

**A Comparative Study of Alternative Methods for Efficiency Measurement
With Applications to the Transportation Industry**

by

CHUNYAN YU

**B.Eng. Northern Jiaotong University, 1982
M.Sc. The University of British Columbia, 1986**

**A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

in

**THE FACULTY OF GRADUATE STUDIES
COMMERCE AND BUSINESS ADMINISTRATION**

**We accept this thesis as conforming
to the required standard**

THE UNIVERSITY OF BRITISH COLUMBIA

January, 1995

© Chunyan Yu, 1995

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

(Signature)

Department of Commerce

The University of British Columbia
Vancouver, Canada

Date Jan 26, 1995

Abstract

This thesis is concerned with measuring and explaining the productive efficiency of firms or organizations. In particular, the study compares three alternative methods of measuring efficiency, namely, the deterministic frontier method, the stochastic frontier method, and the data envelopment analysis method (DEA).

The dissertation consists of two parts. In Part I, the relative merits of the three methods are discussed and evaluated through a Monte Carlo study under certain known conditions. The study focuses on the effects of exogenous variables on efficiency estimates. The results show that the stochastic frontier method generally produces better efficiency estimates than the other two methods. The DEA, however, has a slight advantage in cases where there are weak input substitution and large variations in input variables. In Part II, the three methods are examined empirically through their applications to a panel of 19 railways in OECD countries and a panel of 36 international airlines. Comparison of the three sets of efficiency estimates confirms that on average the stochastic frontier method yields higher efficiency estimates than the other two methods, as indicated by the Monte Carlo results. The efficiency estimates by the two parametric methods are highly correlated, whereas there are considerable differences between the DEA estimates and those from the parametric methods. This is also consistent with the Monte Carlo results. By comparing the alternative efficiency estimates in the two applications, it is found that there is less discrepancy among the three sets of efficiency estimates in the airline case than in the railway case. This can be partly attributed to the fact that there are fewer variations in the

operating environments in the airline case than in the railway case.

The simulation results in Part I provide some general guidelines regarding the relative merits of the three alternative methods under certain known conditions. The two applications of the three methods in Part II serve as examples of how these three methods can be applied to practical problems where no *a priori* knowledge of either the production technology nor the efficiency profile exists. They illustrate some of the problems that may be encountered in empirical applications.

TABLE OF CONTENTS

Abstract	ii
Table of Contents	iv
List of Tables	vii
List of Figures	xi
Acknowledgement	xii
Overview	1
Part I. A Monte Carlo Comparison of Alternative Methods for Efficiency Measurement	
Chapter 1 Introduction	10
1.1 Background	10
1.2 Research Issues	13
1.3 Objective of Part I	17
1.4 Organization of Part I	18
Chapter 2 Literature Review of Comparative Studies on Efficiency Measurement . .	19
2.1 Studies based on Empirical Data	19
2.2 Studies based on Simulated Data	24
Chapter 3 Concept of Efficiency	29
3.1 Farrell's Definition of Efficiency	31
3.2 Efficiency and Production Function	36
Chapter 4 Alternative Methods for Efficiency Measurement	40
4.1 Parametric Approaches	40
4.1.1. Deterministic Frontier Methods	43
4.1.2. Stochastic Frontier Methods	48
4.2 Nonparametric Approaches	57
4.2.1 The CCR Ratio	60
4.2.2 The BCC model	69
4.3 The Effects of Exogenous Variables	71
4.4 Summary	75
Chapter 5 Design of the Monte Carlo Experiments	78
5.1 Specification of the "true" production technology	78

5.2 Determination of Sample Size and Number of Replications	82
5.3 Generation of Inputs, Exogenous Variables, and Error Terms	83
5.4 Generation of the outputs	85
5.5 Conduct of the experiments	85
5.6 Evaluation Criteria	91
Chapter 6 The Results of Monte Carlo Experiments	93
6.1 The Effects of Sample Size	94
6.2 The Effects of Input Range	100
6.3 The Effects of Noise Level	107
6.4 The Effects of Exogenous Variables	114
6.5 The Effects of Outliers	129
6.6 The Effects of Underlying Production Technology	133
Chapter 7 Summary of Part I	145
Part II. Applications to the Transportation Industry	
Chapter 8 The Efficiency of Passenger Railway Systems	152
8.1 Introduction	152
8.2 Model Specification and the Data	155
8.3 The Results	158
8.3.1 The DEA Results	159
8.3.2 The Deterministic Frontier Results	170
8.3.3 The Stochastic Frontier Results	174
8.4 Comparison of Alternative Efficiency Estimates	178
8.5 Summary and Concluding Remarks	181
Chapter 9 The Efficiency of International Airline Industry	184
9.1 Introduction	184
9.2 Model Specification and the Data	189
9.3 The Results	191
9.3.1 The DEA Results	192
9.3.2 The Deterministic Frontier Results	202
9.3.3 The Stochastic Frontier Results	206
9.4 Comparison of Alternative Efficiency Estimates	211
9.5 Summary and Concluding Remarks	213
Chapter 10 Summary and Conclusions	216
Bibliography	224

Appendix A Sample Computer Code	236
Appendix B Data in Monte Carlo Experiments	241
Appendix C Railway Data	250
Appendix D Airline Data	260

List of Tables

Table

6.1.1 The Effects of Sample Size: Means of Efficiency Estimates	95
6.1.2 The Effects of Sample Size: The Mean Absolute Deviations	96
6.1.3 The Effects of Sample Size: Correlation	98
6.1.4 The Effects of Sample Size: Rank Correlation	99
6.2.1 The Effects of Input Range: Input Variable Statistics	101
6.2.2 The Effects of Input Range: Means of Efficiency Estimates	102
6.2.3 The Effects of Input Range: The Mean Absolute Deviations	104
6.2.4 The Effects of Input Range: Correlation	105
6.2.5 The Effects of Input Range: Rank Correlation	106
6.3.1 The Effects of Noise: Noise Term Statistics	107
6.3.2 The Effects of Noise: Means of Efficiency Estimates	108
6.3.3 The Effects of Noise: The Mean Absolute Deviations	110
6.3.4 The Effects of Noise: Correlation	112
6.3.5 The Effects of Noise: Rank Correlation	113
6.4.1 The Effects of Exogenous Variables: Exogenous Variable Statistics	114
6.4.2 The Effects of Exogenous Variables: Means of Efficiency Estimates	115
6.4.3 The Effects of Exogenous Variables: The Mean Absolute Deviations	116
6.4.4 The Effects of Exogenous Variables: Correlation	119
6.4.5 The Effects of Exogenous Variables: Rank Correlation	120
6.4.6 The Effects of Exogenous Variables: Mean Efficiency Estimates with Two- step model	123

6.4.7 The Effects of Exogenous Variables:: The Mean Absolute Deviations by Two-step Model	124
6.4.8 The Effects of Exogenous Variables: Correlation by Two-Step Model	126
6.4.9 The Effects of Exogenous Variables: Rank Correlation by Two-Step Model .	127
6.5.1 The Effects of Outliers: Mean Efficiency Estimates	129
6.5.2 The Effects of Outliers: The Mean Absolute Deviations	130
6.5.3 The Effects of Outliers: Correlation	132
6.5.4 The Effects of Outliers: Rank Correlation	132
6.6.1 The Effects of Underlying Production Technology: Mean Efficiency Estimates	135
6.6.2 The Effects of Underlying Production Technology: Mean Absolute Deviations	137
6.6.3 The Effects of Underlying Production Technology: Correlation	140
6.6.4 The Effects of Underlying Production Technology: Rank Correlation	142
8.1 Railways Efficiency: Definition of Policy and Uncontrollable Variables	157
8.2 Railways Efficiency: DEA Gross Efficiency Index	160
8.3 Railways Efficiency: Tobit Regression Results	164
8.4 Railways Efficiency: Residual Efficiency Index	169
8.5 Railways Efficiency: Deterministic Frontier Production Function	172
8.6 Railways Efficiency: Efficiency Estimates by Deterministic Frontier	173
8.7 Railways Efficiency: Stochastic Frontier Production Function	176
8.8 Railways Efficiency: Efficiency Estimates by Stochastic Frontier	177
8.9 Railways Efficiency Comparison: Mean Efficiency Estimates	178
8.10 Railways Efficiency Comparison: Correlation and Rank Correlation	179

9.1 Airline Efficiency: List of Exogenous Variables	190
9.2 Airline Efficiency: DEA Gross Efficiency Index	193
9.3 Airline Efficiency: Mean DEA Gross Efficiency by Region	195
9.4 Airline Efficiency: Tobit Regression Results	196
9.5 Airline Efficiency: The Residual Efficiency Index	200
9.6 Airline Efficiency: Mean Residual Efficiency Estimates by Region	202
9.7 Airline Efficiency: Deterministic Frontier Production Function	203
9.8 Airline Efficiency: Efficiency Estimates by Deterministic Frontier Method . .	205
9.9 Airline Efficiency: Regional Mean Efficiency Estimates by Deterministic Frontier Method	206
9.10 Airline Efficiency: Stochastic Frontier Production Function	207
9.11 Airline Efficiency: Efficiency Estimates by Stochastic Frontier Method	209
9.12 Airline Efficiency: Regional Mean Efficiency Estimates by Stochastic Frontier Method	210
9.13 Airline Efficiency: Alternative Mean Efficiency Estimates	211
9.14 Airline Efficiency: Correlation and Rank Correlation Coefficients	212
B.1 Summary Statistics for Variables in Monte Carlo Experiments	242
B.2 Determination of Number of Replications: Deterministic Frontier Method results	246
B.3 Hypothesis Test for Deterministic Frontier Method	246
B.4 Determination of Number of Replications: DEA results	247
B.5 Hypothesis Test for DEA	248
B.6 Parameter Values of the Underlying Production Technologies	249
C.1 Lists of Railways	251

C.2 Railway Characteristics	253
C.3 Railway Route Length and Traffic Volume	254
C.4 Railway Cost Recovery Indicators	255
D.1 List of Airlines	261
D.2 Airline Characteristics	262
D.3 Airline Financial Outputs	264

List of Figures

Figure

3.1 Production Function as Frontier	30
3.2 Farrell's Definition of Efficiency	32
3.3 Production Frontier and Inefficiency	38
4.1 Diagrammatic Interpretation of DEA	66

Acknowledgements

I am most grateful to my supervisor, Professor Tae Oum, for all the time, guidance, and help he has given me during all those years, and for his effort of pushing me to finish this dissertation. I am also grateful to the members of my supervisory committee, Professors Trevor Heaver, Maurice Queyranne, and Ken White, for their time, advice, and helpful comments, and to Professor Bill Waters for reading and providing comments on the draft of this dissertation and for his generous help throughout my study at UBC.

I would like to thank my husband, Li, for his patience, support and encouragement throughout the course of my study, and our children, Jimmy and Kathy, for all the joy they have brought into my life. I would like to thank my parents for their constant support and encouragement. This dissertation is dedicated to them.

I gratefully acknowledge the financial support from the Tangkakji Fellowship and the UPS Foundation Fellowship.

Overview

This thesis is concerned with measuring and explaining the productive efficiency of firms or organizations. In particular, the study compares three alternative methods to measure technical efficiency in situations where firms operate under different operating conditions and are subject to the effects of factors which are beyond managerial control.

The dissertation consists of two parts. In Part I, a Monte Carlo study is conducted to compare the relative merits of three alternative methods, i.e. the deterministic frontier method, the stochastic frontier method, and the data envelopment analysis (DEA) method, in measuring efficiency. The study focuses on the effects of exogenous variables and outliers on efficiency estimates. In Part II, the empirical performance of the three methods is examined through their applications to two cases in the transportation industry. The first application is to the passenger railways - a small sample case where the services are mostly provided by highly regulated, nationalized firms. The effects of exogenous variables are expected to be significant, and outliers might be present in the sample. The second application is to the international airlines - a medium sample case where the firms operate in a fairly competitive environment, and they have access to essentially the same technology available even though there is a high degree of diversity in size.

Since Farrell (1957)'s work, the development and refinements of various methods for measuring efficiency have progressed rapidly. However, there are limited studies which compare the relative merits of the different methods. Most of these studies compare results from the application of different methods to the same empirical data (such as Banker, Conrad and Strass, 1986, Bjurek, Hjalmarsson and Forsund, 1990). The main motivation for these

comparative analyses is the desire for the alternative methods to tell consistent stories about the efficiency performance of the firms or Decision Making Units (DMUs) under study. In this regard, these studies are applying the "methodology cross-checking" principle. However, further insight into the relative performance of the various methods can be obtained by applying different methods to a mixture of efficient and inefficient observations generated using a "known" production technology model. Using this approach one can evaluate these different methods not only relative to each other, but also relative to the "known" underlying model.

There have been a few studies which use simulated data to evaluate the relative performance of different methods with the knowledge of the "true" production and inefficiency structures, such as Banker, Charnes, Cooper and Maindiratta (1988), Gong (1987), Li (1991), Gong and Sickles (1992), and Banker, Gadh and Gorr (1993). However, none of these comparative studies have addressed the issue of outliers and data errors explicitly, although one of the main critiques of DEA is its sensitivity to outliers and data errors. Improved knowledge in this aspect could help provide guidelines for the selection of methodology in practical applications. Another issue that has not been addressed in these previous studies is the effect of exogenous factors on efficiency measures. In most practical applications, one would be most likely to have a group of firms (or decision making units) which operate under different conditions. The observed productive performance of a firm is the result of both its productive efficiency and its environment. In making efficiency comparisons, one must separate the effects of the production environment and the effects of the productive efficiency. The role of exogenous variables in measuring the efficiency of

transportation firms is especially important because transportation production and cost depend heavily on network and market characteristics of the firm. In some previous empirical studies, researchers have attempted to take into account the effects of exogenous variables. In general, there are two basic procedures to account for the effects of exogenous variables: (1) an one-step procedure which includes the exogenous variables directly in estimating the parametric frontier production functions, and (2) a two-step procedure which first estimates the relative "gross" efficiencies using only inputs and outputs, then analyzes the effects of the exogenous variables on the "gross" efficiency. It is necessary to examine how well each of these methods incorporates the exogenous variables in order to provide helpful guidelines in selecting an appropriate method for a particular practical problem.

In Part I of this thesis, Monte Carlo experiments are carried out to examine the relative performance of the three alternative methods with respects to the effects of the sample size, the variations in input values, the noise level, the exogenous variables, the data outliers, and the different (underlying) production structures. The results show that the performance of all three methods (in terms of correlations) improves with the sample size, but deteriorates sharply with the noise level. The variations in input ranges do not appear to have much effects on the performance of the alternative methods in the cases where input substitution is over one, however, the performance of the two parametric methods is found to fall noticeably in the cases where weak input substitution exists. The results also show that the magnitude of exogenous variables does not appear to have any significant effects on the performance of the one-step deterministic frontier method and the one-step stochastic frontier method as long as the exogenous variables are correctly identified and accounted for.

However, for the two-step procedure, especially for the DEA and the deterministic frontier method, the effects of the exogenous variables are significant. The stochastic frontier method, as expected, is quite robust with respects to outliers. By comparing their performance in terms of mean efficiency estimates, the DEA method appears to be more robust to the outliers than the deterministic frontier method, although both methods have similar performance in terms of correlations. For the parameter range considered in this study, the performance of the stochastic frontier method does not rely much on the structure of the underlying production technology. However, it is found that the performance of the deterministic frontier method is affected by the presence of input complementarity but not by the returns to scale. The performance of the DEA method, on the other hand, is found to deteriorate when the returns to scale increases and/or input complementarity exists. It is found in general that the stochastic frontier method surpasses the other two methods in all aspects examined.

The Monte Carlo study in Part I is an extension to the works by Gong (1987), Li (1991), and Banker, Gadh and Gorr (1993). It provides some useful information regarding the relative merits of the alternative methods in the presence of exogenous factors and outliers (data errors) as well as in situations where input complementarity and non-constant returns to scale exist.

The applications of the three alternative methods to the railways and the international airlines in Part II of this thesis provide empirical evidence on the relative performance of these methods. The studies of efficiency performance of the sample railways and airlines also provide useful information on the effects of some policy variables and some other

factors on the efficiency performance of the industries.

There are some previous studies examining the relative performance of the railways across different countries, however, none of these studies has examined the effects of public subsidies on the efficiency performance of the railways. In addition, there has been no study comparing the results from applications of the DEA and the parametric methods to the railway industry. Chapter 8 measures the productive efficiency of the railway systems in 19 OECD countries for the period of 1978-89, identifies the effects of government intervention and subsidization on the productive efficiency of those railways, and compares the alternative efficiency estimates from the three methods.

The empirical results from the railway study show that railway systems with high dependence on public subsidies are significantly less efficient than similar railways with less dependence on subsidies, and railways with high degree of managerial autonomy from regulatory authority tend to achieve higher efficiency. Therefore, the institutional and regulatory framework for railway industry must squarely address the question of railways' managerial freedom, and subsidy policy must encourage railways to use normal market mechanisms for improving their cost recovery and to use the subsidies only for improving their services. The empirical results also indicate that efficiency measures may not be meaningfully compared across railways without controlling for the variations in railways' operating and market environments. Comparison of the efficiency estimates from the three alternative methods confirms that on average the stochastic frontier method yields higher efficiency estimates than the other two methods, as indicated by the simulation results in Part I. The efficiency estimates by the two parametric methods are highly correlated although

their mean values are different. There are substantial differences between the efficiency estimates from the DEA-TOBIT and those from the parametric methods even though the policy implications from all three methods are consistent. The main reason for the differences is that the DEA-TOBIT is a two-step procedure whereas the parametric methods utilize an one-step procedure. As shown by the Monte Carlo simulation study in Part I, the one-step procedure would produce different estimates from those using the two-step procedure. Further, the DEA-TOBIT considers two outputs, i.e. passenger kilometers and freight tonne-kilometers, while the one-step parametric production functions consider only one output, i.e. total train kilometers.

The efficiency performance of 36 international airlines during the period of 1980-1992 is examined and analyzed in Chapter 9. Previous studies on the performance of airlines are mostly based on late 1970s and early 1980s data. There has been no comparative study on alternative efficiency estimates for international airlines. This study includes more up-to-date data to evaluate recent changes in efficiency performance of the airlines, and compares the alternative efficiency estimates. Some other factors which have been ignored in the previous studies are also included in this study, such as incidental services. Incidental services refer to catering services, ground handling services and maintenance services performed for other airlines, etc. These services could account for up to 30 percent of total operating revenues for some airlines¹. To properly reflect the "total" output of an airlines,

¹ Good, Nadiri, Röller, and Sickles (1993) does consider the incidental services as one of airlines' outputs. However, their studies are limited to the four largest European airlines and their U.S. counterparts. They attempt to control for the effects of some operating characteristics, but ignore the government ownership variable.

the incidental services are treated as one of the outputs in this study.

The efficiency of the 36 sample airlines are measured and compared, and the effects on efficiency of government ownership and technical progress are identified. The empirical results show that technological progress has improved the overall performance of the international aviation industry, especially for some of the major Asian carriers. The airlines with majority government ownership are shown to be less efficient than other airlines with similar operating characteristics. The results also indicate that the effects of network and market environments should be controlled for in order to measure productive efficiency meaningfully comparable across airlines.

Comparison of results from the alternative methods illustrates that although there are noticeable differences in the levels of efficiency, the overall pattern of the efficiency estimates from the two parametric methods are essentially the same, and the results by the DEA-TOBIT are broadly similar to those from the parametric methods. The policy implications from all three methods are consistent.

By comparing the relative performance of the alternative methods in the railway and the airline cases, it is noted that there is much less discrepancy and hence higher correlations among the alternative efficiency estimates in the airline case than in the railway case. This is partly due to the better data quality for the airlines. The airlines' operating environments have less variations than those for the railways. In addition, the airline data set has a larger sample size and is collected in a more consistent manner than the railway data set. The higher correlations under these circumstances are consistent with the results from the Monte Carlo experiments in Part I which shows that when the magnitude of the exogenous factors

and/or the random noise level are high the efficiency estimates by the three alternative methods may deviate considerably from the "true" efficiency. Thus, larger discrepancies among the alternative efficiency estimates are expected for the railway case. One empirical lesson from this study is that extra care is warranted when examining the data and the estimation results in situations where large variations in the operating environments exist. Although the application of the stochastic frontier method is expected to yield better results in general, computational and data problems, such as multiple outputs, collinearity among the inputs, and the lack of necessary price data, may leave Data Envelopment Analysis as the only choice for certain empirical situations. There is no definite answer to the question of which method is the best method, the answer relies on the particular situation in question. The Monte Carlo results in Part I provide some general guidelines on the relative merits of the three alternative methods under certain known conditions. The applications in Part II, on the other hand, demonstrate some of the problems that may be encountered and also serve as examples of how to deal with the practical problems where there is no prior knowledge of either the production technology nor the efficiency profile.

Part I

A Monte Carlo Comparison of Alternative Methods for Efficiency Measurement

Chapter 1

Introduction

1.1 Background

Since the pioneering work of Farrell (1957), much has been accomplished in the field of efficiency measurement in terms of both theoretical studies and empirical applications. Various methods, such as Data Envelopment Analysis (DEA) methods, stochastic frontier methods, etc., have been developed to measure productive efficiency. Different methods often yield different efficiency rankings among the firms being considered, and may lead to different policy implications on how to improve the efficiency of a particular firm and of the overall industry. Each method has its strengths and weaknesses. Knowledge of these strengths and weaknesses will help researchers and policy analysts to choose the most "suitable" method for a particular situation, and thus to make more accurate measurements of efficiency. These in turn will help policy makers to make appropriate policy decisions. Therefore, it is important to study the relative merits of different methods in terms of their abilities to reveal the structure of production technology and the nature and extent of inefficiency under different conditions.

The idea of productive efficiency proceeds from the concept of the production function. The *production function* specifies the maximum quantities of realizable output, given any level of inputs and, for any given level of output, the minimum quantities of inputs needed for producing these outputs. That is, the production function describes a boundary, or a frontier which sets a bound on the range of possible observations. Any production

situation, given the definition of the frontier, is deemed inefficient if the outputs and inputs are represented by a point below the production frontier. On the other hand, a situation will be deemed efficient if the point is on the production frontier itself¹. The amount by which a production situation lies below its production frontier provides a measure of efficiency. Thus, the production function acts as a criterion, or yardstick, serving as a base for assessing efficiency. Efficiency measurement, therefore, involves two aspects. The first is to define the production frontier and the second is to measure the efficiency using the production frontier as a yardstick.

