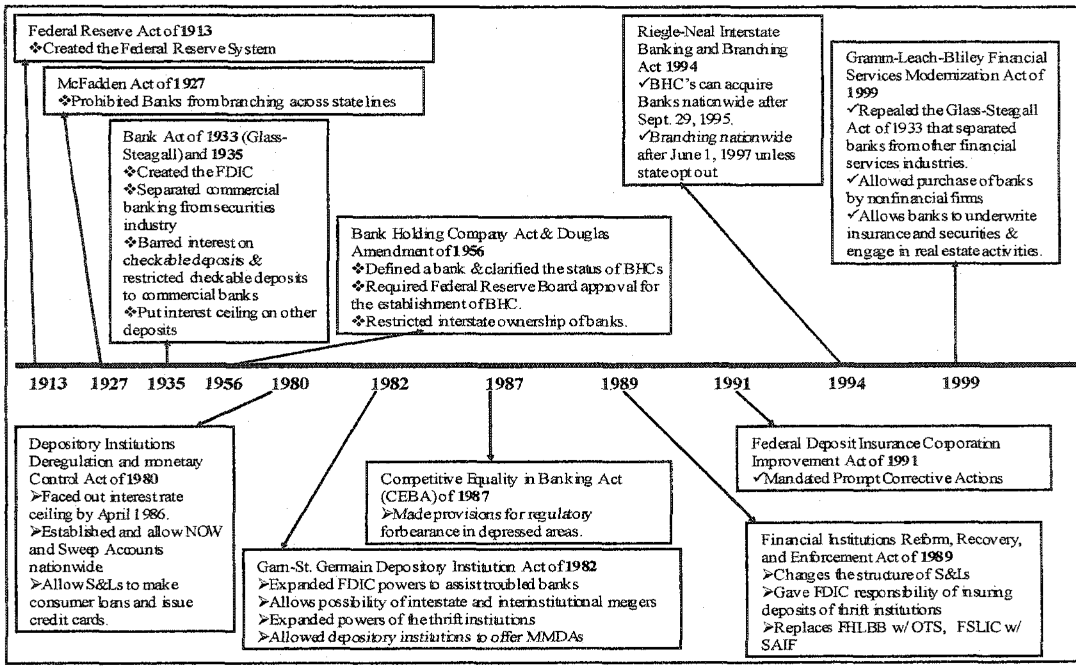


Figure 1.10: Historical Banking Regulatory Developments in the US

Following the banking system's distress during the 1970s and early 1980s, the industry regulators initiated a process of massive deregulation. The Depository Institution Deregulation Act of 1980 provided the banks and thrifts with more freedom to compete with nonbank financial firms and eliminated deposit-rate limits over a period of years. However, the thrifts institutions did not benefit immediately as they fell victim of high inflation rates that resulted in high short-term interest rates (hence leading to massive thrift failures).

The regulators then turned towards market solutions to curb the problem of massive failure. A thrift deregulation Act (Garn-St Germain Act of 1982) gave thrifts more freedom to make loans, purchase funds, participate in underwriting and capital provision, attract deposits, buy and sell in financial markets, etc. Deposits insurance levels were also increased following the Depository Institutions Deregulation and

Monetary Control Act of 1980. Regulators also eased or ignored rules against inter-state banking and market concentration because the dollar volumes of failed thrifts were more than the deposit insurance funds could absorb. This facilitated some viable depository institutions to take over failing or insolvent depositories.

During 1982 and 1984, the Federal regulators also loosened the anti-trust guidelines. Of significance were the changes in the merger guidelines during 1984. During that time, the M&A guidelines were still based on section 1&2 of the Sherman Act of 1890, Section 7 of the Clayton Act of 1914, and section 5 of the Federal Trade Commission Act of 1914. Factors considered in examining mergers were market definition (product and geographic market), market share and concentration, and ease of entry. The most applied factor was the market share and concentration, which was determined by the Herfindahl-Hirschman Index (HHI)[1].

However, during 1984, the federal regulators incorporated a significant change in the merger guidelines. For the first time, the department of justice allowed the antitrust authorities to consider efficiencies as a factor in evaluating merger cases. That is, if convincing evidence existed that the gains from merger would not be achieved but for mergers. Apart from the failing company defense, changing market conditions defense, and the ease of entry into the market defense, banks were now able to present an efficiency defense as a case for merger. Overall, a combination of the banking crisis during the 1980s and the first wave of regulatory overhaul facilitated bank-merger wave as authorities seek to find ways to dispose off insolvent depositories under a 'least cost to the taxpayer' criterion.

---

[1] A complete exposition on the HHI can be found in Rhoades (1993a)

The regulatory authorities assisted in many merger cases during the 1980s. However, as regulation changes it became easier for banks to facilitate mergers without assistance. In addition, many banks began to use mergers to expand their fee-generating multiproduct to customers. As indicated in figure 1.3 above, unassisted mergers increased dramatically during the 1980s and 1990s. During the 1980s, these unassisted mergers accounted for about 1 percent of the bank population every year. However, by the mid-1990s unassisted mergers eliminated more than 50% of the bank population.

One of the primary forces underlying the sustained merger waves during the 1980s and throughout the 1990s was the gradual removal of state and federal restrictions on geographic expansion in banking. Of significance was the passage of Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994 that permitted interstate branching in almost all states. The Act allowed nationwide interstate banking through holding company banks by June 1995, and allowed interstate branch banking by September 30, 1997. This legislation opened more opportunities for bank mergers. Figure 1.11 illustrates the interstate and intrastate bank mergers during the period 1985-2001.

It is evident that before 1995 there were insignificant interstate mergers whereas intrastate mergers constituted more than 90% of all mergers. However, interstate mergers started increasing as a percentage of all mergers, from 4.3% during 1995 to 31% during 1997 and 33% during 2000. At the same time, intrastate mergers started to decline after 1995. In addition, the number of interstate branches operated by commercial banks increased following the enactment of Riegle-Neal Act (figure 1.12).

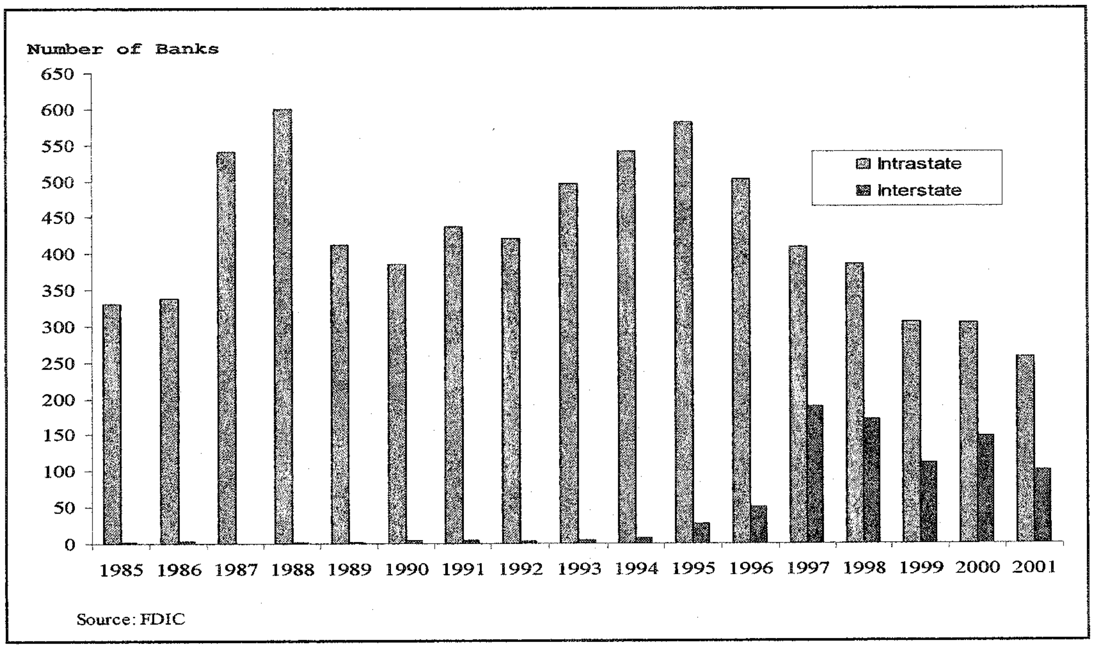


Figure 1.11: Interstate vs. Intrastate Mergers 1985 - 2001

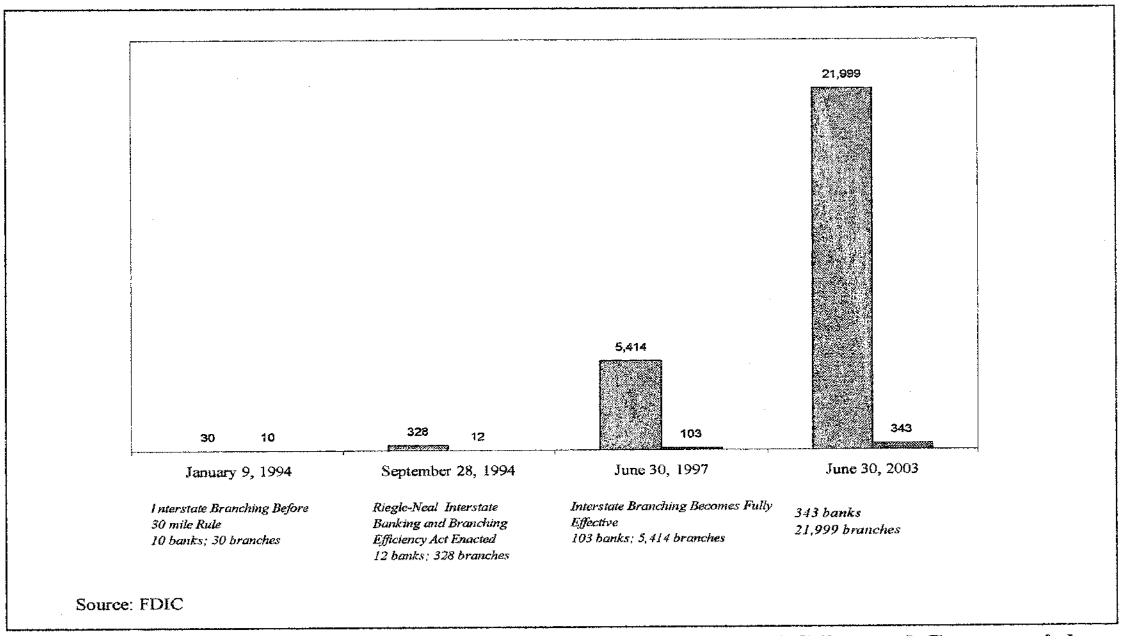
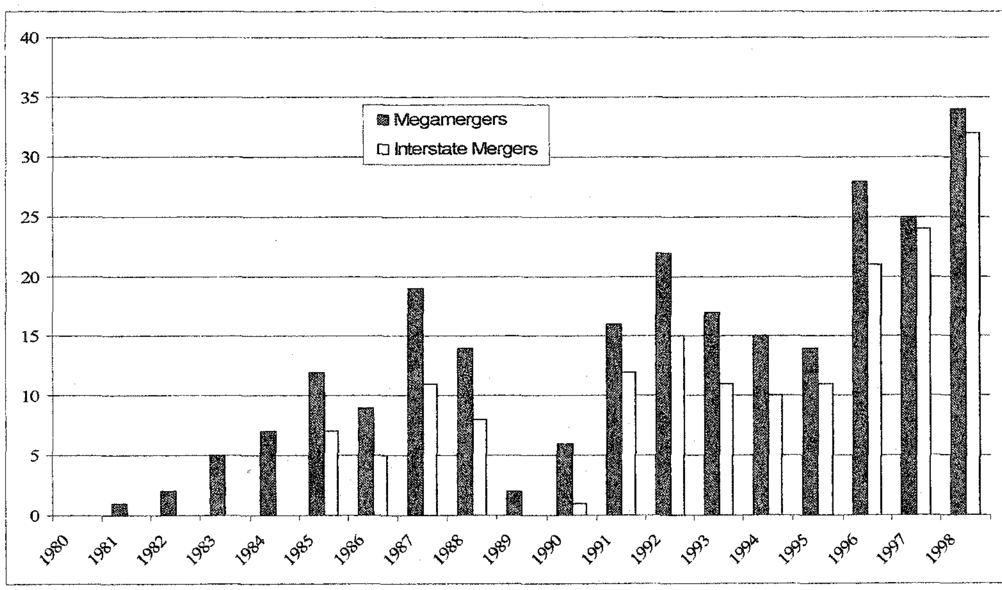


Figure 1.12: Number of Interstate Branches Operated by FDIC-Insured Commercial Banks (1994 – 2003)

Within these mergers, more megamergers (mergers of banks with assets greater than $1 billion) occurred during the 1990s than during the 1980s. Figure 1.13 above shows that during 1990s a number of large interstate mergers (megamergers) overshadowed the increase in the number of total megamergers.



Rhoades 2000
Mergers exclude corporate reorganizations

Figure 1.13: Megamergers – Interstate Mergers 1980 - 1998

Lastly, the **globalization of financial institutions** also contributed to the continued merger wave in the banking industry. Apart from the Riegle-Neal Act of 1994 in the U.S., two other pieces of legislation motivated the global consolidation of the financial institutions: the Single Market Programme in the European Union, and the Gramm-Leach-Bliley Act of 1999 in the U.S. During the 1990s, many banks decided to pursue an upscale retail banking strategy (Berger et al. 1999). These banks focused on integrated financial market operations, where they can aggressively expand their range of financial services. As such, there was a new wave of M&As between large banking organizations and the investment banks, insurance companies, and brokerage houses. The pace of these types of consolidations between banks and nonbanking firms increased dramatically following the passage of the Gramm-Leach-Bliley Act of 1999 that allowed commercial banks to be purchased by nonbank financial firms.

The consolidation of the financial services industry around the globe also intensified competition thereby leading to more and more M&As. Reduction in trade barriers, declines in transportation costs, communication technology advancements, and acceleration in international economic integration are other contributing factors. Most of the cross-border M&As involve large universal type institutions that provides multiple types of financial services in multiple nations. Such megamergers as the Deutsche Bank-Bankers Trust provided a leading European universal bank greater access to wholesale commercial and investment banking resources in the U.S. This stiffened the competition within the U.S. banking industry and many banks decided to merge to cope with this new competition.

1.6: Synopsis of the Study

The dissertation is organized as follows:

Chapter 1 gives a brief background of the study. It also provides an explanation and justification of the study. In addition, it provides a summary of the statement of the problem, the hypotheses, and the objectives of the study. The chapter also highlights the structural changes in the U.S. baking industry especially the trend on M&As.

Chapter 2 places the study in context in terms of the previous literature on bank efficiency. It also summarizes the existing literature on the effects of bank mergers on efficiency under the financial ratio methods, parametric, and nonparametric methods.

Chapter 3 is devoted to the theoretical framework or foundation underlying efficiency studies. It not only characterizes the production technology with production frontier but also provide some definitions of efficiency. It also discusses various approaches and specific methods of measuring efficiency. The stochastic production frontier is also discussed with two distributional assumptions. Finally, this chapter also gives a brief discussion on the profit frontier and profit efficiency.

Chapter 4 discusses the methodology used to analyze the problem stated in the study. Summary statistics of the data is also presented. It also discusses the concept of efficiency pertaining to this study. It also outlines the model specification and the procedure of analyses as applied in the study. The chapter also specifically addresses the issue of modeling inefficiency effects.

Chapter 5 is devoted to the presentation and discussion of the empirical results. It provides various tests for testing the presence of inefficiency, functional forms, etc. of the model. It also presents and discusses the efficiency estimates. Finally, this chapter discusses the sources of technical efficiency as estimated.

Chapter 6 provides the summary and conclusions. That is, it summarizes the main findings of the study and present suggestions for future research.

CHAPTER 2: REVIEW OF THE LITERATURE

2.1: Introduction

We can divide studies on the effects of mergers on bank performances into two broad categories. First, the financial ratio studies that only compared the pre- and post-merger financial ratios. These studies not only looked at the effects of past mergers but also simulated the effects of the potential future mergers. The financial ratio studies mostly concluded that mergers were not associated with significant improvements in cost ratios or other measures of firm performance. Second, the X-efficiency studies that estimates the best practice frontier of banks. These studies measure the degree of inefficiency of banks in the sample relative to the best practice technology. They estimate the cost frontier and/or the profit frontier in analyzing bank performances and such issues as the effects of mergers on bank performance.

Concerning mergers and acquisitions (M&As), by the early 1990s research focused purely on share price effects of announcements. These studies focused on stock market return around the time of acquisition announcements. They also confirm that shareholders of target banks benefited, but did not shed any light on whether total shareholder wealth was improved. The question of whether banks can improve performance by increasing scale through a merger remained unanswered.

2.2: FINANCIAL RATIOS STUDIES

Financial Ratio Studies focused on past mergers and simply compared the pre- and post-merger financial ratios (e.g. operating costs divided by total assets, etc) to evaluate firm performances. These accounting based ratios are still a traditional tool employed by bank analysts to measure firm performances.

Rhoades (1986) examined 413 acquired banks over 1968-1978 and found no significant efficiency benefits from Megamergers. Rhoades (1990) examined the ratio of operating cost to asset before and after acquisition for 13 acquired banks, that each had over $1 billion in assets over 1981-1987. He concluded that there was no significant performance difference in these mergers relative to nonacquired banks three years before or three years after the merger.

Linder and Crane (1992) addressed the question of whether commercial bank mergers improve performance. Their research was motivated by the Treasury department's proposal that commercial banks should be allowed to merge their subsidiaries across state lines. They analyzed annual data of the before-and-after merger performance of all commercial banks mergers in the New England region during the period 1982-1987. If one dominant bank merged with two or more target banks in one calendar year, it was treated as one merger. Their sample consists of 47 mergers that resulted from 81 target banks and 34 different dominant or surviving banks. There were also 22 holding companies. No single dominant bank assets accounted for more than 8.5% of the transactions, and no bank holding companies participated in more than 15% of the transactions (measured in terms of assets).

Linder and Crane (1992) also separated mergers of newly acquired banks from intra-holding company mergers (i.e. mergers of sibling bank subsidiaries that were acquired at least one year earlier). Therefore, of the 47 mergers, 25 were intra-holding company mergers (corporate reorganization). Twenty-two were mergers of newly acquired banks in which the target bank was acquired and merged into the dominant bank during the same calendar year. The authors then compared the performance of merging banks before the transaction to that of the combined entity after the merger. To assess the impact of mergers, the authors focused on operating income and its three components: net interest income, noninterest income, and noninterest expenses. First, they expressed these variables relative to the book values of assets. Secondly, they compared the growth of each variable for merging banks to that of the comparable industry group.

Excluding the intra-holding company mergers, Linder and Crane (1992) found that, on aggregate, mergers of newly acquired banks did not improve performance relative to their competitors in the first two years after a merger (whether performance was measured by operating return on asset or by growth in operating income). The intra-holding mergers improved net interest margins more effectively than newly acquired banks but were less likely to control overhead expenses. In both types of mergers, reduction in expenses did not keep pace with the loss of assets.

Srinivasan and Wall (1992), performed similar analysis as Linder and Crane (1992) using a larger and more general sample. They compared operating cost and performance ratios up to 4 years after the merger. Their sample consisted of 240 bank holding companies with at least $100 million in assets during the period 1982-1986.

The sample also represented combinations from across the entire country. The authors found that the ratio of operating expenses to assets started out significantly low or more efficient than a national composite of banking companies that did not merge during the same period. After the merger, the median operating ratio rose significantly and ended up closer to the median of the national composite. Overall, the authors found that there was no significant cost savings relative to the industry.

However, Srinivasan and Wall's (1992) data was at the holding company level. Therefore, their analysis did not look at the actual units merging nor directly address the question of what happens when two or more banks merge. For instance, an acquired bank that remained as a separate legal entity could not produce the same results as the one that loses its identity or become a branch of the holding company after the merger.

Srinivasan (1992) tested the proposition that bank mergers produced significant savings in cases in which costs are measured as noninterest expenses divided by the sum of net interest revenue and noninterest revenue. The author examined the before-and-after performances of all bank and bank holding company mergers between 1982 and 1986. Using the year-end data from the Report on Condition and Income (Call Report), the sample period incorporated the effects of deregulation of deposit rates. The sample only included transactions between acquiring and target banks that each exceeded $100 million in assets. The author excluded intra-holding company transactions and acquisitions by foreign bank holding companies.

The expense ratio results suggested that the average bank merger produced small but significant declines in the ratio of noninterest expense to the sum of net interest income and noninterest income. Srinivasan (1992) also found that changes in noninterest expenses at the merging banks were not significantly different from industry wide trends during the same period. That is, there was no significant cost savings relative to the industry.

Contrary to most of the studies above, Cornett and Tehranian (1992) came to a different conclusion. They directly examined the changes in economic performance resulting from bank mergers using pre-and post merger accounting data for 15 large interstate and 15 large intrastate bank acquisitions completed between 1982 and 1987. Their sample consisted of mergers where the purchase price exceeded $100 million, and the target bank's common stock was publicly traded. They collected cash flow and accounting data for the sample firms for the period before and after the merger. The authors also used the operating cash flow divided by the market value of assets to evaluate performance. They examined performance changes for the full sample of bank mergers as well as subsamples of intrastate and interstate mergers.

Cornett and Tehranian's (1992) results showed that the ratio of pretax operating return (excluding interest expenses) to the market value of equity plus book value of subordinated debt increased by 1.2%, which was statistically significant relative to the industry. For the three years before the merger, the sample mean return on asset (ROA) was 0.12% above the industry, and for the three years after the merger, it was 0.19% above the industry. That is, even though this 0.07% increase shows that the merged bank outperforms the industry, it was not statistically different from zero. That is, there was no improvement in the ROA, which includes interest expenses. However, Cornett and Tehranian (1992) found that the return on equity (ROE) underperformed the industry by 1.8% before the merger but outperform the industry by 2.1% after the merger. The 3.9% performance increase was statistically significant at 0.05 significant level.

Spindt and Tarhan (1992) corrected for the change in the financial ratios for other firms in the industry to avoid confounding secular change within the industry with the efficiency effects of mergers. They examined a sample of 394 mergers that occurred during 1986 and found that the median difference in ROE between merging and nonmerging banks was negative and statistically significant in the two years before the merger, but was positive and statistically significant in the two years after the merger. The average asset size of the target bank in their sample was $25 million, while their consolidated institutions averaged about $130 million. They found that there were some net average merger benefits on the revenue or output side.

Overall, these studies that simply compared the cost ratios, suggested that the potential for cost efficiency improvement from mergers was generally not realized. There was no evidence of a substantial change in cost performance associated with bank mergers. These cost ratio studies also have limited ability to assess the potential for cost improvements from mergers or the likely efficiency implications of future mergers because of some methodological shortcomings. It is inappropriate to compare the cost ratios of two different banks. Even when comparing a bank's cost ratio to the average cost ratio of a peer group, one need to take extra precautions in constructing such a peer group. One need to ask a question on how identical the two banks are especially with respect to bank size, market conditions, product mix, and other factors that might affect banks expenses.

Berger and Humphrey (1992a) explored some of the limitations of the ratio studies. First, most financial ratio studies exclude interest expenses in their analysis. The performance ratio analysis derives its conclusion from the use of operating cost ratio only. About 70% of all bank costs consist of interest expenses, which often represent operating costs expended elsewhere in the financial system to raise funds (e.g. when a bank pays interest expenses to borrow the excess core deposits of another bank through the federal funds market). Operating expenses account for about 30% of the total bank costs. Larger banks often rely on purchased funds for much of their financing and there are interest expenses incurred in borrowing or raising this type of financing. On the other hand, smaller banks might not be too keen on purchased funds like the larger banks. These small banks might incur more operating expenses due to their method of raising funds. Therefore, operating cost comparison might not give us a clear picture as to the efficiency of these banks.

Second, operating cost comparisons might produce results bias towards showing false benefits from mergers. Berger and Humphrey (1992a:560) indicated that, "as banks grow larger, they tend to substitute interest cost-intensive purchased funds for operating cost-intensive core deposits". The bank's capital structure tends to shifts towards purchased funds when core deposits decline more than the assets. This will lead to a decline in the operating cost ratio, which can be mistaken as an increase in efficiency. Cornett and Tehranian (1992) study outlined above excluded interest expenses in their analysis thereby concluding that there was a significant improvement in the bank performance after the merger in relation to the industry. However, such a conclusion can be misleading unless we analyze both the total operating plus the interest expenses.

Third, the measurement of firm efficiency is a concept that cannot easily be captured by mere accounting ratios. Banks do not necessarily operate under a perfect competitive environment. Beyond the traditional banking activities (taking deposits and making loans), different banks may sell different products, which are all packaged differently depending on the location of the banks (rural, urban, small town, larger city), attitude towards risk, management style, etc. The cost-to-asset ratio assumes that a dollar of assets should entail the same costs irrespective of the asset type and the cost of doing business in that particular area. To the contrary, banks located in fast-growing markets might incur expenses related to expansion. On the other hand, banks operating in economically depressed regions might have to deal with problem loans that might lead to huge administration costs. As Berger and Humphrey (1992a) indicated, cost ratios do not control for product mix and input price. Therefore, a direct measure of efficiency that econometrically controls for these differences would be appropriate.

In addition, when using the cost ratios, there is no clear distinction between X-efficiency and scale efficiency effects. A clear distinction might be important when trying to determine which mergers are likely to succeed in terms of performance improvements. It is also not clear which mergers are more likely to produce cost savings and which ones are not. All mergers, those where an inefficient bank is acquired by an efficient bank or vice versa, are pooled together. However, various bank studies have already demonstrated that mergers where an X-efficient bank acquires a less X-efficient bank are likely to show some performance improvement. Lastly, most of the financial ratio studies do not take full account of the distinction between within-market and across-market mergers.

DeYoung (1997a) indicated that even though it is easy to construct and use these accounting ratios, one still need to take extra precaution as the interpretation is not straightforward. For example, a bank that excessively cut its costs might destroy service quality, earnings, portfolio quality, etc. Reducing the labor force or cutting down on the number of branches does not guarantee cost savings nor does it indicate that the bank is being run efficiently. On the other hand, spending more on these items does not signal inefficiency. In addition, a bank's expense level might be an epitome of its business strategy or market conditions it faces.

2.3: X-EFFICIENCY STUDIES

X-efficiency or frontier analysis has become a viable alternative to accounting-based efficiency ratios. It also helps solve the methodological limitations associated with ratio analysis. These statistics-based efficient frontier approaches, not only provide a more accurate estimates of cost/profit efficiency, but also helps the researcher to estimate the maximum amount a bank could reduce (increase) costs (profits) while still producing the same amount and combination of financial services. Each bank is also compared with a hypothetical best practice benchmark that corresponds exactly to that bank's individual characteristics, thereby alleviating a problem of constructing a peer group.

Berger and Humphrey (1997) surveyed 130 studies that applied frontier efficiency analysis to financial institutions in 21 countries. They compared average efficiency levels both across measurement techniques and across countries, and outlined the primary results of the many applications of efficiency analysis to policy and research issues. The authors found that efficiency estimates from nonparametric and parametric frontier models are similar. However, the nonparametric methods generally yield slightly greater dispersion. Depository financial institutions experienced an average efficiency of around 77% (median 82%).

2.3.1: Cost and Profit Efficiency

Berger and Humphrey (1992b) measured the output and efficiency in the U.S. banking industry during the 1980s. They used annual bank data for all 14000 U.S. banks for the years 1980, 1984, and 1988, which correspond to periods of pre-, mid-, and post-deregulation of the deposit side of banking, respectively. They measured the X-efficiency of banks using the thick frontier method based on a translog cost function. In specifying the thick frontier, they also specified a separate equation for each of the three types of costs: physical operating costs, interest costs on produced deposits, and interest costs on purchased funds. They also decomposed the standard approach to measuring productivity change into two parts: the frontier shifts and the change in dispersion from the frontier. This enables them to examine the shift in the cost frontier over time so as to assess the effects of deregulation and technical changes associated with deregulation of the deposit side of banking during that period.

Berger and Humphrey (1992b) found that the deregulation of deposits in the early 1980s played a part in raising failure rates (i.e. 1074 banks failed during the period 1981-1989 compared to the postwar decades when fewer than ten banks failed per year), raising the costs directly through the removal of interest ceilings on deposits and increasing the competition among banks. The dispersion in bank costs represented inefficiencies such as different input prices, scale of operations, or product mix, rather than market factors. In all banks, excluding the very largest banks, inefficiencies were mainly operational involving the overuse of physical labor and capital inputs, rather than financial. They also found that low-cost banks remain stable over time, and had the highest profits and lowest probabilities of failure during the 1980s.

Berger and Humphrey (1992b) also found that operating costs of low-cost banks rose over the 1980-1988 period but fell during the 1984-1988 period. However, only the largest banks managed to reduce costs while operating costs continued to increase for smaller banks during the period 1984-1988. In addition, deregulation removed a substantial source of monopsony power over depositors for banks. As such, interest costs increased significantly, eliminating deposits as independent profit center in the process. However, deregulation did not make a significant contribution to net technical change or productivity growth in banking in the 1980s.

Berger, Hancock, and Humphrey (1993) derived efficiency estimates from the profit function applied to data on U.S. banking. They also devise new measures of allocative and technical inefficiencies based on the underlying sources of inefficiency. Their data consists of annual observations from 1984 to 1989 on banks in three different regulatory environments (unit banking, limited branching, and statewide branching). They obtained the data from the Call Report. Only banks that were in existence for the full sample were included. The final data set included three panels of 384 to 599 banks each. The authors specified a Fuss normalized quadratic variable profit function.

Inspecting all size classes of banks, Berger, Hancock, and Humphrey (1993) found that inefficiencies were quite large. The total estimated inefficiencies consumed 70%, 42%, and 46% of the potential variable profits of banks in unit, limited, and state regulatory environments, respectively. A weighted average across all three samples yields an overall 49.1% loss, or about half of all potential variable profits lost due to inefficiency. In terms of the ratio of inefficiencies to total assets, the authors found that banks threw away between 2.6% and 3.7% of their assets on average. This measures the losses in terms of the equivalent of assets being destroyed or the approximate loss of ROA due to inefficiency. The industry return on asset, which averages about 1%, could be raised by several folds if banks were to become fully efficient. They also found that, in all size classes, technical inefficiency dominates allocative inefficiencies, and most efficiency occurred on the output side and not the input side.

