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  800/521-0600
RICE UNIVERSITY

Essays on Semiparametric Estimation and Structural Modeling with Applications in the Banking Industry

by

Robert Matthew Adams

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree

Doctor of Philosophy

Approved, Thesis Committee:

Robin C. Sickles, Chairman
Professor of Economics

Bryan W. Brown
Hargrove Professor of Economics

George Karatas
Jesse H. Jones School Professor of Management

Houston, Texas

April, 1997
Abstract

Essays on Semiparametric Estimation and Structural Modeling with Applications in the Banking Industry

by

Robert Matthew Adams

New semiparametric panel estimation methods have been developed, which make minimal distributional or functional form assumptions on the model. These estimators are illustrated in an efficiency analysis of the banking industry. This analysis finds that functional form and distributional assumptions are important in efficiency estimation. Moreover, models of time varying efficiency are relevant and indicate productivity movements in the banking industry.
Acknowledgments

First and foremost, I would like to thank my advisor, Dr. Robin Sickles, for his patient and insightful direction throughout my graduate studies at Rice. Robin has been instrumental to my acquiring the necessary skills for my future work. I would also like to thank Dr. Bryan W. Brown and Dr. George Kana\text\texttext{a}s for their assistance and comments. They too were important to my education at Rice.

My family played an important role in the completion of my dissertation. My siblings, Gwyn, Hope, and David, provided excellent conversation and gave insight into my work. More importantly, my parents, Bob and Carol, played a vital role in their support and encouragement.

My fiance, Jill Koncaba, has helped a great deal and has given me the support during the vital final year. Her help has been invaluable.

Moreover, I would like to thank Allen Berger for the dataset and helpful comments. Ila Semenick and Joya Mukherjee were also extremely helpful. Furthermore, I would like to thank my other fellow students. Specifically, Anthony Postert, Patrik Hultberg, Sly Gaudin, and Ellen Read.

Finally, I would like to thank the secretaries, Patsie, Liz, and Pattie for their service and the economics faculty at Rice for providing me with great education.
Contents

Abstract ii
Acknowledgments iii
List of Tables vi
List of Illustrations vii
Preface viii

1 Data 1

2 Semiparametric Approaches to Stochastic Panel Models 5
  2.1 The Model ............................................. 9
  2.2 Efficient Estimation .................................. 11
  2.3 Nonparametric Regression Estimation ............... 17
    2.3.1 Estimation of Conditional Expectations .......... 18
  2.4 Results ................................................. 19
    2.4.1 Semiparametric Efficient Estimation ............ 21
    2.4.2 Nonparametric Regression-Based Estimation .... 26
  2.5 Conclusion ............................................. 27

3 Computation in Efficient Estimation 30
  3.1 The Model ............................................. 32
  3.2 Efficient Estimation .................................. 33
  3.3 Computation and Smoothing ............................ 36
  3.4 Results ................................................. 40
  3.5 Conclusion ............................................. 44
4 Efficient Distribution Free Estimation of Panel Models  45
  4.1 Model and Statistical Assumptions .......................... 46
  4.2 Efficient Estimation ........................................... 47
    4.2.1 A Semiparametric Efficient Estimator of a Panel Model .... 47
    4.2.2 A Hausman-Taylor Estimator .............................. 50
  4.3 Results ......................................................... 51
  4.4 Appendix A ..................................................... 53
    4.4.1 A Semiparametric Efficient Estimator of a Panel Model ... 53
    4.4.2 A Hausman-Taylor Estimator .............................. 56

5 Convergence of Banking Efficiency ............................. 59
  5.1 An Econometric Motivation for Dynamic Efficiencies ......... 61
  5.2 Data Envelopment Analysis ................................... 64
  5.3 A Semiparametric Model ....................................... 66
    5.3.1 The Model ................................................ 66
    5.3.2 Semiparametric Estimation ............................... 68
  5.4 Battese and Coelli Estimator ................................ 69
  5.5 Results ......................................................... 70
Tables

1.1 Bank Size Description ........................................... 3
1.2 General Data Description ....................................... 4
1.3 Call Report Information ........................................ 4

2.1 Unit Bank Regressions .......................................... 22
2.2 Limit Bank Regressions ......................................... 22
2.3 State Bank Regressions ......................................... 22
2.4 Average Absolute Efficiencies for Unit, State and Limit Banks .... 24
2.5 Average Relative Efficiencies for Unit, State and Limit Banks .... 24
2.6 Nonparametric Regression Based Estimation .................... 26

3.1 Unit Banks Regressions ........................................ 41
3.2 Limit Banks Regressions ....................................... 42
3.3 State Banks Regressions ....................................... 42
3.4 Average Efficiencies for Unit, State and Limit Banks .......... 42

4.1 State Bank GSPE ................................................ 53
4.2 Limit Bank GSPE ................................................ 54
4.3 Unit Bank GSPE ................................................ 54

5.1 Unit Banks ...................................................... 73
5.2 State Banks ..................................................... 74
5.3 Limit Banks ..................................................... 74
Illustrations

2.1 Absolute Efficiencies in Semiparametric Efficient Regression  . . . . . . 25
2.2 Relative Efficiencies in Semiparametric Efficient Regression  . . . . . . 25
2.3 Absolute Efficiencies in Nonparametric Based Regression  . . . . . . 28
2.4 Relative Efficiencies in Nonparametric Based Regression  . . . . . . 28

3.1 Absolute Efficiencies in Semiparametric Efficient Regression  . . . . . . 43
3.2 Relative Efficiencies in Semiparametric Efficient Regression  . . . . . . 43

5.1 Unit Bank Annual Mean Relative Efficiency  . . . . . . . . . . . . . . 75
5.2 State Bank Annual Mean Relative Efficiency  . . . . . . . . . . . . . . 76
5.3 Limit Bank Annual Mean Relative Efficiency  . . . . . . . . . . . . . . 77
Preface

The topics of productivity and efficiency are well established in the economics literature. Productivity and efficiency are important theoretical tools that can be used to describe an industry or compare the firms within an industry. Both notions help researchers better understand the relationships within an industry. This information can be used to assess government policy, competition, and firm production. This dissertation concentrates on the notion of efficiency as a method of describing differences in firm production. Koopmans (1951) provided a formal definition of technical efficiency, where a technically efficient producer can increase output if only if another output is reduced or at least one input is increased. Debreu (1951) and Farrell (1957) both introduced the notion of a radial measure of technical efficiency, where efficiency is measured by the distance to the frontier along a ray from the origin.\footnote{See Lovell (1993).} Today, a rich and voluminous literature has been developed which has extended the notion of efficiency and how it applies to numerous industries. Consequently, extensive advancements have been made in the theory and in the measurement of efficiency. One particular example is duality theory. Duality theorems (Shepherd, 1971; Diewert, 1982; Färe and Primont, 1995) have opened the door for a variety of methods for modeling firm technology by showing the equivalence of the cost, profit and distance functions. These results are important because they give the researcher the option to use different methods of modeling firm technologies. Availability and quality of data often play a significant role in determining which model to use. For example, the cost function requires input price and output data (as well as costs), while the distance function only requires input and output data. By the duality theorems, both methods can capture technical efficiency.
The proper measurement of efficiency is important as successful policies of firms and governments depend on it. The main goal of this dissertation is thus to highlight potential problems of efficiency measurement and provide possible solutions.

Efficiency measurement goes back to von Neuman (1938, 1945), but it was not until the introduction of maximum likelihood estimators for cross sections by Aigner, Lovell, and Schmidt (1977) and Meeusen and Van den Broeck (1977) and linear programming methods such as data envelopment analysis by Charnes, Cooper, and Rhodes (1978) that efficiency measurement has gained in popularity. In the last two decades, numerous methods of efficiency measurement have been introduced, each method adding a little more to the description firm heterogeneities. These approaches follow two general lines of methodology: linear programming and econometric methods. Essentially, two major differences in these methodologies exist. First, linear programming methods are nonstochastic and do not allow for random error. Linear programming technical efficiency measures thus include both actual technical efficiency and random error. Econometric methods, on the other hand, are stochastic and, hence, do allow for random error, so that technical efficiency measures can be separated out from the random error. Second, linear programming methods are nonparametric, where the frontier is determined by creating a piecewised linear hull around the data. Econometric approaches are parametric and assume a functional form for the frontier.

Now, the fundamental question is whether technical efficiency really exists or are there other unaccounted explanations for observed firm heterogeneities. Stigler (1976) notes that efficiency is an empirical construct necessitated by the fact that researchers

---

2Aigner and Chu (1968) introduced a deterministic method of measuring efficiency as well.

3An example includes Cornwell, Schmidt, and Sickles' (1990) introduction of an estimator that allows for time varying efficiencies.

4The econometric approaches measure an 'average frontier', where single observations can lie below (above) the observed frontier because of random error. This is not true for the linear programming methods, where the data is enveloped by the frontier.
do not have complete information on the firms. Researchers could be using an incorrect objective function or not possess all relevant variables. Thus, some important variables such as management quality, output and input characteristics, and firm environment, may be omitted. If a researcher had complete information on the firm, then the observed technical efficiency would not exist. Schmidt (1985) describes efficiency as the amount that some firms produce more than others even after we account for the effects of observable variables. The difference in output is called “efficiency”. Both Stigler and Schmidt indicate the crux of the efficiency measurement problem. The efficiency argument becomes a matter of semantics, because firm differences are captured by a residual, the content of which is not necessarily obvious. Is the observed technical efficiency actually caused by a firm’s inability to produce at the frontier or is the firm producing efficiently given its choices of inputs. Schmidt’s description of efficiency is practical because it allows researchers to identify firm heterogeneities without specific information. These residuals could indicate differences in production caused by unobservable variables such as management quality, market power, and firm environment. However, the matter of efficiency measurement through residuals becomes problematic, if the researcher believes that an important variable such as an input is omitted. In this case, the firm is in fact using another input which would increase production and observed technical efficiency would overstate the true technical efficiency.

Since efficiency measurement centers on the firm specific residual and other possible explanations for these residuals exist, new methods of measurement and new models are needed. This dissertation concentrates on one possible shortcoming of efficiency measurement, model misspecification. Model misspecification is a particular problem of the econometric approaches, because functional form and distributional

---

5 Efficiency is modeled as a Solow-based residual. All omitted variables are captured by this residual.
6 The notion of efficiency as an indication of market power goes back to Hick’s “quiet life” of a monopoly.
assumptions are necessary to estimate firm technology.\textsuperscript{7} I develop semiparametric methods that make minimal assumptions on a portion of the model in an effort to mitigate model misspecification, while the theoretical interpretation of the model remains intact. Similar to the linear programming approaches, these semiparametric methods structuralize the model in order to allow the data to determine relationships and functional form.

I use an output distance function as a method of modeling productivity for a multioutput firm as opposed to the "traditional" cost minimization or profit maximization problem typically applied in the literature.\textsuperscript{8} The output distance function is a method of modeling firm technology for a multioutput firm which does not require price data.\textsuperscript{9} Semiparametric models in a panel data setting are developed for an output distance function, where minimal assumptions are made either on the distribution of the efficiencies or on the functional form of the output distance function.

In this dissertation, I measure efficiency in the banking industry. The banking industry is of particular interest because it has undergone substantial changes in the last two decades. These changes include deregulation of deposits, introduction of new technology and the movement by banks into off balance sheet activities. An economic analysis of banking efficiency will shed light on the overall effect of these changes in the industry. However, it is important to note that these changes effect the productivity in the banking industry, which is not included in the scope of this dissertation, but can be easily measured using similar methods.

The chapters are organized in the following fashion. Chapter 1 describes the data. The dataset is a balanced panel of about 3000 banks and represents an excellent op-

\textsuperscript{7}The linear programming methods are nonparametric and nonstochastic. Hence, model misspecification does not pose a problem.

\textsuperscript{8}This is especially true for the banking efficiency literature, where cost function estimation dominates.

\textsuperscript{9}Price data is available for the banking industry, but it is problematic because of government regulation of the banking industry and product heterogeneity, where output prices depend on risk.
portunity to illustrate new semiparametric techniques. Chapter 2 introduces two semiparametric estimators: one estimator that makes minimal assumptions on the distribution of the effects and estimates the slope parameters efficiently, and another that makes minimal assumptions on the functional form of the inputs in the output distance function. Furthermore, output determination in the banking sector is discussed. Chapter 3 derives another similar semiparametric efficient estimator using slightly different assumptions. More importantly, this chapter discusses the problems such as binwidth and kernel selection which confront a researcher using these techniques. Chapter 4 derives a new semiparametric efficient estimator which not only makes minimal assumptions on the distribution of the effects, but also on the distribution of the random error. Chapter 5 considers the problem of time varying efficiencies and their movements (convergence-divergence) in the banking industry. The results from the three estimators are compared. A new semiparametric estimator which allows for time varying efficiencies and makes minimal functional form assumptions is illustrated.
Chapter 1

Data

This chapter gives a detailed description of the data. The data were collected by Allen N. Berger of the Board of Governors of the Federal Reserve System and Wharton School. This description borrows from Berger’s data description in other papers (Berger, 1991; Berger, 1993; Berger, Hancock, and Humphrey, 1993; Akavein, Swamy, and Taubman, 1996).

The balance sheet data consists of annual observations from 1980 through 1989 of U.S. commercial banks and were collected from the Report of Condition and Income (Call Report) and the FDIC Summary of Deposits. These observation represent averages of three points in time (December, June, previous December) to improve accuracy. All dollar figures are in thousands of 1982 dollars.

In order to account for differences in regulatory environment, the dataset has been separated into three subsamples: unit banking (Unit), limited branching (Limit), and statewide branching (State). Only banks that were in existence for all 10 years and whose state had the same law status in at least 9 of the 10 years were included.¹ All banks with any zero observations for a specific variable were excluded to avoid any statistical problems with boundaries cause by logarithmic transformations.² The endresult is three balanced panel datasets of 796 to 933 banks each or of 7960 to 9330 total observations each. The included banks represent over 45% of all bank assets as of 1989.

¹The exclusion of failed or merged banks could lead to a bias in the efficiency estimates. This is especially relevant because the number of failed banks rose substantially during the 80s. Berger (1993) notes this shortcoming of the dataset.
²Berger adds ones to these observations. Some banks are dropped because of this difference in data.
Inputs include labor, capital and purchased funds. Labor is the number of full
time-equivalent employees on payroll at the end of the current period (round to the
nearest whole number). Capital is measured by premises and fixed assets including
capitalized leases. Purchased funds consists of several posts: deposits greater than
$100,000, foreign debt, federal funds purchased, and other liabilities for borrowed
money.

Measurement of loans and deposits is a more complex matter. The banking indus-
try is a service industry, where the measurement of services to the consumer cannot
be directly observed. Two approaches to measuring services are available: produc-
tion and intermediation approach.3 The production approach gives rise to the notion
that the number of accounts or loans best measure the services produced by banks,
while the intermediation approach assumes that dollar amounts of loan and deposit
stocks are proportional to the services rendered. The data are collected according to
the intermediation approach. Loan types include real estate loans, commercial and
industrial loans, and installment loans. Installment loans include loans to individual
households, family, and other personal expenditures. Demand deposits and time and
saving deposits are listed in the call reports.4

The analysis in this dissertation centers on the traditional banking outputs, i.e.
loans. It should be noted that commercial banks have moved into other non-traditional
markets. One indication of the importance of these other markets to the banking in-
dustry is the percent of noninterest income. For all banks, the percent of noninterest
income rose from 8% in 1980 to 18% in 1989.5 This increase in off balance sheet
activities does represent an important change in the industry, but an analysis which
concentrates on loans is still warranted, because they make up a substantial portion
of bank income.

---

3See Berg et al for a comparison of both approaches.
4Time and saving deposits greater than $100,000 are included in purchased funds.
The bank size variable is a count variables measured by total assets of the banks. The ranges are defined in table 1.1. The table also includes the number and percentage (in parenthesis) of banks in each size category.

<table>
<thead>
<tr>
<th>Size</th>
<th>Total Assets</th>
<th>Unit</th>
<th>State</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0-10 Mil</td>
<td>528 (5.7)</td>
<td>198 (2.3)</td>
<td>294 (3.7)</td>
</tr>
<tr>
<td>2</td>
<td>10-25 Mil</td>
<td>2427 (26)</td>
<td>1250 (14.5)</td>
<td>1796 (22.5)</td>
</tr>
<tr>
<td>3</td>
<td>25-50 Mil</td>
<td>2336 (25)</td>
<td>1835 (21.3)</td>
<td>2228 (27.9)</td>
</tr>
<tr>
<td>4</td>
<td>50-75 Mil</td>
<td>1328 (14.2)</td>
<td>1159 (13.4)</td>
<td>1065 (13.3)</td>
</tr>
<tr>
<td>5</td>
<td>75-100 Mil</td>
<td>785 (8.4)</td>
<td>732 (8.5)</td>
<td>667 (8.4)</td>
</tr>
<tr>
<td>6</td>
<td>100-200 Mil</td>
<td>1265 (13.6)</td>
<td>1226 (14.2)</td>
<td>980 (12.3)</td>
</tr>
<tr>
<td>7</td>
<td>200-300 Mil</td>
<td>310 (3.3)</td>
<td>420 (4.9)</td>
<td>251 (3.1)</td>
</tr>
<tr>
<td>8</td>
<td>300-500 Mil</td>
<td>187 (2.0)</td>
<td>411 (4.8)</td>
<td>216 (2.0)</td>
</tr>
<tr>
<td>9</td>
<td>500-1 Bil</td>
<td>63 (0.7)</td>
<td>437 (5.1)</td>
<td>161 (2.0)</td>
</tr>
<tr>
<td>10</td>
<td>1-2 Bil</td>
<td>38 (0.4)</td>
<td>330 (3.8)</td>
<td>152 (1.9)</td>
</tr>
<tr>
<td>11</td>
<td>2-5 Bil</td>
<td>23 (0.2)</td>
<td>359 (4.2)</td>
<td>120 (1.5)</td>
</tr>
<tr>
<td>12</td>
<td>5-10 Bil</td>
<td>20 (0.2)</td>
<td>127 (1.5)</td>
<td>35 (0.5)</td>
</tr>
<tr>
<td>13</td>
<td>≥ 10</td>
<td>20 (0.2)</td>
<td>136 (1.6)</td>
<td>14 (0.2)</td>
</tr>
</tbody>
</table>

Number and % in each Size

The data description in table 1.2 gives the means for inputs and outputs. The data is already separated into the three subsamples: Unit, Limit, and State. These are individual banks which do or do not belong to a bank holding company. This analysis does not consider the ramifications of bank holding company ownership on banking efficiency.