In practice, the true production frontier is not likely to be known, thus one needs to construct an empirical production frontier using observations on inputs and outputs. The subsequent measurements of efficiency vary according to the chosen functional form and error distribution of the frontier as well as the size and composition of the sample. Therefore, the construction of the production frontier is crucial in evaluating the relative performance of a group of firms in an industry. There are many ways of estimating a production frontier. The various estimation methods can be classified into two basic approaches, parametric and nonparametric, depending on whether or not the frontier can be specified as a function with constant parameters. The *parametric approach* estimates a parametric representation of the production frontier using econometric techniques². Within

¹ Note that the idea of considering the production function as a frontier may be applied to the cost function too. Efficiency can then be defined on the basis of the minimum cost function.

² Some studies use linear programming techniques to estimate a parametric frontier function, such as Aigner and Chu (1968) and Försund and Hjalmarsson (1979).

the parametric approach, the estimation of production frontiers has proceeded along two general paths: (1) *deterministic frontiers*, which force all observations to be on or below the frontier so that any deviation from the frontier is attributed to inefficiency; and (2) *stochastic frontiers*, where a deviation from the frontier is decomposed into a random component reflecting measurement error and statistical noise, and a component reflecting inefficiency.

The *nonparametric approach*, represented by the Data Envelopment Analysis (DEA) method (Charnes, Cooper and Rhodes, 1978), uses mathematical programming techniques to envelop observed input-output vectors as tightly as possible without requiring a prior specification of functional forms for the production frontiers. That is, it does not assume that the underlying technology "belongs to a certain class of functions of a specific functional form which depend on a finite number of parameters, such as the well-known Cobb-Douglas functional form" (Diewert and Parkan, 1983). It only requires the assumption of convexity of the production possibility set and disposability of inputs and outputs. It employs a postulated minimum extrapolation from observed input-output data in the sample. The efficiency of a Decision Making Unit (DMU) is measured relative to all other DMUs in the sample.

In the literature on measuring the performance of production units, the terms of productivity and efficiency are often used interchangeably. In this dissertation, productivity and efficiency are considered as separate but related concepts. *Productivity* refers to the ratio of outputs to inputs. *Efficiency* involves the comparison between observed and optimal values of outputs and inputs of a production unit, and it is a component of productivity. With this distinction, this dissertation will not consider the methodologies for measuring

productivity such as total factor productivity (TFP).

1.2 Research Issues

The development and refinements of the various methods for measuring efficiency have progressed rapidly during the last two decades. However, there are limited studies which compare the relative performance of the different methods in efficiency measurement. Most of these studies compare results from the application of different methods to the same empirical data (such as Banker, Conrad and Strass 1986, Bjurek, Hjalmarsson and Forsund 1990). The main motivation for these comparative analyses is the desire for the alternative methods to tell consistent stories about the efficiency performance of the firms or DMUs under study. In this regard, these studies are applying the "methodology cross-checking" principle. However, further insight into the relative performance of the various methods could be obtained by applying these different methods to a mixture of efficient and inefficient observations generated using a "known" production technology model. In this way, it will be possible to compare these different methods in terms of their accomplishments not only relative to each other, but also relative to the "known" underlying model. There have been a few studies which use simulated data to evaluate the relative performance of different methods with the knowledge of the "true" production and inefficiency structures. Banker, Charnes, Cooper and Maindiratta (1988) compares the performance of the DEA method and a deterministic translog frontier function under two "known underlying technologies" - piecewise loglinear and translog, first for a sample of 500 randomly generated observations and then for the set consisting of the first 100 of the 500 observations. Banker, Gadh and

Gorr (1993) extends the work of Banker, Charnes, Cooper and Maindiratta (1988) by introducing measurement errors and different efficiency distributions. Li (1991) investigates the relative performance of the DEA and the stochastic frontier method in estimating firm specific efficiency levels with regards to three aspects: (1) different inefficiency profiles, (2) different returns to scale, and (3) different noise levels. Gong and Sickles (1992) compares the performance of the stochastic frontier method (with panel data) and the DEA for the cases in which: (1) the complexity and structures of an underlying technology differ, (2) the relative size of technical inefficiency to statistical noise in the stochastic components differs, (3) the forms of the true structure of technical inefficiency vary, (4) input levels and technical inefficiency are allowed to have an arbitrary degree of correlation.

One issue that is not considered in the previous comparative studies is the effect of exogenous factors on the efficiency measures. In any realistic situation, the quantity of output produced by a firm is determined by a large number of factors other than inputs. Usually only a few of these factors are controlled by the firm and can be varied at the discretion of the decision makers. Other factors, not necessarily less important, are exogenously determined and serve as operating constraints within which production decisions are to be made. For example, in the transportation industry, the output level, as measured by revenue passenger kilometers and revenue tonne kilometres, is affected by shipment size, load factor, spatial pattern of networks, the general economic condition, the population density, and the extent of development of other transport modes etc.. Some of these factors,

such as shipment size and load factor, are partially controllable by the carriers³. However, other factors, such as economic condition and population density, are beyond the carriers' control. Consider a firm whose output level is fixed⁴, the firm must minimize its input consumption in producing the given level of output in order to be productively efficient. When the actual level of input consumption exceeds the optimal level of input requirement as specified by the production function, inefficiency occurs. However, this observed "inefficiency" could be attributed to "bad" weather condition, which is out of the firm's control, or to unnecessary waste of certain inputs, which could be avoided by efficient production. Thus, when evaluating public policies and management strategies designed to improve performance, it is essential to separate the effects of the production environment from the effects of productive efficiency. In making efficiency comparisons, one must take into account differences in the firms' environments, otherwise the efficiency measures would reflect not merely the differences in efficiency but also the degree to which the environment of a particular firm is favourable or unfavourable. The role of exogenous variables in measuring the efficiency of transportation firms is especially important because transportation production and cost depend heavily on network and market characteristics of the firm.

The parametric methods can incorporate these exogenous variables directly in estimating the frontier function (Lee and Schmidt, 1993). This is referred to as the one-step procedure in this dissertation. However, the DEA method cannot account for the effects of

³ Some of these decisions are subject to governmental control.

⁴ In many cases output is essentially determined by the market condition, governmental control *etc.*, therefore it can be considered as fixed.

exogenous variables when computing the efficiency indices⁵. It is necessary to use a second stage, such as regression analysis with the DEA index as dependent variable, to account for the variables beyond managerial control and to identify the effects of these exogenous variables on the observed efficiency (Ray, 1991, and Oum and Yu, 1994). This two-step procedure can also be used with the parametric methods as well (Bruning, 1991, and Loeb, 1994). In practice, one may face situations with different degrees of variations in the firms' operating environments. For example, one may have a sample of firms operating in very different environments in the transportation industry. Under such conditions, the two-step procedure might be more useful to the decision makers since it relates the exogenous variables directly to the firms's performance. On the other hand, the differences in firms' operating environments in a manufacturing industry are likely to be to a much lesser degree. The production of a particular sets of products is more likely to follow a clear engineering design where there is better information about the "true" production technology. In such cases, the one-step procedure might work better. These arguments are mostly speculations, as there is not yet obvious methodological and empirical support. Knowledge about the relative merits of the alternative methods under such different situations would help the

⁵ Banker and Morey (1986) takes into account non-discretionary inputs or outputs which are not subject to managerial control. However, in applying their model to incorporate the effects of the exogenous variables, one has to treat these exogenous variables either as nondiscretionary inputs or outputs, thus affecting the resulting efficiency scores. Another problem with the Banker and Morey (1986) model is that it requires an assumption of free disposability of the nondiscretionary inputs. That is, the decision-maker is allowed to under-utilize any available nondiscretionary inputs. This is not necessarily a realistic assumption. For example, weather condition is a nondiscretionary input in transportation, since the carriers have no control over the amount of "bad" weather. Thus, they can have neither more nor less than the exogenously determined level of "bad" weather.

researcher to select the appropriate method for a particular problem. To my knowledge, there have not been any studies comparing the relative performance of the various methods in measuring efficiency while taking into account the effects of exogenous factors. Chapter 6 of this thesis will attempt to investigate this issue using Monte Carlo experiments where the underlying production technology is known.

Another issue that is not considered in the previous comparative studies is the effect of outliers. One of the main criticisms of the DEA is its sensitivity to outliers and data errors (Sexton, Silkman, and Hogan, 1986, and Bauer, 1990) because it is an extremal method. However, there is no solid evidence on how sensitive the DEA is to outliers and data errors, and how it compares to the parametric methods in this regard. Except for Mensah and Li (1993), none of the previous comparative studies addresses the issue of outliers explicitly. Mensah and Li (1993) attempts to evaluate the effects of outliers on efficiency estimates by the DEA and the translog models. They first remove one efficient observation to reestimate the efficiencies, and then compare the new estimates with the original ones. Their study is based on observed empirical data. However, further investigation with the knowledge of the "true" production and efficiency structure will improve our understanding of the effects of outliers and data errors, and this will be discussed in Chapter 6 of this thesis.

1.3 Objective of Part I

The objective of Part I is to evaluate the DEA and the parametric frontier methods in measuring efficiency under certain less than ideal conditions, particularly, when there are

considerable differences among the firms in question, and to provide some guidelines for the selection of methodology in practical applications. This study focuses on the effects of exogenous variables and the effects of outliers (and data errors) on efficiency measurement. To accomplish this objective, the Monte Carlo technique is used. It allows for control of the underlying production technology and the "operating environments". In particular, we evaluate the comparative performance of the alternative efficiency measurement methods in terms of the sample size, different functional forms for the underlying production technology, and in the presence of exogenous variables and outliers. The comparisons are made among three most popularly used models: the deterministic frontier method, the stochastic frontier methods, and the DEA method. This study serves as an extension to the works by Li (1991), Gong and Sickles (1992), and Banker, Gadh and Gorr (1993).

1.4 Organization of Part I

The previous comparative studies are reviewed in Chapter 2. The review is limited to comparative studies because of the large volume of theoretical and empirical literature in the field of efficiency measurement. Chapter 3 presents the basic definitions of efficiency adopted in this dissertation. Chapter 4 then describes the various efficiency measurement methods and reviews the related theoretical literature. Chapter 5 outlines the general design of the Monte Carlo experiments. Chapter 6 discusses the experimental results. Chapter 7 summarizes the general results and provides some guidelines for practical applications.

Chapter 2

Literature Review of Comparative Studies On Efficiency Measurement

The objective of this chapter is to give an overview of previous *comparative* studies on efficiency measurements. The studies are grouped into two categories: studies based on empirical data and studies based on simulated data. The related literatures on the development of various methods are reviewed in Chapter 4 where we discuss the alternative methodologies.

2.1 Studies based on empirical data

Most of the comparative studies on efficiency measurement are based on specific empirical data, and are *ad hoc* in nature. van den Broeck, Førsund, Hjalmarsson and Meeusen (1980) compares the deterministic and stochastic methods for estimation of a frontier production function, based on cross-section data for 28 Swedish dairy plants for the period of 1964-73. They estimate the stochastic frontier using a maximum likelihood estimator, and estimate the deterministic frontier using a combined linear programming and maximum likelihood approach. Their comparison focuses on characteristics of the estimated frontier functions, such as input elasticity and scale elasticity, as well as their relationships to the average production function. They did not compare the firm specific efficiency measures.

Corbo and de Melo (1986) uses the 1967 Chilean manufacturing census data to

compare four different methods: (1) the modified linear programming approach proposed by Aigner and Chu (1968)¹; (2) the deterministic statistical frontier by corrected ordinary least squares (COLS) method (Richmond, 1974); (3) the deterministic statistical frontier by maximum likelihood estimation method (MLE) (Greene, 1980); and (4) the stochastic frontier approach using COLS. They focus their comparison on the sensitivity of the results to model selection, including the selection of functional form for the average production function, and the selection of error structure and specific characteristics of the distribution of error terms. The study finds that within the deterministic frontier methods, the linear programming and the deterministic statistical model yield highly correlated measures of technical efficiency. It also finds that the estimates are sensitive to the selection between deterministic statistical and stochastic formulations, but the choice of error structures has a very small impact on the measurement of inefficiency. The general result of their study is that the different methods to measuring technical efficiency yield broadly similar results. Although they carry out a rather comprehensive comparison of the frontier function approaches in measuring efficiency, the nonparametric methods are not included in their analysis.

Banker, Conrad and Strauss (1986) applies a translog frontier cost function and the Data Envelopment Analysis method to a sample of North Carolina hospitals to examine inferences about the hospital cost and production functions, and to compare the results from the two different methods. The translog results indicate the presence of constant returns to

¹ Schmidt (1976) shows that, in a logarithmic mode, the linear programming procedure is equivalent to the deterministic statistical model with an exponential error structure.

scale, while the DEA estimates suggest both increasing and decreasing returns to scale in different segments of the production functions. The DEA efficiency estimates are found to be closely related to the degree of capacity utilization, but no such relationship is found for the translog estimates. The study is concerned only with the deterministic case, it does not include stochastic frontier models.

Levitt and Joyce (1987) uses both conventional regression analysis and the DEA to examine the relative performance of police authorities in Great Britain. They find that the DEA efficiency rankings and those from the regression analysis, "while not identical, have a good deal in common". Their conclusions are derived from a direct comparison (observation by observation) of the rankings from the two methods.

Bjurek, Hjamarsson and Forsund (1990) compares three different methods of efficiency estimation on the basis of a sample of local social insurance offices of the Swedish social insurance system. The three methods considered are: (1) a Cobb-Douglas deterministic frontier production function, (2) a Quadratic deterministic frontier production function, and (3) the Data Envelopment analysis. They find that the differences in the efficiency estimates between the methods are very small. Stochastic frontier models are not included in their comparison.

Ferrier and Lovell (1990) compares two methods for estimating production economies and efficiencies. One is the stochastic frontier cost function method, and the second is a variant of the DEA method which constructs a production frontier through a series of linear programs. They find that the two methods yield very similar results regarding returns to scale, but dissimilar results regarding estimated costs due to inefficiencies. Their comparison

emphasizes the amount of increased costs due to inefficiency, rather than the firm specific efficiency measures.

Land, Shutler and Kirthisingha (1990) conducts a comparison between the conventional regression method and the DEA in measuring the efficiency of the post offices in the United Kingdom. With the regression analysis, they measure the relative efficiency using the residuals from the regressions and the "unit residuals" which are the regression residuals divided by predicted values. Comparisons are made over two separate data sets, the 1987 data set and 1989 data set. Simple correlation, Spearman rank correlation and Kendall rank correlation are used to compare the efficiency rankings from regression against the DEA efficiency rankings. The results do not show a clear pattern between the efficiency estimates from the two methods. They believe that the two methods are complementary in measuring efficiency and assisting management to improve efficiency.

Forsund (1992) estimates a non-parametric frontier and a deterministic parametric frontier to seek out the most efficient firm among the observed Norwegian Ferries. The study finds that the methods yield very similar results for individual efficiency scores. The differences are mainly to be found at the extreme ends of the size distributions. The estimated scale properties are different from the two methods. The study focuses on similarities or dissimilarities of the distributions of the efficiency scores, and scale properties. The study does not consider stochastic frontier models.

Sickles and Streitierwer (1992) estimates the firm specific levels of efficiency of the U.S interstate natural gas transmission industry using a stochastic frontier model and the DEA. They find that the levels and relative rankings of efficiencies from the two methods

are often not the same. However both methods indicate a downward trend in firm efficiency over the sample period.

Fecher, Kessler, Perelman and Pestieau (1993) examines the productive performance of the French insurance industry using the DEA method and a stochastic frontier model based on a Cobb-Douglas production function. They find that there is a high correlation between estimated efficiency measures from the two methods, and the frequency distributions of firm specific efficiency measures from the two methods also show similar pattern.

In summary, all the comparative studies based on empirical data have one thing in common, that is, they are mainly motivated by the desire for the alternative methods to tell consistent stories about the efficiency performance of the firms or DMUs under study. In this regard, these studies are applying the "methodology cross-checking" principle since most empirical work is hindered by the lack of knowledge of the true structure of production and efficiency. From the discussions above, it is noted that in most cases the alternative methods yield similar estimates regarding the efficiency rankings among the firms or DMUs under consideration, but the level of estimated efficiencies often depends on the choice of method used. Further insight into the relative performance of the various methods could be obtained by applying these different methods to a mixture of efficient and inefficient observations generated using a "known" production technology model. This approach would make it possible not only to compare these different methods in terms of their accomplishments relative to each other, but also make it possible to compare what each accomplishes relative to the "known" underlying model. There have been a few studies which use simulated data to compare the relative performance of different methods with the knowledge of the "true"

production and inefficiency structures.

2.2 Studies based on simulated data

A number of studies compare the performance of different methods based on simulated data, thus they can compare the results from each of the alternative methods with the "true" production technology. Bowlin, Charnes, Cooper and Sherman (1985) uses artificially generated data for a group of hypothetical hospitals to test the DEA relative to output-input ratio analysis (including index numbers) and regression approaches (linear regressions of cost against outputs in particular) in terms of locating sources and estimating amounts of inefficiency in the observed inputs and outputs. Their study finds that the DEA method generally performs better than regression analysis and ratio analysis in identifying sources and amounts of inefficiency in all except a very few cases. They carry out the comparison under one specific "true" production situation, and consider only a deterministic case.

Using the same hypothetical hospitals, Thanassoulis (1993) extends Bowlin, Charnes, Cooper and Sherman (1985) by comparing DEA and regression analysis, on estimates not only of relative efficiencies but also of marginal input or output values and target levels. The study concludes that DEA outperforms regression analysis on the accuracy of estimates but regression analysis offers greater stability of accuracy. Again the results are based on a "specific" sample, and are limited to a deterministic case.

Banker, Charnes, Cooper and Maindiratta (1988) uses simulated data to compare the estimates of technical efficiency, returns to scale, and rates of substitution obtained by the

DEA method and a deterministic translog frontier function. Under two known underlying technologies -piecewise loglinear and translog, the comparison is carried out first for a sample of 500 randomly generated observations (from an uniform distribution) and then for the set consisting of the first 100 of the 500 observations. The study concludes that DEA could approximate the known underlying technology better than the translog frontier function and the accuracy of DEA estimates improves for the larger sample size. However, it is found that the DEA method tends to misclassify inefficient "corner" points² as efficient. The study does not consider the stochastic models.

Banker, Gadh and Gorr (1993) extends the work of Banker, Charnes, Cooper and Maindiratta (1988) by introducing measurement errors in comparing the performance of the corrected ordinary least squares method and the data envelopment analysis method in estimating efficiency. The study considers two piece-wise Cobb-Douglas technologies as the underlying true production technology, four efficiency distributions (two half-normal and two exponential), and four different sample sizes from 25 to 200. They use the mean absolute deviation of true versus estimated efficiencies as the performance criterion of the two alternative estimation methods. The study finds that COLS performs better with half-normal efficiency distribution when sample size is over 50, but DEA performs better in other cases. However, both methods perform poorly with high measurement errors. They also find that COLS fails to decompose deviations into efficiency and noise components in almost all cases. Because of the failure of the COLS procedure, the study is in fact limited to deterministic

² A "corner" point refers to an observation with a very small or very large quantity for at least one of the inputs or the outputs.

methods.

Gong and Sickles (1989) investigates the relative performance of three stochastic frontier estimators in terms of their sensitivity to the complexity and structure of the underlying technology. The three estimators considered are: maximum likelihood, generalized least squares and dummy variables (or the within estimator). They conclude that the ability of the stochastic frontier models to estimate firm-level efficiency is quite sensitive to the complexity and structure of the underlying production technology. The study is concerned only with the stochastic frontier models.

Gong and Sickles (1992) conducts a rather comprehensive comparison of the performance of the stochastic frontier method (with panel data) and the DEA in estimating firm specific technical inefficiency for cases in which: (1) the complexity and structure of the underlying technology differ, (2) the relative size of technical inefficiency to statistical noise in the stochastic components differs, (3) the form of the true structure of technical inefficiency varies, and (4) input levels and technical inefficiency are allowed to have an arbitrary degree of correlation. Their results indicate that, in terms of correlation and rank correlation coefficients between the estimated efficiency level and the "true" efficiency level, DEA is dominated by the stochastic frontier models in most cases except for the case where there is high correlation between inputs and technical efficiencies. However, the study considers only the case of constant returns to scale, and does not address the issues of exogenous variables and outliers explicitly.

Li (1991) examines the effectiveness of the DEA and the stochastic frontier model in estimating firm specific efficiency levels with regards to three aspects: (1) different "true"

inefficiency patterns; (2) different returns to scale specifications; and (3) various levels of random noise. The study finds that the general performance difference between the two methods is quite small, and both methods perform quite well. The study assumes a Cobb-Douglas function with one output and two inputs as the underlying production technology. The results are based on a simulated data set of 100 observation with 5 replications for each experiment. The artificially generated data set simulates a situation where firms or DMUs are fairly homogenous, since the allowed variations of the input variables and the random noise are rather small. Input variables are generated from a uniform distribution over the intervals of (10,20) and (20,30), respectively, and random noise is assumed to have a mean of zero, standard deviations ranging from 0.01 to 0.09, which may explain the small variations among the experimental results.

In summary, there is a limited number of comparative studies based on simulated data where the "underlying" production technology and efficiency structure are known. From the discussions above, it is clear that a researcher can learn more about the performance of the alternative methods in measuring efficiency by applying the methods to hypothetical production situations. Since the characteristics of these hypothetical production technology and structures of efficiency are artificially controlled, the researcher is able to achieve in depth understanding of the abilities of the alternative methods in a particular aspect, such as the effects of the sample size, the effects of random noise, *etc.* These comparative studies have provided useful information about the relative merits of different methods in terms of their abilities to reveal the structure of production technology and the nature and extent of inefficiency under different conditions. However, none of the studies

has looked at the effects of exogenous variables on efficiency estimates. In any realistic situation firms operate under different conditions, and one should not neglect differences in these firms' environments in making efficiency measurement and comparison. Therefore, it is important to consider the effects of the exogenous factors on the observed productive performance of the firms. There is another point that should be mentioned here, that is, the problem of outliers and data errors. Most of the studies in efficiency measurement have recognized the problem of outliers and data errors, especially for the deterministic methods. However, there has been no explicit discussions about the magnitude of the potential effects of outliers and data errors. It would be interesting to see some solid evidence on these potential effects.

