INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313-761-4700  800-521-0500
EFFICIENCY IN FINANCIAL INTERMEDIATION:
A STUDY OF THE CHILEAN BANKING INDUSTRY

DISSertation

Presented in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Graduate School of The Ohio State University

by

Ricardo Shirota, B.S., M.Sc.

The Ohio State University

1996

Dissertation Committee:
Richard L. Meyer, Ph.D.
Cameron S. Thraen, Ph.D.
D. Lynn Forster, Ph.D.

Approved by
Richard L. Meyer
Advisor
Agricultural Economics and Rural Sociology Graduate Program
To

Antonio Shirota
Guinko Shirota
Massatoshi Suzuki (in memorian)
Setsuko Suzuki (in memorian)
ACKNOWLEDGMENTS

Several people and institutions made important contributions toward the completion of my degree at The Ohio State University. I would like to take this opportunity and mention a few of them who were especially significant and influential in my academic life and in this program, in particular. Of course, to list all of them here would be impractical and I apologize for any one missed.

My adviser, Prof. Richard Lee Meyer with his guidance, support, friendship and patience was an extremely positive role model not only in my academic and professional training but for life as well. Prof. Meyer and his wife, Carol, were exceptionally kind to my family and I during our stay in the U.S. They were very important in facilitating our adjustment to a whole new environment and in the process of learning and appreciating a different culture. The positive experience that we had was certainly improved by their generosity.

Many people at OSU, faculty and staff, were helpful and supportive at different stages of my program. In particular, Profs. D. Lynn Forster and Cameron S. Thraen, members of my dissertation committee, made several comments and gave suggestions that improved this work and my understanding of this investigation. Prof. Thraen's support and friendship throughout all my program at OSU is gratefully acknowledged. Mrs. Barbara Lee, in the Rural Finance Group, with professional competence and kindness, was always helpful with her nice and positive attitude.
Several friends and fellow graduate students in the Department of Agricultural Economics with encouragement, support and acute intellect, made my experience at OSU very interesting and fruitful. I especially mention Dr. Geetha Nagarajan, Hidetaka Kawano, Dr. Koroutoumu Ouattara, Dr. Mayada Baydas, Dr. Shiva Makki, and William Carson.

My program at OSU would have not been possible without the contribution of several people in Brazil. Profs. Joaquim José de Camargo Engler and Evaristo Marzabal Neves had decisive influence on my decision to remain in academia and they have been very supportive throughout my professional life. In especial, I wish to thank my dear mentor and friend Prof. Paulo Fernando Cidade de Araujo whose constant encouragement and support were very important in making the resolution to get a degree in the U.S, and also to make it possible. Support from the faculty of the Departamento de Economia e Sociologia Rural at Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo is also acknowledged.

I wish to acknowledge the financial support from the University of São Paulo, the Rural Finance Group at The Ohio State University and Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP at various stage of my program.

Finally, I express my particular gratitude to all my family for their constant support and love. My parents have constantly and unselfishly encouraged my studies even knowing that it could represent a departure from what they had originally planned for their only son. To my wife, Eliza, and our children, Camila, Ricardo and Alexandre I would like to express my apologies for many sacrifices during this program. Their patience, love and encouragement made the hard times much easier to bear.

To all these people and institutions I wish to express my deepest appreciation.
VITA

April 6, 1960

Born - Marialva, State of Paraná, Brazil

1982

B.S. in Agronomy, Escola Superior de Agricultura
Luiz de Queiroz - ESALQ, Universidade de São
Paulo - USP, Piracicaba, State of São Paulo, Brazil

Agricultural Economist, Fundação de Estudos
Agrários Luiz de Queiroz - FEALQ

1986

Agricultural Economist, Escola Superior de
Agricultura Luiz de Queiroz, Universidade de São
Paulo, Piracicaba, State of São Paulo, Brazil

Instructor, Departamento de Economia e Sociologia
Rural, Escola Superior de Agricultura Luiz de
Queiroz, Universidade de São Paulo, Piracicaba,
State of São Paulo, Brazil

1988

M.S. in Agricultural Economics, Escola Superior de
Agricultura Luiz de Queiroz, Universidade de São
Paulo, Piracicaba, State of São Paulo, Brazil

1988 - present

Assistant Professor, Departamento de Economia e
Sociologia Rural, Escola Superior de Agricultura
Luiz de Queiroz, Universidade de São Paulo,
Piracicaba, State of São Paulo, Brazil

1988 - present
FIELDS OF STUDY

Major Field: Agricultural Economics

Studies in:

- Rural Finance
  - Richard L. Meyer
  - Claudio González-Vega
  - Douglas H. Graham
  - Dale W. Adams
  - Carlos Cuevas

- Production Economics
  - Leroy J. Hushak
  - Cameron S. Thraen

- Statistics and Econometrics
  - Cameron S. Thraen
  - Stephen Cosslet
  - C. S. Cameron

- Banking and Finance
  - Edward J. Kane
  - John H. McCullough
  - Francis A. Longstaff
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ................................................................................................... iii

VITA ....................................................................................................................................... v

LIST OF TABLES ............................................................................................................... ix

LIST OF FIGURES ........................................................................................................... x

INTRODUCTION .................................................................................................................. 1
  1.1. PROBLEM STATEMENT AND ITS RELEVANCE ............................................. 3
  1.2. OBJECTIVES ........................................................................................................ 12
  1.3. ORGANIZATION OF THE STUDY .................................................................. 13

BACKGROUND ON THE CHILEAN ECONOMY ........................................................... 15
  11.1. THE CHILEAN ECONOMY.............................................................................. 17
  11.2. THE EVOLUTION OF THE FINANCIAL SECTOR IN CHILE ................... 21
  11.3. MAJOR SOURCES OF THE CHILEAN FINANCIAL CRISIS ...................... 26

EFFICIENCY AT THE FIRM LEVEL ............................................................................... 28
  III.1. THEORETICAL BACKGROUND ................................................................ 29
  III.2. OUTPUT EFFICIENCY ................................................................................... 34
    III.2.1. Scale Economy \[S(y)\] ............................................................................. 37
    III.2.2. Product-Specific Scale Economy \[S_s(y)\] .................................................. 40
    III.2.3. Ray Scale Economy \[S_{\infty}(y)\] ................................................................. 43
    III.2.4. Trans-Ray Convexity \[TRC\] .................................................................. 44
    III.2.5. Expansion Path Scale Economy \[EPSCE(y^A, y^B)\] .............................. 45
    III.2.6. Scope Economy \[S_e(y)\] ...................................................................... 47
  III.3. INPUT EFFICIENCY ......................................................................................... 48
    III.3.1. The Concept of Input Inefficiency ......................................................... 50
    III.3.2. Methods for Estimating the Frontier ..................................................... 56
LIST OF TABLES

Table 1. Selected Macroeconomic Indicators, Chile, 1970 ~ 90. ........................................ 19
Table 2. Average Nominal Interest Rate on Deposits, Loans, their Spread and the Inflation Rate, Chile, 1975 ~ 91.  .............................................................. 23
Table 3. Number of Financial Intermediaries in the Sample, by Type and Nationality. ........................................... 68
Table 4. Sample Statistics of Selected Variables Used in the Analysis. ................................. 74
Table 5. Estimated Coefficients for the Frontier Cost Function*. ................................. 93
Table 6. Test Statistics for Alternative Production Structures. ......................................... 98
Table 7. Estimated Ray Scale Economy for the Financial Institutions in Chile. .............. 100
Table 8. Estimated Scope Economy for the Financial Institutions in Chile. ................... 103
Table 9. Estimated Expansion Path Scale Economy for the Financial Institutions in Chile. ........................................... 105
Table 10. Selected Statistics for the Estimated Overall Economic Efficiency (EE), Allocative Efficiency (AE) and Technically Efficiency (TE). .............. 105
Table 11. Selected Statistics of the Overall Efficiency Measure (EE) by Type and Nationality of the Financial Institution, Chile, March/1984 ~ December/90. ........................................... 115
Table 12. Estimation Results for the Sources of Overall Input Inefficiency (EE). ................. 117
Table 13. Domestic Commercial Banks and Savings and Loans Included in the Study, Chile. ........................................... 127
Table 14. Foreign Commercial Banks Included in the Study, Chile. ............................... 129
### LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Effect of Different Spreads (l-r) on Deposit and Loan Levels.</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Average Cost (AC), Marginal Cost (MC), and Their Relation to the Scale Economy (S) for the Single Output Case.</td>
<td>38</td>
</tr>
<tr>
<td>3</td>
<td>Hypothetical Cost Surface in the Two Output Case.</td>
<td>39</td>
</tr>
<tr>
<td>4</td>
<td>Examples of Change in Scale and Product Mix in the Two Outputs Case.</td>
<td>41</td>
</tr>
<tr>
<td>5</td>
<td>Graphical Representation of Farrell's Analysis of Technical and Allocative Efficiencies.</td>
<td>51</td>
</tr>
<tr>
<td>6</td>
<td>Distribution of Total Operating Costs (OSEWR) Against the Sum of Total Deposits, Loans and Investments (SUMDLI), All Financial Institutions, Chile.</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>Distribution of Total Operating Costs (OSEWR), Against the Sum of Total Deposits, Loans and Investments (SUMDLI), Small Financial Institutions, Chile.</td>
<td>76</td>
</tr>
<tr>
<td>8</td>
<td>Distribution of the Ratio of SUMDLI to OSEWR Against the Sum of Total Deposits, Loans and Investments (SUMDLI), All Financial Institutions, Chile.</td>
<td>77</td>
</tr>
<tr>
<td>9</td>
<td>Box and Whiskers Plot of the Estimated EE, TE and AE levels.</td>
<td>107</td>
</tr>
<tr>
<td>10</td>
<td>Histogram of the Estimated Overall Economic</td>
<td></td>
</tr>
</tbody>
</table>
Efficiency (EE) and the Plot of its Corresponding (smoothed) Normal Distribution. .................................. 108

Figure 11. Histogram of the Estimated Technical Efficiency (TE) and the Plot of its Corresponding (smoothed) Normal Distribution. ...... 109

Figure 12. Histogram of the Estimated Allocative Efficiency (AE) and the Plot of its Corresponding (smoothed) Normal Distribution. ...... 110

Figure 13. Average EE for each Financial Institution in the Study. ...... 112

Figure 14. Average TE for each Financial Institution in the Study. ...... 113

Figure 15. Average AE for each Financial Institution in the Study. ...... 114

Figure 16. Box and Whiskers Plot of the Estimated EE by Type of Financial Institution. ......................................................... 116

Figure 17. Distribution of the Ratio of SUMDLI to the Number of Branches, Chile. ................................................................. 130

Figure 18. Distribution of the Ratio of SUMDLI to the Number of Employees, Chile. ................................................................. 131

Figure 19. Distribution of the Number of Employees per Branch, Chile. .................................................................................... 132
CHAPTER I

INTRODUCTION

Financial intermediation performs a fundamental role in the process of economic development (World Bank (1989)). A complex financial sector is a common characteristic of any modern economy. There are several theories that explain and justify the existence of financial intermediaries and their role in the market. They differ in terms of the basic reasons financial intermediaries are created. However, the fundamental point in all the existing theories is that well-functioning financial intermediaries make the overall economy more efficient.

Unfortunately, there is considerable empirical evidence showing that many financial institutions operating in developed and developing countries present outcomes that are sub-optimal. Indeed, due to many causes, financial institutions seem to fail to take advantage of potential scale and scope economies. Recent literature shows strong evidence that financial institutions operating in the US are not even able to use inputs in the most efficient way. Either, they use inputs in a sub-optimal combination and/or they simply use more inputs than is technically required.

If this is the case, then this sector's contribution to the overall efficiency of the economy and to its growth could be improved. As explained later in this chapter and the

---

1 See Easterwood (1985) for an overview of some of these theories.
chapters that follow, there are important implications resulting from such findings. First, at the firm level, inefficiency has consequences not only for the firms' profitability but also for its very survival in a competitive market. A firm that is not efficient is a prime target for takeovers or may be forced out of the market by more efficient ones.

Second, from a policy perspective, inefficiency has important implications as well. Basically, it indicates that there is waste in the economy. Most likely improvements in these inefficiencies would potentially generate higher growth rates and induce overall gains in productivity in these firms and in the whole economy.

These problems are especially important in developing countries that are struggling to reach high rates of economic growth. In many cases, their domestic economies suffer from many problems such as market imperfections that make it possible for inefficient firms to survive. The measure of the degree of these inefficiencies and the determination of their causes are the first step toward addressing and solving them. This is especially important in the financial sector due to the crucial role it plays in facilitating transactions in the market and in improving the allocation of resources.

In this sense, the Chilean financial market constitutes an interesting example. It suffered the consequences of the political and economic turmoil that this country went through until the early 1980's. Government regulations and interventions plagued not only the financial market but the whole economy during a long period of time [Edwards and Edwards (1989), Corbo et alii (1986), Nauriyal (1993)]. As a consequence, its average annual economic growth was sluggish in the period between the 30's and early 80's.

However, since the military revolution of 1973, the Chilean economy has moved toward liberalization. Indeed, its aggressive shift to a more open and market driven economy is unprecedented for most countries in recent history. In particular, the financial markets were
deregulated and competition was encouraged. Most of the nationalized banks were returned to the private sector and the domestic market was opened to foreign financial institutions. Unfortunately, excessively liberal policies with weak supervision ended up generating a series of financial crisis during the 1970's and early 80's. The financial market suffered an array of interventions from the Central Bank that intervened and foreclosed troubled institutions. Since then, the financial sector in Chile has shown a healthy return to a growth path consistent with the good performance of the rest of the economy. In this sense, the financial markets of Chile constitute a unique case to study in which the effect of regulation on the efficiency of financial institutions may be studied consistently.

The empirical estimation of firm level efficiencies involves many interesting and challenging methodological issues not completely developed in the literature to date. As presented in Chapter 3, this study makes a contribution to the understanding and application of this technique in empirical studies.

I.1. PROBLEM STATEMENT AND ITS RELEVANCE

Efficiency is a fundamental concept in Neoclassical Theory (NCT). Its importance can be analyzed in several ways. From the point of view of a private firm, efficiency is important in determining its competitive viability.\(^2\) At a more aggregate level, efficiency is related to the problem of optimal allocation of resources.

In a competitive environment, the existence of inefficiency at the firm level (in the long-run) is difficult to justify. Inefficient firms are expected to be driven out of the market by more efficient ones and, in the long-run, only the efficient ones will remain. Thus, for the

\(^2\) A firm is said to be competitively viable if there is no other set of firms with different scale and/or product mixes that can produce the same product mix at lower (scale adjusted) cost [Berger et alii (1987)].
decision maker of a firm it is important to know its relative level of efficiency with respect to other firms in the market and to the frontier of possibilities.

At the social level, less than an optimal allocation of resources generates "dead weight" losses. This implies that society is consuming more in terms of resources used than is technically required to obtain the same level of outputs. Also, more output can often be produced with the same quantity of resources used.

Both points described are basically related to the idea that raises doubts about the postulate of maximization behavior. As mentioned above, basic Neoclassical Theory (NCT) under perfect competition cannot explain the existence of inefficiencies. It implicitly assumes an optimization behavior for the unit under consideration (say, a firm) which will take full advantage of all opportunities in terms of gains and will observe all of the constraints. A violation of either assumption will imply higher costs which is not rational under optimization. However, due to the complexities of economic relations, in many instances economic units are not able to take full advantage of opportunities available in the market. Also, their behavior in the market does not seem to be always rational. In these cases, individual economic units will not be operating on the frontier of technical/economic possibilities.

For the economy as a whole, the optimal allocation of resources is an important factor for the determination of its rate of growth. If scarce resources are not used for their most productive ends, it is clear that an economy will grow at a rate less than its potential capacity.

---

3 Leibenstein (1966) developed an explanation for the existence of "free-lunch" which he named "X-(in)efficiency" (X-E).

4 Among other variables, Leibenstein (1978) argues that differences in the interests between principal and agent, the existence of incomplete contracts, moral hazard, etc. are some of the main sources for inefficiencies.

5 It is important to notice, however, that X-E has been a controversial issue in the literature. See, e.g., Stigler (1976) for an argument against this theory and Frantz (1992) for an argument favorable to it.
The standard neoclassical model of competitive equilibrium (that gives the Pareto solution) only holds if there are no transaction costs [Coase (1960)] and markets are perfect. However, a real economy is full of distortions that make market transactions very costly.\footnote{This is particularly true for transactions in the financial sector.} History shows that in the process of economic development of different societies, as they became larger and more complex, different institutional arrangements were created to ease transactions among a growing number of geographically dispersed economic agents. In many cases, the ability to develop these institutional arrangements was a key element that influenced the speed of their growth and ultimately their success or failure. One of these institutions is the financial system [Fry (1988)].

According to Fry (1988), there are two basic economic functions performed by financial intermediaries. First, they create money and administer the payments mechanism. Second, in a broad sense, they bring savers (surplus units/lenders) and investors (deficit units/borrowers) together. More specifically, the World Bank (1989) mentions that finance matters in an economy because it provides the following services: i) a payment system; ii) saving mobilization; iii) credit allocation; and, iv) limit, price, pool and trade the risk resulting from the process of saving mobilization and credit allocation. Basically, the contribution of a financial system to the development of an economy is that it can make the trade of goods and services, and the process of borrowing and lending less expensive and less risky [Bhatt (1983), Easterwood (1985), Long (1983), Patrick (1966), and World Bank (1989)].

Implicit here is the concept of division of labor and the consequent gains from specialization. By creating an efficient payment system, the financial sector makes trade less expensive. In turn, inexpensive trade gives incentives for the division of labor in production
that results in gains from specialization. Effective savings mobilization, credit allocation and risk management allow further specialization between savers and investors which enhances the production possibility set of the economy as a whole. Finally, as the economy becomes more complex, different types of financial services are created and specialization in the financial sector becomes another source of gains in productivity [Bhatt (1983), and Gurley and Shaw (1967)].

Financial intermediation however, draws upon real resources from the economy, and an efficient financial system is expected to perform financial transactions at minimum cost. The magnitude of the financial sector's contribution to the overall efficiency in the economy is directly related to the degree of efficiency with which the financial system works. The higher the quality and quantity of the services provided, the more significant is its contribution. The quality of financial services can be measured in different ways. First, it is related to its ability to provide a stable and widely accepted medium of exchange. Thus, by allowing inexpensive trade in the goods market and encouraging the process of specialization, it contributes to the general gains of productivity in the whole economy. Second, it is associated with the capacity of offering attractive financial assets which encourage the practice of saving. Third, it depends on the success in evaluating the alternative investments and channeling the resources available to them, and in monitoring the borrowers. This activity will ensure an efficient use of the scarce resources available. Finally, it hinges on the activity of portfolio management. An appropriate portfolio management should pool, price and trade the risks involved in financial intermediation. The World Bank (1987) mentions that trade, efficient use of resources, saving, and risk taking are the cornerstone of a growing economy.
The gains attained by efficient financial intermediation come from different sources [Fry (1988)]. First, there are gains from specialization. Second, there are economies of scale arising from different sources. Search costs, risk premium, liquidity factors, etc., would make transaction costs prohibitively high if individual savers (or investors) had to directly search for borrowers (lenders). Also, the cost of information gathering and processing is more cheaply carried out by a centralized institution. In particular, a customer (borrower or lender) that performs many transactions with the same institution would make the average cost of information much smaller for both the financial intermediary and for the customer as well. Portfolio management is better carried out if a large and diversified portfolio is available. The same applies to the efficiency in the process of size and maturity intermediation.

Basically, the argument is that efficient financial intermediaries should operate with a minimum spread between the cost of borrowing/deposits (r) and the rate of lending (l). Figure 1 shows a hypothetical case in which the spread of financial transactions affects the volume of transactions. Assuming a downward sloping demand for loans (represented by LL') and an upper sloping supply of financial resources (represented by DD'), the market equilibrium is reached at interest rate r = l and quantity (volume of transactions) Q, if no transaction costs exist. However, if transaction costs (represented by the difference between l and r on the vertical axis) are not zero, then the volume of transactions is negatively related to this difference. At the limit, if the spread is high enough, there would be no transactions in the financial market.⁷

---

⁷ In the literature related to the cost of production, economy of scale is a concept associated with the characteristic of the production technology by which the scale-adjusted cost of production declines as the level of output is increased. A more precise definition is given in what follows.

⁸ See Bhatt (1983) for a similar argument in the case of two individuals, one with excess supply of financial resources and the other with excess demand for financial resources.
Earlier theories implicitly assumed that the evolution of the financial sector more or less followed the development of the real sector of the economy. Financial intermediaries were assumed to be aware of market demand in order to offer financial services as the demand for them appeared. However, Patrick (1966) first argued that there are many instances in which the financial sector faces constraints that restrict its development. In such cases, Patrick suggested the "supply leading" approach to develop the financial sector.

---

9 Among others, restrictive regulations in the financial sector, religious or cultural barriers for interest charges on lending/borrowing, and other imperfections in the market mechanisms are cited as causes for the apparent non-responsiveness of the financial sector.
Thus, given the importance of the financial system for economic development and the argument that in many LDCs the financial sector could face many problems if left to develop by itself, many national governments (with financial and technical support of international donor organizations) have undertaken efforts to develop their financial sectors. It is unfortunate that in most cases, these efforts have not only failed to reach their final objectives but have also created negative side effects. It seems that there exists a consensus in the literature that most of the agricultural credit programs developed under "supply-leading" schemes did not achieve the objectives for which they were implemented. Also, a more important issue for this research is the problem of the dependence of many financial intermediaries on external funds. First, this dependence is related to stiff administrative and targeting requirements associated with preferential credit lines. Second, the limited emphasis on deposit mobilization, loan repayment, and the problem of low interest rates on loans kept their operations contingent on the periodical influx of new funds to replenish their portfolio. In particular, interest rate controls, targeting of funds, and requirements for different paper work and reports gave little flexibility to the financial intermediaries. Furthermore, public financial intermediaries in general suffered the consequences of political interventions.

In many cases the main objective was not to develop the financial sector per se. Rather, "supply leading" finance was thought to be an appropriate mechanism to reach different policy objectives. Among others, credit was believed to be an appropriate mechanism to promote development, equity and to carry-out compensatory policies [Bourne and Graham (1983)]. In this sense, the development of the financial intermediaries was an instrument to deliver credit to the final borrower.