Table 1.3 describes where the data can be located on a Call Report. It is important to note that the Call Reports have changed and that some of these post number could have changed as well. The ones listed are consistent through the entire sample period.
### Table 1.2 General Data Description

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
<th>Unit</th>
<th>Limit</th>
<th>State</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>Log of average number of employees</td>
<td>3.4525</td>
<td>3.6596</td>
<td>4.4843</td>
</tr>
<tr>
<td>K</td>
<td>Log of physical capital (fixed asset)</td>
<td>6.4089</td>
<td>6.5096</td>
<td>7.4927</td>
</tr>
<tr>
<td>Purf</td>
<td>Log of purchased funds</td>
<td>8.2342</td>
<td>8.2486</td>
<td>9.3359</td>
</tr>
<tr>
<td>DD</td>
<td>Log of demand deposits</td>
<td>8.9507</td>
<td>8.9389</td>
<td>9.92</td>
</tr>
<tr>
<td>OD</td>
<td>Log of retail time and savings deposits (&lt; $100,000)</td>
<td>10.1428</td>
<td>10.4538</td>
<td>10.8430</td>
</tr>
<tr>
<td>cln</td>
<td>Log of commercial and industrial loans</td>
<td>8.5174</td>
<td>8.5129</td>
<td>9.94391</td>
</tr>
<tr>
<td>inln</td>
<td>Log of installment loans</td>
<td>8.3580</td>
<td>8.5605</td>
<td>9.3598</td>
</tr>
<tr>
<td>reln</td>
<td>Log of real estate loans</td>
<td>8.9547</td>
<td>9.2778</td>
<td>9.9308</td>
</tr>
<tr>
<td>size</td>
<td>Size measured as a count variable</td>
<td>3.7066</td>
<td>4.0672</td>
<td>5.1690</td>
</tr>
<tr>
<td></td>
<td>Number of Observations</td>
<td>9330</td>
<td>7960</td>
<td>8620</td>
</tr>
</tbody>
</table>

### Table 1.3 Call Report Information

<table>
<thead>
<tr>
<th>Variable</th>
<th>Call Report Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>reln</td>
<td>RCFD 1410</td>
</tr>
<tr>
<td>clin</td>
<td>RCFD 1600</td>
</tr>
<tr>
<td>inln</td>
<td>RCFD 1975</td>
</tr>
<tr>
<td>DD</td>
<td>RCON 2210</td>
</tr>
<tr>
<td>OD</td>
<td>RCON 2350 - RCON 6645</td>
</tr>
<tr>
<td>Purf</td>
<td>RCON 6645 + RCFD 2800 + RCFD 2835 + RCFD 2910 + RCFD 3200</td>
</tr>
<tr>
<td>L</td>
<td>RCFD 4150</td>
</tr>
<tr>
<td>K</td>
<td>RCFD 2145</td>
</tr>
</tbody>
</table>
Chapter 2

Semiparametric Approaches to Stochastic Panel Models

This chapter relies heavily on a paper of similar title written by Adams, Berger, and Sickles (1996).

There has been great deal of research interest in the topic of efficiency in the banking industry over the past several years (see Berger, Hunter, and Timme 1993 for a review). There are a number of reasons for this attention. First, banking has historically been a regulated industry with limited market entry, which might allow inefficiencies to persist over time. In fact, some recent evidence suggests that the cost inefficiencies created by a lack of market discipline in this industry may substantially exceed the social costs of any mispricing of deposit and loan services created by this market power (see Berger and Hannan, 1994). In any event, the regulation of the industry has brought with it a wealth of publicly available data for researchers to use. In this study, we use panel data with over 25,000 observations of U.S. banks over a 10-year period.

Second, the on-going consolidation of the banking industry raises policy concerns about the efficiency effects of mergers and acquisitions in this industry. From 1979 to 1994, the numbers of banks and banking organizations in the U.S. fell by about one-third. Further consolidation is expected under the Riegle-Neal Interstate Banking and Branching Efficiency Act of 1994, which essentially allows nationwide bank branching as of 1997. Similar consolidation is expected across international boundaries in the European Union over the next several years. The research findings using data from the 1980s generally suggest that there are no cost efficiency gains on average from mergers, but profit efficiencies may have increased on average because of improved
diversification (see Berger and Humphrey, 1992a; Akhvein, Berger, and Humphrey, 1996). Evidence from the 1990s data is not yet conclusive.

Third, the financial services produced by the banking industry create some interesting measurement issues. Because of data limitations, banking services are usually assumed to be proportional to the dollar values of assets or liabilities, but there is no clear consensus on which assets and liabilities constitute outputs. For example, deposits have output characteristics because banks provide important transactions, liquidity, and safekeeping services to depositors. However, deposits also have input characteristics because they supply investable funds that are needed in the production of assets. Three approaches have been used to identify which balance sheet entries should count as outputs: 1) the asset approach (only assets are outputs); 2) the user cost approach (accounts that provide net revenue above opportunity costs are outputs); and 3) the value-added approach (accounts associated with large expenditures of real resources are outputs).6 The key difference among these approaches is whether deposits are inputs or outputs. In the asset approach they are always inputs. In the user cost approach they are sometimes inputs and sometimes outputs. In the value-added approach they are usually outputs. In this paper, we are able to test whether deposits are more appropriately treated as inputs versus outputs, indirectly giving evidence as to which of these approaches is most consistent with (or least violates) the data of the U.S. banking industry. By way of preview, we find that the model specification of deposits as inputs statistically dominates the other model specifications, providing some indirect support for the asset approach.

Finally, study of banking industry efficiency is of academic interest because all of the major efficiency measurement techniques have been applied to banking data over the last several years. The approaches in the literature differ primarily by their method of disentangling inefficiency differences from random error. The econometric

6For a more complete discussion of these approaches see Berger and Humphrey (1992b).
frontier approach (EFA) assumes that inefficiencies follow an asymmetric distribution (e.g., the half-normal) and the random errors follow a symmetric distribution (e.g., the normal), and that both the inefficiencies and random errors are orthogonal to all of the regressors (applied to banking data by Ferrier and Lovell, 1990; Bauer, Berger, and Humphrey, 1993; and others). The thick frontier approach assumes that deviations from predicted costs within a grouping of low-cost banks represent random error, while differences in predicted costs between high- and low-cost groups represent X-inefficiencies (Berger and Humphrey, 1991,1992b; Bauer, Berger, and Humphrey, 1993). No restriction is put on the correlations between inefficiencies and the regressors. DEA and other mathematical programming approaches typically assume no random error and attribute all deviations from estimated frontier to inefficiency (Ferrier and Lovell, 1990; Fixler and Zieschang, 1991; Ferrier, Grosskopf, Hayes, and Yaisawarng, 1993). Finally, the ‘distribution-free’ approach is a panel estimation method that assumes that efficiency differences across banks are constant over the observed time periods, while random error tends to average out over time (Bauer, Berger, and Humphrey, 1993; Berger, 1993; Berger, Hancock, and Humphrey, 1993).

Our study parallels the distribution-free method, using the same data set (with a few minor deletions) as Berger (1993). Berger applied two of the methods outlined in Schmidt and Sickles (1984) to the data: OLS, where inefficiencies and regressors are assumed to be orthogonal and the inefficiencies are obtained by averaging the composite error term; and the dummy variable method, where firm-specific dummy variables identify firm inefficiencies. In the latter case, efficiencies and regressors can freely be correlated. In both cases, some truncation of extreme values was performed to account for very good or bad luck by some banks. Berger’s OLS method has the problem that if the inefficiencies are not orthogonal to the cost function regressors (the output quantities and input prices) as assumed, the inefficiencies may be understated and the scale, scope, and input price effects may be overstated. Berger’s dummy variable method solved this problem by removing the orthogonality restrictions, but
ran into another empirical problem that yielded unreasonable results. The inefficiency measure appeared to pick up too many of the effects of bank scale, since banks differ so much more in terms of scale (some banks are 10,000 times as large as others) than they do in terms of inefficiency.

In this study, we try to resolve these difficulties by introducing a stochastic distance function measure that allows the inefficiencies to be correlated with a subset of the regressors. The distribution-free methods can only be justified if either all the regressors are orthogonal to the inefficiencies or all are correlated with the inefficiencies. We consider the case where only a subset of the regressors is correlated with the inefficiencies in the spirit of Hausman and Taylor (1981) and Cornwell, Schmidt and Sickles (1990). In this case, within estimation will yield consistent results, but more efficient estimators can be used to take advantage of this specific correlation. Hausman and Taylor (1981) suggest an instrumental variables estimator to deal with the correlated regressors and to get a more efficient estimate. Similarly, Amemiya and McCurdy (1984) and Breusch, Mizon and Schmidt (1989) have suggested more elaborate instruments to solve this problem. One disadvantage to IV estimation is that it is not an efficient method of estimation. We suggest a semiparametric method which improves upon IV estimates and achieves a lower bound of variance.

We model the stochastic distance frontier by modifying a semiparametric efficient estimator developed by Park, Sickles, and Simar (1996). We assume that product mix affects the efficiency of firms causing outputs to be correlated with the inefficiencies. We then apply a model that makes minimal assumptions about the parametric form of the stochastic structure of firm inefficiencies.

Furthermore, a more flexible form of the distance function frontier is introduced. where no parametric assumptions are made for inputs. This specification of the distance frontier allows us to apply necessary restrictions on the outputs. In particular, linear homogeneity of the output distance function, and keep the same correlation structure as in the semiparametric case. The procedure involves estimating the con-
ditional expectation using a nonparametric regression estimator of the Nadaraya-Watson type, and then uses these estimates of the conditional expectation to transform the distance frontier function. The inefficiencies are then captured from the transformed model.

In section 1, we describe the basic model to be estimated. In section 2, an efficient estimator in the class of estimators which do not impose distributional assumptions on the model will be developed. This will be done by deriving the asymptotic lower bound for the parameter estimates of interest in the presence of nuisance parameters, while allowing for dependencies between the random effects and the outputs. The efficient semiparametric estimator attains this lower bound. Section 3 describes the nonparametric based estimation procedure. Section 4 compares the results of the estimation. A within estimator and Hausman-Taylor estimator are compared to the estimates of our model.

2.1 The Model
We start with the output distance function as a motivation for our model of a firm with multiple outputs and inputs. The output distance function provides a radial measure of technical efficiency that describes the fraction of aggregated outputs, given chosen inputs. For a particular observation $i$, the output distance function is:

$$D(x, y) = \frac{\prod_{j=1}^{m} y_{ji}^{\frac{1}{m}}}{\prod_{k=1}^{k} x_{ki}^{r_{ki}}} \leq 1$$

(2.1)

where $y_{ji}$ and $x_{ki}$ are the levels of output $j$ and input $k$. The $q_{ji}$ and the $r_{ki}$ are the weights which describe the technology of the firm. When a firm is producing efficiently or when the value of the distance function equals one, then it is not possible to increase the value of output without either decreasing another output or increasing inputs.
It is convenient to linearize equation (1) using logarithms of outputs and inputs and include error terms, which results in the following Cobb-Douglas form:

\[ 0 = \sum_j \gamma_j \ln y_{j, it} - \sum_k \beta_k \ln x_{k, it} + \alpha_i + \epsilon_{it} \] (2.2)

By construction, the output distance function is linearly homogeneous in outputs. We impose linear homogeneity and then normalize with respect to one \( y_j \) to get the following transformation:

\[ -\ln (y_j) = \sum_j \gamma_j \ln \hat{y}_{j, it} - \sum_k \beta_k \ln x_{k, it} + \alpha_i + \epsilon_{it} \] (2.3)

where \( \hat{y} = (\hat{y}_j) = (\frac{y_j}{y_j^*}), \; j = 1, \ldots, J - 1. \)

This is basically a generic panel data model where we estimate the firm specific efficiency levels in a stochastic panel production frontier model. The model fits into the class of production frontier models (Aigner, Lovell, and Schmidt, 1977; Meeusen and van den Broeck, 1977; Sickles and Schmidt, 1984; Cornwell, Schmidt and Sickles, 1990) and gives rise to the analysis of firm efficiencies (heterogeneities). In this class of models, the limitations of single cross sections (or time series) are overcome through the use of panel models, which allow us to directly identify the firm efficiencies.

For the basic framework of the model, we consider a panel frontier model with firm specific heterogeneities, derived from the distance function above. Let \( X_{it} = -\ln x_{k, it}, Y_{it}^* = \ln \hat{y}_{j, it} \) and \( Y_{it} = -\ln y_j \): \n
\[ Y_{it} = X_{it}' \beta + Y_{it}' \gamma + \alpha_i + \epsilon_{it} \] (2.4)

where we assume \( N \) independent observations \( (X_i, Y_i^*, Y_i) \) and \( X_i = (X_{i1}', \ldots, X_{iT}') \), \( Y_i^* = (Y_{i1}^*, \ldots, Y_{iT}^*) \), \( Y_i = (Y_{i1}, \ldots, Y_{iT}) \). The \( X_i \) are i.i.d. \( k \times T \) dimensional random

---

7 This approximation of the output distance function could lead to measurement error. This possible problem is addressed in section 4, where we allow for a flexible functional form of the inputs using a nonparametric regression estimator.

8 We use the Cobb-Douglas form for simplicity. As the number of correlated variables increases, so does the dimension of the kernel estimates of the joint density with the effects. This has an adverse effect on the convergence of the kernel estimates, which in turn, affects the convergence of the semiparametric efficient estimator.

9 See Lovell, Richardson, Travers and Wood (1995) for a complete discussion.
vectors with unknown density function \( g \) and \( \beta \) is a \( k \) dimensional unknown vector. \( \gamma \) is a \( J-1 \) dimensional unknown vector. The \( \epsilon_{it} \) are considered i.i.d. random variables from \( N(0, \sigma^2) \) with unknown \( \sigma^2 \) and the \((\alpha_i, Y_i^*)\) are i.i.d. from unknown density \( h \). It is important to note that the \( \alpha \)'s are independent of the \( X_i \), but not to \( Y_i^* \). We assume the support of the marginal density of \( \alpha \) is bounded above (below). Let \((\alpha, Y^{*,*}, X^*)\)' denote generic observations \((\alpha_i, Y_i^{*,*}, X_i^*)\)' and \( \bar{Y}^* = T^{-1} \sum_{i=1}^{T} Y_i^* \).

Consider the joint distribution of \( f(\alpha, X, Y^*) \). We assume that \( \alpha_i \) and \( X_i \) are conditionally independent and that firm heterogeneities are correlated only with long run movements of \( Y^* \), i.e. only with \( \bar{Y}^* \).\(^{10}\) Thus, the density of \( \alpha, X, Y^* \) can be described in the following manner:

\[
f(\alpha, x, y^*) = h(\alpha, \bar{y}^*)g(x|y^*)
\]

(2.5)

where \( h \) represents the joint distribution of \( \alpha \) and \( \bar{y}^* \). This assumption simplifies the correlation structure, but it avoids the "curse of dimensionality" problem that would arise.

### 2.2 Efficient Estimation

We utilize below a semiparametric efficient estimation procedure which allows for multiple regressors to be correlated with the effects. In this estimation method, efficient estimation of the slope coefficients does not depend on the nuisance parameters, i.e. the nonparametric part of the model. In other words, the slope parameters can be efficiently estimated regardless of the nonparametric part of the model. For an excellent discussion on efficient estimation see Bickel, Klaassen, Ritov and Wellner (1993) as well as Newey (1990) and Robinson (1988). We briefly outline the procedure here and note any differences to Park, Sickles, and Simar (1996).

\(^{10}\)This is model 3 in Park, Sickles, and Simar (1996). Other possible correlation structures are considered in their article.
We start by considering the set of all possible joint distributions of the observable variables \((X, Y^*, Y)\), and as in Park, Sickles, and Simar (1996), assume regularity of the parametric submodel to insure smoothness of the model. Let \(L(X, Y^*, Y; \theta, \eta)\) denote the log likelihood of an observation and \(\ell_\theta(X, Y^*, Y) = \frac{\partial L}{\partial \theta |_{\theta_0, m_0}}\) and \(\ell_{\eta_i}(X, Y^*, Y) = \frac{\partial L}{\partial \eta_i |_{\theta_0, m_0}}\) where \(\eta = (\eta_1, \ldots, \eta_m)\) denote the scores. Let \([\ell_{\eta_i}]\) denote the linear span generated by \(\{\ell_{\eta_i}\}_{i=1}^m\). The nuisance parameters denoted by \(\eta\) are \(\sigma^2\) and \(h(\alpha, y^*)\), the unknown distribution of the effects and \(y^*\) in this case. The information bound for \(\theta\) is then given by:

\[
I(X, Y^*, Y; \theta) = E \ell^* \ell^{*T}
\]

where \(\ell^* = \ell_\theta - \Pi(\ell_\theta | [\ell_\eta])\) and the notation \(\Pi(\ell_\theta | [\ell_\eta])\) denotes the vector of projections of each component of \(\ell_\theta\) onto \(\ell_\eta\). In other words, we project the scores with respect to the slope parameters onto the nuisance parameter tangent space and then purge the scores of these projections to get the efficient score, which is then orthogonal to the nuisance parameters.

To define efficiency, we call an estimator \(\hat{\theta}_N\) efficient if it achieves the semiparametric efficiency bound defined above:

\[
\sqrt{N}(\hat{\theta}_N - \theta) \rightarrow N(0, [I^{-1}(X, Y^*, Y; \theta)])
\]

We now will show the information bound for the model above. Let \(X' = (X_{i1}, \ldots, X_{iT})'\), \(Y^* = (Y^*_{i1}, \ldots, Y^*_{iT})'\) and \(Y_i = (Y_{i1}, \ldots, Y_{iT})'\) for the generic observations \((X, Y^*, Y)\).

Let \(S_i(\theta) = Y_i - X_i \beta - Y_i^* \gamma\) and \(U_i = S_i(\theta) - \bar{S}_i(\theta)\) where \(Y_i\) is the relative output and \(\bar{S}_i(\theta) = T^{-1} \sum_{t=1}^T S_i(\theta)\). Let

\[
\omega(\tilde{s}, \tilde{y}^*) = \int \phi_\sigma(\tilde{s} - u)h(u, \tilde{y}^*)du = \phi_\sigma h(\ldots, \tilde{y}^*)(\tilde{s})
\]

be the joint density of \((\bar{S}(\theta), \bar{Y}^*)\), where \(\sigma = \frac{T}{T}^\text{\#}\) and \(\bar{Y}^* = T^{-1} \sum_{t=1}^T Y_i^*\). Let
\[ I_0 = \int \frac{(\omega')^2}{\omega} (\tilde{s}, \tilde{y}^*) d\tilde{s} d\tilde{y}^* \]

be the Fischer information for location of \( \omega \), where \( \omega' \) is the derivative with respect to the first component of the vector. Consider the within and between covariances, which are defined as follows:

\[ \Sigma_W(X) = \mathbb{E} \left[ T^{-1} \sum_{t=1}^{T} (X_t - \bar{X})(X_t - \bar{X})' \right] \]

\[ \Sigma_W(Y^*) = \mathbb{E} \left[ T^{-1} \sum_{t=1}^{T} (Y_t^* - \bar{Y}^*)(Y_t^* - \bar{Y}^*)' \right] \]

\[ \Sigma_W(X, Y^*) = \mathbb{E} \left[ T^{-1} \sum_{t=1}^{T} (Y_t^* - \bar{Y}^*)(X_t - \bar{X})' \right] \]

\[ \Sigma_B(X|Y^*) = \mathbb{E} \left[ T^{-1} \sum_{t=1}^{T} (\bar{X} - \mathbb{E}(\bar{X}|Y^*))(\bar{X} - \mathbb{E}(\bar{X}|Y^*))' \right] \]

\[ \Sigma_W = \begin{bmatrix} \Sigma_W(X) & \Sigma_W(X, Y^*) \\ \Sigma_W(X, Y^*) & \Sigma_W(Y^*) \end{bmatrix} \]

where \( \bar{X} = T^{-1} \sum_{t=1}^{T} X_t \). We will assume that \( I_0 < \infty \) and that the within and between moment matrices both exist and are nonsingular. The efficient scores and information bound are obtained as:\(^{11}\)

\[ \ell'_3 = \sigma^{-2} \sum_{t=1}^{T} U_t X_t - (\bar{X} - \mathbb{E}(\bar{X}|Y^*)) \left( \frac{\ell'}{\omega} \right) (\tilde{s}, \tilde{y}^*) \]

\[ \ell_* = \sigma^{-2} \sum_{t=1}^{T} U_t Y_t^* \]

\(^{11}\)The proof is the same as theorem 2.5 in Park, Sickles, and Simar (1996), except \( Y^* \) is a vector.
\[ I = \begin{pmatrix} 
\tilde{\sigma}^{-2} \Sigma_w(X) + I_0 \Sigma_B(X|\tilde{Y}^*) & \tilde{\sigma}^{-2} \Sigma_w(X, Y^*) \\
\tilde{\sigma}^{-2} \Sigma_w(X, Y^*) & \tilde{\sigma}^{-2} \Sigma_w(Y^*) \end{pmatrix} \]

To construct an efficient estimator of \( \theta \), define \( \tilde{\ell} = I^{-1} \ell^* \) as the efficient influence function, where \( E \tilde{\ell} = 0 \) and \( E \tilde{\ell} \tilde{\ell}' = I^{-1} \). We proceed in the following manner (see Bickel, Klaassen, Ritov and Wellner, 1993):

(A) Find a \( \sqrt{N} \)-consistent estimator \( \tilde{\theta}_N \) of \( \theta \).