Chapter 3

The Concept of Efficiency

The purpose of this chapter is to discuss the basic concept of efficiency. To measure the relative efficiency performance of firms or organizations in an industry, we need first to define what we mean by efficient production.

The concept of productive efficiency proceeds from the concept of production function. A *production function* specifies the maximal output obtainable from a given input vector. Thus a production function can be interpreted as a *frontier*, delineating the limits of what a firm or an organization can achieve. Figure 3.1 shows the concept of production frontier for a one-input one-output situation. The frontier production function is denoted by $f(x)$ where x denotes the input level. All points on or below it, such as B, C, or D, are deemed realizable, hence can be observed, whereas points beyond it, such as E, are neither realizable nor (in the absence of noise or measurement errors) can be observed. This frontier interpretation of production function leads to the concept of productive efficiency. A productive situation is *inefficient* if its output-input point lies below the frontier, such as D and C in Figure 3.1, since it does not do as well as it could with the same inputs, as specified by the production function. On the other hand, a situation is *efficient* if the output-input point is on the production frontier itself, such as B in Figure 3.1. The "distance" an observed production point deviates from the frontier provides a measure of efficiency for the corresponding firm or organization. Therefore, the frontier production function acts as a criterion, or norm, serving as a base for assessing efficiency.

In his path-breaking paper Farrell (1957) gives an explicit definition of efficiency, and

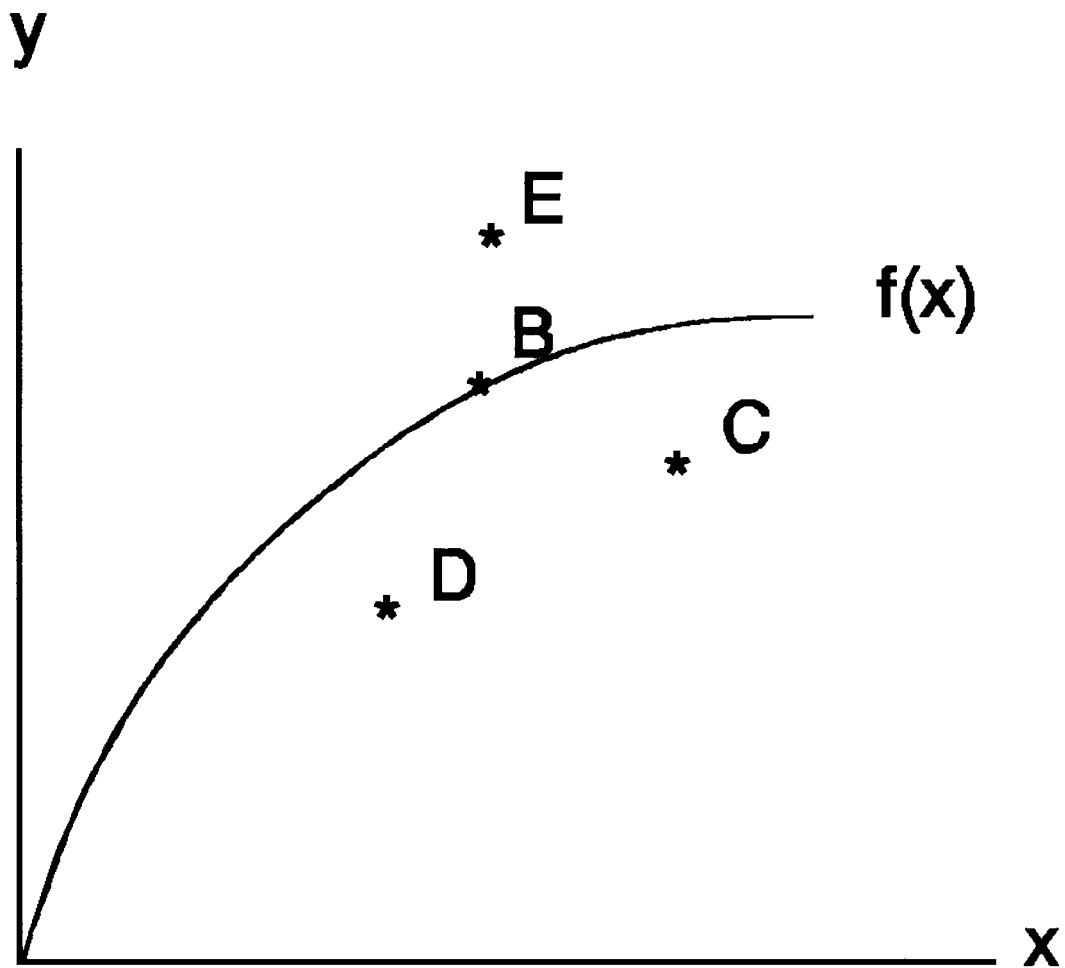


Figure 3.1 A Production Function as a Frontier

provides a computational framework for subsequent studies on productive efficiency. In the remainder of this chapter, we first look at Farrell's definition of efficiency, then relate the efficiency measurement to production function.

3.1 Farrell's Definition of Efficiency

Consider the simple case of a firm producing a single output, y , from two inputs, x_1 and x_2 . Suppose the firm's production function (frontier) may be written as $y=f(x_1, x_2)$. If we also assume that the firm produces under conditions of *constant returns to scale*, then the production function is homogeneous of degree 1 and the equation of the frontier can be written as $1=f(x_1/y, x_2/y)$. This means that we can represent the technical possibilities open to the firm in terms of a unit isoquant such as $ACA'A''$ shown in Figure 3.2. Potential or "maximal" performance is defined along $ACA'A''$, the frontier. No firm could produce at a point below $ACA'A''$ because this would not be technically feasible. On the other hand, a firm producing at any points above $ACA'A''$ uses more of at least one input than that is needed.

Suppose now that the available budget is represented by the line PP' which is tangent to $ACA'A''$, all points along PP' have the same cost, that is, PP' is the current *isocost line*. Its slope reflects the ratio of input prices. Therefore, the firm producing at a point on PP' is using the "optimal" input proportion.

According to Farrell (1957), a firm is *technically efficient* if it chooses an input mix

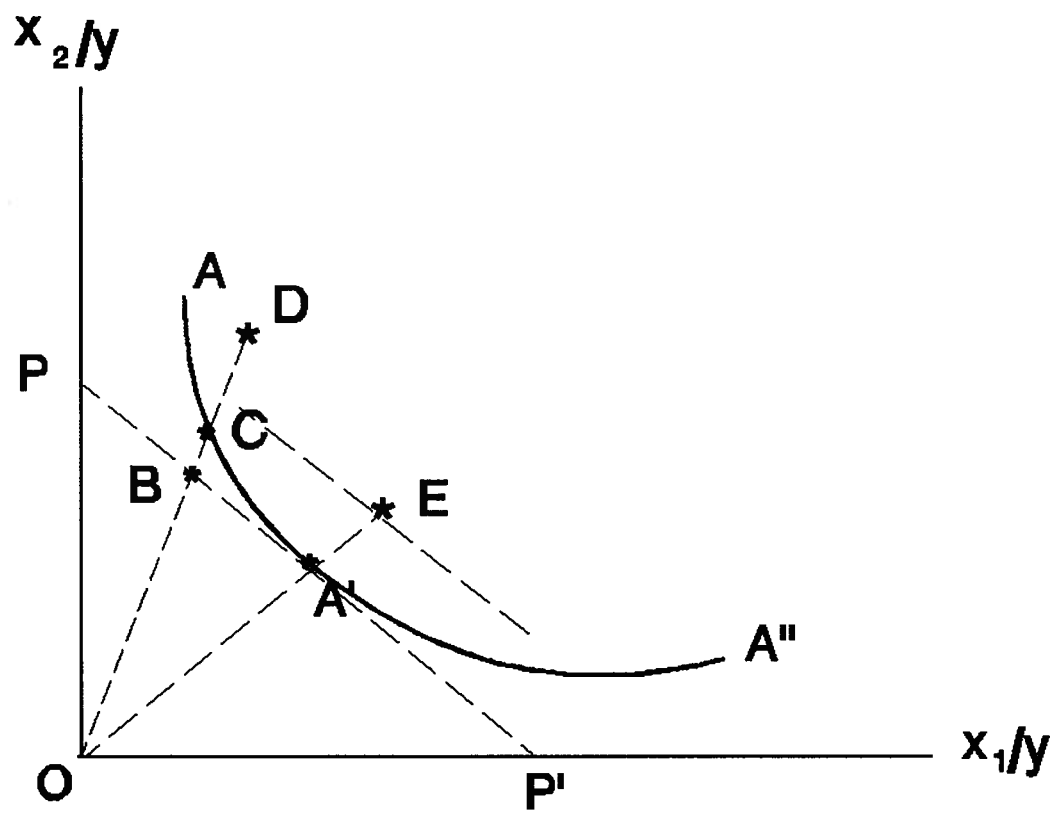


Figure 3.2 Farrell's Definition of Efficiency

on the unit isoquant, and a firm is *allocatively efficient* (price-wise efficient) if the marginal rate of substitution between the two inputs is equal to the input price ratio. In other words, technical inefficiency is due to excessive input usage (given the level of output); allocative inefficiency results from employing inputs in the wrong proportions. Allocative efficiency is significant in that it emphasises that the frontier production *per se* is not sufficient to minimize costs. Full efficiency requires simultaneous technical and allocative efficiency which is obtained at A' . Firm C is technically efficient, but is allocatively inefficient, while firm E is allocatively efficient but technically inefficient.

Suppose a firm is observed using (x_1^0, x_2^0) to produce y^0 , let D in Figure 3.2 represent $(x_1^0/y^0, x_2^0/y^0)$, which cannot lie below $ACA'A''$ by definition. At point D the ratio of x_1 and x_2 is identical to the ratio at all points along OD. On this OD line, C is the most technically efficient point. That is, if one could have chosen the best technology but kept the input proportions of D, one would have chosen the technology represented by C. Comparing the amount of input requirements at C and D can obviously form a basis for efficiency comparisons. Farrell (1957) defines the *technical efficiency* of D as:

$$TE = \frac{OC}{OD} \quad (3.1)$$

TE measures the proportion of (x_1^0, x_2^0) actually necessary to produce y^0 . It is easy to see that Farrell's definition of technical efficiency requires all inputs of an inefficient firm to be

reduced by the same proportion¹. TE equals 1 if the firm is on the frontier $ACA'A''$. As the observed performance of D worsens, the distance between C and D increases so that the technical efficiency ratio falls towards zero. Likewise, as performance improves, the efficiency ratio rises in value to unity. Thus in general:

$$0 \leq TE \leq 1 \quad (3.2)$$

As mentioned above, firm C is technically efficient, but it costs too much because the unit of output could be produced at the cost of OB by substituting x_1 for x_2 . Therefore, firm C has an allocative inefficiency (price inefficiency) caused by non-optimal input proportions. Farrell (1957) defines the *allocative efficiency* (price efficiency) of C as:

$$PE = \frac{OB}{OC} \quad (3.3)$$

As C moves closer to A' , PE rises towards one. That is, PE also lies between zero and unity. Since D has the same input proportion as C, it has an allocative inefficiency of the same amount $PE=OB/OC$. It is noted that measuring allocative efficiency requires information on input prices which is not required for measuring technical efficiency.

Combining the technical and allocative efficiency measures gives an overall measure of the efficiency of D. Following Farrell (1957), the *productive efficiency* (or economic efficiency) of D is defined as:

¹ This could cause mis-identifying an inefficient firm as efficient when the isoquant is not everywhere downward sloping. An alternative efficiency measure, the "Russell measure", was proposed to remedy this problem (see Fare, Grosskopf and Lovell, 1985 for more discussions). However this "Russell Measure" is not very easy to implement.

$$EE = TE * PE = \frac{OB}{OD} \quad (3.4)$$

It should be noted that the above analysis does not consider the optimality of the level of production, since the optimal scale of production is indeterminate in the case of constant returns to scale. However, if the technology has nonconstant returns to scale, then the scale of production will be optimal if and only if at the chosen level of output, price is equal to marginal cost. A firm is said to be *scale efficient* if it chooses a profit maximizing level of production. Extensions to Farrell's original approach to accommodate non-constant returns to scale technologies appear to be cumbersome and without much success (Farrell and Fieldhouse, 1962, and Seitz, 1971).

In empirical studies, one often relies only on information on output and input quantities, thus cannot measure allocative inefficiency. In those cases, both firms A' and C are considered as efficient, while firm E is considered as inefficient even though it is allocatively efficient and achieves the same degree of productive (overall) efficiency as firm C . This dissertation focuses on the technical efficiency as defined by Farrell (1957) although the allocative efficiency is a very important aspect of efficiency analysis. The reason for doing this is that (1) technical efficiency is always desired in any circumstances, and understanding of the structure of technical efficiency may serve as a starting point for further study of a production process including allocative efficiency²; and (2) in many sectors of the economy such as the public service sector, information on prices is either unavailable or

² Allocative efficiency has been investigated in numerous studies, including Schmidt and Lovell (1979), Kopp and Diewert (1982), Kumbhakar (1987, 1989), Kalirajan (1990).

unreliable, thus studies of allocative efficiency are difficult if not impossible, while enhancing technical efficiency appears to be the main avenue to improve production performance.

3.2 Efficiency and Production Function

The technical efficiency (TE) can be defined directly in terms of the production function³. When inefficiency is present, the production function may be written as an inequality:

$$y_j \leq f(X_j) \quad (3.5)$$

where y_j is observed output level of firm j , X_j is a vector of inputs of firm j , and $f(\cdot)$ is the production function and has the interpretation of a frontier, or y_{\max} . At inefficient operations, potential output (y_{\max}) will exceed observed performance y_j . Hence, technical inefficiency implies $\epsilon_j = y_j - y_{\max}$, a residual in the production function, is negative. To preserve the frontier interpretation of $f(\cdot)$, the ϵ_j are always non-positive and truncated at zero such that deviations are only possible below the production frontier. This ensures that observed output cannot exceed potential and that the distribution of the residuals is one-sided.

³ Cost functions can also be used to estimate technical efficiency as well as allocative efficiency. The estimation of cost functions requires data on input prices which may not always be available. For reasons mentioned in section 3.1, this dissertation is focused on technical efficiency estimated in relation to production frontier. However, the basic concept is equally applicable to cost functions.

The addition of the efficiency residuals "balances" the production function⁴:

$$y_j = f(x_j) - \epsilon_j \quad \epsilon_j \leq 0 \text{ for all } j \quad (3.6)$$

This could be illustrated by Figure 3.3 where $f(x)$ represents the production frontier. Firm j produces output y_j using input x_j , denoted by point D, which is less than the frontier output y_{\max} for input x_j . The difference between actual and optimal output, ϵ_j , is negative and hence production at firm j is relatively inefficient. The degree of technical efficiency at firm j may be measured by the ratio of observed to optimal output defined by the production frontier, and written as:

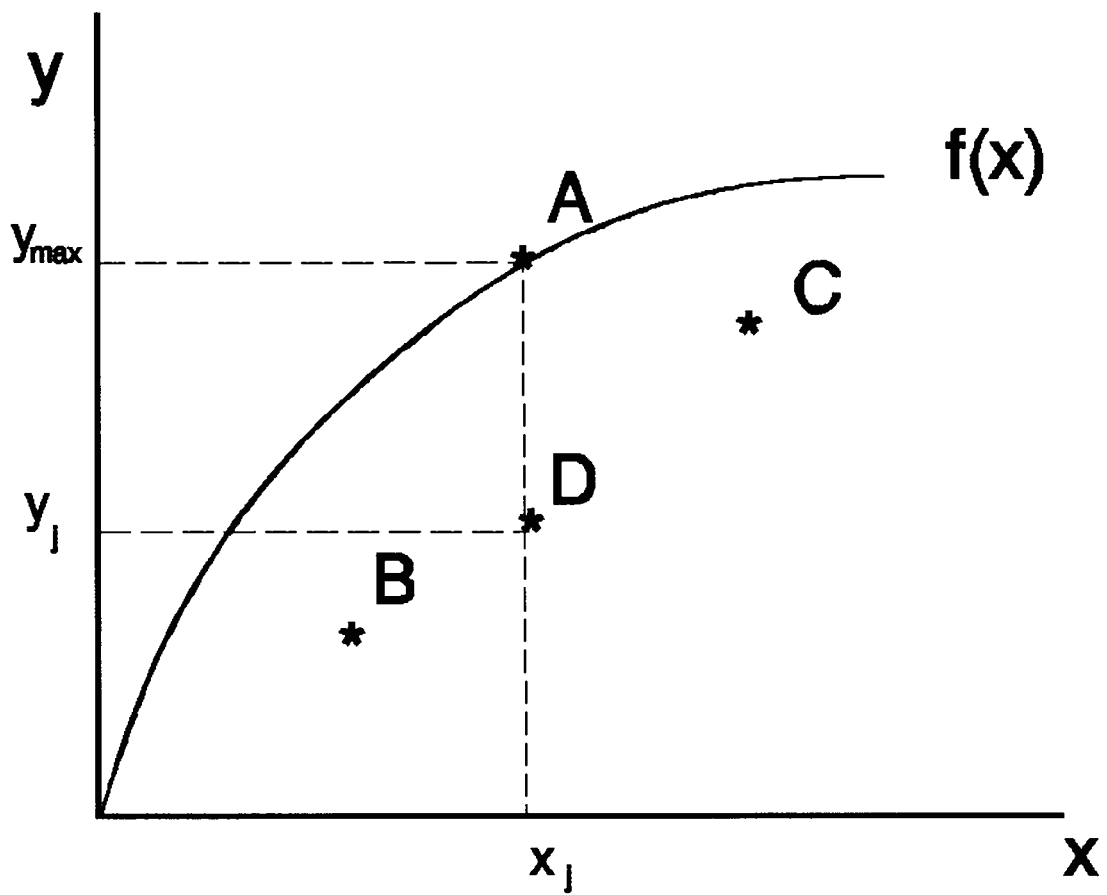
$$TE_j = \frac{y_j}{f(x_j)} \quad (3.7)$$

Notice that efficient production implies that observed and frontier attainments coincide and that the residual equals zero thus TE equals to unity. Equation (3.7) is the econometric version of the Farrell's measure of technical efficiency. The most notable difference between the Farrell efficient isoquant and the frontier functions is the assumption of a specific functional form⁵.

Assume that all the firms have access to the same technology, the deviations of points, B, C, and D in Figure 3.3 from the efficient frontier could be interpreted in two ways. The first interpretation is that some firms are more successful in utilizing the available

⁴ For now it is assumed that there are no statistical noises in the observed production performance.

⁵ Kopp (1981) discusses how Farrell's concept of efficiency measurement related to the frontier functions.

**Figure 3.3 Production Frontier and Inefficiency**

technology than others, that is, all deviation from the frontier is attributed to inefficiency. Equation (3.7) is based on this interpretation, and it corresponds to the deterministic frontier method to be discussed in Chapter 4. The second interpretation is that all firms face the same technology up to a random factor that takes into account the effects on production of measurement errors in the output and input variables and other random shocks outside the firm's control. Thus, the resulting production frontier is stochastic and the departure from this frontier reflects technical inefficiency. This is the rationale for the stochastic frontier method described in the following Chapter.

Chapter 4

Alternative Methods for Efficiency Measurement

As stated in Chapter 1, there are two aspects in efficiency measurement: estimation of the production frontier and measurement of efficiency relative to the frontier. In discussing the basic concept of efficiency in Chapter 3, it is implicitly assumed that the production frontier is known. In practice, however, one has only data - a set of observations corresponding to achieved output levels and input consumption. Thus, the initial problem is to construct an empirical production function or frontier based on the observed data. Then, the relative efficiency of the firms or organizations in question can be measured in relation to this empirical frontier. The purpose of this chapter, therefore, is to describe a number of alternative methods used for the estimation of frontiers and the subsequent measurement of efficiency.

There are many ways of estimating a production frontier. The various methods can be classified according to the assumptions made on the form of the production frontier. A distinction is often made between the parametric and non-parametric methods, depending on whether or not the frontier function can be specified by a particular functional form with constant parameters, such as the Cobb-Douglas function, or the translog function and *etc.*

4.1 Parametric Approaches

Aigner and Chu (1968) made the first attempt to impose a parametric functional form on production frontier within the framework provided by Farrell. They specify a

homogeneous Cobb-Douglas production frontier. In estimating this frontier production function, they force all observations to be on or beneath the frontier through the introduction of an one-sided error term. Aigner and Chu's work introduces explicitly the concept of "frontier" production function into the production theory. Although the frontier production function seems clearly to be in accord with the theoretical definition of a production function, standard statistical techniques had been used to estimate "average" production functions prior to Aigner and Chu's work.