For an extensive literature review of many interesting cases see Von Pischke et alii (1983) and Adams et alii (1984). They contain many papers reporting problems generated by "supply-leading" finance in LDCs.

Among other causes, the very nature of fungibility, divisibility and substitution of funds distorted the intended applications of the resources channeled to the final borrower.

It is not uncommon, for example, to find governmental interventions to write-off loans. In general, this is the case of loans to farmers in years of bad weather, especially if it coincides with electoral

(continued...)
problems have a direct effect on the performance of the financial intermediaries, in general, and on their viability in particular.

These external (direct and indirect) interventions constrain the ability of decision makers at the firm or financial intermediary level to freely adjust the level and mix of outputs and inputs to meet the objective of cost minimization or profit maximization. If these constraints are binding, as seems to be the case of financial institutions in LDCs, then they will affect the efficiency of the firms, i.e., firms will be forced to operate at levels of inputs and outputs that are not optimal. Furthermore, if there are many regulations to comply with, managerial capacity is spent on these activities instead of on the more fundamental objective of improving efficiency.

Actually, the concept of viability of financial intermediaries as used in the literature on Rural Financial Markets is closely related to the idea of efficiency in particular, and market competition, in general. According to Meyer (1988), the viability of the rural financial institution is characterized by:

"... (rural financial institution) that is self-sustaining, that covers its costs, that provides services valued by rural households and businesses, that serves an ever increasing number of rural customers, that is dynamic in providing new financial products and services, and that actively searches for ways to improve efficiency reflected in reduced transaction costs for itself and/or its customers".

The first three characteristics are necessary conditions for the existence of a competitive firm in the long run. Unless there exists a steady inflow of external resources, a firm that is not self-sustainable and that does not cover its costs cannot survive in the market. Also, a financial intermediary must offer attractive services to the public. Otherwise it will not get a base of customers that is large enough to justify its existence. The last two characteristics

(...continued)
campaigns.
are consequences of competition. Indeed, one of most important aspects of firms in competitive markets is the search for innovations that make products more attractive to the public, that increase efficiency, that reduce costs, and that increase profits.

Implicitly, Meyer (1988) is criticizing the financial environment in many LDCs which is characterized by extreme regulation that results in the distortions described above. The intermediaries themselves are either overwhelmed by strict regulations and bureaucratic requirements and/or do not have the correct incentives to compete, innovate, and perform. In these cases input and/or output inefficiencies (to be defined later) may arise.

Until recently, most of the studies of banking efficiency were conducted using the cost function approach to study the underlying technology of production in this industry. Other than addressing the problems related to measurement and estimation issues, they limited their analysis to questions related to scale and scope economies. However, it is interesting to notice that the inefficiency resulting from such sources seems to be relatively small [Evanoff and Israilevich (1991), Berger and Humphrey (1992)]. Indeed, most of the empirical studies show that they result in inefficiencies that are, in general, below 10%. These findings are consequences of relatively flat "U-shaped average" cost surfaces. They indicate that the expansion of bank size, beyond very small ones, would not result in significant cost savings, and the expansion of very large banks would even imply higher average costs.

On the other hand, recent literature using a new technique developed to estimate the inefficiency in the use of inputs has shown that this type of inefficiency is much more significant. The use of inputs in an inappropriate combination or the use of inputs beyond technically required levels implies significantly higher costs. Studies using data from the US banking sector show that this type of inefficiency may represent an average cost increase of about 35%.
It seems, therefore, that the early empirical studies of banks may have addressed a secondary issue in terms of efficiency analysis. In the case of developing countries, this is the case in the majority of studies found in the literature. As a consequence, a study using the new approach applied to developing countries will make an important contribution towards confirming or rejecting the US findings. As mentioned earlier, these findings have important implications not only to the managers of the financial institutions but also to the policy makers in the financial sector, and in the whole economy as well.

I.2. OBJECTIVES

The main objective of this study is to measure the level of output and, in particular, input inefficiency of financial institutions in a developing country. This task will be performed through the application of the NCT to production of banking firms. A particular interest is to assess the factors associated with the inefficiencies. These findings may help determine the conditions under which input inefficiency is most likely to occur.

The basic methodology applied is the cost function approach which is implemented using the frontier function technique. This method allows the incorporation of input inefficiency analysis into the classical cost function method.

The specific objectives of this study include:

i. Estimation of a frontier cost function for a sample of banks in Chile;

ii. Test different assumptions regarding the econometric model used to measure inefficiencies;

iii. Calculate output and input efficiencies across different financial institutions in the sample and study their relative performance regarding the measures suggested; and

---

14 A notable exception is the study of Korean agricultural credit cooperatives by Park (1993).
iv. Analyze the potential causes for the existence of any inefficiencies found in (iii) above.

The results of this study are expected to be of importance for decision makers at the level of financial institutions, for policy makers/regulators, and for those interested in the academic study of financial intermediaries. As is explained below in detail, inefficiencies are of crucial importance for the viability of financial intermediaries. Consequently, decision makers should be aware of the level of inefficiencies of their own bank relative to their competitors in the market. For regulators, information about the characteristics of output and input efficiencies can be valuable. First, it may help them determine the potential impacts of different policies on bank performance. Second, it can enhance their understanding of bank characteristics related to efficiency levels which could, in turn, help set policies designed to improve them. Academically, the method proposed in this study may correct previous problems in the analysis of bank firm efficiency.

1.3. ORGANIZATION OF THE STUDY

This dissertation is organized into five chapters. Chapter 2 gives a brief overview about the evolution of the Chilean economy in the recent past. It gives a description of the early 1970's reforms in the economy and the economic evolution since then. In particular, it emphasizes the most important aspects of the reforms affecting the organization and evolution of the financial sector as background to the analysis of efficiencies that follows.

Chapter 3 contains the theoretical background of the analysis proposed in this study. It presents a detailed description of the concepts of output and input inefficiency and the methodology to measure them. The final section of this chapter briefly discusses the problems related to the estimation of the frontier function and the alternative methodologies available. It justifies the choice of the parametric and stochastic frontier used in this study.
in Chapter 4, the hypotheses underlying this study and the empirical model used are presented. Issues related to the problem of variable definition and treatment of the data, the alternative forms of the composed error term used in the study, the functional form for the frontier cost function and its estimation procedure, and the measures of output and input efficiencies are addressed.

Chapter 5 contains the results of the analysis proposed in this study. It presents the estimated frontier cost function and the results for output and input inefficiencies. Some descriptive analysis of the distribution of the estimated input inefficiencies is also presented. It has a section that includes a discussion of a series of potential factors affecting the overall input efficiency in the banking industry in Chile.

Finally, Chapter 6 closes this dissertation with a summary of the findings and a discussion of the conclusions. It ends with a section that discusses the limitations of the study. Specifically, it presents a series of cautions regarding the findings of the study, especially for policy purposes. Suggestions for future research complete this chapter.
CHAPTER II

BACKGROUND ON THE CHILEAN ECONOMY

Chile is located in the southwestern coast of South America. It has 286,397 square miles (757 thousand sq.km.) distributed in a narrow strip between the Andes and the Pacific Ocean that runs from north to south. In the middle of the period considered in this study, Chile had an estimated population of about 12.5 million [World Bank (1989)].

Chile's average annual rate of population growth between 1965 and 1987 was 1.7%. For a developing country, this a strikingly low rate. Despite the problems Chile has had in the recent past, this factor may have mitigated other serious problems faced by many developing countries in Latin America and other regions in the world, and it may facilitate its development in the near future. As is common today in many developing countries, most of Chile's population is located in urban areas. In the 1980-87 period, the urban population grew at an average annual rate of 2.3% and in this last year it represented about 85% of the total population.

15 To put the information in perspective wherever possible, the data presented here relate to the 1980's which is the period considered in the analysis presented later in this study. In particular, if no reference is made, the data refer to 1987.
Relatively speaking, Chile has good social indicators: its life expectancy at birth was a high 72 years; about 70% of school aged children were enrolled in secondary schools; and the per capita daily supply of calories was 2,579.\footnote{Figures for 1986 [World Bank (1989)].}

Chile’s economy has been a laboratory for a wide range of different economic policies in the last 30 years. Since the early 1970’s, the economy has shown a steady movement towards a more open and market driven economy. The financial markets, in particular, experienced significant reform with the lifting of interest rates and quantity controls, the elimination of subsidies and barriers to entry, and the reforms conducted on the regulatory and supervision structures.

The early results of this liberalization process were a series of economic and financial crises that required governmental interventions at different points in time. However, since the mid 80’s, the Chilean economy has shown healthy growth. The financial sector has not suffered any major crisis since the government interventions in the early 80’s. As a consequence of the liberalization process, the financial sector today consists of a variety of different financial institutions in terms of nationality, size and type. Also, it is expected that the financial sector is more competitive and market driven than it was during the period before these interventions. For this study, this is an important point because one can study the effects of market liberalization on banking efficiency.

Thus, the Chilean case gives a unique opportunity to study the effects of financial reforms on the efficiency of firms in this market. Also, the existence of different types of financial institutions operating in this country can give insights into the differences in their economic efficiencies as well.
Finally, in terms of empirical analysis of financial institutions, Chile is a rare case among developing countries where information on the operation of firms in the market has been consistently published for a considerable period of time. If this fact does not constitute a good reason on its own right, on the other hand, the unavailability of reliable and large enough data set would preclude a plausible analysis.  

This chapter presents a brief description about the evolution of the Chilean economy and its financial sector. It ends with a brief section that presents the major sources that caused the Chilean financial crisis of the early 80's. This presentation seems important since the analysis in this study covers the period that followed the financial crisis and the governmental interventions in the early 80's, and is relevant for the investigation of the efficiency of financial institutions in Chile.

II.1. THE CHILEAN ECONOMY

Despite the interesting and positive results presented by the Chilean economy in the recent past, it has had serious problems in the last 50 years. Different political and economic events, both internally and in the rest of the world, have deeply affected Chile during this period.

The economy of Chile is based on the primary activities of agriculture and mining. In 1986, these activities were responsible for about 80% of the value added in the manufacturing sector. In the following year, the primary commodities responded for more than 90% of the total Chilean merchandise exports [World Bank (1989)]. In particular, it is endowed with large deposits of natural resources and in the early 1900's it was once considered a country with

---

17 This is an important point that should not be disregarded as a secondary issue. Financial institutions, in particular, are known to be extremely wary in releasing information on their operations and, in general, it is very difficult to obtain this type of data.
promising future. However, for a series of reasons briefly explained below, Chile was not capable of capitalizing on its potential. The growth rate of the economy was relatively low compared to its Latin American counterparts of Brazil and Mexico.

In 1987, Chile’s per capita GNP was estimated at US$ 1,310 dollars and it was growing at an average rate of 0.2% in the period of 1965-87. The total external public and private debt was about US$ 18 billion dollars which was larger than its GNP. This represented a big burden on its economy because the annual long-term debt service consumed about 26% of its total annual exports of goods and services (Table 1). Furthermore, its current account balance was running a deficit of US$ 811 million dollars in that year. The data presented above show how Chile’s exports are extremely dependent on primary commodities. Indeed, over 90% of its export revenues are based on them. If the volatility of their prices are taken into account, many of the Chilean external problems can be easily explained.

An additional problem Chile has in common with many other developing countries is its history of chronic high inflation rate. In the period of 1980-87, the average annual inflation rate was 20.6% (Table 1). In fact, in the early 70’s it suffered a very serious inflationary process. At its peak, the annual rate of inflation reached 500%.
Table 1. Selected Macroeconomic Indicators, Chile, 1970 – 90.

<table>
<thead>
<tr>
<th>YEAR</th>
<th>RGDP * GROWTH RATE (%)</th>
<th>INFLATION RATE (%)</th>
<th>UNEMPLOYMENT RATE (%)</th>
<th>DEBT SERVICE / EXPORTS (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1970</td>
<td>1.4</td>
<td>34.9</td>
<td>7.1</td>
<td>32.9</td>
</tr>
<tr>
<td>1971</td>
<td>8.9</td>
<td>22.1</td>
<td>5.5</td>
<td>36.8</td>
</tr>
<tr>
<td>1972</td>
<td>-1.2</td>
<td>163.4</td>
<td>3.7</td>
<td>15.6</td>
</tr>
<tr>
<td>1973</td>
<td>-5.5</td>
<td>508.0</td>
<td>4.7</td>
<td>11.8</td>
</tr>
<tr>
<td>1974</td>
<td>1.0</td>
<td>375.9</td>
<td>9.7</td>
<td>15.1</td>
</tr>
<tr>
<td>1975</td>
<td>-12.9</td>
<td>340.7</td>
<td>14.9</td>
<td>31.2</td>
</tr>
<tr>
<td>1976</td>
<td>3.5</td>
<td>174.3</td>
<td>12.7</td>
<td>39.4</td>
</tr>
<tr>
<td>1977</td>
<td>9.9</td>
<td>63.5</td>
<td>11.8</td>
<td>43.9</td>
</tr>
<tr>
<td>1978</td>
<td>8.2</td>
<td>30.3</td>
<td>14.1</td>
<td>44.6</td>
</tr>
<tr>
<td>1979</td>
<td>8.3</td>
<td>38.9</td>
<td>13.6</td>
<td>41.2</td>
</tr>
<tr>
<td>1980</td>
<td>7.8</td>
<td>31.2</td>
<td>10.4</td>
<td>39.6</td>
</tr>
<tr>
<td>1981</td>
<td>5.5</td>
<td>9.5</td>
<td>11.3</td>
<td>64.2</td>
</tr>
<tr>
<td>1982</td>
<td>-14.1</td>
<td>20.7</td>
<td>19.6</td>
<td>64.4</td>
</tr>
<tr>
<td>1983</td>
<td>-0.7</td>
<td>23.1</td>
<td>14.6</td>
<td>48.7</td>
</tr>
<tr>
<td>1984</td>
<td>6.4</td>
<td>23.0</td>
<td>13.9</td>
<td>50.2</td>
</tr>
<tr>
<td>1985</td>
<td>2.5</td>
<td>26.4</td>
<td>12.0</td>
<td>46.0</td>
</tr>
<tr>
<td>1986</td>
<td>5.6</td>
<td>17.4</td>
<td>8.8</td>
<td>40.6</td>
</tr>
<tr>
<td>1987</td>
<td>5.7</td>
<td>21.5</td>
<td>7.9</td>
<td>26.9</td>
</tr>
<tr>
<td>1988</td>
<td>7.4</td>
<td>12.7</td>
<td>6.3</td>
<td>20.3</td>
</tr>
<tr>
<td>1989</td>
<td>10.0</td>
<td>21.4</td>
<td>20.3</td>
<td>-</td>
</tr>
<tr>
<td>1990</td>
<td>2.1</td>
<td>27.3</td>
<td>20.2</td>
<td>25.9</td>
</tr>
</tbody>
</table>

* Real Gross Domestic Product.

Source: Inflation rates are from Boletim Mensual and the remaining variables are from Nauriyal (1993).
In the last 50 years, Chile's pursuit of economic development has experienced a series of swings and turns in terms of political and ideological approach [Nauriyal (1993)]. In general terms, it is possible to divide the evolution of the Chilean economy into two periods: before and after 1973. In the period before 1973, it is clear that Chile's attempt to industrialize and develop its economy was based on import substitution. Among others: i) it established differential tariffs and trade restrictions to protect selected domestic sectors; ii) it determined price controls and subsidies across a wide range of the economy, including ceilings on interest rates; and, iii) it set multiple exchange rates.

By 1973, the consequences of the strategy followed by Chile were reflected in its poor economic performance. It was facing a serious problem of price distortions that fuelled the black market, generated a chronic and high fiscal deficit, a severe inflationary process which reduced real wages, a negative economic growth and a balance of payments crisis [Edwards and Edwards (1987), Corbo (1985)]. In September of that year, a military coup took over the government and implemented a significant reform of the economy. The basic framework of this reform was a macroeconomic stabilization plan through the deregulation of commodity and factor markets. It included the removal of trade barriers and capital flows with foreign markets. The main objectives were to improve resource allocation, eliminate bottlenecks in the economy and restore growth [Corbo et alii (1986)]. The specific measures undertaken included the removal of price controls, liberalization of interest rates, decentralization of financial intermediation and labor market deregulation.

The short term consequence of these measures was a deep economic and social crisis in 1974 and 1975. An indication of its seriousness was the 12.9% decline in real GDP in this last year. However, this adjustment process seemed to have a positive effect on the economy. In
the period of 1977 - 81, its average annual growth rate was above 7% and these were called the Chilean miracle years.

However, the joint effect of the second oil shock, the plunge in the price of copper (its main export commodity), and the international inflation rate that increased interest rates both internationally and in the domestic market had a costly impact on Chile. The drop in export earnings due to the lower copper prices added to the problem caused by the higher oil prices and the larger cost of servicing its international debt. During this period, the Chilean international debt increased from about US$ 7.5 billion dollars in 1978 to over US$ 17 billions of dollars in 1982. Again, the economy suffered a negative growth rate of -14% and the unemployment rate grew to a high 19% in 1982.

The recovery in world copper prices and the deep adjustment process that followed helped Chile recover from its crisis. In fact, by the second half of the 1980’s it was showing signs of healthy improvements. Chile followed a strict policy of orthodox recommendations that exposed the country to the international markets to a degree rarely seen in Latin America. Despite some social costs, the results obtained are an example worthy of consideration by other developing countries.

II.2. THE EVOLUTION OF THE FINANCIAL SECTOR IN CHILE

In the period just previous to the 1973 reform, the socialist government of Chile operated a fully nationalized banking system. At the top of its structure was the Central Bank of Chile - Banco Central de Chile - which also operated as a commercial bank. The financial system was dominated by the commercial banks. Among them, the historically public sector Banco del Estado de Chile was by far the largest and most important financial institution. At that time, foreign financial institutions were prohibited from operating in Chile. Other

Heavy governmental intervention in the financial sector that determined interest rate ceilings, credit targeting and subsidies, high reserve requirements, etc., created an environment favorable for financial depression. In particular, large interest rate spreads and low or even negative real interest rates on deposits were common place [see Table 2].

The financial reform of 1973 was part of the broad economic program put into place by the new government. The basic actions were privatization of the financial institutions and a gradual liberalization of financial activities including all controls on interest rates, credit quantity limitations and barriers to entry. The Chilean financial markets were opened to foreign banks and the regulatory and supervision system was reformed.

The idea was to create an efficient financial system that would play a major role in the new Chilean economy. A market oriented economy would require that the financial intermediaries be capable of mobilizing deposits and providing the economic agents with a wide range of financial and investment options. In the new environment, competition would determine the performance of the economy and the financial markets would be no exception.

These reforms, however, did not prevent a series of crises that hit the financial sector in the following years. Indeed, the second half of the 70's and first half of the 80's witnessed bankruptcies and bailouts of many banks and savings and loans with deep consequences for the entire economy.

---

18 This brief review is presented here to put the evolution of the financial sector of Chile in perspective for the analysis developed later in this study. For a more detailed description of the Chilean financial sector, see these authors.
Table 2. Average Nominal Interest Rate on Deposits, Loans, their Spread and the Inflation Rate, Chile, 1975 — 91.*

<table>
<thead>
<tr>
<th>YEAR</th>
<th>AVERAGE RATES</th>
<th>INFLATION RATE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DEPOSIT</td>
<td>LOAN</td>
</tr>
<tr>
<td>1975</td>
<td>267.3</td>
<td>411.4</td>
</tr>
<tr>
<td>1976</td>
<td>197.9</td>
<td>350.7</td>
</tr>
<tr>
<td>1977</td>
<td>93.7</td>
<td>163.1</td>
</tr>
<tr>
<td>1978</td>
<td>62.8</td>
<td>86.1</td>
</tr>
<tr>
<td>1979</td>
<td>45.0</td>
<td>62.1</td>
</tr>
<tr>
<td>1980</td>
<td>37.4</td>
<td>47.1</td>
</tr>
<tr>
<td>1981</td>
<td>40.7</td>
<td>52.0</td>
</tr>
<tr>
<td>1982</td>
<td>47.9</td>
<td>63.8</td>
</tr>
<tr>
<td>1983</td>
<td>27.9</td>
<td>42.8</td>
</tr>
<tr>
<td>1984</td>
<td>27.6</td>
<td>38.3</td>
</tr>
<tr>
<td>1985</td>
<td>31.9</td>
<td>40.8</td>
</tr>
<tr>
<td>1986</td>
<td>18.9</td>
<td>26.2</td>
</tr>
<tr>
<td>1987</td>
<td>25.2</td>
<td>32.8</td>
</tr>
<tr>
<td>1988</td>
<td>15.1</td>
<td>21.1</td>
</tr>
<tr>
<td>1989</td>
<td>27.7</td>
<td>35.9</td>
</tr>
<tr>
<td>1990</td>
<td>40.2</td>
<td>48.8</td>
</tr>
<tr>
<td>1991</td>
<td>22.3</td>
<td>28.5</td>
</tr>
</tbody>
</table>

* Expressed in percent per year.

Source: The interest rates are from Nauriyal (1993), and the inflation rates are from Boletim Mensual.

The careless auctioning of the publicly owned banks in the privatization process and a soft regulatory and supervisory structure have been blamed for these problems [Diaz-Alejandro
These auctions, mainly directed to the banks' previous owners, were backed by generous credit arrangements that involved no prudent checks on new entrants, free interest rate financing and an unrestricted and soft supervision on the operations of the savings and loans. Also, the new financial sector arising from the privatization process was highly concentrated in the hands of large financial and non-financial corporations. These non-financial corporations were using their new financial "arms" to fund their own activities, many times disregarding proper precautionary actions.

The consequences of this strategy did not take long to appear. As a consequence of a series of bank failures in the end of 1976 and early 1977, the government decided to establish a minimum capital requirement for new entries. An explicit statement made by the central bank that, except for small amounts, deposits were not guaranteed and that the government would not bail-out troubled financial intermediaries in the future informed the market to be aware of the risks involved in financial transactions.