Berger, Hancock, and Humphrey (1993) inferred that the dominance of the revenue side may be due to the difficulties in expanding loan outputs in the face of geographic restrictions on branching within and across state lines, and the encroachment of nonbanks into traditional bank lending areas that might also be creating excess capacity in banking. The authors found that large firms are considerably more efficient from both the technical and allocative perspectives than are small and medium firms. They found that banks with over $1 billion in assets tend to be much closer to their optimal netput points, given the prices they face and their fixed netputs. Much of the efficiency advantages for the large banks were on the output side. They also found a negative relationship between inefficiency and bank size.

Berg et al. (1993) considered the relative efficiency of the banking industries in Finland, Norway, and Sweden as observed during 1990. They evaluated the differences in average banking productivity between the three countries, and the spread of efficiency levels with each country. The authors employed the Data Envelopment Analysis (DEA) to assess the productivity and productivity differences. They also used the Malmquist indices to characterize the productivity differences between banks in all three countries. The data set consisted of 503 Finnish, 150 Norwegian, and 126 Swedish banks, and was collected from official bank statistics in Finland and Norway, and the Swedish published annual reports.

Berg et al. (1993) found that efficiency spreads between banks was most important in Finland and Norway, but least important in Sweden. While comparing the best practice frontiers of these countries, they found the highest share of the banking industry on the frontier in Sweden, and the lowest in Finland. Sweden had most of the banks in the Nordic best practice frontier. The average Swedish bank was also more efficient than the Finnish and Norwegian average banks. The Norwegian average bank was more efficient than the average Finnish bank. Looking at the pooled Nordic sample, the authors found that the largest Swedish banks were among the most efficient units, followed by only one large Finnish bank. No large Norwegian bank had efficiency scores above 0.9.

Kaparakis et al. (1994) examined the short-run cost inefficiency of all large U.S. commercial banks. They provided efficiency estimates for a large sample of commercial banks that included banks with assets over $1 billion. They obtained data from the 1986 Call Report for banks with total assets above $50 million, a total of 5548 banks. They divided the sample into nine subgroups of banks according to asset size, and then estimated a stochastic cost frontier based on a translog cost function.

Kaparakis et al. (1994) found that average inefficiencies rise with bank size, except for banks between $5 and $10 billion. For banks over $10 billion in assets, the average inefficiency measure was 17%, which was 70% higher than the average for all banks combined (i.e. 9.8%). The largest banks were two times as inefficient as the most efficient group with assets between $75 and $150 million, where the inefficiency measure was only 9.0%. The largest group, with assets in excess of $1 billion, had the highest inefficiencies of all groups. The authors then did an econometric analysis looking at a set of explanatory variables that might explain some intra-industry variations of these inefficiencies. They looked at portfolio variables, bank structure variables, variables reflecting ease of servicing the market, variables reflecting competition from savings institution, and variables reflecting the financial sophistication of bank customers.

Kaparakis et al. (1994) found that some portfolio variables provided valuable information about the management decisions and operational strategies. That is, the ratio of nonaccrual to total loans had a significant positive effect on bank inefficiency.

As bank's portfolio of loans deteriorates in quality, bank inefficiency rises. They also found that bank inefficiency was positively related to the ratio of purchased funds to total deposits. That is, the more aggressive a bank becomes (i.e. increases the level of purchased funds to total deposits), the more inefficient the operation in the short run. Statewide branching states had, on average, higher inefficiency and unit banking states were more efficient. However, given that a bank operates under particular branching regulations, the more branches that it operates, the more efficient it was.

Kaparakis et al. (1994) found that banks that operated foreign branching networks were even more inefficient. However, if the bank operated foreign branches, the larger the number of branches the more efficient the operation. They also found that as the population density increases, ceteris paribus, inefficiency falls. That is, banks can service a larger number of customers for a given branch network as population density rises. An increase in the number of savings institutions and the percentage of deposits held by saving institutions leads to a rise in bank efficiency. Competition for deposits, either through more savings institutions with a given deposit share, or a larger deposit share for a given number of savings institutions make banks more efficient.

Kaparakis et al. (1994) also found that an increase in the percentage of assets held by savings institutions leads to higher inefficiency in banks. That is, as the percentage of assets (deposits) held by savings institutions rises, holding the percentage of deposits (assets) constant, the total equity in banks falls (rises). Therefore, banks with lower levels of equity tend to be inefficient. That is, inefficient banks tend to have lower levels of equity. Also, an increase in per capita income, holding per capita deposit constant or a decrease in per capita deposit holding per capita income constant means that bank customers are more efficient in deposit use (or they are financially more sophisticated). That is, a rise in per capita income or a fall in per capita deposits reduces bank inefficiency.

Kwan and Eisenbeis (1996) examined the properties of X-efficiencies and the relations of X-efficiency with risk-taking and stock returns of U.S. banking firms. The authors employed a semiannual data for a sample of 254 bank holding companies from 1986 to 1991. They grouped the banks into size-based quartiles to allow for different production technologies for each size class. They obtained the semiannual bank holding company data from the Federal Reserve FRY-9C Bank Holding Company Reports. Their sample consisted of bank holding companies with total consolidated assets of $150 million or with more than one subsidiary bank. They obtained the daily stock price data from the Center for Research in Security Prices (CRSP) at the University of Chicago.

Of the 254 bank holding companies, 174 had complete time-series data from 1986 through 1991. The authors then used the average total assets of the firms with complete time series of observations to sort all the 174 firms into size-based quartiles. The remaining 80 firms with incomplete time-series of observations were then classified into respective size classes using the quartile break points established by the 174 firms at matching time periods.

Kwan and Eisenbeis (1996) then estimated separate cost functions for each size quartile using the method of maximum likelihood. They measured the X-efficiency of individual banking firms using the stochastic frontier efficiency method based on a multiproduct translog cost function. They found that substantial inefficiencies exist in banking, averaging about 10% to 20% of total costs. They then control for scale differences and found that smaller bank holding companies were less efficient than the larger ones, and exhibit larger variations in X-inefficiency. There was also some evidence of a gradual decline in X-inefficiency estimates from 1986 to mid-1990s. They attributed this gradual decline to increased competition due to market and regulatory changes in banking during the 1980s.

On the efficiency and bank risk-taking, Kwan and Eisenbeis (1996) found that X-inefficiency was positively and significantly correlated with the total risk and the idiosyncratic risk of the banking firms, regardless of firm size. X-inefficiency was also negatively correlated with market value capitalization for firms in the first three quartiles (quartile one being the smallest firms and quartile four being the largest firms), and negatively correlated with the book value capitalization for all firms. They also found a significant positive correlation between X-inefficiency and loan charge-offs. Overall, Kwan and Eisenbeis (1996) found strong evidence that X-inefficiency is associated with bank risk-taking behavior. Inefficient firms tend to have higher stock return variance, higher idiosyncratic risk in stock returns, lower capitalization, and higher loan losses. The authors finally found that, for smaller firms, stock returns were negative and significantly related to firm-specific X-inefficiency.

Allen and Rai (1996) examined the global cost function for international banks in order to test for both input and output inefficiencies. They used 1988-1992 data from the Global Vantage database published by Standard and Poor's Compustat Service. They obtained usable data for 194 banks in 15 countries with a total of 785 observations over 5 years. They looked at countries' banking industry in terms of whether they operate functionally integrated banking (universal banking) or functionally separated banking. They also divided the banks into two groups, large and small. They considered a bank large if its asset size falls above the median for all banks in that country. These two classes of banks (large or small) were again grouped under all banks, universal banking, and separated banking.

Allen and Rai (1996) estimated both the stochastic cost frontier model and the distribution free model based on the multiperiod translog functional cost function. They estimated stochastic cost frontiers and subsequent measures of inefficiencies using separate cost functions for universal and separated banks. When analyzing the stochastic cost frontier measures of X-inefficiency, the authors found that all firms display significant levels of input inefficiency ranging from 14.9% for large banks in universal banking countries to 27.5% for large banks in separated markets. However, the universal banks showed significantly lower levels of X-inefficiency than did separated banks. The X-inefficiency for all banks (large or small) in all categories (universal or separated) was the same, averaging around 15% of cost. However, large banks in separated banking countries showed significantly more X-inefficiency than did all other banks.

For their country specific results, countries with relatively efficient universal banks were Australia, Austria, Canada, Switzerland, Germany, Denmark, Spain (for large banks only), Italy (for small banks only) and Sweden. Japan and large banks in Belgium were the only separated banks with above average efficiency levels. Countries with least efficient financial institutions were Finland (X-inefficiency levels of 15.3% and 15.5% for large and small banks respectively), France (15.7% and 26.6%), Italy (20.1%) and UK (16.1%), and the U.S. (29.8% and 17.7%). Allen and Rai (1996) also found that large banks were significantly more (less) X-inefficient than small banks in Australia, Canada, Italy, Japan, and the U.S. (Austria, Switzerland, Spain and France).

Allen and Rai (1996) further examined the empirical regularities in the inefficiency measure by regressing the firm-specific inefficiency measures against various bank and market characteristics (performance, capitalization, portfolio composition, and size). Overall, they found that firms with greater total cost exhibit higher levels of inefficiency for large banks. Greater efficiency was connected with higher profitability. They also found that more aggressive banks (those engaging in more lending activity), tend to be more efficiently operated. Increase in stockholder's equity lead to increase in X-inefficiency for small universal banks and large banks in separated banking countries only. In addition, access to stable base of deposits significantly increases efficiency for small universal banks only, and small banks in all countries showed significant levels of economies of scale.

Berger and Mester (1997) examined possible sources of efficiency differences in financial institutions. Their analysis included differences in efficiency concept and measurement method. They also looked at possible efficiency correlates under various categories, such as the number of banks, market, and regulatory characteristics. They estimated efficiency of 6000 U.S commercial banks that were in continuous existence over the six-year period 1990-1995 and had no missing or questionable data on any of the variables used. They employed three economic efficiency concepts (cost, standard profit, and alternative profit efficiencies).

Berger and Mester (1997) used the distribution-free approach based on the Fourier-Flexible functional form to estimate efficiencies under all three efficiency concepts. They found that cost efficiencies were about 0.868. That is, 13.2% of costs were wasted on average relative to a best-practice firm. Mean efficiencies for the standard and alternative profit functions show that about half of potential profits that could be earned by a best-practice firm were lost to inefficiency. They also found that the alternative profit efficiency ratios were lower on average than the standard profit efficiency ratios, which may be due to service quality or market power. Overall, they found that measurement of each of the efficiency concepts does add some independent informational value. Also, the choices concerning measurement technique and functional form made very little difference in terms of either average industry efficiency or the rankings of individual firms.

While analyzing the potential correlates of bank efficiency, Berger and Mester (1997) found mixed results. Costs efficiency estimates did not vary much across bank size classes. That is, cost efficiencies were 2.5% higher at the largest banks than the smallest banks. However, profit efficiency for small banks showed the greatest level of efficiency. Also, banks in holding companies tend to have higher levels of profit efficiency than independent banks, and their cost efficiencies were significantly greater. Banks with high loan-to-asset ratios showed higher profit efficiency. Banks that used purchased funds had lower profit efficiencies than others. Banks with more variable returns showed lower profit efficiencies and lower cost efficiencies. The authors also found that market power was negatively related to cost efficiency but positively related to alternative profit efficiency.

Bikker (2001) measured the efficiency of banks within the European Banking Industry. He evaluated efficiencies of nine European countries: Belgium, France, Germany, Italy, the Netherlands, Spain, UK, Luxembourg, and Switzerland. His sample included all banking categories within the European banking industry (Commercial banks, Co-Operative banks, Savings banks, Investment banks, Medium and Long term credit institutions, Real Estate and Mortgage banks, and Specialized Government Credit banks) during the period 1989-1998. He sourced the data from IBCA-Fitch database. Where banks are part of bank holding company, he only included the consolidated figures. The sample consisted of 3085 banks and 14751 bank-year observations.

Bikker (2001) estimated the X-efficiency using the stochastic cost frontier approach based on a multiproduct translog cost function. A Europe wide estimate of average level of efficiency was about 0.381. He asserts that this low level of efficiency might indicate an underestimation of actual efficiency, which might be due to the exclusion of input prices in estimating the frontier. The author found that, on average, French, Italian, and Spanish banks were less efficient than banks in Germany, the Netherlands, and the UK. Banks in Luxembourg, Belgium, and Switzerland were the most efficient. There were also large differences in average X-efficiencies and cost-levels between countries. Spain was 40% above and Luxembourg was about 35% below the European average.

Bikker (2001) then extended the Europe-wide model by including dummy variables to account for country specific differences. This resulted in an increase on the average efficiency from 0.381 to 0.460. The classification of the efficiency of Belgium banks fell from high to low and the status of Italian banks rose from low to high. Bikker then estimated the Europe-wide efficiency for separate bank categories. The weighted average of these efficiency estimates for separate categories was 0.622, higher than when all categories were taken together. Banking industry in Spain, Italy and France tended to be less efficient when evaluating banks under separate categories. Germany, the Netherlands, and the UK took medium position while banks in Switzerland, Belgium and Luxembourg were the most efficient. Overall, Bikker found that the differences in average cost-levels and X-inefficiency between EU countries were huge. The average inefficiency was probably higher than the 20% found in the U.S. literature.

2.3.2: Mergers and Cost Efficiency

Berger and Humphrey (1992a) looked at the cost efficiency implications of megamergers in the U.S. banking industry. They analyzed M&As where both organizations exceeded $1 billion in assets at the time of the merger over the period 1981-1989. They identified 114 such mergers and banks within a holding company, and treated them as a single unit. They collected annual data on banks from the Call Report from 1980 to 1990. The authors further computed performance efficiency variables over three time interval- years after the merger for the consolidated bank, and years prior the merger for both the acquiring and the target bank. They excluded acquisitions that involved a foreign bank or thrift institution.

Berger and Humphrey (1992a) narrowed the 114 identified mergers to 57 cases. They specified a translog cost function with five outputs and four input prices for all U.S. banks with at least $1 billion in banking assets for at least one of the years from 1980 to 1990. They excluded data on the year of the merger. They then employed the distribution-free approach to measure the X-efficiency variables of these mergers. This technique measures X-efficiency of each bank over the given time period before or after a merger by comparing the bank's average cost function residual over time with the average residuals for each of the members of its peer group over the same time interval.

Berger and Humphrey (1992a) found that there was no significant cost efficiency benefits from the megamergers of the 1980s on average, even though there appeared to be some successful individual mergers as well as some failures. Neither of the key hypotheses about ex ante efficiency differences or market overlap explain the ex post performance from these megamergers. So, despite all the potential cost efficiency gains from megamergers (e.g. efficiency of the acquiring bank exceeded that of the target, within-market mergers or substantial deposit market overlap, etc.) megamergers were not very successful on average in improving cost efficiency. The average X-efficiency improvement was less than 5 percentage points and statistically insignificant. The authors also found that the scale diseconomies created by the megamergers were more than large enough to offset the slight X-efficiency gains, so that the total cost efficiency actually declined slightly.

Rhoades (1993b) took on Srinivasan and Wall (1992) findings of no efficiency gain associated with the degree of market overlap, and analyzed horizontal bank mergers (mergers in which there was some local market overlap prior to the merger) to determine if there was any efficiency improvement relative to other firms. This analysis was motivated by the notion that horizontal (within-market) bank mergers would permit the reduction of cost through consolidation of back offices and administrative functions, and the closing of overlapping offices. The sample consists of 898 banks involved in horizontal mergers and all other banks from each of the years 1981-1986. He used various expense-to-asset ratios to measure efficiency of merged firms relative to other firms between the three years before merger, and 4^{th} – 6^{th} year after the merger.

In his analysis, Rhoades (1993b) compared the change in the expense ratio of merging and nonmerging banks, and the likelihood of a change in efficiency quartiles by merged banks relative to others. He included three years of data before and after the merger to average out unusual expense items and appropriate long-term expense performance. The post-acquisition analysis also ensure that there was sufficient time for the merged firm to achieve reductions in operating costs and for the expense ratios (averaged over three years) to approximate the new long-term expense performance of the merged firm. He sourced the data from the FDIC, Federal Reserve Bulletin, the Federal Reserve's analysis of Competition, and various files on banking structure.

Nonmerging or other banks were included if they were not involved in an acquisition during the period under investigation, and were not subsidiaries of any bank holding company making acquisitions during the year. The average asset size of the acquiring bank was \$1.8 - \$2.5 billion, and the target was \$100 million - \$500 million. Rhoades' (1993b) data showed that the acquired firms consistently had higher noninterest expenses to assets than the acquiring firms. That is, on average, these mergers are generally regarded as ones most likely to yield efficiency gains (more efficient firm acquiring a less efficient firm). After the merger, the consolidated firms had higher noninterest expense ratio than the consolidated firms before the merger. Overall, Rhoades' (1993b) data showed that the acquiring firms had lower noninterest expenses to asset than the target firm, but the acquirers were not generally successful in reducing the expense ratios of the combined firms. He ran two sets of tests: first, the ordinary least squares (OLS) analysis of changes in expense ratios of merged and other banks before and after the merger. Secondly, the logit analysis of the likelihood of a change in efficiency quartiles from before to after merger by merged banks relative to others.

Rhoades (1993b) found no indication of efficiency gains from horizontal bank mergers based on total expense ratios. In addition, the horizontal mergers with a relatively large degree of overlap did not result in efficiency gains as measured by the ratio of total expenses to assets. He then used efficiency quartiles from the thick frontier approach to determine if efficiency gains were made. He found that merged firms were no more likely to experience an increase in efficiency than other firms were, and the degree of deposit overlap had no significant influence on the likelihood of gains (or losses) in efficiency.

Shaffer (1993) took a different approach. Instead of examining the efficiency effects of past mergers, like other studies before him, he examined the range of potential future efficiency changes and the sources of these changes. His analysis was advantageous in the sense that it offered a platform to address the shortfall of efficiency studies of past mergers. That is, one cannot rely on efficiency studies of past mergers if the motivation for the past and present mergers were not the same. Therefore, Shaffer simulated hypothetical mergers between pairs of large U.S. commercial banks with assets exceeding $1 billion. He attempted to distinguish among separable contributions of scale and product mix economies, branch closings and X-efficiency.

Shaffer's (1993) sample included banks that had a continuous charter throughout the sample period (1984-1989) and at least the five prior years. These banks accounted for 57% of all U.S. bank assets as of year-end 1988. He estimated a multiproduct cost function across the sample and then used the function to simulate mergers between pairs of banks by comparing the sum of predicted costs for a given pair of banks with the predicted cost of the merged entity whose balance sheet was the simple sum of the quantity-weighted input prices facing each bank. The model, would then predict a reduction in total costs relative to independent operation of the two banks if the predicted costs for a given pair of banks exceeds the predicted costs of the merged entity.

Shaffer (1993) used the thick frontier approach to simulate the potential effects of scale and product mix, branch closings, and X-efficiency changes on cost. The thick frontier technique involves partitioning the sample into quartiles by average cost with the low-cost quartile defining the frontier and the high-cost quartile defining the least efficient banks. Rather than the average cost, the author defined the quartiles in terms of the residual in the estimated cost function. This approach takes account of variation in input prices and X variables.

The simulation results showed that 57% of mergers would increase costs assuming no branches were closed. However, closing all branches of the target bank resulted in about 2-percentage point reduction in post-merger costs. Therefore, there was limited scope for reducing costs. On the other hand, if the efficient bank merges with the least efficient bank, costs would fall by 21% if the managerial efficiencies can be transferred to the acquired bank. Efficiency losses of as much as 21% would be realized if the least efficient bank imposes its X-efficiency on the combined organization. Shaffer (1993:432) stated that, "86% of all possible mergers were found to reduce cost by more than 10% if the low-cost bank's X-efficiency prevailed in each pair, while 96% would increase cost by more than 10% if the high cost bank's X-efficiency prevailed". Overall, X-efficiency dominates scale, product mix, and branching effects, and that cost reduction is possible if the low cost bank export its X-efficiency to the combined entity.

Fixler and Zieschang (1993) examined the change in relative productivity arising from bank mergers. They used a superlative index number approach to measure technical efficiency. After constructing the Tornqvist productivity indices for every bank in their sample and using these as measures of relative efficiency, they defined the measure of inefficiency as the distance between the frontier of an individual bank and the frontier of the reference bank. They then measured this distance as relative productivity (i.e. ratio of an index of relative output to an index of relative input). The relative output and input indexes are Tornqvist indexes between a particular bank and a reference bank (i.e. they are the weighted geometric means of the ratio of the relevant quantities for a bank and the reference bank, where the weights are average shares in revenue or cost). Their method is no different from assuming that each bank has its own technologies and then measuring the efficiency by comparing the quality of these technologies.

Fixler and Zieschang's (1993) sample contains 2000 banks in each year of the sample period, 1984-1988. They considered mergers approved by the office of the Comptroller of Currency (i.e. national banks), and occurring in 1986. The sampled banks had an average asset size of $780 million in 1984. They examined the productivity trend of the acquiring banks, relative to the industry, for two years prior to the merger and two years after. They collected the data from the Call Report.

Fixler and Zieschang (1993) found that acquiring banks were about 40-50% more efficient than other banks prior to merging, and that they maintained about the same advantage over other banks in the years after the merger (hence acquisition activity may boost the productivity and efficiency of the industry). They found that the median productivity of an acquiring bank in a merger exceeded the full sample median productivity in 1984, 1985, 1986, 1987, and 1988 by 40.3%, 53.4%, 63.8%, 41.9% and 55% respectively. Overall, the authors found that mergers in and of themselves do not generally increase the efficiency of a bank's operations. Of significance, Fixler and Zieschang's (1993) study included revenue effects through their output index as well as cost effects through their input effect (something missing in most of the previous cost studies of bank mergers).

Vander Vennet (1996) examined the performance effects of horizontal acquisitions and mergers between EC credit institutions (commercial banks) over the period 1988-1993. He analyzed the profitability and operational efficiency of merging credit institutions relative to their industry peers in the period surrounding the mergers and acquisitions. His analysis covered a period starting three years before and three years after the merger and acquisitions. Criterion for inclusion was that through the acquisitions, the acquiring bank exceeded the threshold of 50% of the voting stock of the target institution. He excluded takeovers involving a failed credit institution. The initial sample contained a total of 1164 transactions taken from various bank supervisors and private databases. Due to differences in accounting practices and publication customs within the EC, the author performed some functional regrouping and screening of published statements. This resulted in a sample reduction up to a total of 492 domestic mergers and 70 cross-border transactions where full financial information about all participating credit institutions was available.

The analysis consisted of a univariate comparison of pre- and post-merger performance of merging banks. The ROE and ROA measured profitability before and after the merger. Vander Vennet (1996) also calculated the overall level of operational efficiency as the deviation of a bank's total costs from a stochastic cost frontier based on a translog cost function. He found that from the majority of domestic acquisitions, target banks tend to exhibit inferior performance than the acquirer. However, the acquirers were unable to remedy the situation. Therefore, these types of takeover fail to purge the system from inefficient bank. Post merger efficiency showed a tendency to deteriorate. Therefore, efficiency theories could not explain domestic acquisitions. The author also found that domestic mergers of equal partners showed impressive recovery of operational efficiency, which leads to improved profit. In addition, labor costs and operational expense ratios dropped to a level below the industry averages. Frontier efficiency measure also showed a significant improvement. However, he could not translate these results into the cross-border acquisitions.

Peristiani (1997) analyzed the post-merger performance of acquiring banks that participated in a merger during the period 1980-1990. He considered all 4900 individual transactions that occurred during that period. He obtained a list of mergers from the Board of Governors of the Federal Reserve System and the rest of the data was from the quarterly Call Report, and the Annual Summary of Deposits. He classified mergers into three categories; intracompany consolidations, FDIC assisted mergers, and intrastate combinations of non-affiliated solvent banks. He measured the pre-merger and post-merger performance of banks by selecting the available financial history of the acquirer and target up to sixteen quarters before and after the merger. The final sample included 786 multiple mergers and 395 onetime mergers.

Peristiani (1997) used the distribution-free approach based on the translog flexible functional form to estimate the cost structure of banks and then derived the measures of efficiency. He found that banks that participated in a merger realized a small but significant decline in pro forma X-efficiency two to four years after the merger. Therefore, during the 1980s, mergers were not beneficial to banks in terms of X-efficiency. This decline in X-efficiency was significant for Bank Holding Companies consolidations. Peristiani also used regression analysis to analyze cross-sectional differences for several measures of performance. He used as the dependent variables the pro forma change in X-efficiency, and in scale efficiency. However, he found that no factor appear to contribute to scale- and X-efficiency of these firms. In addition, in-market mergers did not yield significant improvements in post-merger performance.

De Young (1997b) tested whether the pattern of pre-and post-merger X-efficiency observed in banks was consistent with an efficient market for corporate control in which well-managed banks acquire poorly managed banks and subsequently improve their performances. He also tried to identify the merger characteristics most conducive to converting pre-merger X-inefficiency into post-merger cost savings. His sample included 348 single target bank mergers approved by the office of the Comptroller of Currency (OCC) in 1987 and 1988. In 88% of the mergers, the acquiring bank was larger than its target. Only about 3% of the mergers were megamergers. The author sourced the data from year-end Call Report.

De Young (1997b) used the thick frontier methodology to estimate the pre- and post-merger X-efficiency for acquiring and target banks based on the translog cost function with associated factor share equations. He generated X-inefficiency for acquiring banks, target banks, and for combined post-merger banks. He estimated the pre-merger X-inefficiency for both acquiring banks and target banks in the year before the merger (1986 or 1987), and post-merger X-inefficiency for merged banks, three years after the merger (1990 or 1994). He also partitioned the bank population into ten equal-sized groups by asset size. The lowest quartile of banks in each asset decile or the 25 percent of the banks with the most negative residual values were assumed to be the most cost efficient or "best practices" banks.

DeYoung (1997b), like other previous studies, found that acquiring banks were more efficient than their targets, a result consistent with an efficient market for corporate control in which efficient firms acquire inefficient firms. He found that banks that made multiple acquisitions showed post-merger cost-efficiency improvement. However, cost efficiency improved in only about half of the mergers in which the acquiring bank made only a single acquisition during the sample period. The author asserted that experience effects could improve bank's chances to glean efficiencies from their targets. He also found that the post-merger gains in X-efficiency were not related to an efficiency advantages that acquirers held over their targets. Overall, the author did not find results that support the traditional market for corporate control in which well-managed firms acquire poorly managed firms and subsequently improve their performance.

Berger (1998) extended the past efficiency analysis of M&A by using data from 1990-1995 to analyze the effects of mergers on U.S. banks. He not only examined bank mergers of all size but, also mergers in which two or more bank charters were combined during a single year into a consolidated bank. Thus, he excluded acquisitions in which banks retain separate charters. He employed three separate X-efficiency concepts – cost efficiency, standard profit efficiency, and alternative profit efficiency. The author then measured whether mergers on average improved these three types of efficiency. He also determined if there were any reliable ex ante conditions that might help predict which types of M&As were most likely to be successful in improving efficiency.

The sample consisted of 639 mergers from 1991 through 1994. Cost and alternative profit efficiencies were calculated for all 639 mergers, but standard profit efficiency was measured for only 546 mergers because one or more of the output prices of the other 93 banks were judged to be insufficiently accurate. Berger (1998) found that there was no significant improvement on average in cost efficiency rank. There was, however, a significant improvement on average profit efficiency rank. He also found that ex ante efficiencies of participating banks played a major role. That is, when merging banks were below median efficiency, the average improvement in efficiency rank was between 5% and 21% and statistically significant for all types of efficiency applied in all merger categories. In addition, efficiency rank tends to decline significantly after mergers for banks that were relatively efficient ex ante.

Berger (1998) tested the diversification hypothesis, in which the improved diversification of risks associated with the merger from an improved mix of geographical areas, industries, loan types, or maturity structures might allow consolidated banks to shift their output mixes from securities towards both consumer and business loans, thereby raising expected revenue (i.e. improved profit efficiency may be manifested in a portfolio shift from securities into consumer loans and business loans). He found that the data was consistent with this hypothesis. The author also examined how well merger efficiency improvements can be predicted using information available before the merger. He found that the data was not generally consistent with the relative efficiency hypothesis under which it is important that the acquiring bank be more efficient than the acquired bank so that superior management abilities, policies, and procedures may be spread over more resources. He, however, found that the data was consistent with the low efficiency hypotheses, under which greater merger efficiency gains are predicted if either or both of the participating banks are less efficient than other banks before the merger.

Lang and Welzel (1999) analyzed the size and X-efficiency effects of mergers among the Bavarian cooperative banks. These banks were typically small but together made up about 70% of all German universal banks dominant in merger wave at that time. Their sample not only covered 283 mergers between 1989 and 1997, but they also looked at the full sample of 1220 Bavarian Cooperative banks, which covered about 20% of all German universal banks and more than a quarter of cooperative banking segment. The maximum number of banks participating in one single merger was four, and the merged unit was assumed to represent a new independent bank. The acquiring bank was also bigger than the acquired institution.

Lang and Welzel (1999) used a time-variable stochastic frontier analysis based on the multi-output translog cost function to estimate firms' X-inefficiency. They distinguished between pre-merger and post-merger considerations, and also considered both the case of no branch closures and the case of closure of all branches of the acquired bank. That is, they infer from Shaffer (1993) that the magnitude of size effects depends heavily on the extent to which branch offices are closed in the post-merger phase. They also compared size effects from observed mergers to those of all hypothetical mergers, and examined differences in the management quality between acquiring and acquired banks.

Lang and Welzel (1999) found that in 47% of the 283 mergers the acquired bank was more X-efficient than the acquiring bank. Efficiency differences were significantly positive in 70 of the 283 merger cases. They also found that only 38.9% of all merger cases provided cost savings due to a size effect, if there were no branch closures. However, there was no evidence of dramatic cost reductions due to the size effect of merging if all offices of the acquired banks were closed. The authors then extended their analysis by calculating the size effects for all hypothetical pair-wise mergers and compared the results to the actual observations. They found that the size effects of actual mergers were slightly better than the size effects of hypothetical mergers.

On the question of whether there existed a positive relationship between the difference in X-efficiency before the merger and the after-merger performance, Lang and Welzel (1999) found that the transfer of superior management quality to a badly managed part of newly merged bank does not work well or at least does not work quickly enough. Therefore, the pre-merger X-efficiency advantages of acquiring banks were not the main driving force behind the mergers observed. The authors also found that if the acquired bank was more X-efficient, outperforming efficiency growth was more frequent than underperformance or deterioration. That is, there was no evidence of ex ante X-efficiency advantages to transform into superior performance ex post.