To estimate the parametric production frontiers, Aigner and Chu (1968) suggests minimizing the sum of absolute residuals from the logarithm of the production function (a linear programming problem), or alternatively minimizing the sum of squared residuals (a quadratic programming problem), while constraining all residuals to be non-negative. If we write the production function for the j -th firm as $y_j = f(x_j, \beta) - \epsilon_j$, here $f(x_j, \beta)$ is the maximum output obtainable from x_j , a vector of (non-stochastic) inputs, y_j is the observed output, ϵ_j is the one-sided (non-negative) error term, and β is an unknown parameter vector to be estimated. Then Aigner and Chu's method can be expressed, in mathematical terms, as following:

$$\min \sum_{i=1}^n |y_i - f(x_i, \beta)|, \quad s.t. \quad y_i \leq f(x_i, \beta), \quad i=1, \dots, n \quad (4.1.1)$$

or

$$\min \sum_{i=1}^n [y_i - f(x_i, \beta)]^2 \quad \text{s.t.} \quad y_i \leq f(x_i, \beta), \quad i=1, \dots, n \quad (4.1.2)$$

If $f(x_i, \beta)$ is linear in β , then equation (4.1.1) is a linear programming problem and equation (4.1.2) is a quadratic programming problem. This approach assumes all deviations from the production frontier to be the result of technical inefficiency and thereby constrains the residuals to be of one sign. Their measure of technical efficiency can be specified as $y_i / f(x_i, \beta)$. With such mathematical programming techniques the choice of the functional form for the production function has a strong influence on the conclusions reached about the degree of technical inefficiency. This is clear from the consideration that the number of observed units that appear to be fully efficient technically is generally only as large as the number of parameters of the production function to be estimated. Another problem with this approach is that the estimated frontier is supported by a subset of the data and thus is sensitive to outliers and data errors. This leads to the development of so-called "probabilistic" frontiers by Timmer (1971) which proposes a modified linear programming problem allowing a certain percentage (arbitrarily determined) of observations to lie above the frontier. The third problem with this mathematical programming approach is that the estimators lack identifiable statistical properties. This third problem could be solved by making further distributional assumptions about the one-sided errors and estimating the production frontier using statistical methods.

Afriat (1972) was the first to impose a statistical assumption on the one-sided errors,

and to introduce the concept of "distribution of technical inefficiency", which makes statistical estimation of frontier functions possible. Since Afriat's work most of the studies of the parametric frontiers have been focused on the statistical approach. In fact, Schmidt (1976) shows that if ϵ_j is exponential, then Aigner and Chu's linear programming procedure is maximum likelihood, while their quadratic programming procedure is maximum likelihood if ϵ_j is half-normal. Therefore, this dissertation considers only the parametric frontier models using econometric/statistical techniques. For more discussions on the various parametric frontier models, readers are referred to the survey papers by Forsund, Lovell and Schmidt (1980), Lovell and Schmidt (1988), Bauer (1990), and Greene (1993).

Within the parametric approach, a distinction is made between two different methods: the deterministic frontier method and the stochastic frontier method. The main difference between the two methods is that the deterministic method attributes all deviations from the frontier to inefficiency while the stochastic method distinguishes the deviations into a random component capturing statistical noise and an inefficiency component. These two methods are discussed in the following subsections.

4.1.1 Deterministic Frontier Methods

The deterministic frontier methods assume that the discrepancies between the estimated frontier function and the observed production situations exclusively capture technical inefficiencies, that is, the discrepancies are one-sided. In estimating the deterministic frontiers, all observations are forced to be on or below the frontier so that all deviation from the frontier is attributed to inefficiency. A basic model of the deterministic

frontier is specified as:

$$y = f(x)e^{-u} \quad (4.1.3)$$

where y is observed output, $f(x)$ is the production frontier, and $u \geq 0$ thus $0 \leq e^{-u} \leq 1$ which captures the degree of inefficiency. This frontier function may be estimated by either the maximum likelihood method (MLE) or the "corrected" ordinary least squares method (COLS).

The maximum likelihood method was first suggested by Afriat (1972) and first used by Greene (1980a) and Stevenson (1980). MLE requires the specification of a particular distribution for the one-sided residual u , and it is then implemented by estimating all of the production parameters and the parameters of the distribution of u . The MLE frontier envelops all observations, and the MLE residuals can then be used to provide a measure of estimated efficiency:

$$e_j = \frac{y_j}{f(x)} = e^{-u_j} \quad (4.1.4)$$

It should be noted that the choice of a distribution for u is important since the maximum likelihood estimates depend on it in a fundamental way - different distributional assumptions lead to different estimates. There do not appear to be good *a priori* arguments for any particular distribution. Afriat (1972) assumes a two-parameter beta distribution for e^{-u} (a

gamma distribution for u). Greene (1980a) shows that a gamma distribution for u satisfies the conditions for the maximum likelihood estimators to be consistent and asymptotically efficient. Other distributions have also been assumed in the literature, such as the half-normal and exponential distributions (Aigner, Lovell, and Schmidt, 1977), truncated normal distribution (Stevenson, 1980), and the two parameter gamma distribution by Greene (1990).

Another way to estimate the parameters of the frontier function is by the corrected ordinary least squares method (COLS). COLS was first proposed by Richmond (1974) and improved by Greene (1980a,b). COLS first estimates the technology parameters of equation (4.1.3) by OLS, which gives unbiased and consistent estimates of the parameters except for the constant term in log-linear form. The OLS intercept can be corrected by either of two ways. The first way¹ to "correct" the OLS intercept requires the assumption of a specific functional form for the one-sided error term u . The OLS intercept is corrected by shifting it up by adding the sample mean of u which can be estimated consistently from the second or higher moments of the OLS residuals. The OLS residuals are consequently modified in the opposite direction. There is no guarantee that this technique shifts the estimated intercept up far enough to cover all the observations, and if an observation has a sufficiently large positive OLS residual, it is possible that $e_j > 1$.

A second way to "correct" the OLS intercept is simply shifting it upward until all corrected residuals are nonpositive and at least one is zero. With this method there is no need to assume a specific distribution function for the one-sided error term, one needs only

¹ This technique is termed as "modified" ordinary least squares by Lovell (1993).

assume that u is independently and identically distributed from an unspecified one-sided distribution. This dissertation adopts this method in the estimation of the deterministic frontier models because it is the easiest one to estimate and less restrictive.

The basic procedure of the second variant of COLS is as follows: first, one estimates an average production function through OLS estimation, then the constant term is "corrected" by shifting it up until no residual is positive and at least one is zero. This corrected production function is considered as the production frontier, and used as reference point in efficiency measurement. This method provides a consistent estimate of all the parameters of the frontier function. The efficiency is measured by the ratio of the actual output value Y_j divided by the "optimal" output value Y_j^F obtained from the estimated production frontier, given a certain level of input. The firm (observation) which has the highest positive residual from the OLS regression is by definition 100% efficient.

If we take the logarithm of both sides of equation (4.1.3), then an OLS estimation gives the following equation:

$$\ln y_j = \hat{\alpha} \ln f(x_j) + \varepsilon_j \quad (4.1.5)$$

where α is a vector of regression coefficients. Letting $E = \text{Max}(\varepsilon_j)$, one obtains the following frontier function:

$$\ln y_j^F = \hat{\alpha} \ln f(x_j) + E \quad (4.1.6)$$

which allows for measuring the efficiency of the firms (observations) as:

$$e_j = y_j / y_j^F = e^{\varepsilon_j - E} \leq 1 \quad (4.1.7)$$

The deterministic frontier method has two major drawbacks. The first one is that this method is entirely deterministic with no allowance for noise, measurement error and the like, thus the estimated frontier depends on a small number of observations that might be inaccurately measured or otherwise abnormal. That is, it tends to be sensitive to outliers and data errors. The second drawback is its inability to deal easily with multiple outputs unless the dual cost frontier can be estimated directly².

The deterministic frontier methods assume that all firms share a common production frontier, and all variations in firm performance are attributed to variations in firms efficiencies relative to this common frontier. This assumption ignores the very real possibility that a firm's performance may be affected by factors entirely outside its control, such as bad weather, labour dispute, and so on, as well as by factors under its control (inefficiency). To distinguish the effects of "noise", including exogenous random shocks, measurement errors, misspecification of production functions, and so on, from inefficiency, Aigner, Lovell, and Schmidt (1977) and Meeusen and van den Broeck (1977) introduce the stochastic frontier model which is discussed in the following section.

² All deviation from the cost frontier is attributed to cost inefficiency. Kopp and Diewert (1982) and Zieschang (1983) show how to decompose cost efficiency into its technical and allocative components.

4.1.2 Stochastic Frontier Methods

The basic idea behind the stochastic frontier method is that the deviation from the frontier is composed of two parts. A symmetric component permits random variation of the frontier across firms and captures the effects of measurement error, statistical "noise", and exogenous random shocks outside the firm's control. An one-sided component captures the effects of inefficiency. This method was first proposed by Aigner, Lovell and Schmidt (1977) and Meeusen and van den Broeck (1977), and has been extended by Jondrow, Lovell, Materov, and Schmidt (1982), and Battese and Coelli (1991) among others. Specific distributional assumptions about the disturbance terms must be made in order to obtain estimates of individual firm efficiencies³. The statistical noise is generally assumed to be identically independently distributed (*iid*) normal, while a number of distributions have been assumed for the one-sided (inefficiency) term. Aigner, Lovell and Schmidt (1977) proposes half-normal and exponential distributions which have been widely used. Stevenson (1980) proposes a truncated normal distribution, while Greene (1990) proposes a two-parameter Gamma distribution⁴. Further, other distributions could be constructed following Greene's (1990) methodology.

The basic stochastic frontier model is given by:

³ When panel data are available, estimates of the inefficiency disturbances can be obtained without assuming a particular distribution for these terms (Schmidt and Sickles, 1984). However, one must specify how efficiency changes over time instead. Most of these studies assume that inefficiency is time-invariant.

⁴ There are a number of survey papers which cover alternative frontier procedures, such as Greene (1993), Bauer (1990), Lovell and Schmidt (1988), Schmidt (1986) and Forsund, Lovell and Schmidt (1980).

$$y = f(x, \beta) e^v e^{-u}, \quad u \geq 0 \quad (4.1.8)$$

where y represents output, $f(x, \beta)$ is the deterministic core of the frontier production function, β are the parameters to be estimated, v is a random variable that takes values over the range $(-\infty, +\infty)$ and represents the effects of measurement errors, non-observable explanatory variables and random shocks, and u is a random variable that takes nonnegative values which captures inefficiency. In other words, $f(x, \beta)e^v$ is the stochastic frontier while e^{-u} is the inefficiency. The condition $u \geq 0$ ensures that all observations lie on or below the stochastic production frontier.

The stochastic frontier method allows room for errors of observations, thus avoids the deterministic techniques' high sensitivity to errors in the data. However, the problem remains of what distribution to assign to the residual components v and u ⁵.

Direct estimates of the stochastic frontier functions may be obtained by either maximum likelihood (MLE) or corrected least square (COLS). Olson, Schmidt and Waldman (1980) presents Monte Carlo evidence which indicates that COLS generally performs as well as MLE for the normal/half-normal case. Whether the model is estimated by MLE or COLS, the distribution of u must be specified.

⁵ The central limit theorem can be invoked to warrant assuming a normal distribution for v . However, no such consideration stands out to support any particular assumption about the one-sided residuals u , and thus statistical convenience and general plausibility continue to rule the roost. Half normal and exponential distributions have often been used. There are limited evidence which suggests that these two assumptions give rise to quite similar parameters for the estimated production frontier (Aigner, Lovell, and Schmidt, 1977). Corbo and de Melo (1986) indicates that the estimated inefficiency levels based on the two assumptions are highly correlated.

The COLS is similar to the deterministic COLS as described earlier. We first estimate the parameters, β , in $f(x, \beta)$ through OLS. We can then estimate the variance σ_u^2 and σ_v^2 by:

$$\hat{\sigma}_u^2 = \left[\sqrt{\frac{\pi}{2}} \left(\frac{\pi}{\pi-4} \right) \hat{\mu}_3' \right]^{\frac{2}{3}} \quad (4.1.9)$$

$$\hat{\sigma}_v^2 = \hat{\mu}_2' - \frac{\pi-2}{\pi} \hat{\sigma}_u^2 \quad (4.1.10)$$

where $\hat{\mu}_2'$ and $\hat{\mu}_3'$ are the second and third moments of the OLS residuals. Next, we "correct" the intercept by adding to the OLS intercept the negative of the estimated bias which is the sample mean of $\varepsilon = v - u$, $\mu = -\sqrt{2/\pi} \hat{\sigma}_u$. Note that the COLS estimator of all elements of β except the intercept is the same as the OLS estimator. These estimates are consistent and unbiased but not asymptotically efficient.

The COLS is simpler in terms of computation. However, there is one problem with COLS, that is, the estimates may not exist in some samples. If the third moment of the OLS residuals turns out to be positive, then $\hat{\sigma}_v^2 < 0$, the estimation procedure breaks down. The probability of this occurrence depends on the value of μ_3' , the third moment of the disturbance. When μ_3' is near zero, the probability of $\hat{\mu}_3'$ being positive may be substantial.

This is a problem mainly when $\lambda = \sigma_u / \sigma_v$ is small. According to Olson, Schmidt, and Waldman (1980), as $\lambda \rightarrow 0$ ($\sigma_u^2 \rightarrow 0$) the probability of $\hat{\mu}_3'$ being positive approaches (approximately) 1/2. There is another possibility when the estimation procedure may break down. This occurs when $\hat{\sigma}_e^2 < ((\pi-2)/\pi)\hat{\sigma}_u^2$, which implies $\hat{\sigma}_v^2 < 0$. This may occur with non-negligible probability when λ is large. Banker, Gadh, and Gorr (1993) finds that the two failure types are complementary, there is almost always one present. In their Monte Carlo experiments, only two out of 640 estimates have no failure. Because of these problems, COLS procedure is not used in this study.

The maximum likelihood estimator is obtained by the (numerical) maximization of the log likelihood function of equation (4.1.8) with respect to the parameters in $f(\cdot)$ and the distribution functions of v and u , after specific assumptions being made about the distributions of the two error terms.

Assume that the elements of v are identically independently distributed (*iid*) as $N(0, \sigma_v^2)$, the elements of u are half-normal taking absolute values from variables which are *iid* as $N(0, \sigma_u^2)$, and v and u are independent of each other and are also independent of x . Let $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $\lambda = \sigma_u / \sigma_v$. The density function of u is given by:

$$f(u) = \frac{\sqrt{2}}{\sqrt{\pi} \sigma_u} \exp\left[-\frac{u^2}{2\sigma_u^2}\right] \quad u \geq 0 \quad (4.1.11)$$

Suppose that after taking logarithms, equation (4.1.8) becomes linear in parameters and can

be specified as⁶:

$$Y = X\beta + \varepsilon \quad (4.1.12)$$

where $\varepsilon = v - u$. Aigner, Lovell and Schmidt (1977) shows that the log likelihood function of equation (4.1.12) is:

$$L = \frac{N}{2} \ln(2/\pi) - N \ln \sigma + \sum_{j=1}^N \ln[1 - F(\varepsilon_j \lambda \sigma^{-1})] - \frac{1}{2\sigma^2} \sum_{j=1}^N \varepsilon_j^2 \quad (4.1.13)$$

where N is the total number of observations, j indexes the observations, and F is the cumulative distribution function of the standard normal distribution. Details may be found in Aigner, Lovell, and Schmidt (1977). Note that the parameter λ embodies the model of inefficiency. The simple OLS regression model results from $\lambda = 0$. The implication would be that every firm operates on its frontier. The MLE estimates are obtained by the maximization of equation (4.1.13) with respect to the parameters (β, λ, σ) . The MLE estimates are consistent and asymptotically efficient.

Several alternative distributional assumptions have been suggested in the literature. The log-likelihood function and associated results under some of these alternative distributional assumptions, for example, may be found in Aigner, Lovell, and Schmidt (1977) for exponentially distributed inefficiency, Stevenson (1980) for truncated normal distribution, and Greene (1980), Greene (1990) for gamma distributions among others.

⁶ In nearly all applications, after transformation, the functional form of the model to be estimated is linear in the logs of output and a set of independent variables, so equation (4.12) is assumed as the general form of the production frontier.

Since half-normal distribution is the most popular function among empirical works, this study considers the normal/half normal situation.

As mentioned at the beginning of this chapter, efficiency measurement involves two aspects: estimation of the "empirical" production frontier and the measurement of efficiency relative to the estimated frontier. The COLS and MLE procedures described in the previous paragraphs accomplish the first aspect, that is, the estimation of the production frontier from the observed data. The next step is to measure the efficiency level of each observation in relation to this estimated production frontier.

Since the residuals from equation (4.1.12) or equation (4.1.8) estimate $\epsilon_j = v_j - u_j$, not u_j , the efficiency component of the composed error terms has to be estimated indirectly. Jondrow, Lovell, Materov and Schmidt (1982) suggests the use of either the mean or the mode of the efficiency term, u , conditional on the estimate of ϵ as a measure of observation specific estimates of efficiency. They derive an explicit form for the half-normal model⁷as:

$$E(u_j/\epsilon_j) = \frac{\hat{\sigma}_u^2 \hat{\sigma}_v^2}{\hat{\sigma}^2} \left[\frac{f(\hat{\epsilon}_j \hat{\lambda} / \hat{\sigma})}{1 - F(\hat{\epsilon}_j \hat{\sigma})} - \hat{\epsilon}_j \frac{\hat{\lambda}}{\hat{\sigma}} \right] \quad (4.1.14)$$

where $f(\cdot)$ is the standard normal density function.

In Chapter 3 the efficiency is defined as the ratio of observed output over frontier output for a given level of input consumption. Therefore, $\exp(-u_j)$ should be considered as the efficiency measure instead of u_j . Again, in practice one needs to estimate the $\exp(-u_j)$

⁷ Jondrow, Lovell, Materov and Schmidt (1982) also derives the expression for the exponential case. Greene (1993) presents the expressions for gamma models and truncated normal model as well.

given the residuals from equation (4.1.12). Battese and Coelli (1988) suggests the use of conditional expectations of $\exp(-u_i)$, given $\epsilon_j = e_j$, as the firm specific efficiency estimate, where e_j is the residual from equation (4.1.12). With the half normal assumption, this efficiency measure can be derived as follows.

The density function of u_j is given by equation (4.1.11). As $\epsilon = v - u$, the joint density of u and ϵ is given by:

$$f(u, \epsilon) = \frac{1}{\pi \sigma_u \sigma_v} \exp \left[-\frac{1}{2\sigma_u^2} u^2 - \frac{1}{2\sigma_v^2} (u^2 + \epsilon^2 + 2u\epsilon) \right] \quad (4.1.15)$$

and the density function of ϵ is given by:

$$f(\epsilon) = \frac{2}{\sqrt{2\pi} \sigma} (1 - F(\epsilon \lambda / \sigma)) \exp \left[-\frac{1}{2\sigma^2} \epsilon^2 \right] \quad (4.1.16)$$

Therefore, the conditional density of u given ϵ is the ratio of equations (4.1.15) to (4.1.16), which can be written as:

$$f(u|\epsilon) = \frac{1}{\sqrt{2\pi} \sigma_*} \frac{1}{1 - F(\epsilon \lambda / \sigma)} \exp \left[\frac{-1}{2\sigma_*^2} u^2 - \frac{1}{\sigma_v^2} u\epsilon - \frac{\lambda^2}{2\sigma^2} \epsilon^2 \right] \quad u \geq 0 \quad (4.1.17)$$

where $\sigma_*^2 = \sigma_u^2 \sigma_v^2 / \sigma^2$. Note that equation (4.1.17) is the density of a $N(\mu_*, \sigma_*^2)$ variable truncated at zero, where $\mu_* = -\sigma_u^2 \epsilon / \sigma^2$. Then the conditional expectation of $\exp(-u)$, given $\epsilon = e$, is defined by:

$$\begin{aligned}
E(\exp(-u)/\varepsilon) &= \int_0^{\infty} \exp(-u) f(u/\varepsilon) du \\
&= \frac{1 - F(\sigma_* - \varepsilon/\sigma_v^2)}{1 - F(-\varepsilon/\sigma_v^2)} \exp \left[\mu_* + \frac{\sigma_*^2}{2} \right]
\end{aligned} \tag{4.1.18}$$

The operational estimator of equation (4.1.18) may be obtained by substituting the relevant parameters by their maximum likelihood estimators (or COLS estimators) as in equation (4.1.14).

There are some problems with both COLS and MLE estimators as described above. The first problem is associated with the distributional assumptions of the error terms. Different distributional assumptions may lead to different efficiency measures. The second problem is that inefficiency and input levels are assumed to be independent which may not always be the case. The third problem is that the estimator used to compute the firm specific inefficiency is not consistent although it is unbiased.

To avoid assuming a particular distribution for the error terms, Schmidt and Sickles (1984) proposes a number of methods of estimating individual firm efficiency level given that panel data is available on sample firms. However, inefficiency is assumed to be time-invariant in their panel data model while statistical noise is assumed to vary over firms and time. In particular, they consider three estimators: (1) A "fixed effects" model, or the "within" model, which assumes that the inefficiency error term u is firm specific constant. The frontier function is estimated by using dummy variables in OLS. This model does not assume the independence between inefficiency and inputs, and it does not require the

normality assumption. Further, the estimated efficiency is consistent over the time period. (2) A generalized least squares (GLS) estimator, which makes no distributional assumption on the inefficiency term, but assumes that the efficiency and inputs are independent. This model allows for the inclusion of time-invariant firm specific attributes, such as the capital stock of a firm which is not growing, location, or some other characteristics. (3) A MLE estimator, which makes both distributional and independent assumptions. Gong and Sickles (1989) compares the three estimators in a Monte Carlo study, and finds that efficiency estimates from the three estimators are very similar.

The time-invariant assumption may be plausible for short panels⁸. However, it would be desirable to examine how the performance of a particular firm changes over time. Thus the time-invariant assumption would appear to be restrictive. To allow efficiency to vary over time, Cornwell, Schmidt and Sickles (1990) extends the models by Schmidt and Sickles (1984) to relax the time-invariant assumption by imposing certain structure on how inefficiency varies over time. Their model assumes that the one-sided firm effects are a quadratic function of time in which the coefficients vary over firms according to the specification of a multivariate distribution. For large N , this model presents a fairly cumbersome problem of estimation since the number of additional variables to capture the time-varying firm specific effects would be substantial.

Battese and Coelli (1992) proposes a time-varying model for unbalanced panel data in which the efficiency is assumed to be distributed as truncated normal. This generalizes

⁸ By short panel, we mean that production performance is observed over a short period of time.

a number of previous stochastic frontier models such as Aigner, Lovell and Schmidt (1977), Stevenson (1980), Pitt and Lee (1981), among others. This model assumes that the efficiency is an exponential function of time. This assumption appears to be restrictive.

From the foregoing discussions, we can see that there is a trade-off between imposing a specific distributional assumption on the error terms and imposing an explicit functional assumption on how the efficiency of a particular firm varies over time, while both bringing the potential effects of misspecification. The choice of a specific model in any empirical study will depend on the particular situation in question and the availability of necessary data.