(B) Consider \( \tilde{\ell} \) as \( \tilde{\ell}(x, y^*, y, \theta, \sigma^2, h) \) and construct a suitable estimator

\[ \hat{\ell}(\ldots, \theta; X_1, Y_1^*, Y_1, \ldots, X_N, Y_N^*, Y_N) \]

of \( \ell(\ldots, \theta, \sigma^2, h) \) where we note that a suitable estimator retains the regularity conditions imposed on the distributions and has the proper rate of convergence needed for \( \sqrt{N} \)-consistency.

(C) Form the efficient estimator:

\[ \hat{\theta}_N = \tilde{\theta}_N + N^{-1} \sum_{i=1}^T \hat{\ell}(X_i, Y_i^*, Y_i, \theta; X_1, Y_1^*, Y_1, \ldots, X_N, Y_N^*, Y_N) \]

In our model, we need to find a \( \sqrt{N} \)-consistent estimator as a preliminary estimator \( \hat{\theta}_N \) of \( \theta \). One possibility is the within estimator obtained by regressing the mean deviations of \( Y_{it} \) on the mean deviations of \( X_{it} \) and \( Y_{it}^* \) using ordinary least squares. Park and Simar (1995) as well as Park, Sickles, and Simar (1996) show that within is efficient when all regressors are correlated with the effects. Within is \( \sqrt{N} \)-consistent and thus can be used as a preliminary estimator. However, when only a subset of the regressors is correlated with the effects, a more efficient IV estimator is available as a preliminary estimator, the Hausman-Taylor estimator. The Hausman-Taylor estimator is a \( \sqrt{N} \)-consistent IV estimator of a random effects model with instrument set \( (X_{it}, Q_v) \), where \( Q_v \) are the mean deviations of all regressors in the model. It is obtained by the least squares regression on a transformed model defined in the following manner:
\[ P_A \Omega^{-1} Y_{it} = P_A \Omega^{-1} X_{it} \beta + P_A \Omega^{-1} Y_{it}^* \gamma + P_A \Omega^{-1} \epsilon_{it} \]

where \( P_A \) is the orthogonal projection operator onto the column space of the instruments defined above.

To construct an efficient estimator of \( \omega \) define the kernel estimator:

\[
\omega(\tilde{s}, \tilde{y}^*, \theta) = N^{-1} \sum_{i=1}^{N} K_{s,n}(\tilde{s} - \tilde{s}_i) \prod_{j=1}^{J-1} K_{h,n}(\tilde{y}_{ij}^* - \tilde{Y}_{j, it}) + c_N
\]

where \( K_h = K(t/h)/h \). We use a higher order normal kernel defined as: \((1.5 - 0.5z^2)\phi(z)\), where \( \phi(z) \) is a normal kernel.\(^{12}\) For simplicity, we assume the same bandwidth and kernel for all variables.

Define the estimates of \( \sigma^2 \) and \( \Sigma_B(X|\tilde{Y}^*) \) in the following manner:

\[
\hat{\sigma}(\theta) = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} U_{it}^2(\theta)}{N(T - 1)}
\]

To estimate \( E(X|\tilde{Y}^*) \) at \( \sqrt{N} \)-rate in \( \ell_2^* \) and \( \Sigma_B(X|\tilde{Y}^*) \), we use a parametric form for \( E(X|\tilde{Y}^*) \). As an example, we use a linear model:

\[
E(X|\tilde{Y}^*) = a + A\tilde{Y}^*
\]

where \( a \) is a \( p \times 1 \) and \( A \) is a \( p \times T(J - 1) \) matrix.

Then, the model above can be estimated at \( \sqrt{N} \)-rate by standard least square methods. Define estimates by \( \hat{a} \) and \( \hat{A} \) and let \( \hat{p}(\tilde{Y}^*) = a + A\tilde{Y}^* \). \( \hat{p}(\tilde{Y}^*) = \hat{a} + \hat{A}\tilde{Y}^* \).

We can then estimate \( \hat{\Sigma}_B(X|\tilde{Y}^*) \) in the following manner:

\[
\hat{\Sigma}_B(X|\tilde{Y}^*) = \frac{1}{N} \sum_{i=1}^{T} (X - p(\tilde{Y}^*))(X - p(\tilde{Y}^*))'
\]

The semiparametric efficient estimator is then defined by:

\(^{12}\)Higher order kernels are used to insure proper convergence of the multidimensional kernel estimates.
\[ \hat{\theta}_N = \hat{\theta}_N + N^{-1} \hat{I}^{-1} \left[ \sum_{i=1}^{N} \left[ \sum_{t=1}^{T} \hat{\sigma}^{-2} \hat{U}_{it} X_{it} - (\bar{X}_t - \hat{p}(\bar{Y}^*)) \left( \frac{\tilde{w}_i}{\tilde{w}^*} \right) (\tilde{S}_i, \tilde{Y}_i^*, \hat{\theta}_N) \right] \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\sigma}^{-2} \hat{U}_{it} Y_{it}^* \right] \]

where

\[
\hat{I} = \begin{pmatrix}
T \hat{\sigma}^{-2} \hat{\Sigma}_W(X) + \hat{I}_0 \hat{\Sigma}_B(X|\bar{Y}^*) & T \hat{\sigma}^{-2} \hat{\Sigma}_W(X, Y^*) \\
T \hat{\sigma}^{-2} \hat{\Sigma}_W(X, Y^*) & T \hat{\sigma}^{-2} \hat{\Sigma}_W(Y^*)
\end{pmatrix}
\]

\[
\hat{I}_0 = N^{-1} \sum_{i=1}^{N} \left( \frac{\tilde{w}_i}{\tilde{w}^*} \right)^2 (\tilde{S}_i, Y_i, \hat{\theta}_N)
\]

where \( \hat{\theta}_N \) is the Hausman-Taylor estimator, and where \( \hat{\sigma}^2, \hat{U}_{it} \) and \( \tilde{S}_i \) are used for \( \hat{\sigma}^2(\hat{\theta}_N), \hat{U}_{it}(\hat{\theta}_N), \tilde{S}_i(\hat{\theta}_N) \).

We now focus on the predictors of the individual effects, given the efficient estimator \( \hat{\theta}_{N,T} \). Park, Sickles, and Simar (1996) investigate the asymptotic properties of these predictors and of the estimator of the level of the frontier function in more detail. We discuss some of the major results pertaining to the predictors of the individual effects. The individual effects, \( \alpha_i \), are predicted by:

\[ \hat{\alpha}_i = \tilde{S}_i(\hat{\theta}_{N,T}) \]

Under the assumptions of the model above Park, Sickles, and Simar (1996) prove that as \( T \to \infty \) and \( T \hat{\sigma}^2_{N,T} \to \infty \):

\[ \sqrt{T}(\hat{\alpha}_i - \alpha_i) \to N(0, \sigma^2) \]

Relative technical inefficiencies of the i-th firm with respect to the j-th firm can be predicted by: \( \hat{\alpha}_i - \hat{\alpha}_j \). We normalize with respect to the most efficient firm: \( \max_{j=1,\ldots,N}(\hat{\alpha}_j) \). Relative technical efficiencies are then derived by taking the distance of each firm's \( \alpha_i \) from the most efficient firm. The predictor of the relative
technical efficiencies \((\hat{\alpha}_i - \max_{j=1,...,N}(\hat{\alpha}_j))\) has an asymptotic \(N(0,2\sigma^2)\) distribution, when normalized by \(\sqrt{T}\).

2.3 Nonparametric Regression Estimation

The semiparametric efficient estimator of the basic model (1) is used to estimate a linearized form of the distance frontier.\(^{13}\) It is not apparent that the log linearized form of the distance frontier is an appropriate approximation. A more flexible, nonlinear form may be warranted. The semiparametric efficient estimator does add flexibility to the linearized form in that the joint distribution of the effects and a subgroup of regressors is not specified. In this section, we propose a another semiparametric procedure in the spirit of Robinson (1988) to estimate a distance frontier, where the functional form of the inputs is not parameterized. In this procedure, we use a nonparametric regression to find the conditional expectation with respect to the inputs. Kneip and Simar (1995) suggest a fully nonparametric regression estimator to find the firm effects in the case of a single output. Their estimator is based on a two step method that first uses a kernel estimator to find the conditional mean and then OLS to generate the effects. In a multiproduct setting, their procedure becomes intractable, because appropriate restrictions for the distance frontier such as linear homogeneity of outputs cannot be applied transparently to the functional form. Our semiparametric procedure, though similar to Kneip and Simar (1995), allows us to impose the appropriate restrictions to the distance function and also to apply the same correlation structure to the model as in the semiparametric efficient estimator. Kneip and Simar's (1995) estimator assumes that firm effects are uncorrelated with the regressors, while our estimator allows for correlation.\(^{14}\)

\(^{13}\)See section 2.

\(^{14}\)In the banking efficiency literature, McAllister and McManus (1993) apply nonparametric regression as well as other nonparametric methods to estimate a cost frontier.
The partially nonparametric distance function is defined in the following manner as in equation (1):

$$D(X, Y) = \frac{\prod_{i} Y_{ij}^*}{f(X)} \leq 1$$ (2.6)

where $Y_{ij}$ and $X$ are defined as before. $f(X)$ is a smooth function that is at least twice differentiable.\textsuperscript{15} By the same manipulations as before, the distance frontier becomes:

$$Y_{it} = f(X) + Y_{it}^* \gamma + \alpha_{i} + \epsilon_{it}$$ (2.7)

where $Y_{it}$ and $Y_{it}^*$ are defined as before. Assuming that the inputs are not correlated with the effects, the conditional expectation for the distance frontier function is:

$$E[Y|X] = f(X) + E[Y^*|X]$$

where the means of the random effect $\alpha_{i}$ are assumed to be zero. Subtracting this conditional expectation from the distance function provides us with the model to be estimated:

$$Y_{it} - E[Y|X] = (Y_{it}^* - E[Y^*|X]) \gamma + \alpha_{i} + \epsilon_{it}$$ (2.8)

$$f(X) = E[Y|X] - E[Y^*|X] \gamma$$ (2.9)

We apply dummy variables or the within estimator to the transformed model to find the parameters, $\gamma$,\textsuperscript{16} and firm effects, $\alpha_{i}$, can be estimated as in section 3.

### 2.3.1 Estimation of Conditional Expectations

To estimate the conditional mean,\textsuperscript{17} we use a kernel based nonparametric regression. Although, different nonparametric regression methods are available, we apply the

\textsuperscript{15}The function contains an intercept, which is cancelled out in the following steps.

\textsuperscript{16}In this procedure, the correlation between the effects and outputs is retained, because the within estimator allows for correlation between the regressors and effects.

\textsuperscript{17}We show this only for the conditional expectation of $Y$ given $X$, but it is exactly the same for $Y^*$ given $X$. 
Nadaraya-Watson estimator. A more detailed discussion of this nonparametric estimator can be found in Haerdle (1990), Scott (1992), Kneip and Simar (1995). For a discussion of asymptotic properties and rates of convergence of the nonparametric estimators, we refer the reader to the articles mentioned. The derivation starts by defining the conditional mean function $E[Y|X]$ in the following manner:

$$E[Y|X] = \int Y \, \nu(Y|X = x_0) dY$$

(2.10)

$$= \frac{\int Y \, \nu(Y, X) dY}{\int \nu(Y, X) dY}$$

$$= \frac{(nh^k) \sum_{i=1}^{N*} Y_i K(\frac{X_i - x_0}{h})}{(nh^k) \sum_{i=1}^{N*} K(\frac{X_i - x_0}{h})}$$

where $\nu(X, Y)$ is the joint distribution of $(X, Y)$, $h$ the bandwidth, and $k$ the dimension of $X$. This can further be reduced to the following form:

$$E[Y|X] = \frac{\sum_{i=1}^{N*} K(\psi_i) Y_i}{\sum_{i=1}^{N*} K(\psi_i)}$$

(2.11)

where $\psi_i = \frac{X_i - x_0}{h}$

The nonparametric regression estimate of $E[Y|X]$ is thus a weighted sum of those $Y_i$'s that correspond to $X_i$ in a neighborhood of $X$. This estimator is also called a linear smoother, because it is linear in $Y_i$.

2.4 Results

We use a Cobb-Douglas stochastic panel distance function (Schmidt and Sickles, 1984: Cornwell, Schmidt, and Sickles, 1990) for estimation, where $y = \ln($real estate loans$)$; $X = \ln($labor$), \ln($capital$), \ln($purchased funds$), \ln($demand deposits$), \ln($retail time and savings deposits$), \ln($size$); Y^* = \ln($commercial and industrial loans$), \ln($installment loans$). English, et al. (1993) estimate a distance function using a linear programming method, where they classify both deposit types as inputs. They assume a translog functional form of the distance function. Fixler and Zieschang (1991) estimate a
distance function for revenue share equations using the user cost approach. They observe that deposits and other specified types of securities showed qualities of outputs and inputs, where output and input status changed from year to year (especially for deposits).

Other studies like Berger (1993) and Bauer, Berger and Humphrey (1993) estimate panel cost frontiers, where their model specification includes both deposit types as outputs. They base their model specification on the fact that a large percentage of operating costs (i.e. labor and capital expenses) is absorbed by demand and retail time and savings deposits (Berger and Humphrey 1992). The cost function frontier setup allows them to model the dual characteristics of these regressors by including deposit interest. Furthermore, Berger (1993) allows for heterogeneities in the cost function by allowing the coefficients to vary from year to year. Bauer, Berger, and Humphrey (1993) estimate the model using GLS. Within and MLE, where the coefficients do not vary over time. They also estimate MLE and allow the coefficients and inefficiencies to vary over time. Large differences in inefficiencies are reported between GLS and the other three models, indicating problems with the assumption that regressors are uncorrelated with the effects. Measurement error due to model misspecification is dealt with in the nonparametric based estimator.

Differences in inefficiency estimates can be attributed more to differences in correlation structures between regressors and effects (Bauer, Berger, and Humphrey 1993). Our methods allow for a subgroup of regressors to be correlated to the effects, as opposed to other methods in the banking literature which assumes that all regressors are either correlated or uncorrelated with the effects. The next sections describe the results obtained from both our estimation procedures.

---

18Interestingly, there is little change between the two different MLE estimates. MLE by year has an average inefficiency of 15.3% while MLE (panel) is 16.6%.

19The estimator allows for a flexible functional form of the inputs, but not the outputs as described above.
2.4.1 Semiparametric Efficient Estimation

A Hausman test for orthogonality, which tests the performance of within estimation relative to random effects and thus the correlation assumptions of the firm heterogeneities, was performed on each of the data sets. In each test, the null hypothesis of no correlation was rejected at the 95% level.

The efficient semiparametric estimator of $\theta$ is based on a bandwidth $h$ chosen using a specification defined in Silverman (1986) in the following manner: $h_{opt} = \left(\frac{4}{(2d+p)^{1.5}}\right)\frac{1}{n(2d+p)}$ where $d$ is the dimension of the kernel estimates, $p$ is the order of the kernel estimator, and $n$ is sample size.\textsuperscript{20}

In each sample, we considered various model specifications and estimated a distance frontier utilizing the methods above. In every case, we assume that all loan types are outputs, while demand and savings deposits are specified as either outputs or inputs. We tested the different specifications by using a Hausman-Wu error specification test and compared them relative to within.\textsuperscript{21} We find the case where demand and savings deposits are specified as inputs performs the best against within.\textsuperscript{22} This result supports the asset approach used by English, et al. (1993) in their estimation of an output distance function. The within, Hausman-Taylor and efficient semiparametric estimates and standard errors for the stochastic panel distance function, where demand and savings deposits are inputs, are presented in tables 2.1, 2.2, and 2.3 for unit, limit and state banks.

We now move on to describe our technical efficiency results obtained from semiparametric efficient estimation. Average relative technical efficiency ranges from 61% to 68.5%, where limit banks are on average more efficient than unit or state banks.

\textsuperscript{20}Park, Sickles and Simar (1996) and Park and Simar (1995) suggest a method of cross validation to determine the optimal bandwidth. This method was applied to the data with very similar results.

\textsuperscript{21}We consider both Hausman-Taylor and the semiparametric efficient estimators versus within to test model specification.

\textsuperscript{22}In all samples, the semiparametric efficient estimator cannot be rejected when tested against within.
### Table 2.1 Unit Bank Regressions

<table>
<thead>
<tr>
<th>Estimated Slope Coefficients</th>
<th>Within</th>
<th>Hausman-Taylor</th>
<th>Semiparametric Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>-0.2144 (0.0148)</td>
<td>-0.1971 (0.0141)</td>
<td>-0.1987 (0.0107)</td>
</tr>
<tr>
<td>K</td>
<td>-0.1335 (0.0081)</td>
<td>-0.1369 (0.0077)</td>
<td>-0.1349 (0.0059)</td>
</tr>
<tr>
<td>Furf</td>
<td>-0.0948 (0.0046)</td>
<td>-0.1022 (0.0044)</td>
<td>-0.1008 (0.0034)</td>
</tr>
<tr>
<td>DD</td>
<td>-0.1366 (0.0104)</td>
<td>-0.2931 (0.0099)</td>
<td>-0.2881 (0.0076)</td>
</tr>
<tr>
<td>OD</td>
<td>-0.3764 (0.0110)</td>
<td>-0.3883 (0.0106)</td>
<td>-0.3873 (0.0084)</td>
</tr>
<tr>
<td>clin</td>
<td>0.3595 (0.0047)</td>
<td>0.3593 (0.0047)</td>
<td>0.3593 (0.0044)</td>
</tr>
<tr>
<td>inln</td>
<td>0.2954 (0.0051)</td>
<td>0.2893 (0.0051)</td>
<td>0.2893 (0.0050)</td>
</tr>
<tr>
<td>size</td>
<td>-0.0017 (0.0074)</td>
<td>0.0142 (0.0072)</td>
<td>0.0080 (0.0059)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.9282</td>
<td>0.9272</td>
<td>0.9243</td>
</tr>
<tr>
<td>F-Test</td>
<td>17240</td>
<td>16991</td>
<td>16273</td>
</tr>
</tbody>
</table>

### Table 2.2 Limit Bank Regressions

<table>
<thead>
<tr>
<th>Estimated Slope Coefficients</th>
<th>Within</th>
<th>Hausman-Taylor</th>
<th>Semiparametric Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>-0.1443 (0.0170)</td>
<td>-0.2009 (0.0161)</td>
<td>-0.1880 (0.0132)</td>
</tr>
<tr>
<td>K</td>
<td>-0.2577 (0.076)</td>
<td>-0.2119 (0.0073)</td>
<td>-0.2355 (0.0062)</td>
</tr>
<tr>
<td>Furf</td>
<td>-0.0735 (0.0046)</td>
<td>-0.0753 (0.0043)</td>
<td>-0.0679 (0.0036)</td>
</tr>
<tr>
<td>DD</td>
<td>-0.2854 (0.0108)</td>
<td>-0.2848 (0.0104)</td>
<td>-0.2921 (0.0089)</td>
</tr>
<tr>
<td>OD</td>
<td>-0.3973 (0.0127)</td>
<td>-0.4088 (0.0121)</td>
<td>-0.3946 (0.0100)</td>
</tr>
<tr>
<td>clin</td>
<td>0.3345 (0.0047)</td>
<td>0.3390 (0.0047)</td>
<td>0.3250 (0.0044)</td>
</tr>
<tr>
<td>inln</td>
<td>0.3474 (0.0061)</td>
<td>0.3484 (0.0060)</td>
<td>0.3484 (0.0059)</td>
</tr>
<tr>
<td>size</td>
<td>-0.0017 (0.0074)</td>
<td>0.0510 (0.0079)</td>
<td>0.0477 (0.0066)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.9469</td>
<td>0.9461</td>
<td>0.9460</td>
</tr>
<tr>
<td>F-Test</td>
<td>20274</td>
<td>19985</td>
<td>19924</td>
</tr>
</tbody>
</table>

### Table 2.3 State Bank Regressions

<table>
<thead>
<tr>
<th>Estimated Slope Coefficients</th>
<th>Within</th>
<th>Hausman-Taylor</th>
<th>Semiparametric Efficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>L</td>
<td>-0.1464 (0.0160)</td>
<td>-0.1561 (0.0149)</td>
<td>-0.1279 (0.0112)</td>
</tr>
<tr>
<td>K</td>
<td>-0.1383 (0.0097)</td>
<td>-0.1316 (0.0090)</td>
<td>-0.1442 (0.0068)</td>
</tr>
<tr>
<td>Furf</td>
<td>-0.1223 (0.0049)</td>
<td>-0.1258 (0.0047)</td>
<td>-0.1230 (0.0035)</td>
</tr>
<tr>
<td>DD</td>
<td>-0.3076 (0.0107)</td>
<td>-0.3019 (0.0102)</td>
<td>-0.3216 (0.0082)</td>
</tr>
<tr>
<td>OD</td>
<td>-0.4433 (0.0102)</td>
<td>-0.4428 (0.0098)</td>
<td>-0.4546 (0.0075)</td>
</tr>
<tr>
<td>clin</td>
<td>0.3866 (0.0048)</td>
<td>0.3901 (0.0048)</td>
<td>0.3901 (0.0047)</td>
</tr>
<tr>
<td>inln</td>
<td>0.2780 (0.0053)</td>
<td>0.2778 (0.0053)</td>
<td>0.2778 (0.0050)</td>
</tr>
<tr>
<td>size</td>
<td>0.0581 (0.0071)</td>
<td>0.0605 (0.0067)</td>
<td>0.0649 (0.0051)</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.9600</td>
<td>0.9609</td>
<td>0.9429</td>
</tr>
<tr>
<td>F-Test</td>
<td>2954</td>
<td>30236</td>
<td>20340</td>
</tr>
</tbody>
</table>
Tables 2.4 and 2.5 shows average absolute and relative efficiencies for Unit, Limit and State banks. Berger found that banks in statewide branching states (State) tend to be less efficient than banks in unit branching (Unit) or limited branching (Limit) states, where untruncated average efficiencies\textsuperscript{23} are 67.8\%, 57.6\% and 58.6\% for the three bank types. These differences can be attributed to the differing correlation structure and coefficient heterogeneity. Bauer, Berger, and Humphrey (1993) report a higher average efficiency, 93\% for random effects and 83.4\% and 84.7\% for the MLE estimates. English, et. al. (1993) report a mean efficiency of 75.4\% in their DEA estimates.\textsuperscript{24} Our estimates lie close to inefficiencies obtained by these and other methods (Berger 1993).\textsuperscript{25}

The similarity of our estimates with those of Berger (1993), Bauer, Berger, and Humphrey (1993) and English, et. al. (1993) leads us to believe that differences in correlation structure rather than in coefficient heterogeneity cause large differences in efficiency estimates.