The cost efficiency studies are limited to the fact that they all take output as given and ignore the revenue effects of mergers. That is, because total output of a consolidated firm changes after the merger, it is difficult to determine from the cost analysis alone whether the cost changes are greater or less than the revenue changes. Therefore, we cannot tell by looking at the state of the cost efficiency (improved or worsened) whether the firm has become more or less efficient or become more profitable.

Looking only at the cost side of the bank may create some bias against banks with high quality service as their competitive strategy. That is, these banks will show relatively high costs per unit of output because of high quality services, but it might very well be that because of these services, the revenue per unit of output is relatively high as well. Efficiency estimates based on cost function approach would also concentrate on the higher costs, resulting in a lower level of efficiency. Besides, banks are usually considered profit-maximizing entities, and, hence, an efficiency estimate based on profits will be more consistent with actual bank behavior. Therefore, profit efficiency might be better suited to analyze the efficiency benefits of bank mergers rather than cost efficiency.

2.3.3: Mergers and Profit Efficiency

Vander Vennet (1996), as referenced above, found that domestic mergers of equals had a positive impact on profitability, which was driven by improvements in operational efficiency. However, his analysis did not control for the correlation of many of the balance sheet variables that he examined. Berger (1998), also referenced above, calculated the standard and alternative profit efficiencies for 639 mergers, and found a statistical significant increase of a few percent in profit efficiency rank on average relative to other banks.

Akhavein et al. (1997) applied a frontier profit function to data on bank megamergers to examine the efficiency and price effects of mergers. They investigated both the profit efficiency and market power effects of mergers whilst trying to identify the ex ante conditions that predict when either is likely to be increased. They also compared the effects of efficiency and market power, and show the differences between cost and profit efficiency. The authors found an improvement in profit efficiency of large U.S. banks after M&As but little change in cost efficiency. Profit efficiency increased especially if both merger participants were relatively inefficient prior to the merger. They also found that after merging, banks tend to shift their portfolios from securities to loans. This lead to increased profit efficiency as loans creates more value than purchasing securities (consistent with the diversification hypothesis).

Akhavein et al. (1997) collected data on all mergers of U.S. banking organizations during 1981-1989 in which both partners had at least $1 billion in assets. They treated the high holding company (the holding company that is not owned by any other holding company) as the decision-making unit that tries to maximize profits. They, therefore, sum together all the commercial banks that are jointly owned through the holding company structure, and treat them as a single profit-maximizing unit. That is, the authors excluded M&As in which the banks involved remained separate corporate entities after the consolidation, but included mergers between banks that were subsidiaries of the same holding company (corporate reorganizations).

While evaluating and comparing all three potential sources of increased operating profits from mergers (cost efficiency, profit efficiency, and market power), Akhavein et al. (1997) specified a modified Fuss normalized quadratic variable profit function. However, their measure of profit efficiency did not account for changes in risk likely to result from a portfolio switch found. They just assumed that the equity market would recognize and account for any changes in the portfolio composition after the merger. Gains in profit efficiency were also attributed to benefits of risk diversification. That is, larger banks had more diversified loan portfolios and lower equity-asset ratios. This can be attributed to the sample period that coincided with the gradual lifting of the Riegle-Neal Interstate Banking and Branching Efficiency Act, which allowed banks from different states, each with geographically concentrated portfolios, to merge and thus diversify their holdings. The authors also found that changes in market power associated with megamergers were very small on average and not statistically significant. This was consistent with the hypothesis that antitrust policy had been successful in preventing mergers that would bring about large increases in market power.

Focarelli et al. (2002) analyzed the characteristics and performance of Italian banks involved in M&A during the period 1985-1996 using balance sheet data. They sourced the data from the Banking Supervision Register at the Bank of Italy, Central Credit Register, and the Census of Banks. They concentrated on retail commercial banks thereby excluding bank associations' clearinghouses, banks specializing in medium and long-term lending, branches of foreign banks, and mutual banks. The authors also separated transactions in terms of whether it was a merger (when a previously independent bank loses its charter and becomes part of an existing bank, with one headquarters and a unified branch network) or acquisition (when a bank purchases the majority of the voting rights of another bank without combining the assets of the two).

Focarelli et al. (2002) found that acquisitions resulted in long-run improvement in profitability for the acquired bank. The acquired bank was faced with the reassessment of the loan portfolio and more conservative lending strategy from the acquiring bank. As a result, there was a long-run decrease in the ratio of lending to total financial assets. They also found that the bad loan ratio increases by 8 percentiles in year t = 0. Subsequently, the quality of loan portfolio increases permanently because of the improvement in credit risk management brought by the acquiring bank. They found a long-run decrease in 11 percentiles of bad loan ratio. This was due to a decline in bad loans. Acquisitions did not lead to any permanent effect on the cost structure of the acquired banks.

2.4: Conclusion

This chapter put the study into context in terms of previous literature on bank efficiency. We grouped these literatures under two broad categories: financial ratio studies and the X-efficiency studies. Financial ratio studies only focus on past mergers and acquisitions and simply compare the pre- and post-merger financial ratios. While only comparing the financial cost ratios, these studies concluded that the potential for cost efficiency improvement from mergers was generally not realized. These studies did not find any evidence of a substantial change in cost performance associated with bank mergers. However, most of these studies only examined the cost or input side. The only exceptions were Cornett and Tehranian (1992) and Spindt and Tarhan (1992) who found most of the merger benefits on the revenue or output side.

Even after the introduction of X-efficiency concept in the 1960s (Leibenstein 1966), financial ratio studies did not attempt to measure the X-efficiencies of financial institutions. This was irrespective of the implications from the main findings of the previous financial studies efficiency that the X-inefficiency appeared to be the most important source of efficiency differences among financial institutions. Most financial efficiency studies were also limited to mergers and acquisitions during the 1980s. Berger and Humphrey (1992a) listed other methodological difficulties of the financial ratio studies.

The X-efficiency studies not only estimate the best practice frontier of banks, but also help solve the methodological limitations associated with the ratio analysis. Studies that analyzed the cost-efficiency implications of mergers were limited as they took output as given and ignored the revenue effects of mergers. These studies did not find any efficiency benefits on average following a bank merger. They also could not determine whether the cost changes were greater or less than the revenue changes by looking at the cost analysis alone. That is, the state of cost efficiency (improved or worsened) cannot tell if a firm has become more or less efficient or become more profitable after a merger. One of the implications of the cost efficiency analysis of bank mergers was that future research should apply the profit function in all merger efficiency analysis.

Studies that applied the profit efficiency concept in analyzing bank mergers found a statistically significant increase of a few percent in profit efficiency rank on average relative to other banks. Gains in profit efficiency were also attributed to benefits of risk diversification. Overall, Berger, Hunter, and Timme (1993) summarized the implications of the main findings of previous studies on bank efficiency. They indicated that studies should move away from the use of simple financial ratios. Also, it is important to determine factors that predict merger efficiency loss or gains, and the profit function (not the cost function) should be applied in all merger analysis. This study employs the stochastic frontier approach based on the flexible translog functional form in analyzing the effects of mergers on bank efficiency. In line with the implications of the main findings of previous studies, this study applies the profit function instead of the cost function.

CHAPTER 3: THEORETICAL FRAMEWORK

3.1: Introduction

This chapter presents the theoretical framework and analytical foundation of producer theory and efficiency measurement. The second section characterizes the production possibilities set and the production frontier. This section also provides definitions of the component of productive efficiency (technical and allocative efficiency). That is, we trace the microeconomic theory of efficiency measurement back to Farrell (1957) who defined a measure of firm efficiency within the context of technical, allocative, and productive efficiency.

Section 3 outlines the efficiency measurement techniques. This section also discusses the two primary methodologies developed to measure economic efficiency: mathematical programming approach and the econometric approach. Section 4 pays special attention to the theoretical foundation of the stochastic production frontier. This section also presents two distributional assumptions associated with the stochastic production frontiers: normal-half normal distribution and the normal-truncated normal distribution. Section 5 presents an analysis of the profit function and profit efficiency concepts, which are adopted in this study. Section 6 summarizes the chapter by outlining the developments in the efficiency measurement theory.

3.2: Definitions and Measures of Efficiency

3.2.1: Production Possibility Set

Mas-Colell et al. (1995) defined a production possibility set as a set of all production vectors that constitute feasible plans for the firm. It is a set that describes a pattern of inputs and outputs that are technologically feasible. We can model a production technology that transforms factors of production or vector of inputs $\tilde{x} = (x_1, \cdots, x_n) \in \mathbb{R}_+^n$ to produce a vector of net outputs $\tilde{y} = (y_1, \cdots, y_n) \in \mathbb{R}_+^m$ as

$$\mathrm{T} = \left\{ (\mathrm{x,y}) : x \in \mathbb{R}_+^n, y \in \mathbb{R}_+^m, \text{ x can produce y} \right\} \tag{3.1}$$

We can also model the production possibility set by the input requirement set $y \to V(y)$ or inversely by the output possibility set (output correspondence) $x \to P(x)$. The input requirement set, which is a special case of a restricted production possibility set, represents a combination of all input vectors $\tilde{x} = (x_1, \cdots, x_n) \in \mathbb{R}_+^n$ that yield at least the output vector. We can represent the input requirement set as

$$V(y) = \left\{ x \in \mathbb{R}_+^n : (y, -x) \in T \right\} \tag{3.2}$$

For a given T, the production function $f(x) : \mathbb{R}_+^n \to \mathbb{R}_+$ is defined as the maximum level of all outputs $\tilde{y} = (y_1, \cdots, y_n) \in \mathbb{R}_+^m$ produced from the given inputs $\tilde{x} = (x_1, \cdots, x_n) \in \mathbb{R}_+^n$. That is,

$$f(x) = \max_y \left\{ y : (x, y) \in T \right\} \tag{3.3}$$

Figure 3.1 below illustrates the production possibilities sets.

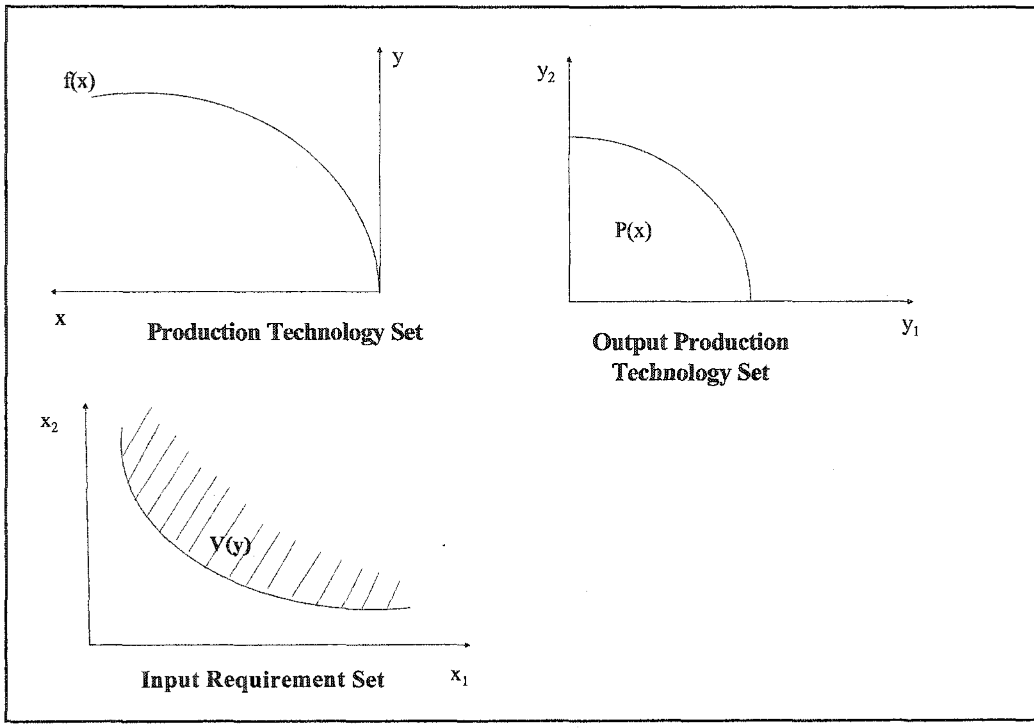


Figure 3.1: Production Possibilities Sets

3.2.2: Production Frontiers

We can define a production frontier in terms of the input sets V(x) and output sets P(x). The input and output correspondence can be obtained from one another by means of the relationships

$$\begin{aligned} V(y) &= \{x : y \in P(x)\} \\ P(x) &= \{y : x \in V(y)\} \end{aligned} \tag{3.4}$$

The graph of the production technology is the collection of all feasible input-output vectors:

$$\begin{aligned} PT &= \{(x,y) \in \mathbb{R}_+^{N+M} : y \in P(x), x \in \mathbb{R}_+^N\} \\ &= \{(\mathrm{x,y}) \in \mathbb{R}_+^{N+M} : x \in V(y), y \in \mathbb{R}_+^M\} \end{aligned} \tag{3.5}$$

Therefore, a production frontier is a function

$$\begin{aligned} f(x) &= \max \{y : y \in P(x)\} \\ &= \max \{y : x \in V(y)\} \end{aligned} \tag{3.6}$$

From figure 3.2 below, the production frontier f(x) describes the maximum output that can be produced with any given input vector. The Production frontier not only serves as a standard against which to measure technical efficiency but also provides the upper boundary of the production possibilities. The producer's input-output combination is located on or beneath the production frontier.

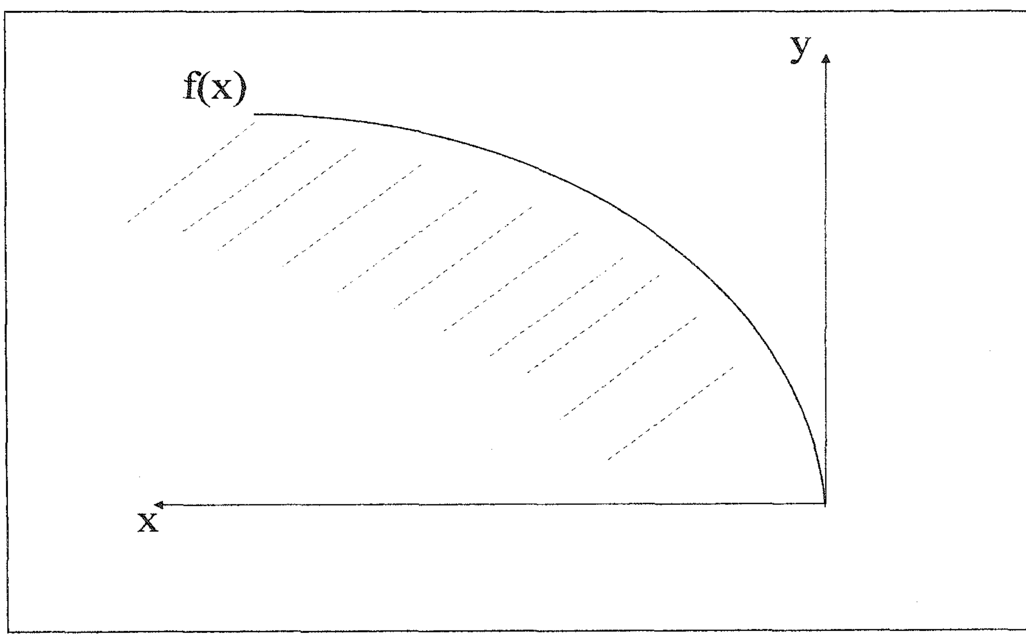


Figure 3.2: Production Frontier

3.2.3: Definitions of Efficiency

Productive efficiency consists of technological (or technical) and allocative efficiency. A production unit that employs its resources allocatively and technologically efficiently can be referred to as having achieved its economic efficiency. The technological or technical efficiency focuses on the level of inputs relative to output levels. To be technologically or technically efficient a firm must minimize its inputs given outputs or maximize outputs given inputs. Derived technical efficiency scores tend to be higher (as technical efficiency exclude allocative efficiency) compared to economic efficiency scores. Technical efficiency also requires only the input and output data.

On the other hand, economic efficiency is a broader concept than technical efficiency as it involves optimally choosing the levels of mixes of inputs and/or outputs based on reaction to market prices. Therefore, economic efficiency requires both technical and allocative efficiency, which mean that the optimal inputs and/or outputs are chosen, based on both the production technology and their relative prices in the market. Therefore, a firm can be technically efficient but economically inefficient or vice versa. Economic efficiency sets a higher standard that include allocative efficiency. Therefore, its scores might be lower than the technical efficiency scores.

We can trace the microeconomic theory of efficiency measurement back to Farrell (1957). He defined a measure of firm efficiency within the context of technical, allocative, and productive efficiency. According to Farrell, the efficiency of a firm consisted of the technical efficiency (which is the ability of a firm to maximize output from a given set of inputs) and allocative efficiency (which is the ability of a firm to use these inputs in optimal proportions, given respective prices). Therefore, productive efficiency equals a combination of both allocative and technical efficiency. Figure 3.3 below illustrate Farrell's argument.

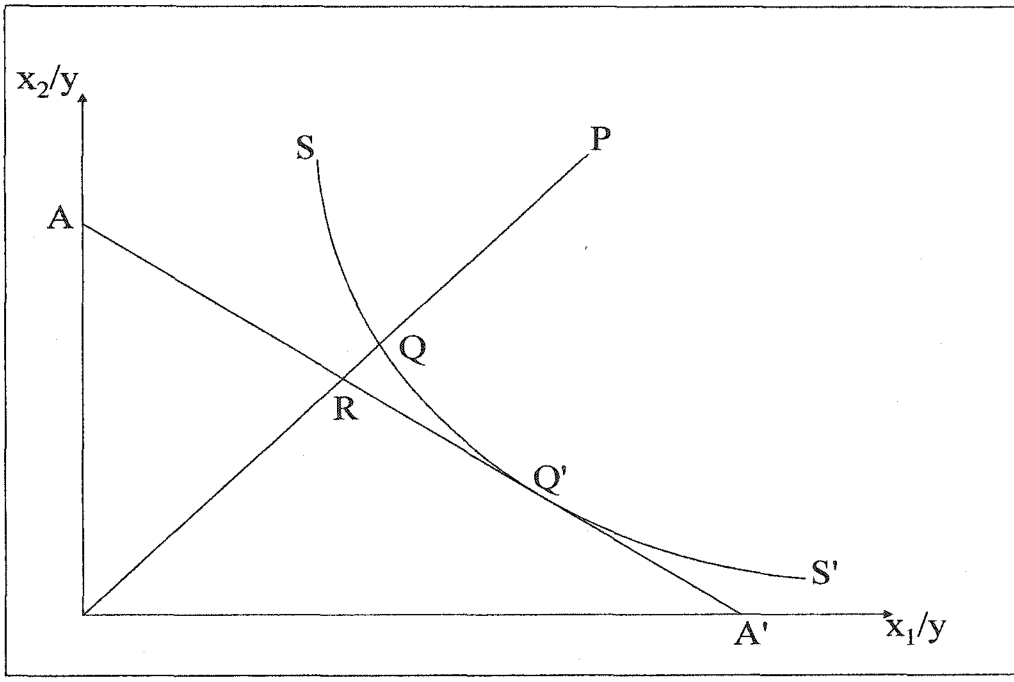


Figure 3.3: Technical, Allocative, and Productive Efficiency

We assume that there are two inputs, x_1 and x_2 used to produce a single output y. The production frontier is modeled as y = f (x_1, x_2). Under the assumption of constant return to scale, 1 = f (x_1/y, x_2/y). That is, as inputs increases, the relationship between the inputs (x_1 and x_2) and the output y does not change. The isoquant SS' shows various combinations of the two input factors that a firm employs to produce a unit of output. The ratio OQ/OP define the level of technical efficiency (or the ability to maximize output from a given set of inputs) for a firm using inputs (x_1^*, x_2^*) defined by point $P\left(\frac{x_1^*}{y}, \frac{x_2^*}{y}\right)$ to produce a unit of output, y*. That is, point Q represents an efficient firm using input factors (x_1, x_2) in the same ratio as point P. Therefore, the ratio OQ/OP measures the proportion of (x_1, x_2) necessary to produce y*. It follows that the ratio OQ/OP measures the technical efficiency of the production unit of a firm operating at P. Therefore, 1 – OQ/OP (which measures technical inefficiency of the production unit) measures the proportion by which $\left(x_1^*, x_2^*\right)$ could be reduced (holding x_1/x_2 constant) without reducing output. That is, it measures the possible reduction in the cost of producing y*.

Point Q lies on the efficient isoquant. If the input price ratio that is represented by the slope of the isocost line AA' is known, then we can calculate the allocative efficiency (Farrell referred to as the Price Efficiency). A ratio OR/OQ (which shows the level of allocative efficiency) indicates the production unit's ability to use inputs in optimal proportions, given the respective prices at point P. Therefore, 1 – OR/OQ is the allocative inefficient point. The distance RQ represents the reduction in production costs that would occur if production were to occur at the allocatively and technically efficient point Q' rather than Q (which is the technically efficient but allocatively inefficient point).

Therefore, if the production unit was perfectly efficient (i.e. both technically and allocatively), then the total economic or productive efficiency (which Farrell -1957- referred to as the Overall Efficiency) would be defined by the ratio OR/OP. The total inefficiency is therefore 1 – OR/OP. We can interpret the distance RP in terms of the cost reduction achieved by moving from the observed point P to the cost minimizing point Q'.

Farrell (1957) defined these efficiency measures based on the assumption that the efficient production function is known. That is, he assumed that there are methods of computing the observed performance of a production unit with some postulated standard. However, this is not usually the case. Therefore, efficient isoquant must be estimated using sample data. He, therefore, suggested the use of mathematical programming approach (or a nonparametric piecewise-linear convex isoquant) or a parametric or econometric approach. These approaches (further discussed in full below) use different techniques to envelope the observed data, and therefore make different accommodations for random noise and for flexibility in the structure of the production technology.

3.3: Efficiency Measurement Techniques

Recent academic research on performance of institutions in various industries has shifted away from standard performance measures (nonfrontier) to the adoption of frontier efficiency or X-efficiency. Frontier efficiency measures deviations in performance from that of best-practice firms on the efficient frontier, ceteris paribus. There are two primary methodologies developed to measure economic efficiency: the mathematical programming approach and the econometric approach. These methodologies estimate a "best practice" frontier with the efficiency of a specific decision making unit measured relative to the frontier. The frontier efficiency of an institution measures how well that institution performs relative to the predicted performance of the best firms in the industry market conditions.

As Bauer et al. (1998:86) indicated, these frontier efficiency measures "use programming or statistical techniques to try to remove the effects of differences in prices and other exogenous market factors affecting the standard performance ratios in order to obtain better estimates of the underlying performance of the managers". The two approaches differ in the assumptions made regarding the shape of the efficient frontier, the existence of random error, and (if random error is allowed) the distributional assumptions imposed on the inefficiencies and random error in order to disentangle one from the other. These approaches also differ in whether the underlying concept to be analyzed is technological efficiency (that focuses on the level of inputs relative to outputs level) or economic efficiency (a broader concept that also involves optimally choosing the levels of mixes of inputs and/or outputs on reactions to market prices).

3.3.1: Mathematical Programming Approach

Also referred to as the data envelopment analysis (DEA), this approach usually measures technological or economic efficiency. The DEA is a method that calculates technical inefficiencies for decision-making units (DMUs) as the ratio of the sum of virtual outputs to the sum of virtual inputs or vise versa. The DEA (first formulated by Charnes, Cooper, and Rhodes – 1978, 1989) focuses primarily on the technological aspects of production correspondences. It is not sensitive to input or output measurement and does not place restrictions on technology. It also does not assume maximizing behavior of firms. Therefore, it can be used to estimate technical and scale efficiency without requiring estimates of input and output prices.

This nonparametric approach (which also includes the Free Disposal Hull - FDH) uses linear programming techniques to measure efficiency. In this framework, an efficient firm is one in which there is no other firm or linear combination of firms that produces as much or more of every output for given inputs or uses as little or less of every inputs for given outputs. This undominated firm is considered efficient if it minimizes the cost of producing its observed outputs given the best practice technology and input prices.

The main advantage of the DEA is that it places less structure on the frontier and it is nonstochastic. That is, it does not require an explicit specification of a functional form such that it imposes very little structure on the shape of the efficient frontier. However, the DEA suffers from major shortcomings. Because the DEA does not require any restrictive assumptions about technology (except about convexity) and does not require any distributional assumption about the efficiency, nor is there any stochastic specification imposed, all variation between production units or any departure from the frontier is interpreted as inefficiency. It is also not possible to test the statistical significance of the parameters associated with the inputs. Therefore, the DEA is susceptible to data measurement errors, and it is sensitive to the number of firms and the number of inputs and outputs employed.

The ability to distinguish efficient DMUs diminishes as the number of inputs and outputs increases. Although the DEA allow us to assume returns to scale, this affects the ability to discriminate between efficient and inefficient DMUs and to rank the inefficient ones. The DEA cannot distinguish between technical inefficiency and the effect of external factors outside the control of management (e.g. weather, luck, strikes, etc.). The DEA scores may also be affected by data transformation. For example, the logarithmic transformation of the DEA data makes the scores closer to one thereby giving an impression of very little technical inefficiencies in firms. Because the DEA method is designed to measure technical inefficiency of each DMU, additional estimation is required to evaluate and explain allocative inefficiency of each DMU. That is, the DEA approach provides a partial and incomplete analysis of inefficiency.

3.3.2: Econometric Approach

The econometric approach consists of the stochastic frontier analysis (SFA), the thick frontier analysis (TFA), and the distribution-free analysis (DFA). All these three analysis usually measure economic efficiency (which include both the technological and allocative efficiency). The parametric approach specifies a functional form for the cost, profit or production frontier. It is also stochastic such that firms can be off the frontier either because they are inefficient or because of random shocks or measurement errors. Therefore, efficiency is measured by separating the efficiency component from the overall error term. Therefore, the primary advantage of the econometric approach is its ability to accommodate and separate random noise in efficiency estimation. The main shortfall of the parametric approach is that it imposes more structure on the shape of the frontier by specifying a functional form for the frontier. The challenge is in determining how best to separate random error from inefficiency, as neither of them is observed.

To illustrate the econometric approach to measuring efficiency, we look at a single equation cross-sectional model. We also limit the discussion to the estimation of technical efficiency under the assumption that a producer uses inputs $x \in \mathbb{R}_+^n$ to produce a scalar output $y \in \mathbb{R}_+^m$ for each of the N producers. We present a production frontier model as

$$y_i = f(x_i;\beta)\cdot TE_i \tag{3.7}$$

where y_i = scalar output of producer i, i=1,...,N; x_i = vector of K inputs used by producer i; $f(x_i;\beta)$= the production frontier; and β = vector of technology parameters to be estimated. TE_i is the Debreu-Farrell output-oriented technical efficiency of producer i.

Following Kumbhakar and Lovell (2000), we can represent the TE as a function

$$TE(x,y) = \left[\max\left\{\phi : \phi y \le f(x)\right\}\right]^{-1} \tag{3.8}$$

Thus,

$$TE_i = \frac{y_i}{f(x_i;\beta)} \tag{3.9}$$

From equation (3.9), technical efficiency is the ratio of observed output to maximum feasible output where maximum feasible output is given by the production frontier $f(x_i;\beta)$. y_i achieves its maximum feasible value of $f(x_i;\beta)$ if and only if $TE_i = 1$. However, $TE_i < 1$ represent the measure of shortfall of observed output from maximum feasible output (inefficiency).

The production frontier $f(x_i;\beta)$, as specified in equation (3.7) is deterministic as it uses a purely one-sided error. Consequently, in equation (3.9), any production below the frontier $f(x_i;\beta)$ is called technical inefficiency. This specification assumes that all conceivable inputs or external events have a maximal possible or bounded effect. The problem is that the deterministic frontier ignores the fact that output can be affected by random shocks outside of the control of a producer. All the factors outside the control of the producer are taken as inefficiency. In addition, any error or imperfection in the specification of the model could likewise translate into increased inefficiency measures.

However, a more realistic formulation is one that recognizes that any particular producer faces their own production frontier and that, given a set of inputs, there is a maximal output level that is possible, but this maximal level is random rather than exact. This assumes that some other inputs or external effects have maximal possible effects, but others have potentially unbounded effects. For example, the effects of weather and other external events might be regarded as normally distributed (and thus unbounded). Therefore, we can rewrite equation (3.7) as:

$$y_i = f(x_i;\beta)\cdot e^{v_i}\cdot TE_i \tag{3.10}$$

where $f(x_i;\beta)\cdot e^{v_i}$ is the stochastic production frontier. This stochastic production frontier is determined by the structure of production technology (the deterministic production frontier) and the external events, positive and negative, beyond the control of the producer. That is, $\exp\{v_i\}$ captures the effect of random shocks on each producer. The stochastic frontier expresses maximal output, given a set of inputs, as a (normal) distribution rather than a point.

Therefore, the econometric version of Debreu-Farrell output-oriented technical efficiency is

$$TE_i = \frac{y_i}{f(x_i;\beta)\cdot e^{v_i}} = e^{u_i} \tag{3.11}$$

which, defines technical efficiency as the ratio of observed output to maximum feasible output in an environment characterized by $\exp\{v_i\}$. Therefore, y_i will achieve its maximum value of $f(x_i;\beta)\cdot\exp\{v_i\}$ if and only if $TE_i = 1$. However, $TE_i < 1$

provides a measure of the shortfall of observed output from maximum feasible output in an environment characterized by stochastic elements that is allowed to vary across producers. We can estimate technical efficiency using either the deterministic production frontier or the stochastic production frontier. Table 3.1 summarizes the three econometric approaches to efficiency estimation.