These promises were not kept by the Chilean government, however, in the several years that followed. Already in 1977, it intervened in Banco Osorno, a medium-sized bank and rescued its depositors. The economic depression that the economy was suffering precipitated and deepened the crisis that was just starting. In 1981 CRAV, a conglomerate of diverse interests centered in the sugar industry, filed for bankruptcy and its effect reverberated throughout the rest of the banking system. As a consequence, the Central Bank of Chile was forced to intervene and rescue four banks whose assets represented over 35 % of the whole banking sector [Nauriyal (1993)].

Unfortunately, the economic crisis deepened and the sum of internal and external factors worsened the financial sector problems. The second wave of rescues and interventions began in the second semester of 1982 and culminated in January of 1983. Overall, the Central
Bank was forced to intervene in five banks, liquidate five others and take over the direct supervision of two more.

The clean-up of the problems faced by the financial sector, especially those related to bad assets and investments, and the actions that followed this last round of financial crises in Chile seemed to have had the desired effects. Since then, no further widespread interventions have been necessary, and the growth of the financial sector has been quite significant.

II.3. MAJOR SOURCES OF THE CHILEAN FINANCIAL CRISIS

It seems that the crisis that affected the Chilean financial sector during the late 1970's and early 80's was a consequence of a series of inter-related factors [Diaz-Alejandro (1985), Edwards and Edwards (1987), Nauriyal (1993)]. The major factors explaining the financial crisis in Chile during this period are as follows.

First, the explicit statement that the government would not guarantee other financial intermediaries did not change the general public expectation of implicit government guarantee. Basically, these threats were not taken seriously by the market. This seems to be especially true with foreign lenders. They facilitated a massive capital inflow, especially in 1981, when despite the prevailing fixed exchange rate, a low currency risk with a stable nominal rate was perceived. An important negative effect of this problem is related to the risk taking behavior of the depositors and foreign lenders as well as by the financial institutions themselves.

In the previous years, domestic inflation was running at rates higher than 30% per year. But as a part of the government economic plan, the exchange rate was fixed in July of 1979. Furthermore, the relaxation of the convertibility and capital movements only made the inflow larger [Diaz-Alejandro (1985)].

Of course, the cost of such behavior is borne by the public when the troubled financial institution is rescued by the government.
Second, the lack of proper precautionary regulation and supervision of the agents operating in the financial sector allowed them to capitalize on the market's perception that an implicit government guarantee was in place leading them to riskier behavior. Perhaps this was one of the most important factors generating the whole problem.

Third, the privatization process allowed new entrants without any previous experience in the financial markets to control an important part of the industry. This fact in combination with the soft regulatory and supervision structure became a time bomb in the market.

Fourth, the structure of the financial market that resulted from the privatization process was concentrated in a small number of large financial and non-financial conglomerates. They ended up having significant oligopolistic power. In particular, as mentioned before, many of the banks were funneling funds to the non-financial sector of their conglomerates without attention to adequate financial practices.

Fifth, the liberalization of interest rates along with the relaxation of other controls did not result in larger financial intermediation beyond short-term credits. A significant part of the financial "boom" that followed the reform of the early 70's was based on contracts with short-term maturities (duration shorter than 6 months). In fact, the new policies may have destroyed most of the long-term intermediation arrangements that existed previously. To make things worse, as is usual in developing countries, the stock market did not become an option for long term financing.\(^2\)

Sixth, the liberalization process presented mixed results regarding the generation of financial savings. Also, its impact on aggregate investment is not clear.

---

\(^2\) The major reason for the lack of public interest in the stock market is the widespread belief that there is rampant manipulation and fraud in this market. Furthermore, usually minority stockholders are poorly protected in developing countries.
Finally, external factors such as the sudden oil price increase in the late 70's and the consequent fall in Chile's terms of trade and the increase in international interest rates burdened its already difficult external situation. Facing these problems, it is not surprising that the Chilean government did not discourage the large external capital inflow that occurred at the time.
CHAPTER III

EFFICIENCY AT THE FIRM LEVEL

In a competitive market economy, microeconomic theory shows that efficiency in production is a fundamental factor in determining the survival of a firm. Unless it takes full advantage of the technology available and operates on the frontier of production possibilities, a firm in a competitive market will be driven out of the market.

The theory of production economics addresses the problem of efficiency in different ways. In this study the efficiency of financial intermediaries is analyzed in terms of the level and mix of outputs and inputs. In particular, there is a specific interest in the multi-output firm, as this is the case of most financial intermediaries.22

This chapter first presents the theoretical background underlying the analysis that follows. Next, it introduces the different concepts and measurements used concerning output and input efficiencies.

---

22 A typical financial intermediary has several outputs - loans, deposits, and other services. However, some financial intermediaries only offer loans. This is often the case of many development banks.
III.1. THEORETICAL BACKGROUND

As a first step toward the measurement of the inefficiency of financial institutions, it is necessary to model their behavior and precisely define the meaning of an efficient firm. The results from such a model are then used as a benchmark to compare the performance of all firms analyzed. This task will be performed through the application of the NCT of production under competition. Initially, the optimization behavior of a firm is considered and a definition of efficiency is given. Deviations from the efficient outcome are explained and measurements of the resulting inefficiencies are then suggested.

The theoretical model used to analyze bank efficiency is derived from the microeconomic analysis of the behavior of an individual firm in a competitive market. As any rational economic agent operating in the market, it is assumed that financial institutions react to economic incentives. Thus, their behavior is analyzed using microeconomic production theory.

Basically, the analysis proposed expands the theory of the firm to incorporate the efficiency measures presented below. First, it is postulated that firms produce a vector of outputs \( y \) out of a vector of inputs \( x \). The mathematical notation for these vectors are:

\[
y = [y_1, y_2, \ldots, y_m] \in \mathbb{R}^m
\]

\[
x = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^n
\]

where:
- \( \mathbb{R}^k \) is the k-dimensional, non-negative orthant of the Euclidean space; and,
- \( \mathbb{N} \) is the set of natural numbers.
Second, in this production process, firms face technological constraints. There are some combinations of \( (y,x) \) that are technically feasible and some others that are not. Specifically, there is a maximum output that a firm can produce out of a given quantity of inputs. There are alternative (and equivalent) mathematical representations for this technology [see, e.g., Varian (1992)]. One of them is the input requirement set, \( V(y) \), which is well defined in the multi-output case. The usual representation for \( V(y) \) is given by:

\[
V(y) = \{ x \in \mathbb{R}^n \mid (y,x) \text{ is feasible} \}
\]  

(4)

Third, firms face market constraints. In general, this assumption is represented by price taking behavior. Firms face the demand curve for their products/outputs and supply curves for factors of production that are perfectly elastic and, thus, take input and output prices as given.  

Fourth, firms are assumed to have an objective function that determines their behavior. In production economics the standard cases are output maximization, cost minimization, and revenue and profit maximizations [Varian (1992)]. In each case, these alternative assumptions also determine the variables that are taken as exogenous and endogenous by the firms. The

---

23 It is important to notice that, in practice, this technology is not directly observable.

24 At least for the demand side, this may not be the case in the present analysis. A given financial institution, in a small market, may face a downsloping demand curves for its services and it could behave as a monopolist in that market. However, on the input side, it is reasonable to assume that the supply curves are horizontal for individual financial institutions. To some degree, firms may be able to set output prices but they will take input prices as given. If this is the case, the cost function approach suggested below is still appropriate since it considers outputs as given. Rational (cost minimizing) firms will choose the combination of inputs that minimize costs for a given output levels and input prices [see, e.g., Varian (1992)].

25 If the model is to reflect the firms' behavior, it has to appropriately consider the variables that they take as given and those that are the choice variables. The latter are the ones whose levels result from the optimizing behavior.
choice among them should be based on the underlying firms' behavior. The model chosen for analysis should reflect the constraints faced by the decision makers at the firm level.

For the present work, a maintained assumption is that firms minimize costs. Thus, their behavior can be represented by the following mathematical model:

\[
\begin{align*}
\text{Min} & \quad w \cdot x \\
\text{s.t.} & \quad x \in V(y)
\end{align*}
\]  

where:

- \( y, x, \) and \( V(y) \) are defined above;
- \( w = [w_1, w_2, \ldots, w_n] \in \mathbb{R}^n_+ \) is the vector of input prices; and,
- \( \mathbb{R}^n_+ \) is the n-dimensional, strictly positive orthant of the Euclidean space.

Under some regularity conditions for the technology (monotonic, convex, nonempty and closed input requirement set), it is possible to prove that the result from such an optimization model is a well behaved cost function, defined over the output levels and the input prices [Varian (1992), Chambers (1988)]. This function gives the minimum cost of \( y \) given the technology represented by \( V(y) \) and the input prices \( w \). Moreover, it is possible to show that such a cost function contains all the economically relevant information regarding the underlying technology. That is, it is possible to recover all economically relevant information with respect to the technology directly from the cost function. In particular, if the production function is regular, there is a one-to-one correspondence between the production function and the derived cost function. The usual notation for the cost function is given by:

\[
C' = C(y', w')
\]
There are several reasons why this approach is the one chosen in the present study. First, the usual production function approach cannot handle multiple outputs, which is the case of most financial institutions. Empirically, the production function presents other problems as well. For example, it is known that in most cases, the input quantities are highly correlated among each other, making the direct statistical estimation of the production function difficult.

Second, in the present case, the profit function approach has many shortcomings as well. It assumes that firms maximize profits over the vector of inputs and outputs, given some technological constraint and the vectors of input and output prices. This may not be a good representation of the behavior of financial institutions operating in various developing countries. Many of them were created by governments and other organizations to pursue specific social and developmental goals. Usually, these goals are not compatible with the objective of profit maximization. Furthermore, banks may pursue still other objectives such as larger market share. In these cases, the profit function approach would not be appropriate. In terms of output efficiency analysis, this method has yet another problem. Since it is defined over the vector of input and output prices, it is not possible to derive the output efficiencies directly from the profit function. Empirically, it is much harder to find reliable data on firm profits. On the other hand, cost values are much easier to obtain and they can usually be defined precisely.

Third, compared to the above mentioned alternatives, the cost function presents several advantages: i) it is reasonable to assume that for any level of output and input prices, firms will behave in order to minimize costs; ii) the cost function method is the most popular approach in the literature on banking production economics and the results from this type of study can be directly compared with the findings previously reported; and iii) as shown below,
this method is amenable to measurement not only of output efficiency but input efficiency as well.

It is always interesting to remember that the cost function, as defined above, has some important and well known properties with respect to the input prices. These properties result directly from the optimization process and the characteristics of the underlying technology. They are [Varian (1992)]:

i) $C(y,w)$ is nondecreasing in $w$;

ii) $C(y,w)$ is homogeneous of degree one in $w$;

iii) $C(y,w)$ is concave in $w$; and,

iv) $C(y,w)$ is continuous in $w$.

Furthermore, if competition is introduced, then firms will operate not only on the cost function at any level of output $y$, but they will also choose the output level $y$ that minimizes the "overall" cost of production. If not, they will be driven out of the market by the firms that do so [Baumol et alii (1982), Bailey and Friedlander (1982)]. This point is important for the analysis of efficiency. First, in the general case, if perfect competition holds firms are forced to select the output level $y$ which gives the minimum of the average cost curve. This is related to the concept of output efficiency given below. Second, at any output level $y$, firms have to operate on the cost function. They will use the combination of $x$ that gives the minimum cost of producing $y$. This is associated with the concept of input efficiency.

Thus, in this study, the efficiency of financial intermediaries is analyzed in terms of the level and mix of outputs and inputs. In what follows, precise definitions of output and input

---

26 This terminology is rather imprecise. It is shown later that the problem is a little more complicated in a multi-output industry. A precise description of the problem is given below.
efficiencies are given. Furthermore, the deviations from the efficient outcome are explained and measurements for the degree of inefficiencies are suggested.

III.2. OUTPUT EFFICIENCY

The idea of output efficiency is related to the "correct" choice of $y$. Generally speaking, a firm is said to be efficient in the output dimension if it chooses a scale of operation that produces $y$ in the cheapest way. In the single product case, this is the point where the average cost curve is minimum.

There are basically two main reasons for the existence of output efficiency. First, it may exist due to the nature of the production process. Second, it may be due to restrictions that firms face in the market. These two main sources interact with each other and they affect the shape of the cost function, which is used to measure output efficiency.

The literature reports several (specific) theoretical reasons for the existence of output efficiency for firms in general [Baumol et al. (1988), Bailey and Friedlander (1982)], and for financial institutions in particular [Berger et al. (1987), Evanoff and Israilevich (1991)]. First, in many instances, there are gains from specialization. These gains are derived from the characteristic of production processes in which firms are able to produce more output per unit of input as the level of output increases. As a consequence, productivity increases and the unit cost of the output declines. Second, output efficiency is also gained due to the existence of inputs that are available in discrete quantities only. That is, it may result from the indivisibility of some of the inputs. If this is the case, at certain ranges output can be increased without

\footnote{Examples of this type of inputs in banking are: bricks and mortar, teller machines, computer services, etc.}
requiring more of this type of input. \(^2\) Thus, the output unit cost from these inputs declines in that range. Closely related to this is the case of public inputs \(^3\) [Baumol et alii (1988)].

In the case of financial institutions there are additional sources of output efficiencies that are specific to this type of firm. Before they are examined, however, a few considerations are required with respect to some particular characteristics of the output in this industry. \(^4\) Traditionally, there are three basic types of activities in a typical banking firm: deposits, loans, and other services. The first two do not require much explanation. Deposit services are related to the process of opening and servicing checking, savings, and time deposit accounts for economic agents with cash surpluses. Credit services are those associated with the operations of accepting desired applications for, analyzing, servicing and collecting loans from those economic agents with deficits in their cash flows. Finally, banks also provide their customers with other types of services, such as payment services, funds transfer, safe keeping, etc. These services are offered to regular customers through the payment of a fee or, sometimes, free of charges. Non-regular customers have access to these services through the payment of a fee.

Except for the last case where the services are provided with an up-front fee payment, the two first financial activities have a special characteristic in that the market transaction between the customer (depositor and/or borrower) and the financial intermediary is not complete until the deposit is drawn or the loan is paid. That is, there is a time span between

\(^2\) Notice that, in multi-output industry, this kind of outcome can result either from an increase in all outputs (in a given proportion) or from an increase in one output, keeping the remaining constant. As is shown below, the first one is related to the concept of scale economy and the second one is related to the concept of scope economy.

\(^3\) In financial markets, this is the case with information. Once it is collected for a given customer, it can be used in the future for further transactions with the same customer.

\(^4\) In fact, the problem of output definition and measurement in this sector is more complex than conceptually presented here. For the sake of simplicity, the detailed discussion of this matter is left to the empirical model.
the beginning and the end of each transaction which involves a multidimensional risk component. In the case of deposits, banks face the risk of early withdrawal which forces them to keep a reserve in cash. At the limit, a "run" would force a bankruptcy. For loans, the risk is that borrowers will not pay the full loan on time.

Financial institutions try to minimize these risks through a series of actions. On the deposit side, they offer different types of accounts with different maturities and liquidities over a large base of customers in order to minimize the probability that a major part will be withdrawn at once. Also, they usually work under some kind of scheme such as deposit insurance that guarantees the value of the deposit for their owners. This will diminish the risk of a "run" on the deposits of the bank. On the credit side, besides diversification, banks try to minimize the risk of default of each account with good screening, monitoring and collection efforts. It is not difficult to understand that those activities involve increased expenses as well.

Thus, on the one hand, the expansion of output in this industry will increase costs as a result of higher demand for inputs. On other hand, due to the spread of overhead and information costs, the reduction of the risk factor due to portfolio diversification, the smaller interest paid on deposits as a consequence of customers' conveniences that are captured by the intermediary, etc., there is a reduction in the "unit cost" of the output generated as output increase in the financial sector. In this sense, the net effect whether there is a decline or increases of the "per unit" cost of output as output is expanded seems to be an empirical question.

In what follows, the notion of output efficiency is presented for multi-output cases. First, the definition of scale economy is given for the case of a single output firm. Then, it is shown that such a concept is not appropriate if there are two or more outputs. For these cases, alternative measures are suggested.
III.2.1. Scale Economy [$S(y)$]

In the single output case, efficiency in the output dimension can be analyzed in terms of a well known concept: the scale economy. However, in the multi-output case this concept must be expanded and some other measures must be defined to properly capture the implications of output efficiency. This is the task undertaken in this section.

The standard textbook definition for scale economy, usually referring to the single output case, is given in terms of the marginal change in the quantity of output due to a marginal and proportional change in all inputs [Varian (1992)]. However, for the purpose of this study, it is more convenient to use an alternative definition of scale economy expressed in terms of the cost curve.\(^{31}\) This is not a new concept in economics but it was significantly developed in the literature of industrial organization.

Scale economy is said to exist when the unit cost of production decreases as the scale of production on output level is increased. In the case of a single output, convenient measure for scale economy at any level of output $y$ is given by the following definition:\(^{32}\)

\(^{31}\) Chambers (1988) identifies the first one as economy of scale and the second as economy of size. However, following Baumol et alii (1988), Chamber's economy of size is called economy of scale (or scale economy) in the present study.

\(^{32}\) In what follows, the concepts and notations are close to the ones used in Baumol et alii (1988) and Berger et alii(1987).
Definition 1. Let \( y \in \mathbb{R}_+ \) be the single output of a firm. Then, the degree of scale economy at the output level \( y \), \( S(y) \), is given by:

\[
S(y) = \frac{AC(y)}{MC(y)}
\]  

(8)

where:

- \( AC(y) \) is the average cost at \( y \); and
- \( MC(y) \) is the marginal cost at \( y \).

The returns to scale at \( y \) are said to be increasing, constant or decreasing if \( S(y) \) is greater, equal or smaller than one. This concept is shown graphically in Figure 2.

\[\text{Figure 2. Average Cost (AC), Marginal Cost (MC), and Their Relation to the Scale Economy (S) for the Single Output Case.}\]
The meaning of this definition of scale economy has an important economic implication. Whenever there are increasing returns to scale, total costs increase at a slower rate than output. The opposite is true for the range of $y$ in which there are decreasing returns to scale. Thus, the minimum average cost is achieved at $S(y) = 1$ which is referred to as constant returns to scale (CRTS). This concept is very intuitive in the single output case where there are only two ways of changing the output: either increase or decrease.

However, this is not the situation in the multi-output case ($y \in \mathbb{R}^m$, $m \in \mathbb{N}$ and $m \geq 2$). Here, the cost function is represented by a hyper-surface in the output dimension (see Figure 3 for an example of the cost surface with two outputs). Now, an increase or decrease in output has a less precise meaning unless (as explained below) this change is better qualified. To address this problem, the different types of output change (i.e., different ways of changing scale) are carefully defined.

Figure 3. Hypothetical Cost Surface in the Two Output Case.
Figure 4 shows the four basic ways in which output can change in the two dimensional case. First, a change in "scale" from V to T is the simplest one and represents an increase in one output (y₁) while keeping the other (y₂) constant (Product-Specific Expansion). Second, OC is a change that keeps the output proportions fixed (Ray Expansion). Third, RS is a change that represents an increase in one output as the other one decreases at a linear proportion (Trans-Ray). Finally, fourth, the change AB is a move from an arbitrary point A to another B (Expansion Path) [Berger et alii (1987)].

Following these ideas, the literature basically reports five measurements for the behavior of the multi-output cost surface at a given output level y, according to different directions or types of output changes. They are: Product Specific Scale Economy [S₁(y)], Ray Scale Economy [S₂(y)], Expansion Path Scale Economy [EPSCE(y^A, y^B)], Trans-Ray Convexity (TRC), and Scope Economy [S₃(y)]. Each is individually defined and explained.

III.2.2. Product-Specific Scale Economy [S₁(y)]

Product-specific scale economy is measured along the line VT in Figure 4. It is given by the ratio between an appropriately defined measure of average cost and the marginal cost for a given output yᵢ.

33 The two dimensional case is used here due to the ease in its graphical presentation. However, the concepts developed can be generalized for the n-output case (n > 1).
First, consider the following definitions:

Definition 2. Let $y \in \mathbb{R}^m$ be a vector of output such that $y_i > 0$, $\forall i = 1, 2, \ldots, m$. Let $y_i$ be the vector $y$ with zero output for $y_i$. Then the incremental cost of $y_i$, $IC_i(.)$ evaluated at $y$, is defined as:

$$IC_i(y) = C(y) - C(y_{-i})$$  \hspace{1cm} (9)$$

where:

- $C(.)$ is the cost of producing the argument in the function.
Definition 3. The average incremental cost of output $i$, $AIC_i(.)$ evaluated at $y$ is defined as:

$$AIC_i(y) = \frac{IC_i(y)}{y_i}$$  \hspace{1cm} (10)$$

Then, $S_i(y)$ is defined by:

Definition 4. Product-Specific Scale Economy is defined by:

$$S_i(y) = \frac{IC_i(y)}{y_i, C_i'(y)}$$  \hspace{1cm} (11)$$

where:

- $C(y)$ is the partial derivative of $C(y)$ with respect to the $i^{th}$ output, and evaluated at $y$.

i.e.:

$$C_i'(y) = \frac{\partial C(.)}{\partial y_i}$$  \hspace{1cm} (12)$$

The interpretation of $S_i(y)$ is straightforward. It gives the behavior of the cost surface as output is changed along the line parallel to the $y_i$-axis that passes through $y$, i.e., line VT on Figure 4. If $S_i(y) > 1$, there is increasing RTS in product $i$, and this is reflected in a less than proportional rise in costs as $y_i$'s level of production is increased, holding all other outputs fixed at the level given in $y$. However, if $S_i(y) < 1$, the cost increases more than proportional to the increase in $y_i$.

Notice that this measure is not unique in a sense that there are an infinite number of vectors like VT in Figure 4. That is, for any level of $y_i$, there is a line VT over which the
product-specific scale economy can be measured. The same applies for \( y_2 \). Also, \( S_i(y) \) in one vector \( V_T' \) cannot be compared to another \( S_i(y) \) measured on a different vector \( V_T'' \). It is not difficult to realize that this measure is not very useful, unless some criteria can be found to restrict the number of vectors for analysis.