Furthermore, the rankings of bank types by average efficiency parallels Berger’s (1993). According to Berger (1993), these results contradict the usual expectation that greater competition in statewide branching states results in higher efficiencies. Total average efficiency loss for a firm in each bank type lies at $4.98$, $38.28$, and $5.26$ million for Unit, State and Limit banks over the entire 10 year period. These figures lie close to those obtained in other papers (Berger 1993).

We now turn to our distributional results on efficiencies. The figures 2.1 and 2.2 show the empirical distribution of the absolute and relative efficiencies for unit, limit and state banks. We use a naïve kernel\textsuperscript{26} to estimate the density of the absolute

\textsuperscript{23}Berger truncates the efficiencies to allow for extremely lucky or unlucky firms. Also, we report his method 1, which assumes effects are uncorrelated with regressors.

\textsuperscript{24}DEA assumes that regressors are not correlated with effects.

\textsuperscript{25}We do not truncate the inefficiencies to account for extreme outliers.

\textsuperscript{26}A naïve kernel estimate is used, because we do not want to oversmooth the bounded relative efficiencies.
and relative efficiencies. A naïve kernel is based on a histogram and is defined for a kernel, K(t), as: $\frac{1}{2}$ for $|t| < 1$, 0 otherwise. As in Berger (1993), the data suggest a normal distribution rather than the usual half-normal assumption. Bauer, Berger, and Humphrey (1993) observe a distribution, which is consistent with the half-normal.

The Shapiro-Wilk variance ratio statistic is used to test for normality. The W statistic results range from 0 to 1, where values close to one indicate closeness to normality. Our numbers - 0.972, 0.975, 0.957, respectively for unit, limit, and state banks - indicate that the distributions are near normal, which is consistent with the results obtained in Park, Sickle, and Simar (1996) as T becomes large. We also check the skewness coefficient for the three samples, where positive values indicate skewness to the right. We observe values of 0.592, 0.529, and 0.864. This implies that all three distributions are skewed to the right as is the case in the half-normal assumption, where firms are grouped close to the most efficient firm. We use a Kolomogrov-Smirnoff test to test these distributions against folded half-normal, which resulted in rejection in all three cases.
Figure 2.1 Absolute Efficiencies in Semiparametric Efficient Regression

Figure 2.2 Relative Efficiencies in Semiparametric Efficient Regression
2.4.2 Nonparametric Regression-Based Estimation

We next apply our new nonparametric regression based estimator to the data assuming demand and savings deposits are inputs. We used a higher order normal kernel and determined the bandwidth using the method in section 6.1. The results are listed in the table 2.6:

The parameter estimates for the outputs are very close to those obtained in the estimation of the linearized form of the distance frontier. Standard deviations are slightly higher, noting a loss in estimation efficiency.

Absolute Efficiencies are derived in the following manner:

\[ \alpha_i = \hat{Y}_i - \hat{Y}_i^* \gamma \]

where \( \hat{Y}_it = Y_{it} - E[Y|X] \) and \( \hat{Y}_it^* = Y_{it}^* - E[Y^*|X] \). The results are similar to those obtained by the semiparametric efficient estimator in the linearized model. Again, the data suggest a normal distribution: 0.971, 0.968, 0.948 for unit, limit and state banks for the Kolmogorv-Smirnoff test. The skewness coefficient differs somewhat, where we observe values of 0.618, 0.488 and 0.868 accordingly. The distributions of the absolute efficiencies are now more similar to each other than those in the semiparametric efficient estimation case. An analysis of average relative efficiencies\(^{27}\) shows that, contrary to our previous results, unit banks are now on average less efficient than both state and limit bank. Average efficiencies range from 64% for

\begin{table}[h]
\centering
\caption{Nonparametric Regression Based Estimation}
\begin{tabular}{|c|c|c|}
\hline
\multicolumn{3}{|c|}{Estimated Slope Coefficients} \\
\hline
 & Unit Banks & Limit Banks & State Banks \\
\hline
clin & 0.3608 (0.0049) & 0.3189 (0.0049) & 0.3750 (0.0048) \\
lin & 0.2914 (0.0053) & 0.3275 (0.0062) & 0.2663 (0.0053) \\
\hline
\end{tabular}
\end{table}

\(^{27}\)See table 2.5.
unit banks to 76.5% for limit banks. Average efficiency for state and limit banks has increased, while unit bank average efficiency did not change. By allowing for a flexible functional form for the inputs, we are reducing possible measurement error that can occur due to model misspecification. The reduction in average efficiency in all bank types strengthens this notion of reducing measurement error. However, the fact that our efficiency scores did not greatly change strengthens the idea that the correlation structure is more relevant to the efficiency estimates. Total average amount of inefficiency is $5.00, $26.62, and $3.92 million for unit, state and limit banks over the entire 10 year period. Figures 2.3 and 2.4 show the empirical distributions.

2.5 Conclusion

In this study, I introduce a method to estimate efficiencies for multiple output and input firms using a stochastic distance frontier. The stochastic distance frontier has the advantage that it avoids possible endogeneity of prices, while still giving estimates of efficiencies economic interpretation, which are comparable to cost and profit frontiers.

I generalize the semiparametric estimator derived in Park, Sickles, and Simar (1996) to estimate a stochastic distance frontier in the banking industry, where a subgroup of the regressors are correlated with the effects. Furthermore, we introduced an extended semiparametric estimator based on a nonparametric regression, where no functional form is imposed on inputs. In both cases, our results on efficiencies were similar to those obtained in Berger (1993) and reveal a model assuming demand and savings deposits are inputs as a statistically dominate specification in banking production.

It does not appear that the model specification of the distance function has a direct affect on the efficiency estimates as opposed to other findings in the literature (Berger 1993, Bauer, Berger, and Humphrey 1993). A more flexible form like the
Figure 2.3 Absolute Efficiencies in Nonparametric Based Regression

Figure 2.4 Relative Efficiencies in Nonparametric Based Regression
nonparametric based estimator introduced in this paper shows no drastic change in efficiency estimates. For a more informative interpretation, this estimator needs to be integrated in a cost or profit frontier to better draw comparison to the literature, where it has been determined that a translog approximation of the cost function is problematic (Berger, 1993). Moreover, the same should be done for the semiparametric efficient estimator.
Chapter 3

Computation in Efficient Estimation

This chapter draws heavily on an article published by Adams, Berger, and Sickles (1997).

Recently in the efficiency literature, several semiparametric efficient methods have been developed that estimate the slope parameters efficiently while making minimal assumptions on the distribution of the effects and regressors. Park and Simar (1995) introduced a semiparametric efficient estimator which makes minimal assumptions on the distribution of the effects. Park, Sickles, and Simar (1996) extend the Park and Simar (1995) semiparametric efficient estimator to allow for different, more complex correlation patterns. Both of these articles center on the derivation of the semiparametric efficient estimator, but they do not discuss the computational methods used in estimation. This article extends the work in Park and Simar (1995) and Park, Sickles, and Simar (1996) by allowing for a subgroup of the regressors to be correlated with the effects in the spirit of Hausman and Taylor (1991) and by discussing the computational issues involved. The new estimator is then illustrated by estimating efficiency in the banking industry.

There has been a great deal of research interest in the topic of efficiency in the banking industry over the past several years (See Berger, Hunter, and Timme 1993 for a review). Banking represents an interesting topic of research for several reasons. First, banking has historically been a regulated industry with limited market entry, which might allow inefficiencies to persist over time. Second, the financial services produced by the banking industry create some interesting measurement issues. Specifically, the determination of outputs in the banking industry posses a problem for research. Deposits posses both input and output characteristics. Finally, a study
of the banking industry is of academic interest because all of the major efficiency measurement techniques have been applied to banking data over the last several years. In this study, we use panel data of over 25,000 observations of U.S. banks over a 10-year period.

I discuss the different computational methods available to estimate the semiparametric efficient estimator. We choose kernel-based methods to estimate the joint density between the effects and correlated regressors. Computation of the kernels involves choice of a kernel function as well as binwidth selection, which ultimately affects the convergence of the semiparametric efficient estimator. This paper discusses kernel choice and binwidth selection methods.

I model the stochastic distance frontier by modifying a semiparametric efficient estimator developed by Park, Sickles, and Simar (1996). We assume that product mix affects the efficiency of firms causing outputs to be correlated with effects. We then apply a model that makes minimal assumptions about the parametric form of the stochastic structure of firm inefficiencies.

In section 1, we describe the basic model to be estimated. In section 2, an efficient estimator in the class of estimators which do not impose distributional assumptions on the model is developed. This is done by deriving the asymptotic lower bound for the parameter estimates of interest in the presence of nuisance parameters, while allowing for dependencies between the random effects and the outputs. The efficient semiparametric estimator attains this lower bound. Section 3 discusses computation of the semiparametric estimator. Section 6 compares the results of the estimation. Results using a within estimator and Hausman-Taylor estimator are compared to the estimates of our model.
3.1 The Model

We start with the output distance function as a motivation for our model of a firm with multiple outputs and inputs. The output distance function provides a radial measure of technical efficiency that describes the fraction of aggregated outputs, given chosen inputs. For a particular observation i, the output distance function can be approximated by:

\[ D(x, y) = \frac{\prod_j y_{ji}}{\prod_k x_{ki}} \leq 1 \]  

(3.1)

where \( y_{ji} \) and \( x_{ki} \) are the levels of output j and input k. The \( q_j \) and the \( r_k \) are the weights which describe the technology of the firm. When a firm is producing efficiently or when the value of the distance function equals one, then it is not possible to increase the value of output without either decreasing another output or increasing inputs.

It is convenient to linearize (1) using logarithms of outputs and inputs and include error terms, which results in the following Cobb-Douglas form:

\[ 0 = \sum_j \gamma_j \ln y_{j, it} - \sum_k \beta_k \ln x_{k, it} + \alpha_i + \epsilon_{it} \]  

(3.2)

By construction, the output distance function is linearly homogeneous in outputs. Normalizing with respect to one \( y_j \) linear homogeneity results in the following transformation:\(^{28}\)

\[ -\ln(y_{J}) = \sum_j \gamma_j \ln \tilde{y}_{j, it} - \sum_k \beta_k \ln x_{k, it} + \alpha_i + \epsilon_{it} \]  

(3.3)

where \( \tilde{y} = (\tilde{y}_j) = (\frac{y_j}{y_J}), \ j = 1, \ldots, J - 1. \)

This is basically a generic panel data model where we estimate the firm specific efficiency levels in a stochastic panel production frontier model. The model fits into the class of production frontier models (Aigner, Lovell, and Schmidt 1977; Meeusen and van den Broeck 1977; Sickles and Schmidt 1984; Cornwell, Schmidt and Sickles 1990) and gives rise to the analysis of firm efficiencies (heterogeneities). For the basic framework of the model, we consider a panel frontier model with firm specific

\(^{28}\)See Lovell, Richardson, Travers and Wood (1995) for a complete discussion.
heterogeneities, derived from the distance function above. Let $X_{it} = -\ln x_{k,it}$, $Y_{it}^* = \ln \tilde{y}_{j,it}$ and $Y_{it} = -\ln (y_J)$:

$$Y_{it} = X'_{it} \beta + Y_{it}^* \gamma + \alpha_i + \epsilon_{it}$$

(3.4)

where I assume $N$ independent observations $(X_i, Y_i^*, Y_i)$ and $X_i = (X'_{i1}, \ldots, X'_{iT})$, $Y_i^* = (Y_{i1}^*, \ldots, Y_{iT}^*)$, $Y_i = (Y_{i1}, \ldots, Y_{iT})$. The $X_i$ are i.i.d. $k \times T$ dimensional random vectors with unknown density function $g$ and $\beta$ is a $k$ dimensional unknown vector. The $\epsilon_{it}$ are considered i.i.d. random variables from $N(0, \sigma^2)$ with unknown $\sigma^2$ and the $(\alpha_i, Y_i^*)$ are i.i.d. from unknown density $h$. It is important to note that the $\alpha$'s are orthogonal to the $X_i$, but not to $Y_i^*$. We assume the support of the marginal density of $\alpha$ is bounded above(below). Let $(\alpha, Y^*, X')'$ denote generic observations $(\alpha_i, Y^*_i, X'_{it})'$ and $\hat{Y}^* = T^{-1} \sum_{t=1}^{T} Y_{it}^*$.

Consider the joint distribution of $(\alpha, Y^*)'$. We assume that firm heterogeneities are correlated only with long run movements of $Y^*$, i.e. only with $\hat{Y}^*$. Thus, the density $h$ can be described in the following manner:

$$h(\alpha, Y^*) = h_M(\alpha, \hat{Y}^*)p(Y^*)$$

(3.5)

where $h_M$ represents the joint distribution of $\alpha$ and $\hat{Y}^*$ and $p(Y^*)$ represents an arbitrary function.

### 3.2 Efficient Estimation

I utilize the estimation procedure derived in Park, Sickles, and Simar (1996) and allow for multiple regressors to be correlated with the effects. In this estimation method, efficient estimation of the slope coefficients does not depend on the nuisance parameters, i.e. the nonparametric part of the model. In other words, the slope parameters can be efficiently estimated regardless of the nonparametric part of the model. For an excellent discussion on efficient estimation see Bickel, Klaassen, Ritov, and Wellner (1993) as well as Newey (1990) and Robinson (1988). For expositonal purposes, we present the estimator as an extension of Park, Sickles. and Simar (1996).
The derivation of the estimator follows along the same lines as in Park, Sickles, and Simar (1996).

We now will show the information bound and estimator for the model above. Let \( X' = (X_{i1}', \ldots, X_{iT}')' \), \( Y^{*,*} = (Y_{i1}^{*,*}', \ldots, Y_{iT}^{*,*}')' \) and \( Y_i = (Y_{i1}, \ldots, Y_{iT})' \) for the generic observations \((X, Y^{*,*}, Y)\). Let \( S_i(\theta) = Y_i - X_i\bar{\beta} - Y_i^{*,*}\gamma \) and \( U_i = S_i(\theta) - \bar{S}_i(\theta) \) where \( Y_i \) is the relative output and \( \bar{S}_i(\theta) = T^{-1} \sum_{t=1}^{T} S_i(\theta) \). Let

\[
\omega(\bar{s}, \bar{y}^{*}) = \int \phi_{\sigma}(\bar{s} - u)h_M(u, \bar{y}^{*})du = \phi_{\sigma}h_M(., \bar{y}^{*})(\bar{s})
\]

be the joint density of \((\bar{S}(\theta), \bar{Y}^{*})\), where \( \sigma = \frac{s}{\sqrt{T}} \) and \( \bar{Y}^{*} = T^{-1} \sum_{t=1}^{T} Y_i^{*,*} \). Let

\[
I_0 = \int \frac{(\omega')^2}{\omega}(\bar{s}, \bar{y}^{*})d\bar{s}d\bar{y}^{*}
\]

be the Fischer information for location of \( \omega \), where \( \omega' \) is the derivative with respect to the first component of the vector.

The efficient scores and information bound are obtained (see Theorem 3.5, Park, Sickles, and Simar (1996)).

\[
\ell_{\beta}^{*} = \sigma^{-2} \sum_{t=1}^{T} U_i X_t - (\bar{X} - \text{E}(\bar{X})) \left( \frac{\omega'}{\omega} \right) (\bar{S}, \bar{Y}^{*})
\]

\[
\ell_{\gamma}^{*} = \sigma^{-2} \sum_{t=1}^{T} U_i Y_t^{*,*}
\]

\[
I = \begin{pmatrix}
\sigma^{-2} \Sigma_W(X) + I_0 \Sigma_B(X) & 0 \\
0 & \sigma^{-2} \Sigma_W(Y^{*,*})
\end{pmatrix}
\]

where \( \bar{X} = T^{-1} \sum_{t=1}^{T} X_t \) and where \( \Sigma_W \) and \( \Sigma_B \) denote the within and between covariance operator.

\(^{29}\)The proof is the same as in Park, Sickles, and Simar (1996), except \( Y^{*,*} \) is a vector and \( I \) is block diagonal.
To construct an efficient estimator of \( \theta \) (Park, Sickles, and Simar (1996)).\(^{30}\) First, find a \( \sqrt{N} \)-consistent estimator \( \hat{\theta}_N \) of \( \theta \). In our model, we choose the Hausman-Taylor estimator. Second, consider \( \hat{\ell} \) as \( \hat{\ell}(x, y^*, y, \theta, \sigma^2, h) \) and construct a suitable estimator \( \hat{\ell}(\ldots, \theta; X_1, Y_1^*, Y_1, \ldots, X_N, Y_N^*, Y_N) \) of \( \hat{\ell}(\ldots, \theta, \sigma^2, h) \) using the computational methods discussed in the next section. The efficient estimator is then formed by including \( \hat{\ell} \) in a one step Newton-Rhapson iteration:

\[
\hat{\theta}_N = \tilde{\theta}_N + N^{-1} \sum_{i=1}^{T} \hat{\ell}(\ldots, \theta; X_1, Y_1^*, Y_1, \ldots, X_N, Y_N^*, Y_N)
\]

Specifically, construction of the nonparametric part of \( \hat{\ell} \) follows by estimating \( \omega \), the unknown distribution of the random effects and \( \bar{Y}_{i*} \). The efficient estimator of \( \omega \) is defined as:\(^{31}\)

\[
\omega(\bar{s}, \bar{y}^*, \theta) = N^{-1} \sum_{i=1}^{N} K_{sN}(\bar{s} - \bar{S}_i) \prod_{j=1}^{J} K_{sN}(\bar{y}_j^* - \bar{Y}_{ji}) + c_N
\]

where \( K_s = K(t/s)/s \). We use a higher order normal kernel defined as: \((1.5 - 0.5z^2)\phi(z)\), where \( \phi(z) \) is a typical normal kernel.\(^{32}\) We assume the same bandwidth and kernel for all variables and simplify the kernel by using a product kernel as opposed to a multivariate kernel.