Table 3.1
Summary of the Parametric /Econometric Approaches

Stochastic Frontier Approach (SFA)
❖ Employs a composite error, $\varepsilon = u + v$. ✓ $u \geq 0$ represents inefficiency and follows a half-normal asymmetric distribution, while v represents random error and behaves according to a symmetric normal distribution. ✓ u and v are assumed to be orthogonal to the input prices, output quantities, and any other cost function regressors specified. ❖ Provides point estimates of efficiency for all individual firms. ❖ It rank firm inefficiencies in the same order as their cost function residual, no matter which specific distributional assumptions are imposed.
The Thick Frontier Approach (TFA)
❖ Take the same functional form as the SFA. ❖ Regression is based only on the best performers in the data set (those in the lowest average cost quartile for their size class). ❖ Does not provide point estimates of efficiency for all individual firms. ❖ The levels of efficiency generated are based on an arbitrary assumption – that the lowest average cost quartile within each size class is an adequate thick frontier of efficient firms. ❖ Efficiency orderings are consistent with the SFA and other measures of performances.
The Distribution Free Approach (DFA)
❖ Requires a panel data set, and specifies a functional form like the SFA and TFA. ❖ Does not impose a specific shape on the distribution of efficiency. ❖ A cost or profit function is estimated for each period of a panel data set. ❖ Residual in each separate regression is composed of both inefficiency, lnu, and random error, lnv. ❖ Random component, lnv, is assumed to average out over time, so that the average of bank's residuals from all of the regression, $\ln\hat{u}$, will be an estimate of the inefficiency term, lnu. ❖ For banks with very low or very high $\ln\hat{u}$, an adjustment (truncation) is made to assign less extreme values of $\ln\hat{u}$, since extreme values may indicate that random error, lnv, has not been completely purged by averaging. ❖ The resulting $\ln\hat{u}$ for each bank is used to compute its core efficiency. ❖ Cannot capture X-efficiency changes within a firm over time.

3.4: The Stochastic Production Frontiers

The stochastic production frontier model was simultaneously introduced by Aigner, Lovell, and Schmidt (1977), and Meeusen and van den Broeck (1977). The model is motivated by the idea that deviations from the production frontier may not be entirely under the control of the production unit under scrutiny. As Kumbhakar and Lovell (2000:72) indicated, "the great virtue of stochastic production frontier models is that the impact on output of shocks due to variation in labor and machinery performance, vagaries of the weather, and just plain luck can at least in principle be separated from the contribution of variation in technical efficiency".

While allowing for technical inefficiency, these models also acknowledge the fact that random shock outside the control of producers can affect output. They account for measurement error and other factors (e.g. weather, luck, etc.) on the value of the output of unspecified input variables in the production process. We can formulate a stochastic frontier model in terms of a general production function for the i^{th} production unit as,

$$\begin{aligned} y_i &= f(x_i, \beta) + v_i - u_i \\ &= f(x_i, \beta) + \varepsilon_i \quad i=1,\ldots,N \end{aligned} \tag{3.12}$$

where v_i = two-sided noise component, and $u_i \geq 0$ is the nonnegative technical inefficiency components of the error term. The model is such that the possible production, Y_i, is bounded above by the stochastic quantity, $f(x_i;\beta)\cdot\exp(v_i)$, (hence, the term stochastic frontier). The noise component v_i is assumed to be independently and identically distributed (iid) and symmetric, distributed independently of u_i. The error term $\varepsilon_i = v_i - u_i$ is, therefore, asymmetric since $u_i \geq 0$. Following Battese (1992), figure 3.4 below illustrates the basic structure of the stochastic frontier model.

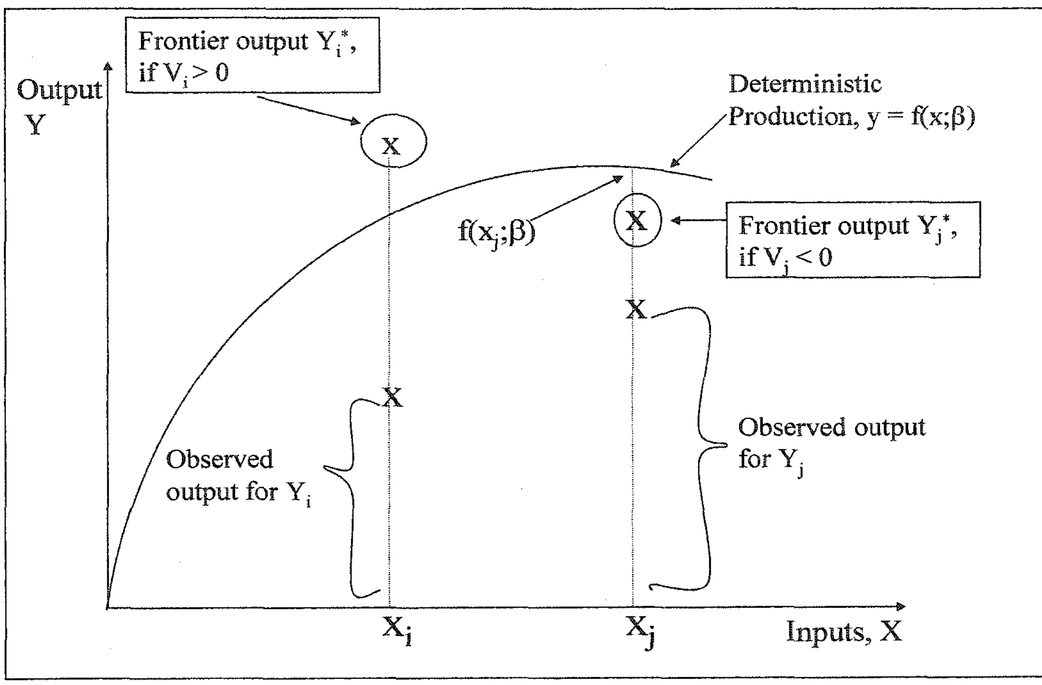


Figure 3.4 Stochastic Frontier Production Function

Figure 3.4 illustrates the productive activities of two firms, i and j. Firm i use inputs $\underset{\sim}{x}_i$ to produce output Y_i. However, firm i's productive activity is associated with favorable conditions for which the random error V_i is positive. Therefore, firm i's frontier output Y_i^* exceeds the value on the deterministic production function, $f(x_i;\beta)$. However, firm j uses inputs $\underset{\sim}{x}_j$ to produce output, Y_j. The corresponding frontier output, Y_j^*, is less than the value on the deterministic production function,

$f(x_i;\beta)$, because firm j's productive activity is associated with unfavorable conditions for which the random error, V_j, is negative. In both firms, the observed production values are less than the corresponding frontier values:

$$\begin{aligned} Y_i &< Y_i^* \\ Y_j &< Y_j^* \end{aligned} \tag{3.13}$$

However, the unobservable frontier production values would lie around the deterministic production function associated with the firms involved. Also, if $V_i > U_i$, both the observed and frontier production values, Y_i and $Y_i^* \equiv f(x_i;\beta)\cdot\exp(v_i)$, would lie above the corresponding value of the deterministic production function, $f(x_i;\beta)$.

If we assume that v_i and u_i are distributed independently of x_i, then the estimation of (3.12) by OLS provides consistent estimates of the βs except β_0. That is, $\mathrm{E}(\varepsilon_i) = -\mathrm{E}(u_i) \le 0$. The OLS estimation also provides a simple test for the presence of technical efficiency in the data. That is, if $u_i > 0$, then $\varepsilon_i = v_i - u_i$ is negatively skewed and there is evidence of technical inefficiency in the data.

However, the OLS does not provide estimates of producer-specific technical efficiency, which is the ultimate objective of the estimation process in addition to obtaining estimates of the production technology parameters β in $f(x_i;\beta)$. This requires an extraction of separate estimates of statistical noise v_i and technical inefficiency u_i from the estimates of ε_i for each producer. Therefore, to estimate the technical efficiency of each producer, we require the distributional assumptions. The next sections illustrate only two distributional assumptions: the normal-half normal, and the normal-truncated normal distributions.[2]

3.4.1: Normal-Half Normal Stochastic Frontier Model

Considering the stochastic production frontier model given in equation (3.12), we can make the following distributional assumptions,

i. $v_i \sim$ iid $N(0,\sigma_v^2)$

ii. $u_i \sim$ iid $N^+(0,\sigma_u^2)$ i.e. nonnegative half-normal

iii. v_i and u_i are distributed independently of each other, and of the regressors.

The density functions for $u \geq 0$ and for v are given by

$$f(u) = \frac{2}{\sqrt{2\pi}\sigma_u} \cdot e^{\left\{-\frac{1}{2}\left(\frac{u}{\sigma_u}\right)^2\right\}}$$

$$\text{and} \tag{3.14}$$

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot e^{\left\{-\frac{1}{2}\left(\frac{v}{\sigma_v}\right)^2\right\}}$$

Given that u and v are distributed independently, the joint density function of u and v is the product of their individual density functions, and is given as

$$f(u,v) = \frac{2}{2\pi\sigma_u\sigma_v} \cdot e^{\left\{-\frac{1}{2}\left[\left(\frac{u}{\sigma_u}\right)^2 + \left(\frac{v}{\sigma_v}\right)^2\right]\right\}} \tag{3.15}$$

Since $\varepsilon = v - u$, the joint density function of u and ε is

$$f(u,\varepsilon) = \frac{2}{2\pi\sigma_u\sigma_v} \cdot e^{\left\{-\frac{1}{2}\left[\left(\frac{u}{\sigma_u}\right)^2 + \left(\frac{(\varepsilon+u)}{\sigma_v}\right)^2\right]\right\}} \tag{3.16}$$

[2] Green ($1980_{a,b}$), Stevenson (1980), and Green (1990) provide extended analysis on the exponential and the gamma distributions.

Integrating u out of f(u, ε), we obtain the marginal density function of ε,

$$
\begin{aligned}
f(\varepsilon) &= \int_0^{\infty} f(u,\varepsilon)du \\
&= \frac{2}{\sqrt{2\pi}\sigma}\cdot\left[1-\Phi\left(\frac{\varepsilon\lambda}{\sigma}\right)\right]\cdot e^{\left\{-\frac{1}{2}\left(\frac{\varepsilon}{\sigma}\right)^2\right\}} \qquad (3.17) \\
&= \frac{2}{\sigma}\cdot\phi\left(\frac{\varepsilon}{\sigma}\right)\cdot\Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right)
\end{aligned}
$$

where $\sigma = \left[\sigma_u^2 + \sigma_v^2\right]^{\frac{1}{2}}$, $\lambda = \frac{\sigma_u}{\sigma_v}$, $\Phi(.)$, and $\phi(.)$ are the standard normal cumulative distribution and density functions, respectively. Aigner, Lovell, and Schmidt (1977) provide more details of the derivatives. The parameter λ embodies the model of inefficiency. The normal-half normal model is parameterized in terms of λ and σ^2. This provides some convenience as we can interpret λ as an indicator of the relative variability of the two sources of random error that distinguish firms from one another.

As $\lambda \to 0$ either $\sigma_v^2 \to +\infty$ and /or $\sigma_u^2 \to 0$. That is, the symmetric error component dominates the one-sided error component in determining ε. This yields an OLS production function model with no technical inefficiency. That is, every firm operates on its frontier. As such, it is difficult to test for inefficiency through the usual means (e.g. Lagrange Multiplier). That is, because the polar value λ = 0 is on the boundary of the parameter space and not in the frontier. As $\lambda \to +\infty$ either $\sigma_u^2 \to +\infty$ or $\sigma_v^2 \to 0$. That is, the one-sided error component becomes the dominant source of random variation in the model. Thus, reverting to a deterministic production frontier model with no noise. The marginal density function f(ε) is asymmetrically distributed around zero, with mean and variance

$$E(\varepsilon) = -E(u) = -\sigma_u \sqrt{\frac{2}{\pi}}$$

$$V(\varepsilon) = \frac{\pi - 2}{\pi} \sigma_u^2 + \sigma_v^2 \tag{3.18}$$

Aigner, Lovell, and Schmidt - ALS (1977) suggested [1 – E(u)] as an estimator of the mean technical efficiency of all producers. However, Lee and Tyler (1978) proposed

$$E\left(e^{-u}\right) = 2\left[1 - \Phi(\sigma_u)\right] \cdot e^{\left\{\frac{1}{2}\sigma_u^2\right\}} \tag{3.19}$$

which is preferred to [1 – E(u)] since [1-u] include only the first term in the power series expansion of exp{-u}. E(exp{-u}), unlike [1 – E(u)], is consistent with the definition of technical efficiency given in equation (3.11). The log likelihood function for a sample of N producers is

$$\ln L\left(y \mid \beta, \lambda, \sigma^2\right) = -N \ln \sigma - \text{constant} + \sum_i \ln \Phi\left(-\frac{\varepsilon_i \lambda}{\sigma}\right) - \frac{1}{2\sigma^2} \sum_i \varepsilon_i^2 \tag{3.20}$$

where $\varepsilon_i = y_i - \beta' x_i$, $\lambda = \sigma_u / \sigma_v$, $\sigma^2 = \sigma_v^2 + \sigma_u^2$, and $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.[3]

By maximizing the log likelihood function with respect to the parameters (σ^2, λ, β), we obtain the maximum likelihood estimates of all parameters. These estimates are consistent as $N \to +\infty$. After estimating these parameters, the next step will be to estimate the technical inefficiency of each producer. The estimates of $\varepsilon_i = v_i - u_i$ contain information on u_i. However, the problem is extracting information that ε_i

[3] Aigner, Lovell, and Schmidt (1977) provide the first and second-order derivations of lnL with respect to $\beta, \lambda,$ and σ^2 for use in computational algorithms and as a basis for estimating asymptotic standard errors.

contain on u_i. That is, the inefficiency component u_i must be observed indirectly. A solution can be found by looking at the conditional distribution of $(u_i \mid \varepsilon_i)$ which contains whatever information ε_i contain concerning u_i. Jondrow, Lovell, Materov, and Schmidt (JLMS) (1982) present an explicit form for the normal-half normal model. They showed that if $u_i \sim N^+\left(0, \sigma_u^2\right)$, the conditional distribution of u given ε is

$$\begin{aligned} f(u|\varepsilon) &= \frac{f(u,\varepsilon)}{f(\varepsilon)} \\ &= \frac{1}{\sqrt{2\pi}\sigma_*} \cdot \exp\left\{ -\frac{1}{2}\left(\frac{u-\mu_*}{\sigma_*}\right)^2 \right\} \Big/ \left[1-\Phi\left(-\frac{\mu_*}{\sigma_*}\right) \right] \end{aligned} \tag{3.21}$$

where $\mu_* = -\dfrac{\varepsilon\sigma_u^2}{\sigma^2}$ and $\sigma_*^2 = \dfrac{\sigma_u^2\sigma_v^2}{\sigma^2}$. Since $f(u \mid \varepsilon) \sim N^+\left(\mu_*, \sigma_*^2\right)$, the mean or the mode of this distribution can serve as a point estimator of u_i. The mean is given by

$$\begin{aligned} E(u_i|\varepsilon_i) &= \mu_{*_i} + \sigma_* \left[\frac{\phi\left(\dfrac{-\mu_{*_i}}{\sigma_*}\right)}{1-\Phi\left(\dfrac{-\mu_{*_i}}{\sigma_*}\right)} \right] \\ &= \sigma_* \left[\frac{\phi\left(\dfrac{\varepsilon_i\lambda}{\sigma}\right)}{1-\Phi\left(\dfrac{\varepsilon_i\lambda}{\sigma}\right)} - \left(\frac{\varepsilon_i\lambda}{\sigma}\right) \right] \end{aligned} \tag{3.22}$$

After deriving the point estimates of u_i, we can obtain the estimates of the technical efficiency for each producer. Thus,

$$TE_i = e^{-\hat{u}_i} = e^{E(u_i|\varepsilon_i)} \tag{3.23}$$

However, the estimates of technical efficiency are inconsistent because the variation associated with the distribution of $(u_i \mid \varepsilon_i)$ is independent of i.

Thus far the discussion focused on analysis of stochastic frontiers on the assumption that $u \sim N^{+}\left(0, \sigma_u^2\right)$. The following section looks at the normal-truncated normal model. Due to lack of a priori justification for selecting a particular distributional form for the technical inefficiency effects, the normal-half normal (and the exponential) distributions are the arbitrary selection. These distributions have a mode at zero, which implies that there is the highest probability that the inefficiency effects are in the neighborhood of zero. This also implies that technical efficiency is high, which might not be evident in practice. Attempts to address these shortfalls lead to the development of the normal-truncated normal and the gamma distributions. The normal-truncated normal and the gamma distributions allow for a wider range of distributional shapes, which include ones with non-zero modes.

3.4.2: Normal-Truncated Normal Stochastic Frontier Model

The normal-half normal model assume a distribution of u which has a mode at u = 0 (u represents the level of inefficiency). This specification by Aigner, Lovell, and Schmidt (1977) is based on the implicit assumption that the likelihood of inefficient behavior monotonically decreases for increasing levels of inefficiency. However, Stevenson (1980) argued that economic agents are humans or human institutions such that the possibility of a nonzero mode for the density function of u should seem a more tenable presumption. Stevenson (1980) generalized Aigner, Lovell, and Schmidt (1977) specification so as to permit a non-zero mode for density function of u, and to enable the testing of the special case of a zero mode. His model is reasonably general without regard to the pattern of efficiency distribution throughout the sample (Hence the introduction of the normal-truncated normal formulation). We can make the following distributional assumptions:

i. $v_i \sim$ iid $N(0,\sigma_v^2)$

ii. $u_i \sim$ iid $N^+(\mu,\sigma_u^2)$

iii. v_i and u_i are distributed independently of each other, and of the regressors.

This normal-truncated normal distribution differs from the one-parameter half-normal distribution in that it allows the latter, to have a non-zero mode. Kumbhakar and Lovell (2000) pointed out that the truncated normal distribution contains an additional parameter μ to be estimated. Therefore, the truncated-normal model is a more flexible representation of the pattern of efficiency in the data.

The density function for u and v are:

$$f(u) = \frac{1}{\sqrt{2\pi}\sigma_u \Phi\left(\frac{\mu}{\sigma_u}\right)} \cdot e^{\left\{-\frac{1}{2}\left(\frac{u-\mu}{\sigma_u}\right)^2\right\}} \quad \text{for } u \geq 0 \tag{3.24}$$

$$= 0 \quad \text{otherwise}$$

$$f(v) = \frac{1}{\sqrt{2\pi}\sigma_v} \cdot e^{\left\{-\frac{1}{2}\left(\frac{v}{\sigma_v}\right)^2\right\}} \quad \text{for all } v \tag{3.25}$$

where v is assumed to be normally distributed with zero mean and variance σ_v^2, u is assumed to be distributed as a truncated normal with mode μ, and $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable. f(u) is the density of a normally distributed variable with possibly nonzero mean μ, truncated below at zero. If $\mu = 0$, the density function in equation (3.24) collapses to the half-normal density function. The truncated normal distribution, as specified by Stevenson

(1980) is a two-parameter distribution, which depends on the placement and spread parameters μ and σ_u. The joint density function of $\varepsilon = u - v$ is given as:

$$f(u,v) = \frac{1}{2\pi\sigma_u\sigma_v\Phi\left(\frac{\mu}{\sigma_u}\right)} \cdot e^{\left\{-\frac{1}{2}\left[\left(\frac{u-\mu}{\sigma_u}\right)^2 + \left(\frac{v}{\sigma_v}\right)^2\right]\right\}} \tag{3.26}$$

The joint density of u and ε is

$$f(u,\varepsilon) = \frac{1}{2\pi\sigma_u\sigma_v\Phi\left(\frac{\mu}{\sigma_u}\right)} \cdot e^{\left\{-\frac{1}{2}\left[\left(\frac{u-\mu}{\sigma_u}\right)^2 + \left(\frac{\varepsilon+u}{\sigma_v}\right)^2\right]\right\}} \tag{3.27}$$

which can be integrated to get the marginal density of ε as

$$\begin{aligned} f(\varepsilon) &= \int_0^\infty f(u,\varepsilon)\,du \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v\Phi\left(\frac{\mu}{\sigma_u}\right)} \cdot e^{\left\{-\frac{1}{2}\left[\left(\frac{u-\mu}{\sigma_u}\right)^2 + \left(\frac{\varepsilon+u}{\sigma_v}\right)^2\right]\right\}} du \\ &= \frac{1}{\sqrt{2\pi}\sigma\Phi\left(\frac{\mu}{\sigma_u}\right)} \cdot \Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right) \cdot e^{\left\{-\frac{1}{2}\left(\frac{\varepsilon+\mu}{\sigma}\right)^2\right\}} \\ &= \frac{1}{\sigma}\phi\left(\frac{\varepsilon+\mu}{\sigma}\right) \cdot \Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right) \cdot \left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1} \end{aligned} \tag{3.28}$$

where $\sigma = \left(\sigma_u^2 + \sigma_v^2\right)^{\frac{1}{2}}$ and $\lambda = \frac{\sigma_u}{\sigma_v}$, ϕ is the standard normal density function evaluated at $\left(\frac{\varepsilon+\mu}{\sigma}\right)$. Note that $f(\varepsilon)$ collapses to the half-normal marginal density function if $\mu = 0$ (similar to Aigner, Lovell, and Schmidt, 1977). $f(\varepsilon)$ is asymmetrically distributed, and its mean and variance are respectively given as:

$$
\begin{aligned}
E(\varepsilon) &= -E(u) \\
&= -\frac{\mu a}{2} - \frac{\sigma_u a}{\sqrt{2\pi}} \cdot e^{\left\{-\frac{1}{2}\left(\frac{\mu}{\sigma_u}\right)^2\right\}} \\
V(\varepsilon) &= V(u) + V(v) \\
&= \mu^2 \frac{a}{2}\left(1 - \frac{a}{2}\right) + \sigma_u^2 \frac{a}{2}\left(\frac{\pi - a}{\pi}\right) + \sigma_v^2
\end{aligned}
\tag{3.29}
$$

where $a = \left[\Phi\left(\frac{\mu}{\sigma_u}\right)\right]^{-1}$.

This normal-truncated normal distribution has three parameters: a placement parameter μ and two-spread parameters, σ_u and σ_v. The log likelihood function for a sample of N producers can be represented as:

$$
\ln L = \text{constant} - N \ln\sigma - N \ln \Phi\left(\frac{\mu}{\sigma_u}\right) + \sum_i \ln\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon_i \lambda}{\sigma}\right) - \frac{1}{2}\left(\frac{\varepsilon_i + \mu}{\sigma}\right)^2 \tag{3.30}
$$

where we write σ_u as $\sigma\left(\lambda^{-2}+1\right)^{-\frac{1}{2}}$. Maximizing the log likelihood function with respect to the parameters β, λ, σ^2, and u will yield the maximum likelihood estimates.

Due to limitations of the normal-half normal model stated above, this study adopts the normal-truncated normal model because of its flexible representation of the pattern of efficiency in the data. The normal-truncated normal distribution, like the gamma distribution, also allows a wider range of distributional shapes, which include ones with nonzero modes.

3.5: Profit Frontiers and Profit Efficiency

We consider a producer who uses inputs $x \in \mathbb{R}_{+}^{n}$ available at prices $w \in \mathbb{R}_{++}^{n}$ to produce $y \in \mathbb{R}_{+}^{m}$ for sale at prices $p \in \mathbb{R}_{++}^{m}$ so as to maximize profit $(\max \pi = \mathrm{p'y} - \mathrm{w'x})$. Kumbhakar and Lovell (2000) defined a measure of profit efficiency as a function

$$\pi\mathrm{E}(y,x,p,w) = \frac{p'y - w'x}{\pi(p,w)} \quad \text{provided } \pi(\mathrm{p,w}) > 0 \tag{3.31}$$

Profit efficiency requires (either input-oriented or output-oriented) technical efficiency and both input allocative efficiency and output allocative efficiency. These efficiencies are not collectively sufficient for profit efficiency, since profit efficiency also requires technical efficiency and both types of allocative efficiency to be achieved at the proper scale. Thus, profit efficiency also requires a type of scale efficiency. From equation (3.31), profit efficiency is a ratio of actual profit to maximum profit. It is also obvious that $\pi\mathrm{E}(y,x,p,w) \le 1$.

From figure 3.5, a producer facing prices (p', w') achieves $\pi\mathrm{E}(y,x,p,w) = 1$ at output-input combination E, and $\pi\mathrm{E}(y,x,p,w) < 1$ for all other feasible output-input combinations. The measure of profit efficiency is bounded above by unity, and it achieves its upper bound if and only if a producer actually adopts a profit-maximizing combination of inputs and outputs. However, the measure of profit efficiency is not bounded below by zero, since negative actual profits are possible.

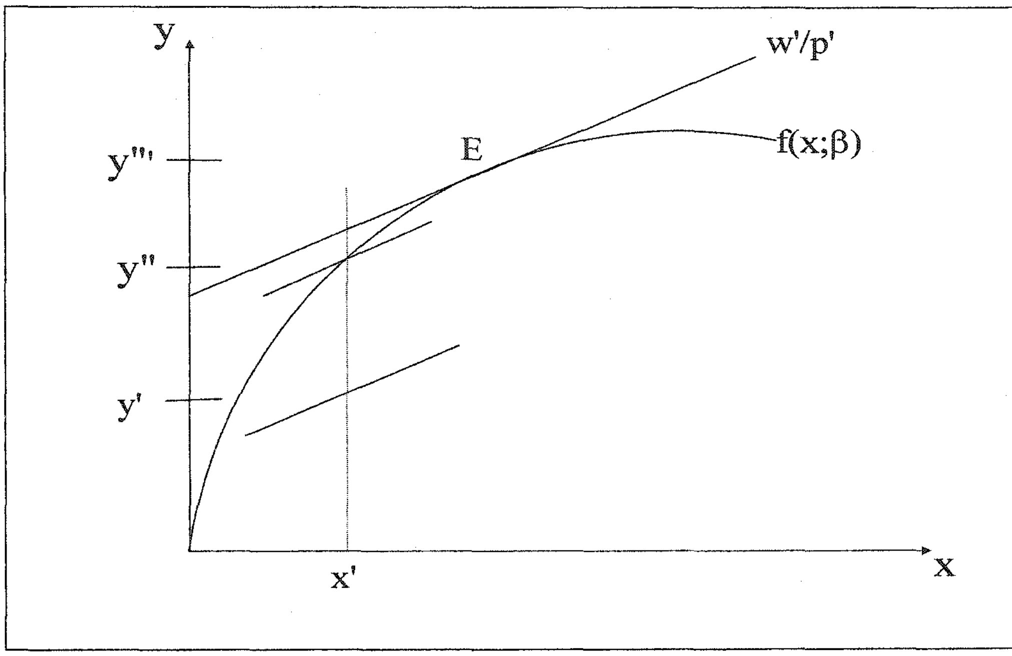


Figure 3.5: Measurement of Profit Efficiency

Following Kumbhakar and Lovell (2000), the profit efficiency satisfies the following properties:

i. $\pi E(y,x,p,w) \leq 1$, with $\pi E(y,x,p,w) = 1$ iff $y = (p,w)$, $x = (p,w)$ so that $(p'y - w'x) = \pi(p,w)$

ii. $\pi E(\lambda y,x,p,w) \geq \pi E(y,x,p,w)$, $\lambda \geq 1$ nondecreasing in output

iii. $\pi E(y,\lambda x,p,w) \leq \pi E(y,x,p,w)$, $\lambda \geq 1$ nonincreasing in input

iv. $\pi E(y,x,\lambda p,\lambda w) = \pi E(y,x,p,w)$, $\lambda > 0$ homogeneous of degree zero in output prices and input prices collectively (but not generally homogeneous in output prices and input prices separately)

If we consider an output-oriented measure of profit efficiency and a producer facing prices (p', w') and using input vector x' to produce output y', then increasing output-oriented technical efficiency by increasing output radially to y'' will increase profit and profit efficiency. However, output allocative inefficiency can remain at y'', and input vector x' may exhibit input allocative inefficiency. Both output allocative inefficiency at y'' and input allocative inefficiency at x' reduce profit beneath $\pi(p', w')$. Therefore, they both contribute to profit inefficiency. Even after accounting for technical inefficiency and both types of allocative inefficiency, scale inefficiency can remain.

From the graph, (y'', x') occurs at an inefficiently small scale to maximize profit. As such, scale inefficiency constitutes the final source of profit inefficiency. Following all these components of the measure of profit efficiency, the attainment of maximum profit requires technical efficiency (use of the right input mix in light of the input price w), production of the right output mix in light of the output price p, and operation at the right scale in light of (p, w).

3.6: Conclusion

There has been a rapid methodological development in frontier production estimation since the seminal articles by Farrell (1957) and Aigner and Chu (1968). There exist, today, a multitude of approaches (deterministic as well as stochastic) used to estimate production function. Production function, as presented in any standard economic textbook, gives the maximum possible output, which can be produced from given quantities of a set of inputs. A cost function gives the minimum level of cost at which it is possible to produce some level of output, given input prices. Also, a profit function gives the maximum profit that can be attained, given output price and input price. For all these functions (production, cost, profit) the concept of maximality or minimality is important. These functions set a limit to the range of possible observations, such that we can meaningfully apply the concept of frontier in each case. Accordingly, a measure of inefficiency is characterized by the amount by which a firm lies below its production and profit frontiers, and the amount by which it lies above its cost frontier.

Frontier production, cost, and profit functions have been applied in the analysis of financial institutions efficiency. These frontier functions permitted sophisticated analysis of technical efficiency. The economic theory of production is not only based on efficient subsets of production sets, but also on the value duals such as the minimum cost functions and maximum profit functions. The envelope properties that yield the cost-minimizing input demand functions and profit maximizing output supply and input demand functions play a significant role. We, therefore, place emphasis on efficient production and its consequences thereof while also applying the concept of frontier on these bounded functions.

The concept of technical efficiency date back to Koopmans (1951). His definition of technical efficiency (p. 60) states, "a producer is technically efficient if an increase in any output requires a reduction in at least one other output or an increase in at least one input, and if a reduction in any input requires an increase in at least one other input or a reduction in at least one output". Therefore, a producer producing the same outputs with less of at least one input or using the same inputs to produce more of at least one output is considered technically inefficient. Debreu (1951) and Farrell (1957) introduced a measure of technical efficiency. Farrell's (1957) seminal work provided a foundation for redirecting attention from the production function specifically to the deviations from that function thereby respecifying the regression and the techniques accordingly.

Farrell's (1957) analysis was based on the assumption that the efficient production in known. That is, he assumed that there are methods of computing the observed performance of a production unit with some postulated standard, which is not usually the case. He, therefore, suggested the use of mathematical programming approach or the econometric approach. The econometric approach is stochastic (and so attempts to distinguish the effects of noise from the effects of inefficiency), parametric, and confound the effects of misspecification of functional form (of both technology and inefficiency) with inefficiency. The programming approach is nonstochastic, and lumps noise and inefficiency together and calls the combination inefficiency. It is also nonparametric and less prone to the type of specification error confronting the parametric approach. This study adopts the parametric approach (the stochastic frontier analysis). The stochastic frontier approach (SFA), while allowing for technical inefficiency, acknowledges that random shock outside the control of producers can affect output. The SFA is also notable for its internal consistency and ease of implementation.