III.2.3. Ray Scale Economy \([S_m(y)]\)

At this point it should be clear that the major problem with the measurement of output efficiency in multi-output firms is the absence of a single scalar by which the total cost could be divided to give some sort of average cost, as in the case of single output firms. Also, the marginal cost in this case is defined over the vector of \( m \) outputs and, thus, is a vector itself (the gradient vector). Consequently, an appropriate scalar measure of scale economy is required for multi-output cases. Baumol et alii (1988) use the following definition:

Definition 5. In a multi-output firm, the degree of scale economy at the output level \( y \in \mathbb{R}^m \), \( S_m(y) \), is given by:

\[
S_m(y) = \frac{C(y)}{y \cdot \nabla C(y)} = \frac{C(y)}{\sum_{i=1}^{m} y_i C_i(y)}
\]

(13)

where:

- \( \nabla C(y) \) is the gradient vector of \( C(.) \) evaluated at point \( y = [y_1, y_2, \ldots, y_m] \);
- \( y_i \) is the \( i^{th} \) element of \( y \); and
- \( C(y) \) is the \( i^{th} \) element of \( \nabla C(y) \), and is given by equation (12) above.
Again, as in the case of a single output, it is said that there is economies of scale, constant returns to scale and diseconomies of scale whenever $S_m(y)$ is greater, equal to and smaller than one, respectively.

The reason this measure is called Ray Scale Economy (RSCE) is clear from Figure 4 [see, for example, Baumol et alii (1988), Bailey and Friedlander (1982), Berger et alii (1987)]. Actually, the fact that it is measured on a ray passing through the origin imposes the restriction that output quantities are kept in fixed proportions. In this sense, RSCE reduces to the case of single output where the "single" output is a composite one. This is a well defined measure of scale economy for the multi-output case. However, some difficulties still remain in the sense that the degree of scale efficiency is measured along a ray from the origin that passes through the vector of outputs $y$. The problem, again, is that there are an infinite number of such vectors. As it stands, this measure of scale economy is not useful to compare efficiencies among two points on different rays.  

III.2.4. Trans-Ray Convexity [TRC]

As mentioned above, a Trans-Ray change in output is given by the line RS in Figure 4. In the two output case, it represents an increase in one of them at a cost of decreasing the other. It gives a measure of the behavior of the cost surface as the output mix changes.

34 More specifically, $S_m(y)$ is a local concept defined on a given ray. It only makes sense to compare this measure of scale economy among outputs lying on the same ray $t \cdot y$, $t \in \mathbb{R}_{++}$. It is important to notice that in the multi-output case there may potentially exist an infinite number of points for which $S_m(.) = 1$. Moreover, the distance from the origin to the point of $y$ that gives this condition may be different. Thus, this concept has strong limitations if the overall cost surface is to be considered.

35 For three (or more) outputs this concept becomes a little more complicated because the graph of the cost surface over the output dimension is a hyperplane in the three (or more)-dimensional Euclidean space. The general idea is to keep some sort of scale measure fixed. This is a useful concept in many empirical cases [see Thraen and Hahn (1992) for an application].
keeping constant some measure of total output or scale. Actually, TRC has a nice intuitive rationality. In the case of financial intermediaries, the natural scale (to keep constant in TRC analysis) is the total amount of financial resources intermediated and the mix could be the different types of deposits and/or loans made. That is, for a given scale of operation, changes in the mix of outputs could potentially result in cost reduction. Notice, however that there are an infinite number of such lines in the output dimension, one for each level of the output.

TRC is given by:

**Definition 6.** A cost function $C(y)$ is said to be trans-ray convex through some point $y^* \in \mathbb{R}^n$, if for every two output vectors $y^*, y^* \in \mathbb{R}^n$ that lie on the hyperplane $\sum w_i y_i = w_0$ through point $y^*$ we have:

$$C(k y^* + (1-k) y^*) \leq k C(y^*) + (1-k) C(y^*)$$

for any $k$ such that $0 < k < 1$.

III.2.5. Expansion Path Scale Economy [EPSCE($y^\Delta$, $y^\Phi$)]

This is another concept developed to study the behavior of cost surfaces in the multi-output case. It is particularly helpful in addressing two problems. The first is in the case of a comparison of the cost between two different points in the output space that are not on the same ray (e.g., between points A & B in Figure 4). If competitive viability is at issue, Berger et alii (1987) argue that potential competition in the market is likely to come from those firms already operating in the market rather than from hypothetical firms on the ray. As mentioned earlier, $S_{m}$ is not properly defined in these cases. The second problem solved by EPSCE is in the

---

* Their argument for the banking industry is that existing banks rarely have the same product mix.
analysis of scope economies. It is shown below that the usual definition of scope economy involves the determination of costs at zero outputs. Two problems arise here. First, zero outputs may not be observed in practice, and the estimation of costs at such points may be far out of the range of the sample observations. Second, some functional forms such as the popular translog are not defined at zero outputs and further problems arise in the estimation of scope economy in such cases.

To solve these problems, Berger et alii (1987) developed the concept of expansion path scale economy. It is a measure of cost change between two points such as A and B (in Figure 4) with vectors $y^A$ and $y^B$, as follows:

Definition 7. The Expansion Path Scale Economy from some output $y^A$ to some other output $y^B$, $\text{EPSCE}(y^A, y^B)$, is given by:

$$\text{EPSCE}(y^A, y^B) = \frac{\partial}{\partial \ln \ell} \ln \left\{ C(y^A + \ell(y^B - y^A)) - C(y^A) \right\}$$

(15)

$\text{EPSCE}(y^A, y^B)$ measures the elasticity of incremental cost with respect to incremental output along the vector from $y^A$ to $y^B$. The interpretation of EPSCE is similar to the one for $S_M$. The difference now is that the cost behavior is measured along the ray from $y^A$ to $y^B$. An analogy to the single output case can be made in Figure 2 if the horizontal axis is interpreted as

37 Basically, this is the same as $S_M(y^B)$ if $y^A$ is taken as the origin. In this sense, $S_M(\cdot)$ is a special case of $\text{EPSCE}(\cdot)$, where $y^A = 0$. This is easier to see if equation (15) is rewritten as:

$$\text{EPSCE}(y^A, y^B) = \sum_i \left\{ \frac{y_i^B - y_i^A}{r_i^B - r_i^A} \right\} \left\{ \frac{C(y^A + \ell(y^B - y^A)) - C(y^A)}{c_\ell^A} \right\}$$
a composite output that changes along the ray AB. The average and marginal costs, represented on the vertical axis, should also be properly redefined to give these measures with respect to the composite output. As mentioned above, this has strong intuitive appeal. It can be used to compare the change in costs along the ray that connects two firms producing different output levels and/or output mix.

III.2.6. Scope Economy [$S,(y)$]

Scope economy is a measure of the cost advantage of a single firm producing a set of outputs as compared to the same output produced by two or more firms, each producing a single product. Scope economy is said to exist whenever an output $y \in \mathbb{R}^m$, can be produced cheaper by a single firm vis-a-vis two or more separate firms.

For the two output case, scope economy can be represented by:

$$C(y) = C(y_1, y_2) \leq C(y_1, 0) + C(0, y_2)$$

And the degree of scope economy is given by:

**Definition 8.** Let $y$ be the vector of outputs, $y \in \mathbb{R}^m$. The degree of scope economy is defined as:

$$S,(y) = \frac{\sum_{i=1}^{t} C[y_i, 0]}{C(y)}$$

where:

- $C[y_i, 0]$ is the cost of producing $y_i$ only.
It is clear from the definition that \( S(y) \) can be positive or negative. If \( S(y) < 0 \), then the cost of producing \( y \) is smaller for specialized firms. If \( S(y) > 0 \), then a single multiproduct firm can produce \( y \) at a lower cost than many single-output firms.

### III.3. INPUT EFFICIENCY

By symmetry to the output analysis, efficiency in the input dimension is related to the level and mix of inputs used in the production process. Up to now the analysis of output efficiency implicitly assumed that firms are rational economic agents, such that for any output \( y \) the levels and mix of inputs are used in some optimal amount. However, as explained later, there are many instances in which firms seem to miss this optimal outcome. In these cases, it is important that a measure of the departures from the optimal point is available and to find reasons that explain why they occur.

The basic idea underlying efficiency analysis from the input side is the concept of an optimal frontier function. This is a requirement since a standard is needed against which to measure the efficiency of actual outcomes. In economics, it has an intuitive appeal due to the usual definitions of some important economic concepts. This is the case of production, cost and profit functions. They are all defined as frontier concepts. For example, production functions give the maximum output possible given some level of inputs. Similar ideas apply to the other two concepts. The importance of this approach in the analysis of efficiencies is that deviations from these frontiers can be interpreted as inefficiencies.

---

38 In finance, scope economy is related to the cost advantage of specialized financial institutions offering one or only a few services as compared to one that offers a variety of services.

39 This is implicit in most of the standard analysis of production, cost and profit functions. In each case, firms are assumed to maximize an objective function given a set of constraints. The specific objective function and the restrictions change according to whether the production, cost or profit approach is used.
Except for a recent and growing literature, most of the empirical analysis of these functions ignored the frontier nature of their definitions and instead estimated mean functions. In traditional analysis firms are assumed to conform to the competitive paradigm in that they operate on the frontier of the technical and/or economic possibilities. Deviations from these frontiers were usually presumed to be a consequence of random outcomes out of the firm's control and/or errors of measurement on the dependent variable. This implies accepting both positive and negative errors from the frontiers which is, strictly speaking, inconsistent with their original definitions. Furthermore, the estimated residuals were assumed to be random deviations from the function and carry no additional information with respect to the underlying technology and to the actual firm's performance. In particular, they cannot address the problem of inefficiencies in input use.

The frontier function method is an approach that can correct this problem. As shown later, in this approach the deviations from the frontier are modeled in such a way as to reflect input inefficiencies as well.40

There are two basic problems with this method. First, it is necessary to estimate/determine the frontier function itself. Second, an appropriate measure of inefficiency for departures from the frontier must be defined.41 In what follows, the concept of input

---

40 For example, most of the literature on output efficiency was developed in this fashion.

41 In fact, there are many different alternative models reported in the literature. As explained with further details below, the basic differences between them are related to the approach used in the estimation procedure as well as the way inefficiencies are modeled.

42 This is a fundamental point not always mentioned in the empirical literature. It is important to keep in mind that the frontier estimated is dependent on the information contained in the sample observations. If, for some reason, all the firms in the sample are constrained (e.g., by some sort of regulation) then the frontier estimated also reflects this regulation in the sense that the most efficient firms (those on the frontier) may not be operating in their "full efficiency" level. As will be mentioned later, this may have important implications for the analysis of financial intermediaries in regulated

(continued...)
efficiency is presented first. The problem of estimation of the frontier is considered next. Finally, the measurement issue is treated last. In particular, it concentrates on the problem of measurement of departures from this frontier and its decomposition into technical and allocative inefficiencies.

III.3.1. The Concept of Input Inefficiency

The easiest way of presenting the notion of input efficiency is using the original Farrell definition [Farrell (1957)]. To make the graphical presentation possible, assume two inputs \((x_1, x_2)\) with prices represented by \((w_1, w_2)\), a production frontier \(y = f(x_1, x_2)\), and a constant returns to scale technology. Thus, the efficient unit isoquant can be defined as:

\[
1 = \frac{f(x_1, x_2)}{y}
\]

This isoquant is shown by \(y-y'\) in Figure 5. Now, assume a firm with unit output but at some point A, strictly inside \(V(y)\). Under these conditions, Farrell defined two measures of inefficiencies. First, Technical Inefficiency (TE) is defined as the ratio of the distance OB to the distance OA. It is a measure of the potential reduction in the use of all inputs if the firm was producing the same output on the frontier.

Second, Allocative Inefficiency (AE) is given by the ratio of the distance OC to the distance OB and is a measure of the inefficiency arising from the suboptimal mix of inputs, given the prices represented by \((w_1, w_2)\). This makes intuitive sense since the line \(w_1w_2\) is the markets.

\footnote{It is interesting to notice from Figure 5 that there are an infinite number of ways in which the firm at point A can move toward the frontier. The shift to point B is specific in that the proportions of input use are kept constant.}
isocost curve and the ratio above can be interpreted as the increase in cost due to the "wrong" combination of inputs.

Farrell's indices are radial in a sense that they are measured along the straight line from the origin to the point A. This keeps the proportion of input use constant and allows for a nice cost interpretation of the estimated inefficiencies.

Figure 5. Graphical Representation of Farrell's Analysis of Technical and Allocative Efficiencies.
However, it should be mentioned that Farrell's index of TE caused some controversies in the literature. In particular, Färe and Lovell (1978) pointed that there are two alternative definitions for TE at a point like A. First, inefficiency can be measured in terms of actual input use versus the minimal input technically required to produce output $y$. Second, it can be measured in terms of actual output produced $y$ versus the maximum output technically possible given the input use at A. These authors also note that Farrell's index of TE is considered with respect to the isoquant of production and not to the efficient set of production. Under the restricted conditions imposed in Farrell's original analysis, these problems are not relevant. However, if more general conditions are considered, then the two measures of TE may give different results of inefficiencies for the same set $(y, x)$. And TE, as defined by Farrell, becomes difficult to interpret. To solve these problems they defined an alternative measurement of technical inefficiency named the "Russell measure" which mitigated these issues at the cost of introducing some other problems.\(^4\)

Furthermore, if one considers a technology that does not have CRTS, then the production function cannot be written as in equation (18) above and the definition of TE and AE should be appropriately adjusted (see definitions below). Also, if homotheticity is not assumed, then ex-ante information available at the decision time (rather than ex-post) are the relevant ones and the use of ex-post information may bias the analysis [Fersund et alii (1980), Schmidt (1985), and Bauer (1990)].

Despite some weaknesses in Farrell's original index of TE, it has interesting properties not possessed by alternative measures suggested in the literature [Russell (1985)]. Thus, in

what follows the inefficiency problem is evaluated using the concepts of TE and AE in Farrell’s sense which are given by the next two definitions:

**Definition 9.** Let \((y, x^a) \in Y\) (the production possibility set) be a set of output and input vectors for a given firm \(A\). Let \(x^a\) be a point on the isoquant \(Q(y)\) with the same input proportions as in \(x^a\). Point \(x^a\) is on the intersection between the straight line from the origin to point \(x^a\) and \(Q(y)\). Then technical efficiency of firm \(A\), \(TE(A)\) is defined as:

\[
TE(A) = \frac{||x^a||}{||x^d||}
\]

where:

- \(||x||\) is the norm of vector \(x\).

Thus, as in the case of TE, AE is defined as:

**Definition 10.** Let \(x^a\) and \(x^b\) be as in Definition 9. Let \(x^c\) be on the intersection point between the straight line connecting the origin and \(x^a\) and the isocost line \(w_1, w_2\). Then the allocative inefficiency of firm at point \(A\), \(AE(A)\) is defined as:

\[
AE(A) = \frac{||x^c||}{||x^a||}
\]

Finally, the overall economic efficiency (EE) of a firm at point \(A\) is defined as:

**Definition 11.** Given definitions of TE and AE above, the firm's overall economic efficiency, \(EE\), is given by:

\[
EE(A) = TE \times AE = \frac{||x^c||}{||x^a||}
\]

Consequently, the basic information needed to calculate TE, AE and EE are the vectors of input quantities at points \(A\), \(B\), and \(C\) in Figure 5. The problem is to devise a method to find \(x^a\), \(x^b\), and \(x^c\).

The concepts just presented were developed using the notion of isoquant and the isocost curves, i.e., the frontier benchmark (the isoquant) is derived from the production side of the
firms' activity. It is interesting to notice that the same idea of efficiency can be derived using a different frontier function. Cost and profit frontiers are the natural candidates. In this study the estimation of a cost function is chosen. A cost function is modeled so to incorporate a component to reflect the inefficient part of a firm's performance in the market. This component is then disaggregated to find $x^A$, $x^B$, and $x^C$. In this way, the notions developed above can be used without modification. The methodology, which is presented below, was originally developed by Kopp and Diewert (1982) and simplified in the present work.

The input use at point $A$, $x^A$, is given by the sample observations. Consequently, the problem is in finding $x^B$ and $x^C$. First, $x^C$ can be found as follows. Using Shephard's Lemma, it is known that the optimal level of input use, $x^E$, is given by:

$$x^E = \nabla_w C(y^*, w')$$

where:

$y^* = y^A$ for a firm at point $A$; and

$w^*$ is the vector of actual input prices.

Now, since $x^C$ is on the straight line from the origin to the point $A$, it can be represented by:

$$x^C = \lambda^C \cdot x^A$$

where:

$\lambda^C \in (0, 1]$ is an unknown scalar.

Since $x^C$ is on the line $w_1, w_2$ passing through $x^E$, $\lambda^C$ is given by:
\[
\lambda^c = \frac{C(y^*, w^*)}{w^* \cdot x^A} \tag{24}
\]

Turning to \(x^b\), the same way as in the case of \(x^c\), it is possible to write:

\[
x^b = \lambda^b \cdot x^A \tag{25}
\]

where:

\(\lambda^b \in (0, 1]\) is an unknown scalar.

Furthermore, the vector \(x^b\) is on the isoquant \(yy'\), so that it is a solution for the Shephard's Lemma for some unknown price vector \(w^b\), i.e.:

\[
x^b = \nabla_w C(y^*, w^b), \text{ for some } w^b. \tag{26}
\]

Consequently, equations (25) and (26) represent a system of 2n equations in 2n+1 unknowns, \((x^b, w^b, \text{ and } \lambda^b)\). The required condition to reduce this system is:

\[
w^b = 1 \tag{27}
\]

That is, the price vector is normalized by the \(n^{th}\) element of the vector of input prices,\(^{41}\) reducing the system to 2n equations in 2n unknowns. Now, this system can be solved and \(x^b\) can be found [Kopp and Diewert (1982)]. Actually, these authors fail to recognize that it is possible to substitute equation (25) into equation (26) and, along equation (27), the system reduces to n equations in n unknowns \([w_1, w_2, ..., w_{n-1} \text{ and } \lambda^b]\).\(^{44}\) Of course, \(x^b\) is easily

\[\text{\footnotesize \(41\) It should be clear that the choice of which price to use as a norm is irrelevant to the analysis.}
\]

\[\text{\footnotesize \(44\) This point is mentioned, but not developed, by Zieschang (1983). In fact, it is shown in the empirical part of the present study that the system can be further reduced into a system of (n-1) equations in (n-1) unknowns.}\]
calculated from $\lambda^b$ and $x^A$. Once $x^A$, $x^b$, and $x^c$ are available, TE, AE, and EE of a firm at point A can be estimated.  

III.3.2. Methods for Estimating the Frontier

As mentioned before, the estimation of inefficiency requires a standard or optimal outcome against which to compare actual outcomes. The literature reports several alternative methods to perform this estimation. In what follows, these alternatives are separated into broad groups according to their underlying techniques and a brief review is presented for each.

Most of the recent work on efficiency analysis and frontier estimation begins with a reference to Farrell (1957). In his pioneering work, this author suggests two alternatives to find the frontier function. The first is to determine the frontier from engineering information. The second is to estimate the frontier from empirical data. He recognized the complexities of economic activities in the real world, which would be difficult to accurately capture in engineering models. Thus, the alternative of estimating the frontier from empirical data is the approach used by Farrell and the majority of authors that followed in the field. This is also the choice adopted in this research.

Several, alternative methods were developed to empirically estimate frontier functions using different techniques. Basically, they can be grouped into two main approaches: deterministic and stochastic frontier methods. Each can also be divided into parametric and non-parametric frontier techniques. Thus, they can be divided into four large groups:

---

47 The method described above was developed originally by Kopp and Diewert (1982). Zieschang (1983) proposed an alternative method using the concept of a distance function, which reduces the system of equations to $n$ equations in $n$ unknowns. The results obtained from either methods are equivalent, but the second method is much less intensive in computation. The first method is presented here due to its easier interpretation and because it is possible to simplify the first one into a system of $(n-1)$ equations in $(n-1)$ unknowns.
deterministic and non-parametric (DNP); deterministic and parametric (DP); stochastic and non-parametric (SNP); and stochastic and parametric (SP).  

The DNP method is an extension of several non-parametric tests developed in the theory of production. One particularly interesting and widely used approach in this group is the "data envelopment analysis" method (DEA). It uses a mathematical programming technique (specifically, linear programming) to determine the lower bound of the input requirement set [see Seiford and Thrall (1990) for a review of this method]. It generates the frontier and gives a measure of the inefficiency for each observation in the data set.

The advantage of this method and of all non-parametric approaches as compared to parametric methods is that it does not impose a functional form on the technology, and thus is much more flexible. However, there are important drawbacks as well. First, it does not consider statistical noise in the data set. For example, it implicitly assumes that the data are exact and that the list of inputs is exhaustive. It is clear that such assumptions are too strong for most empirical applications. It is widely accepted in the econometric literature that data

---

48 Several references are available on each of these methods. For the DNP method, see e.g., Aigner and Chu (1968), Timmer (1971), Afriat (1972), Schmidt (1976), and Seiford and Thrall (1990). For the SP method, see, e.g., Fersund et alii (1980), Schmidt (1986), and Bauer (1990).

49 It should be mentioned that there is yet another method called "thick-frontier". It can be considered stochastic and parametric but differs from the SP approach in a fundamental way. Basically, this method divides the sample into quartiles according to a given efficiency criteria. The frontier is then estimated using the traditional "mean-function" method based on the data in the most efficient quartile. The efficiency indices are measured against the rest of the sample observations. There are several weaknesses in this method. First, a criteria is needed to rank the firms in the sample according to some measure of efficiency. This is used to select the observations in each quartile. Second, there is no statistical criteria to select the size of the most efficient sub-sample to perform the frontier estimation. Third, unless the information contained in the remaining observations is of no value for the frontier estimation, ignoring them would make the estimation inefficient.

50 Indeed, many non-parametric tests were developed to check different conditions of the production theory. In particular, tests for the existence of a well defined production function, for constant returns to scale, for homotheticity of the production function, and for violation of cost minimization are some of the examples available.
sets are plagued with noise and that it is not practical to expect information on an exhaustive list of inputs. This method is also particularly sensitive to outliers which can significantly affect the shape of the frontier being estimated and, ultimately, the errors can be interpreted as efficiencies or inefficiencies. The second problem is that the outcome does not have probabilistic statements: firms are either efficient or not, without any range of variation.