For the Monte Carlo study in this dissertation, we consider the basic model proposed by Aigner, Schmidt and Lovell (1977) with the assumptions of normally distributed statistical noise and an efficiency term following a half-normal distribution. This model is chosen, because it is the most popular one used in empirical applications. The firm specific (or observation specific) efficiency is measured by the conditional expectation of $\exp(u_i)$, given the residuals from the estimated frontier function, as defined by equation (4.1.18).

4.2 Nonparametric Approach - The Data Envelopment Analysis Method

The nonparametric approach, represented by the Data Envelopment Analysis method, uses mathematical programming techniques to construct production frontiers. The Data Envelopment Analysis (DEA) method, introduced in Charnes, Cooper and Rhodes (1978), involves an application of linear programming to observed data to locate frontiers which can then be used to evaluate the efficiency of each of the firms or organizations responsible for

the observed output and input quantities. In DEA, the entities responsible for converting inputs into outputs are referred to as decision making units (DMUs) which may represent any kind of firms and organizations or their subdivisions as long as they perform the same or similar tasks. DEA utilizes a sequence of linear programs, one for each DMU, to construct a piecewise linear production frontier, and to compute an efficiency index relative to the frontier. Units that lie on the production frontier are deemed efficient. Units that do not lie on the production frontier are termed inefficient and the analysis provides a measure of their relative efficiency.

DEA is non-parametric in the sense that it does not assume that the underlying technology has a specific functional form with a finite number of parameters, such as Cobb-Douglas functional form. It only requires the convexity assumption about the production frontier. Note that DEA is also "non-statistical" because it makes no explicit assumption on the probability distribution of the "errors" in the production function.

DEA can accommodate multiple outputs and multiple inputs with each being stated in different units of measurement. The relative efficiency of a DMU is defined as the ratio of its total weighted output (virtual output) to its total weighted input (virtual input). The weights (virtual multiplier) required to incorporate the multiple outputs and multiple inputs are determined by linear programming optimization. DEA assumes that each DMU will select the weights that maximize its own efficiency score, that is, it will evaluate each input and each output in such a way as to maximize the ratio of its own weighted output to its own weighted input. Because different DMUs use different combinations of inputs to produce different combinations of outputs, they are expected to select sets of weights that reflect this

variety. Generally, DMUs will place higher weights on the inputs that they use least and on the outputs they produce most. In this sense DEA shows each DMU in its best possible light. The efficiency ratio generated by DEA is consistent with a frontier interpretation of performance. A ratio of unity implies that observed and potential performance coincide in which case the corresponding DMU is said to be efficient or "best practice". Where observed performance is lower than potential a DMU receives an efficiency ratio of less than unity which implies that its performance is poorer than that of a combination of some of its peer DMUs and so it is relatively inefficient.

Extensions to the original Charnes, Cooper and Rhodes (CCR) model have resulted in a variety of alternative formulations all sharing the principle of envelopment. In this study, we consider two of these models⁹, the CCR ratio model (Charnes, Cooper and Rhodes, 1978) and the BCC model (Banker, Charnes and Cooper, 1984). These two formulations have been widely used in empirical applications. The CCR model yields an index of overall efficiency, it points out the observable differences in performance among DMUs¹⁰. The BCC model takes into account the effect of "returns to scale" within the analyzed group of DMUs. It distinguishes between technical and scale inefficiencies by estimating pure technical efficiency at a given scale of operation.

⁹ For more discussions on the various DEA models, please refer to Charnes and Cooper (1985), Banker, Charnes, Cooper, Swarts and Thomas (1989), Seiford and Thrall (1990), and Ali and Seiford (1993).

¹⁰ Some of the subsequent models incorporate some of the explanations to the observed efficiency differences into models themselves.

4.2.1 The CCR ratio

Assuming convexity of production possibility sets, Charnes, Cooper and Rhodes (1978) defines the DEA efficiency measure as the maximum of a ratio of weighted outputs to weighted inputs subject to the condition that similar ratios for every DMU be less than or equal to unity. In mathematical form, the efficiency of the k -th DMU is

$$\begin{aligned}
 h_k^* = \text{Max} \quad & \frac{\sum_{r=1}^s u_r Y_{rk}}{\sum_{i=1}^m v_i X_{ik}} \\
 \text{st.} \quad & \frac{\sum_{r=1}^s u_r Y_{rj}}{\sum_{i=1}^m v_i X_{ij}} \leq 1 \quad j=1,2,\dots,n \\
 & u_r, v_i \geq \epsilon \quad r=1,\dots,s; i=1,\dots,m
 \end{aligned} \tag{4.2.1}$$

where the Y_{rj} , X_{ij} are the known outputs and inputs of the j -th DMU, the u_r , v_i are the weights (virtual multipliers) to be determined by the solution of the problem, and ϵ represents a small positive non-Archimedean quantity introduced to ensure that all of the observed inputs and outputs will have "some" positive value assigned to them. This value serves as a lower limit for the values that can be assigned to the weights u_r and v_i as shown by the final constraints in equation (4.2.1). In this model, $h_k^* = 1$ if and only if the k -th DMU is efficient relative to the other DMUs.

Note that the above model is a fractional programming problem. This fractional program is not used for actual computation of the efficiency scores because it has intractable

non-linear and non-convex properties (Charnes, Cooper and Rhodes, 1978). However, it can be converted into a linear programming problem by using a simple transformation (Charnes and Cooper, 1962, 1973). Since $\sum_{i=1}^m v_i X_{ij} > 0$, and the problem is invariant under positive scalar multiplier, we may let $\sum_{i=1}^m v_i X_{ik} = 1$, the problem can then be replaced by the following linear programming problem:

$$\begin{aligned}
 h'_k &= \text{Max} \sum_{r=1}^s u_r Y_{rk} \\
 \text{st.} \quad &\sum_{r=1}^s u_r Y_{rj} - \sum_{i=1}^m v_i X_{ij} \leq 0 \quad j=1, \dots, n \\
 &\sum_{i=1}^m v_i X_{ik} = 1 \\
 &u_r, v_i \geq \epsilon \quad r=1, \dots, s; \quad i=1, \dots, m
 \end{aligned} \tag{4.2.2}$$

and hence solutions can be obtained by repeated application of a linear programming software to each of the observations in the sample data.

The dual for formulation (4.2.2) constructs a piecewise linear approximation to the true frontier by minimising the quantities of the m inputs required to meet stated levels of the r outputs. The dual program is formulated as:

$$\begin{aligned}
 h_k &= \text{Min} \quad \theta_k - \epsilon \left(\sum_r \sigma_r + \sum_i s_i \right) \\
 \text{s.t.} \quad &\sum_j y_{rj} \lambda_j - \sigma_r = y_{rk} \\
 &\sum_j x_{ij} \lambda_j - \theta_k x_{ik} + s_i = 0 \\
 &\lambda_j \geq 0, \quad j\text{-DMUs}, j=1, \dots, n \\
 &\sigma_r \geq 0, \quad r\text{-outputs}, \quad r=1, \dots, s \\
 &s_i \geq 0, \quad i\text{-inputs}, \quad i=1, \dots, m
 \end{aligned} \tag{4.2.3}$$

Note that the choices of $\epsilon > 0$ are defined so that the optimal value of θ_k will not be

affected by any value that may be assigned to the slack variables associated with ϵ in the objective function of the dual. The dual problem has one variable λ_j for each DMU in the sample. Variable θ_k is unconstrained. It indicates the potential of a proportional reduction in all the inputs of the k -th DMU. Variables σ_r are output slacks, and variables s_i are input slacks. If the k -th DMU is efficient, then the optimal dual solution will have all λ_j , σ_r and s_i equal to zero except for λ_k , and θ_k^* is equal to 1. When the k -th DMU is inefficient, there is a hypothetical DMU on the empirical production frontier which serves as the reference point for the measurement of the inefficiency for the k -th DMU. The input and output levels of this hypothetical DMU are linear combinations of the input and output levels of the DMUs in the efficient reference set of the k -th DMU. The optimal λ_j from the dual problem are used as the coefficients for these linear combinations. To become efficient, all of the inputs of the inefficient DMU _{k} must be reduced to $\theta_k^* x_{ik}$, where θ_k^* is less than 1. If, in addition, any input slack s_i is not zero for DMU _{k} then a further reduction by the amount of this slack must also be made from the i -th input used by DMU _{k} without altering any other inputs or outputs.

The DEA efficiency measures depend on the number of degrees of freedom that are available. There are $m + s$ constraints to be satisfied in the dual formulation and n observations, one for each of the $j = 1, \dots, n$ DMUs, that form the possible combinations from which efficiency estimation can be secured. From degrees of freedom considerations, the number of variables λ_j used for the solutions in the problem on the left should be at least as great as the number of constraints. Thus, the number of DMUs for which there are

observations should be greater than the number of constraints. According to Banker, Charnes, Cooper, Swarts and Thomas (1989), it is generally advisable to have $n \geq 3(m+s)$.

The name *Data Envelopment Analysis* is obtained from the dual formulation in the following manner. An optimal solution will envelop outputs of DMU_k from above via

constraints of the form $y_{rk} \leq \sum_{j=1}^n y_{rj} \lambda_j^*$ with at least one of these $r = 1, \dots, s$ constraints

satisfied as an equation. Thus, there will be at least one "touching" of an observed output for DMU_k by the solution associated with an optimal choice of λ^* values. Similarly, the

inputs of DMU_k are enveloped from below via the constraints $\theta_k^* x_{ik} \geq \sum_{j=1}^n x_{ij} \lambda_j^*$ with at least

one of these input constraints satisfied as an equation. Thus, the term *Data Envelopment Analysis* is used because the output and input data of DMU_k are enveloped from above and below in the manner just described.

An analogous linear programming formulation of equation (4.2.2) can be obtained by setting the numerator of the objective function of fractional program equal to unity, and minimizing the weighted inputs for DMU_k :

$$\begin{aligned}
h_k'' &= \text{MIN} \sum_{i=1}^m v_i X_{ik} \\
\text{st. } &\sum_{i=1}^m v_i X_{ij} - \sum_{r=1}^s u_r Y_{rj} \geq 0 \quad j=1, \dots, n \\
&\sum_{r=1}^s u_r Y_{rk} = 1 \\
&u_r, v_i \geq \epsilon \quad r=1, \dots, s; \quad i=1, \dots, m
\end{aligned} \tag{4.2.4}$$

This formulation determines the output efficiency of a DMU for a given set of inputs.

The corresponding dual is

$$\begin{aligned}
h_k^* &= \text{MAX} \ z_k + \epsilon (\sum_r \sigma_r + \sum_i s_i) \\
\text{s.t. } &\sum_j X_{ij} \lambda_j + s_i = X_{ik} \\
&Y_{rk} z_k - \sum_j Y_{rj} \lambda_j + \sigma_r = 0 \\
&\lambda_j, s_i, \sigma_r \geq 0
\end{aligned} \tag{4.2.5}$$

where z_k is unconstrained.

In many circumstances outputs are partly determined by the market condition and governmental control. It may therefore not make much sense to suggest that output be raised to increase efficiency. On the other hand, efficient use of inputs is desirable in any case. Therefore in this study we use the input minimization formulation given by equations (4.2.2) and (4.2.3). To have a better understanding of the basic idea underlying DEA, a diagrammatic interpretation of the dual for input minimization is provided.

The estimated dual technology is not smooth but constructed out of a series of intersecting linear facets. Each of these facets represents a constraint in the optimal solution to the dual. Collectively they intersect to form a convex production set which is closed and

bounded from above. The frontier for efficiency comparisons is the lower convex hull of the possibility set which is shown in Figure 4.1. Figure 4.1 illustrates a hypothetical frontier technology based on 5 firms producing a single output, Y , from 2 inputs, X_1 and X_2 . Firms G, F and E, lying on the frontier, are the "best practice". Thus, no other firms or linear combination of firms in this sample can be identified which is producing the same level of output for less of either or both inputs. These firms have unity efficiency scores and zero slacks in the solution to the dual. For example, the solution of the dual for firm F:

$$h_F^* = 1$$

and the constraints are:

$$\text{input 1} \quad X_{1F}h_F^* - 0 = X_{1F}\lambda_F^*$$

$$\text{input 2} \quad X_{2F}h_F^* - 0 = X_{2F}\lambda_F^*$$

$$\text{and on the output} \quad Y_{1F} + 0 = Y_{1F}$$

The left- hand side of the constraints defines the "target", which in this case is equal to actual performance on the right-hand side of the constraints because best-practice implies $\lambda_F^* = 1$. The peer group¹¹ drops out of the RHS of the constraints and for an efficient firm is none other than that firm itself since $\lambda_F^* = 1$ and $\lambda_j^* = 0, j \neq F$.

Firms B and D are inefficient relative to frontier performance. That is, for the same

¹¹ The peers are defined by those firms that have non-zero weights in the optimal solution in the dual.

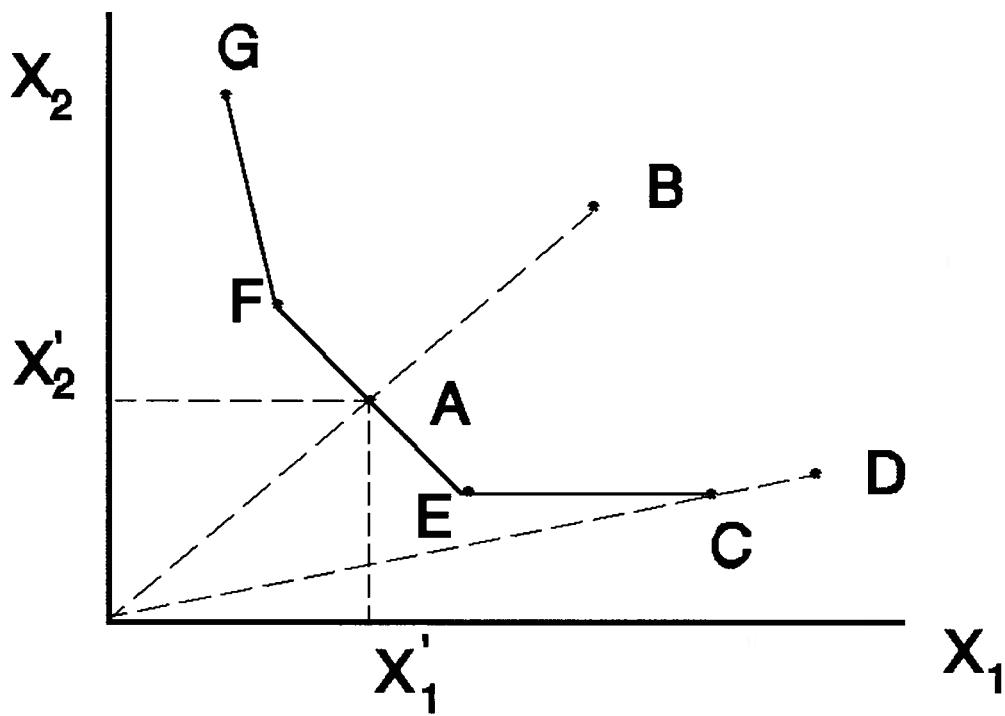


Figure 4.1 Diagrammatic Interpretation of DEA

level of output it is possible to find a firm, or a linear combination of firms, which is using less of at least one of the inputs. Consider firm B, for example, with an efficiency ratio OA/OB which is less than unity. This reflects the fact that a linear combination of firms E and F is producing at least as much output as B with less of X_1 and X_2 . That is, the existing input consumption at firm B can be adjusted by the efficiency ratio to X'_1 and X'_2 in figure 4.1 while maintaining its current level of output. Therefore, the peer group (or the reference firms) for firm B are firms E and F. For firm B, the optimal solution in the dual is:

$$h_B^* = OA/OB < 1$$

and the constraints are:

$$\text{input 1} \quad X_{1B}h_B^* - 0 = X_{1E}\lambda_E^* + X_{1F}\lambda_F^*$$

$$\text{input 2} \quad X_{2B}h_B^* - 0 = X_{2E}\lambda_E^* + X_{2F}\lambda_F^*$$

$$\text{and on output} \quad Y_{1B} + 0 = Y_{1E}\lambda_E^* + Y_{1F}\lambda_F^*$$

Target performance for B, $X_{iB}h_B^*$, $i=1,2$, is equal to a linear combination of performance at firms E and F where $\lambda_E^*, \lambda_F^* > 0$ and the weights on the other firms are all zero: $\lambda_j^* = 0$, $j \neq E, F$.

One can see that the solution for firm B has all input and output slacks equal to zero. However, firm D has a non-zero slack on input X_1 . The efficiency ratio for D is OC/OD which defines an initial radial contraction in both inputs. However at point C, firm E is

producing the same output for less of X_1 and the same amount of X_2 . Hence D is not fully efficient until it reduces its consumption of X_1 by the horizontal distance C to E. This distance is given by a non-zero slack s_1^* in the final solution of the dual for firm D:

$$h_D^* = OC/OD$$

and the input constraints are

$$\text{input 1} \quad X_{1D}h_D^* - s_1^* = X_{1E}\lambda_E^*$$

$$\text{input 2} \quad X_{2D}h_D^* - 0 = X_{2E}\lambda_E^*$$

$$\text{and on output} \quad Y_{1D} + 0 = Y_{1E}\lambda_E^*$$

The target for D is radial contraction in both inputs given by h_D^* plus the additional reduction in X_1 , given by s_1^* . Its peer group is firm E alone since its target coincides exactly with performance observed at this best-practice firm. Thus $\lambda_E^* = 1$ and $\lambda_j^* = 0$ for $j \neq E$.

From the above discussions, we can see that there are two aspects to the target in the dual. The input constraints define a radial (or equi-proportionate) reduction in inputs given by the efficiency ratio, h_p^* , plus any further reductions in inputs suggested by non-zero slacks. In addition, however, the presence of non-zero output slacks may require adjustments to outputs.

4.2.2 The BCC model

The basic CCR model assumes constant returns to scale. That is, proportional changes in all input levels result in changes of equal proportion in output level. In practice, one may find this assumption too restrictive. Banker, Charnes and Cooper (1984) extends the original CCR formulation to incorporate the effect of returns to scale on the efficiency. The BCC model adds an additional restriction to the envelopment requirements. It requires that the reference point on the production frontier for DMU_k be a convex combination of the observed efficient DMUs. It is formulated as:

$$\begin{aligned}
 h_k &= \text{Max} \sum_{r=1}^s u_r Y_{rk} - u_0 \\
 \text{st.} \quad &\sum_{r=1}^s u_r Y_{rj} - \sum_{i=1}^m v_i X_{ij} - u_0 \leq 0 \quad j=1, \dots, n \\
 &\sum_{i=1}^m v_i X_{ik} = 1 \\
 &u_r, v_i \geq \epsilon \quad r=1, \dots, s; \quad i=1, \dots, m
 \end{aligned} \tag{4.2.6}$$

The corresponding dual problem is formulated as:

$$\begin{aligned}
 h_k &= \text{Min} \quad \theta_k - \epsilon (\sum_r \sigma_r + \sum_i s_i) \\
 \text{s.t.} \quad &\sum_j y_{rj} \lambda_j - \sigma_r = y_{rk} \\
 &\sum_j x_{ij} \lambda_j - \theta_k x_{ik} + s_i = 0 \\
 &\sum_j \lambda_j = 1 \\
 &\lambda_j \geq 0, \quad j\text{-DMUs}, j=1, \dots, n \\
 &\sigma_r \geq 0, \quad r\text{-outputs}, r=1, \dots, s \\
 &s_i \geq 0, \quad i\text{-inputs}, i=1, \dots, m
 \end{aligned} \tag{4.2.7}$$

Comparing formulation (4.2.3) with formulation (4.2.7), we can see that the only difference between the two formulations is the additional constraint in formulation (4.2.7). This additional constraint ensures that a DMU is evaluated only by reference to original data points and their "convex combinations" on the efficiency frontier. The new variable, u_0 , in the primal problem is unconstrained in sign, and is interpreted by BCC as an indicator of returns to scale. Banker, Charnes and Cooper (1984) shows that the returns to scale at the referent efficient point are estimated by the sign of the variable u_0 : $u_0 < 0$ indicates increasing returns to scale; $u_0 > 0$ indicates decreasing returns to scale; $u_0 = 0$ indicates constant returns to scale¹².

It should be noted that the returns to scale indicated by u_0 are "local" in the sense that they are applicable only to the facet on the efficiency frontier where the reference point for the efficiency evaluation is positioned.

From the discussions above, a number of features of DEA become apparent. First, the DEA efficiency ratios are solely dependent on observed best practice in the sample. One of the consequences of measuring efficiency relative to observed best practice is that DEA is usually considered to be sensitive to extreme outliers and measurement errors. Secondly, since the weights for each DMU are chosen so as to give the most favourable efficiency ratio possible subject to the specified constraints, DEA evaluates a DMU as efficient if it has the best ratio of any one output to any one input. Therefore, the DEA efficiency ratios could be sensitive to the selection of inputs and outputs included in the analysis.

¹² Note that u_0 is added to correct for the effect of non-constant returns to scale, but the magnitude of this parameter is not directly interpretable as a measure of returns to scale.

4.3 The Effects of Exogenous Variables

There are mainly two motivations for studying the frontier techniques: (1) the desire to measure inefficiency; and (2) the desire to see how efficiency is related to observable characteristics of the firm and production environment. Both the parametric approach and the non-parametric approach discussed in this chapter focus on how to construct a production frontier and how to measure efficiency relative to the estimated frontier. The resulting efficiency measures are based on observed units of outputs and inputs, and thus ignore the effects of variations (except for the statistical noise with the stochastic frontier models) in the market, operating, institutional and regulatory policy environments, and other specific factors which may affect the observed production of the DMUs. Therefore, these efficiency scores may represent factors other than efficiency. In order to make meaningful comparisons about the relative performance of the firms (DMUs), and to identify the sources of observed efficiency differentials, additional analysis would be necessary in order to purge these efficiency scores of influences of these "exogenous variables".