In estimating technical efficiency of each producer, we require the distributional assumptions (normal-half normal, normal-truncated normal, exponential, and gamma distributions). The normal-half normal distribution has been the arbitrary choice in most bank efficiency studies, due to a lack of a priori justification for selecting a particular distributional form for the technical inefficiency effects. The normal-half normal distribution has a mode at zero, which implies that there is the highest probability that the inefficiency effects are in the neighborhood of zero. This also implies that technical efficiency is high, which might not be evident in practice. These shortfalls led to the development of the normal-truncated normal distribution adopted in this study.

CHAPTER 4: MODEL SPECIFICATIONS

4.1: Introduction

There are varying views as to the relevant approach used to define and measure bank output. In banking, many balance sheet items have both input and output characteristics. For example, deposits are input to an extent that they provide loanable funds. However, they can also be treated as output because they provide liquidity, payments, and safekeeping services to depositors. Various approaches have been applied in banking literature to try to address this dilemma. These approaches are asset approach, the user cost approach, and the value added approach. However, in this analysis we employ the profit function that alleviates any misspecification and/or mismeasurement problems associated with the definition and measure of bank output.

This chapter present the models specified and estimated in this dissertation. Section 1 presents an analysis of the profit efficiency concept adopted in the study. This section also presents the textbook analysis of the importance of the profit function compared to the cost function. Section 2 outlines the alternative profit efficiency concept and its specification. This section also outlines the pros and cons of the alternative profit efficiency concept (as it pertains to bank efficiency studies) compared to the traditional profit efficiency concept. Section 3 presents the stochastic alternative profit function as specified in a flexible translog functional form.

This section also addresses the advantages and disadvantages of the translog functional form, and compares the translog functional form with the Fourier-Flexible form. We also discuss the variables used to measure the alternative profit frontier.

Section 4 presents a model for inefficiency effects as applied in this dissertation. This section discusses the shortfalls of the two-stage approach adopted by other banking studies, and outlines the importance of the single-stage maximum likelihood estimation as adopted in the study. This section also presents and discusses variables that determine technical inefficiencies, and variables employed to test the diversification hypothesis. Lastly, section 5 discusses the sample data employed in the study.

4.2: Efficiency Concept

In measuring financial institution efficiency, the choice of measurement concept is crucial. There are three main economic efficiency concepts (cost, standard profit, and alternative profit efficiencies) that are based on the economic optimization in reaction to market prices and competition. Berger and Mester (1997) gave a detailed explanation of these three efficiency concepts. For the study of mergers and acquisitions (M&A), the analysis of profit efficiency has proven to be the most appropriate. Various studies[4] have since applied the profit efficiency in analyzing bank efficiency. However, few studies have applied the profit efficiency concept in analyzing M&As.

---

[4] Berger and Mester (1997), Berger, Hancock, and Humphrey (1993), Berger (1998), Akhavein, Berger, and Humphrey (1997), Berger and Mester (2003).

The profit efficiency concept is based on the assumption that a firm's objective is to maximize variable profits subject to some exogenous business conditions. Therefore, profit X-efficiency is a general concept that includes the cost X-efficiency effects of the merger plus any revenue and cost effects of changes in output. Therefore, in evaluating the overall performance of the firm, the profit efficiency is superior to the cost efficiency.

From a textbook analysis (e.g. Varian 1992, Mas-Colell et al. 1995), given a price vector $p \gg 0$ and a production vector $y \in \mathbb{R}^n$, the profit generated by implementing y is $p \cdot y = \sum_{i=1}^{n} p_i y_i$ (i.e. the total revenue minus total cost). Therefore, given the technological constraints represented by its production set Y, the firm's profit maximization problem is

$$\begin{aligned} &\underset{y}{\text{Max}}\ p \cdot y \\ &\qquad \text{s.t. } y \in Y \end{aligned} \tag{4.1}$$

that can be restated as

$$\begin{aligned} &\underset{y}{\text{Max}}\ p \cdot y \\ &\qquad \text{s.t. } F(\cdot) \leq 0 \end{aligned} \tag{4.2}$$

where $F(\cdot)$ is the transformation function that describes Y. Thus, given a production set Y, the firm's profit function $\pi(p)$ associates to every p the amount $\pi(p) = \max\{p \cdot y : y \in Y\}$, the value of the solution to the maximization problem. If the transformation function $F(\cdot)$ is differentiable, then we can use the first-order conditions (FOC) to characterize the solution to the profit maximization problem. If $y^* \in y(p)$, then for some $\lambda \geq 0$, y^* must satisfy the first order conditions

$$p = \lambda \nabla F(y^*) \tag{4.3}$$

where y(p) is the firm's supply correspondence at p. It is important to note that, if the production set Y is convex, then the FOC in equation (4.3) is not only necessary but also sufficient for the determination of a solution to the profit maximization problem.

One important implication of the firm choosing a profit-maximizing production plan is that there is no way to produce the same amount of outputs at a lower total input costs. That is, the cost minimization is a necessary condition for profit maximization. The cost minimization problem can be stated as

$$\min_{x \geq 0} \ w \cdot x \qquad \text{s.t. } f(x) \geq y \tag{4.4}$$

where x is the nonnegative vector of inputs, f(x) is the production function, y is the amount of output, and $w \gg 0$ is the vector of input prices. The cost function $c(w, y)$ gives the optimized value of the cost minimization problem. Also, if x^* is the optimal solution in the cost minimization problem, and if the production function $f(\cdot)$ is differentiable, then for some $\lambda \gg 0$

$$w \geq \lambda \, \nabla \, f(x^*) \tag{4.5}$$

must hold for every input. From the analysis of both the profit maximization problem and the cost minimization problem, we can infer that, profit maximization implies that input choices are cost minimizing for the chosen output level y. In addition, the cost function does not contain more information than the profit function. Under the convexity restrictions, there is a one-to-one correspondence between profit and cost functions. That is, we can recover the production set from either function, and then derive the other function. Therefore, we can use the cost function to restate the firm's profit maximizing production level as

$$\underset{y \geq 0}{Max}\, py - c(w, y) \tag{4.6}$$

where the necessary FOC for y^* to be profit maximizing is

$$p - \frac{\partial c(w, y^*)}{\partial y} \leq 0, \text{with equality if } y^* > 0 \tag{4.7}$$

That is, at an interior optimum (if $y^* > 0$), price equals marginal cost. We also note that the FOC of the cost minimization problem coincides with the FOC of the profit maximization if and only if $\lambda = p$. Also noting that λ (the multiplier on the constraint in the cost minimization problem), is equal to $\frac{\partial c(w, y)}{\partial y}$.

Berger, Hancock, and Humphrey (1993:317) stated that, "the primary advantage of the profit function is that it allows for measurement of inefficiencies on the output side as well as the input side of the firm". That is, the profit function incorporates the revenue effects of producing at incorrect levels or mixes of outputs, in addition to the cost effects of employing the wrong levels or mixes of inputs.

Other advantages of the profit function include the reduction of misspecification and mismeasurement problems prevalent in banking studies. That is, there is a lack of clarity on whether a balance sheet item should be classified as input or output. However, the profit function specifies inputs and outputs almost symmetrically, thereby lessening any specification or measurement error. Profit function also allows us to pinpoint better the sources of inefficiency.

Looking at M&As, Akhavein et al. (1997:01) indicated that, "mergers could improve cost efficiency if the reconfiguration of output associated with the merger increases costs, or it reduces costs more than it reduces revenue". It is therefore appropriate to apply profit efficiency in an analysis of mergers rather than cost efficiency. As noted in chapter 3, the underlying assumption for the profit frontier $\pi(p, w, \beta)$ is that, prices are exogenous and producers seek to maximize profit by selecting outputs and inputs under their control. That is, because producers operate in competitive markets, prices are, therefore, exogenous.

However, if producers have some degree of control or monopoly power in their product markets, the demands would be exploited to determine output prices and quantities jointly, such that only input prices will be exogenous. Therefore, the standard cost frontier (which treats output as being exogenous) and the standard profit frontier (which treats output prices as being exogenous) cannot provide an appropriated framework to evaluate producer performance. Therefore, various authors[5] introduced the notion of an "alternative" profit frontier to bridge the gap between a cost frontier and a profit frontier.

4.2.1: Alternative Profit Efficiency

We employ the alternative profit function rather than the standard profit function in measuring bank efficiencies. The alternative profit function uses the same dependent variable as the standard profit function, but the same explanatory variables as the cost function. The alternative profit frontier, however, lacks some of the advantages prevalent in the standard profit function. Unlike the standard profit

[5] Berger, Cummins, and Weiss (1997), Hasan and Hunter (1996), Akhavein, Berger, and Humphrey (1997), Berger and Mester (1997), Humphrey and Pulley (1997), Lozano Vivas (1997), and Berger and Mester (2003).

function, the alternative profit function requires choosing whether deposits are inputs or outputs.

Because this study adopts the alternative profit efficiency concept, we also follow the "value-added" approach (Berger and Humphrey 1992b) in defining and measuring bank outputs. The value-added approach defines output as those activities that have substantial value added as judged by using an external source of operating cost allocations. That is, activities that have large expenditure on labor and physical capital. It considers all liability and asset categories to have some output characteristics rather than distinguishing inputs from outputs in a mutually exclusive way.

The value-added approach also explicitly uses operating cost data as part of the return or cost not accounted for by the difference between measured financial flows and marginal opportunity costs. Therefore, the value-added approach is considered the best approach for accurately estimating changes in bank technology and efficiency over time. The value-added approach, as applied in this study, identifies the major categories of produced deposits (demand, time and savings) and loans (real estate, commercial, and installment) as important outputs because they are responsible for the great majority of value added.

The alternative profit efficiency also assumes exogenous output quantities unlike output prices in the standard profit function. It also assumes exogenous input prices, i.e. no monopsony power. Under the alternative profit efficiency, bank efficiency is measured by the difference between a bank's actual profits and the maximum profits given its output levels rather than output prices. That is, how close a bank is to earning maximum profits given its output levels rather than its output prices. Following Berger and Mester (1997), the alternative profit function is specified in log form as:

$$\begin{gathered}\text{Ln}\,(\pi + \theta) = \text{f}_{a\pi}(\text{X}_\pi) + \ln \text{v}_{a\pi} + \ln \xi_{a\pi} \\ \text{where } \text{X}_\pi \equiv (\ln w, \ln y, \ln z, \ln r)\end{gathered} \tag{4.8}$$

where π = The profits of the firm; θ = Constant added to every firm's profit so that the natural log is taken of a positive number; y = Vector of variable outputs; w =Vector of variable input prices; r = Altman Z-score or measure of bank insolvency risk; z = Fixed Input Quantities; Ln $\xi_{a\pi}$ = Captures both Technical and Allocative X-inefficiency; Ln $V_{a\pi}$ = Represents the random error; Vs are assumed to be iid ~ $N(0,\sigma^2_v)$ incorporated in the model to reflect the random disturbance that is independent of the explanatory variable and the ξs. The ξs are nonnegative random disturbances that capture the degree of technical inefficiency in production. We also assume that the ξs (degree of inefficiency) are independently distributed and truncated at zero $N\left(m'\ \delta, \sigma^2_\xi\right)$.

The alternative profit efficiency is then expressed as a ratio of predicted actual profits to the predicted maximum profits for a best practice firm.

$$\begin{aligned}\text{Alt}\,\pi\,\text{EFF}^b &= \frac{a\hat{\pi}^b}{a\hat{\pi}^{\max}} \\ &= \frac{\left[e^{\left[\hat{f}\left(y^b, w^b, z^b, r^b\right)\right]} \times e^{\ln \hat{u}^b_{a\pi}}\right] - \theta}{\left[e^{\left[\hat{f}\left(y^b, w^b, z^b, r^b\right)\right]} \times e^{\ln \hat{u}^{\max}_{a\pi}}\right] - \theta}\end{aligned} \tag{4.9}$$

We employ the alternative profit concept even if firms do not actually take their outputs as given and maximize profits as implied by the alternative profit specification. The alternative profit concept not only bridges the gap between the standard cost and profit function, but also becomes more relevant if one or more assumptions underlying the cost efficiency and the standard profit efficiency do not hold. Berger and Mester (1997) identified some violations of these assumptions such that the alternative profit concept becomes more useful and provides valuable information in efficiency measurement. Therefore, the alternative profit efficiency is appropriate if one or more of the following conditions hold:

First, there is the presence of substantial unmeasured differences in the quality of banking services. The competitiveness of the output markets might lead to customers willing to pay more for higher quality. Therefore, the banking firm should be able to extract higher revenue to compensate for the extra costs of producing these high quality services. This might lead to different categorization of banks, which are sorted into market niches that differ by service quality or intensity. Because the alternative profit efficiency adds revenue to their dependent variable, it is able to account for this additional revenue earned through the provision of high quality services, thereby offsetting the additional costs. However, the cost and standard profit efficiency may consider these unmeasured improvements in quality as deterioration in performance, which might penalize high quality banks.

Second, variable outputs are not completely variable, which might lead to scale bias. The standard profit efficiency treats both small and large banks as if they should have the same variable outputs when facing the same input and output prices, and fixed netputs as specified in the standard profit function. This leads to scale bias where large banks can have higher profits that are not explained by the exogenous variables. Smaller banks might not be able to match the output levels of the bigger banks, such that bigger banks are considered to have higher standard profit efficiency. The alternative profit efficiency addresses this shortfall by evaluating performance at the existing output levels as it compares the ability of banks to generate profits for the same levels of output. Statistically, outputs are held constant in the alternative profit function.

Third, standard profit efficiency assumes that output prices are exogenous. That is, banks can sell as much output as they prefer without altering the prices. This might lead to an understatement of the standard profit efficiency. That is, if a firm's output is below an efficient scale, it might have to lower its prices in order to raise output. However, the alternative profit efficiency might measure the increase in the exercise of market power as in improvement in best practice. That is, a firm with market power may be able to increase revenue more than costs by increasing service quality due to absence of competitors in that quality niche. Berger and Hannan (1989) also showed that banks in more concentrated local markets have some control over prices as they pay lower rates to small depositors. Hannan (1991) showed that banks charge higher rates to small borrowers, which implies that prices are not necessarily exogenous.

Lastly, in most banking studies output prices are not accurately measured (this study in not immune to this shortfall). That is, the Call Reports data does not include a measure of prices. As such, we normally construct prices as ratios of revenue flows to stock of assets, which is subject to many noises. Therefore, it is difficult to measure the standard profit function accurately. If output prices are measured accurately, they should be strongly related to profits and should also explain a substantial portion of the variance of profits in the standard profit function. However, if not well measured, the predicted part of the standard profit function would explain less of the variance of profits and yield more errors in the estimation of the efficiency term. The alternative profit efficiency might be an appropriate alternative as it can measure some components of the change in profitability that is independent of output prices. Berger and Mester (1997), and Humphrey and Pulley (1997) found that the alternative profit function fits the data better than the standard profit function for every year during 1984-1997 period.

4.3: Empirical Stochastic Frontier Profit Model

The choice of a functional form in an empirical study is of prime importance, since the functional form can significantly affect the results. A flexible functional form is generally preferred since it does not impose general restrictions on the parameters nor the technical relationships among inputs. In this analysis, we choose a flexible translog functional form. Apart from being simple, the translog functional form provides consistent inefficiency rankings relative to other functional forms. Therefore, our specification of the stochastic frontier alternative profit function is:

$$
\begin{aligned}
\ln\left[\frac{\pi}{z_2}+\theta\right] = {} & \alpha_0+\sum_{i=1}^{4}\alpha_i\ln\left(\frac{y_i}{z_2}\right)+\frac{1}{2}\sum_{i=1}^{4}\sum_{r=1}^{4}\delta_{ir}\ln\left(\frac{y_i}{z_2}\right)\ln\left(\frac{y_r}{z_2}\right) \\
& +\sum_{j=1}^{4}\beta_j\ln\left(w_j\right)+\frac{1}{2}\sum_{j=1}^{4}\sum_{k=1}^{4}\gamma_{jk}\ln\left(w_j\right)\ln\left(w_k\right) \\
& +\sum_{i=1}^{4}\sum_{j=1}^{4}\rho_{ij}\ln\left(\frac{y_i}{z_2}\right)\ln\left(w_j\right) \\
& +\varphi_0\ln R+\frac{1}{2}\varphi_1\left(\ln R\right)^2 \\
& +\sum_{i=1}^{4}\kappa_i\ln\left(\frac{y_i}{z_2}\right)\ln R+\sum_{j=1}^{4}\mu_j\ln\left(w_j\right)\ln R \\
& +\omega_0\ln\left(\frac{z_1}{z_2}\right)+\frac{1}{2}\omega_1\ln\left(\frac{z_1}{z_2}\right)^2 \\
& +\sum_{i=1}^{4}\phi_i\ln\left(\frac{y_i}{z_2}\right)\ln\left(\frac{z_1}{z_2}\right)+\sum_{j=1}^{4}\eta_j\ln\left(w_j\right)\ln\left(\frac{z_1}{z_2}\right) \\
& +\ln v_\pi+\ln\xi_\pi
\end{aligned}
\tag{4.10}
$$

Berger and Mester (1997) stated that one of the drawbacks with studies on bank efficiency before the introduction of the Fourier-Flexible form was the reliance on the translog frontier functions. As such, recent literature on scale economies and cost efficiencies in the banking industry adopted the Fourier-Flexible form following Bauer and Ferrier's (1996) claim that it is the global approximation capable of providing a better fit to bank data.[6]

Most of these studies indicated that the translog function does not fit the data well especially if the data are far from the mean in terms of output size or mix. McAllister and McManus (1993), and Mitchell and Onvural (1996) pointed out that variation or differences in results on scale economies across studies might be due to the ill fit of the translog function across a wide range of bank sizes. Therefore, the

[6] Other studies include McAllister and McManus (1993), Berger, Cummins, and Weiss (1997), Berger and DeYoung (1997), Berger and Mester (1997), Berger and Humphrey (1997).

Fourier-Flexible form was considered more flexible than the translog. It was also considered a global approximation to virtually any cost and profit function. Berger and DeYoung (1997) found that measured inefficiencies were about twice as large when the translog was specified in place of the Fourier-Flexible form.

Shaffer (1998) while highlighting some of the shortcomings of the translog functional form indicated that the translog form cannot incorporate zero quantities of any output and typically exhibits substantial multicollinearity among the various terms and their squares and cross products. He also indicated that the translog form tend to impose spurious U-shaped average cost structure in the face of monotonically declining true average cost data. That is, the translog form in practice generally cannot portray monotonically declining average costs. However, Shaffer (1998) also indicated that the translog form has the theoretical advantage of being able to fit exactly the level, first derivative, and second derivative of an arbitrary function at a point.

Altunbas and Chakravarty (2001) urged some caution on the growing use of the Fourier-Flexible specification of the frontier function to investigate bank efficiency. In their analysis of the Fourier-Flexible form as a better fit to the data compared to the translog function, they found that the translog functional form does a better job in the prediction and forecasting for the largest 5% of the banks in their sample. Thus, the goodness of fit criterion is not necessarily an indication of goodness in prediction or a reliable indicator of the claim that the Fourier-Flexible functional form is a global approximation to any cost or profit function, and that it fits the data for financial institutions better than the translog.

In addition, the optimal number of estimated coefficients in a flexible Fourier form is equal to the number of sample observations to the 2/3 power. In this study, we have 347 observations. According to the standard requirement of the Fourier flexible form, we need about $(347)^{2/3} \approx 49$ estimated coefficients. Our translog flexible form, as specified in equation (4.10), has 65 coefficients, which is more than what is required by the Fourier flexible form. Therefore, a properly specified Fourier form would be no more flexible (in terms of the number of estimated coefficients) than our translog.

In equation (4.10), the dependent variable of the alternative profit function is specified as

$$\ln\left[\frac{\pi}{z_2} + \left|\left(\frac{\pi}{z_2}\right)^{\min}\right| + 1\right] \tag{4.11}$$

where $\left|\left(\frac{\pi}{z_2}\right)^{\min}\right|$ is the absolute value of the minimum value of $\left(\frac{\pi}{z_2}\right)$ over all banks during the same period. Thus, we add a constant $\theta = \left|\left(\frac{\pi}{z_2}\right)^{\min}\right| + 1$ to the dependent variable in every firm so that we can take natural logs of a positive number.

Following Berger and Mester (1997), we also specify all of the profit, variable output quantities, and fixed input quantities as ratios to the fixed equity capital input, z_2. This helps address various shortfalls. For example, smallest firms have profits many times smaller than those of the largest firms and vice versa. As such, large firms would have random errors with much larger variances than the small firms would in the absence of normalization. Therefore, normalization helps alleviate this heteroscedasticity problem. We should also note that we are deriving the inefficiency term $\ln \xi_{\pi}$ from the composite residuals and this might make the variance of the efficiencies dependent on the size of the bank in the absence of normalization.

By normalizing the variable output, the dependent and the independent variables would be of the same order of magnitude rather than been skewed towards large banks. Therefore, scale bias is also reduced as we can now express profits or asset per dollar of equity, which alleviate differences in profits and asset sizes between large and small banks. That is, large banks are prone to have higher profits for given set of prices, while small banks cannot achieve those profits levels, especially in the short run. However, small banks can easily attain the profits per dollar of equity and assets per dollar of equity of the large banks. Berger and Mester (1997) also indicated that normalization by equity capital has economic meaning. That is, the dependent variable becomes the return on equity (ROE) or a measure of how well banks are using their scarce financial capital. That is, banking is the most highly financially leveraged industries. Shareholders are mostly interested in their rate of return on equity (ROE), which is a measure closer to the goal of the bank than maximizing the level of profits.

Normalization by the financial equity capital also follows from the choice of equity capital as a fixed input quantity. That is, equity capital is very difficult and costly to change substantially except over the long run. Although we specify physical capital (premises and equipment) and equity capital as fixed input quantities, fixed assets are very small in banking, only about 20% (Akhavein et al. 1997) as large as equity, and can be increased much more quickly and easily than equity. Therefore, equity capital is preferred as a normalization variable besides being the fixed input quantity. Also, if equity were not specified as fixed, the largest banks may be measured as the most profit efficient simply because their higher capital levels allow them to have the most loans.

4.3.1: Variables

Table 4.1 below gives definitions of the variables in the alternative profit function, their sample means, medians, and standard deviations.

Table 4.1
Variables employed in measuring the Alternative Profit Efficiency - means, medians and standard deviations for 1997

Symbol	Definitions	Mean	Median	Std. Dev.
Dependent Variable ($1000)				
π	Profits	355,072	14,545	706,202
Variable output quantities ($1000)				
Y_1	Demand Deposits	2,921,078	143,075	5,610,506
Y_2	Real Estate Loans	8,528,412	471,123	16,556,212
Y_3	Consumer Loans	2,007,645	94,117	3,495,424
Y_4	Business Loans (C&I)	5,778,665	137,968	11,947,830
Variable input prices				
W_1	Unit price of Purchased Funds	1.010702	1.010102	0.00421
W_2	Unit price of Core Deposits	0.025028	0.025023	0.00624
W_3	Unit price of Physical Capital	0.32146	0.26722	0.64981
W_4	Unit price of Labor	37.16756	37.51563	9.30837
Fixed input quantities ($1000)				
Z_1	Physical Capital	395,317	21,140	778,050
Z_2	Financial Equity Capital	3,006,808	126,553	6,258,621
R	Z-Score: Measure of Insolvency risk	20.27502	18.68850	9.05659

- All stock values are quantities as of December Call Reports and all prices are flows over the year divided by these stocks. All of the continuous variables that can take on the value 0 have 1 added before taking logs. This applies to the y's. For π, an additional adjustment was made because profits can take a negative value (see text).

Although the continuous variables are generally expressed in natural logs in the profit function, we show the means, medians, and standard deviations of the levels. The variable R (i.e. measure of risk) is the Altman Z-Score, which is an indicator of the probability of bankruptcy.

Bankruptcy is defined as the case in which $\tilde{\pi} < -E$ or losses (negative profits) exceed equity. If A = Total Assets, $\tilde{r} = \frac{\tilde{\pi}}{A}$ and $k = \frac{-E}{A}$, we can express the probability of bankruptcy as

$$P(\tilde{\pi} < -E) = P(\tilde{r} < k) = \int_{-\infty}^{k} \phi(r)dr \tag{4.12}$$

where P(.) is a probability and ϕ(.) is the probability density function of r. If r is normally distributed, then we can rewrite (4.12) as,

$$p(\tilde{r} < k) = \int_{-\infty}^{z} N(0,1)dz \tag{4.13}$$

$z = \frac{k-\rho}{\sigma}$, where ρ = true mean of the $\tilde{r}$ distribution, σ = Standard deviation, and z = number of standard deviation below the mean by which profits would have to fall in order to eliminate equity. Z is the risk measure used here except that sample estimates are substituted for ρ and σ, and the estimated value of –z (z is a negative number) is labeled Z.

High values of Z are associated with low probabilities of failure. Z increases with the ratio of equity to assets (-k) and with the mean rate of return on assets ρ. Z decreases with the volatility of assets returns, σ. It is also important not to compute risk measures based on industry aggregate profits, assets, and equity as this would result in averaging rates of return across firms within an industry. That is, we should compute Z for each firm using individual firm rate of return time series[7].

4.4: Modeling Inefficiency Effects

Because efficiency scores are performance indicators, researchers are interested in explaining their distribution. Therefore, the efficiency scores become attractive dependent variable in a second-stage regression analysis. Most empirical studies investigated the determinants of technical inefficiencies among firms in the banking industry by regressing the predicted inefficiency effects observed from an estimated (stochastic) frontier upon a vector of firm-specific factors (e.g. bank size, financial ratios, etc.) in a second-stage analysis. The results of the second-stage analysis are then used to guide public policy or managerial decision-making. In none of the previous studies in banking are the determinants of technical inefficiency used jointly with other variables of the model.

The two-stage analysis is outlined as follows: The **first stage** involves the specification and estimation of the stochastic frontier production (cost and profit) functions and the prediction of the technical inefficiency effects, (ξ_i). In this stage, we assume that the inefficiency effects (ξ_i) are independently and identically distributed. This allows us to employ Jondrow et al. (1982) approach to predict the values of the technical inefficiency effects. The estimation of the inefficiency effects in this stage ignores the fact that inefficiencies might be a function of some other variables.

---

[7] Even if $\tilde{r}$ is not normally distributed, z is still a useful risk measure as long as ρ and σ exists. We can invoke the Bienayme-Tchebycheff inequality and $p(r \lesssim k) \leq [\frac{\sigma}{\rho - k}]^{\frac{1}{2}}$. Then z is the upper bound or worst-case probability of bankruptcy.

In addition, the efficiency scores derived in the first stage analysis are bounded, either by zero and one or below by one. Therefore, they must be transformed prior to use as dependent variable in the second stage regression. Otherwise, we should employ appropriate limited dependent variable regression techniques. For example, we can transform the technical inefficiencies into the logit form: $\ln\left[\frac{|E|}{1-|E|}\right]$ where E = technical inefficiency. Then we can regress the transformed variable on several other variables hypothesized to contribute to technical inefficiency. This logit transformation assures a two-sided distribution of efficiency rather than the half-normal assumed into the stochastic frontier approach. The ordinary least squares estimates of the parameters derived are, therefore, consistent.

The **second stage** analysis involves the specification of a regression model for the predicted technical inefficiency effects. That is, we assume that the predicted inefficiency effects are a function of a number of firm-specific factors. Once the inefficiencies are estimated, they are further regressed on a set of variables that are supposed to explain differences in technical inefficiency among firms. The parameters of this second stage inefficiency model have been generally estimated by using ordinary least squares (OLS). By regressing the estimated inefficiency effects against firm-specific variables, we imply that these inefficiency effects are not identically distributed unless all the coefficients of the factors are simultaneously equal to zero. The two-stage analysis poses other serious problems:

1. Technical inefficiency may be correlated with inputs causing inconsistent estimates of the parameters as well as technical inefficiency.

2. Standard ordinary least squares (OLS) results in the second stage may not be appropriate since technical inefficiency (the dependent variable) is one-sided.
3. The estimated value of the technical inefficiency should be nonpositive for all observations.
4. The meaning of the residual term in the second-stage regression is not clear.

On this second-stage analysis, the nature of the production paradigm underlying the two-stage model is also of concern. As Lovell (1993) pointed out, explanatory variables in this formulation influence the efficiency with which inputs generate outputs. They, however, do not influence the transformation process itself. If we characterize the first-stage production relationship as

$$y_i = f\left(x_i; \beta\right) \cdot e^{\{v_i - \xi_i\}} \qquad \mathrm{i} = 1, \cdots, \mathrm{N} \tag{4.14}$$

then, the second-stage explanatory relationship can be expressed as

$$\mathrm{e}^{\xi_i} = g\left(\mathrm{m_i}\ ; \delta\right) \cdot e^{\tau_i} \qquad \mathrm{i} = 1, \cdots, \mathrm{N} \tag{4.15}$$

where $\exp\{\xi_i\} = TE_i$ is to be explained with a vector of variables z_i and $\tau_i \sim N\left(0, \sigma_\tau^2\right)$ captures statistical noise. The formulation is sensible if z_i and x_i are uncorrelated. However, if z_i and x_i are correlated, estimates of β are biased, and presumably so too are estimates of TE_i that are to be explained in the second stage (Lovell 1993).

It is also important to think about what type of variables (inputs and outputs) belongs into the first stage analysis, and what explanatory variables belong in the second stage regression. Lovell (1993) suggested that variables over which the decision maker has no control during the time period under consideration (e.g. quasi-fixed variables, site-specific characteristics, demographic characteristics, weather, etc.) belong in the second stage regression. Variables under the control of the decision maker during the period under consideration belong in the first stage analysis.

The limitations of the two-stage analysis prompted further research into better ways of modeling these inefficiency effects. Several authors noted the inconsistency of the two-stage analysis and specified stochastic frontier models in which they defined the inefficiency effects to be explicit functions of some firm-specific factors. These authors also estimated all parameters in a single-stage maximum likelihood procedure.

Pitt and Lee (1981) investigated the sources of technical inefficiency by specifying that firm intercepts in the stochastic frontier are a function of firm characteristics. The authors regress the estimated firm intercepts on the specified characteristics or incorporated the firm characteristics into the production frontier and jointly estimate the parameters involved. Kumbhakar, Ghosh, and McGuckin (1991) assumed that the technical inefficiency effects are non-negative truncation of a normal distribution with mean, which is a linear function of exogenous factors whose coefficients are unknown, and an unknown variance. Reifschneider and Stevenson (1991) proposed a model for the inefficiency effects of the stochastic frontier production function involving the sum of a non-negative function of relevant explanatory variables and a non-negative random variable, which is assumed to have half-normal, exponential or gamma distribution.