As the name implies, SNP methods use non-parametric methods with stochastic components. This method is still in an early stage of development and only a few references can be found in the literature.\textsuperscript{51}

DP methods assume a parametric representation for the underlying production technology. If the cost frontier is considered, it can basically be represented by:\textsuperscript{52}

\[
C \geq C(y, w, \theta) \quad \rightarrow \quad C = C(y, w, \theta) + u, \quad u \geq 0
\]  

(28)

where:

- \(C\) is the dependent variable (actual costs);
- \(C(y, w, \theta)\) is the deterministic cost frontier function defined over the variables \((y, w)\) and parameterized by \(\theta \in \mathbb{R}^k\); and,
- \(u\) are deviations from the frontier (inefficiency).

An example of this model used for production functions is Aigner and Chu (1968) in which the frontier function \(f(x, \theta)\) is assumed to be linear and the approach reduces to a linear programming problem. A statistical version of the problem can be obtained by assuming that the one-sided error is stochastic, that is:

\textsuperscript{51} See comments in Schmidt (1986).

\textsuperscript{52} The cost function frontier is represented here because this is the approach proposed in this study. The production and profit frontier can be presented in similar fashion. For further details, see some of the references above.
Further, if a specific distributional assumption can be made about $u$, then maximum likelihood estimation (MLE) can be performed. Greene (1980) gives the necessary and sufficient conditions on the distribution of $u$ for the MLE to have the usual desirable properties. Manski (1984) suggests an estimation method which estimates the parameters and the distribution of $u$ simultaneously (adaptive estimator). However, this method is still not operational in empirical problems.

Since $u$ are one-sided deviations from the frontier function, they represent inefficiencies and the same criticism as in DNP analysis applies. Furthermore, as compared to non-parametric models, now it is necessary to impose a functional form for the frontier function itself plus a distributional assumption for $u$ (the statistical model). Basically, one has to consider the trade-off between structure and flexibility in choosing among these alternatives.

The parameter estimation can be performed using two different techniques: the Corrected OLS (COLS) or MLE.\(^{53}\) A more detailed discussion on the estimation issues is left to the next chapter which treats the empirical model. Once the cost frontier is estimated, it is possible to use the methodology presented previously to calculate the technical and allocative inefficiencies.

\(^{53}\) See Olson et alii (1980) and Greene (1992) for a description of both methods and also for a comparison between them using a Monte Carlo study.
CHAPTER IV

DATA AND EMPIRICAL MODEL

As discussed in the previous chapter, the efficiency analysis of a banking firm can be conducted using different methods. For this study, the econometric stochastic and parametric method of a composed error term was chosen. Furthermore, a firm's efficiency can be measured through alternative forms of technological description. Production, cost, revenue and profit functions have been suggested in the literature. In this study the cost function approach was chosen due to the reasons mentioned in Chapter 3.

The empirical model specification starts with the following general form of the stochastic and parametric frontier cost function:

\[ C_i = C(y_i, w_i | \theta) + \epsilon_i \]  \hspace{1cm} (30)

where:

- \( C_i \) is the observed cost for the i-th observation;
- \( C(.) \) is a parametric version of the cost function presented in Equation (7);
- \( y_i \in \mathbb{R}^n \) is the vector of output quantities for the i-th observation;
- \( w_i \in \mathbb{R}^n_+ \) is the vector of input prices for the i-th observation;
- \( \theta \in \mathbb{R}^k \) is the vector of k parameters implicitly defined in \( C(.) \);
- \( \epsilon_i = v_i + u_i \);
\[ v_i = N(0, \sigma^2_v) \; ; \text{and} \]
\[ u_i \geq 0, \text{a given random variable.} \]

The empirical implementation of this econometric formulation requires additional information. First, it is necessary to define the relevant variables. That is, one has to specify and measure \( C, y \) and \( w \).

Second, it is necessary to determine a functional form for \( C(.) \), which is the parametric representation of the cost function. The problem here is that other than some general results from the cost minimization process, economic theory does not give any clue regarding the form of the technology. Also, for practical purposes, the underlying production technology is not directly observable. Consequently, neither the functional form of the production function nor the corresponding cost function can be explicitly observed. In these cases, the literature suggests that the use of a flexible functional form may be an appropriate solution.

Third, an econometric approach requires that the one-sided error term, \( u \), be modeled. As explained below, in this study a given one-sided distributional form is proposed.

Fourth, once the issues above are addressed it is necessary to find a method to perform the estimation of the vector \( \Theta \). Also, it is essential to find a proper testing procedure for the estimation performed.

Finally, after presenting the estimation process for the frontier function, the last section of this chapter describes the actual procedures used to estimate the output and input efficiencies. This chapter treats each one of these topics in detail.
IV.1. GENERAL HYPOTHESES

The general modeling hypothesis underlying the present study assumes that the production structure of the financial institutions in Chile can be captured by its dual, multi-output frontier cost function. This study uses a Translog functional form with linear homogeneity in input prices and symmetry conditions built in the cost function.

In terms of their production structure, the main consideration in this study is the estimation of output efficiencies which are represented by the following measures: product specific scale economy, ray scale economy, scope economy and expansion path scale economy. On the input side, it deals with the estimation of overall economic efficiency as well as its disaggregation into the technical and allocative components.

On the output side, this study specifically tests for the presence of cost complementarity and economy of scale, at least in some range of the output vector. The production structure is believed to have the following characteristics: jointness, non-separability and homotheticity. Furthermore, the Cobb-Douglas form is considered non-satisfactory in capturing the characteristics of the banking industry in Chile.

Finally, this study analyses the hypothesis that non-trivial input efficiencies exist in the financial sector in Chile. In particular, it undertakes the task of not only estimating input efficiency but also disaggregating its value into the technical and allocative components. Additionally, once the overall economic efficiency is estimated, it tests the hypothesis that the level of inefficiency is associated with several variables that are expected to represent the financial institutions' characteristics and the environment in which they operate.

If the financial markets in Chile are competitive, one would not expect to find significant inefficiencies. On the other hand, if a significant level of inefficiency is found, then the existence of these inefficiencies over the long run may be an indication of market
imperfections. Furthermore, the level of inefficiencies could be explained by differences in the environment in which banks operate or by their own characteristics. Of course, highly regulated markets with binding prices, quantity and other controls are likely to allow inefficient firms to survive. However, as regulations are lifted and greater competition is allowed, one would expect that the inefficient firms would either adjust and increase efficiency or disappear from the market.

Finally, firms with different characteristics are likely to present different levels of inefficiencies. Better management and administration would improve efficiency. In fact, even different technology could affect the firm's performance.

IV.2. THE PROBLEM OF VARIABLE CHOICE IN BANKING STUDIES

Before starting the discussion regarding variable choice, it is important to keep in mind the material presented in the early part of chapter 2. That is, in a study of cost functions of a firm or industry it is necessary to use a proper set of variables that affect its structure. The wrong choice may induce a biased analysis. This constitutes the central point of the discussion that follows.

There are two controversial tasks in choosing the variables for a study of this nature of the banking sector. The first is the choice of a proper set of variables. Second, once the set of variables is chosen, it is necessary to decide on their metric. The literature reports different approaches, and each follows a well reasoned rationale but, unfortunately, there has not been a definite settlement of these issues. To justify the choice made in the present study as well as to keep in mind the limitations thereof, some of the most relevant points are briefly presented in this section.
The problem of selecting the vector of outputs and their metrics is one of the most contentious issues in the estimation of the cost function in banking. This is, in part, due to the nature of financial intermediation. It seems that there is no argument that most of firms in the financial markets, and banks in particular, are multioutput in nature. They offer deposits, credits/loans and other services to their customers. The difficulty arises from the lack of a strong consensus in the literature about which one of them can be considered as output in this industry. The central question regarding this choice reduces to whether deposits should be considered inputs or outputs.

Among the different approaches, the ones developed in bank production analysis are the most appropriate for this study. Basically, there are three different ways to view outputs in banking and, especially, how deposits are treated. The first one is the so called production approach. Here, output is defined as those activities that use a significant amount of real resources. Banks are viewed as firms that employ labor and capital to produce financial products. This approach implicitly considers that deposits in banks provide valuable services demanded by depositors and, as a consequence, they should be regarded as outputs [Benston and Smith (1976)]. In a cost function analysis, the number of different types of financial accounts are taken as outputs and capital and labor are considered inputs [Benston et alii (1982), Berger and Humphrey (1992)]. Variable costs are represented by operating costs only, not including the interest banks pay on deposits.

Berger and Humphrey (1992) and Hancock (1991) present good reviews of this topic.

Recent literature has challenged the conventional definition of output in the banking literature [see the comments by Frank C. Wyckoff and by Jack E. Triplett in Berger and Humphrey (1992)]. Unfortunately, these new ideas are in the early stages of development and are neither theoretically mature nor empirically operational.
This approach presents some weaknesses. First, it does not consider the interest costs which represent a large part of banking expenses, and it does not distinguish between produced deposits and purchased funds. Furthermore, it fails to recognize some bartered services that are not explicit in the financial market. As a consequence, it may not capture some features of the financial production technology.

An alternative point of view is the intermediation approach (sometimes called asset approach). Here, it is argued that, in most of the financial institutions, deposits are an important source of funds that are needed to provide credit/loan services. Consequently, there are justifications to consider them as inputs [Mester (1987)]. This differs from the production approach in the sense that the metric used is the monetary values of the financial accounts instead of their number. To be consistent with the definition of deposit as an input, the interest expenses are included in the costs [Humprey (1992)].

Critics contend that the choices made in this approach are somewhat arbitrary and different authors select distinct set of variables as inputs with no internal mechanism to solve the argument (see footnote 56). The important services implicit in deposit accounts mentioned above are also ignored by this procedure. Additional problems in this approach result from the fact that the amount of deposits and their prices (interest cost of deposits) appear in the vector of explanatory variables. The first one is included in the vector of outputs and the second in the vector of input prices. The first problem is theoretical in the sense that it violates the condition for duality of the cost function. Second, it is reasonable to suspect that the interest rate on deposits is likely to affect the quantity of deposits [Nauriyal (1993)]. This endogeneity

\footnote{See the comments by Tripplet in Berger and Humprey (1992).}
problem may introduce bias in the econometric estimation of the parameters of the cost function [Judge et alii (1980)].

The third option is to make the decision based on how much each type of service adds to the bank's total aggregate value or profit. Those services that provide small aggregate value are considered as inputs. Those with a large contribution to aggregate value or profit are considered outputs [Berger et alii (1987), Hancock (1991), Fixler and Zieschang (1992)]. Of course, the advantage of this approach is that it permits statistical tests regarding the different hypotheses. But, it not difficult to see that there is an implicit and significant source for bias in this method too. For example, consider time and demand deposits. The first one usually compensates the holders with an explicit interest payment. However, generally, demand deposits do not pay any explicit return to the depositors. Instead, they receive in return different financial services as payment, much like in a barter economy. As a consequence, any empirical analysis that does not take these differences into account may give biased results (see footnote 56).

Of course, from an empirical point of view, it is necessary to keep in mind that data availability may determine the value of these approaches in each case. Unfortunately, some information such as the number of transactions, and the revenue and expenses figures for each account type, are not readily available [Berger and Humphrey (1992)]. These limitations may preclude the implementation of some of these alternatives.

Once a decision is made, one way or another, a related issue that is very important in empirical analysis is the choice of metric for the output vector. One body of literature (usually in the production approach) uses the number of different deposit accounts and the number of credits/loans outstanding as the metric. On the other hand, some authors use the monetary values of these variables as the appropriate metric. This is usually the case in the
intermediation approach. Unfortunately, both methods have shortcomings. First, both
implicitly assume homogeneity in the financial market. However, it is not difficult to realize
that differences in risk and services provided by the firms in this market are widely diverse and
constitute important variables perceived by customers. Second, both measures denote stock
concepts and do not capture the flow of transactions associated with the accounts which are the
most important source of operating expenses in this industry [Cuevas (1989)].

An additional important observation regarding this type of study is related to the
aggregation of variables. In empirical analyses of cost functions, it is usual that different
variables taken into account are in fact results of aggregation. This may result from the
unavailability of data or problems related to the statistical estimation procedure. And, in the
case of firms in the banking sector the problem of data availability on a exhaustive list of
variables is not the norm and aggregation is unavoidable.

Notwithstanding these caveats, the present study adopts a version of the production
approach. Consequently, the differences in the portfolio composition of the source of funds
among the financial intermediaries in the data may affect some of the results. In particular,
those with a larger proportion of purchased funds may show higher efficiencies. The
unavailability of information regarding the composition of banks' deposit portfolio
unfortunately precludes this type of inference.

IV.3. DATA SET AND THE DEFINITION OF EMPIRICAL VARIABLES

The main data set containing the information used in the empirical analysis was
obtained from various issues of Información Financiera, published by the Superintendencia de
Bancos e Instituciones Financieras and Boletín Mensual, published by the Central Bank of Chile. Some additional data were collected from the IMF.

The analysis was performed on the main data set which contains monthly information on 32 domestic and foreign banks and 4 domestic savings and loans operating in the Chilean financial market (Table 3). The data cover the period between March, 1984 and December, 1991. In total, there are 94 monthly observations for each of the financial institutions which makes an overall sample size of 3384 observations.

Table 3. Number of Financial Intermediaries in the Sample, by Type and Nationality.

<table>
<thead>
<tr>
<th>Nationality</th>
<th>Type of F.I.</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Banks</td>
<td>Savings &amp; Loans</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Domestic</td>
<td>14</td>
<td>4</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Foreign</td>
<td>18</td>
<td>0</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Latin American</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>U.S.</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>32</strong></td>
<td><strong>4</strong></td>
<td><strong>36</strong></td>
<td></td>
</tr>
</tbody>
</table>

I am indebted to Dr. Bharat Nauriyal, who made me aware of the existence of these publications and kindly provided a significant part of the data used in this dissertation. A precise description of the data and the criteria for handling some shortcomings are presented in Nauriyal (1993). His data were appended and adjusted to conform to the model used in this dissertation.

The name list of all financial institutions in the sample and their codes is presented in Table 13 and Table 14.
All the monetary values were deflated by the monthly Consumer Price Index (CPI) for the Chilean economy. Several adjustments were made to the variables as described below.

i) Cost: According to the production approach used, this variable is defined as the total operating costs including: rents, use of equipment and material, wages (including benefits), and the cost associated with the risk exposure of each bank. The latter item was measured in a non-traditional way. The idea here is to determine a figure that reflects the cost incurred by the financial institutions from their risk exposure. In the absence of better information, this cost is proxied by the cost of the amount each financial institution holds as provisions for doubtful loans plus the total default portfolio. The cost of risk is calculated as the product of this sum and the banks' average cost of deposits. In this study, the variable representing the operating costs is named OSEWR.

Corrections to the data were also made as described in Nauriyal (1993). The most important adjustment was related to the cost figures during the early period of the analysis. In Chile, Información Financiera publishes the monthly information regarding the costs incurred by the financial institutions in terms of cumulative values within a given year. For example, the value for operating costs published for the month of June is the cumulative values since the beginning of the year. The value of operating costs for May is then subtracted from this value to calculate the operating cost for the month of June. However, as mentioned before, in the early 80's, the Central Bank of Chile made several interventions in the financial sector. It not only closed and took control of many financial institutions but also forced others to clean up their accounting statements. As a consequence, for selected months and financial institutions, the values obtained became negative. Instead of dropping the entire observation, these nonsensical values were replaced by the linear interpolation of the values immediately before and after such occurrence.
Another correction was also performed due to the fact that the data originate from accounting statements. In selected months and for specific banks, the monthly values for some cost values misbehaved in a systematic way. Either, the values for a given month were roughly twice the previous months and dropped to almost zero in the following month or they behaved in the opposite way. That is, they were almost zero in a given month and twice the average value in the month that immediately followed. First, this type of behavior is difficult to explain from a theoretical or operational perspective. Second, due to the way those if left uncorrected, this problem may have serious implications for the analysis. In fact, it could not only affect the estimation of the cost function itself but also bias the estimation of the efficiencies. In this case, a firm can be highly inefficient in one month and highly efficient the following month or vice-versa. Or, which is worse, if this affects wages, for example, a firm can be allocatively inefficient in both months for it would be using too much labor in one month and too little in the following month. Instead of dropping these observations, the values for the suspect months were substituted by their average values.

Furthermore, a distinctive behavior of the published values for labor cost required an additional correction for the variable operating costs. The reason and process of this adjustment is made clear in the section describing the input prices below.

ii) Outputs: The multioutput nature of this industry is taken into account by considering the monthly values of total deposits (TDEP), total loans (TLON) and total investments (TINV) for each bank.

iii) Input prices: The inputs considered in this study are labor and capital. The price of labor (WAGE) is determined by the ratio of total expenditures on wages and benefits for each month divided by the corresponding number of employees. The total expenses in personnel is available monthly, however, a problem here is that the data on the number of
employees are available quarterly only for the months of March, June, September and December of each year. Thus, the data give figures for wages in only these months. Furthermore, for selected banks and years, the observations show a systematic behavior for some months, especially in December. In these months, the estimated wages are roughly twice the value of the other previous and following months. This suggests that these banks paid some extra amount of money to their employees in that month. This fact has important implications. If left unadjusted, the estimated cost function may give biased results in the sense that the extra expenses incurred in these months are related to the whole year and not to this month only. More importantly, the analysis of input efficiency may give seriously biased results as well. The wage figures for these months would be inflated and, their values would be smaller than the correct cost of labor for the other months. The correction procedure adopted was to spread the observed extra cost (above the average wage of the rest of the year) over the whole year. Since the expenses on labor constitute an important part of the firms' total operating costs, an adjustment was also made in the cost figures.

The second problem concerns the absence of observations for the number of employees for three quarters of the year. One option is to just calculate the wages for the months with information on employees and drop the remaining months from the analysis. However, this will considerably reduce the sample size and ignore the information contained in the other variables that are available monthly. Another alternative is to try to fill the observations that

59 In many countries it is common to find legislation that requires employers to pay an extra monthly salary to their employees once a year, usually in December. Also, in many other countries, firms pay bonuses in December.

60 Notice that it is possible that the firms could have paid the extra compensation in the months for which the number of employees are not observed. This should be kept in mind when interpreting the results.
are missing. Using the statistical theory of missing observations,⁶¹ in principle, there are two alternatives to quantify the observed values. First, one could use the available information to fill in the values for the number of employees that are missing. A problem with this approach is that the actual number of employees is likely to change from month to month in a non-systematical way. The second method is to try to fill in the wages directly. This seems to be a better way since wages should be much more stable and change in a systematic fashion. As a consequence, the unobserved wages were calculated using a regression on the observed values.

The price of the second input, capital (PCAP), cannot be observed directly as in the case of labor. Since it is a combination of a series of different inputs, its price must be calculated indirectly. The method adopted here starts with the estimation of the expenditure on capital by taking the difference between the total operating expenses and the expenses on labor. To determine a proxy for price, it is necessary to divide this value by some measure of total capital used. Two common alternatives found in the literature are the total assets of the bank and some measure for the total activities undertaken by the financial intermediary. One option is the sum of total deposits, loans and investments. This is the alternative chosen in this study.

iv) Input quantities: As is clear in chapter 3, the analysis of input efficiency requires information about input use for each observation point (the $x^4$ vector). The choices adopted in this study were described above. Due to the problem of missing observations for the number of employees, technical and allocative efficiencies were estimated only for the months for which this variable was observed. To control for randomness and outliers, the corresponding quantity of capital used in this estimation was the quarterly average instead of the actual value observed. Furthermore, this procedure would incorporate some of the information contained in

⁶¹ See, for example, Little and Rubin (1987).
the other months which would be lost otherwise. The variables quantity of labor and capital used by the financial institutions are, respectively, represented by LABOR and CAPITAL in this study.

All these adjustments were necessary, either to correct several problems found in the dataset or to make it suitable for the estimation proposed. Of course, they were carefully executed so as not to violate the theory underlying this study.

All the monetary values are expressed in million of pesos. As mentioned before, these figures are expressed in real values deflated by the CPI.

IV.4. DESCRIPTIVE PROFILE OF THE FINANCIAL INSTITUTIONS

The sample statistics for the variables defined above are presented in Table 4. Notice that the financial institutions included in this study present a wide distribution regarding these variables. The values for the total operating expenses and the three outputs considered show a large range of variation. Therefore, as expected, the same occurs for the input quantities. On the other hand, the wide variation observed for the price of inputs is somewhat unexpected. In a competitive market, these prices would not be expected to vary over such a wide range. However, the standard deviation calculated for WAGE is quite small and the extreme values likely reflect some problems in the data as described previously. PCAP, on the other hand, seems to be more problematic. Probably part of the variation is a consequence of the complex definition of this variable. As described above, this variable is defined indirectly and it is an aggregation of all inputs other than labor. This important point should be kept in mind when evaluating the process of estimation and interpretation of results.
Table 4. Sample Statistics of Selected Variables Used in the Analysis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number</th>
<th>Minimum</th>
<th>Average</th>
<th>Maximum</th>
<th>Std Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OSEWR</td>
<td>3384</td>
<td>4.8</td>
<td>253.3</td>
<td>3376.4</td>
<td>434.8</td>
</tr>
<tr>
<td>Outputs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDEP</td>
<td>3384</td>
<td>151.9</td>
<td>57573.1</td>
<td>547307.4</td>
<td>92778.6</td>
</tr>
<tr>
<td>TLON</td>
<td>3384</td>
<td>800.8</td>
<td>83167.4</td>
<td>696513.4</td>
<td>132096.2</td>
</tr>
<tr>
<td>TINV</td>
<td>3384</td>
<td>6.3</td>
<td>57849.0</td>
<td>875274.4</td>
<td>127677.3</td>
</tr>
<tr>
<td>Input prices</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WAGE</td>
<td>3384</td>
<td>0.07754</td>
<td>0.22257</td>
<td>1.37406</td>
<td>0.08225</td>
</tr>
<tr>
<td>PCAP</td>
<td>3384</td>
<td>0.00011</td>
<td>0.00152</td>
<td>0.22724</td>
<td>0.00409</td>
</tr>
<tr>
<td>Input quantities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LABOR</td>
<td>1116</td>
<td>26</td>
<td>846.1</td>
<td>8434.0</td>
<td>1391.8</td>
</tr>
<tr>
<td>CAPITAL</td>
<td>3384</td>
<td>2151.7</td>
<td>198589.5</td>
<td>1961476.5</td>
<td>339018.9</td>
</tr>
</tbody>
</table>

A better idea of the size distribution of the firms in the sample is given by Figure 6. It contains a plot of the average OSEWR against a measure of the average total operation [sum of TDEP, TLON and TINV (SUMDLI)] for each bank. It shows that there are a couple of extremely large banks, several medium size ones, and a cluster of many small banks. The legend in this graph also illustrates that most of the large financial institutions are domestic commercial banks.