As pointed out by Nerlove (1965), the observed differences in production performance among a group of firms may be attributed to three general sources: (1) ability to maximize short-run profits, given a particular production function and in a given environment; (2) the production function itself which summarizes the state of technical knowledge and the possession of fixed factors; and (3) the environment. Within the framework of efficiency measurement discussed earlier we assume that the firms have the same production technology, thus we can ignore the second source while the first source may be broadly thought of as inefficiency. As for the environmental factors, some efforts must

be made to standardize environment, ideally through the use of some meaningful quantitative variables, in order to permit measurement of relative efficiency since firms cannot physically be transferred from one environment to another.

The observed inefficiency may be related to various factors that could explain them, such as environmental conditions, administrative structures, social constraints, the quality of production factors, *etc.* These factors are referred to as exogenous variables in this dissertation since they are mostly beyond the control of the firms (DMUs) in question. The relation between the observed efficiency level and the exogenous variables may be estimated after having assessed the production frontier and measured the observed efficiency. This approach is often associated with the non-parametric methods such as in Ray (1991), McCarty and Yaisawarng (1993) and Oum and Yu (1994), but has also been used with the parametric methods as well such as in Bruning (1991) and Loeb (1994). This approach of incorporating exogenous variables is referred to as the two-step procedure in this dissertation. Alternatively, the relation may be estimated while assessing the production frontier in which case the frontier and the functional relationship explaining efficiency differences are simultaneously estimated. This approach is often associated with the parametric efficiency estimation methods for obvious reasons. An example of using this second approach, referred to as the one-step procedure, can be found in Lee and Schmidt (1993). Since with the one-step parametric methods, the exogenous variables can be incorporated directly in the estimation of the production function simply by introducing additional variables, it is not necessary to make any methodological modifications with frontier models described in section 4.1. Therefore, this section focuses on how to account

for the effects of exogenous variables through the use of the two-step procedure, particularly with the nonparametric methods. The two-step parametric methods follow the same procedure as the two-step DEA-regression procedure described below.

Both CCR and BCC methods measure efficiency based on units of outputs and inputs, but ignore variations in the market, operating, institutional and regulatory policy environments, and other factors which may affect the observed performance of the DMUs. Therefore, the DEA efficiency indices from CCR and BCC models may represent factors other than efficiency. To derive meaningful inferences about the relative performance of the DMUs, additional analysis is necessary to purge the DEA efficiency indices of influences of these "exogenous variables". A second stage regression analysis can be used to identify the effects of these variables and to measure the "residual" efficiency¹³ which is a closer indicator of relative performance of the DMUs than the gross DEA efficiency index. A number of studies have used this DEA-regression procedure in empirical applications including Ray (1991), Fazel and Nunnikhoven (1992), and McCarty and Yaisawarng (1993).

In the second stage, the DEA efficiency scores are used as dependent variable in a regression on the exogenous variables. There is a methodological problem with such a regression. As defined in sections 4.2.1 and 4.2.2, the DEA efficiency index falls between 0 and 1 ($0 \leq h_k^* \leq 1$), making it a limited dependent variable. Consequently, an OLS regression of h_k^* would produce biased and inconsistent parameter estimates. In order to

¹³ The term residual efficiency is used here instead of "true" efficiency, because in practice one may not be able to identify all potential influential factors thus the efficiency measures from the regression analysis may still reflect the influences of some neglected factors.

treat the limited dependent variable properly, the following form of the Tobit model (Tobin, 1958; Amemiya, 1985) is adopted ¹⁴:

$$Lh_j = \begin{cases} Z_j\beta_j + \eta_j & \text{if } Lh_j < 0, \quad j=1,\dots,n \\ 0 & \text{otherwise} \end{cases} \quad (4.3.1)$$

where Lh_j is the logarithm of the DEA efficiency index for firm j , Z_j is a vector of the logarithms of the variables potentially influencing the DEA efficiency scores, β is a vector of coefficients to be estimated, and $\{\eta_j\}$ are assumed to be independently identically distributed error terms which can take on negative, zero, or positive values. The residual efficiency is defined as:

$$Eff_j = EXP(\eta_j - \eta_{\max}) \quad (4.3.2)$$

where Eff_j is the residual efficiency score for the j -th DMU, and η_{\max} is the maximum of $\{\eta_j\}$. Note that the DMU with the largest positive η_j will be considered as the most efficient DMU, and given a residual efficiency score of one. By definition, Eff_j also falls between zero and one.

This two-step procedure accomplishes two tasks. First, it identifies the effects of potential influential factors on the DEA efficiency index. Second, it allows one to compute the "residual" efficiency index from the residuals of the Tobit regression.

¹⁴ The Tobit Model allows one to incorporate only one bound on the dependent variable while the DEA efficiency index is constrained between zero and one. By taking the logarithm of the DEA efficiency index, one can convert the dependent variable to have only one (upper) bound, zero.

4.4 Summary

This section summarizes the comparative merits of the three alternative methods based on their theoretical and methodological differences.

- The parametric methods use econometric techniques to estimate the frontier function, thus the estimators have identifiable statistical properties. However, they impose a particular functional form on the estimated frontier function and confounds the effects of misspecification of functional form with inefficiency. On the other hand, the data envelopment analysis method does not impose an explicit functional form on the frontier and is less prone to misspecification error in this regard. Therefore, DEA may have a comparative advantage over the parametric methods in situations where the underlying production technology does not meet the classical assumptions. The parametric methods are likely to have a comparative advantage when the classical assumptions are met. Further, the DEA method assumes that the production possibility set is convex. This assumption is violated in the case of increasing returns to scale. Therefore, the DEA method has an inherent disadvantage in situations with increasing returns to scale.
- The parametric methods are statistical methods, they normally require a relatively large sample size (depending on the particular functional form being estimated). DEA on the other hand is less restrictive in terms of the sample size since the feasibility of the DEA analysis depends on only whether there are enough observations to span the convex cones in input and output space. However, all three methods are expected to yield better estimates as the sample size increases.

- The deterministic frontier method and the data envelopment analysis are both deterministic, and attribute all deviation from the frontier to inefficiency. No allowance is made for noise, measurement error, and the like. Therefore, a single errant observation (efficient outlier) can have profound effects on the estimates. There is no *a priori* answer for which method is more sensitive to the effects of outliers and data errors. Any conclusion regarding the relative merits of these two methods in this aspect would be drawn from empirical evidence. Unlike the two deterministic methods, the stochastic frontier method attempts to distinguish the effects of random noise from the effects of efficiency and is less sensitive to potential outliers or other extraordinary behaviour of observations. However, additional structure is imposed on the distribution of inefficiency. Thus, specification error of the efficiency term may affect the accuracy of the efficiency estimates. Nonetheless, the stochastic frontier method is expected to yield more robust efficiency estimates than the other two methods especially when there are large variations in the firms' operating environments.
- The DEA method is designed to deal with production technologies with multiple inputs and multiple outputs. On the other hand, the parametric methods may have difficulty estimating production frontier functions in situations involving multiple outputs unless the dual cost frontier functions could be estimated.
- The deterministic frontier method appears to combine the bad features of the econometric and programming approaches to frontier construction: it is deterministic and parametric. It is still used in practical applications because it is simple and easy

to apply. The DEA method requires repeated solutions of linear programming problems, thus usually incurs more computational costs than the parametric methods especially for large samples. The stochastic frontier method involves complex estimation procedures and may not be easy to use. However, special computer software is available for estimating all three methods.

- The one-step procedure for incorporating the effects of exogenous variables is expected to produce efficiency estimates closer to the "true" values if the exogenous variables can be correctly identified and accounted for, and specification errors (related to the exogenous variables) are not very serious. The two-step procedure is intuitively more appealing to decision makers since it relates firms' performance directly to the potentially influential factors. However, cumulative specification errors from both steps are likely to have a negative effect on the accuracy of efficiency estimates.

The foregoing discussions indicate that one cannot always draw solid conclusions regarding how well each of the methods performs in comparison with the other alternative methods under different scenarios on the basis of their conceptual and methodological differences. Therefore, empirical evidence is necessary to help provide useful guidelines for selecting the appropriate methodology in practical applications. The Monte Carlo experiments described in the following chapter are intended to provide such information.

Chapter 5

Design of the Monte Carlo Experiments

The Monte Carlo experiments are often used to "solve" those analytical problems that are technically intractable. Monte Carlo experimentation solves these problems by substituting an equivalent stochastic problem and solving this simulated problem. As described in Chapter 1, Monte Carlo experiments are used herein to examine the relative performance of the three alternative efficiency measurement methods with respects to the sample size, the variations in input values, the noise level, the exogenous variables, the data outliers, and the different (underlying) production structures. This chapter outlines the basics of the Monte Carlo experiments.

For each Monte Carlo experiment, we need to specify: (1) the "true" underlying production technology $y = f(X)$, (2) a sample size N , and (3) values for σ_u^2 and σ_v^2 , the variances of the two error terms, and we also need to generate a $N \times I$ input data matrix in which I is the number of inputs specified in the "true" production technology.

5.1 Specification of the "true" underlying production technology

The first step in conducting the Monte Carlo experiments is to specify the "true" underlying production technology. This "true" technology must satisfy the following regularity conditions:

- (1) The production of an output always requires the use of at least one input.
- (2) Outputs are finite for all finite inputs.
- (3) An increase in inputs can not lead to a reduction in outputs.

- (4) A reduction in outputs remains producible with no change in inputs.
- (5) The production possibility set is closed.

In practical situations, the choice of a particular functional form depends largely on *a priori* information about the underlying technology. Without such information, the choice of functional form is usually based on its flexibility. In this study, we assume the CRESH (Constant Ratio of Elasticity of Substitution, Homothetic) production function as the underlying production technology. The CRESH function was developed by Hanoch (1971). The well-known CES (Constant Elasticity of Substitution) function as well as its limiting forms (the Cobb-Douglas, Leontief, and linear functions) are special cases of CRESH. That is, CRESH function nests a number of commonly used functional forms. Moreover, the CRESH function allows different patterns of substitution or complementarity among three or more inputs, and different returns to scales for changing input proportions. This makes it possible to examine the performance of the alternative methods under different degrees of input substitution and returns to scale. Because of these advantages, this functional form has been used in a number of previous Monte Carlo studies as the underlying production technology, including Guilkey and Lovell (1980), Guilkey, Lovell and Sickles (1983), Gong and Sickles (1989), and Gong and Sickles (1992). CRESH function is thus selected as the underlying true production technology in this study. The CRESH production function is specified as follows:

$$ye^{\theta y} = \left(\sum_{i=1}^n \delta_i x_i^{-\rho_i} \right)^{-\gamma/\rho} \quad (5.1)$$

where $\theta \geq 0$, $\gamma > 0$, $\delta_i > 0$, for all i , and $\sum \delta_i = 1$. Function (5.1) has variable returns

to scale $r(x)$, and variable Allen-Uzawa partial elasticities of substitution $\sigma_{ij}(x)$, which are given by:

$$r(x) = \frac{\gamma}{1+\theta y} \frac{\sum_{i=1}^n \rho_i \delta_i x_i^{-\rho_i}}{\sum_{i=1}^n \rho_i \delta_i x_i^{-\rho_i}} \quad (5.2)$$

$$\sigma_{ij}(x) = \frac{1}{(1+\rho_i)(1+\rho_j)} \frac{\sum_{k=1}^n \rho_k \delta_k x_k^{-\rho_k}}{\sum_{k=1}^n \left(\frac{\rho_k \delta_k}{1+\rho_k}\right) x_k^{-\rho_k}}, \quad i \neq j \quad (5.3)$$

Under certain parametric restrictions function (5.1) becomes one of a number of well known production functions. For example, if $\theta = 0$ then equation (5.1) is an almost homogeneous CRES (Constant Ratio of Elasticity of Substitution) function; if $\rho_1 = \dots = \rho_n$ it is a homothetic CES (Constant Elasticity of Substitution) function; if $\theta = 0$ and $\rho_1 = \dots = \rho_n$ it is a homogeneous CES function, and it has the homothetic Cobb-Douglas function as a limiting form as $(\rho_1 = \dots = \rho_n) \rightarrow 0$.

The discussions above consider simply the relation between inputs and output. In practice, firms operate under different conditions. The observed productions are influenced by factors outside firms' control. To simulate closer to a real situation, therefore, the effects of operating environments should be incorporated into the "known" underlying production

situation. Let us assume that all firms in an industry have the same underlying production function, but operate in different environments with different levels of efficiency¹. Thus the observed production of a representative firm may be described by the following functions:

$$Y = f(x) \exp(\omega) \quad (5.4)$$

$$\omega = g(z) + \nu + \mu \quad (5.5)$$

where Y is output, x is a vector of inputs, $f(x)$ is the industry's underlying production technology, ω represents the aggregate deviation from the deterministic core of the production frontier as specified by $f(x)$, z is a vector of exogenous variables reflecting firm characteristics and operating environments, $g(z)$ is the firm specific effects on the observed production², ν is a random disturbance which captures the statistical noises, and μ represents the efficiency level of an individual firm.

The industry underlying production technology $f(x)$ is assumed to take the functional form specified by function (5.1). An appropriate form for $g(z)$ is not obvious. Herein we assume $g(\cdot)$ is a linear function of z . The disturbance ν is typically assumed to be normally

¹ In this study, we focus only on technical efficiency thus assume that the firm is allocatively efficient. The definitions of allocative efficiency and technical efficiency are given in Chapter 3. See Farrell (1957) for more discussions on these concepts.

² Reifschneider and Stevenson (1989) consider $g(z)$ as the systematic influences on the firm's inefficiency level. We assume here that the firm does not have direct control over the variables z and we want to examine how the changes in z affect the estimated efficiency scores while the "true" efficiency levels unchanged. Therefore, we choose to separate $g(z)$ from the efficiency term u .

distributed and represents the statistical noise which is not captured by $g(z)$. The disturbance μ is assumed to be a nonpositive error term reflecting technical inefficiency and follow the half normal distribution. The half-normal distribution is the most popular distributional assumption for the efficiency component in the empirical literature on efficiency measurements (Lovell, 1993).

5.2 Determination of Sample Size and Number of Replications

As a general principle, the larger the sample size and the more the replications, the better the results. However, computation cost increases rapidly as the sample size and number of replications increase especially for the DEA models. The DEA method requires the solution of one linear programming problem for EACH observation point. For example, for a sample with 100 observations, DEA needs solving 100 different linear programming problems to produce a set of efficiency scores. In addition, each additional observation introduces another variable into the dual formulation of the DEA model which further reduces the computation speed of the LP problem³. Due to these reasons, for most of the experiments, the sample size is set at 250 observations which reflects the "real" sample size often found in the empirical literature. However, a set of experiments are conducted to examine the effects of sample size on the relative performance of the three alternative methods. Each experiment is repeated 25 times. Appendix B tests the hypotheses whether

³ Each additional observation will add one more constraint in the primal formulation, which will have even greater effect on the computation speed of the LP program. See Appendix A for more discussions. Appendix A also lists a sample computer code for computing DEA efficiency index.

or not there are any significant differences between the results with 25 replications and the results with 50 and 100 replications. The test results indicate that there is no significant difference. Therefore, 25 replications are adequate for this study.

5.3 Generation of Inputs, Exogenous Variables and the Two-component Errors

For the sake of simplicity, we consider an one-output three-input production technology as specified by equations (5.1), (5.4) and (5.5). The inputs are drawn randomly and independently from a log-normal distribution, and fixed throughout an experiment. The log-normal assumption is considered as a reasonable approximation to reality when a random variable is regarded as representing the joint effects of a number of factors. It also has the advantage that the possibility of generating negative values is avoided. It has been used for input generation in previous simulation studies such as Wales (1977), Guilkey and Lovell (1980), and Gong and Sickles (1992). The procedure used to generate the log-normal random variables and the summary statistics of the inputs are given in Appendix B.

One exogenous variable is assumed. The generality of the results will not be lost, since this variable may be considered as an aggregate indicator reflecting all identifiable exogenous factors. The exogenous variable is generated by αz where z is drawn from $N(0,1)$, and the value for α varies according to the requirement of a particular experiment. Consequently, the effects of the exogenous variables on the observed production output, as reflected by $\exp(g(z))$, follow a log-normal distribution. There have been very limited simulation studies which incorporate exogenous variables in the production or cost functions. Nerlove (1971) includes an uniformly distributed exogenous variable in his Monte Carlo

study. Normal distribution is preferred to the uniform distribution here because of the same reasons as discussed for the input variables.

The noise component of the error term is drawn randomly and independently from a normal distribution $N(0, \sigma_v^2)$ over the range $(-\infty, +\infty)$. The variance σ_v^2 varies across experiments according to the requirement of a particular experiment. Normality is routinely assumed for the statistical noise, and it is considered as a reasonable approximation to reality when the sample size is large enough to rely on the Central Limit Theorem. The inefficiency component of the error term is drawn randomly and independently from a half-normal distribution which takes absolute values from $N(0, 0.36)$. The half-normal distribution is used here because it is the most popular distributional assumption for the inefficiency term in empirical studies. The values for this inefficiency term are fixed throughout experiments. The choice of this particular value for σ_u^2 would not likely to affect the inferences from the experimental results in any significant way. This is examined by conducting additional experiments assuming $\sigma_u^2 = 1$. The results show slight increases in the correlations between the "true" and the estimated efficiencies for all three methods (the mean efficiency estimates fall as expected). However, the relative performance of the alternative methods does not appear to be affected by the changes in σ_u^2 . This is also confirmed by Gong and Sickles (1992). They show that the correlations between the true and estimated efficiency rise as σ_u^2 increases (for given σ_v^2) for both the stochastic frontier method and the DEA method with the stochastic frontier method dominating DEA for all values of σ_u^2 examined.

5.4 Generation of the output observations

We have fixed $\theta=0$, $\gamma=1$, $\delta_1=\delta_2=0.3$, and $\delta_3=0.4$ in equation (5.1) throughout the experiments. These values are assigned to the parameters mainly due to computational convenience. Other parameters are given different values for different experiments so that the underlying production technologies exhibit various degrees of factor substitution and returns to scale. Therefore, it is adequate to fix θ , γ , δ_i at this particular set of values. This is done mainly to keep the amount of computation requirements within a manageable range. It should not have any dramatic effect on the generality of the experiment results.

After the inputs, exogenous variable, and the two error terms are generated, the corresponding values for the "observed" output can be generated by means of equations (5.4) and (5.5). This provides us with a sample data base with 250 observations on output, three inputs and one exogenous variable.

5.5 Conduct of the Experiments

Six sets of experiments are conducted to examine the relative merits of the three alternative methods from different aspects. In the first five sets of experiments, we limit the range of technologies to those which exhibit constant returns to scale in order that the comparison between the alternative methods is not distorted by the treatment of scale economies. In particular, three production technologies are assumed with input substitution at 3.03, 1.03 and 0.333 respectively. The BCC model is not included in these experiments since this variant of the DEA methods is intended to deal with the problem of non-constant returns to scale. In experiment six, non-constant returns to scale and input complementarity

are introduced into the underlying production technology to examine how each of the three alternative methods performs under complex technologies. In this last set of experiments, the BCC model is compared with the other three models. The experiments are described as follows.

Set 1: The Effects of Sample Size

The first set of experiments examines whether sample size (N) has any effect on the performance of the three alternative methods. In practice, people may encounter small samples with less than 100 observations, or large samples with over 500 observations. If the relative performance of the alternative methods depends heavily on the sample size, then people working with empirical data would need to take this into consideration in selecting the appropriate method for their particular problems.

In these experiments, the standard deviation for the noise term (σ_v) is set at 0.15. To avoid the distortion caused by the effects of exogenous variables and to simplify the computation, the effects of exogenous variables are assumed to be negligible, that is, α is set to 0. Three sets of sample data are generated with sample size of 100, 250, and 500, respectively. The alternative methods are then applied to each of the three data sets.

Set 2: The Effects of Input Range

This set of experiments is intended to examine how sensitive the alternative methods are to the homogeneity of the DMUs or observations in the sample. The less variations in inputs, the more homogenous the observations will be. In some data sets there are very

small variations in the values of input variables among the sample firms, while in other data sets one may face observations with input levels varying over a wide range. If a method is very sensitive to the homogeneity assumption, one may need to choose an alternative method in situations where there are large variations among the DMUs or observations. This set of experiments is also intended to verify whether the input specifications used in other sets of experiments are reasonable since the input variables are fixed throughout all the experiments except for SET 2.

Inputs (X) are assumed to follow a log-normal distribution. X are generated as $X = e^w$, where $w = 3 + \tau x_0$, and x_0 is drawn from $N(0,1)$. The values of input variables for all other sets of experiments are generated assuming τ equal to 1⁴. In this set of experiments, τ is given the values of 0.1, 1, 1.5. Thus, three sets of sample data are generated. In all three data sets, sample size (N) is fixed at 250, and the standard deviation for the noise term (σ_v) is set at 0.15, and the exogenous variable is ignored by setting α equal to 0.

Set 3: The Effects of Noise Level

Set 3 of the experiments examines the effects of noise level on the relative performance of the alternative methods. The more robust the method, the less the effects of noise level on its performance. In situations where there are large variations in the operating environments, one would probably need to use a method which is less sensitive to the level of noise.

⁴ See Appendix B for more discussions and summary statistics for the input variables.

Throughout set 3, the sample size (N) is fixed at 250, and as in SET 1, α is fixed at 0 assuming the effect of quantifiable exogenous factors is negligible. By varying the standard deviation of the noise term (σ_v) at 0.03, 0.15, 0.25, 0.50, and 0.75, five sets of sample data, with different levels of noise, are generated. The three alternative methods are then applied to each of these data sets.

Set 4: The Effects of Exogenous Variables

In most practical situations, there is a large number of exogenous factors⁵ affecting the productive performance of the firms or organizations. Some of these exogenous factors can be quantitatively identified, some cannot. The statistical noise reflects the aggregate effects of the exogenous factors which cannot be quantitatively identified. Experiment 4 is intended to examine how well each of the alternative methods deals with those exogenous factors which can be quantitatively identified.

In these experiments, sample size (N) is again fixed at 250, and the standard deviation for noise (σ_v) is set at 0.15. The extent of the effects of exogenous factors is reflected by the values of α which is set at 0.05, 0.10, 0.25, 0.50, 0.75, 1.00, and 2.00.