Huang and Liu (1994) considered a stochastic frontier production function in which the non-negative technical inefficiency effects are a linear function of variables involving firm characteristics. The additive random error of the inefficiency model is assumed to be the truncation of a normal distribution with mode zero, whose point of truncation is dependent on the firm characteristics, such that the inefficiency effects are non-negative. The authors assumed that their explanatory variables in the inefficiency model are a function of firm-specific variables and the explanatory variables of the stochastic frontier. This makes their model a non-neutral shift of the traditional average response function, in that the marginal product of inputs and marginal rate of technical substitution depend on the firm-specific variables in the inefficiency model.

Battese and Coelli (1995) proposed a model for technical inefficiency effects in a stochastic frontier production function for panel data. Their model permits the estimation of both technical change in the stochastic frontier and time-varying technical inefficiencies, given that inefficiency effects are stochastic and have a known distribution.

Hjalmarsson, Kumbhakar, and Heshmati (1996) while comparing several models in the Data Envelopment Analysis, Deterministic Parameter Frontier Analysis, and the Stochastic Frontier Analysis (SFA) categories (with respect to magnitude of efficiency scores, efficiency ranking of units, and time pattern of efficiency), specified three variation of the stochastic frontier analysis in an attempt to model the inefficiency effects. First, the regular panel data model with time-varying technical inefficiency. The second model introduces some plant and production characteristic variables that can be used as controls instead of regular inputs. Lastly, they specified a model where technical efficiency is explained by variables other than the regular inputs.

All these analysis lead to the proposal of a one-stage model of the form

$$y_i = h(x_i; m_i; \delta) \cdot e^{\{v_i - \xi_i\}} \qquad i = 1, \cdots, N \qquad (4.16)$$

where m_i influences y_i directly rather than indirectly through ξ_i. m_i is a $K \times 1$ vector of firm specific inefficiency explaining variables, and δ is a $1 \times p$ parameter vector to be estimated[8].

In this paper, we shall model the error term, ε_i, as a two-component error structure. That is, $\varepsilon_i = V_i - \xi_i$. The symmetric random error component V_i is assumed to be iid $N(0, \sigma_V^2)$, independently distributed of the ξ_is. The inefficiency error component, ξ_i, is assumed to be independently distributed, such that ξ_i is obtained by truncation (at zero) of the normal distribution with mean $z_i\delta$ and variance σ^2; z_i is a (1xM) vector of firm-specific variables; and δ is an (Mx1) vector of unknown coefficients of the firm-specific inefficiency variables.

If ξ_i and V_i are independent, the joint density functions of $\varepsilon_i = V_i - \xi_i$ is

$$f(\varepsilon_i \mid \sigma^2, \lambda) = \sqrt{\frac{2}{\pi}} \cdot \frac{1}{\sigma} \cdot e^{\frac{1}{2}\left(\frac{-\varepsilon_i}{\sigma}\right)^2 \left[1 - F^*\left(\frac{-\varepsilon_i \lambda}{\sigma}\right)\right]} \qquad (4.17)$$

where $\sigma^2 = \sigma_v^2 + \sigma_\xi^2$; $\lambda = \sigma_\xi / \sigma_V$; f^* and F^* are the standard normal and standard normal cumulative density functions respectively.[9] Given the p.d.f.'s of ε_i, the log-likelihood functions of the observed random variable $\ln\left[\frac{\pi}{z_2} + \theta\right]$ can be expressed as

---

[8] However, Lovell (1993) pointed out that the choice of one-stage analysis does not solve the problem of correlation between m_i and x_i. It only replaces the omitted variable problem (two-stage model) with a multicollinearity problem.

$$\ln L\left[\ln\left(\frac{\pi}{z_2}+\theta\right) \mid \text{parameters}, \sigma_\xi^2, \lambda\right]$$
$$= \frac{n}{2}\ln\frac{2}{\pi} - n\ln\sigma - \frac{1}{2}\left(\frac{\sum\varepsilon_i}{\sigma}\right)^2 + \sum_1^n\left[\ln\left(1-F^*\left(-\frac{\varepsilon_i\lambda}{\sigma}\right)\right)\right] \qquad (4.18)$$

where $\varepsilon_i = \ln\left[\frac{\pi}{z_2}+\theta\right] - \underset{\sim}{X}_i'\beta$, and $\underset{\sim}{X}_i$ is a vector of the right hand side variables in (4.10). In equation (4.18) above, as λ goes to zero, the frontier approaches the OLS estimate.

4.4.1: Inefficiency Model

To investigate the existence of systematic influences on the inefficiency disturbances we incorporate firm-specific effects in the model of the inefficiency error. We assume that the inefficiency disturbance is composed of two factors: a factor reflecting systematic influences, and a random factor. Therefore, we can specify the inefficiency effects, ξ_i, in the stochastic frontier model (4.10) as

$$\xi_i = z_i\delta + \tau_i \qquad (4.19)$$

where ξ_i is the composite inefficiency level, Z_i is a vector of firm-specific inefficiency explanatory variables, and τ_i is the unexplained component of inefficiency error.

We assume that the random variable, τ_i, has the same general distribution as previously assumed for ξ_i. That is, we define τ_i by the truncation of the normal distribution with zero mean and variance σ^2 such that the point of truncation is $-z_i\delta$,

---

[9] Meeusen and van den Broeck (1977), and Stevenson (1980) provide the derivations for this expression.

i.e., $\tau_i \geq -z_i\delta$. These assumptions are consistent with the ξ_is being non-negative truncations of the $N(z_i\delta, \sigma^2)$-distribution.

From the model in equation (4.10), technical inefficiency of production for the i-th firm is defined by

$$TE_i = \exp(-\xi_i) = \exp(-z_i\delta - \tau_i) \tag{4.20}$$

In this inefficiency frontier function (4.10) and (4.19), the τ-random variables are not identically distributed. They could be negative if $-z_i\delta > 0$, i.e., $\tau_i \geq -z_i\delta$, but are independent truncations of the normal distributions with zero mean and variance σ^2.

The explanatory variables in the inefficiency model (4.19) may include some input variables in the stochastic frontier, provided the inefficiency effects are stochastic. If the first z-variable has value one and the coefficients of all other z-variables are zero, the model represent that specified by Stevenson (1980) and Battese and Coelli (1988, 1992). If all elements of the δ-vector are equal to zero, then the technical inefficiency effects are not related to the z-variables. We, therefore, obtain Aigner, Lovell, and Schmidt's (1977) model. If we include any interaction between firm-specific variables and input variables as z-variables, we obtain the non-neutral stochastic frontier model proposed by Huang and Liu (1994).

We propose a method of maximum likelihood to, simultaneously, estimate the parameters of the stochastic frontier and the model for the technical inefficiency effects. The technical efficiencies are predicted using the conditional expectation of $\exp[-\xi_i]$ given the composed error term of the stochastic frontier. The program FRONTIER 4.1 (Coelli 1992, 1994) and STATA 8 produces predictions for these technical efficiencies for all firms during 1997.

4.4.2: Sources of Technical Inefficiencies

Table 4.2 presents variables employed as determinants of technical inefficiency. In this section, we relate the efficiency estimates to various aspects of the banks, their market conditions and other factors that are potential correlates to efficiency. That is, we examined factors that may explain some of the differences in measured efficiencies. The model for technical efficiencies is given by:

$$\xi_i = \delta_0 + \sum_{i=1}^{20} z_i \delta + \tau_i \tag{4.21}$$

Bank size variables (Small, Medium, Large, and Mega) are measured with dummy variables to allow for nonmonotonicity and nonlinearities in the relationship between bank size and efficiency. The Herfindahl index (HERF) and the share of local market deposits (SHARE) are two measures of market structure. These variables help check whether measured inefficiency might mistakenly reflect differences in market power rather than true inefficiency. That is, an indirect test of exogeneity of prices. The HERF also measures the degree of local deposit market concentration and proxies for the bank's market power. Market power is hypothesized to be positively related to alternative profit efficiency but negatively related to inefficiency. Banks in less competitive markets can charge higher prices for their services but might feel less pressure to keep costs down.

Table 4.2
Variables Employed as Determinants of Technical Inefficiency

Variables	Description	Mean	Std. dev	Expected Sign
Bank Size Variables				
Small	Dummy equals one if bank has GTA below $100 million	58,696	25,191	Positive
Medium	Dummy equals one if bank has GTA of $100 million to $1 billion	443,457	293,936	Positive
Large	Dummy equal one if bank has GTA of $1 billion to $10 billion	4,132,751	2,388,090	Positive
Mega	Dummy equals one if bank has GTA over $10 billion	125,419,926	87,434,046	Negative
Bank Characteristics				
HHI_{REV}	Revenue HHI Index	0.67217	0.09091	Positive
HHI_{NON}	Noninterest Income HHI Index	0.76249	0.19442	Positive
HHI_{LOAN}	Loan Portfolio HHI index	0.41187	0.13514	Positive
ROA	Return on Asset	0.0112	0.0055	Negative
HERF	Herfindahl-Hirschman Index	0.3780	2.9101	Negative
SHARE	Local Market Concentration	0.5293	3.1180	Negative
RISK	Altman Z-scores: Measure of Bank Insolvency Risk	2.9616	0.2676	Negative

- GTA is the Gross Total Assets. The GTA equals to total assets plus loan and lease loss reserves and allocated risk reserve (reserve for certain foreign loans). It does not depend on the performance status of the assets, and is therefore a superior measure of bank size to total assets.
- HERF is the weighted-average Herfindahl index of local deposit market concentration across the bank's markets, where each weight is the bank's deposit share in the market. Let d^{jk} = bank j's deposits in market k, then HERF for bank j is: $HERF^{j} = \sum_{market=k} SHARE^{jk} \times MKTHERF^{k}$

  where $SHARE^{jk} = \dfrac{d^{jk}}{\sum_{bank=i} d^{ik}}$ and $MKTHERF^{k} = \sum_{bank\ i} \left(SHARE^{ik}\right)^{2}$
- HERF is calculated at the state levels.

Other bank characteristic variables include measures of bank diversification[10]. There are many ways banks can diversify their operations. The broadest indicator of bank diversification that spans all activities undertaken by a bank is the diversification

[10] These are bank-specific variables, and Stiroh (2003) has applied them in bank studies.

between the traditional banking business of lending (and holding other liquid assets) and nontraditional ones (like fiduciary operations, trading activities, and other noninterest income generating business). As Stiroh (2003:08) indicated, "a traditional bank could diversify by expanding into new types of loans or a nontraditional bank could diversify by expanding the number and type of noninterest income business lines. If these activities are negatively or only weakly correlated, all three diversification strategies could yield standard portfolio benefits that improve the risk/return performance as the bank expands operations."

We constructed three Hirschman-Herfindahl Index (HHI) measures of bank diversification.[11] First, we constructed the revenue HHI for each bank from revenue flows associated with each of the bank's major activities:

$$HHI_{REV} = \left(\frac{NON}{NETOP}\right)^2 + \left(\frac{NET}{NETOP}\right)^2 \qquad (4.22)$$
$$\text{NETOP} = \text{NON} + \text{NET}$$

where NON is the noninterest income, NET is the net interest income, and NETOP is the net operating revenue. As the HHI rises, the bank becomes more concentrated and less diversified thereby becoming more technically inefficient and worsening the alternative profit.

Second, we constructed the noninterest income HHI, which measures diversification within nontraditional activities for each bank. Thus,

$$HHI_{non} = \left(\frac{FID}{NON}\right)^2 + \left(\frac{SRV}{NON}\right)^2 + \left(\frac{TRADE}{NON}\right)^2 + \left(\frac{OTHER}{NON}\right)^2 \qquad (4.23)$$
$$\text{NON} = FID + SRV + TRADE + OTHER$$

[11] One shortcoming of the HHI is that it does not provide information about the type of activity. For example, a bank with all its revenue from noninterest sources would have the same HHI_{REV} as a bank with all its revenue from net interest sources.

where FID is fiduciary income; SRV is a service charge; TRADE is trading revenue; and OTHER is fees and other noninterest income. A higher HHI_{NON} value indicates increased concentration, less diversification, more technical inefficiency and worsened profits.

Lastly, we constructed the loan portfolio HHI, which measures diversification within lending activities for each bank. Thus,

$$HHI_{LOAN} = \left(\frac{RLOAN}{LOAN}\right)^2 + \left(\frac{BLOAN}{LOAN}\right)^2 + \left(\frac{CLOAN}{LOAN}\right)^2 + \left(\frac{OLOAN}{LOAN}\right)^2 \qquad (4.24)$$
$$\mathrm{LOAN} = \mathrm{RLOAN} + \mathrm{BLOAN} + \mathrm{CLOAN} + \mathrm{OLOAN}$$

where RLOAN is the real estate loans, BLOAN is the commercial and industrial loans, CLOAN is the consumer/individual loans, OLOAN is other loans, and LOAN is total loans. Therefore, a higher value of HHI_{LOAN} indicates increased concentration, less diversification, less technical efficiency, and worsened profits.

We also employ the Altman Z-score incorporated as an input in the profit function above. A higher Z-Score indicates improved risk-adjusted performance, less technical inefficiency and improved alternative profit.

4.4.3: Testing the Diversification Hypothesis

Table 4.3 below provides variables employed in testing the diversification hypothesis. That is, as the banking organizations increase in size (through mergers), do their risk-return tradeoffs improves, presumably because of better diversification of portfolio risk? Under this hypothesis, improved diversification of the loan portfolio owing to broader coverage of geographic areas, industries, loan types, or maturity structures might allow consolidated banks to shift their output mixes from lower-yielding securities toward higher-yielding loans without raising their costs of funding.

Table 4.3
Diversification Hypothesis Ratios - means, standard deviations and expected sign for 1997

Variables	Descriptions	Mean	Std. Dev	Expected signs♣
Loans/GTA	Total Loans Divided by Gross Total Assets	0.60495	0.12921	Positive
GTA/Equity	Gross Total Assets Divided by Equity	11.6472	2.35704	Positive
Loans/Equity	Total Loans Divided by Equity	7.09835	2.13003	Positive
CLoans/GTA	Consumer Loans Divided by Gross Total Assets	0.08863	0.06784	Positive
BLoans/GTA	Business Loans Divided by Gross Total Assets	0.12491	0.06486	Positive
Cloans/Equity	Consumer Loans Divided by Equity	1.05106	0.84210	Positive
BLoans/Equity	Business Loans Divided by Equity	1.47600	0.86372	Positive
PF/GTA	Purchased Funds Divided by Gross Total Assets	0.07709	0.08814	Negative
PF/Equity	Purchased Funds Divided by Equity	0.92221	1.03325	Positive

♣These variables are expected to have opposite signs when measured with respect to technical inefficiency

A better diversified bank can absorb the rates of more loans without any increase in equity capital or any increase in the cost of uninsured bank debt because the diversification helps protect the equity and debt holders from risk. A shift from securities to loans under the diversification hypothesis would improve the standard profit efficiency rank on average by increasing revenues for a given set of output price. However, this revenue improvement would not affect cost efficiency, because cost efficiency excludes revenue entirely. The improvement would only affect alternative profit efficiency to the extent that output specified in the alternative profit function are imperfect in controlling for changes in product mix.

Banks that merge tend to increase their proportions of loans in the asset portfolio and more assets and loan per dollar of equity. That is, merging banks are able to shift towards loans without the capital market requiring more equity, consistent with a more diversified portfolio.

Merging banks also show a shift into both consumer loans and business loans and out of securities. Merging banks are also able to raise additional uninsured purchased funds, and that they buy those funds at reduced rates. It appears that mergers allow the participants to move from paying relatively high rates for purchased funds to about average rates, and to buy more of these funds, consistent with capital markets evaluating these banks as safer.

Reliance on purchased funds could also be related to efficiency, since the cost of purchased funds differs from that of core deposits over the business cycle. Banks that use more of these funds tend to have lower profit efficiencies than other banks. Banks with higher loan-to-asset (Loan/GTA) ratios tend to have higher profit efficiency. This might reflect that the bank's loan product is more highly valued than securities, or it could reflect higher market power that exists in loan markets compared to the other product markets in which the bank operate.

4.5: Data

We apply the specified models to U.S. banking mergers of 1997. The choice of the period was motivated by the enactment of the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994. The Act permits adequately capitalized and managed bank holding companies to acquire banks in any state one year after enactment. Beginning June 1, 1997, the Act allows interstate mergers between adequately capitalized and managed banks, subject to concentration limits, state laws, and Community Reinvestment ACT (CRA) evaluations. We collected data of all mergers irrespective of asset size. There were 762-recorded mergers during 1997. Because the majority of mergers reported were merely corporate reorganizations and will have less meaning in this study, we analyzed only bank mergers that consolidated under common ownership operating banks formerly independent of one another. That is, we exclude corporate reorganizations. As a result, the number of mergers cases decline to 347. We collected data from the Federal Reserve's Annual Report of Condition and Income (Call Report) during 1997.

CHAPTER 5: EMPIRICAL RESULTS

5.1: Introduction

In this chapter, we present an analysis of the main findings of the presence of inefficiency in merged banks. First, we estimated an overall stochastic frontier alternative profit function assuming a truncated normal distribution and excluding the inefficiency effects model. Then we reestimated the same model simultaneously with the inefficiency effects model. That is, we fit a conditional mean model in which the mean of the truncated-normal distribution in the profit frontier is modeled as a linear function of a set of covariates. As discussed in the previous chapters, this model affords us an opportunity to avoid an assumption that the data are identically distributed.

Second, we tested some of the structural or methodological properties of the alternative profit frontier model with inefficiency effects. We estimated the Cobb-Douglas functional form to see if it fits the data better than the flexible translog specified in this study. We also tested if the half-normal distribution is an adequate representation of the data given the specified normal-truncated normal distribution. Lastly, we evaluated the impact of the inefficiency correlates to see if they have a positive or negative impact on bank technical efficiencies.

5.2: Estimates of the Alternative Profit Frontier Model

Table 5.1 presents the maximum likelihood estimates for some of the parameters of the stochastic alternative profit function for 1997 (that is, the output table was large such that only selected variables are included in table 5.1).

Table 5.1*
Maximum-Likelihood Estimates for Some Parameters of the Stochastic Alternative Profit Function

Variable	Parameter	Coefficient	Std. Err.
DDEPOSIT	β_1	1.2801	0.0374*
RLOAN	β_2	-0.2062	0.0294*
CLOAN	β_3	0.2256	0.0150*
BLOAN	β_4	-0.3781	0.0567*
UFUNDS	β_5	-49.7530	-
UCDEP	β_6	0.0998	0.0424*
UCAP	β_7	-1.1907	0.0422*
ULAB	β_8	0.1434	0.1168*
RISK	β_9	0.4008	0.0462*
Z_1	β_{10}	-0.2625	0.0126*
Variance Parameters			
	σ^2	0.0159	0.0043
	γ	1.0000	0.0000
	σ^2_v	0.0000	0.0000
	σ^2_ξ	0.0159	0.0043
	Mu	-0.4195	0.1317
	ILGTγ	23.7003	10.9548
	$\ln\sigma^2$	-4.1445	0.2683
Log-Likelihood Function		**829.516**	
N = 347			
H_0: No Inefficiency Component:		**Z = -1.974**	**Prob <= 0.024**

*Significant at 0.05 level of significance; the significance is tested based on the Z-statistics reported by STATA. *The output file was too big, so only selected variables are included in this table and other tables below.

We estimated the results using STATA v8. Almost all variables estimated in the alternative profit function were significant at 5% level of significance. Of the specified variable outputs, the demand deposits (DDEPOSIT) and the consumer loans (CLOAN) are positively and significantly related to the alternative profit (as expected). They are significant at the 0.05 level of significance. Therefore, a bank that is issuing more consumer loans in this case should improve its alternative profit. This is consistent with the diversification hypothesis, under which merged banks tend to shift away from securities into consumer loans as they have more diversification opportunities. For a complete picture, we also need to evaluate the proportion of consumer loans in relation to total assets or equity capital (to be discussed later in the study). Also, banks that are taking more demand deposits improve their alternative profit, as demand deposits are positive and statistically significant in determining alternative profit. We normalized the parameters of the model by the fixed input quantity, financial equity capital. The specification of equity capital as a fixed input helps alleviate the possibility that the largest banks may be measured as the most profit efficient simply because their higher capital levels allow them to have the most loans.

The coefficient for real estate loans (RLOAN) is negative and significantly different from zero at 0.05 level of significance. This implies that banks that increase their real estate loans portfolio experience a decline in their alternative profit. The coefficient for business loans (commercial and industrial loans – BLOAN) is also negative and statistically different from zero at 0.05 level of significance. Therefore, an increase in the bank's business loan portfolio also leads to a decline in the alternative profit.

We can interpret these results in line with the 'value added approach' (Berger and Humphrey 1992b) used to define and measure bank outputs in this study. As explained in the previous chapter, the value added approach employs, as important outputs, categories of the bank's financial statements that have substantial value added (as judged using an external source of operating cost allocations). Those identified as major categories include produced deposits (demand, time and savings) and loans (real estate, commercial, and installment). Our results confirm that demand deposits and the consumer loans do add value on average to the firm, as they are positive and statistically significant in determining alternative profit. To the contrary, the real estate loans and the business loans do not add value on average to the banks as expected. An increase in the portfolio of the real estate loans and the business loans leads to a decrease in alternative profit.

The coefficient for the unit price of core deposits (UCDEP) is positive and statistically significant (at 0.05 level of significance) in determining alternative profit. The unit price of core deposits is not only statistically significant in determining alternative profit but an increase in these costs is associated with more activities that add value to the firm, thus leading to an increase in alternative profit.

The coefficient for the unit price of labor (ULAB) is also positive and statistically significant (at 0.05 level of significance) in determining alternative profit. This result is contrary to what is asserted by most bank analysts who measure efficiency in terms of spending on overhead (e.g. physical plant and bank personnel) relative to the amount of financial services produced. From the bank analysts' perspective, the unit price of labor should be inversely related to bank profit. That is, we should expect the banking industry to not only improve their profits but also record some impressive efficiency gains because reducing overhead is a stated goal in many bank mergers and bank holding company reorganizations.

Our results also allude to the misleading nature of the accounting based ratio analysis widely employed by bank analysts in assessing bank efficiencies and cost structures. We indicated in chapter 1 that even though the number of banks continued to decline since the 1980s due to mergers (figure 1.1), the number of branch offices continued to increase (figure 1.6). This result should, however, be interpreted with caution. That is, the unit price of labor is calculated as the ratio of salaries and benefits to the number of people employed during the period in question (1997) in the banking industry. An increase in the unit price of labor might be a result of an increase in the salaries and benefits while the number of people employed is constant or declining. Therefore, an increase in expenditure on labor might not necessarily mean that banks are hiring more employees. Humphrey (1994) reported that the number of bank locations (main offices, branches, and ATMs) per person in the U.S. tripled between 1973 and 1992. DeYoung (1997a) indicated that from the mid-1980s to the mid-1990s total employment in commercial banks fell by about 5% (about 13% per dollar of assets). However, this was offset by a 19% increase in real salaries and benefits per employee. This implies that the unit price of labor (the way we measure it) should increase.

It should be noted that an increase in the cost of labor might be associated with value added in the banking industry. That is, higher wages might lead to the production of more financial services per worker. However, higher wages might also be a result of the production of more financial services. DeYoung (1997a) also indicated that a large employee turnover might make a bank healthier if additional workers are put to work monitoring loans. Our results that the coefficient of the unit price of labor is positive and statistically significant in determining alternative profit should also indicate to the bank analysts that efficiency and cost cutting might not necessarily be one and the same thing. Therefore, we can expect the industry-wide expenditures on labor to increase at a time when the most inefficient banks are exiting the industry.

The coefficient for the unit price of physical capital (UCAP) is negative and statistically significant (at 0.05 level of significance) in determining alternative profit. That is, an increase in the unit costs associated with physical capital depletes value on average thereby leading to a decline in the alternative profit. In this case, reducing the physical overhead should add value on average into the bank in line with what most bank analysts infer. This result is also confirmed by the coefficient of the input fixed quantity (physical capital – Z_1), which is also negative and statistically significant in determining alternative profit (at 0.05 level of significance). Therefore, alternative profit should improve in conjunction with a decline in the physical overhead.

Another variable of interest is the Altman Z-score (RISK), which measures the bank's probability of bankruptcy. From the public policy perspective, the risk of failure of banks is of primary concern regarding banks' product line expansion.

This is also the reason this study incorporates this measure of risk (Z-score) as we expect the merged banks to not only increase their geographic reach, but also expand their product line. In addition, Boyd and Graham (1989) indicated that one of the views of proponents of bank holding company expansion is that increases in volatility or rates of return (as represented by the standard deviation of ROA) would be offset by increases in rates of return thereby resulting in lowered risk of failure. The measure of risk, Z-score, offers us an opportunity to, directly, test this view in this study.

The coefficient for the measure of risk (RISK) is positive and statistically significant (at 0.05 level of significance) in determining the alternative profit. We expect this measure of risk to decrease with the volatility of assets returns. That is, losses push a firm towards insolvency but these loses are cushioned by the firm's equity capital. Therefore, the Z-score accounts for both the mean level of bank profits and mean equity ratio such that a higher coefficient of RISK indicates improved risk-adjusted performance in the bank. This will in turn lead to improvement in alternative profit.

Apart from the reported coefficients, Table 5.1 also reports the estimates for the parameters $\sigma_v^2, \sigma_\xi^2, \gamma, \mu$, etc. Gamma is the estimate of $\gamma = \sigma_\xi^2 / \sigma_s^2$, and σ^2 is the estimate of $\sigma_s^2 = \sigma_v^2 + \sigma_\xi^2$. Since γ must be between 0 and 1, the optimization is parameterized in terms of the inverse logit of γ, and this estimate is reported as ILGTγ. Also, since σ_s^2 must be positive, the optimization is parameterized in terms of $\ln(\sigma_s^2)$, whose estimate is reported as $\ln\sigma^2$. Mu (μ) is the mean of the truncated-normal distribution.

From table 5.1, the generalized log-likelihood ratio test for the presence of the inefficiency term has been replaced with a test based on the third moment of the OLS residuals. That is, if $\mu = 0$ and $\sigma_\xi = 0$, then the truncated-normal model reduces to a linear regression model with normally distributed errors. However, in this case, the distribution of the test statistic under the null hypothesis is not well established as it becomes impossible to reliably evaluate the log-likelihood as σ_ξ approaches zero. Therefore, we cannot use the likelihood-ratio test in this case.

Coelli (1995) noted that the presence of an inefficiency term would cause the residuals from an OLS regression to be negatively skewed. Therefore, by identifying the negative skewness in the residuals with the presence of an inefficiency term, he defined a one-sided test for the presence of the inefficiency term. The result is presented at the bottom of table 5.1. This result affords us an opportunity to test the hypothesis that determines if we should incorporate the inefficiency component in the frontier model (i.e. is there evidence of inefficiency in the model?). Thus,

Hypothesis 1:

H_0: $\gamma = 0$ (No inefficiency component in the model)

H_A: $\gamma > 0$ (Inefficiency component is present)

From the results presented at the bottom of table 5.1, we can reject the null hypothesis at 5% level of significance. Therefore, the inefficiency component needs to be incorporated into the frontier model. We shall retest this hypothesis after estimating both the stochastic frontier and the inefficiency effects models.

After fitting the truncated-normal on the data and detected a statistically significant inefficiency term in the model, we might question whether the inefficiency term is identically distributed over all firms or whether there might be heterogeneity across firms. The stochastic frontier approach provides us with an extension to the truncated normal model by allowing the mean of the inefficiency term (ξ_i) to be modeled as a linear function of a set of covariates (which is also the main objective of this study). We, therefore, reestimated the model in the next section including variables that measures various aspects of the bank characteristics. We also retest the presence of inefficiency and other hypotheses of this study.

Finally, to confirm the presence of inefficiency or inefficiency component in the model, table 5.1 also reports the results of γ. We observe that the estimated gamma is 1.0000 and its standard error is 0.0000. This result indicates that the vast majority of the residual variation is due to the inefficiency effect (ξ_i), and that the random error, V_i, is approximately zero. We also estimated the same model using FRONTIER 4.1 and the results show that based on the LR test, the stochastic frontier is statistically different from the OLS estimation. That is, the estimated γ is significantly different from zero, suggesting that the auxiliary equation (the technical efficiency equation) plays an important role in the estimation of the profit function. This also suggests that the concentration should be on the maximum likelihood estimation.

Most previous studies on bank efficiency employed the ordinary least squares (OLS) in estimating the parameters of the frontier function. These studies obtained the inefficiency residuals that are then regressed against some firm-specific variables or potential efficiency correlates. However, results in this study indicated that the maximum likelihood estimation, not OLS, should be the focus when estimating the parameters of the frontier function.

5.3: Hypotheses Testing on the Estimates and Structure of the Model

It is important to note that the model for inefficiency effects can only be estimated if the inefficiency effects are stochastic and have a particular distributional specification. Hence, there is growing interest to test the null hypothesis that the inefficiency effects are not stochastic; the inefficiency effects are not present; and the coefficients of the variables in the model for the inefficiency effects are zero. These and other null hypotheses are of interest in this study and they are tested using the generalized likelihood-ratio statistic, λ, and the Wald test (Green 2003, Gallant 1997).[12]

Table 5.2 presents the tests for these hypotheses. Table 5.2 also presents the results for a retest of our first and main hypothesis (*that is, does the merged bank show any evidence of inefficiencies?*). We know that banks are not perfectly efficient such that some level of inefficiencies exist. We, however, present the results and procedures followed in administering the test.

---

[12] The generalized likelihood-ratio test statistic, λ, is calculated as $\lambda = -2\ln\left[\frac{L(H_0)}{L(H_1)}\right] = -2\left\{\ln\left[L(H_0)\right] - \ln\left[L(H_1)\right]\right\}$. If the null hypothesis, H_0, is true, then λ is asymptotically distributed as a Chi-square (or mixed Chi-square) random variable with parameters equal to the number of parameters assumed to be equal to zero in the null hypothesis, H_0.