It is interesting to note that, except for small deviations, the total operating expenses seem to increase more or less in a linear fashion regarding SUMDLI. This is consistent with previous empirical evidence of "flat" average cost curves in the financial sector. On the other hand, this graph seems to contradict the significant economies of scale found in a study of these same Chilean banks [Nauriyal's (1993)].
Unfortunately, due to the clustering of the small banks and the scaling of the axis in Table 4, it is difficult to get a clear picture of their distribution. Figure 7 shows a plot for financial institutions with less than $100 million pesos in terms of SUMDLI. These financial institutions show a much wider distribution, and the banks in the US and OTHER categories are located in the lower range of OSEWR. This indicates that they report lower operating costs. On the other hand, S&L are scattered in the upper range suggesting that for the same level of operation, they bear higher costs.
Figure 7. Distribution of Total Operating Costs (OSEWR), Against the Sum of Total Deposits, Loans and Investments (SUMDLI), Small Financial Institutions, Chile.

The next graph, Figure 8, plots the ratio of SUMDLI to OSEWR against SUMDLI. This ratio is a measure of the level of operation of each financial institution relative to each peso spent in operational costs. The most "efficient" banks show a ratio above $1,500 pesos of SUMDLI for each peso spent in OSEWR. The least efficient ones have a ratio of less than 500. This is a strong indication that there are indeed wide differences in the operational efficiencies among financial institutions operating in Chile. If the inverse of this ratio is taken as a rough measure of average costs, then this figure shows the existence of a significant scale
economies for the small banks at the frontier. But as the size increases, the average cost surface becomes flat, indicating constant returns to scale for larger bank sizes.

Figure 8. Distribution of the Ratio of SUMDLI to OSEWR Against the Sum of Total Deposits, Loans and Investments (SUMDLI), All Financial Institutions, Chile.

Notice again, that the S&L are located in the lower range of Figure 8 and the upper range is determined by those in OTHER and US. The middle range is occupied mainly by DOMESTIC and LATIN, indicating non-trivial differences in their operating efficiencies. Similar plots for the ratio of SUMDLI to the number of branches and to the number of employees and the ratio of number of employees to the number of branches are presented in the
Appendix (respectively, Figure 17, Figure 18, and Figure 19). They all show results that are similar to the ones presented above.

IV.4. SPECIFICATION FOR THE COMPOSED ERROR TERM

The stochastic and parametric method of measuring firm efficiency was originally suggested by Aigner et alii (1977) and Meeusen and Van den Broeck (1977) in the context of frontier production function. As mentioned earlier, the basic difference of this approach from the traditional analysis of production function is the introduction of a one-sided error term, \( u \), in the frontier function being estimated besides the normal random error term \( v \) [see equation (30)]. The error term \( u \) is interpreted as deviations from the frontier which reflect inefficiencies and the inability of some firms to reach the underlying technical and/or economic frontier.

In this method, the estimation is usually performed through the specification of a given distribution for \( u \) and \( v \). The two sided random error terms, represented by the vector \( v \), are assumed to be independent and identically distributed normal random variables with mean 0 and variance \( \sigma_v^2 \). However, there are many different suggestions about the distributional form of the one-sided error term in the frontier cost function. Some of these alternatives are presented below.

One of the first one-sided distributions suggested in the literature and still the most commonly used is the half-normal (the absolute value of a normal random variable), i.e.:

\[
\begin{align*}
    u &= w, \quad w > 0 \\
    u &= 0, \quad \text{otherwise}
\end{align*}
\]

where:
In this case, the error term \( \varepsilon \) in equation (30) is a random variable resulting from the sum of normal and half-normal random variables. Following the derivation by Weinstein (1964), Aigner et alii (1977) present the probability density function (pdf) and the log-likelihood function of such a distribution for the production function. The correspondent pdf and log-likelihood function for the cost function model are\(^{62}\):

\[
\begin{align*}
 f(\varepsilon | \sigma, \lambda) &= \frac{2}{\sigma} \phi \left( \frac{\varepsilon}{\sigma} \right) \left[ 1 - \Phi \left( \frac{-\varepsilon \lambda}{\sigma} \right) \right], \quad -\infty \leq \varepsilon \leq +\infty \\
 \ln L(\theta, \sigma, \lambda | \varepsilon) &= \frac{N}{2} \ln \left( \frac{2}{\pi} \right) - N \ln \pi + \sum_{i=1}^{N} \ln \Phi \left( \frac{\varepsilon_i \lambda}{\sigma} \right) - \frac{1}{2} \ln \left( \frac{\varepsilon_i^2}{\sigma^2} \right)
\end{align*}
\]

where:

- \( \varepsilon_i \) is the \( i \)th element of the error vector \( \varepsilon \) in Equation (30);
- \( \beta \) is the parameter vector implicit in Equation (30);
- \( \Phi(.) \) is the distribution function of the standard normal distribution function; and
- \( \lambda = \frac{\sigma \varepsilon}{\sigma_w} \) \hspace{1cm} (35)
- \( \sigma = \sqrt{\sigma_v^2 - \sigma_w^2} \) \hspace{1cm} (36)

\(^{62}\) Notice that in production functions (as well as in profit and revenue functions) \( \varepsilon = v - u \). However, in cost functions \( \varepsilon = v + u \). As a consequence the pdfs of these models are different.
Equations (35) and (36) are convenient parameterization for the likelihood function presented. They incorporate the one-sided inefficiency component of the model [see, e.g., Aigner et alii (1977), Bauer (1990), and Greene (1992)].

Despite its convenience, the half-normal assumption for the one-sided error component has received some criticism. In particular, it is interesting to notice that distinctly from the two-sided normally distributed error term of the standard regression models in econometrics (which can be justified through large sample arguments), the same claim does not apply in the present case [Greene (1992)]. Also, the half-normal distribution, implicitly imposes the restriction that its mode is equal to zero.

In this study, the truncated normal distribution [Stevenson (1980), Greene (1992)] is adopted to represent the error structure. Suppose that the random variable \( w \) presented in equation (32) has the following distribution:

\[
  w \sim N(\mu_w, \sigma_w^2)
\]

Now, the distribution function and the likelihood function of \( \varepsilon \) are given, respectively, by:

\[
  f(\varepsilon | \sigma, \lambda, \mu_w) = \frac{1}{\sigma} \phi \left( \frac{\varepsilon - \mu_w}{\sigma} \right) \left[ \Phi\left( \frac{\mu_w + \varepsilon \lambda}{\sigma} \right) \sigma \right]^{-1}
\]

and, 

\[
  \text{Var}(u) = \sigma_u^2 = \sigma_w^2 \left( \frac{\pi}{2} - 1 \right)
\]
A relative advantage of this formulation with respect to the half-normal distribution is its flexibility in the sense that the mode of the one-sided distribution is not imposed \textit{a priori}. Instead, the data will "reveal" its value. Furthermore, the half-normal model is nested in the truncated-normal distribution. As a consequence, it is possible to nicely test for these alternative distributions.

\textbf{IV.5. FUNCTIONAL FORM}

The problem of choosing a proper functional form in empirical estimations of production, cost and profit functions has been the topic of a large body of economic literature. An early popular form was the Cobb-Douglas and it has been extensively studied and applied in empirical analysis. However, despite many convenient properties possessed by this and other initially popular functions, they impose certain structures on the model under study. As an example, the study of scale economy using a traditional Cobb-Douglas function is not possible since it imposes constant elasticity of scale.

From the theoretical point of view, one should look for a form that does not impose any \textit{a priori} restrictions on the cost structure other than the ones resulting from the duality relationships and it should allow the data to reveal the underlying measures of output efficiencies. Furthermore, it would be interesting to have a form that permits statistical testing of the different characteristics of the production process [Gilligan et alii (1984)].

To address these problems, a more recent literature has suggested functional forms that are "flexible". These are the case of the (Separable) Quadratic form and the Translog.
specifications. But, again, these forms have their limitations. The classical problem of the Translog function, for example, is its inability to handle variables which have zero values. In the specification of the cost function, and the analysis of output efficiencies, this is a restrictive property and an appropriate adjustment is required to estimate scope economies, for example.

Baumol et alii (1988) present a list of desirable properties that a functional form of a multi-output cost function should have. First, it is required to be a "proper" cost function. Basically, this means that it should be consistent with the underlying theory of cost minimization presented in Chapter 3. Second, the functional form chosen should be able to handle zero values for some of the outputs without imposing implausible cost values. Third, the functional form should not impose any ex-ante cost properties that are relevant to the analysis under consideration. The form chosen should be malleable enough to let the data determine the characteristic of the underlying technology. In many empirical applications, this condition has been claimed for the use of functional forms that in their general form do not impose a priori restrictions on the value of the first and second partial derivatives regarding the relevant variables. Functions satisfying this condition can be interpreted as a second order local approximations to an arbitrary twice continuously differentiable cost function [Boisvert (1982), Hertel (1984), Chambers (1988)]. Fourth, the functional form should be parsimonious in the number of parameters to be estimated.

There are many alternatives that conform to these conditions. In particular, Baumol et alii (1988) propose the Hybrid Translog, the Quadratic and the CES as potentially interesting formulations. Another option that seems appealing is the Composite cost function suggested by Pulley and Braunstein (1992). It is a flexible form which incorporates a log-quadratic structure for the input prices and a quadratic structure for the output quantities.
Unfortunately, most of these options violate one or more of the properties in the desiderata above. Furthermore, some of them generate problems in other areas. For example, the Hybrid Translog and the Composite functions are more demanding in the estimation process and have not become popular in the literature.

On the other hand, the regular Translog form possesses most of the qualities listed by Baumol et alii (1988) and its use has became very popular. Furthermore, its main shortcoming of having problems handling zero outputs is not relevant in the present case for this does not occur in the sample under analysis. The scope economy measure is re-defined so as to avoid the requirement of evaluating the estimated cost function at zero outputs. Finally, this choice is also convenient in the sense that it allows a direct comparison with many results already reported in the literature.

The general form for the Translog cost function can be represented by:

\[
\ln[C(y,w)] = \alpha_0 + \sum_{i=1}^{m} \alpha_i \ln y_i + \sum_{k=1}^{n} \beta_k \ln w_k + \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_{ij} \ln y_i \ln y_j \\
+ \sum_{k=1}^{n} \sum_{l=1}^{n} \beta_{kl} \ln w_k \ln w_l + \sum_{i=1}^{m} \sum_{k=1}^{n} \gamma_{ik} \ln y_i \ln w_k
\]

(40)

where:

- \(\alpha\)'s, \(\beta\)'s and \(\gamma\)'s are the parameters; and,
- \(y\) and \(w\) are vector of outputs and input prices.

By Young's theorem, the following parametric restrictions are imposed:\*:

---

\* Some alternatives have been presented in the literature. In their original papers Aigner et alii (1977) and Meuusen and Van den Broeck (1977) suggested the exponential error structure as a possible alternative. Lee (1983) proposed the so called four parameters Pearson family of distribution. It incorporates many desirable theoretical properties at a cost of great complexity. In fact, due to its difficulty no empirical application using this formulation have been found in the literature. Of course, this is mathematically equivalent to the symmetry condition of the cost function.
\( \alpha_{ij} = \alpha_{ji}; \)

\( \beta_{ki} = \beta_{ik}; \) and

\( \gamma_{ik} = \gamma_{ki}. \)

Furthermore, to be consistent with the duality of the cost function, homogeneity of degree one in input prices is also imposed. This condition is fulfilled by the parametric restrictions given by:

\[
\sum_i \beta_i = 1 \\
\sum_i \beta_{ij} = 0, \quad \forall j \\
\sum_j \gamma_{ij} = 0, \quad \forall i
\]

(41)

One last *caveat* is in order. The flexibility property mentioned above holds for the general Translog form presented in equation (40). However, in many cases, further parametric restrictions are imposed on the cost function for different purposes.\(^{45}\) In such cases, the flexibility property breaks down and one cannot claim that these transformed Translog forms possess the capacity to approximate an arbitrary technology [Hertel (1984), Chambers (1988)]. A more pragmatic approach is to accept this limitation and acknowledge that this kind of empirical estimation requires some sort of functional form. The use of Translog forms should

\(^{45}\) This is the case of homogeneity in prices and the symmetry conditions imposed above. Another instance of this procedure is the dropping of some variables due to statistical problems.
be justified in terms of their ability to place fewer restrictions on the estimation than other competing forms [Chambers (1988)].

IV.6. ESTIMATION PROCEDURE FOR THE COST FUNCTION

Among the different alternatives for estimating the parameters in the models above, this study discusses the Corrected Ordinary Least Squares (COLS) and the Maximum Likelihood (MLE) methods.

Under the assumptions presented above, it is possible to show that all the slope parameters of the cost function can be consistently estimated by simple OLS. The problem is the intercept term which is biased due to the effect of the one-sided disturbance term in the composed random variable $\varepsilon$. This results from the fact that $E(\varepsilon) = 0$. However, consider the following second and third central moment equations of $\varepsilon$:

$$\sigma_\varepsilon^2 = \sigma_\nu^2 + \left( \frac{\pi - 2}{\pi} \right) \sigma_\nu^2$$

$$\mu_\varepsilon^3 = \sqrt{\frac{2}{\pi}} \left( \frac{4}{\pi} - 1 \right) (\sigma_\nu^2)^{\frac{3}{2}}$$

Thus, using the second and third sample moments of $\varepsilon$ it is possible to solve these system of equations for $\sigma_\varepsilon^2$ and $\sigma_\nu^2$. Also, it is possible to show that:

$$E(\mu) = \left( \frac{2}{\pi} \right)^{\frac{1}{2}} \sigma_\nu$$

These equations can be used to correct the intercept term in the model above. The corrected intercept term is:
As mentioned above, COLS gives consistent results but they are not efficient. However, they are useful as initial values for the MLE procedure described below.

The MLE estimation can be performed using the likelihood functions presented above. They are non-linear in the parameters and the estimation requires a numerical algorithm. It is possible to prove that the parameter estimates from these models have all the nice properties common to the MLE's [Greene (1992)]. For these reasons, this is the main method adopted in the present study.

Using the general MLE results, several testing procedures were performed. They can be basically divided into two categories. The first ones are those related to the econometric estimation of the general model. The second group of tests is related to alternative production structures. These tests are briefly discussed in the next chapter that presents the empirical results.

**IV.7. ESTIMATION OF OUTPUT EFFICIENCIES**

Once the appropriate frontier cost function is estimated, it is possible to calculate the output efficiencies. The interest here is to study the shape of the frontier cost surface in the

\[ \alpha_{\text{COLS}} - \alpha_{\text{OLS}} = \left( \frac{2}{\pi} \right)^{\frac{1}{2}} \sigma_{\text{OLS}} \]  

---

66 Greene (1980) presents the necessary and sufficient conditions for a proper maximum likelihood estimation.

67 The MLE estimator for the model used in this study is not implemented in the standard econometric packages typically available as such implementation required programming the exact likelihood function in GAUSS®, Aptech System Inc.
output dimension. Among the different concepts of output efficiency presented previously,
this study actually estimates the following: Ray Scale Economy \( S_m(y) \); Scope Economy
\( S_c(y) \); and, Expansion Path Scale Economy \( \text{EPSCE}(y) \).

i) **Ray Scale Economy:** This measure of output efficiency is calculated using the
reciprocal of equation (13). This choice is followed for two reasons. First, it is the most
common measure of Ray Scale Economy found in the literature. Second, it can be
conveniently written as:

\[
S_m(y) = \sum_{i=1}^{m} \left[ \frac{\partial \ln C(.)}{\partial \ln y_i} \right]
\]  

(45)

If \( S_m(y) = 1 \), then at point \( y \), there are constant returns to scale. However, if the
values of \( S_m(y) < 1 \), there are economies of scale at point \( y \). And, of course, it is said that
diseconomies of scale exist if \( S_m(y) > 1 \). This equation is evaluated at three different points in
the output space: at the median of the first two quartile; at the geometric mean; and, at the
median of the last two quartile of the sample variables.

ii) **Scope Economy:** It was mentioned before that a problem exists with this measure
in the case of the Translog cost function. This results from the fact that this form is unable to
handle variables with zero values, and \( S_c \) explicitly requires that the cost function be evaluated
at zero outputs. Solutions to this problem using small but strictly positive values for the output
do not solve the problem. In fact, it can even bias the analysis [Berger et alii (1987;1). Thus,
the alternative followed here is to evaluate the cost function not at zero outputs but at the
minimum output reported in the data. This is a convenient solution because no zero values for
any of the outputs is observed in the sample. Furthermore, this approach avoids the problem of
evaluating the cost function outside the sample observations. Again, this measure is calculated for the same three points used to calculate the Ray Scale Economy.

iii) Expansion Path Scale Economy: As an alternative to calculating a measure of scale economy along a ray from the origin, EPSCE gives the same measure along a ray connecting any arbitrary points A and B in the output dimension [Berger et alii (1987)]. However, to make sense, these points have to be chosen with a criteria. Berger et alii (1987) perform the analysis using nine classes according to the size of outputs. In this study, the financial institutions are roughly divided into small banks and large banks. Thus, EPSCE is calculated along a ray connecting the mean values of these size classes. The actual EPSCE is reduced to a system of n equations in n unknowns by equating equations (25) and (26), as follows:

$$\lambda_i^8 = \frac{\nabla w_i C(y^*, w^8)}{x_i^4}, \quad i = 1, 2, \ldots, n. \quad (46)$$

Along with the normalization of the price vector, this equation represents a system estimated using the equation that results from the expansion of equation (15) and is presented in footnote 37. This measure is evaluated at two different points. The first one is at the geometric mean along the ray that connects the median value of the first two quartiles and the geometric mean. The second point of evaluation is at the median of the last two quartiles along the ray that connects this point to the geometric mean.

---

* This was mentioned by Zieschang (1983).
IV.8. ESTIMATION OF THE INPUT EFFICIENCIES

As presented previously, the input inefficiencies (TE, AE, and EE) were defined in terms of the input quantities at points A, B and C in Figure 5. They are represented by $x^A$, $x^B$ and $x^C$ in equations (19), (20) and (21). This section presents the method used in the study to find these values.

The vector of actual input quantities used, $x^A$, is obtained from the data. The problem is to obtain the unobserved values $x^B$ and $x^C$.

The vector $x^C$ is obtained from equation (23). In fact, due to the radial nature of these concepts, the problem is obtaining $\lambda^C$ in equation (24). In the present case, this equation is represented by:

$$\lambda^C_i = \frac{C(y^*, w^*, \theta)}{w^*_i \cdot x^*_i}$$

where:

- $C(.)$ is the translog cost function used in the present study;
- $y^*$, $w^*$ and $x^*$ are, respectively, the observed values for the vectors of outputs, input prices and input quantities for the i-th observation; and,
- $\theta$ is the vector of estimated parameters implicit in $C(.)$.

Notice that from equation (23) and the Pythagorean theorem, it is easy to see that $\lambda^C$ is in fact equal to the overall economic efficiency (EE) given by equation (21). That is:

$$\lambda^C = \frac{x^C}{x^A} = \frac{\|x^C\|}{\|x^A\|} = EE$$
This fact will make the actual estimation of input efficiencies a little easier.

The estimation of $x^n$ is more complicated. It was shown that the method developed by Koop and Diewert (1982) involves a solution of a system of $2n$ equations in $2n$ unknowns represented by equations (25), (26) and (27). However, notice that this system can be further reduced to a system of $n$ equations in $n$ unknowns. The unknowns are the $(n-1)$ relative prices at point B and $\lambda^n$. In fact, it can be further simplified into a system of $(n-1)$ equations into $(n-1)$ unknowns substituting $\lambda^n$ defined in the first equation of system (46) into the remaining equations of this system. This will result in the following equations:

$$\frac{\nabla_{w_i} C(y^*, w^*)}{x_i^A} = \frac{\nabla_{w_i} C(y^*, w^*)}{x_i^A}, \quad i = 2, 3, \ldots, n \quad (49)$$

Of course, once the system is solved for the $(n-1)$ relative prices, $\lambda^n$ can be calculated by one of the equations in (46), and $x^n$ can be estimated using equation (25). In the present case, with two inputs ($n=2$), system (49) reduces to a single equation in a single unknown relative price. Its derivation gives:

$$\frac{x_i^A}{x_2^A} \left\{1 - B^* + 2 \left( B_{11} \right) \ln \left( \frac{w_1}{w_2} \right) \right\} - \exp \left\{-\ln \left( \frac{w_1}{w_2} \right) \right\} = 0 \quad (50)$$

where:

- $(w_1 / w_2)$ is the relative input price;
- $(\beta_{11} / 2)$ is a parameter from the translog cost function presented above; and
- $B^* = \beta_1 + \gamma_{11} \ln(y_1) + \gamma_{21} \ln(y_2) + \gamma_{31} \ln(y_3)$

(51)
Again, the empirical solution of equation (50) is not trivial. In the present study, this equation was programmed and used with a commercial computer software that solved it numerically for each observation point. The solution for this highly non-linear equation are the (n-1) relative prices. The same computer program use these estimated relative prices to calculate $\lambda^B$. Notice that the vector of input quantities at point B and C are not necessary for the estimation of input inefficiencies. In fact, they were used previously to develop the concepts of TE, AE and EE. However, using equations (19), (25) and the Pythagorean theorem it is possible to see that $\lambda^B$ is equal to TE. Furthermore, from equations (21) and (48) it is easy to derive that:

$$AE = \frac{\lambda^C}{\lambda^B}$$  \hspace{1cm} (52)

Thus, EE, TE and AE can be estimated directly from $\lambda^B$ and $\lambda^C$. This is the procedure used in this study.
CHAPTER V

EMPIRICAL RESULTS

This chapter presents the results obtained from the empirical application of the model proposed in this study. The first section presents the econometric results and the tests that validate the estimates. The second section contains the estimation of the output efficiencies calculated from the estimated frontier cost function. It also contains the results of a selected testing procedure for alternative production structures. In the third section, the estimates of the input efficiencies are presented.