Set 5: The Effects of Outliers

This set of experiments examines how sensitive each of the three alternative methods is to the "efficient" outliers or the measurement errors in the sample data. In measuring the

⁵ For the sake of computational convenience, it is assumed that productive performance has no effects on those exogenous variables.

relative efficiencies of the DMUs or organizations, the "best practice" DMUs are used as the reference points for other DMUs. If one of the "best practice" DMUs happens to be an "efficient" outlier, then all the other DMUs whose performance is evaluated relative to this "efficient" outlier would have a misleading performance rating. On the other hand, an "inefficient" outlier would only lead to an under-estimated performance rating for itself.

An "efficient" outlier may be an indicator of using inappropriate input and output variables or model misspecification, or may be evidence of an extraordinary event or condition, or may simply be due to data errors. There are two general strategies for dealing with outliers. The first strategy requires that outliers be detected and their causes investigated. The offending observation can then often be corrected, deleted for good reason, or otherwise given individual attention. The second strategy is to use robust methods so that outliers have little influence over any inferences or conclusions. If the second strategy is chosen, it is necessary to know how robust the alternative methods are in order to select the appropriate method. Both the DEA and the deterministic frontier method have been criticized as being sensitive to outliers. This set of experiments is intended to find out how sensitive these methods actually are to the outliers.

Again the sample size (N) is fixed at 250 in these experiments, and the standard deviation of noise (σ_v) is set at 0.15. The exogenous variables are assumed to be negligible, thus α is set at 0. To minimize computational requirements, it is assumed that there is only one outlier in the data set. To create an "efficient" outlier, the output level of an efficient

observation is then increased by 5%, 25%, 50%, 75% and 100%⁶.

Set 6: The Effects of the Underlying Production Technology

The first five sets of experiments consider only technologies of constant returns to scale, so that the relative performance of the alternative methods is not distorted by different treatments of returns to scale. This restriction is relaxed in this set of experiments. The parameters of the underlying production technology are chosen such that the returns to scale, $r(x)$, and the Allen-Uzawa partial elasticities of substitution, $\sigma_{ij}(x)$, fall within the ranges of $0.92 < \gamma < 1.57$, and $-0.27 < \sigma_{ij} < 0.54$. These ranges for returns to scale and input substitutions are determined from reviews of empirical studies on production and cost characteristics in transportation industry. See, for example, Borger (1991) and McGeehan (1993) for railways, Gillen, Oum and Tretheway (1985) and Keeler and Formby (1994) for airlines. The transportation industry is considered here because in the second part of this dissertation the alternative methods are applied to a railway data set and an airline data set to examine the efficiency performance of railways and international airlines.

As in the other sets of experiments, the sample size N is fixed at 250, the standard deviation of the noise term (σ_v) is set at 0.15, α is set at 0 assuming the absence of the effects of exogenous factors.

⁶ Note that the observed "efficient" outlier may actually be inefficient.

5.6 Evaluation Criteria

The performance of the three alternative methods is evaluated using the following statistics: the mean efficiency estimates, the mean absolute deviations (MAD) of the efficiency estimates from the true values, the correlation coefficient, and the rank correlation coefficient between the estimated level of efficiency and the "true" level of efficiency.

The mean efficiency estimates and the MAD examine the ability of the alternative methods to approximate the actual level of efficiency. The MAD measures the degree of deviations, and the mean efficiency estimates provide an indicator as to which direction the estimates deviate from their true values, eg. overestimation vs underestimation.

Pearson product-moment correlation coefficient is used here to measure the strength of relatedness between the estimated and "true" efficiency levels. It is a number that varies from -1 to +1. A correlation of +1 denotes a perfect positive relationship. A correlation of -1 denotes an inverse relationship. A zero value of correlation means that there is no linear relationship. The Pearson correlation coefficient for the estimated efficiency ee and the "true" efficiency EE is measured as:

$$\gamma = \frac{\Sigma(ee - \bar{ee})(EE - \bar{EE})}{\sqrt{\Sigma(ee - \bar{ee})^2 \Sigma(EE - \bar{EE})^2}} \quad (5.6)$$

where the \bar{ee} and \bar{EE} are the average of ee and EE respectively.

The rank correlation coefficient measures the correlation between the rankings of the estimated efficiency level and the "true" efficiency level. In this study, the Spearman's rank

correlation coefficient is used, it is derived from the Pearson product-moment correlation by using the ranks of the two variables considered instead of the raw data. The Spearman's rank correlation also falls within the interval of $(-1, +1)$. A value of $+1$ indicates a strong positive association between the rankings in which case the rankings of both groups are similar. A value of -1 indicates a strong negative association between the rankings where the rankings of the two groups would seem opposite. If the value is zero, one will observe no pattern between the rankings. The formula for calculating a Spearman's rank correlation is as follows:

$$\gamma_s = 1 - \frac{6\sum d^2}{n(n^2-1)} \quad (5.7)$$

where n is the number of pairs being correlated, and d is the difference in the ranks of each pair.

Chapter 6

The Experiment Results

This chapter reports and discusses the results from the Monte Carlo experiments. The sample data for the experiments are generated by SHAZAM 6.2 (White, Wong, Whistler, and Haun, 1990) on a P.C. The deterministic frontier models are estimated using SHAZAM on a UNIX computer. The DEA methods are also implemented on UNIX. The stochastic frontier models are estimated on a P.C. using the program FRONTIER developed by Coelli (1991). Appendix A gives a more detailed discussion on the computer program for the DEA and the stochastic frontier models. Appendix B gives a detailed description of the data generated.

As mentioned in Chapter 5, the choice of a particular functional form mostly depends on its flexibility when there is no prior information on the true production technology¹. For the deterministic frontier function and the deterministic core of the stochastic frontier function, therefore, we assume the following translog functional form:

$$\ln y = \beta_0 + \sum_{i=1}^n \beta_i \ln x_i + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \beta_{ij} \ln x_i \ln x_j + \sum_{k=1}^m \gamma_k \ln z_k \quad (6.1)$$

where y is the output, x_i are the inputs, z_k are the exogenous variables, and β , γ are the coefficients to be estimated. The translog function has been shown to be quite flexible, and has been widely used in empirical applications. In this study, we use a second order

¹ The simulated data sets in the Monte Carlo experiments are treated as "real world" data sets where information on the underlying production technology and error distributions is unknown.

approximation on inputs and a first order approximation on exogenous variables².

6.1 The Effects of Sample Size

The first set of experiments is concerned with how well the three methods perform when applied to samples of different sizes. Three sample sizes are examined: 100, 250, and 500. For each of the three sample sizes, three sets of data are generated by assuming three different underlying production technologies. The three technologies are specified by setting ρ in equation (5.1) equal to -0.67, -0.25, and +2.0, with the corresponding input substitution at 3.03, 1.33, and 0.33, respectively. Thus, nine sets of sample data are generated, and the three methods are applied to each of these nine data sets.

The mean efficiency estimates are listed in Table 6.1.1 while the mean absolute deviations (MAD) of the efficiency estimates are listed in Table 6.1.2. By comparing the mean efficiency estimates by the stochastic frontier method and the true means in Table 6.1.1 and looking at the corresponding MAD values in Table 6.1.2, we can see that the efficiency estimates by the stochastic frontier method approximate the true efficiency levels very well, and remain quite stable over different sample sizes. By comparing the MAD values for the DEA and the deterministic frontier method (Table 6.1.2), it is found that the DEA efficiency estimates are closer to the true efficiency values than those by the deterministic frontier method. It is also noted that the deterministic frontier method appears to underestimate the efficiency level for all three sample sizes, while the DEA method appears to overestimate

² Note that equation (6.1) specifies the frontier function to be estimated by the one-step procedure. As for the first stage of the two-step procedure, we simply ignore the exogenous variables.

Table 6.1.1
The Effects of Sample Size
The Means of Estimated Efficiency

Sample Size	N=100	N=250	N=500
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$: True Mean	0.6567 (0.0186)	0.6561 (0.0157)	0.6563 (0.0092)
Deterministic	0.5295 (0.0469)	0.4983 (0.0281)	0.4851 (0.0281)
Stochastic	0.6586 (0.0330)	0.6566 (0.0218)	0.6549 (0.0129)
DEA	0.6898 (0.0279)	0.6257 (0.0190)	0.5913 (0.0146)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$: True Mean	0.6567 (0.0186)	0.6561 (0.0157)	0.6563 (0.0092)
Deterministic	0.5293 (0.0456)	0.5041 (0.0275)	0.4861 (0.0292)
Stochastic	0.6639 (0.0305)	0.6569 (0.0219)	0.6556 (0.0132)
DEA	0.7141 (0.0234)	0.6502 (0.0188)	0.6138 (0.0131)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$: True Mean	0.6567 (0.0186)	0.6561 (0.0157)	0.6563 (0.0092)
Deterministic	0.4899 (0.0526)	0.4149 (0.0715)	0.3853 (0.0562)
Stochastic	0.6599 (0.0755)	0.6591 (0.0259)	0.6574 (0.0179)
DEA	0.7219 (0.0238)	0.6582 (0.0221)	0.6197 (0.0142)

Note that standard deviations are in parenthesis.

Table 6.1.2 The Effects of Sample Size The Mean Absolute Deviations			
Sample Size	N=100	N=250	N=500
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$:			
Deterministic	0.1435 (0.0340)	0.1628 (0.0275)	0.1737 (0.0248)
Stochastic	0.0836 (0.0125)	0.0755 (0.0045)	0.0695 (0.0030)
DEA	0.1113 (0.0012)	0.1075 (0.0102)	0.1097 (0.0081)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$			
Deterministic	0.1431 (0.0322)	0.1576 (0.0275)	0.1726 (0.0266)
Stochastic	0.0808 (0.0103)	0.0745 (0.0043)	0.0685 (0.0028)
DEA	0.1164 (0.0111)	0.1026 (0.0083)	0.1004 (0.0073)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$:			
Deterministic	0.1807 (0.0469)	0.2453 (0.0679)	0.2727 (0.0564)
Stochastic	0.1047 (0.0477)	0.0905 (0.0067)	0.0833 (0.0055)
DEA	0.1143 (0.0093)	0.0959 (0.0074)	0.0941 (0.0064)

Note that standard deviations are in parenthesis.

the efficiency level for the sample size of 100 but underestimate it for the sample size of 500.

As the sample size increases, the mean efficiency levels estimated by all three methods tend to fall as shown in Table 6.1.1. The most noticeable drop is observed in the estimates by the DEA method. This is because that the percentage of efficient observations in the sample would decrease as the sample size increases, and thus results in a decrease in the mean efficiency level. In terms of the MAD values, both the DEA and the stochastic frontier method show a slight improvement in their performance as the sample size increases. However, the opposite is true for the deterministic frontier method.

By comparing the mean efficiency estimates and the MAD values under the three different production technologies, it is noted that the performance of a particular method does vary with the underlying production technology. The stochastic frontier method is the least sensitive to the underlying production technology, it does not show any significant changes when the underlying technology is changed. However, when the elasticity of input substitution is less than unity the performance of the deterministic frontier method deteriorates substantially in terms of the MADs with the sample size. On the other hand, the performance of the DEA appears to improve slightly in terms of the MAD values as the elasticity of input substitution of the underlying production technology decreases. However, the relative performance of the three methods is not critically affected by assuming different underlying production technologies.

The correlation coefficients between the estimated and the true efficiencies are listed in Table 6.1.3 while the rank correlations between the true and estimated efficiency levels

Table 6.1.3
The Effects of Sample Size
Correlation Between the True and Estimated Efficiency

Sample Size	N=100	N=250	N=500
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$			
Deterministic	0.8357 (0.0271)	0.8609 (0.0195)	0.8781 (0.0119)
Stochastic	0.8657 (0.0353)	0.8900 (0.0128)	0.9028 (0.0087)
DEA	0.7394 (0.0619)	0.7818 (0.0451)	0.8228 (0.0234)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$			
Deterministic	0.8381 (0.0253)	0.8644 (0.0193)	0.8820 (0.0116)
Stochastic	0.8718 (0.0314)	0.8935 (0.0128)	0.9066 (0.0077)
DEA	0.7464 (0.0587)	0.7884 (0.0412)	0.8281 (0.0208)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$			
Deterministic	0.7845 (0.0341)	0.7989 (0.0234)	0.8201 (0.0195)
Stochastic	0.8304 (0.0411)	0.8453 (0.0178)	0.8610 (0.0154)
DEA	0.7659 (0.0520)	0.8080 (0.0384)	0.8407 (0.0202)

Note that standard deviations are in parenthesis.

Table 6.1.4 The Effects of Sample Size Rank Correlation Between the True and Estimated Efficiency			
Sample Size	N=100	N=250	N=500
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$			
Deterministic	0.8385 (0.0365)	0.8714 (0.0223)	0.8892 (0.0111)
Stochastic	0.8496 (0.0372)	0.8764 (0.0168)	0.8913 (0.0095)
DEA	0.7087 (0.0765)	0.7889 (0.0427)	0.8384 (0.0212)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$			
Deterministic	0.8415 (0.0347)	0.8742 (0.0220)	0.8929 (0.0104)
Stochastic	0.8545 (0.0348)	0.8804 (0.0171)	0.8952 (0.0083)
DEA	0.7042 (0.0743)	0.7855 (0.0395)	0.8354 (0.0195)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$			
Deterministic	0.7972 (0.0431)	0.8290 (0.0266)	0.8471 (0.0165)
Stochastic	0.8147 (0.0461)	0.8453 (0.0178)	0.8459 (0.0161)
DEA	0.7208 (0.0689)	0.8015 (0.0385)	0.8458 (0.0193)

Note that standard deviations are in parenthesis.

are listed in Table 6.1.4. Since we are more interested in the relative efficiency rankings among the DMUs, we rely more on the rank correlation coefficients in evaluating the relative performance of the three methods. The results in Table 6.1.3 and Table 6.1.4 clearly show that in terms of correlation and rank correlations all three methods perform better with a larger sample. The rank correlations for the stochastic frontier method and the deterministic frontier method are very close. That is, the stochastic frontier method and the deterministic frontier method would give very similar efficiency rankings of the DMUs. The rank correlations (and the correlation coefficients) for the DEA method are noticeably lower than those for the parametric methods, indicating that the two parametric frontier methods perform better than the DEA method. However, the gaps in their performance become smaller as sample size increases. A closer look at the results in Table 6.1.3 and Table 6.1.4 also shows that the performance of the two parametric methods deteriorates as the elasticity of input substitution of the underlying production technology departs from unity, but the performance of the DEA method improves as the elasticity of input substitution decreases.

6.2 The Effects of Input Range

In this section, the relative performance of the three methods is examined in relation to the variations in input variables. The input variables, X , are generated as $X = e^w$, where $w = 3 + \tau x_0$, and x_0 is drawn from $N(0,1)$. By varying the values of τ , one can vary the ranges over which input variables are taking values. In this set of experiments, τ is given the values of 0.1, 1, and 1.5. The summary statistics for input variables are listed in Table 6.2.1. All three input variables are assumed to vary over the same value ranges. Note that

the entries in Table 6.2.1 are the averages over 25 replications of the respective statistics.

Table 6.2.1 The Effects of Input Range The Statistics of Input Variables			
	$\tau = 0.1$	$\tau = 1.0$	$\tau = 1.5$
Mean of X	20.17	33.51	60.82
Standard Deviation of X	2.02	43.27	149.45
Maximum of X	26.63	382.84	1674.0
Minimum of X	15.08	1.28	0.40

The mean efficiency estimates are listed in Table 6.2.2. For the deterministic frontier method, it can be seen that the mean efficiency level tends to decrease as τ increases. This is because the efficiency estimates by the deterministic frontier method are measured against the largest OLS residual, and the OLS residuals are affected by the input variations. The larger the variations in the input variables, the larger the variations in the OLS residuals, and thus results in a lower average efficiency estimates by the deterministic frontier. On the other hand, the results in Table 6.2.2 shows that the mean efficiency estimates by the DEA method tend to rise with the increasing variations in input variables. This is because that in maximizing the ratio of weighted output over weighted input for a particular observation, the DEA method might give higher efficiency ratings to those observations which have inputs with very small values even when there are excess consumptions of other inputs. Therefore, the average efficiency estimates by the DEA could be higher when there are large variations in input variables. The mean efficiency estimates by the stochastic frontier method is

Table 6.2.2
The Effects of Input Range
The Means of Estimated Efficiency

	$\tau = 0.1$	$\tau = 1.0$	$\tau = 1.5$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.5071 (0.0364)	0.4983 (0.0281)	0.4699 (0.0411)
Stochastic	0.6528 (0.0218)	0.6566 (0.0218)	0.6597 (0.0214)
DEA	0.5962 (0.0252)	0.6257 (0.0190)	0.6407 (0.0191)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.5071 (0.0364)	0.5041 (0.0275)	0.4991 (0.0271)
Stochastic	0.6528 (0.0217)	0.6569 (0.0219)	0.6614 (0.0192)
DEA	0.5968 (0.0251)	0.6502 (0.0188)	0.6781 (0.0185)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.5071 (0.0363)	0.4149 (0.0715)	0.3135 (0.0721)
Stochastic	0.6528 (0.0218)	0.6591 (0.0259)	0.6678 (0.0789)
DEA	0.5995 (0.0249)	0.6582 (0.0221)	0.6653 (0.0199)

Note that the standard deviations are in the parenthesis.

relatively stable compared with those by the DEA and the deterministic frontier method. However, Table 6.2.2 indicates a slight increase in the mean efficiency estimates by the stochastic frontier method. One explanation for the slight increase in the mean efficiency estimates by the stochastic frontier method is that when there are large variations in the inputs the production possibility set is scattered over a larger area (or space). As a result, the stochastic production frontier may have a wider "band", and the resulting average efficiency level would hence be higher.

Table 6.2.3 presents the mean absolute deviations (MAD) of the efficiency estimates for different values of τ . There are very small fluctuations in the MAD values for the DEA when τ increases. Therefore, the DEA is considered as being rather robust to the variations in input variables in terms of the MADs. Under the first two production technologies where the elasticity of input substitution is over one, the stochastic frontier method shows very little changes in its MAD values as the variations in input variables change, but the deterministic frontier method sees slight rises in its MAD values as τ increases. Under the third production technology where the elasticity of input substitution is less than one, however, the performance of the two parametric methods deteriorates noticeably when the variations in input variables become larger.

Examination of the correlation coefficients in Table 6.2.4 and the rank correlations in Table 6.2.5 draws similar conclusions: (1) the DEA method does not appear to be very sensitive to variations in the input variables; (2) the two parametric methods yield relatively stable estimates when the elasticity of input substitution of the underlying production function is over one; and (3) when the elasticity of input substitution is set at 0.33, the degree of

Table 6.2.3 The Effects of Input Range The Mean Absolute Deviations			
	$\tau = 0.1$	$\tau = 1.0$	$\tau = 1.5$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$			
Deterministic	0.1545 (0.0287)	0.1628 (0.0275)	0.1902 (0.0366)
Stochastic	0.0714 (0.0049)	0.0755 (0.0045)	0.0793 (0.0054)
DEA	0.1169 (0.0149)	0.1075 (0.0102)	0.1024 (0.0097)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$			
Deterministic	0.1545 (0.0287)	0.1576 (0.0276)	0.1613 (0.0222)
Stochastic	0.0720 (0.0055)	0.0745 (0.0043)	0.0725 (0.0051)
DEA	0.1166 (0.0148)	0.1026 (0.0083)	0.1001 (0.0095)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$			
Deterministic	0.1545 (0.0286)	0.2453 (0.0679)	0.3455 (0.0703)
Stochastic	0.0721 (0.0054)	0.0905 (0.0067)	0.1214 (0.0442)
DEA	0.1150 (0.0144)	0.0959 (0.0074)	0.0987 (0.0070)

Note that the standard deviations are in the parenthesis.

Table 6.2.4
The Effects of Input Ranges
Correlation Between the True and Estimated Efficiency

Sample Size	$\tau = 0.1$	$\tau = 1.0$	$\tau = 1.5$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$			
Deterministic	0.8777 (0.0167)	0.8609 (0.0195)	0.8434 (0.0222)
Stochastic	0.9005 (0.0145)	0.8900 (0.0128)	0.8754 (0.0193)
DEA	0.7865 (0.0449)	0.7818 (0.0451)	0.7909 (0.0362)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$			
Deterministic	0.8777 (0.0167)	0.8644 (0.0193)	0.8701 (0.0173)
Stochastic	0.9005 (0.0146)	0.8935 (0.0128)	0.8984 (0.0153)
DEA	0.7869 (0.0448)	0.7884 (0.0412)	0.7992 (0.0370)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$			
Deterministic	0.8777 (0.0167)	0.7989 (0.0234)	0.6914 (0.0449)
Stochastic	0.9005 (0.0146)	0.8453 (0.0178)	0.7443 (0.0320)
DEA	0.7886 (0.0447)	0.8080 (0.0384)	0.8022 (0.0300)

Note that standard deviations are in parenthesis.

Table 6.2.5
The Effects of Input Ranges
Rank Correlation Between the True and Estimated Efficiency

Sample Size	$\tau = 0.1$	$\tau = 1.0$	$\tau = 1.5$
1. $\sigma_{12} = \sigma_{13} = \sigma_{23} = 3.03$			
Deterministic	0.8835 (0.0201)	0.8714 (0.0223)	0.8553 (0.0250)
Stochastic	0.8872 (0.0180)	0.8764 (0.0168)	0.8601 (0.0238)
DEA	0.8046 (0.0386)	0.7889 (0.0427)	0.7891 (0.0432)
2. $\sigma_{12} = \sigma_{13} = \sigma_{23} = 1.33$			
Deterministic	0.8835 (0.0201)	0.8742 (0.0220)	0.8778 (0.0218)
Stochastic	0.8872 (0.0181)	0.8804 (0.0171)	0.8838 (0.0213)
DEA	0.8047 (0.0387)	0.7855 (0.0395)	0.7817 (0.0466)
3. $\sigma_{12} = \sigma_{13} = \sigma_{23} = 0.33$			
Deterministic	0.8835 (0.0201)	0.8290 (0.0266)	0.7544 (0.0342)
Stochastic	0.8873 (0.0181)	0.8453 (0.0178)	0.7444 (0.0362)
DEA	0.8051 (0.0385)	0.8015 (0.0385)	0.7905 (0.0331)

Note that standard deviations are in parenthesis.

input variations has considerable negative effects on the performance of the two parametric methods. In particular, at $\tau = 1.5$, the DEA method even outperforms the stochastic frontier method under the third production technology whereas in all other cases the DEA is dominated by the stochastic frontier method.