The semiparametric efficient estimator is then defined by:

\[
\hat{\theta}_N = \tilde{\theta}_N + N^{-1} \hat{\ell}^{-1} \left[ \sum_{i=1}^{N} \left[ \sum_{t=1}^{T} \hat{\sigma}^{-2} \hat{U}_{it}X_{it} - (\bar{X}_t - \bar{X})(\frac{\omega^*}{\omega})(\bar{S}_i, \bar{Y}_i^*, \tilde{\theta}_N) \right] \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\sigma}^{-2} \hat{U}_{it}Y_{it}^* \right]
\]

where

\(^{30}\)See also Bickel, Klaassen, Ritov, and Wellner (1993).

\(^{31}\)See next section for a discussion.

\(^{32}\)Higher order kernels are used to insure proper convergence of the multidimensional kernel estimates.
\[
\hat{I} = \begin{pmatrix}
T\hat{\sigma}^{-2}\hat{\Sigma}_W(X) + \hat{\Sigma}_B(X) & 0 \\
0 & T\hat{\sigma}^{-2}\hat{\Sigma}_W(Y^*)
\end{pmatrix}
\]

\[
\hat{I}_0 = N^{-1} \sum_{i=1}^{N} \left( \frac{\hat{\omega}'}{\hat{\omega}} \right)^2 (\hat{S}_i, \bar{Y}_i, \hat{\theta}_N)
\]

where \( \hat{\theta}_N \) is the Hausman-Taylor estimator, and where \( \hat{\sigma}^2, \hat{U}_i \) and \( \hat{S}_i \) are used for \( \hat{\sigma}^2(\hat{\theta}_N), \hat{U}_i(\hat{\theta}_N), \hat{S}_i(\hat{\theta}_N) \).

Given the efficient estimator \( \hat{\theta}_{N,T} \), \( \alpha_i \) are predicted by:

\[
\tilde{\alpha}_i = \hat{S}_i(\hat{\theta}_{N,T})
\]

Relative technical inefficiencies of the i-th firm with respect to the j-th firm can be predicted by: \( \tilde{\alpha}_i - \tilde{\alpha}_j \). We normalize with respect to the most efficient firm: \( \max_{j=1,...,N}(\tilde{\alpha}_j) \). Relative technical efficiencies are then derived by taking the distance of each firm’s \( \alpha_i \) from the most efficient firm.

### 3.3 Computation and Smoothing

As shown in the previous section, estimates of the joint density of effects and a subgroup of the regressors are embedded in the semiparametric efficient estimator. Computation of this density should involve nonparametric procedure, so that a large class of densities can be covered and ensure that the true density is included. Several computational methods are available to nonparametrically compute the joint distribution. These methods include smoothing by convolution, orthogonal series approximation, and kernel estimators. We choose kernel estimators because of their generality (Scott 1992).

When using kernel estimators, two decisions arise: kernel choice and optimal binwidth selection. Kernel choice can be determined by finding any functional form that meets the necessary assumptions (sum to one, nonnegativity, moment restrictions.
etc.). Generally speaking, density functions fit these requirements. The literature (see Scott 1992, Silverman 1986) discusses optimal kernels. However, the selection of kernel made in this paper is based on the desire to improve convergence of the density estimator. Higher order kernels allow us to increase the convergence at the cost of violating the nonnegativity of the kernel estimate. The higher order kernel imposes restrictions on the higher moments of the kernel function. These restrictions can be generalized in the following manner:

$$\int K = 1, \int w^i K = 0 \quad i = 1, \ldots, p - 1 \quad \text{and} \quad \int w^p K \neq 0$$

where $p$ is the order of the kernel. It can be shown that these restrictions increase the convergence of the estimates as the order increases. This can be seen by considering the bias for a $p$th order kernel in the univariate case:

$$\text{BIAS}\{\hat{f}(x)\} = \frac{1}{p!} s^p \mu_p f^{(p)}(x)$$

which is $O(s^p)$. We use a higher order normal kernel ($p = 4$), because the dimensions of the kernel estimates have a direct bearing on the convergence of the semiparametric efficient estimator. Higher order kernels represent one possible method for dealing with the 'curse of dimensionality'.

A more important aspect of computation in the semiparametric efficient estimator is binwidth selection. Binwidth selection is important because the smoothing parameter determines the levels of bias and variance. At the optimal binwidth, the effects

---

33. The kernel estimator can be described as a minimization problem. One minimizes roughness of the estimates given several constraints. See Scott (1992).

34. We consider the univariate case here for expositional purposes. It is easy to generalize the bias for the multivariate case. See Scott (1992).

35. The $p = 3$ and $p = 4$ higher order normal kernels are equivalent, because the normal kernel is already symmetric.

36. We have applied the normal kernel to the estimator. The higher order kernels improved the results giving evidence to the dimensionality problem. See Scott (1992) for a discussion of other possible methods of handling the 'curse of dimensionality'.
bias and variance are equalized and mean integrated squared error of the estimates is minimized.

Several methods of optimal binwidth selection are available. These methods include plug in methods, biased cross validation, unbiased cross validation, and bootstrapped cross validation, as well as the bootstrap method proposed by Park and Simar (1995). These methods differ in their approaches to approximate the roughness of the density estimate, where Park and Simar's method centers on the estimation of the slope parameters. The roughness of the density estimate is defined as: \( R(f) = \int f^2(x) dx \). In this paper, we discuss plug in methods and compare them to the bootstrap method of Park and Simar. We do not consider the other methods because the theory regarding these methods in the multivariate case is incomplete in the literature.\(^\text{37}\)

Plug in methods represent a computation free method of binwidth selection. They generally are calculated by making additional assumption about the distribution of the data to find the optimal binwidth. For multivariate product kernels, the binwidth value that asymptotic minimizes mean integrated squared for the bivariate case is:

\[
s^*_i = s_i (1 - \rho^2)^{\frac{1}{12}} (1 + \frac{\rho^2}{2})^{-\frac{1}{6}} n^{-\frac{1}{6}}
\]

where \( \rho \) is the correlation between the variables and \( s \) differs across variables. One can see that the optimal binwidth depends on the correlation between variables in higher dimensional cases. One plug in method can be derived by assuming independence of the variables and by applying a normal kernel. The 'normal reference rule' is described by the following:

\[
s^*_i = \left( \frac{4}{d + 2} \right)^{\frac{1}{2(d+1)}} \sigma_i n^{-\frac{1}{(d+1)}}
\]

\(^{37}\text{See Scott (1992).}\)
where \( d \) is the dimension of the kernel and \( n \) is the sample size. Silverman (1986) suggests a more simplified version of this plug in method. Simplifying the kernel to have the same binwidth across all variables, we get the following plug in method:

\[
 s_{opt} = \left( \frac{4}{(2d + 1)} \right)^{\frac{1}{(d+1)}} \sigma n^{-\frac{1}{(d+1)}} \]

where \( \sigma = d^{-1} \sum_i \sigma_i \) is the average marginal variance. Even though these methods make additional assumptions on the data to select the optimal binwidth, they do simplify computation of the joint density and can be considered an approximation for the binwidth. We use the plug in method that accounts for the higher order kernel. This plug in method is derived in the similar manner as above:

\[
 s_{opt} = \left( \frac{4}{(2d + p)} \right)^{\frac{1}{(d+2+p)}} \frac{1}{n^{\frac{1}{(d+2+p)}}} \]

where \( p \) is the order of the kernel. This method is derived from the Taylor series expansion of the higher order kernel.

Park and Simar (1995) suggest the following binwidth selection method for their semiparametric efficient estimate, \( \hat{\theta} \). The bootstrap choice of \( s^* = \arg_s \min C(s) \):

\[
 C(s) = \left( \frac{1}{M} \right) \sum_{m=1}^{M} \left[ \hat{\theta}_{N,T}^{(m)}(s) - \hat{\theta}_{N,T}(s) \right]' \left[ \hat{\theta}_{N,T}^{(m)}(s) - \hat{\theta}_{N,T}(s) \right]
\]

where \( \hat{\theta}_{N,T}^{(m)}(s) \) denotes the m-th pseudo sample bootstrap version of \( \hat{\theta}_{N,T}(s) \) using binwidth \( s \). This binwidth selection criteria is geared towards optimizing the semiparametric efficient estimator. The method centers on computing the most efficient results for the slope parameters. We are interested in the slope parameters, given the fact that the distribution of the effects and a subgroup of the regressors is unknown.
3.4 Results

As an illustration of the estimator and computational methods, we apply the semi-parametric efficient estimator to a banking data set used in Berger (1993) and Akh-vein, Swamy, Taubman (1996). The data are separated into three groups according to branching regulations: statewide branching (State), limit branching (Limit), and no branching (Unit). We use a Cobb-Douglas stochastic panel distance function (Schmidt and Sickles 1984; Cornwell, Schmidt and Sickles 1990) for estimation, where \( y = \ln(\text{real estate loans}) \); \( X = \ln(\text{labor}), \ln(\text{capital}), \ln(\text{purchased funds}), \ln(\text{demand deposits}), \ln(\text{retail time and savings deposits}), \text{size}; \ Y^* = \ln(\text{commercial and industrial loans}), \ln(\text{installment loans}) \). Along the lines of the asset approach to output determination (assets are outputs), we assume demand and savings deposits are inputs.

A Hausman test for orthogonality, which tests the performance of within estimation relative to random effects and thus the correlation assumptions of the firm heterogeneities, was performed on each of the data sets. In each test, the null hypothesis of no correlation was rejected at the 95% level.

The efficient semiparametric estimator of \( \theta \) is based on a bandwidth \( s \) chosen using a specification defined in Silverman (1986) in the following manner: \( s_{opt} = \left( \frac{4}{(2d+p)} \right)^{\frac{1}{d+2p^2}} n^{\frac{d+2p^2}{2}} \) where \( d \) is the dimension of the kernel estimates, \( p \) is the order of the kernel and \( n \) is sample size. We use this specification for optimal bandwidth as opposed to the method in Park and Simar (1995) and Park, Sickles. and Simar (1996) for its simplicity. The optimal binwidths for Unit, Limit, and State banks are: 0.5104, 0.5178, 0.5141. 39

---

38 A detailed description of the data can be found in Berger (1993).
39 Park and Simar's (1995) and Park, Sickles, and Simar's (1996) method was applied to the data with similar end results. The method finds a minimum in the same regions as the plug in method.
The within, Hausman-Taylor and efficient semiparametric estimates and standard errors for the stochastic panel distance function are presented in tables 3.1, 3.2, 3.3 for unit, limit, and state banks.

We now move on to describe our technical efficiency results obtained from the semiparametric efficient estimation. Figure 3.1 shows absolute technical efficiencies for the three bank types and figure 3.2 show relative technical efficiencies. We use a na"ive kernel\textsuperscript{40} to estimate the density of the relative and absolute efficiencies. A na"ive kernel is based on a histogram and is defined for a kernel, K(t), as: \( \frac{1}{2} \) for \( |t| < 1 \), 0 otherwise. Table 3.4 shows means and variances for absolute and relative technical efficiencies. Kolomogrov-Smirnoff tests against normality indicate that the distributions of the absolute efficiencies are close to normality.

Average relative technical efficiency ranges from 61\% to 68.5\%, where limit banks are on average more efficient than unit or state banks. Table 4 shows average efficiencies for Unit, Limit and State banks. These results are similar to Berger's, who found that banks in statewide branching states (state) tend to be less efficient than banks in unit branching (unit) or limited branching (limited) states, where untrun-
### Table 3.2  Limit Banks Regressions

<table>
<thead>
<tr>
<th></th>
<th>Estimated Slope Coefficients</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within</td>
<td>Hausman-Taylor</td>
<td>Semiparametric Efficient</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-0.1943 (0.0170)</td>
<td>-0.2009 (0.0161)</td>
<td>-0.1879 (0.0132)</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-0.2237 (0.076)</td>
<td>-0.2119 (0.0073)</td>
<td>-0.2356 (0.0061)</td>
<td></td>
</tr>
<tr>
<td>Purf</td>
<td>-0.0703 (0.0046)</td>
<td>-0.0753 (0.0043)</td>
<td>-0.0678 (0.0035)</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>-0.2864 (0.0108)</td>
<td>-0.2848 (0.0104)</td>
<td>-0.2920 (0.0088)</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>-0.3973 (0.0127)</td>
<td>-0.4058 (0.0121)</td>
<td>-0.3948 (0.0099)</td>
<td></td>
</tr>
<tr>
<td>clin</td>
<td>0.3245 (0.0047)</td>
<td>0.3250 (0.0047)</td>
<td>0.3250 (0.0044)</td>
<td></td>
</tr>
<tr>
<td>inin</td>
<td>0.3474 (0.0061)</td>
<td>0.3484 (0.0060)</td>
<td>0.3484 (0.0059)</td>
<td></td>
</tr>
<tr>
<td>size</td>
<td>-0.0017 (0.0074)</td>
<td>0.0510 (0.0079)</td>
<td>0.0477 (0.0066)</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.9469</td>
<td>0.9461</td>
<td>0.9460</td>
<td></td>
</tr>
<tr>
<td>F-Test</td>
<td>20274</td>
<td>19985</td>
<td>19706</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.3  State Banks Regressions

<table>
<thead>
<tr>
<th></th>
<th>Estimated Slope Coefficients</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Within</td>
<td>Hausman-Taylor</td>
<td>Semiparametric Efficient</td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>-0.1464 (0.0160)</td>
<td>-0.1561 (0.0149)</td>
<td>-0.1292 (0.011)</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>-0.1383 (0.0097)</td>
<td>-0.1316 (0.0090)</td>
<td>-0.1438 (0.0068)</td>
<td></td>
</tr>
<tr>
<td>Purf</td>
<td>-0.1223 (0.0049)</td>
<td>-0.1258 (0.0047)</td>
<td>-0.1226 (0.0034)</td>
<td></td>
</tr>
<tr>
<td>DD</td>
<td>-0.3076 (0.0107)</td>
<td>-0.3019 (0.0102)</td>
<td>-0.3211 (0.0081)</td>
<td></td>
</tr>
<tr>
<td>OD</td>
<td>-0.4433 (0.0102)</td>
<td>-0.4428 (0.0098)</td>
<td>-0.4548 (0.0074)</td>
<td></td>
</tr>
<tr>
<td>clin</td>
<td>0.3895 (0.0048)</td>
<td>0.3901 (0.0048)</td>
<td>0.3901 (0.0042)</td>
<td></td>
</tr>
<tr>
<td>inin</td>
<td>0.2780 (0.0053)</td>
<td>0.2778 (0.0053)</td>
<td>0.2778 (0.0050)</td>
<td></td>
</tr>
<tr>
<td>size</td>
<td>0.0581 (0.0071)</td>
<td>0.0605 (0.0067)</td>
<td>0.0650 (0.0051)</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.9600</td>
<td>0.9609</td>
<td>0.9435</td>
<td></td>
</tr>
<tr>
<td>F-Test</td>
<td>29554</td>
<td>30236</td>
<td>11126</td>
<td></td>
</tr>
</tbody>
</table>

### Table 3.4  Average Efficiencies for Unit, State and Limit Banks

<table>
<thead>
<tr>
<th>Bank Type</th>
<th>Average Efficiency</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Relative</td>
<td>Absolute</td>
<td></td>
</tr>
<tr>
<td>Unit</td>
<td>0.644 (0.051)</td>
<td>0.940 (0.107)</td>
<td></td>
</tr>
<tr>
<td>State</td>
<td>0.611 (0.067)</td>
<td>1.28 (0.140)</td>
<td></td>
</tr>
<tr>
<td>Limit</td>
<td>0.685 (0.0595)</td>
<td>1.01 (0.1026)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.1  Absolute Efficiencies in Semiparametric Efficient Regression

Figure 3.2  Relative Efficiencies in Semiparametric Efficient Regression
cated average efficiencies\(^{41}\) are 67.8%, 57.6% and 58.6% for the three bank types. These differences can be attributed to the differing correlation structure and coefficient heterogeneity. Furthermore, the rankings of bank types by average efficiency parallels Berger’s (1993). According to Berger (1993), these results contradict the usual expectation that greater competition in statewide branching states results in higher efficiencies.

3.5 Conclusion

In this study, I extend the semiparametric estimator derived in Park, Sickles, and Simar (1996) to allow for a subgroup of regressors to be correlated with the effects. We discussed the computational issues involved with estimation using a semiparametric efficient estimator. These issues in the case of kernel estimation involve kernel choice and binwidth selection. Both of which effect the convergence of the estimates. Our study shows that the plug in methods can be effective in estimating the kernels for the semiparametric efficient estimator.

The estimator is illustrated using a stochastic panel distance frontier in the banking industry, which allows for multiple outputs and inputs. Our results on efficiencies were similar to those obtained in Berger (1993).

---

\(^{41}\)Berger truncates the efficiencies to allow for extremely lucky or unlucky firms.
Chapter 4

Efficient Distribution Free Estimation of Panel Models

Semiparametric efficient estimation has been extensively used in the statistics and econometrics literature. Bickel (1982), Newey (1990), Bickel, Klaassen, Ritov, and Wellner (1992) as well as others have considered semiparametric efficient methods and examples. In their article on semiparametric efficient estimation, Park and Simar (1995) introduce a semiparametric efficient estimator for the specific problem of a panel data model, where the distribution of the firm specific heterogeneity is unknown. In the derivation of their estimator, they assume normality of the symmetric error as well as independence of the regressors and effects. Park, Sickles, and Simar (1996) extended their model in that they allowed a regressor to be correlated with the effects and explore the impacts of various correlation patterns among effects and regressors on the form of the semiparametric efficient estimator. Adams, Berger, and Sickles (1996) allow for correlation between a subgroup of the regressors and the effects. The distributional assumptions, normality of the symmetric error and independence of the effects and regressors, have a direct bearing on the form of the efficient score and information bound (the center pieces of the estimator). Furthermore, these distributional assumptions allow the authors to concentrate on the unknown distribution of the effects and also draw similarities to other estimators, such as the within estimator.  

The distributional assumptions, normality and especially independence, are important for derivation of the semiparametric efficient estimator, because they deter-

---

42 Park, Sickles, and Simar (1996) show that within is efficient when all regressors are correlated with the effects.
mine the form of the efficient score through the likelihood function or through the nuisance parameter space. A change in these assumptions results in a change in the log likelihood function and nuisance parameter space and, hence, in the semiparametric efficient estimator. In this paper, we generalize the semiparametric efficient estimator derived by Park and Simar (1995) and by Adams, Berger, and Sickles (1996) in that the distribution of the symmetric error as well as the effects are not parameterized.

In Section 1, I outline the general panel model. In Section 2, I derive our semiparametric efficient estimator, allowing for free distributions in both symmetric errors and effects. Section 3 considers monte carlo results and illustrates the estimators with an application using banking data. Section 4 concludes.

4.1 Model and Statistical Assumptions

We consider the following model which assumes N independent observations \((X_i, Y_i)\):

\[ Y_{it} = X_{it} \theta + \alpha_i + \varepsilon_{it} \]  
(4.1)

where \(X_i = (X'_{i1}, \ldots, X'_{iT})'\) and \(Y_i = (Y_{i1}, \ldots, Y'_{iT})'\). Each \(X_{it}\) is a d-dimensional random vector, \(X_i\) are iid dT-dimensional random vectors with unknown density function \(g\), and \(\theta\) is a d-dimensional unknown vector. The \(\varepsilon_{it}\) are iid from an unknown density. We assume \(X_{it}\) and \(\alpha_i\) are independent of the symmetric error, \(\varepsilon_{it}\). The \(\alpha_i\) are iid from an unknown density. An additional assumption is needed to facilitate the separation of between and within residual densities: \(\varepsilon_{it} - \bar{\varepsilon}_i\) is independent of \(\alpha_i + \bar{\varepsilon}_i\).