V.1. ESTIMATION OF THE FRONTIER COST FUNCTION

The parameter estimates resulting from the MLE procedure are presented in Table 5. The gradient vector of the likelihood function with the respect to each parameter is equal to zero, as required for a maximum. This table shows that except for $\gamma$, and $(\alpha/2)$, the estimates are highly significant. The joint significance of the parameters in the frontier cost function is tested using a Likelihood Ratio Test (LRT) [Judge et alii (1980), Cramer (1986)]. It demonstrates that they are jointly significant at 1% confidence. These results indicate that the estimates are sound from an statistical point of view.
Table 5. Estimated Coefficients for the Frontier Cost Function*.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Variable</th>
<th>Estimates</th>
<th>Standard Error</th>
<th>Estimates/Std. Error</th>
<th>Probability (t-distr.)</th>
<th>Gradientb</th>
</tr>
</thead>
<tbody>
<tr>
<td>σ</td>
<td></td>
<td>0.5170</td>
<td>0.0152</td>
<td>34.0500</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>λ</td>
<td></td>
<td>1.0908</td>
<td>0.1133</td>
<td>9.6270</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>α₀</td>
<td>Constant</td>
<td>-0.4105</td>
<td>0.0262</td>
<td>-15.6560</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>α₁</td>
<td>ln y₁</td>
<td>0.7614</td>
<td>0.0242</td>
<td>31.4970</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>α₂</td>
<td>ln y₂</td>
<td>0.1841</td>
<td>0.0220</td>
<td>8.3540</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>α₃</td>
<td>ln y₃</td>
<td>0.0272</td>
<td>0.0122</td>
<td>2.2270</td>
<td>0.0130</td>
<td>0.0000</td>
</tr>
<tr>
<td>β₁</td>
<td>ln w₁</td>
<td>0.4829</td>
<td>0.0112</td>
<td>43.1540</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>(α₁₁)/2</td>
<td>ln (y₁)²</td>
<td>0.3186</td>
<td>0.0277</td>
<td>11.5030</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>α₁₂</td>
<td>ln y₁ ln y₂</td>
<td>-0.3293</td>
<td>0.0482</td>
<td>-6.8300</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>α₁₃</td>
<td>ln y₁ ln y₃</td>
<td>-0.3021</td>
<td>0.0254</td>
<td>-11.9110</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>γ₁₁</td>
<td>ln y₁ ln w₁</td>
<td>-0.1089</td>
<td>0.0231</td>
<td>-4.7100</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>(α₁₁)/2</td>
<td>ln (y₂)²</td>
<td>0.0699</td>
<td>0.0233</td>
<td>2.9990</td>
<td>0.0014</td>
<td>0.0000</td>
</tr>
<tr>
<td>α₂₃</td>
<td>ln y₂ ln y₃</td>
<td>0.2822</td>
<td>0.0251</td>
<td>11.2600</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>γ₂₁</td>
<td>ln y₂ ln w₁</td>
<td>0.0282</td>
<td>0.0200</td>
<td>1.4040</td>
<td>0.0801</td>
<td>-0.0000</td>
</tr>
<tr>
<td>(α₂₂)/2</td>
<td>ln (y₃)²</td>
<td>0.0032</td>
<td>0.0033</td>
<td>0.9490</td>
<td>0.1713</td>
<td>-0.0000</td>
</tr>
<tr>
<td>γ₃₁</td>
<td>ln y₃ ln w₁</td>
<td>0.1549</td>
<td>0.0137</td>
<td>11.3000</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td>(β₁₁)/2</td>
<td>ln (w₁)²</td>
<td>-0.0480</td>
<td>0.0063</td>
<td>-7.5690</td>
<td>0.0000</td>
<td>-0.0000</td>
</tr>
</tbody>
</table>

σ² =
σ₀² = 0.1453
Ln-lik Function (mean value) = -0.5485
No. cases = 3384
LRT for the cost function parametersc (15 d.f.) = 8463.98

* This model imposes homogeneity of degree one in input prices and symmetry conditions.
* This is the gradient vector of the Likelihood Function with respect of each parameter. Any interior solution for local maxima requires that all the elements in this vector be equal to zero.
* This test statistic has an asymptotic χ² distribution. The critical value (1% significance) is equal to 30.58 (15 d.f.).
The second validation check for the estimated cost function is related to the meaningfulness of the parameters estimates with respect to their theoretically expected signs. One would like to obtain results that do not violate a series of important conditions. For example, the parameter estimates should give a cost function that is nondecreasing in input prices, that reflects positive marginal costs, etc. In the Translog form, however, some of these conditions cannot be easily checked. In particular, this form gives conditions that depend on the point of evaluation. However, as a consequence of data normalization around the geometric mean and the logarithm form of the Translog function, it is easy to check many of these conditions at the mean point.

In the Translog case, the marginal cost with respect of each output is given by:

\[
\frac{C(.)}{\partial y_i}_{mean} = \frac{\partial \ln C(.)}{\partial \ln y_i}_{mean} = \left\{ \alpha_i + \sum_{j=1}^{k} \alpha_{ij} \ln y_j + \sum_{j=1}^{k} \gamma_{ij} \ln w_j \right\} \frac{C(.)}{y_i}_{mean} \tag{53}
\]

Of course, the evaluation of equation (53) at the mean values gives:

\[
\frac{\partial C(.)}{\partial y_i}_{mean} = \alpha_i \frac{C(.)}{y_i} \tag{54}
\]

Since \(C(.)\) and \(y_i\) are always positive, the sign of marginal costs will depend on the sign of parameters \(\alpha_i\) at the mean values. As expected in a proper cost function, Table 5 shows that the estimates of these parameters (the linear terms in outputs) are all positive and consistent with theoretical expectations.
A second condition implied by theory is that the cost function should be concave in input prices. A function is concave with respect to a set of variables if the matrix of the second derivatives (Hessian) is negative semi-definite [Chiang (1984)]. And, it is possible to prove that a matrix is negative definite if the determinants of its ordered principal minors have the sign equal to \((-1)^k\), where \(k\) is the order of the minor. At the geometric mean, one can check this condition by the signs of parameters \(\beta_j\)'s. This is a consequence of the fact that:

\[
\frac{\partial^2 C(.)}{\partial w_i^2} \bigg|_{\text{mean}} = -\beta_i \frac{C(.)}{w_i^2}
\]

Thus, it is not difficult to see that:

\[
\frac{\partial^2 C(.)}{\partial w_i^2} \bigg|_{\text{mean}} = -\beta_i \frac{C(.)}{w_i^2}
\] (56)

Since both, \(C(.)\) and \(w = [w_1, w_2, ..., w_n]\) are always positive, the sign of equation (56) depends on the sign of \(\beta_i\). In this study, the estimated \(\beta_i\) is equal to 0.4829 and \(\beta_2 = 1 - \beta_1\). Consequently, the condition above is satisfied for both inputs and the estimated cost function is indeed concave in \(w\). As a result, the estimated frontier cost function seems to conform to the expected form at least around the geometric mean of the data.

An important feature of the approach considered here is the postulation of the existence of a frontier function whose parametric representation is different from the average cost function. Several alternatives could be used to test for this condition [Schmidt and Lin (1984)]. In the present case, it is tested using a standard Likelihood Ratio Test (LRT) based on the restriction that the variance of the normal distribution underlying the one-sided error term is
equal to zero. Parametrically, this is represented by the condition that $\sigma^*_i = 0$. The estimated value for the LR statistic is equal to 17.7309. For one degree of freedom, the critical value for 1% confidence is 6.635 which rejects the null hypothesis that the estimated frontier function is statistically equal to the "mean" cost function. This result shows that the data support the frontier function approach used in the present study.

V.2. TEST RESULTS FOR SELECTED ALTERNATIVE PRODUCTION STRUCTURES

As mentioned in chapter 3, in general, the production technology cannot be directly observed in practice. This creates a series of problems when empirically estimating the cost function. The widespread use of a certain functional form implies certain parametric restrictions on the underlying technology being estimated. Unfortunately, even the popular flexible functional forms can only partially address this problem. Furthermore, data limitations and the requirements of enough degrees of freedom for the estimation procedures and sound statistical inferences usually imply an aggregation of variables. Of course, this can have serious implications for the estimations and one should be aware of the implications of the alternative chosen.

To mitigate some of these limitations, this study performed a series of tests for alternative production structures. They should give indications regarding the appropriateness of the choice made. Among others, tests for non-jointness, separability, homogeneity and Cobb-Douglas production structures were performed. Each of them is briefly discussed and their results are presented in what follows.

67 It is not difficult to see that this condition will reduce the likelihood functions of the composed error term (presented in chapter 3) to the standard likelihood function of the cost function fitted through the mean of the data.
i) **Non-jointness:** Jointness of production technology means that a set of outputs is jointly produced. As mentioned before, there are strong conceptual reasons that lead to the belief that this is the case in the banking industry. An attractive feature of the multi-output Translog cost function is its ability to test for this characteristic. This is a consequence of the fact that if the production technology is non-joint, then the marginal cost of a given output should not be affected by the level of any other output (non-jointly) produced. This condition can be tested from the estimated cost function for it is equivalent to:

\[
\frac{\partial^2 C(.)}{\partial y_i \partial y_j} = 0, \quad \forall \ i \neq j
\]  

(57)

Parametrically, this is the same as the condition that \( \alpha_{ij} = 0, \forall \ i \neq j \) and it can be easily tested using, e.g., the LRT which has an asymptotic \( \chi^2 \) distribution with degrees of freedom equal to the number of restrictions imposed. Table 6 shows that the calculated LRT statistic for non-jointness is equal to 149.91, with 3 d.f. This is much larger than the critical value of 11.34 (1% confidence) which indicates that non-jointness should be rejected in favor of the Translog form used in the present study.

ii) **Separability:** A production process is said to be input-output separable if a change in the input prices affects the optimal product mix. In the cost function, it implies that a change in the input prices does not affect the ratio of marginal costs. In terms of the cost function, separability implies that:

\[
\frac{\partial}{\partial \ln w} \left\{ \frac{\partial \ln C(.)/\partial \ln y_j}{\partial \ln C(.)/\partial \ln y_i} \right\} = 0
\]  

(58)
Again, it is not difficult to see that it can be easily tested using the cost function used here. The parametric restriction given by $\gamma_{ii} = 0, \forall \ i,j$ is equivalent to equation (58) and can be tested using the LRT. With three degrees of freedom, the calculated LRT statistic is equal to 165.10 [Table 6]. As a consequence, separability should be rejected in favor of the Translog form used in this study.

<table>
<thead>
<tr>
<th>Production Characteristic</th>
<th>Test Statistic*</th>
<th>Degrees of Freedom</th>
<th>Critical Value$^b$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-jointness</td>
<td>3 149.91</td>
<td>3</td>
<td>11.34</td>
</tr>
<tr>
<td>Separability</td>
<td>165.10</td>
<td>3</td>
<td>11.34</td>
</tr>
<tr>
<td>Homotheticity</td>
<td>289.57</td>
<td>4</td>
<td>13.28</td>
</tr>
<tr>
<td>Cobb-Douglas</td>
<td>811.10</td>
<td>10</td>
<td>23.21</td>
</tr>
</tbody>
</table>

* Likelihood Ratio Test.

$^b$ $\chi^2$ distribution at 1% significance.

iii) **Homotheticity**: A production technology is said to be homothetic if its expansion path is a straight line. Furthermore, a homothetic production technology is said to be homogeneous of degree $k$ if a t-fold increase in all inputs results in a output increase of $t^k$. This is reflected in the cost function in that the cost elasticity of each output is a constant.

Using the Translog form, the cost elasticity of output $y_i$ can be written as:
The parametric restriction equivalent to equation (59) is given by:

\[
\frac{\partial \ln C(.)}{\partial \ln y_i} = \text{constant}, \quad \forall \ i
\]  

(59)

Again, the LRT on these parametric restrictions reject an homogeneous production technology in favor of the functional form used in this study [Table 6].

iv) **Cobb-Douglas Production Technology**: It is known that the once popular and much used Cobb-Douglas function unfortunately imposes too restrictive parametric limitations on the technology being studied. However, it is very convenient due to its simplicity and ease for use in both theoretical analysis and empirical applications. Thus, a test statistic is performed on the Translog form originally proposed above for the Cobb-Douglas form. This form is obtained from the Translog function by setting all the parameters in the square and cross-product terms equal to zero. A LRT on the null hypothesis that the cost function has a Cobb-Douglas form is rejected.

These results give statistical assurance that the Translog functional form used in the present study is more appropriate than alternatives that include functional forms implying joint, separable and homothetic production technologies. Also, the use of a more simple Cobb-Douglas form is not warranted.
V.3. OUTPUT EFFICIENCIES

The methodology used in this study allows for the measurement of output efficiency using the estimated frontier cost function. As described in chapter IV, this study calculates three different measures of output efficiencies: Ray Scale Economy; Expansion Path Scale Economy; and, Scope Economy.

i) Ray Scale Economy \([S_M(y)]\)

The estimated measures of Ray Scale Economy at three different points in the output dimension are presented in Table 7. It is surprising to observe that the estimated results show a cost surface with significant convexity. The variation in the calculated \(S_M\), as the point of evaluation changes, is quite notable. For the median value of the first two quartiles, it shows significant economies of scale. This means that an increase of scale at this point results in a less than proportional expansion in costs. At the geometric mean of size, the \(S_M\) shows slight economy of scale. At the median bank size of the last two quartiles, the \(S_M\) indicates significant diseconomies of scale.

<table>
<thead>
<tr>
<th>Bank Size (y)</th>
<th>Ray Scale Economy ([S_M(y)])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median 1(^{*})</td>
<td>0.34656</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td>0.97270</td>
</tr>
<tr>
<td>Median 2(^{b})</td>
<td>1.74646</td>
</tr>
</tbody>
</table>

\(^{*}\) Median bank size for the first two quartiles.

\(^{b}\) Median bank size for the last two quartiles.

\(^{70}\) It is important to keep in mind, however, that the estimated values of output efficiencies are relative to the frontier cost function and not necessarily to any specific firms in the sample.
These results are consistent with an U-shaped "average" cost curve. However, they are somewhat different from the previous results found in the literature. The studies of the banking industry in the US show "average" cost curves with important scale economies for small banks, which is consistent with the results in this study. However, the banking sector in the US shows a flat "average" cost curves for larger banks with either slight economies or diseconomies of scale [Berger et alii (1987), Berger and Humprey (1992), Berger and Humprey (1992), Evanoff and Israilevich (1991), Mester (1994)]. This seems to be the case for financial institutions in developing countries as well [Cuevas (1984), Nauriyal (1993), Park (1993), Srinivasan (1988)].

In particular, the results obtained by Nauriyal (1993) in Chile are particularly relevant because they are related to the same financial institutions and the differences found may have important methodological and empirical consequences. Nauriyal's (1993) results show significant scale economies for the entire range of observation studied. This conflicts with the results obtained in the present study, indicating that there are major methodological differences between these studies. In particular, the mean function method used by Nauriyal cannot capture the input inefficiencies and the shape of the cost surface may be biased in this sense.

On the other hand, the diseconomies of scale found in this study for larger banks are not consistent with the descriptive results presented in Figure 8. Indeed, as expected by the empirical results above, the raw measure of average costs presented in this figure seems to indicate that there are significant economies of scale for small banks. As mentioned before, for the small banks on the frontier, this measure of average cost declines rapidly with the increase in size. However, for larger banks, this ratio becomes flat, which is inconsistent with the large diseconomies of scale found in Table 7. This may be indication of problems with the translog

71 These results seem to be consistent across different methodologies of estimation as well as across distinct data and variable definitions.
functional form [McAllister and McManus (1993)]. Indeed, in certain situations, the translog form is not sufficiently flexible and it may give cost surfaces that have the incorrect shape over some range of the data.

ii) Scope Economy \([S_c(y)]\)

The estimated \(S_c\) between the vector of minimum observed outputs and three different points in the output dimension is presented in Table 8. It shows significant scope economies for all three points of evaluation. The meaning of these findings is that there are strong cost complementarities among these three outputs in the financial industry in Chile. That is, banks offering multiple outputs present cost advantages vis-a-vis specialized firms. In fact, in the range calculated, the scope economies are not exhausted even for the larger banks. Actually, \(S_c(y)\) becomes larger with \(y\) indicating that the cost of specialized firms grow faster than the multi-output counterparts.

In the US, large scope economies were found for smaller banks [Berger et alii (1987)]. For larger banks, these authors found scope diseconomies. Unfortunately, comparative analysis of scope economies measures estimated in this study with other developing countries cannot be made because not many previous research in these countries report this measure.\(^7\) Specifically, Nauriyal (1993) did not present this measure which precludes the comparison of scope economies calculated using the frontier function with those resulting from mean functions.

---

\(^7\) In part, this is due to the early problems related to this measure in translog forms (see chapter III).
Table 8. Estimated Scope Economy for the Financial Institutions in Chile.

<table>
<thead>
<tr>
<th>Bank Size (y)</th>
<th>Scope Economy ( S_\delta (y) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median 1(^a)</td>
<td>2.05740</td>
</tr>
<tr>
<td>Geometric Mean</td>
<td>4.71152</td>
</tr>
<tr>
<td>Median 2(^b)</td>
<td>5.74352</td>
</tr>
</tbody>
</table>

\(^a\) Median bank size for the first two quartiles.
\(^b\) Median bank size for the last two quartiles.

iii) Expansion Path Scale Economy \( \text{EPSCE}(y^\delta, y^\circ) \)

As explained previously, \( \text{EPSCE}(y^\delta, y^\circ) \) is similar to \( S_\delta (y) \) with the exception that it is measured along the ray that connects points \( y^\delta \) and \( y^\circ \), instead of the ray passing through the origin. This measure was calculated for two points: at the geometric mean along the ray that connects it to the median valued for the two first quartiles; and, at the median output for the last two quartiles along the ray that connects it to the geometric mean. As Table 9 shows, there is a slight EPSCE resulting from the expansion of outputs from the median output of the first two quartiles to the geometric mean.

On the other hand, there are cost diseconomies at the median output of the last two quartiles when EPSCE is calculated along the ray that connects this point to the geometric mean. Of course, this indicates that larger firms are not able to operate with a cost advantage in comparison to those operating closer to the mean output vector.

The estimated EPSCE is close to the ones obtained by Berger et alii (1987) in the US. Unfortunately, the lack of previous results in the financial sector of developing countries precludes a comparative analysis with the estimates presented here.
Table 9. Estimated Expansion Path Scale Economy for the Financial Institutions in Chile.

<table>
<thead>
<tr>
<th>Bank Size (y^a)</th>
<th>Exp. Path S. Econ. [EPSCE(y^a, y^b)]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Geometric Mean*</td>
<td>0.93152</td>
</tr>
<tr>
<td>Median 2^b</td>
<td>1.26373</td>
</tr>
</tbody>
</table>

* Measured along the ray that connects this point to the median output vector for the first two quartiles.

^b Measured along the ray that connects this point to the geometric mean of the output vector.

V.4. INPUT EFFICIENCIES

As described previously, the method used in this study allows for the estimation of input efficiencies at each observation point. In fact, due to the reasons explained in Chapter 4, they were calculated for 1116 points instead of the 3384 points used in the estimation of the frontier cost function.

Before entering the discussion, a caveat is in order. Some of the estimated inefficiencies were shown to be larger than one. This was especially true for allocative efficiency (AE), where about 10% of the results were in this range. A handful of them were as high as 1.5. This problem was not as bad in the case of technical efficiency (TE) and it was negligible for the overall economic efficiency (EE). The suspicion is that the estimation procedure used in this study allows some of the observations to fall below the frontier cost function. This occurs when the effect of the two sided error term is negative and overcomes the effect of the one-sided inefficient term. In this case, the estimated composite error is negative. Strictly speaking, this is not conceivable from the theoretical point of view. However, it is perfectly acceptable in empirical applications when the modeling of the composed error term is taken into account.
This inconsistency was handled by setting their efficiency levels to one whenever the estimated values were larger than one. This seems to be appropriate because the source of this inconsistency is likely to be generated by those firms near of or on the frontier. Any measurement error in the proper direction, for example, would show that these firms are operating outside the implicit input requirement set.

V.4.1 Magnitude and Characterization of the Estimated EE, TE and AE

As expected, the empirical results show quite large input inefficiencies in the banking industry in Chile. Selected statistics for the estimated efficiency measures are presented in

Table 10. Selected Statistics for the Estimated Overall Economic Efficiency (EE), Allocative Efficiency (AE) and Technically Efficiency (TE).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>EE</th>
<th>AE</th>
<th>TE</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Cases</td>
<td>1116</td>
<td>1116</td>
<td>1116</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.127</td>
<td>0.192</td>
<td>0.204</td>
</tr>
<tr>
<td>Mean</td>
<td>0.737</td>
<td>0.886</td>
<td>0.838</td>
</tr>
<tr>
<td>Maximum</td>
<td>1.000</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Std.Deviat.</td>
<td>0.202</td>
<td>0.167</td>
<td>0.186</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.313</td>
<td>-1.349</td>
<td>-1.066</td>
</tr>
<tr>
<td>Median</td>
<td>0.733</td>
<td>1.000</td>
<td>0.904</td>
</tr>
</tbody>
</table>

Table 10. On average, the financial institutions in the sample show an input inefficiency of about 0.74. This means that these firms are incurring a cost of production that is above the
frontier cost function by about 35%. Alternatively, it is possible to conclude that if these firms were 100% technically and allocatively efficient, the same output quantities could be produced with 35% less input costs. Compared with figures previously reported in the literature, this result is about 10% higher than the ones found for the US financial market where the estimated EE range was in the interval around 20 ~ 25%.