Table 5.2
Tests of Hypotheses for Coefficients of the Explanatory Variables for the Technical Inefficiency Effects in the Stochastic Frontier Profit Function

Null Hypothesis	Log- Likelihood Function	Test Statistic λ	Critical Value	Decision
$H_0: \gamma = \delta_0 = \ldots = \delta_{25} = 0$	764	3.64	2.706	Reject
$H_0: \beta_{ij} = 0, i \leq j = 1{:}54$	676.61	175.54	75.35	Reject
$H_0: \delta_1 = \ldots = \delta_{20} = 0$	829	191.9	37.57	Reject
$H_0: \mu = 0$	764.38	3.65	2.706	Reject

- The critical values for the tests involving $\gamma = 0$ are obtained from Table 1 of Kodde and Palm (1986) where the degrees of freedom are q + 1, where q is the number of parameters which are specified to be zero but which are not boundary values.
- If the null hypothesis involves $\gamma = 0$, then λ has mixed Chi-square distribution because $\gamma = 0$ is a value on the boundary of the parameter space for γ (Coelli, 1995, provides more details).

Hypothesis 1: *Do the merged banks show any evidence of inefficiencies? We conduct a generalized likelihood ratio test to test the null hypothesis that these merged banks are completely efficient against the alternative hypothesis that they are not efficient.*

$H_0: \gamma = \delta_0 = \ldots = \delta_{25} = 0$ (No inefficiencies)

$H_A: \gamma > 0$ (Inefficiencies exist)

The null hypothesis specifies that the inefficiency effects are absent from the model (all banks are efficient) against the alternative hypothesis that inefficiencies are present. This null hypothesis (no technical inefficiency effects in the model) can be conducted by testing the null and alternative hypothesis, $H_0: \sigma^2 = 0$ vs. $H_A: \sigma^2 > 0$. In this case, σ^2 is the variance of the normal distribution, which is truncated at zero to obtain the distribution of ξ_i. If this variance is zero, then all the ξ_is are zero, which implies that all firms are fully efficient. The Wald statistic, which involves the ratio of the maximum likelihood estimator of σ^2 to its estimated standard error, has been used to test this null hypothesis.

Coelli et al. (1998) pointed out that another set of hypotheses, H_0: $\lambda = 0$; vs. H_A: $\lambda > 0$; or H_0: $\gamma = 0$ vs. H_A: $\gamma > 0$ can be considered depending upon the parameterization used in the estimation of the stochastic frontier model. In this study, we adopt the Battese and Corra (1977) parameterization such that the hypotheses involving γ are considered (see chapter 3). Considering the Wald test, we calculate the ratio of the estimate for γ to its estimated standard error. If H_0: $\gamma = 0$ is true, this statistic is asymptotically distributed as a standard normal random variable.

Coelli et al. (1998) also indicated that this test must be performed as a one-sided test because γ cannot take negative values. However, following footnote 12, under H_0: $\gamma = 0$, the model is equivalent to the traditional average response function without the technical inefficiency effects. Coelli (1995) pointed out that the difficulties arise in testing H_0: $\gamma = 0$ because $\gamma = 0$ lies on the boundary of the parameter space γ. Therefore if, H_0: $\gamma = 0$ is true, the generalized likelihood-ratio statistic, LR, has asymptotic distribution, which is a mixture of chi-square distributions, viz. $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ where χ_0^2 is the unit mass at zero.

Therefore, from table 5.2, the log-likelihood function (calculated from FRONTIER 4.1) for the full stochastic frontier model and the inefficiency effects model is 764.38 and the value for the OLS fit of the profit function is 762.56, which is less than that for the full frontier model. This implies that the generalized likelihood-ratio statistic for testing the absence of the technical inefficiency effects from the frontier is calculated to be:

$$LR = -2\{762.56 - 764.38\} = 3.64$$

This value is calculated by FRONTIER and reported as the "LR test of the one-sided error". This value is also significant, because it exceeds 2.706, which is the critical value obtained from Table 1 of Kodde and Palm (1986) for the degrees of freedom equal to 1. Hence, we reject the null hypothesis of no technical inefficiency. Therefore, there is evidence of inefficiencies in these merged banks, which also confirm the results from the previous section that inefficiency component should be incorporated into the frontier model. It should be noted that we are only measuring inefficiency at a point in time (1997). The next section will draw comparison of the estimated efficiency scores with those of other previous bank efficiency studies. *The rest of the hypotheses in table 5.2 test the structural properties of the model.*

The second null hypothesis (table 5.2), H_0: $\beta_{ij} = 0$, $i \leq j = 1{:}54$ state that the 2nd order coefficients of the translog frontier are simultaneously equal to zero. That is, we test whether the Cobb-Douglas functional form is an adequate representation of the data, given the specifications of the translog model. The output file for estimating the Cobb-Douglas model using FRONTIER gives the value of the log-likelihood function as 676.61. Hence the value of the generalized likelihood-ratio statistic for testing the null hypothesis, H_0: $\beta_{ij} = 0$, is calculated to be

$$LR = -2\{676.61 - 764.38\} = 175.54$$

This value is compared with the upper one percent point for the χ^2_{54} distribution, which is 75.35. Thus, we reject the null hypothesis that the Cobb-Douglas frontier is an adequate representation of the data, given the specifications of the translog function. This is not a surprising result as the translog offers more flexibility than the Cobb-Douglas functional form.

The third null hypothesis (table 5.2), H_0: $\delta_1 = \ldots = \delta_{20} = 0$ specifies that all the coefficients of the explanatory variables in the model for the technical inefficiency are equal to zero (and hence that the technical inefficiency effects have the same truncated normal distribution). That is, the inefficiency effects are not a linear function of the δ_i's. The calculated generalized likelihood ratio test is 191.9. We compare this value with the upper one percent point of the χ^2_{20} distribution. We, therefore, reject the null hypothesis that the inefficiency effects are not a linear function of the δ_is. This implies that the joint effects of these variables on inefficiencies are significant although the individual effects of one or more variables may not be statistically significant.

The fourth hypothesis (table 5.2), H_0: $\mu = 0$ specifies the distributional assumption of the model. This hypothesis tests whether the simpler half-normal model is an adequate representation of the data, given the specification of the generalized truncated normal model. That is, we test whether the technical inefficiency effects have a half-normal distribution or follow the normal-truncated normal distribution. The value of the likelihood-ratio statistic for testing this null hypothesis is 3.65. We compare this value with the upper ten percent point for the χ^2_1 distribution, which is 2.706. Thus, we reject the null hypothesis that the half-normal distribution is an adequate representation of the inefficiency effects given the normal-truncated normal distribution.

This is an important result as it shows that, contrary to other bank efficiency studies, we should be applying other distributional assumptions than the simpler normal-half normal that is an arbitrary choice. The normal-half normal distribution specified by other bank efficiency studies has a mode at zero. This implies that there is the highest probability that the inefficiency effects are in the neighborhood of zero. That is, technical efficiency could be unnecessarily high, which might not be evident in practice. The normal-truncated normal distribution (like the gamma distribution) addresses this shortfall by allowing for a wider range of distributional shapes, which include ones with non-zero modes. That is, the normal-truncated normal distribution allows the normal-half normal distribution to have a nonzero mode. Berger and Mester (1997) used the stochastic frontier approach where the inefficiencies were assumed to be half-normally distributed. The data, however, did not appear to fit that distribution very well. They indicated that the skew of the data was not consistent with the half-normal assumptions in a number of cases.

5.4: Predicted Technical Efficiencies

In this section, we present the results of the predicted technical efficiencies of the merged banks during 1997. One of the focuses of this dissertation is to measure the bank efficiency with respect to risk and return. This also provides us with an opportunity to test the second hypothesis of the dissertation.

Thus,

Hypothesis 2: *Does the degree of measured efficiency improve (or otherwise change) when we correct for variations in bank risk?*

H_0: No Improvement in Measured Efficiencies

H_A: There is Evidence of Improvement

We shall compare the predicted efficiencies in this study with other previous studies in bank efficiency. As indicated, previous studies found that banks on average are very inefficient with respect to profit. These results should be surprising as studies that incorporated the revenue or output effects of banks (mergers) showed that there were some improvements in profit efficiency. Therefore, does incorporating a measure of risk in analyzing bank mergers lead to an improvement in measured efficiencies?

Table 5.3 presents the summary statistics of the predicted technical efficiencies of bank mergers during 1997, together with results from Berger and Humphrey's (1997) survey of bank efficiency studies. In this study, all banks have predicted technical efficiencies greater than 0.90 (and some very close to 1.0). From the combined model (the frontier model and the inefficiency effects model) the mean efficiency was 0.9809 (median = 0.9818). This implies that, the average inefficiency was $(1 - 0.98)/0.98 = 2\%$. That is, if the average firm were producing on the frontier of the current location, then only 98% of the resources currently being used would be necessary to produce the same output. The last column indicates the summary statistics of the predicted technical efficiencies as summarized by Berger and Humphrey (1997).

Table 5.3
Summary Statistics of Predicted Technical Efficiencies

Statistics	Combined Model*	Cobb-Douglas	U.S. EFF.*
Mean	0.9809	0.9633	0.72 (0.84)
Median	0.9818	0.9656	0.74 (0.85)
Std. Deviation	0.0084	0.2313	0.17(0.06)
Skewness	-2.1235	-2.2893	
Minimum	0.9358	0.8288	0.31(0.61)
Maximum	0.9946	0.9930	0.97(0.95)
Mode	0.9816	0.9656	

*Berger and Humphrey 1997; Parametric estimates in parenthesis.
*Average Inefficiency ~ (1 – 0.98)/0.98 = 2%. That is, If the average firm were producing on the frontier instead of the current location, then only 98% of the resources currently being used would be necessary to produce the same output.

Berger and Humphrey (1997) surveyed about 130 studies that applied frontier efficiency analysis to financial institutions in 21 countries. The studies surveyed differ in terms of the techniques applied in their analysis (parametric or nonparametric approach), functional forms and whether they applied a profit or cost function. Berger and Humphrey (1997) noted that efficiency estimates from nonparametric techniques were slightly lower and seem to have greater dispersion than those from the parametric techniques (as indicated in the last column of table 5.3). The authors' analysis further indicated that from the parametric studies, those that applied the stochastic frontier approach had mean efficiencies that ranged from 0.81 to 0.95.

This dissertation applies the parametric technique in analyzing the profit efficiency of bank mergers. Therefore, we expect the predicted technical efficiencies to be higher as evident from table 5.3 above. However, it should be noted that most of the studies surveyed by Berger and Humphrey (1997) estimated cost efficiency of banks. This dissertation presents technical efficiencies of bank mergers estimated from the profit function. The profit efficiency is measured in terms of best-practice profits, which are typically much smaller than costs, inputs, or output levels used in conventional studies. Therefore, we cannot easily compare the predicted technical efficiencies of this study with those surveyed in Berger and Humphrey (1997). To give a better indication of the distribution of the individual efficiencies, we plot the frequency distributions of these efficiencies in figure 5.1 below.

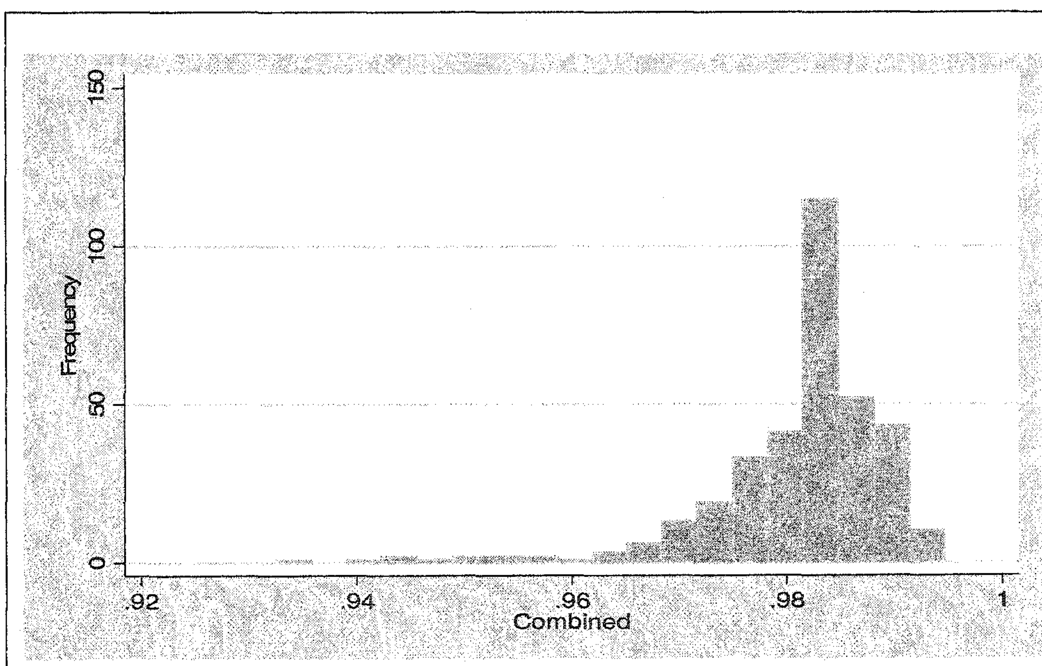


Figure 5.1: Frequency Distribution of Predicted Efficiencies

The plot from the combined model has a thin tail that gradually rises to a maximum in the 0.97 to 0.98 interval and then drops sharply in the 0.99 to 1.0 interval. The fact that the mode of the distribution is not in the final interval offers support for the use of more a general distribution (rather than the normal-half normal distribution applied in most bank studies) like the normal-truncated normal distribution applied in this dissertation.

5.5: Estimates of the Parameters of the Frontier and the Inefficiency Effects Model

In this section, we assess the estimates of both the stochastic frontier profit function that is estimated simultaneously with the inefficiency effects model. We also evaluate whether the coefficients of the alternative profit function maintain their explanatory power in this model as in the model presented in section 5.1.

Table 5.4 below presents the Maximum-Likelihood estimates of the stochastic frontier profit model estimated simultaneously with the inefficiency effects model. Model 1 is the general or unrestricted model and model 2 (restricted model) excludes profit function variables that are not significant in model 1. Results in table 5.4 differ with those in table 5.1 because table 5.1 presents maximum-likelihood estimates of the stochastic frontier model without the inefficiency effects model.

The coefficient of the consumer loans (CLOAN) is positive as expected. That is, the merged entities that increase their consumer loans portfolio, probably due to more diversification opportunities, are able to increase their alternative profit. However, the coefficient is not statistically different from zero. As indicated, the output quantities are normalized by the financial equity capital to avoid scale bias towards large banks.

Table 5.4
Maximum-Likelihood Estimates for some Parameters of the Stochastic Frontier Alternative Profit Function with Inefficiency Effects

Stochastic Frontier Profit Model			
Variable	**Parameter**	**Model 1 Coefficient**	**Model 2 Coefficient**
DDEPOSIT	β_1	0.8900 (-)	-
RLOAN	β_2	0.4053 (-)	-
CLOAN	β_3	0.0244 (0.1043)	-
BLOAN	β_4	-0.6095 (0.1707)*	-0.0597 (0.0270)**
UCDEP	β_{16}	-0.5290 (0.0655)*	-0.0322 (0.0257)
R	β_{45}	-0.3673 (-)	-
Z_1	β_{55}	-0.9249 (-)	-0.9250 (-)
Variance Parameters			
	σ_s^2	0.0027 (0.0008)	0.0003 (0.0000)
	γ	0.9999 (0.0000)	0.4780 (0.0825)
	$\mathrm{Ln}\sigma^2$	-5.9411 (0.2965)	-8.1619 (0.0760)
	ILGTγ	13.7554 (1.2360)	-0.0880 (0.3305)
Log-Likelihood Function		925.48202	1032.3477

*Significant at 0.05 level of significance; **Significant at 0.10 level of significance. Standard errors are in parenthesis.

The Coefficient of the business loans (BLOAN), which was negative in the model presented in section 5.1, is still negative and statistically significant in determining alternative profit in both model 1 and 2 at 0.05 and 0.10 level of significance, respectively. Thus, merged banks that increase their business loans portfolio experience a decline in their alternative profit. This is also not consistent with the diversification hypothesis, where we expect banks to shift out of securities to business and consumer loans due to more diversification opportunities.

The result for the coefficient of the unit price of core deposits (UCDEP) is negative and statistically significant (at 0.05 level of significance) in determining alternative profit in model 1 but negative and not significant in model 2. Therefore, an increase in the unit costs related to core deposits activities of the bank lead to a decline in alternative profit. The coefficients of the unit price of labor and the unit price of physical capital were positive and negative, respectively. However, the standard errors were not reported, so we could not determine their significance.

5.6: Empirical Investigation of the Potential Efficiency Correlates

In this section, we relate efficiency estimates to variables that define various aspects of the banks and their markets. These factors are at least partially exogenous and may explain some of the differences in measured efficiencies. In this section, we also test the last two hypotheses of the dissertation:

Hypothesis 3: *Diversification Hypothesis – As the banking organization increase in size (through merger), does their risk-return tradeoff improve, because of better diversification of portfolio risk?*

H_0: Diversification Hypothesis does not hold

H_A: Hypothesis holds

According to this hypothesis, banks that increase in size through mergers should experience an improved risk-return tradeoff that would not only improve their alternative profit but also reduce technical inefficiencies. We, therefore, look at several diversification hypothesis ratios to determine if this hypothesis holds. We evaluated these diversification hypothesis ratios in relation to technical inefficiencies. We expect a bank that takes advantage of all the diversification opportunities to experience less technical inefficiency. We found that the coefficient HHI_{NON} (an index that measure a bank's ability to diversify within nontraditional banking activities) was positive and statistically significant in determining banks' technical inefficiencies as expected. That is, a higher value of HHI_{NON} indicates an increase in concentration and less diversification. Therefore, a bank that increased its HHI_{NON} becomes more technically inefficient. Contrary to the diversification hypothesis, the coefficient HHI_{LOAN} (an index that measures a bank's ability to diversify within lending activities), and the coefficient HHI_{REV} (an index that measures the bank's ability to diversify within each of the bank's major activities) were negatively related to technical inefficiency. That is, banks that increased their HHI_{LOAN} and HHI_{REV} experienced increased efficiency. However, these coefficients were not statistically different from zero.

We also found that the coefficient of the ratio of gross total assets to equity (GTAEQUITY), the coefficient of the ratio of purchased funds to equity (FUNDSEQUITY), and the coefficient of the ratio of business loans to gross total assets (BLOANGTA) were negatively related to technical inefficiency. The coefficient of the ratio of purchased funds to gross total assets (FUNDSGTA) was positively related to technical inefficiency. All these coefficients were statistically different from zero, and consistent with the diversification hypothesis.

Contrary to the diversification hypothesis, we found that the results for the coefficients of the ratio of total loans to gross total assets (LOANGTA), the ratio of total loans to equity (LOANEQUITY), and the ratio of consumer loans to gross total assets were not consistent with the diversification hypothesis. However, these coefficients were not statistically different from zero. Overall, we rejected the null hypothesis that the diversification hypothesis does not hold in favor of the alternative that the hypothesis holds.

Hypothesis 4: *Do these inefficiency effects variables maintain their explanatory power when measuring efficiency with respect to both risk and return?*

H_0: Variables do not maintain their explanatory power

H_A: Variables maintain their explanatory power

The test for this hypothesis presented mixed results. Some variables significantly maintain their explanatory power while others did not. Overall, we rejected the null hypothesis that inefficiency effects variables do not maintain their explanatory power in this study. From section 5.1, the coefficient for the measure of risk (Risk – Altman Z-score) was positive and statistically significant in determining alternative profit. That is, an improvement in the bank's risk-adjusted performance, which is associated with a higher value of the Z-score, lead to an increase in alternative profit. The variable RISK is the only variable in the study that is included as an explanatory variable in both the alternative profit function model and the inefficiency effects model. Table 5.5 presents the maximum likelihood estimates of the parameters of the inefficiency effects model.

Table 5.5
Maximum-Likelihood Estimates of the Parameters of the Inefficiency Effects Model

Variable	Parameter	Model 1 Coefficient	Model 2 Coefficient
Loan/GTA	δ_1	0.1504 (0.1788)	-0.0415 (0.0675)
GTA/Equity	δ_2	-0.0047 (0.0152)	-0.0112 (0.0040)*
Loan/Equity	δ_3	-0.0046 (0.0206)	0.0061 (0.0054)
Cloan/GTA	δ_4	-0.0067 (0.7625)	0.1737 (0.1362)
Bloan/GTA	δ_5	-0.1703 (0.8608)	-0.2903 (0.1736)**
Cloan/Equity	δ_6	-0.0310 (0.0522)	-0.0113 (0.0125)
Bloan/Equity	δ_7	-0.0201 (0.0771)	0.0101 (0.0135)
Funds/GTA	δ_8	0.0667 (0.8790)	0.2960 (0.1624)**
Funds/Equity	δ_9	-0.0580 (0.0831)	-0.0242 (0.0142)**
Medium	δ_{10}	0.0098 (0.0187)	0.0062 (0.0052)
Large	δ_{11}	0.0558 (0.0223)*	0.0075 (0.0067)
Mega	δ_{12}	-0.4394 (-)	0.0035 (0.0095)
$HHI_{R\ EV}$	δ_{13}	0.0102 (0.0677)	0.0133 (0.0227)
HHI_{NON}	δ_{14}	0.0965 (0.0373)*	0.0092 (0;0085)
HHI_{LOAN}	δ_{15}	-0.0896 (0.1076)	-0.0157 (0.0236)
Risk	δ_{16}	-0.1041 (0.0476)**	-0.1320 (0.0177)*
ROA	δ_{17}	-0.6469 (3.7644)	-9.7262 (0.4161)*
HERF	δ_{18}	-0.0411 (0.0682)	-0.0031 (0.0072)
Share	δ_{19}	0.0084 (0.0391)	0.0025 (0.0859)
Variance Parameters			
	σ^2	0.0027 (0.0008)	0.0003 (0.0000)
	γ	0.9999 (0.0000)	0.4780 (0.0825)
	σ_v^2	0.0000 (0.0000)	0.0001 (0.0000)
	σ_ξ^2	0.0026 (0.0008)	0.0001 (0.0000)
Log-Likelihood Function		925.48202	1032.3477

*Significant at 0.05 level of significance, **significant at 0.10 level of significance. Variable Small was dropped in all models due to collinearity. Standard errors are in parenthesis.

Table 5.6
Statistics for Tests of Hypotheses Involving some Coefficients of the Inefficiency Effects Model.

		Critical Value			
Null Hypothesis	χ^2	$\chi^2_{0.95}$	$\chi^2_{0.90}$	$\chi^2_{0.99}$	**Decision**
H_0: $\delta_4 = 0$	1.89	3.84	2.71	6.63	Do not Reject
H_0: $\delta_5 = 0$	3.04	3.84	2.71	6.63	Reject
H_0: $\delta_8 = 0$	4.47	3.84	2.71	6.63	Reject
H_0: $\delta_9 = 0$	3.47	3.84	2.71	6.63	Reject
H_0: $\delta_{13} = 0$	0.79	3.84	2.71	6.63	Do not Reject
H_0: $\delta_{15} = 0$	0.57	3.84	2.71	6.63	Do not Reject
H_0: $\delta_{16} = 0$	65.25	3.84	2.71	6.63	Reject
H_0: $\delta_{17} = 0$	726.97	3.84	2.71	6.63	Reject

From table 5.5, the result for the coefficient of risk is robust in both models. That is, the coefficient of risk (Altman Z-score that measures the probability of bankruptcy) is negative and statistically different from zero at 0.10 level of significance in model 1, and 0.05 level of significance in model 2. This implies that as the Z-score increases, technical inefficiency declines. That is, a higher Z-score indicates an improved risk-adjusted performance of the bank, which also implies improved profit efficiency. This can also be alluded to the fact that the merged banks have a better diversification of portfolio risk, which is also consistent with the diversification hypothesis. Therefore, as we correct for variations in bank risk, the average degree of measured efficiency in the merged banks should improve. The individual test of significance of the coefficient RISK was also highly significant at the upper one percent level of the χ_1^2 distribution (table 5.6). Therefore, we rejected the hypothesis that the coefficient RISK is zero or does not have an impact in measured technical inefficiency (H_0: $\delta_{16} = 0$ vs. H_A: $\delta_{16} \neq 0$)

The result of the coefficient RISK is important in this study as we directly incorporate a measure of bank risk in both the frontier model and the inefficiency effects model. Previous studies did not incorporate risk within the frontier measure as they only incorporated risk as a potential efficiency correlate in the second-stage regression.

We also evaluated the effects of the coefficient of return on assets (ROA) to technical inefficiency. This coefficient is negative in both model and statistically significant in model 2 (at 0.05 level of significance). This implies that banks with high return on assets are more technically efficient or experience less technical inefficiency. That is, merged banks are able to increase diversification in various aspects of their operations owing to such issues as broader coverage of geographic areas, etc. These banks might be able to shift their output mix from lower- to higher-yielding products without raising costs. Therefore, we expect an increase in the return on assets to lead to a decrease in technical inefficiency, thereby leading to improved alternative profit. An individual test for the coefficient of return on assets (ROA) was also highly significant compared with the upper one percent of the χ_1^2 distribution. That is, we rejected that hypothesis that ROA does not statistically influence technical inefficiency (H_0: $\delta_{17} = 0$ vs. H_A: $\delta_{17} \neq 0$). These results indicate that the diversification hypothesis holds. That is, we expect the merged banks to improve their risk-adjusted performance and also yield more return on assets due to greater diversification opportunities that a merger present.

From model 1 (table 5.5), the coefficient for the dummy variable Large is positive and statistically significant at 0.05 level of significance. This implies that large banks (those with \$1 billion to \$10 billion in assets) are more technically inefficient. This is contrary to Berger, Hancock, and Humphrey (1993) who found that inefficiencies are negatively related to bank size. However, Kaparakis et al. (1994) found that average inefficiencies rise with bank size except for banks between \$5 billion and \$10 billion in assets. In model 2 (table 5.5), the dummy variable for large bank is still positive but statistically insignificant.

The coefficient for dummy variable Mega (banks with assets over \$10 billion) is negative in the general model as expected. That is, mega banks are more technically efficient. However, this coefficient is positive in model 2. It is statistically not different from zero in both cases. The coefficient for dummy variable Medium (banks with assets between \$100 million and \$1 billion is positive in both models, which would imply that as banks increase in size to this category, they also become more technically inefficient. This result would be in line with Kaparakis et al. (1994). However, these coefficients are not statistically different from zero in both models.

The coefficient for the Herfindahl-Hirschman Index (HERF[13]), which measures market concentration, is negative in both models. That is, the more concentrated the market is, the more efficient the banks are. Because HERF is also a proxy for bank's market power, this result also implies that banks that increase their

---

[13] HERF is calculated at state levels.

market power experience less technical inefficiency or become more technically efficient. This might explain the assertion in other studies that market power is positively related to alternative profit. That is, banks in less competitive markets can charge higher prices for their services but might feel less pressure to keep their costs down. However, this coefficient is not significant in both models. We cannot conclusively infer that the higher predicted technical efficiencies found in this study are evidence of higher market power of banks or higher market concentration.

The coefficient, SHARE, which measures a bank's share of local market deposit, (and is a robustness check for HERF) is positive in both models. That is, as banks increase their share of local market deposits they also become more technically inefficient. This is in contrast with the results of the Herfindahl-Hirschman Index that show that the more concentrated the market becomes, the more technically efficient the bank becomes. However, the coefficient SHARE is also not statistically different from zero in both models.

Under the diversification hypothesis, we also look at variables that determine if the merged bank has become more or less diversified. If a bank shows sign of increased diversification in their different portfolios, we should expect an improved alternative profit and less technical inefficiency. From model 1, the coefficient for HHI_{NON}, an index that measures a bank's ability to diversify within nontraditional activities, is positive and statistically different from zero at 0.05 level of significance. That is, a higher HHI_{NON} value indicates increased concentration and less diversification. Therefore, we expect the HHI_{NON} to be positively related to technical inefficiency. That is, merged banks that experience a higher value of HHI_{NON} become more technically inefficient as expected. The coefficient for HHI_{LOAN}, an index that measures a bank's ability to diversify within the lending activities, is negative in both models. A higher HHI_{LOAN} value also implies increased concentration and less diversification. In this case, banks that increase their HHI_{LOAN} value become less technically inefficient, which is contrary to expectations as these banks are also becoming more concentrated and less diversified. However, the coefficient HHI_{LOAN} is not significant in both models. An individual test of the hypothesis that HHI_{LOAN} is zero (H_0: $\delta_{15} = 0$ vs. H_A: $\delta_{15} \neq 0$) could not be rejected (table 5.6).

The coefficient HHI_{REV}, an index that measures a bank's ability to diversify within major banking activities is positive in both models as expected. That is, as the HHI_{REV} increase, which leads to more concentration and less diversification, banks are also becoming more technically inefficient. However, this coefficient is not statistically different from zero and we could not reject an individual test that HHI_{REV} is not statistically different from zero (H_0: $\delta_{13} = 0$ vs. H_A: $\delta_{13} \neq 0$).

From model 2, the coefficient for the ratio of gross total assets to equity capital (GTAEQUITY) is negative and statistically different from zero at 0.05 level of significance. That is, banks that have more assets per dollar of equity are more technically efficient. This is in line with expectations, and consistent with the diversification hypothesis. That is, we expect a merged bank to increase its proportion of assets per dollar of equity due to better diversification opportunities, thereby improving performance and reducing inefficiency.

The coefficient of the ratio of total loans to gross total assets (LOANGTA) is also negative in model 2. That is, banks that increase their proportion of loans in the asset portfolio are more technically efficient. This is also consistent with the implications of the diversification hypothesis. That is, banks' loan product is more valued than securities. Therefore, a merged entity should be able to increase their proportions of loans to total assets, and their profit efficiency should also improve. This result might also reflect high market power that these banks command in the loan market compare with other markets. However, this coefficient is not statistically different from zero.

The coefficient of the ratio of total loans to equity (LOANEQUITY) is positive. That is, a bank that increases its proportion of loans to total equity due to better diversification opportunities is more technically inefficient. That is, the more loans per dollar of equity the bank has, the less efficient it is. This is contrary to expectation where, consistent with the diversification hypothesis, we expect this bank to be more technically efficient. However, this coefficient is not statistically significant.