To place our estimator into a particular modeling scenario we also assume that the support of the \(\alpha_i\) bounded above (below). This is the well-known stochastic panel frontier model (Schmidt and Sickles, 1984) motivated from the problem of measuring production inefficiency.\(^{44}\)

\(^{43}\)See also Park, Sickles, and Simar (1996).

\(^{44}\)See Adams, Berger, and Sickles (1996) for motivation in the multiple output case using the output distance function.
4.2 Efficient Estimation

4.2.1 A Semiparametric Efficient Estimator of a Panel Model

The notion of efficient estimation in semiparametric models is discussed in detail in Bickel (1982), Begun, Hall, Huang and Wellner (1983), Newey (1990), Bickel, Klaassen, Ritov, and Wellner (1992) and Pagan and Ullah (1995). We refer the reader to these articles for a more detailed description.

The first step in deriving the semiparametric efficient estimator is the derivation of the efficient score. Let \( L(X,Y,\theta, \eta) \) be the loglikelihood function \( \ell_\theta(X,Y) = \frac{\partial L}{\partial \theta} \) and \( \ell_\eta(X,Y) = \frac{\partial L}{\partial \eta} \), the scores with respect to the slope parameters, \( \theta \), and the nuisance parameters, \( \eta_j \), i.e. the unspecified parameters in the model. The efficient score is defined by

\[
\ell^*_\theta = \ell_\theta - \Pi(\ell_\theta | \ell_\eta) \tag{4.2}
\]

\( \ell^*_\theta \) is called the efficient score with respect to \( \theta \). \( [\ell_\eta] \) is the linear span generated by the \( \ell_\eta \)'s and \( \Pi(\ell_\theta | \ell_\eta) \) is the projection of \( \ell_\theta \) onto the linear span. The next step in semiparametric efficient estimation is the construction of the information bound. The information bound for \( \theta \) is given by:

\[
I(\theta) = \text{E} \ell''(X,Y) \tag{4.3}
\]

An estimator \( \hat{\theta}_N \) is called efficient if as \( N \to \infty \),

\[
\sqrt{N}(\hat{\theta}_N - \theta) \to N(0, I^{-1}(\beta))
\]

To show the efficient score and information bound for the model, let \( Y = (Y_1, \ldots, Y_T)' \) and \( X = (X_1', \ldots, X_T')' \) for the generic observations \((X,Y)\). Let \( S_t(\theta) = Y_t - X_t'\theta \) and \( U_t(\theta) = S_t(\theta) - \bar{S}(\theta) \), where \( \bar{S}(\theta) = T^{-1} \sum_{t=1}^{T} S_t(\theta) \). Let

\[
\omega(\bar{s}) = \ln \int f_\theta(\bar{s} - u)h(u,X)du
\]

be the density of \( \bar{S}(\theta) \) and \( X \). Define total and between variance as:

\[
\Sigma_T = ET^{-1} \sum_{t=1}^{T} \left[ [X_t - \bar{X}] - \text{E}[X - \bar{X}] \right]' \left[ [X_t - \bar{X}] - \text{E}[X - \bar{X}] \right]
\]

\[
\Sigma_B = ET^{-1} \sum_{t=1}^{T} \left[ [X_t - \bar{X}] - \text{E}[X - \bar{X}] \right]' \left[ [X_t - \bar{X}] - \text{E}[X - \bar{X}] \right]
\]
and
\[ \Sigma_B = ET^{-1} \sum_{t=1}^{T} (X_t - E(\bar{X}))(X_t - E(\bar{X}))' \]
where \( \bar{X} = T^{-1} \sum_{t=1}^{T} X_t \). It is assumed that \( I_0 < \infty \) and that both \( \Sigma_T \) and \( \Sigma_B \) exist and are nonsingular.

Under the assumption of regularity conditions as described in Park and Simar (1995) and Park, Sickles, and Simar (1996), we get the following theorem for the efficient score and information, where we drop the argument \( \beta \) in \( U_i(\beta) \) and \( \bar{S}_t(\beta) \):

The efficient score function and the information bound for estimating \( \beta \) for the panel model above are given by \(^{45}\)

\[ \ell_\theta = \left[ (X - E(X)) - (\bar{X} - E(\bar{X})) \right] f'(U_i) \]
\[ - \left[ \bar{X} - E(\bar{X}) \right] \frac{w'}{w}(\bar{S}, X) \quad (4.4) \]

and
\[ I = \Sigma_T \left( \frac{f'}{f} \right)^2 + \Sigma_B \left( \frac{w'}{w} \right)^2 \quad (4.5) \]

To construct an efficient estimator of \( \theta \), define the efficient influence function as \( \hat{\ell} = I^{-1} \ell \). \(^{46}\) The efficient influence function has the following properties: \( E\hat{\ell} = 0 \) and \( E\hat{\ell}\hat{\ell}' = I^{-1} \). The efficient estimator is then derived by using a one step Newton-Rhapson iteration, where the preliminary estimator needs to be \( \sqrt{N} \)-consistent. From the preliminary estimator and an estimate of the efficient influence function, the semiparametric efficient estimator has the following form:

\[ \hat{\theta}_N = \bar{\theta}_N + N^{-1} \sum_{i=1}^{N} \hat{\ell}(X_i, Y_i, \bar{\theta}; X_i, Y_i, \cdots, X_N, Y_N) \]

where \( \hat{\ell} \) is an estimate of the efficient influence function, \( \hat{\ell} \).

\(^{45}\)The proof can be found in Appendix A.

\(^{46}\)See Newey (1990) and Bickel, Klaassen, Ritov, and Wellner (1993).
As a preliminary estimator, Park and Simar (1995) suggest the within estimator obtained from using OLS on a transformed model of mean deviations or

$$\hat{\theta}_N = (NT\hat{\Sigma}_W)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)(Y_{it} - \bar{Y}_i)$$

where $\hat{\Sigma}_W = (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} (X_{it} - \bar{X}_i)(X_{it} - \bar{X}_i)'$ and is an estimate of $\Sigma_W$. It is easy to show that the within estimator is $\sqrt{N}$-consistent. Park, Sickles, and Simar (1995) show that within is efficient, when all regressors are correlated with the effects. It is important to note that other $\sqrt{N}$-consistent estimators can be used. For example, Park, Sickles, and Simar (1996) and Adams, Berger, and Sickles (1996) both use the Hausman-Taylor estimator on a generalized form of this estimator that allows a subgroup of regressors to be correlated with the effects.

Define the estimates of $\Sigma_T$ and $\Sigma_B$ as:

$$\hat{\Sigma}_T = ET^{-1} \sum_{i=1}^{T} [(X - E(X)) - (\bar{X} - E(\bar{X}))]' [X - E(X) - (\bar{X} - E(\bar{X}))]$$

and

$$\hat{\Sigma}_B = \sum_{i=1}^{N} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'/N$$

where $\bar{X} = \sum_{i=1}^{N} \sum_{t=1}^{T} \frac{X_{it}}{NT}$. Given the efficient score and information bound from theorem 2.1, the efficient estimator is now defined by:

$$\hat{\theta}_N = \hat{\theta}_N + N^{-1} \hat{\Sigma}^{-1} \sum_{i=1}^{N} \left[ [(X - E(X)) - (\bar{X} - E(\bar{X}))] \frac{\hat{f}'}{\hat{f}} (U_i) - [\bar{X} - E(\bar{X})] \frac{\hat{\omega}'}{\hat{\omega}} (\hat{S}) \right]$$

(4.6)

where

$$\hat{\Sigma} = \hat{\Sigma}_T \left( \frac{\hat{f}'}{\hat{f}} \right)^2 + \hat{\Sigma}_B \left( \frac{\hat{\omega}'}{\hat{\omega}} \right)^2$$

The $\frac{\hat{f}'}{\hat{f}}$ and $\frac{\hat{\omega}'}{\hat{\omega}}$ are estimated using a kernel estimator as described in Park and Simar (1995) as well as Adams, Berger, and Sickles (1997). These kernel estimates are then inserted into the estimator above.\(^{47}\)

\(^{47}\)Adams, Berger, and Sickles (1997) analyze the impact of kernel functions and of variance of optimal bandwidth selection criteria on the semiparametric estimator.
Park, Sickles, and Simar (1996) show that within is efficient, when all of the regressors are correlated with the effects.\footnote{This result makes sense, because OLS is MLE under the assumption of normality of the residuals. It is also efficient.} We drop the normality assumption and we find within is no longer efficient. Our results show that the semiparametric efficient estimator changes as the distribution of the symmetric residuals moves away from normality.

4.2.2 A Hausman-Taylor Estimator

We next consider a new model, where a subgroup of the regressors is correlated with the effects. The model can be written as:\footnote{See Adams, Berger, and Sickles (1996) for a motivation of the model. Park, Sickles, and Simar (1996) consider a similar case.}

\[ Y_{it} = X_{it}\beta + Z_{it}\gamma + \alpha_i + \varepsilon_{it} \] (4.7)

where \( X_{it} \) are independent of the symmetric error and conditionally independent of the effects. \( Z \) and \( \alpha_i \) are independent of the symmetric error term, but their joint distribution with the effects is unknown. As a starting point, we assume that effects are correlated in long run movements of \( Z_{it} \), i.e. only with \( \bar{Z} \). Thus, the density of \( \alpha, X, Z \) is:

\[ f(\alpha, X, Z) = h(\alpha, Z)g(X|Z) \] (4.8)

Given the new model and assumptions, we get, \textit{mutis mutandis}, the following theorem:

The efficient score function and the information bound for estimating \( \beta \) and \( \gamma \) for the panel model above are given by\footnote{The proof can be found in Appendix A.}

\[
\ell_{\beta}^* = \left[ X - E(X) - \left( X - E(\bar{X}) \right) \right] \frac{f'}{f}(U_i) - \left[ \bar{X} - E(\bar{X}) \right] \frac{w'}{w}(\bar{S}, Z)
\] (4.9)
\[ \xi_t = \left[ [Z_t - \bar{Z}] - \mathbb{E} [Z - \bar{Z}] \right] \frac{f'}{f}(U_t) \]

and

\[ I = \begin{bmatrix} \Sigma_T \left( \frac{f'}{f} \right)^2 + \Sigma_B \left( \frac{w'}{w} \right)^2 & \Sigma_XZ \left( \frac{f'}{f} \right)^2 \\ \Sigma_XZ \left( \frac{f'}{f} \right)^2 & \Sigma_TZ \left( \frac{f'}{f} \right)^2 \end{bmatrix} \]

where

\[ \Sigma_TZ = ET^{-1} \sum_{t=1}^{T} [(Z - \mathbb{E}(Z)) - (\bar{Z} - \mathbb{E}(\bar{Z})))'[ (Z - \mathbb{E}(Z)) - (\bar{Z} - \mathbb{E}(\bar{Z}))] \]

and

\[ \Sigma_XZ = ET^{-1} \sum_{t=1}^{T} [(X - \mathbb{E}(X)) - (\bar{X} - \mathbb{E}(\bar{X})))'[ (X - \mathbb{E}(X)) - (\bar{X} - \mathbb{E}(\bar{X}))] \]

### 4.3 Results

We conducted monte carlo experiments based on model 2 using regressors and correlation structures similar to those in our empirical illustration. We introduce nonnormal error, uniform distribution, for the symmetric error term. In these experiments, we vary N and T from 10 to 300 and utilize higher order bias reducing kernels. There is no evidence of bias with sample sizes of N = 10 and T = 10. Root mean squared error and absolute mean deviations based on 500 simulations range around 5% of the value of the true parameter estimates. These fall to between 3.0% to 1.5% for N = 100 and T = 10.

To illustrate both estimators, we estimate the models using a banking panel data set described in Adams, Berger, and Sickles (1996) as well as in Berger (1993). The data set consists of about 2500 banks from 1980-1989. It is divided into three subsamples based on different regulatory environments: statewide branching (State), limited branching (Limit), and no branching (Unit). These samples contain 792 - 933 banks each. The variables\(^51\) found in the data are defined as follows: \( Y = \) real estate loans; \( X = \) labor, capital, purchased funds, size, demand deposits, and retail time and savings deposits; \( Z = \) commercial and industrial loans, installment loans. We refer the

---

\(^{51}\) All variables are in logarithm form.
reader to Adams, Berger, and Sickles (1996) as well as Lovell, Richardson, Travers, and Wood (1990) for motivation using an output distance function and proper transformation of the variables. Y and Z represent outputs, while X represents inputs.\footnote{52}{We use the asset approach to output determination. See Berger (1993) and Adams, Berger, and Sickles (1996) for a discussion of the different approaches to output determination.} Following Adams, Berger, and Sickles (1996), we assume that long run movements in Z are correlated with the effects.

Binwidth selection becomes more important in the generalized semiparametric efficient estimator, because of the nonparametric form of the efficient score and information bound. This is especially true for those regressors that are independent of the effects, where two kernel estimates are embedded in the efficient score and information bound. Plug-in rules discussed in Adams, Berger, and Sickles (1997) and Scott (1992) are somewhat successful. We use a mixture of methods to determine optimal binwidth. The binwidth for the multivariate kernel of the density of $(\bar{S}, Z)$, w, is selected by the plug-in method in Adams, Berger, and Sickles (1996).\footnote{53}{This is a simplification and other methods can be applied.} The binwidth for the density of $U_i$, f, is chosen by the method suggested in Park and Simar (1995) and Park, Sickles, and Simar (1996). In other words, the optimal $h$ is found by bootstrapping the following: $h^* = \arg\min_h C(h)$ where $C(h) = \left( \frac{1}{M} \right) \sum_{m=1}^{M} \left[ \hat{\theta}^{(m)}_{N,T}(h) - \hat{\theta}^*_N,T(h) \right] \left[ \hat{\theta}^{(m)}_{N,T}(h) - \hat{\theta}^*_N,T(h) \right]$ and $\hat{\theta}^{(m)}_{N,T}(h)$ denotes the m-th pseudo sample bootstrap version of $\hat{\theta}^*_N,T(h)$ using binwidth $h$. Our estimates are based on $M = 1000$ and consisted of a grid search in the interval [0.1, 2.0]. The optimal binwidth for $h^*$ was 0.44 for unit banks, 0.52 for state banks, and 0.48 for limit banks. We then reestimated the $\theta$ using the optimal binwidth on the original data.

We estimated the model using within and the generalized semiparametric efficient estimator (case 2). Tables 4.1, 4.2, 4.3 show the results. The generalized efficient estimator improves on the IV estimator, within, in estimation efficiency. Variance
of the point estimates fall by as much as 50%. Using the semiparametric efficient estimator in Adams, Berger, and Sickles (1996), which assumes normality, we test the appropriateness of the normality assumption by using a Hausman test.

4.4 Appendix A

4.4.1 A Semiparametric Efficient Estimator of a Panel Model

With the notation and assumptions made in section 5.1, the density of \((\epsilon_{it}, \alpha_i, X_i)\) can be described in the following manner:

\[
 f(\alpha_i, \epsilon_{it}, X) = f_w(\epsilon_{it} - \tilde{\epsilon}_i) h(\alpha_i + \tilde{\epsilon}_i, X) \\
= f_w(U_i(\beta)) h(\bar{S}(\beta), X)
\]

The log likelihood function is:

\[
\mathcal{L}(Y, X, \alpha, \epsilon; \theta, f_w, f_b) = \ln f_w(U_i(\beta)) + \ln \int f_b(\bar{S}(\beta) - u) h(u, X) du
\]

Let \(w = \ln \int f_b(\bar{S}(\beta) - u) h(u, X) du\) be the joint density of \((\bar{S}(\beta), X)\) on \(\mathbb{R}^{1+Td}\) and \(f = \ln f_w(U_i(\beta))\). Furthermore, \(w' = \frac{\partial}{\partial \bar{s}} w(\bar{s}, x)\) and \(f' = \frac{\partial}{\partial U_i} f(U_i)\). The score with respect to the parameters is:

\[
\ell_\beta = -(X - \bar{X}) \frac{f'}{f}(U_i(\beta)) - \bar{X} \frac{w'}{w}(\bar{S}(\beta)) 
\]  

Table 4.1 State Bank GSPE

<table>
<thead>
<tr>
<th></th>
<th>Within S.E.</th>
<th>GSPE S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.146</td>
<td>-0.136</td>
</tr>
<tr>
<td>Capital</td>
<td>-0.138</td>
<td>-0.146</td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.122</td>
<td>-0.117</td>
</tr>
<tr>
<td>Size</td>
<td>0.038</td>
<td>0.050</td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.308</td>
<td>-0.308</td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.443</td>
<td>-0.442</td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.390</td>
<td>0.391</td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.278</td>
<td>0.275</td>
</tr>
</tbody>
</table>
Table 4.2  Limit Bank GSPE

<table>
<thead>
<tr>
<th>Limit Banks</th>
<th>Within S.E.</th>
<th>GSPE S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.174</td>
<td>-0.194</td>
</tr>
<tr>
<td>Capital</td>
<td>-0.294</td>
<td>-0.226</td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.051</td>
<td>-0.070</td>
</tr>
<tr>
<td>Size</td>
<td>0.039</td>
<td>0.047</td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.313</td>
<td>-0.286</td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.345</td>
<td>-0.367</td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.333</td>
<td>0.325</td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.360</td>
<td>0.287</td>
</tr>
</tbody>
</table>

Table 4.3  Unit Bank GSPE

<table>
<thead>
<tr>
<th>Unit Banks</th>
<th>Within S.E.</th>
<th>GSPE S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.214</td>
<td>-0.213</td>
</tr>
<tr>
<td>Capital</td>
<td>-0.136</td>
<td>-0.137</td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.088</td>
<td>-0.102</td>
</tr>
<tr>
<td>Size</td>
<td>-0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.276</td>
<td>-0.271</td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.367</td>
<td>-0.366</td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.360</td>
<td>0.363</td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.290</td>
<td>0.285</td>
</tr>
</tbody>
</table>

Now, we move onto the derivation of the efficient score. Define the nuisance parameter tangent space as:

\[ V_1 = \{a(\tilde{S}(\beta), X); a \in L_2(p), Ea(\tilde{S}(\beta), X) = 0\} \]

\[ V_2 = \{b(U(\beta)); b \in L_2(p), Eb(U(\beta)) = 0\} \]

Lemma [A.1.]

\[ E \left[ \frac{w}{w'}(\tilde{S}(\beta))|X, U_t \right] = 0 \]

Proof: The conditional pdf of \( \tilde{S}(\beta) \) given \( X=x \) is given by

\[ \int f_\beta(\tilde{S}(\beta) - u)h(u|X)du \]

where \( h(u=x) \) denotes the conditional pdf of a given \( X=x \). We can write

\[ E \left[ \frac{w}{w'}(\tilde{S}(\beta))|X, U_t \right] = E \left[ \frac{w}{w'}(\tilde{S}(\beta))|X \right] \]
\[
\begin{align*}
= & \int \left\{ \frac{\partial}{\partial S} \left\{ \int f_b(\tilde{S}(\beta) - u)h(u,X)du \right\} \int f_b(\tilde{S}(\beta) - u)h(u|X)du \right\} ds \\
= & \int \frac{\partial}{\partial S} \left\{ \int f_b(\tilde{S}(\beta) - u)h(u|X)du \right\} ds \\
= & \frac{\partial}{\partial S} \left[ \int \left\{ \int f_b(\tilde{S}(\beta) - u)h(u|X)du \right\} ds \right] \\
= & \frac{\partial}{\partial S} [1] \\
= & 0
\end{align*}
\]

Lemma [A.2]
\[
E\left( \frac{f'}{f} | \tilde{S}(\beta), X \right) = 0
\]

**Proof:** By assumption, \( U_t = \epsilon_t - \tilde{\epsilon} \) is independent of \( \tilde{S}(\beta) = \alpha_t + \tilde{\epsilon} \) and \( X \). Hence,
\[
E\left( \frac{f'}{f} | \tilde{S}(\beta), X \right) = E\left( \frac{f'}{f} \right) = 0
\]

This follows from Lemma 1. Now from Lemma 1 and Lemma 2, we get \( E(\ell_\beta) = 0 \).