It is also interesting to notice that, as expected, the distribution of the inefficiencies are skewed to the left. This can be confirmed by the negative values for the measure of skewness and also by the values of the mean relative to the value of the median. A more detailed picture of their distributions can be seen in Figure 9. This figure presents the Box and Whiskers plot of the estimated EE, TE and AE.

---

73 Park (1993) studying the Korean agricultural credit cooperatives found an average inefficiency of only about 4%. This is a surprisingly small number. However, as mentioned previously, one has to keep in mind that the frontier being estimated is related to the firms included in sample and his study was limited to a very narrow part of the financial market. Methodological differences may also explain some of the differences as well.

74 For a better interpretation of this graph, see e.g., Systat for Windows (1992), pp. 188-203.
An alternative presentation of a distribution is the histogram. It gives further information not provided by the Box and Whisker plots. Figure 10, Figure 11 and Figure 12 present, respectively, the individual histograms for the estimated EE, TE and AE. These figures also contain the smoothing normal distribution underlying each of them.\textsuperscript{75} Again, the skewness to the left of the estimated efficiencies is clearly seen in these figures. The truncation nature of them is also visible in the sense that many observations are accumulated at 1.0.

\textsuperscript{75} The smoothing normal curve is plotted using the mean and standard deviation calculated from the estimates.
Figure 10. Histogram of the Estimated Overall Economic Efficiency (EE) and the Plot of its Corresponding (smoothed) Normal Distribution.
Figure 11. Histogram of the Estimated Technical Efficiency (TE) and the Plot of its Corresponding (smoothed) Normal Distribution.
Consistent with the results reported in the literature, the estimated TE are larger than the AE: 19.3% against 12.9%. However, the differences are less striking in the case of Chile than the ones reported for banks in the US [Evanoff and Israilevich (1991)]. Indeed, most of the empirical studies conducted in the American financial markets show that the estimated TE are typically several fold larger than the AE. Of course, the meaning of these findings is that the managers of banks in the US are better able to use the inputs in their correct proportions. However, once the level of input quantities are determined, they seem to be unable to use them effectively.

Again, in Chile, these results imply that the managers of financial institution seem to be relatively more successful in using the correct input proportions than in using the available inputs productively. However, the estimated AE show that, on average, the incorrect input
proportion increases the actual operating costs by about 13%. Furthermore, the median value of the estimated AE is equal to one. This has an important implication for the firms that are allocatively inefficient because in more than 50% of the sample points the financial institutions in Chile were using the inputs in their correct ratio. Thus, the extra costs resulting from AE burdens less than half of the firms. Finally, it is interesting to notice that the average AE in Chile is within the upper limit of the reported AE for the banks operating in US market which range in the 1.0 ~ 17.1% interval [Evanoff and Israilevich (1991)].

On the other hand, the estimated TE in the present study are in the mid range of estimates found in US. Thus, in spite of the fact that the average TE reflects in an additional cost of about 19%, the managers of financial institutions in Chile are as successful as their US counterparts in using the inputs available to them. Also, it is interesting to notice that the characteristics of the distribution for the estimated TE is slightly different than found for AE. Both the average inefficiency from TE and its variance are larger than for AE, and the skewness and the median are smaller. These results mean that the distribution of the estimated TE is more spread among the firms in the study than the AE. Furthermore, a comparison of Figure 11 and Figure 12 shows that not as many firms were able to be 100% TE efficient. In fact, about 60% of the firms have AE equal to one compared to about 45% in the case of TE.

An even more disaggregate picture of the distribution of the estimated EE, TE and AE is presented in the following three figures (Figure 13, Figure 14 and Figure 15). They show the average inefficiencies for each bank in the study. It is interesting to notice that the three measure of efficiencies do not present a high correlation. That is, a bank with a high TE does not necessarily present a high AE and vice-versa. In fact, these two measures have a slight

---

76 The complete list of the full name of all banks included in this study is presented in the Appendix.
negative correlation coefficient of -0.178. On the other hand, as expected, EE is positively correlated with TE and AE.

Figure 13. Average EE for each Financial Institution in the Study.
Figure 14. Average TE for each Financial Institution in the Study.
V.4.2. Factors Affecting EE

Once the inefficiencies are estimated and they are found to be significant, it is important to try to identify the factors that explain them. As mentioned before, if markets were perfect, inefficient firms would not last. They would be driven out of the market by the efficient ones. However, the evidence previously presented in the literature that technical and allocative inefficiencies exist in the financial sector is confirmed in the empirical results above. Indeed, these results support the idea that both types of inefficiencies seem to be quite significant in the Chilean financial system.

For example, the Box and Whiskers plot of EE by type and nationality of financial institutions presented in Figure 15, show that there are non-trivial differences among them.
The banks included in the category of Other foreign banks (OTHER) are shown to be the most efficient. Their mean and median overall inefficiencies were estimated to be 0.826 and 0.867, respectively. They are closely followed by the US banks (USA) which have an average EE of 0.791 and a median EE of 0.826. Next come the Domestic banks (D) with 0.713 and 0.709, respectively. On average, the Foreign Latin American banks presented an even lower EE with mean of 0.699 and median of 0.635. Finally, the Savings and Loans (FIN) are the least efficient with a mean of 0.576 and a median of 0.592 (Table 11).

<table>
<thead>
<tr>
<th>Type of F.I.*</th>
<th>Cases</th>
<th>Min</th>
<th>Average</th>
<th>Median</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>434</td>
<td>0.147</td>
<td>0.713</td>
<td>0.709</td>
<td>1.000</td>
</tr>
<tr>
<td>LAT</td>
<td>124</td>
<td>0.384</td>
<td>0.699</td>
<td>0.635</td>
<td>1.000</td>
</tr>
<tr>
<td>OTHER</td>
<td>217</td>
<td>0.127</td>
<td>0.826</td>
<td>0.867</td>
<td>1.000</td>
</tr>
<tr>
<td>USA</td>
<td>217</td>
<td>0.249</td>
<td>0.791</td>
<td>0.826</td>
<td>1.000</td>
</tr>
<tr>
<td>FIN</td>
<td>124</td>
<td>0.334</td>
<td>0.610</td>
<td>0.576</td>
<td>1.000</td>
</tr>
</tbody>
</table>

* D = domestic banks; LAT = foreign Latin American banks; OTHER = foreign banks other than from Latin-America and US; USA = foreign banks from US; and, FIN = savings and loans (financieras).
These results are a clear indication that there are significant market imperfections in the Chilean financial markets. The savings and loans represented in this study, for example, are able to survive with an average cost that is almost twice the other banks in the study. Maybe this fact explains part of the bankruptcies that were witnessed in the late 70's and early 80's in Chile.

In an effort to better understand the possible sources of these inefficiencies, the estimated EE for each bank is regressed against a set of variables that could possibly affect this outcome. The explanatory variables chosen were: trend, type/nationality of financial institution, the volume of operation, the number of bank employees and the number of bank branches. The five classes of financial institutions were represented by four binary (dummy)
variables. The volume of operations was represented by the sum of total deposits, total loans and total investments (SDLI). EE, SDLI, employees (EMPLOY) and number of branches (BRCH) were normalized by their respective means. The estimated results are presented in Table 12.

Table 12. Estimation Results for the Sources of Overall Input Inefficiency (EE).

| Variable | Estimate | Standard Error | t-value | Prob > |t| |
|----------|----------|----------------|---------|---------|---------|
| CONSTANT | 1.089491 | 0.032284       | 33.746723 | 0.000   |
| TREND    | 0.003106 | 0.001274       | 2.438058  | 0.015   |
| LATIN    | -0.181534| 0.042205       | -4.301257 | 0.000   |
| US       | 0.207965 | 0.035272       | 5.896079  | 0.000   |
| OTHER    | 0.222182 | 0.034897       | 6.366832  | 0.000   |
| FIN      | -0.498173| 0.042338       | -11.766435| 0.000   |
| SDLI     | 0.140583 | 0.009400       | 14.955061 | 0.000   |
| EMPLOY   | -0.107118| 0.015194       | -7.050022 | 0.000   |
| BRCH     | -0.034639| 0.012032       | -2.879041 | 0.004   |

Depend. var. : EE
Valid cases : 1116
Total SS : 272.378
R-squared : 0.415
Residual SS : 159.454
F(8,1107) : 97.996 Probab. of F : 0.000

The dummy variable for the domestic banks was dropped from the analysis to avoid singularity in the matrix of explanatory variables. Obviously, the interpretation of the estimation results for the remaining financial institutions should be relative to the inefficiency of these banks.
Interestingly, except for the variable representing trend, all the estimated parameters for the remaining variables are highly significant. Even the estimate for trend is significant at 5%. The F-test shows that the estimated coefficients are jointly significant, which gives some assurance for the results as well. Furthermore, the calculated R-square is relatively high for panel data indicating that these results have significant explanatory power. This is an unusual outcome if the previous results in the literature are taken into account. In fact, Aly et alii (1990) and Park (1993) were relatively unsuccessful in finding strong explanatory factors for the inefficiencies estimated in their research. For example, they found R-squares of only 0.06 and 0.085, respectively.  

The positive coefficient for the trend variable indicates that the level of overall economic efficiency has been slowly improving in the Chilean financial markets. However, it is clear that these improvements have been obtained at a sluggish pace. Furthermore, if one considers that the average overall economic inefficiency level is relatively high, it should be clear that the financial reforms of the late 70’s and early 80’s failed to establish a free and competitive environment in this market.

As expected from the results presented in Table 11, the estimated parameters for the dummy variables indicate that there are significant differences in the efficiencies among the different financial institutions. In fact, the relative EE presented in these two tables are perfectly compatible. This is strong empirical evidence showing that the management of these financial institutions varies in its ability to be successful in administering the banks. Of course, access to different technologies may play some role in generating these results but there was no feasible way to capture that effect in this analysis.

---

78 It should be mentioned, however, that both used methodologies to estimate the inefficiencies and explanatory variables that are different from the ones considered in the present study.
An interesting result is the positive coefficient estimate for bank size (SDLI). It indicates that the larger financial institutions present lower degrees of inefficiency. That is, they are able to operate closer to the cost frontier than the smaller firms.

However, as one would expect, the negative coefficient estimates for EMPLOY and BRCH indicates that financial institutions with a larger number of employees and branches are less efficient. This is strong evidence of the difficulties encountered in managing larger firms.
CHAPTER V

CONCLUSIONS

In the past 30 years, the Chilean economy and its financial markets in particular have gone through a series of adjustment processes. Sharp changes in domestic policies as well as the consequences of fluctuations in the international markets have hit the economy hard. Chile went through a difficult and painful adjustment process that resulted from the last round of economic and financial crises it suffered in the late 70's and early 80's.

As a consequence, Chile entered a process of significant deregulation of its economy and moved toward a freer market structure. The liberalization process included a trend toward the removal of price control mechanisms and incentives to compete. Old policies of strict labor and financial market regulations, interest rate and exchange rate controls, etc. were gradually eliminated. And, it has opened its economy to the international markets. The idea behind these changes was that the deregulation of the commodity and factor markets with the removal of barriers to trade and capital flows would allow improvements in resource allocation, eliminate bottlenecks in the economy and generate higher growth rates.

The financial sector was subjected to significant governmental intervention directed toward cleaning-up and liquidating troubled institutions. Also, the Central Bank of Chile opened the country's financial market to foreign institutions.
In the last few years, Chile has experienced a significant "boom" in its economy and in its foreign trade. The positive results obtained from these economic and political change, and the favorable environment in the international markets make Chile an example to be followed by other developing countries.

In the financial sector, after many years of instability and government regulations and interventions, the remaining institutions seem to have recovered from the crisis and have resumed a process of expansion. The old regulated environment not only caused significant financial repression in Chile but also made inefficient firms viable. More seriously, the distortions induced by these regulations made these financial institutions extremely risky. The new deregulated environment was expected to bring competition and efficiency into the market. It was expected that the financial institutions would not only try to take advantage of economies of scale and scope with rearrangements in their portfolios of assets and liabilities but also reduce waste in factor use.

V.1 SUMMARY AND POLICY IMPLICATION

To empirically analyze some of these questions, this study estimated a cost function using the frontier function method. On the one hand, this method takes advantage of the dual relationship between the cost and production technology to capture its underlying characteristics and study output efficiency. On the other hand, the frontier nature of the cost functions is amenable to the study of input efficiency as well.

This study has four major objectives. First to estimate a frontier cost function for a set of financial institutions operating in Chile using the MLE method. Second, estimate the output efficiencies in the banking industry in Chile. Third, estimate their input efficiencies. And, fourth, explain the major sources associated with input efficiencies.
To accomplish this task the economic behavior of the financial institutions was modeled through a multi-output and flexible Translog frontier function. Alternative composed error structures are tested.

The results show that the estimated output efficiencies are consistent with previous studies found in the banking literature. However, the cost surface seems to present interesting differences. In general, it does not seem to be as flat as previously reported and the estimated ray scale economies change considerably with output levels. This result is significantly different from previous results reported in the literature. Also, as expected, there seems to exist significant cost complementarity among the outputs considered.

In terms of input efficiencies, it was found that the financial institutions in the study incur a considerable degree of extra cost due to either pure waste of factors of production, represented by their inability to use them in the most productive way, or the use of factors in incorrect proportions, or both. The estimated values seem to vary considerably among different financial institutions, and their distributions are consistent with theory. This result raises doubts about the competitiveness of the Chilean financial markets. If the markets were competitive, one would not expect to find such a large degree of input inefficiency. In particular, it is interesting to see that the reform policies implemented by the Central Bank of Chile in the late 70's and early 80's did not make the market completely efficient.

A regression using the estimated overall input inefficiency (EE) as the dependent variable was estimated and it confirmed that a series of factors are associated with input inefficiencies. Trend, nationality and type of the financial institutions, scale of operation and quantity of input use all showed statistical explanatory significance. This regression showed that the financial institutions have slowly improved during the period of analysis, and the effects of the financial deregulation mentioned above did not make a significant difference as
far as input efficiencies are concerned. It is striking that some firms with an average EE of less than 0.5 can still survive in the market. Of course, this is an interesting result from a policy perspective as well as for the bank managers, especially for the inefficient banks.

For the regulators and policy makers, this research suggests that there is significant room for improvements in bank efficiency. More importantly, the competitive pressure in the Chilean financial markets is not that strong and some imperfections permit the survival of inefficient firms.

For the management of these inefficient financial institutions, these results carry crucial information. First, these institutions are not making as much profit as they could if they were operating closer to the frontier function. A series of consequences can be inferred from this fact. For example, they may become prime targets for takeovers. Alternatively, their position within the institution is in jeopardy. Once this fact becomes known, the owners may want to change bank administration. Second, further changes in the regulatory or environmental conditions toward a more competitive market could threaten their very survival. Unless they move toward the frontier, they would become unprofitable.

The differences among the types of financial institutions and their nationalities are also significant. The foreign banks from countries other than Latin America are able to operate closer to the cost frontier than the domestic banks, savings and loans, and the banks from other Latin American countries. This may be a consequence of the fact that these banks bring better trained managers from their own headquarters and/or use superior technology.

The positive coefficient for the scale of operation (variable SDLI) indicates that financial institutions with larger portfolios are able to operate more efficiently. This supports the argument that a larger and more diversified portfolio can increase the operational efficiency
of financial institutions and reduce their costs as discussed in Chapter 3. This may indicate that in the future greater concentration will occur in the Chilean financial markets.

On the other hand, growth in bank size measured by input use seems to have the expected negative effect on the level of input inefficiency. It is certainly more difficult to take correct decisions, and achieve higher productivity in larger firms. This seems to be indicated by the negative coefficients for the variables representing number of employees and the number of branches.

**V.2. CONTRIBUTION OF THE STUDY, ITS LIMITATIONS AND SUGGESTIONS FOR FUTURE RESEARCH**

This study adds to the literature concerning the efficiency analysis of financial institutions and is part of a small body of research devoted to this topic in the financial sector of developing countries. It used the stochastic and parametric method to estimate a frontier cost function using alternative error structures. This method is amenable to both analyzing output efficiencies using the frontier function and to studying the firms' deviations from the cost frontier as input inefficiencies. These cost inefficiencies can be desaggregated into their technical and allocative components.

Among other limitations it must be recalled that the data used in this analysis were generated from accounting information. Especially in the early part of the period studied, the Central Bank intervened several times in the market and adjusted some accounting statements of many institutions. On the other hand, the data used represent a large part of the financial institutions operating in Chile during the period. Thus, the results can be considered to be a good representation of the Chilean financial markets.
The problem related to input aggregation may be non-trivial. As explained in Chapter 4, the variable representing capital inputs is a sum of expenses for all inputs except labor and, as a consequence, its price has to be imputed rather than observed. This may generate some bias. In future studies, as more data becomes available, a more disaggregated input vector could correct this problem and improve the analysis.

Similar aggregation procedures were adopted in the output vector where the total value of deposits, loans and investments were considered. In fact, the definition of output in this industry is still an ongoing debate and it does not seem to be solvable in the near future.

This study was limited to the application of one method of efficiency analysis, namely the stochastic and parametric form. It would be interesting to apply other methods to the Chilean banking sector and comparatively analyze their results. Other alternative frontier functions like the profit function could be tested as well. However, this approach requires profit data from the financial institutions which are not available in the data used here.
APPENDIX

Identification and Performance Indicators of Chilean Banks
Table 13. Domestic Commercial Banks and Savings and Loans Included in the Study, Chile.

<table>
<thead>
<tr>
<th>CODE</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>AEDWARDS</td>
<td>Banco de A. Edwards</td>
</tr>
<tr>
<td>BHIF</td>
<td>Banco Bhif</td>
</tr>
<tr>
<td>BICE</td>
<td>Banco Bice</td>
</tr>
<tr>
<td>CHILE</td>
<td>Banco de Chile</td>
</tr>
<tr>
<td>CONCEP</td>
<td>Banco Concepcion</td>
</tr>
<tr>
<td>CREDITO</td>
<td>Banco de Credito e Inversiones</td>
</tr>
<tr>
<td>DESSAR</td>
<td>Banco del Desarrollo</td>
</tr>
<tr>
<td>ESTADO</td>
<td>Banco del Estado de Chile</td>
</tr>
<tr>
<td>INTL</td>
<td>Banco Internacional</td>
</tr>
<tr>
<td>OHIO</td>
<td>Banco O’Higgins</td>
</tr>
<tr>
<td>OSORNO</td>
<td>Banco Osorno</td>
</tr>
<tr>
<td>PACIFIC</td>
<td>Banco del Pacifico</td>
</tr>
<tr>
<td>SANTIAGO</td>
<td>Banco de Santiago</td>
</tr>
<tr>
<td>SUD</td>
<td>Banco Sud Americano</td>
</tr>
</tbody>
</table>

- **COMMERCIAL BANKS**

- **SAVINGS AND LOANS (Financieras)**

<table>
<thead>
<tr>
<th>CODE</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATLAS</td>
<td>Corporacion Financiera Atlas S.A.</td>
</tr>
<tr>
<td>COMTAN</td>
<td>Financiera Comercial</td>
</tr>
<tr>
<td>CONDELL</td>
<td>Financiera Condell S.A.</td>
</tr>
<tr>
<td>FUSA</td>
<td>Financiera Fusa S.A.</td>
</tr>
</tbody>
</table>

* Rechartered as a bank (ABN Tanner Bank) in April of 1990. Because it remained as a savings and loans for the most of the period in consideration, it is included in this group.
Table 14. Foreign Commercial Banks Included in the Study, Chile.

<table>
<thead>
<tr>
<th>CODE</th>
<th>NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Latin American</td>
<td></td>
</tr>
<tr>
<td>ARGENT</td>
<td>Banco de La Nacion Argentina</td>
</tr>
<tr>
<td>BRASIL</td>
<td>Banco do Brasil</td>
</tr>
<tr>
<td>REAL</td>
<td>Banco Real</td>
</tr>
<tr>
<td>SAO</td>
<td>Banco do Estado de Sao Paulo</td>
</tr>
<tr>
<td>• Countries other than Latin American and the US</td>
<td></td>
</tr>
<tr>
<td>BOT</td>
<td>The Bank of Tokyo Ltd.</td>
</tr>
<tr>
<td>CENTRO</td>
<td>Centrobanco</td>
</tr>
<tr>
<td>CONT</td>
<td>Banco Continental</td>
</tr>
<tr>
<td>ESPSAN</td>
<td>Banco Santander Chile *</td>
</tr>
<tr>
<td>EXTER</td>
<td>Banco Exterior S.A.</td>
</tr>
<tr>
<td>HNGKNG</td>
<td>The Hongkong and Shangai Banking Corporation</td>
</tr>
<tr>
<td>SUDAMERI</td>
<td>Banco Sudameris</td>
</tr>
<tr>
<td>• US</td>
<td></td>
</tr>
<tr>
<td>AMEX</td>
<td>American Express Bank Ltd.</td>
</tr>
<tr>
<td>BOA</td>
<td>Bank of America</td>
</tr>
<tr>
<td>BOSTON</td>
<td>The First National Bank of Boston</td>
</tr>
<tr>
<td>CHASE</td>
<td>The Chase Manhattan Bank S.A.</td>
</tr>
<tr>
<td>CHICAGO</td>
<td>Chicago Continental Bank</td>
</tr>
<tr>
<td>CITI</td>
<td>Citibank N.A.</td>
</tr>
<tr>
<td>NEWYORK</td>
<td>Republic National Bank of New York</td>
</tr>
</tbody>
</table>

* Changed its name to Banco Espanol Chile in November of 1989.
Figure 17. Distribution of the Ratio of SUMDLI to the Number of Branches, Chile.
Figure 18. Distribution of the Ratio of SUMDLI to the Number of Employees, Chile.
Figure 19. Distribution of the Number of Employees per Branch, Chile.
REFERENCES


Boisvert, R.N., The Translog Production Function: its properties, its several interpretations and estimation problems. A.E.Res. 82-28, Department of Agricultural Economics, Cornell University Agricultural Experiment Station, Cornell University, Ithaca - NY, September/1982.


Corbo, V. "Reforms and Macroeconomic Adjustments in Chile during 1974-84". World Development 13(8), 1985.