The coefficient of the ratio of consumer loans to gross total asset (CLOANGTA) is also positive in model 2. That is, banks that increase their proportion of consumer loans to total assets are more technically inefficient. According to the diversification hypothesis, we expect merged bank to be able to shift into consumer loans and away from securities due to access to larger market and better diversification opportunities. Therefore, this result is in contrast with expectations as we expect a bank that increases their proportion of consumer loans to gross total asset to be more technically efficient (i.e. negative coefficient of CLOANGTA). However, the coefficient is not statistically significant. From the results in section 5.1, we found that the coefficient for consumer loans (CLOAN) was positive and statistically significant in determining alternative profit. Therefore, we expect the coefficient of CLOANGTA to be negative and statistically significant in relation to technical inefficiency. In addition, an individual test of the hypothesis that CLOANGTA is not statistically different from zero could not be rejected (H_0: $\delta_4 = 0$ vs. H_A: $\delta_4 \neq 0$)

The coefficient of the ratio of business loans to gross total assets (BLOANGTA) is negative in both models but statistically different from zero at 0.10 level of significance in model 2. That is, banks that increase their proportion of business loans to gross total assets are more technically efficient. This is consistent with the diversification hypothesis that merging banks increases the proportion of loans in their assets portfolios thereby leading to improved efficiencies. Therefore, as the merged bank increases its proportion of business loans to the asset portfolio, it also becomes more technically efficient.

The coefficient of the ratio of purchased funds to gross total assets (FUNDSGTA) is positive in both models but statistically different from zero at 0.10 level of significance in model 2. That is, banks that raise the additional uninsured purchased funds in relation to their gross total assets are more technically inefficient. This confirms the assertion that banks that use more purchased funds tend to have lower profit efficiencies than other banks. Even though merger participants might be able to buy more purchased funds at a reduced rate, an increase in the proportion of these funds, in relation to gross total assets, lead to increased technical inefficiencies. We also rejected an individual hypothesis test that FUNDSGTA is zero or does not have an impact on technical inefficiency (H_0: $\delta_8 = 0$ vs. H_A: $\delta_8 \neq 0$).

However, the coefficient of the ratio of purchased funds to equity (FUNDSEQUITY) is negative in both models but only statistically different from zero in model 2 at 0.10 level of significance. That is, banks that increase their proportions of purchased funds to total equity capital are more technically efficient. In this case, it appears that mergers allow the participants to move from paying relatively high rates for purchased funds to about average rates, which lead to a decline in technical inefficiency. An individual hypothesis test that FUNDSEQUITY is zero was also rejected (H_0: $\delta_9 = 0$ vs. H_A: $\delta_9 \neq 0$) at 0.10 level of significance.

5.7: Conclusion

In this chapter, we estimated the stochastic alternative profit frontier. In line with the "value added approach" (Berger and Humphrey 1992b) employed to define and measure bank outputs and inputs, we found that demand deposits and consumer loans add value into the firm, thereby leading to improvement in alternative profit. To the contrary, a bank that increased its business loans and real estate loans portfolio experienced a decline in alternative profit. A bank that increased activities related to core deposits and labor experienced increased alternative profit. That is, the unit price of core deposits and unit price of labor were positive and statistically significant in determining alternative profit. Of significance, this study incorporated a measure of risk (Altman Z-score that measures the probability of bankruptcy) into the profit frontier and the inefficiency effects model. From the profit frontier estimation, we found that the coefficient of the measure of risk (RISK) was positive and statistically significant in determining alternative profit. That is, a bank that increased its risk-adjusted profit experienced an improvement in alternative profit.

In testing if the inefficiency component is present or should be incorporated into the profit frontier, we found that the auxiliary equation (inefficiency effects model) is important in estimating the profit function. We also found that the Cobb-Douglas form is not an adequate representation of the data given the flexible translog form. In addition, the simpler normal-half normal distribution was not an adequate representation of the data given the generalized normal-truncated normal distribution. This confirmed Berger and Mester's (1997) results that the normal-half normal distribution did not fit well the banking data when estimating the stochastic frontier function.

We reestimated the profit frontier model simultaneously with the inefficiency effects model and tested the null hypothesis that there are no inefficiencies in the merged banks. However, we rejected the null hypothesis in favor of the alternative that there is evidence of inefficiencies. The predicted efficiencies were higher compared to those of the previous studies on bank efficiency, as expected. The mean predicted efficiency was 0.9809 (median = 0.9818). Consistent with the diversification hypothesis, we found that merged banks that increased the ratios of gross total assets to equity, business loans to gross total assets, and purchased funds to equity become more technically efficient. Also, banks that increased the ratio of purchased funds to gross total assets became more technically inefficient. Large banks (those with assets between $1 billion and $10 billion) were technically inefficient.

Banks that increases their HHI_{NON} (and index that measure a bank's ability to diversify within nontraditional activities) become more technically inefficient. That is, an increased value of HHI_{NON} leads to more concentration and less diversification of the bank. Banks that increase their return on assets also become technically efficient. That is also consistent with the diversification hypothesis where greater diversification opportunities enabled a bank to raise it ROA. We also incorporated a measure of risk into the inefficiency effects model. We found that the coefficient RISK was negative and statistically different from zero in determining technical inefficiency. That is, banks that increased their risk-adjusted performances become more technically efficient. These results were consistent with the diversification hypothesis and, overall, the variables maintained their explanatory powers.

CHAPTER 6: SUMMARY AND CONCLUSIONS

6.1: Conclusion and Policy Recommendations

Despite all the research effort in examining the efficiencies of financial institutions, there are still varying opinions about the differences in measured efficiencies. In addition, research has concentrated mostly on measuring cost efficiencies of banks rather than profit efficiencies. There is still lack of research in examining the effects of mergers on bank efficiencies from a profit function perspective. Efficiency studies have also failed to explicitly present a model for inefficiency effects, especially in the financial industry. Previous studies on bank efficiencies also concentrated on evaluating performance during the 1980s rather than the 1990s. This dissertation attempts to address these issues.

The primary objective of this dissertation was to evaluate the impact of mergers and acquisitions on the U.S. commercial bank efficiency during 1997. We employ a stochastic frontier approach based on the flexible translog profit function. We examine only one efficiency concept, the alternative profit efficiency. Like other efficiency concepts, the alternative profit efficiency helps us examine how well a bank performs relative to different economic optimization programs thereby giving insights about each bank's efficiency. Analysis of this type of X-efficiency is also a better way to examine if bank mergers are in the public interest. That is, we can distinguish increases in social value added through improvements in efficiency from increases in private value derived from better opportunities of exercising market power in setting prices.

Another objective of the dissertation was to, explicitly, present a model of inefficiency effects. Unlike previous studies on bank efficiency that estimated a frontier function and then regress the estimated inefficiency effects against a set of explanatory variables, this study simultaneously estimates the profit frontier and the inefficiency effects model. The study also assumes that the inefficiency effects term, ξ_i, follows a normal-truncated normal distribution rather than the half-normal distribution adopted by other bank efficiency studies. The normal-truncated normal distribution generalizes the one parameter half-normal distribution, by allowing the normal distribution, which is truncated below at zero, to have a nonzero mode.

The study then tested the distributional assumption of the inefficiency effects term. That is, we conducted a test that the simpler half-normal distribution is an adequate representation of the data, against the alternative hypothesis that the normal-truncated normal distribution was appropriate. The results showed that the normal-truncated normal distribution was an adequate representation of the data than the half-normal distribution. This result confirms Berger and Mester's (1997) results. While estimating a stochastic frontier function, they found that the normal-half normal distribution does not fit the banking data well.

The study could have adopted different functional forms to estimate the profit function. We avoided the specification error related to the functional form by conducting a log-likelihood ratio test to determine whether the second order coefficients of the translog form are simultaneously equal to zero (i.e. Cobb-Douglas technology is an adequate representation of the data) against the alternative hypothesis that they are not. The results of the test indicated that the specified flexible translog form was a significantly better representation of the data than the Cobb-Douglas technology.

This dissertation employs a cross-sectional data that looks at mergers and acquisitions (M&As) during 1997. The choice of the period, 1997, was motivated by the enactment of the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994. The Act permits adequately capitalized and managed bank holding companies to acquire banks in any state one year after enactment. Beginning June 1, 1997, the Act allows interstate mergers between adequately capitalized and managed banks, subject to concentration limits, state laws, and CRA evaluations. There were 762 reported M&As during 1997. The number of M&As declined to 347 cases after excluding all corporate reorganizations. We collected all the financial data from the Federal Reserve's annual Report on Condition and Income (Call Report).

The study then estimated the alternative profit frontier, and assessed if the specified outputs and input prices have an effect on the alternative profit. In line with the "value added approach" (Berger and Humphrey 1992b) used to define and measure bank outputs, the study found that demand deposits and the consumer loans are positive and statistically significant in determining alternative profit. That is, these items do add value to the bank performance. Contrary to the value added approach, the real estate loans and business loans were found to be negative and statistically significant in determining alternative profit. This implies that a merged bank is more likely to add value if it increases the demand deposits and consumer loans portfolio and reduce the real estate and business loans portfolio. Therefore, during these merger activities, bank management should not just assume that better diversification opportunities would improve efficiency, but should pay attention to the composition of the portfolios of different bank activities. That is, if these findings generalize beyond the sample used here.

The coefficients for the unit price of core deposits and unit price of labor were positive and statistically significant in determining alternative profit. That is, costs associated with increased core deposits portfolio and more personnel adds value into the bank, which also leads to improved performance. This result sheds some light to both bank management and antitrust authorities as reducing overhead is often a stated goal in many bank mergers and bank holding company reorganizations. To the bank management, it was also important to realize that the unit price of physical capital is negative and statistically significant in determining alternative profit. Therefore, cutting physical overhead seems to improve performance.

Another important aspect of this study was the incorporation of a measure of bank risk (RISK) when measuring bank performance. We use, as a measure of bank risk, the Altman Z-score that measures bank's probability of bankruptcy. We not only incorporated this measure in estimating the profit frontier but we also incorporated RISK as an explanatory variable in the model for inefficiency effects. From the profit frontier estimation, the measure of risk was positive and statistically significant in determining alternative profit. That is, merged banks showed a lower volatility of return of assets thereby improving their risk-adjusted performances, which also improved their alternative profit.

The study then estimated the profit frontier and the inefficiency effects model (simultaneously) to determine if the degree of measured efficiency show some improvement when we corrected for bank risk. These results were compared with measured efficiencies from other previous studies of bank efficiency. We found that the predicted technical efficiencies were greater than 0.90 (and some were close to 1.0). The average mean efficiency was 0.9809 (median = 0.9818). These efficiency scores were higher than the efficiency scores summarized by Berger and Humphrey (1997). However, we could not draw direct comparison with studies summarized in Berger and Humphrey (1997) as most of their studies summarized estimated technical efficiencies based on the cost function. This dissertation estimates technical efficiencies based on the profit function. Berger and Humphrey (1997) also indicated that the parametric approaches yielded higher efficiency scores than nonparametric approaches. Within the parametric approaches, the stochastic frontier approach yielded higher efficiency scores than the distribution-free and the thick frontier approach. Therefore, higher efficiency scores in this study did not come as a surprise as we are employing the stochastic frontier approach based on the translog profit function.

We also tested if these banks that increase in size (due to mergers) experience an improved risk-return tradeoff. The results show that the measure of risk was negative and statistically significant in determining technical inefficiency. That is, greater diversification opportunities experienced following a merger increases the Z-score thereby leading to a reduction in technical inefficiency. Therefore, a better diversification of portfolio risk improves bank efficiency.

The study also found that the return on asset (ROA) was negatively related to technical inefficiencies. That is, banks that experience high return on assets turn to be more technically efficient. As banks increase diversification in various aspects of their operations or various geographic areas, they are able to earn more dollars per assets, which improves their efficiencies. However, the results on ROA should be interpreted carefully as this might be attributed to higher market power that banks command in a particular market. The study then tested the Herfindahl-Hirschman Index (HHI – measured at the state level) that measures market concentration or acts as a proxy for bank market power. The results suggests that the more concentrated the market is, at the state levels, the more technically efficient the banks become. In addition, the more market powers the bank commands, the more technically efficient it becomes. These results might be troubling to the antitrust authorities who still use the HHI as one of the determining factor in merger cases. However, this result was not statistically significant, thereby suggesting that the ROA results do not necessarily imply higher market power.

The study also analyzed other variables to determine if the diversification hypothesis still holds. The results showed that merged banks increased their ratio of gross total asset to equity (GTAEQUITY), and business loans to equity (BLOANEQUITY), which was consistent with the diversification hypothesis. This in turn led to improved technical efficiencies. Contrary to the diversification hypothesis, the study also found that an increase in the ratio of loans to gross total assets

(LOANGTA), ratio of consumer and business loans to total equity capital (CLOANEQUITY, BLOANEQUITY) lead to increased technical inefficiency. According to the diversification hypothesis, we expect merged banks to be able to shift out of securities to consumer and business loans as more diversification opportunities arise. However, these results were not statistically significant.

Overall, when we measure bank efficiency correcting for variations in bank risk, the study found that banks showed a decrease in technical inefficiencies that is attributed to more and better diversification opportunities. To the bank management and the antitrust authority, a better quantification of different kinds of risks facing the bank offers a better insight on the prospect of a successful merger case. Therefore, to an extent that a merger would improve bank diversification, we should expect mergers and acquisitions to improve bank efficiency.

6.2 Direction for Future Research

Further studies are needed along these lines to test the robustness of our results. A cross-sectional study assessing firm performances has many limitations, as it cannot measure improvements over a period of time. For bank mergers, studies have indicated that one should measure performance at least two years before the merger and two years after the merger. It will therefore be left to future research to employ a panel data set where we can assess efficiency improvement over a period of time.

Another important area of future research is in modeling the inefficiency effects. The dissertation presented a single-step maximum likelihood procedure designed to obtain consistent parameter estimates and identify determinants of technical inefficiency. The two-step procedure presented in other efficiency studies gives inconsistent parameter estimates. This inconsistency may lead to estimates of technical inefficiency that are inappropriate. Therefore, future research should adopt the single-step maximum likelihood procedure to alleviate this problem.

Future research also needs to apply the one-step procedure in a panel data model that permits the estimation of both technical change and time-varying technical inefficiency, given that inefficiency effects are stochastic and have a known distribution. Researchers should also evaluate a model that combines both the stochastic frontier regression and a truncated regression to estimate the production, cost or profit frontier with non-neutral shifting of the average production function. That is, we need to consider a model where the efficiency correlates can also be specified as interacts of the variables in the stochastic frontier function. This might shed some light on further exploring differences in measured efficiencies.

Another significant result in this dissertation was the specification of the normal-truncated normal distribution. Previous studies on bank efficiency adopted the normal-half normal distribution. Berger and Mester (1997) indicated that the banking data did not fit the normal-half normal distribution well. In this dissertation, we found that the normal-half normal distribution was not an adequate representation of the data given the generalized normal-truncated normal distribution. Therefore, future research should also not only test the distributional assumptions of the model but also apply a more general distribution rather than a restrictive one.

The sample data in this study excludes all corporate reorganizations. Future research should also evaluate if corporate reorganizations are better or worse than acquisitions with respect to efficiency.

BIBLIOGRAPHY

Aigner, D.J., S.F. Chu, (1968), "On Estimating the Industry Production Function", *American Economic Review* 58, 226-239.

Aigner, D.J., C.A.K. Lovell, P. Schmidt, (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", *Journal of Econometrics* 6(1), 21-37.

Akhavein, J.D., A.N. Berger, D.B. Humphrey, (1997), "The Effects of Megamergers on Efficiency and Prices: Evidence from a Bank Profit Function", *Review of Industrial Organization* 12, 95-139.

Allen, L., A. Rai, (1996), "Operational Efficiency in Banking: An International Comparison", *Journal of Banking and Finance* 20, 655-672.

Altunbas, Y., S.P. Chakravarty, (2001), "Frontier Cost Functions and Bank Efficiency", *Economic Letters* 72, 233-240.

Battese, G.E., (1992), "Frontier Production Functions and Technical Efficiency: A Survey of Empirical Application in Agricultural Economics", *Agricultural Economics* 7, 185-208.

Battese, G.E., T.J. Coelli, (1988), "Prediction for Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data", *Journal of Econometrics* 38, 387-399.

Battese, G.E., T.J. Coelli, (1992), "Frontier Production Function, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India", *Journal of Productivity Analysis* 3, 153-169.

Battese, G.E., T.J. Coelli, (1995), "A Model for Technical Inefficiency Effects in a Stochastic Frontier Production Function for Panel Data", *Empirical Economics* 20, 325-332.

Battese, G.E., G.S. Corra, (1977), "Estimation of a Production Frontier Model: With Application to the Pastoral Zone of Eastern Australia", *Australian Journal of Agricultural Economics* 21(3), 169-179.

Bauer, P.W., A.N. Berger, G.D. Ferrier, D.B. Humphrey, (1998), "Consistency Conditions for Regulatory Analysis of Financial Institutions: A Comparison of Frontier Efficiency Methods", *Journal of Economics and Business* 50, 85-114.

Bauer, P.W., A.N. Berger, D.B. Humphrey, (1993), "Efficiency and Productivity Growth in U.S. Banking", in: Fried, H.O., C.A.K. Lovell, S.S Schmidt, (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, (Oxford: Oxford University Press), 386-413.

Bauer, P.W., G.D. Ferrier, (1996), "Scale Economies, Cost Efficiencies, and Technological Change in Federal Reserve Payments Processing", *Journal of Money, Credit and Banking* 4, 1004-1039.

Berg, S.A., F.R. Forsund, L. Hjalmarsson, M. Suominen, (1993), "Banking Efficiency in the Nordic Countries", *Journal of Banking and Finance* 17, 371-388.

Berger, A.N., (1998), "The Efficiency Effects of Bank Mergers and Acquisition: A Preliminary Look at the 1990s Data", in Amihud, Y., Miller, G., (Eds.), *Bank Mergers & Acquisitions*, Boston: Kluwer Academic Press, 79-111.

Berger, A.N., J.D. Cummins, M.A. Weiss, (1997), "The Coexistence of Multiple Distribution Systems for Financial Services: The Case for Property Liability Insurance", *Journal of Business* 79(4), 515-546.

Berger, A.N., R.S. Demsetz, P.E. Strahan, (1999), "The Consolidation of the Financial Services Industry: Causes, Consequences, and Implications for the Future", *Journal of Banking and Finance* 23(2-4), 135-194.

Berger, A.N., R. DeYoung, (1997), "Problem Loans and Cost Efficiency in Commercial Banking", *Journal of Banking and Finance* 21, 849-870.

Berger, A.N., D. Hancock, D.B. Humphrey, (1993), "Bank Efficiency Derived from the Profit Function", *Journal of Banking and Fi*nance 17, 317-347.

Berger, A.N., T.H. Hannan, (1989), "The Price-Concentration Relationship in Banking", *Review of Economics and Statistics* 71, 291-299.

Berger, A.N., D.B. Humphrey, (1992a), "Megamergers in Banking and the Use of Cost Efficiency as an Antitrust Defense", *Antitrust Bulletin* 37, 541-600.

Berger, A.N., D.B. Humphrey, (1992b), "Measurement and Efficiency Issues in Commercial Banking", in: Griliches, Z., (ed), *Output Measurement in the Service Sectors*, National Bureau of Economic Research (University of Chicago Press, Chicago, IL), 245-279.

Berger, A.N., D.B. Humphrey, (1997), "Efficiency of Financial Institutions: An International Survey and Directions for Future Research", *European Journal of Operational Research* 98, 175-212.

Berger, A.N., W.C. Hunter, S.G. Timme, (1993), "The Efficiency of Financial Institutions: A Review, and Preview of Research Past, Present and Future", *Journal of Banking and Finance* 17(2/3), 221-249.

Berger, A.N., L.J. Mester, (1997), "Inside the Black Box: What Explains Differences in the Efficiencies of Financial Institutions?" *Journal of Banking and Finance* 21, 895-947.

Berger, A.N., L.J. Mester, (2003), "Explaining the Dramatic Changes in Performance of U.S. Bank: Technological Change, Deregulation, and Dynamic Changes in Competition", *Journal of Financial Intermediation* 12, 57-95.

Bikker, A.J., (2001), "Efficiency in the European Banking Industry: An Explanatory Analysis to Rank Countries", *Cahiers Economiques De Bruxelles*, 170: 4th trimester.

Boyd, J.H., S.L. Graham, (1989), "Bank Holding Company Risk", in Gup, B.E., *Bank Mergers: Current Issues and Perspective*, (Kluwer Academic Publishers), 197-226.

Charnes, A., W.W. Cooper, E. Rhodes, (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operational Research* 2(6), 429-444.

Charnes, A., W.W. Cooper, E. Rhodes, (1989), "Evaluating Program and Managerial Efficiency: An Application of DEA to Program Follow Through", *Management Science* 27(6), 668-696.

Coelli, T.J., (1992), "A Computer Program for Frontier Production Function Estimation: Frontier Version 2.0", *Economics Letters* 39, 39-32.

Coelli, T.J., (1994), "A Guide to FRONTIER Version 4.1: A Computer Program for Stochastic Frontier Production and Cost Estimation", *Mimeo*, Department of Econometrics, University of New England, Armidale.

Coelli, T.J., (1995), "Estimators and Hypothesis Tests for a Stochastic Frontier Function: A Monte Carlo Analysis", *Journal of Productivity Analysis* 6, 247-268.

Coelli, T.J., D.S. Prasada Rao, G.E. Battese, (1998), *An Introduction to Efficiency and Productivity Analysis*, Kluwer Academic Publishers, Massachusetts, USA.

Cornett, M.M., H. Tehranian, (1992), "Changes in Corporate Performance Associated with Bank Acquisitions", *Journal of Financial Economics* 31, 211-234.

Debreu, G., (1951), "The Coefficient of Resource Utilization", *Econometrica* 19(3), 273-292.

DeYoung, R., (1997a), "Measuring Bank Cost Efficiency: Don't Count on Accounting Ratios", *Financial Practice & Education* 7(Spring/Summer), 20-31.

DeYoung, R., (1997b), "Bank Mergers, X-efficiency, and the Market for Corporate Control", *Managerial Finance* 23(1), 32-47.

Farrell, M.J., (1957), "The Measurement of Productive Efficiency", *Journal of the Royal Statistical Society*, Series A (General), 120 (3), 253-89.

Federal Deposit Insurance Corporation, *Historical Statistics on Banking*, Various years.

Federal Reserve Bank of Chicago, (1997), *Annual Report of Condition and Income* (Call Report).

Fixler, D.J., and K.D. Zieschang, (1993), "An Index Number Approach to Measuring Bank Efficiency: An Application to Mergers", *Journal of Banking and Finance* 17, 437-450.

Focarelli, D., F. Panetta, C. Salleo, (2002), "Why do Banks Merge?", *Journal of Money, Credit, and Banking* 34(4), 1053-1066.

Gallant, A.R., (1997), *An Introduction to Econometric Theory: Measure-Theoretic Probability and Statistics with Application to Economics*, Princeton University Press, Princeton, New Jersey.

Green, W.H., (1980a), "Maximum Likelihood Estimation of Econometric Frontier Functions", *Journal of Econometrics* 13(1), 27-56.

Green, W. H., (1980b), "On the Estimation of a Flexible Frontier Production Model", *Journal of Econometrics* 13(1), 101-115.

Green, W. H., (1990), "A Gamma-Distributed Stochastic Frontier Model", *Journal of Econometrics* 46(1/2), 141-164.

Green, W. H., (2003), *Econometric Analysis*, 5th ed., Upper Saddle River, NJ: Prentice Hall.

Hannan, T.H., (1991), "Bank Commercial Loan Markets and the Role of Market Structure: Evidence from the Survey of Commercial Lending", *Journal of Banking and Finance* 15, 133-149.

Hasan, I., W.C. Hunter, (1996), "Efficiency of Japanese Multinational Banks in the U.S.", *Research in Finance* 14 (Chan, A.H., Ed.) Greenwich, CT: JAI Press, 157-173.

Hjalmarsson, L., S.C. Kumbhakar, A. Heshmati, (1996), "DEA, DFA, and SFA: A Comparison", *Journal of Productivity Analysis* 7(2/3), 303-327.

Huang, C.J., J-T. Liu, (1994), "Estimation of a Non-Neutral Stochastic Frontier Production Function", *Journal of Productivity Analysis* 5, 171-180.

Humphrey, D.B., (1994), "Delivering Deposit Services: ATM versus Branches", Federal Reserve Bank of Richmond, *Economic Quarterly* 80(2), 59-81.

Humphrey, D.B., L.B. Pulley, (1997), "Banks' Responses to Deregulation: Profits, Technology, and Efficiency", *Journal of Money, Credit, and Banking* 29, 73-93.

Jondrow, J., C.A.K. Lovell, I. Materov, P. Schmidt, (1982), "On the Estimation of Technical Inefficiency in the Stochastic Production Function Model", *Journal of Econometrics* 19(2/3), 233-238.

Kaparakis, E.I., S.M. Miller, A.G. Noulas, (1994), "Short-run Cost Inefficiency of Commercial Banks: A Flexible Stochastic Frontier Approach", *Journal of Money, Credit and Banking* 26(4), 875-893.

Kodde, D.A., F.C. Palm, (1986), "Wald Criteria for Jointly Testing Equality and Inequality Restrictions", *Econometrica* 54(5), 1243-1248.

Koopmans, T.C., (1951), "An Analysis of Production as an Efficient Combination of Activities", in T.C. Koopmans (ed), *Activity Analysis of Production and Allocation*, Cowles Commission for Research in Economics, Monograph, No. 13, New York: John Wiley and Sons, INC.

Kumbhakar, S.C., S. Ghosh, J.T. McGuckin, (1991) "A Generalized Production Frontier Approach for Estimating Determinants of Inefficiency in U.S. Dairy Farms", *Journal of Business and Economic Statistics* 9(3), 279-286.

Kumbhakar, S.C., C.A.K. Lovell, (2000), *Stochastic Frontier Analysis*, UK: Cambridge University Press.

Kwan, S.H., R.A. Eisenbeis, (1996), "An Analysis of Inefficiencies in Banking: A Stochastic Cost Frontier Approach", *Federal Reserve Bank of San Francisco Economic Review* 2, 16-26.

Lang, G., P. Welzel, (1999), "Mergers among German Cooperative Banks: A Panel-Based Stochastic Frontier Analysis", *Small Business Economics* 13, 273-286.

Lee, L-F., W.G. Tyler, (1978), "The Stochastic Frontier Production Function and Average Efficiency", *Journal of Econometrics* 7(3), 385-389.

Leibenstein, H., (1966), "Allocative Efficiency vs. X-Efficiency", America Economic Review 56, 392-415.

Linder, J.C., D.B. Crane, (1992), "Bank Mergers: Integration and Profitability", *Journal of Financial Services Research* 7, 35-55.

Lovell, C.A.K., (1993), "Production Frontiers and Productive Efficiency", in Fried, H.O., Lovell, C.A.K., Schmidt, S.S. (Eds.), *The Measurement of Productive Efficiency: Techniques and Applications*, New York: Oxford University Press.

Lozano Vivas, A., (1997), "Profit Efficiency for Spanish Savings Banks", *European Journal of Operational Research* 98(2), 381-394.

Mas-Colell, A., M.D. Whinston, J.R. Green, (1995), *Microeconomic Theory*, New York: Oxford University Press.

McAllister, P.H., D. McManus, (1993), "Resolving the Scale Efficiency Puzzle in Banking", *Journal of Banking and Finance* 17, 389-405.

Meeusen, W., J. van den Broeck, (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", *International Economic Review* 18, 435-444.

Mitchell, K., N.M. Onvural, (1996), "Economics of Scale and Scope at Large Commercial Banks: Evidence from the Fourier-Flexible Functional Form", *Journal of Money, Credit, and Banking* 18, 178-199.

Moore, R.R., T.F. Siems, (1998), "Bank Mergers: Creating Value or Destroying Competition?", *Federal Reserve Bank of Dallas, Financial Industry Issues*, third quarter.

Peristiani, S., (1997), "Do Mergers Improve the X-Efficiency and Scale Efficiency of U.S. Banks: Evidence from the 1980s", *Journal of Money, Credit, and Banking* 29(3), 327-337.

Pitt, M.M., L-F. Lee, (1981), "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry", *Journal of Development Economics* 9, 43-64.

Reifschneider, D., R. Stevenson, (1991), "Systematic Departures from the Frontier: A Framework for the Analysis of Firm Inefficiency", *International Economic Review* 32(3), 715-23.

Rhoades, S.A., (1986), "The Operating Performance of Acquired Firms in Banking Before and After Acquisition", *Staff Economic Studies 149*, Board of Governors of the Federal Reserve System, Washington, D.C.

Rhoades, S.A., (1990), "Billion Dollar Bank Acquisitions: A Note on the Performance Effects", Working paper, *Board of Governors of the Federal Reserve System*, Washington, D.C.

Rhoades, S.A., (1993a), "The Herfindahl-Hirschman Index", *Federal Reserve Bulle*tin 79, 188-189.

Rhoades, S.A., (1993b), "Efficiency Effects of Horizontal (in-market) Bank Mergers", Journal of Banking and Finance 17, 411-422.

Rhoades, S.A., (2000), "Bank Mergers and Banking Structure in the United States, 1980-1998", *Staff Study 174*, Board of Governors of the Federal Reserve System, Washington, D.C.

Shaffer, S., (1993), "Can Megamergers Improve Bank Efficiency? *Journal of Banking and Finance* 17, 423-436.

Shaffer, S., (1998), "Functional Forms and Declining Average Costs", Journal of Financial Services Research 14(2), 91-115.

Spindt, P.A., V. Tarhan, (1992), "Are there Synergies in Bank Mergers?" *Working paper*, Tulane University, New Orleans, LA.

Srinivasan, A., (1992), "Are there Cost Savings from Bank Mergers?" *Economic Review*, Federal Reserve Bank of Atlanta, 17-28.

Srinivasan, A., D. Wall, (1992), "Cost Savings Associated with Bank Mergers", Federal Reserve Bank of Atlanta, *Working paper 92-2* (February).

Stevenson, R.E., (1980), "Likelihood Functions for Generalized Stochastic Frontier Estimation", *Journal of Econometrics* 13(1), 57-66.

Stiroh, J.K., (2003), "Do Community Banks Benefit from Diversification?", Paper Presented at the "*Whither the Community Bank Conference*", Federal Reserve Bank of Chicago, March.

U.S. Department of Justice (1984), "*Merger Guidelines*", Washington D.C., Government Printers.

Vander Vennet, R., (1996), "The Effects of Mergers and Acquisitions on the Efficiency and Profitability of EC Credit Institutions", *Journal of Banking and Finance*, 20, 1531-1558.

Varian, H. R., (1992), *Microeconomic Analysis*, 3rd edition, New York: W.W. Norton & Company.