Thus,
\[
\Pi(\ell_\beta| V_1 + V_2) = E(\ell_\beta| U_t(\beta), \tilde{S}(\beta), X) - E(\ell_\beta)
\]
\[
= -E \left[ X - \bar{X} \right] \frac{f'}{f} (U_t(\beta)) - E \left[ \bar{X} \right] \frac{w'}{w} (\tilde{S}(\beta), X)
\]

The efficient score is:
\[
\ell^*_\beta = - \left[ X - E(X) - (X - E(X))' \frac{f'}{f} (U_t(\beta)) \right] \frac{w'}{w} (\tilde{S}(\beta), X)
\]

The information bound is
\[
I(\beta) = E\ell^* \ell^*
\]
\[
= \Sigma_T \left( \frac{f'}{f} \right)^2 + \Sigma_B \left( \frac{w'}{w} \right)^2
\]

where
\[
\Sigma_T = ET^{-1} \sum_{t=1}^{T} \left[ [X_t - \bar{X}] - E \left[ X - \bar{X} \right] \right]' \left[ ([X_t - \bar{X}] - E \left[ (X) - \bar{X} \right]) \right]
\]
and
\[
\Sigma_B = ET^{-1} \sum_{t=1}^{T} (X_t - E(\bar{X}))(X_t - E(\bar{X}))'
\]
4.4.2 A Hausman-Taylor Estimator

Let $\theta = [\beta \gamma]$. The setup is basically the same except now the loglikelihood of $(X, Z, Y)$ is:

$$
\mathcal{L}(X, Z, Y, \alpha, \epsilon) = \ln g(X|Z) + \ln f_w(U_t(\theta)) \\
+ \ln \int f_b(\bar{S}(\theta) - \mu) h(\mu, Z) d\mu
$$

(4.14)

Let $w = \ln \int f_b(\bar{S}(\theta) - \mu) h(\mu, Z) d\mu$ be the density of $\bar{S}(\theta)$ and $Z$. The score are as follows:

$$
\ell_{\beta} = - (X_t - \bar{X}) \frac{f'}{f}(U_t(\theta)) - \bar{X} \frac{w'}{w}(\bar{S}(\theta), Z)
$$

$$
\ell_{\gamma} = - (Z_t - \bar{Z}) \frac{f'}{f}(U_t(\theta)) - \bar{Z} \frac{w'}{w}(\bar{S}(\theta), Z)
$$

In this model, the nuisance parameter space can be defined as follows:

$$V_1 = \{a(X, Z); a \in L_2(p), E(a(X, Z)|Z) = 0\}$$

$$V_2 = \{b(\bar{S}(\beta), Z); b \in L_2(p), E(b(\bar{S}(\beta), Z) = 0\}$$

$$V_3 = \{c(U(\beta)); c \in L_2(p), E(c(U(\beta)) = 0\}$$

Lemma [A.3.]

$$E \left[ \frac{w'}{w}(\bar{S}(\beta), Z)|X, Z, U_t \right] = 0$$

**Proof:** Since $X$ and $\alpha$ are independent conditionally on $Z$ and $\alpha$. $X$ and $Z$ are independent of $U_t$.

$$E \left[ \frac{w'}{w}(\bar{S}(\beta), Z)|X, Z, U_t \right] = E \left[ \frac{w'}{w}(\bar{S}(\beta), Z)|Z \right]$$

The proof follows as in Lemma A.1.

Lemma [A.4.]

$$E(\frac{f'}{f}(\bar{S}(\beta), X, Z) = 0$$

**Proof:** This proof is the same as Lemma A.2.
Note

\[
\ell_{\beta}^* = \ell_{\beta} - \Pi(\ell_{\beta}|V_1 + V_2 + V_3) \\
= \ell_{\beta} - \Pi(\ell_{\beta}|V_1 + V_3) - \Pi(\ell_{\beta} - \Pi(\ell_{\beta}|V_1 + V_3)|W_1)
\]

where

\[
W_2 = V_1^\perp \cap (V_1 + V_2 + V_3) \\
= \{b(\bar{S}(\beta), Z) - E\left(b(\bar{S}(\beta), Z)|Z\right); b \in L_2(p)\}
\]

By Lemma A.3. and Lemma A.4., \(\ell_{\beta} \perp V_1\) hence

\[
\Pi(\ell_{\beta}|V_1) = 0
\]

(4.17)

Then by (A.1.) and (A.2.)

\[
\ell_{\beta}^* = \ell_{\beta} - \Pi(\ell_{\beta} - \Pi(\ell_{\beta}|V_3)|W_1)
\]

(4.18)

From Lemma A.4.

\[
-(X_t - \bar{X}) \frac{f'}{f}(U_t(\theta)) \perp W_1
\]

Thus,

\[
\Pi(\ell_{\beta} - \Pi(\ell_{\beta}|V_3)|W_1) = \Pi(-(X_t - \bar{X}) \frac{f'}{f}(U_t(\theta)) - \bar{X} \cdot \frac{w'}{w}(\bar{S}(\theta), Z)|W_1)
\]

\[
= E\left(-\bar{X} \cdot \frac{w'}{w}(\bar{S}(\theta), Z)|\bar{S}(\theta), Z\right) - E\left(-\bar{X} \cdot \frac{w'}{w}(\bar{S}(\theta), Z)|Z\right)
\]

By Lemma 3,

\[
E\left(-\bar{X} \cdot \frac{w'}{w}(\bar{S}(\theta), Z)|Z\right) = 0.
\]

Since \(X\) and \(\alpha\) are independent conditionally on \(Z\),

\[
E\left(-\bar{X} \cdot \frac{w'}{w}(\bar{S}(\theta), Z)|\bar{S}(\theta), Z\right) = -E(\bar{X}|Z) \cdot \frac{w'}{w}(\bar{S}(\theta), Z)
\]

Both imply that

\[
\Pi(\ell_{\beta}|W_1) = E(\bar{X}|Z) \cdot \frac{w'}{w}(\bar{S}(\theta), Z)
\]

Moreover, we find that

\[
\Pi(\ell_{\beta}|V_3) = E\left(X - \bar{X}\right) \frac{f'}{f}(U_t(\theta))
\]
The efficient score and information bound are:

\[ \ell_\theta = - \left[ x - \mathbb{E}(x) - (\bar{x} - \mathbb{E}(\bar{x})) \right] \frac{f'(U_i(\theta))}{f(U_i(\theta))} - \left[ \bar{x} - \mathbb{E}(\bar{x}|z) \right] \frac{w'(s(\theta), z)}{w(s(\theta), z)} \]

\[ \ell_\gamma = - \left[ z - \mathbb{E}(z) - (\bar{z} - \mathbb{E}(\bar{z})) \right] \frac{f'(U_i(\theta))}{f(U_i(\theta))} \]

and

\[ I = \begin{bmatrix} \Sigma_T \left( \frac{r}{f} \right)^2 & \Sigma_B \left( \frac{w}{w} \right)^2 & \Sigma_XZ \left( \frac{r}{f} \right)^2 \\ \Sigma_XZ \left( \frac{r}{f} \right)^2 & \Sigma_TZ \left( \frac{r}{f} \right)^2 \end{bmatrix} \]

where

\[ \Sigma_{TZ} = \mathbb{E} \sum_{i=1}^T \left( (z - \mathbb{E}(z)) - (\bar{z} - \mathbb{E}(\bar{z})) \right)' \left( (z - \mathbb{E}(z)) - (\bar{z} - \mathbb{E}(\bar{z})) \right) \]

and

\[ \Sigma_{XZ} = \mathbb{E} \sum_{i=1}^T \left( (x - \mathbb{E}(x)) - (\bar{x} - \mathbb{E}(\bar{x})) \right)' \left( (x - \mathbb{E}(x)) - (\bar{x} - \mathbb{E}(\bar{x})) \right) \]
Chapter 5

Convergence of Banking Efficiency

There has been substantial research devoted to the area of technical efficiency in industries with banking being one of the more interesting applications.\textsuperscript{55} The banking industry has historically been regulated with limited market entry, where average inefficiencies have been observed to persist over time. Regulation has brought with it a wealth of publicly available data for researchers to use. In this study, we use panel data with over 25000 observations of U.S. banks. This dataset represents an excellent opportunity to determine movements in banking technical efficiency, a neglected topic in the banking efficiency literature.

The notion of time varying efficiency becomes important in light of the substantial changes that have taken place in banking technology, environment, and especially regulation. Technical and financial innovations have caused dramatic changes in the banking industry (Berger, Kashyap, and Scalise, 1995). For example, the introduction of more advanced computer technology has led to changes in customer services (ATMs) and in information costs to banks. Also, nonfinancial institutions have moved into traditional banking markets (Berger, 1993), while deregulation in the 80’s (which continues in the 90’s) has allowed banks more flexibility in their business. Adjustments in banking regulations include the deregulation of deposit accounts, several major changes in capital requirements, and reductions in reserve requirements.

These new trends in the banking technology, environment, and regulations have lead to movements in banking efficiency. While the existence and persistence in banking inefficiencies has been shown in numerous articles (Berger, 1993; Bauer, Berger, and Humphrey, 1993; Ferrier and Lovell, 1990; Berger, 1995), the time series

\textsuperscript{55}See Berger, Hunter, and Timme (1993) for a review.
properties has not been thoroughly examined. Although some do allow for changes in technology in an effort to overcome possible model misspecification (Berger, 1993; Bauer, Berger, and Humphrey 1993), these studies only estimate average inefficiencies. In an evolving industry, it is important not only to measure the average level of efficiency, but also to analyze the change in efficiencies. Dynamic efficiency models such as Cornwell, Schmidt and Sickles (1990, hereafter CSS), Kumbhakar (1991), Lee and Schmidt (1993), and Battese and Coelli (1992) have been developed to allow for changes in efficiencies. However, these models typically parameterize the production technology. In a changing industry, it is important to allow for changes in technology. We use estimators that make minimal functional form assumptions and allow for time varying efficiencies.

In this paper, we apply three techniques to estimate annual firm technical efficiencies in the banking industry. Two of these approaches make minimal functional form assumptions. The first approach, Linear programming methods, has been extensively used to measure technical efficiency. Data envelopment analysis (DEA) is one such linear programming procedure that has been applied to banking data (Ferrier and Lovell, 1990; English et al., 1993; Kwan, Eisenbeis, and Ferrier, 1995). One advantage of DEA is that it allows for a flexible functional form. DEA's disadvantage is that it does not allow for random error, because it is deterministic. The second approach is a statistical method that allows for flexible functional form in the inputs and also allows for random error. We extend the semiparametric estimator in Adams, Berger, and Sickles (1996, hereafter ABS)\textsuperscript{56} to allow for time varying technical efficiency in the spirit of CSS.\textsuperscript{57} Finally, the third method is the maximum likelihood estimator developed by Battese and Coelli (1992), which models efficiency as an exponential function of time. The Battese and Coelli estimator is fully parametric in that both functional form and distribution of effect and error are specified.

\textsuperscript{56}We use the nonparametric regression based estimator.

\textsuperscript{57}We also the CSS estimator, where we assume a Cobb-Douglas functional form for the inputs.
We apply these three techniques to the same data utilized in ABS. These methodologies differ in their assumptions regarding the functional form of the estimated frontier. Most authors have dealt with this problem by allowing regression coefficients to change from year to year (Bauer, Berger, and Humphrey 1993; Berger 1993). The dynamic nature and changing technology of the banking industry indicate that not only the underlying technology but also the level of inefficiency is changing over time. We allow for flexible functional forms using both DEA and panel methods and then compare them to the parametric stochastic panel distance frontier. A comparison of these results should be robust against possible model misspecification.

Section 1 motivates using models with time varying efficiencies in a brief econometric analysis. Section 2 describes the model and introduces the semiparametric estimation procedure with time varying efficiencies. Section 3 discusses DEA and section 4 discusses stochastic distance frontier. Section 5 presents the results and determines the movements in banking efficiency.

5.1 An Econometric Motivation for Dynamic Efficiencies

In this section, we examine the implications of introducing a firm effect which accounts for omitted variables known to decision-makers, but is not observable by the researcher. Although considerable interest has focused on the specification of the model and effects, little work has been completed on the stochastic implications of representing omitted variables as firm effects in an effort to analyze firm differences.58 Some authors consider the consequences of including random error in utility or cost functions (Brown and Walker, 1989, 1992), but this literature does not consider the case of including a firm effect in the function. Our results indicate that measuring average efficiency could lead to biased parameter estimates.

58We consider only omitted variable problem, because of its meaning in the efficiency literature. Other problems such as model and error misspecification can be considered with similar results.
The centerpiece of efficiency analysis is typically the model of firm cost minimization or profit maximization. For expositional purposes, we consider the output distance function, which through duality (Dievort, 1982) is an equivalent method of modeling efficiency.\textsuperscript{59} The output distance function can be represented by:

\[ D(X, Y) = \frac{g(Y)}{f(X)} \leq 1 \]  

(5.1)

A firm is producing efficiently, when the output distance function equals one. In this case, the firm cannot increase output without increasing the inputs or decreasing another output. Random error and firm effects could enter the output distance function in any number of ways. We model the output distance function with an exponential function of random error and firm effects. A geometric mean representation of the inputs and outputs is assumed for simplification:

\[ D(X, Y) = \frac{\prod_{i=1}^{J} Y_{ij}^{\gamma_i} \exp(\alpha_i(Z_{it}; \theta, \delta) + \epsilon_{it})}{\prod_{i=1}^{K} X_{kit}^{\beta_i}} = 1 \]  

(5.2)

where \( \gamma_i \) and \( \beta_k \) are weights, \( \theta = [\beta \; \gamma] \), and \( \alpha_i \) is a function of the omitted variables and the slope parameters, \( \theta \) and \( \delta \). The output distance function can be transformed by using logarithms and imposing linear homogeneity of the outputs to get the following standard panel model:

\[ Y_{it} = X_{it} \beta + Y_{it}^* \gamma + \alpha_i(Z_{it}; \theta, \delta) + \epsilon_{it} \]  

(5.3)

It becomes apparent that our firm effects depend on omitted variables, \( Z_{it} \). Assume the true model takes on the following form:

\[ Y_{it} = X_{it} \beta + Y_{it}^* \gamma + Z_{it} \delta + \epsilon_{it} \]  

(5.4)

Estimation of specified model (5.3) follows from the distributional assumptions on the effects and random error and their correlation with the regressors.\textsuperscript{60} We consider two cases, when the omitted variables are correlated (Within) or uncorrelated with the other regressors (Random Effects).

\textsuperscript{59} An analysis of cost or profit functions would be just as instructive. However, we are limiting ourselves to a simple case.

\textsuperscript{60} See Schmidt and Sickles (1984) for an overview.
The within estimator proceeds by using mean deviations as instruments to identify the slope parameters. In reference to the true model (5.4), the model is defined by:

\[
(Y_{it} - \bar{Y}_i) = (X_{it} - \bar{X}_i)\beta + (Y_{it}^* - \bar{Y}_i^*)\gamma + (\epsilon_{it}^* - \bar{\epsilon}_i^*)
\]

where \((\epsilon_{it}^* - \bar{\epsilon}_i^*) = (Z_{it} - \bar{Z}_i)\delta + (\epsilon_{it} - \bar{\epsilon}_i)\). It is important to note that the residuals are heteroskedastic. Let \(W = [X Y^*]\). Within estimation of the slope parameters follows from least squares on the transformed model above:

\[
\hat{\theta} = (W'QW)^{-1}W'QY
\]

where \(Q = I_T - \frac{1}{T}ii'\) and \(i\) is a vector of ones. It can be shown that the estimates will be biased, when \(E[W_{it}QZ_{it}] \neq 0\). In other words, the slope parameters will be biased if the mean deviations of the omitted variables is correlated with the mean deviations of the included regressors. The slope parameter estimates account for the within deviation of the omitted variable through this correlation. The effects are now defined by:

\[
\alpha_i = \bar{Y}_i - \bar{X}_i\beta + \bar{Y}_i^*\gamma =^\text{true} = \bar{Z}_i\delta
\]

It becomes apparent that the firm effect accounts for the firm systematic portion of the omitted variable.\(^61\) The bias of the effects also depends on the same correlation as the slope parameters. It is only scaled by \(\bar{X}\) and \(\bar{Y}^*\).

In random effects estimation, the effects are uncorrelated with the regressors. It is safe to assume that the regressors are uncorrelated with the omitted variables. Random effects estimation relies on the specified form of the covariance(\(\alpha_i, \epsilon\)). Typically, the covariance matrix has the following form: \(I_N\sigma_i^2 + I_T\sigma_Z^2\). If we replace the firm effect with the omitted variable, we get a different covariance matrix: \(I_N\sigma_i^2 + E[Z'\delta'\epsilon Z]\). The parameter estimates will be consistent and unbiased, but inefficient, because an inappropriate error structure is applied. In this case, heteroskedasticity tests (White type) could be used to find the appropriate covariance structure.

\(^61\)When the omitted variable is time invariant, then within estimation can consistently identify the slope parameters and firm effects without any problems of biasedness or heteroskedasticity.
In this short analysis of the statistical issues in panel efficiency modeling, several issues need to be resolved. First, if the omitted variable is time invariant, then the panel model with firm effects is an effective way of dealing with firm heterogeneities (regardless if the model is estimated with within or random effects). Second, if the omitted variable is time varying, within estimates could be biased and inconsistent and random effects estimators apply an inappropriate covariance matrix. These notions of bias, correlation, time invariance, and heteroskedasticity are important in identifying firm heterogeneities and technology.

Furthermore, when a residual analysis is the best method of industry efficiency measurement available, this econometric analysis indicates that modeling firm heterogeneities as time invariant is problematic. Allowing firm heterogeneities to change over time is not only statistically a superior, but also it can be more informative about the industry in question. Hence, this chapter considers methods of efficiency measurement that allow firm heterogeneities to change over time.

5.2 Data Envelopment Analysis

Data envelopment analysis (DEA) is a linear programming approach to efficiency measurement. It generates a non-parametric, piecewise linear convex frontier or envelope of the input and output data. Two advantages of DEA have been noted. First, it does not impose any particular functional form for the technology in question. Second, no distributional assumptions are needed. In this sense, DEA is more flexible than either splines, which imposes restrictions on higher derivatives for continuity, or the Nadaraya-Watson estimator, which imposes iid assumptions on the data. However, DEA is deterministic and does not allow for random error. All deviations from the frontier are considered inefficiency.

---

62 This is the true nature of efficiency measurement.
The first step in DEA is to construct a representation of the technology which will serve as the point of reference for measuring relative efficiency. The input requirement set \( L(y) \) characterizes the technology (production possibilities) and can be constructed from observed data in the following manner:

\[
L(l'_1, \ldots, l'_m) = \{ (x_1, \ldots, x_m) : l'_i \leq \sum_{s=1}^{S} \mu_s l'_{is}, \ i = 1, \ldots, m \}
\]

\[
x_j \geq \sum_{s=1}^{S} \mu_s x_{js}, \ j = 1, \ldots, n;
\]

\[
z_h \leq \sum_{s=1}^{S} \mu_s z_{hs}, \ h = 1, \ldots, k_r;
\]

\[
z_h \geq \sum_{s=1}^{S} \mu_s z_{hs}, \ h = k_r+1, \ldots, k;
\]

\[
\mu_s \geq 0, \ s = 1, \ldots, S;
\]  

(5.8)

where \( \mu = (\mu_1, \ldots, \mu_S) \) is an intensity vector that forms convex combinations of observed input vectors and output vectors. The constraints on the first \( k_r \) elements of \( z \) imply that observations would like to expand these variables as much as possible given inputs, and the constraints on the last \( (k - k_r) \) elements of \( z \) imply that the observation would like to limit these variables given outputs. Relative to this bounding technology the technical efficiency of each observation is calculated by solving \( S \) linear programming problems of the following form:

\[
\max \Theta_{is},
\]

s.t. \( \Theta_{is} l'_{jis} \leq \sum_{i=1}^{M} \sum_{s=1}^{S} \mu_{is} l'_{jis}, \ j = i, \cdots, J \)

\[
\sum_{i=1}^{I} \sum_{s=1}^{S} \mu_{is} x_{kis} \leq x_{kis}, \ k = 1, \cdots, K
\]

\[
\mu_{is} \geq 0, \ s = 1, \cdots, S, \ i = 1, \cdots, M;
\]  

(5.9)

The solution to this maximization problem, \( \Theta_{is}^* \), shows the amount by which an observation can multiply its output vector and still utilize no more of any input. For
an efficient firm, where it is not possible to produce any more outputs with the same
input vector, $\Theta_{is}^* = 1$.

5.3 A Semiparametric Model

Several linear programming techniques and semiparametric statistical methods are
available to identify firm inefficiencies. Data envelopment analysis and other linear
programming techniques have been applied throughout the efficiency literature. In
the banking efficiency literature, several examples exist such as Ferrier and Lovell
adapted to a number of different settings and cases, such as multiproduct firms,
input-based and output-based models, and differing returns to scale. However, DEA
does not allow for random error. This is a shortcoming that the semiparametric
methods overcome, while allowing for a flexible functional form in the model. These
semiparametric methods, based on kernel methods like Nadaraya-Watson, have not
been extensively applied in the efficiency literature, especially for multioutput firms.
Exceptions include McAllister and McManus (1993) and ABS. The latter introduce
a semiparametric estimator for a multiple output setting, which allows for flexible
functional forms in inputs.

5.3.1 The Model

As a motivation for a model, consider the output distance function of a firm with
multiple outputs and inputs. The output distance function provides a radial measure
of technical efficiency that describes the fraction of aggregated outputs, given chosen
inputs. For a particular observation $i$, the output distance function can be described
by:

$$D_0(x, y) = \min_{\theta} \{ \theta : (x, y/\theta) \in T \} \leq 1 \quad (5.10)$$
where $T = \{(x, y^\prime) : x \in R_+^N, y^\prime \in R_+^M, x$ can produce $y^\prime\}$ is the technology set. This definition of the output distance function holds in the multioutput setting (Färe and Primont, 1995). This general definition of the output distance function can be approximated in the following manner:

$$D_0(x, y^\prime) = \frac{g(y^\prime)}{f(x)} \leq 1$$  \hspace{1cm} (5.11)

This formulation does not allow for the necessary constraints, though it does allow for a flexible functional form in outputs and inputs. To impose the necessary restrictions, the outputs are linearized leading to a new approximation of the output distance function:

$$D(x, y^\prime) = \frac{\prod_{1}^{J} y_{ji}}{f(x_{ki})} \leq 1$$  \hspace{1cm} (5.12)

where $y_{ji}$ and $x_{ki}$ are the levels of output $j$ and input $k$. The specification of the output distance function is semiparametric, where the outputs are approximated by a weighted linear combination of $y_{ji}$ and the functional form of the inputs, $f(x_{ki})$, is unknown. The necessary restrictions, such as linear homogeneity and convexity in outputs, on the output distance function can be imposed, while making minimal assumptions on the functional form of the inputs. When a firm is producing efficiently, i.e. the value of the distance function equals one, then it is not possible to increase the value of output without either decreasing another output or increasing inputs.

The model specification can be linearized using logarithms. We impose linear homogeneity and include error terms and effects to get the following form, where we normalize with respect to one $y_{j}$:

$$-\ln y_{j} = \sum_{j} \gamma_{j} \ln \tilde{f}_{j, it} - \ln f(x_{ki}) + \alpha_{it} + \epsilon_{it}$$  \hspace{1cm} (5.13)

where $\tilde{f} = (\tilde{f}_{j}) = (\frac{y_{j}}{y_{j}^*}), \ j = 1, \ldots, J - 1$. Define $Y_{it} = -\ln y_{j}$ and $Y_{it}^* = \ln \tilde{f}_{j, it}$. Let $f(X_{it}) = -\ln f(x_{ki})$. The distance frontier now becomes:

$$Y_{it} = f(X_{it}) + Y_{it}^* \gamma + \alpha_{it} + \epsilon_{it}$$  \hspace{1cm} (5.14)

This is a stochastic panel frontier model in the sense of Aigner and Chu (1980), where firm specific effects represent efficiency. This model differs from the traditional
panel model in that it makes minimal assumptions on the functional form of the input index.

5.3.2 Semiparametric Estimation

In the semiparametric estimation, we include an additional assumption to account for the time series properties of efficiencies. Several authors have allowed efficiencies to change over time. CSS model efficiencies as a function of time, while Kumbhakar (1990) models efficiencies as an exponential function of time. Others include Battese and Coelli (1992), and Lee and Schmidt (1993) who allow for other model specifications. In this illustration, we model efficiencies as a function of time in the spirit of CSS:

\[
\alpha_{it} = \delta_1 t + \delta_2 t^2 + \delta_3 t^3 \quad (5.15)
\]

This semiparametric estimation method proceeds in the following manner. Assuming that the inputs are not correlated with the effects, the conditional expectation for the distance frontier function is:

\[
E[Y_{it} | X_{it}] = f(X_{it}) + E[Y_{it}^* | X_{it}]
\]

where the means of the random effects, \( \alpha_{it} \), are uncorrelated with the inputs. Subtracting this conditional expectation from the distance function provides us with the model to be estimated:

\[
Y_{it} - E[Y_{it} | X_{it}] = (Y_{it}^* - E[Y_{it}^* | X_{it}])' \gamma + \alpha_{it} + \epsilon_{it} \quad (5.16)
\]

\[
f(X_{it}) = E[Y_{it} | X_{it}] - E[Y_{it}^* | X_{it}] \gamma \quad (5.17)
\]

This model is estimated in two steps: first, the conditional expectations are estimated using the Nadaraya-Watson estimator; second, the transformed model (sempir) is estimated using the CSS estimator. We then use the residuals from (5.16) to find the time-varying efficiencies by regressing them in (5.15) as in CSS. We also apply the CSS estimator to a model, where we assume a Cobb-Douglas functional form for the inputs.
5.4 Battese and Coelli Estimator

As a point of reference, we apply a fully parameterized model introduced by Battese and Coelli (1992) to the data.\textsuperscript{63} This model differs from the standard stochastic panel frontiers model in that it allows for time varying technical efficiency.

The stochastic distance frontier is motivated from the output distance function which is fully parameterized by approximating the output distance function and including effects and error in exponential form:

\[
D(x, v) = \frac{\prod_j \sqrt[j]{q_j^* \exp(\alpha_{it} + \epsilon_{it})}}{\prod_k \sqrt[k]{x_{ki}^*}} = 1
\]

(5.18)

where \(q_j^*\) and \(x_{ki}^*\) are the levels of output \(j\) and input \(k\). The \(q_j\) and the \(r_k\) are the weights which describe the technology of the firm. \(\alpha_{it}\) are the firm effects and \(\epsilon_{it}\) the random error.

The model can be transformed using logarithms and imposing linear homogeneity:

\[
Y_{it} = X\beta + Y_{it}^*\gamma + \alpha_{it} + \epsilon_{it}
\]

(5.19)

where \(\alpha_{it} = \eta_{it}\alpha_i = (\exp[-\eta(t - T)])\alpha_i\).

This model represents a stochastic panel distance frontier with time varying effects. The \(\beta\) and \(\gamma\) are unknown \(K \times 1\) and \(J - 1 \times 1\) parameter vectors; The \(\epsilon_{it}\) are iid \(N(0, \sigma^2)\) random errors; The \(\alpha_i\) are iid nonnegative truncations of a \(N(0, \sigma^2)\); \(\eta\) is an unknown scalar parameter.

Note that the variation of the firm effects depends on the parameter, \(\eta\). The effects decrease, remain constant, or increase as \(t\) increases. if \(\eta > 0\), \(\eta = 0\), \(\eta < 0\). One obtains the standard time invariant technical efficiency measure when \(\eta = 0\). The model is estimated using Frontier4.1, a program developed by Tim Coelli.

\textsuperscript{63}This approach is similar to Berger's Economic Frontier Approach.
5.5 Results

We estimate relative technical efficiencies in the banking industry using DEA, the semiparametric estimator, CSS, and Battese and Coelli estimators. The dataset is further separated according to three regulatory environments: unit banking (UNIT), limited branching (LIMIT), and statewide branching (STATE). In all cases, we allow technical efficiencies to change over time, in order account for the changing environment, regulation, and technology in the banking industry.

Results are presented in Figures 5.1, 5.2, and 5.3 and Tables 5.1, 5.2, 5.3, 5.4, and 5.5.

DEA technical efficiency scores were derived for each year, where the data for solely each year was used to determine the frontier. DEA results indicate that banking mean relative technical efficiency range from 92% to 95% from 1980 to 1989. LIMIT banks tend to have a higher mean than the UNIT or STATE banks, except in 1980 and 1984, where UNIT banks have the highest mean. This result have been observed in other studies of time invariant technical efficiency. (Berger, 1993; ABS); STATE banks have the lowest mean relative technical efficiency, but their standard deviation is lower than both other types in all years.

More importantly, all three branching types observe the very same movements in relative technical efficiency. Relative DEA scores were regressed on time and time2, where the parameter estimates are close to those in CSS results. Generally speaking, in all bank environments, technical efficiency falls in 1980 to 1984 and then increases from 1985-1989. This observation is true for all estimators except the Battese and Coelli estimator, which doesn’t allow for nonlinear movements.

This result is not surprising considering the fact that the industry underwent substantial changes during the 80s. This is particularly true for the early 80s, where some deposit types were deregulated, capital requirements were defined, and banks
began to move into nontraditional banking markets.\textsuperscript{64} Also, bank failures increased in 1985 to 1989.\textsuperscript{65}

At this point, the cause of these movements is uncertain. In an effort to explain the changes in mean relative technical efficiencies, we regress mean relative efficiencies on demand variables such as state specific income growth and unemployment. We find that a 1\% change in state growth can explain a 0.08 to 0.11 change in relative technical efficiencies.\textsuperscript{66} Unemployment has a much lower effect of 0.006 to 0.009 change in relative technical efficiency. It becomes apparent that one cause of variation in technical efficiency can be attributed to differences in market demand. These methods identify production, where the outputs are assumed exogenous. These results indicate that a structural model that endogenizes output demand should be considered to find the true level of technical efficiency.

Semiparametric estimation starts by first determining the optimal binwidth for the conditional expectations. A grid search over [0.01,2] interval over the regression space. The optimal binwidth is selected by the following criterion: \[ \min \frac{1}{N} \sum (y - f(x))^2. \] This procedure is applied for each regression. Hence, we allow the smoothing parameter to differ for each dependent variable \((Y, Y^*)\). the optimal binwidth parameters range from 0.68 to 1.69.\textsuperscript{67}

In both the parameterized CSS, assuming Cobb-Douglas functional form of the inputs, and the semiparametric CSS, movements in technical efficiency are very similar as noted by the coefficients for time and time\textsuperscript{2}. This result and the DEA results robustify the movements in banking efficiency. In the first half of the sample period (1980-84), banks were diverging. In the second half of the sample period (1985-1989), banks were converging.

\textsuperscript{64}During the 80s, off balance sheet activities increased. Noninterest income was 8\% (3.5\% for small banks) of total income in 1980 and 18\% (4.4\% for small banks) in 1989.

\textsuperscript{65}Note: The dataset does not include any failed banks, which could bias our results.

\textsuperscript{66}Berger and Mester (1996) observe a much lower correlation between growth and technical efficiency.

\textsuperscript{67}Rule of thumbs methods of binwidth selection were also used with similar results.
It is important to note that the parameter estimates for the parameterized CSS differ from estimates using the typical within estimator. We observe a significant change in the coefficients for capital, purchased funds, and demand deposits. These changes in coefficient estimates indicates a possible bias caused by the fact that within estimation only allows for average technical efficiency.

Finally, the results from the Battese and Coelli representation of the stochastic distance frontier indicate different trends across the bank types. Unit banks show a divergence in mean relative technical efficiency from the frontier, where $\eta = -0.112$. State and Limit banks converge to the frontier with $\eta = 0.214$ and 0.235 respectively. While the Battese and Coelli estimator does not impose the same assumptions on the model as the semiparametric estimator, it does indicate that technical efficiency is changing over time. Also, it is important to note that the Battese and Coelli estimator does not allow for a time specific or firm specific $\eta$ and a changing technology (the slope parameters remain constant over time), which would give a more robust interpretation of the trends. Moreover, a nonlinear time trend would be insightful as well.

In short, the banking industry has undergone substantial changes during the 80’s. As a result, technical efficiency in the banking industry has changed during the observed period as well. The movements in mean relative technical efficiency and, hence, absolute technical efficiency indicate that banks become more similar and converge in the second half of the sample period. This result can be observed with both parametric and semiparametric methods.

Furthermore, our results on output and demand variables indicate that output endogeneity does exist and should be account for in a structural model. Some of the differences in efficiency can be attributed to differences in demand.
### Annual Relative Technical Efficiencies: DEA

<table>
<thead>
<tr>
<th>Year</th>
<th>Unit</th>
<th>State</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.945 (0.0334)</td>
<td>0.935 (0.0320)</td>
<td>0.942 (0.0330)</td>
</tr>
<tr>
<td>1981</td>
<td>0.937 (0.0352)</td>
<td>0.930 (0.0326)</td>
<td>0.935 (0.0361)</td>
</tr>
<tr>
<td>1982</td>
<td>0.933 (0.0357)</td>
<td>0.926 (0.0326)</td>
<td>0.932 (0.0372)</td>
</tr>
<tr>
<td>1983</td>
<td>0.931 (0.0359)</td>
<td>0.921 (0.0331)</td>
<td>0.931 (0.0368)</td>
</tr>
<tr>
<td>1984</td>
<td>0.946 (0.0370)</td>
<td>0.937 (0.0330)</td>
<td>0.945 (0.0393)</td>
</tr>
<tr>
<td>1985</td>
<td>0.936 (0.0357)</td>
<td>0.929 (0.0317)</td>
<td>0.939 (0.0360)</td>
</tr>
<tr>
<td>1986</td>
<td>0.935 (0.0353)</td>
<td>0.930 (0.0315)</td>
<td>0.938 (0.0377)</td>
</tr>
<tr>
<td>1987</td>
<td>0.935 (0.0359)</td>
<td>0.935 (0.0320)</td>
<td>0.940 (0.0374)</td>
</tr>
<tr>
<td>1988</td>
<td>0.939 (0.0364)</td>
<td>0.940 (0.0319)</td>
<td>0.945 (0.0367)</td>
</tr>
<tr>
<td>1989</td>
<td>0.942 (0.0363)</td>
<td>0.944 (0.0312)</td>
<td>0.950 (0.0353)</td>
</tr>
</tbody>
</table>

### Annual Relative Technical Efficiencies: Stochastic Distance Frontier

<table>
<thead>
<tr>
<th>Year</th>
<th>Unit</th>
<th>State</th>
<th>Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>0.958 (0.00579)</td>
<td>0.886 (0.0341)</td>
<td>0.875 (0.0411)</td>
</tr>
<tr>
<td>1981</td>
<td>0.953 (0.00643)</td>
<td>0.907 (0.0280)</td>
<td>0.897 (0.0345)</td>
</tr>
<tr>
<td>1982</td>
<td>0.948 (0.00714)</td>
<td>0.925 (0.0229)</td>
<td>0.915 (0.0287)</td>
</tr>
<tr>
<td>1983</td>
<td>0.942 (0.00793)</td>
<td>0.940 (0.0185)</td>
<td>0.930 (0.0238)</td>
</tr>
<tr>
<td>1984</td>
<td>0.936 (0.00879)</td>
<td>0.952 (0.0149)</td>
<td>0.943 (0.0196)</td>
</tr>
<tr>
<td>1985</td>
<td>0.928 (0.00974)</td>
<td>0.962 (0.0120)</td>
<td>0.954 (0.0161)</td>
</tr>
<tr>
<td>1986</td>
<td>0.920 (0.0108)</td>
<td>0.970 (0.00959)</td>
<td>0.962 (0.0132)</td>
</tr>
<tr>
<td>1987</td>
<td>0.912 (0.0119)</td>
<td>0.975 (0.00765)</td>
<td>0.969 (0.0108)</td>
</tr>
<tr>
<td>1988</td>
<td>0.902 (0.0132)</td>
<td>0.981 (0.00610)</td>
<td>0.975 (0.00878)</td>
</tr>
<tr>
<td>1989</td>
<td>0.891 (0.0145)</td>
<td>0.985 (0.00485)</td>
<td>0.980 (0.00714)</td>
</tr>
</tbody>
</table>

### Table 5.1 Unit Banks

<table>
<thead>
<tr>
<th>Variable</th>
<th>CSS</th>
<th>Semiparametric CSS</th>
<th>DEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.280 (0.0140)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>-0.0266 (0.00648)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.0635 (0.00353)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.0163 (0.00562)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.344 (0.00711)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.380 (0.00716)</td>
<td>0.082 (0.0062)</td>
<td></td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.296 (0.00454)</td>
<td>0.629 (0.0058)</td>
<td></td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.289 (0.00477)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.0653 (0.0029)</td>
<td>0.054 (0.0058)</td>
<td>0.356 (0.0017)</td>
</tr>
<tr>
<td>time2</td>
<td>-0.00107 (0.000253)</td>
<td>-0.0037 (0.00051)</td>
<td>-0.028 (0.00023)</td>
</tr>
</tbody>
</table>
### Table 5.2 State Banks

<table>
<thead>
<tr>
<th>Variable</th>
<th>CSS</th>
<th>Semiparametric CSS</th>
<th>DEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.138 (0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>-0.0760 (0.0076)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.0970 (0.0037)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.0496 (0.0044)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.309 (0.0085)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.385 (0.0081)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.347 (0.0082)</td>
<td>0.31 (0.0061)</td>
<td></td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.302 (0.0050)</td>
<td>0.363 (0.0064)</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.0229 (0.0030)</td>
<td>0.029 (0.0042)</td>
<td>0.354 (0.0018)</td>
</tr>
<tr>
<td>time2</td>
<td>-0.00203 (0.00027)</td>
<td>-0.002 (0.0074)</td>
<td>-0.028 (0.00022)</td>
</tr>
</tbody>
</table>

### Table 5.3 Limit Banks

<table>
<thead>
<tr>
<th>Variable</th>
<th>CSS</th>
<th>Semiparametric CSS</th>
<th>DEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Labor</td>
<td>-0.176 (0.014)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Capital</td>
<td>-0.0745 (0.0067)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Purchased Funds</td>
<td>-0.0341 (0.0045)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>-0.0490 (0.0062)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demand Deposits</td>
<td>-0.297 (0.0093)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Retail Time and Savings Deposits</td>
<td>-0.433 (0.0087)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Commercial and Industrial Loans</td>
<td>0.246 (0.0056)</td>
<td>0.292 (0.0052)</td>
<td></td>
</tr>
<tr>
<td>Installment Loans</td>
<td>0.386 (0.0065)</td>
<td>0.316 (0.0067)</td>
<td></td>
</tr>
<tr>
<td>time</td>
<td>0.0388 (0.0038)</td>
<td>0.0383 (0.0038)</td>
<td>0.357 (0.0018)</td>
</tr>
<tr>
<td>time2</td>
<td>-0.0034 (0.00034)</td>
<td>-0.0034 (0.00031)</td>
<td>0.028 (0.00023)</td>
</tr>
</tbody>
</table>
Figure 5.1  Unit Bank Annual Mean Relative Efficiency
Figure 5.2  State Bank Annual Mean Relative Efficiency

State Mean Relative Efficiency 1980–1989
Figure 5.3  Limit Bank Annual Mean Relative Efficiency
Bibliography