6.3 The Effects of Noise

This section examines the effects of statistical noise on the relative performance of the three methods. As stated in Chapter 5, the noise term is generated from a normal distribution with zero mean. By giving the noise term different standard deviations, we are able to create different noise levels. In particular, the standard deviation is given the values of 0.03, 0.15, 0.25, 0.50, and 0.75. The corresponding statistics for the noise term are listed in Table 6.3.1. The entries in Table 6.3.1 are the averages over 25 replications of the mean statistics from each replication. Note that the means of e^v are sample means, thus they are not necessarily equal to one.

Table 6.3.1 The Effects of Noise The Statistics of Noise Term					
	$\sigma_v = 0.03$	$\sigma_v = 0.15$	$\sigma_v = 0.25$	$\sigma_v = 0.50$	$\sigma_v = 0.75$
Mean of e^v	1.001	1.013	1.036	1.136	1.310
St.Dev. of e^v	0.030	0.154	0.267	0.594	1.080
Maximum of e^v	1.089	1.542	2.068	4.121	7.722
Minimum of e^v	0.915	0.670	0.512	0.249	0.126

The mean efficiency estimates are shown in Table 6.3.2. As the noise level

Table 6.3.2 The Effects of Noise The Means of Estimated Efficiency					
	$\sigma_v=0.03$	$\sigma_v=0.15$	$\sigma_v=0.25$	$\sigma_v=0.50$	$\sigma_v=0.75$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.5818 (0.0355)	0.4983 (0.0281)	0.4005 (0.0454)	0.2471 (0.0513)	0.1627 (0.0410)
Stochastic	0.6674 (0.0155)	0.6566 (0.0218)	0.6720 (0.0284)	0.6895 (0.1185)	0.6804 (0.1621)
DEA	0.7051 (0.0160)	0.6257 (0.0190)	0.5403 (0.0314)	0.3970 (0.0362)	0.3032 (0.0377)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.5928 (0.0307)	0.5041 (0.0275)	0.4011 (0.0437)	0.2497 (0.0504)	0.1624 (0.0419)
Stochastic	0.6715 (0.0115)	0.6569 (0.0219)	0.6717 (0.0278)	0.6833 (0.1151)	0.6829 (0.1668)
DEA	0.7225 (0.0154)	0.6502 (0.0188)	0.5643 (0.0290)	0.4153 (0.0367)	0.3162 (0.0376)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$ True Mean	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
Deterministic	0.4458 (0.0722)	0.4149 (0.0715)	0.3652 (0.0566)	0.2391 (0.0568)	0.1596 (0.0405)
Stochastic	0.6465 (0.0197)	0.6591 (0.0259)	0.6780 (0.0498)	0.7024 (0.1345)	0.6824 (0.1601)
DEA	0.7273 (0.0149)	0.6582 (0.0221)	0.5724 (0.0298)	0.4185 (0.0390)	0.3154 (0.0401)

Note that standard deviations are in parenthesis.

increases, the mean efficiency estimates by the DEA and the deterministic frontier method fall dramatically, but the mean efficiency estimates by the stochastic frontier method remain relatively stable. This is expected since both the DEA and the deterministic frontier method treat all the deviation from the "best practice" frontier as inefficiency, thus the average efficiency level will fall as the noise level rise. On the other hand, the stochastic frontier method allows random variation of the frontier across the observations, thus it is much less sensitive to the noise level than the DEA and the deterministic frontier method. Another observation from Table 6.3.2 is that the deterministic frontier method tends to underestimate the "true" efficiency level in all cases. The DEA method appears to overestimate the "true" efficiency level when the noise level is low due to inner envelope property of the DEA's linear frontier³, whereas it tends to underestimate the efficiency level when the noise level is increased. The stochastic frontier method appears to overestimate the "true" efficiency level in almost all cases, however, the actual levels of mean efficiency estimates slightly fluctuate as the noise level changes and do not follow a clear upward or downward pattern.

As shown in Table 6.3.3, high noise level causes large MAD values for all three methods, that is, the high noise level is shown to impose considerable negative effects on the accuracy of the efficiency estimates, especially for the DEA and the deterministic frontier method. For all noise levels, the efficiency estimates by the stochastic frontier method have the least absolute deviation from the "true" efficiency level, whereas those by the deterministic frontier method have the largest absolute deviation.

³ It is only relevant when the noise level is very low.

Table 6.3.3
The Effects of Noise
The Mean Absolute Deviations

	$\sigma_v=0.03$	$\sigma_v=0.15$	$\sigma_v=0.25$	$\sigma_v=0.50$	$\sigma_v=0.75$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$					
Deterministic	0.0807 (0.0336)	0.1628 (0.0275)	0.2600 (0.0455)	0.4130 (0.0508)	0.4973 (0.0429)
Stochastic	0.0319 (0.0049)	0.0755 (0.0045)	0.1055 (0.0065)	0.1681 (0.0501)	0.2042 (0.0562)
DEA	0.0534 (0.0088)	0.1075 (0.0102)	0.1806 (0.0251)	0.3178 (0.0297)	0.4089 (0.0339)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$					
Deterministic	0.0695 (0.0297)	0.1576 (0.0276)	0.2592 (0.0437)	0.4104 (0.0495)	0.4976 (0.0439)
Stochastic	0.0282 (0.0048)	0.0745 (0.0043)	0.1049 (0.0070)	0.1656 (0.0483)	0.2068 (0.0586)
DEA	0.0682 (0.0089)	0.1026 (0.0083)	0.1678 (0.0221)	0.3042 (0.0283)	0.3990 (0.0333)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$					
Deterministic	0.2142 (0.0681)	0.2453 (0.0679)	0.2949 (0.0565)	0.4210 (0.0561)	0.5003 (0.0418)
Stochastic	0.0656 (0.0056)	0.0905 (0.0067)	0.1175 (0.0187)	0.1778 (0.0589)	0.2036 (0.0547)
DEA	0.0723 (0.0099)	0.0959 (0.0074)	0.1594 (0.0213)	0.2972 (0.0294)	0.3965 (0.0331)

Note that standard deviations are in parenthesis.

The correlation coefficients under different noise levels are listed in Table 6.3.4 while the rank correlations are list in Table 6.3.5. In terms of the correlations and the rank correlations, the two parametric methods perform marginally better than the DEA method. The stochastic frontier method shows slight advantage over the deterministic frontier method in terms of the correlation coefficients, however, their performances are very close in terms of the rank correlations. The DEA appears to be a rather competitive alternative to the two parametric methods in terms of the rank correlations especially when the noise level is high.

In general, the performances of all three methods deteriorate noticeably as the noise level increases. When the statistical noise level is relatively low ($\sigma_v < 0.25$), all three methods perform reasonably well, and can be considered as reliable in estimating firm (or observation) specific efficiencies. However, in the presence of high noise levels, the results from any one of the three methods would need careful examination to prevent misleading results.

Table 6.3.4
The Effects of Noise
Correlation Between The True and Estimated Efficiency

	$\sigma_v=0.03$	$\sigma_v=0.15$	$\sigma_v=0.25$	$\sigma_v=0.50$	$\sigma_v=0.75$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$					
Deterministic	0.9576 (0.0156)	0.8609 (0.0195)	0.7352 (0.0299)	0.4894 (0.0558)	0.3189 (0.0514)
Stochastic	0.9786 (0.0051)	0.8900 (0.0128)	0.7943 (0.0278)	0.5724 (0.0562)	0.3985 (0.0599)
DEA	0.9156 (0.0355)	0.7818 (0.0451)	0.6422 (0.0631)	0.4018 (0.0755)	0.2547 (0.0707)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$					
Deterministic	0.9642 (0.0157)	0.8644 (0.0193)	0.7383 (0.0288)	0.4917 (0.0552)	0.3199 (0.0516)
Stochastic	0.9853 (0.0045)	0.8935 (0.0128)	0.7968 (0.0267)	0.5737 (0.0545)	0.3997 (0.0594)
DEA	0.9136 (0.0321)	0.7884 (0.0412)	0.6522 (0.0614)	0.4133 (0.0735)	0.2617 (0.0704)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$					
Deterministic	0.8769 (0.0235)	0.7989 (0.0234)	0.6977 (0.0312)	0.4736 (0.0571)	0.3129 (0.0487)
Stochastic	0.9088 (0.0165)	0.8453 (0.0178)	0.7638 (0.0302)	0.5582 (0.0582)	0.3926 (0.0610)
DEA	0.9210 (0.0319)	0.8080 (0.0384)	0.6685 (0.0560)	0.4285 (0.0768)	0.2710 (0.0704)

Note that standard deviations are in parenthesis.

Table 6.3.5
The Effects of Noise
Rank Correlation Between The True and Estimated Efficiency

	$\sigma_v=0.03$	$\sigma_v=0.15$	$\sigma_v=0.25$	$\sigma_v=0.50$	$\sigma_v=0.75$
1. $\sigma_{12}=\sigma_{13}=\sigma_{23}=3.03$					
Deterministic	0.9579 (0.0171)	0.8714 (0.0223)	0.7680 (0.0367)	0.5460 (0.0660)	0.3756 (0.0587)
Stochastic	0.9750 (0.0058)	0.8764 (0.0168)	0.7707 (0.0362)	0.5471 (0.0656)	0.3756 (0.0582)
DEA	0.9006 (0.0391)	0.7889 (0.0427)	0.6817 (0.0575)	0.4748 (0.0735)	0.3304 (0.0636)
2. $\sigma_{12}=\sigma_{13}=\sigma_{23}=1.33$					
Deterministic	0.9630 (0.0167)	0.8742 (0.0220)	0.7703 (0.0357)	0.5472 (0.0650)	0.3768 (0.0585)
Stochastic	0.9824 (0.0056)	0.8804 (0.0171)	0.7736 (0.0350)	0.5481 (0.0644)	0.3767 (0.0579)
DEA	0.8954 (0.0347)	0.7855 (0.0395)	0.6798 (0.0564)	0.4767 (0.0720)	0.3324 (0.0625)
3. $\sigma_{12}=\sigma_{13}=\sigma_{23}=0.33$					
Deterministic	0.8991 (0.0206)	0.8290 (0.0266)	0.7372 (0.0399)	0.5336 (0.0665)	0.3700 (0.0594)
Stochastic	0.8995 (0.0188)	0.8284 (0.0222)	0.7383 (0.0400)	0.5339 (0.0665)	0.3701 (0.0589)
DEA	0.9038 (0.0373)	0.8015 (0.0385)	0.6918 (0.0515)	0.4887 (0.0703)	0.3411 (0.0623)

Note that standard deviations are in parenthesis.

6.4 The Effects of Exogenous Variables

In the presence of exogenous variables, the observed productive performance reflects the combined outcome of both efficiency and environments. The usefulness of a particular efficiency measurement method depends on its ability to distinguish the effects of environments from the effects of efficiency.

This section examines the relative performance of the two-step DEA-TOBIT procedure and the parametric methods in measuring efficiency in the presence of exogenous variables. The exogenous variable is generated as αz where z is from $N(0,1)$. In this set of experiments, α is given the values of 0.05, 0.10, 0.25, 0.50, 0.75, 1.0, and 2.0. Some of these cases, such as when $\alpha = 1, 2$, are probably not realistic, they are included here just to show how dramatically the results might be affected. The statistics of the exogenous variable are given in Table 6.4.1.

Table 6.4.1 The Effects of Exogenous Variables The Statistics of the Exogenous Variables							
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1.00$	$\alpha=2.00$
Mean of $e^{\alpha z}$	1.003	1.005	1.035	1.238	1.320	1.615	7.658
St Dev $e^{\alpha z}$	0.050	0.101	0.262	0.613	1.136	2.036	34.68
Max. $e^{\alpha z}$	1.148	1.316	2.000	4.203	8.245	17.59	447.9
Min $e^{\alpha z}$	0.873	0.749	0.501	0.247	0.127	0.062	0.004

First, we look at the one-step parametric methods in comparison with the DEA-Tobit procedure. The mean efficiency estimates are presented in Table 6.4.2 and the mean

Table 6.4.2 The Effects of Exogenous Variables The Means of Efficiency Estimates							
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
1. $\sigma_y=3.03$							
True Mean	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)
Determ	0.5006 (0.031)	0.4974 (0.029)	0.4946 (0.026)	0.4990 (0.030)	0.4959 (0.028)	0.4960 (0.026)	0.5004 (0.027)
Stoch	0.6563 (0.025)	0.6556 (0.024)	0.6546 (0.023)	0.6570 (0.023)	0.6551 (0.024)	0.6561 (0.022)	0.6575 (0.022)
DEA1	0.6194 (0.020)	0.6004 (0.021)	0.5317 (0.029)	0.3850 (0.034)	0.2864 (0.036)	0.2181 (0.040)	0.1026 (0.026)
DEA2	0.5751 (0.040)	0.5198 (0.046)	0.3715 (0.050)	0.1757 (0.044)	0.0889 (0.033)	0.0958 (0.022)	0.0488 (0.094)
2. $\sigma_y=1.33$							
True Mean	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)
Determ	0.5029 (0.029)	0.5040 (0.028)	0.5008 (0.025)	0.5046 (0.029)	0.5032 (0.027)	0.5030 (0.025)	0.5049 (0.026)
Stoch	0.6556 (0.024)	0.6552 (0.023)	0.6546 (0.023)	0.6572 (0.022)	0.6547 (0.024)	0.6565 (0.022)	0.6576 (0.022)
DEA1	0.6449 (0.020)	0.6268 (0.022)	0.5560 (0.029)	0.4028 (0.034)	0.2977 (0.034)	0.2243 (0.041)	0.1033 (0.027)
DEA2	0.5971 (0.040)	0.5389 (0.044)	0.3882 (0.054)	0.1841 (0.046)	0.0926 (0.035)	0.0962 (0.021)	0.0495 (0.010)
3. $\sigma_y=0.33$							
True Mean	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)
Determ	0.4145 (0.071)	0.4140 (0.072)	0.4138 (0.072)	0.4166 (0.072)	0.4132 (0.070)	0.4132 (0.070)	0.4169 (0.071)
Stoch	0.6603 (0.031)	0.6604 (0.030)	0.6589 (0.028)	0.6604 (0.031)	0.6596 (0.030)	0.6574 (0.026)	0.6583 (0.026)
DEA1	0.6538 (0.023)	0.6353 (0.022)	0.5667 (0.035)	0.4066 (0.037)	0.2989 (0.031)	0.2174 (0.043)	0.1009 (0.027)
DEA2	0.6058 (0.040)	0.5496 (0.043)	0.3965 (0.055)	0.1832 (0.044)	0.0918 (0.034)	0.0917 (0.018)	0.0473 (0.010)

Note that the standard deviations are in the parenthesis.

DEA1 -- Gross DEA efficiency measures

DEA2 -- Residual DEA efficiency measures from DEA-TOBIT

Table 6.4.3 The Effects of Exogenous Variables The Mean Absolute Deviations							
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
1. $\sigma_y=3.03$							
Determ	0.1612 (0.029)	0.1639 (0.028)	0.1662 (0.027)	0.1625 (0.028)	0.1650 (0.026)	0.1647 (0.027)	0.1613 (0.027)
Stoch	0.0760 (0.005)	0.0761 (0.004)	0.0757 (0.004)	0.0755 (0.004)	0.0760 (0.005)	0.0764 (0.004)	0.0757 (0.005)
DEA1	0.1126 (0.011)	0.1300 (0.014)	0.1952 (0.023)	0.3302 (0.030)	0.4245 (0.033)	0.4874 (0.034)	0.5940 (0.026)
DEA2	0.1293 (0.022)	0.1680 (0.036)	0.2974 (0.052)	0.4861 (0.042)	0.5733 (0.040)	0.6112 (0.036)	0.6479 (0.020)
2. $\sigma_y=1.33$							
Determ	0.1588 (0.027)	0.1578 (0.028)	0.1605 (0.027)	0.1576 (0.028)	0.1586 (0.026)	0.1584 (0.025)	0.1572 (0.027)
Stoch	0.0751 (0.005)	0.0749 (0.004)	0.0749 (0.004)	0.0747 (0.004)	0.0750 (0.005)	0.0753 (0.004)	0.0746 (0.005)
DEA1	0.1068 (0.009)	0.1212 (0.011)	0.1823 (0.020)	0.3175 (0.029)	0.4154 (0.031)	0.4824 (0.034)	0.5930 (0.027)
DEA2	0.1182 (0.018)	0.1536 (0.032)	0.2813 (0.056)	0.4777 (0.043)	0.5697 (0.041)	0.6100 (0.034)	0.6502 (0.014)
3. $\sigma_y=0.33$							
Determ	0.2458 (0.067)	0.2463 (0.068)	0.2466 (0.067)	0.2438 (0.068)	0.2471 (0.065)	0.2468 (0.067)	0.2435 (0.068)
Stoch	0.0911 (0.007)	0.0909 (0.007)	0.0901 (0.006)	0.0911 (0.001)	0.0907 (0.007)	0.0907 (0.006)	0.0907 (0.007)
DEA1	0.1000 (0.008)	0.1123 (0.009)	0.1746 (0.020)	0.3107 (0.029)	0.4107 (0.030)	0.4846 (0.035)	0.5935 (0.028)
DEA2	0.1088 (0.016)	0.1418 (0.030)	0.2728 (0.057)	0.4786 (0.042)	0.5704 (0.041)	0.6126 (0.030)	0.6503 (0.014)

Note that the standard deviations are in the parenthesis.

DEA1 -- Gross DEA efficiency measures

DEA2 -- Residual DEA efficiency measures from DEA-TOBIT

absolute deviations are presented in Table 6.4.3. Recall from Chapter 4, the DEA-TOBIT procedure first estimates a gross efficiency index using the DEA method, and then uses the TOBIT regression to identify the effects of exogenous variables on this gross efficiency index and to compute a residual efficiency index. The gross efficiency estimates and the MADs from the first stage are denoted by *DEA1* in Table 6.4.2 and Table 6.4.3, and the residual efficiency estimates and the MADs from the second stage are denoted by *DEA2*. *Determ* and *Stoch* denote the one-step deterministic and stochastic frontier methods, respectively. In particular, the one-step procedure estimates the production frontier function specified by equation (6.1), and the influence of exogenous variables are controlled for while measuring efficiency.

The results in these two tables draw essentially the same conclusion. They indicate that the magnitude of exogenous variables does not have any significant effects on the efficiency estimates by the two parametric methods as long as the exogenous variables are correctly identified and are accounted for in the estimation. In such cases, there is no misspecification related to those identifiable exogenous variables, hence, they do not contribute to additional disturbance in measuring efficiency. On the other hand, the mean gross efficiency estimates from the DEA-Tobit procedure (*DEA1*) fall dramatically as the magnitude of exogenous variable increases especially when $\alpha > 0.25$. The residual efficiency estimates from the DEA-TOBIT procedure (*DEA2*) is even more underestimated than the gross DEA efficiency measures (*DEA1*). This could partly be explained by the fact that in the second stage of the DEA-TOBIT procedure, only the observation with the largest positive residual from the TOBIT regression is considered as efficient, and consequently

average efficiency estimates would be lower. However, the DEA-Tobit procedure performs slightly better than the deterministic frontier when the magnitude of exogenous variables is low ($\alpha < 0.10$).

Table 6.4.4 lists the correlation coefficients and Table 6.4.5 lists the rank correlations for different magnitudes of exogenous variables. For the two parametric methods, there is essentially no change in the correlations and the rank correlations over the range of α examined. That is, the exogenous variables have little effects on the efficiency estimates by the two parametric methods. However, the performance of the DEA deteriorates rapidly in terms of the correlations and the rank correlations as α increases, especially when α is larger than 0.25. The second stage regression does not appear to be effective in improving the performance of the DEA in terms of correlation coefficient when α is over 0.75, but it does improve the DEA's performance considerably in terms of rank correlation. When the magnitude of exogenous variables is modest, the residual efficiency estimates from the second stage of the DEA-TOBIT procedure do approximate the "true" efficiencies better than the gross DEA efficiency estimates. Overall, the DEA-Tobit procedure performs reasonably well in terms of the rank correlations except for the cases where the magnitude of exogenous variables is set unrealistically high ($\alpha = 1, 2$).

In summary, if the exogenous variables can be correctly identified and incorporated in estimating the production frontiers, the parametric methods have a natural advantage over the DEA-TOBIT procedure in dealing with the effects of exogenous variables. The DEA-TOBIT procedure appears to be a reasonable competitor when the effects of the exogenous variable are modest. When the magnitude of exogenous variables is high, the DEA-TOBIT

Table 6.4.4 The Effects of Exogenous Variables Correlation Between The True and Estimated Efficiency							
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
1. $\sigma_{\hat{y}}=3.03$							
Determ	0.8586 (0.019)	0.8582 (0.020)	0.8585 (0.019)	0.8587 (0.020)	0.8584 (0.019)	0.8589 (0.019)	0.8578 (0.021)
Stoch	0.8884 (0.015)	0.8877 (0.015)	0.8885 (0.014)	0.8888 (0.013)	0.8884 (0.014)	0.8880 (0.013)	0.8889 (0.014)
DEA1	0.7701 (0.046)	0.7348 (0.053)	0.5938 (0.051)	0.3705 (0.073)	0.2433 (0.062)	0.1812 (0.064)	0.0551 (0.084)
DEA2	0.7731 (0.046)	0.7527 (0.055)	0.6597 (0.065)	0.4051 (0.108)	0.2027 (0.102)	0.1677 (0.087)	0.0277 (0.078)
2. $\sigma_{\hat{y}}=1.33$							
Determ	0.8621 (0.019)	0.8617 (0.020)	0.8619 (0.019)	0.8622 (0.019)	0.8619 (0.019)	0.8623 (0.019)	0.8611 (0.021)
Stoch	0.8911 (0.016)	0.8907 (0.015)	0.8912 (0.015)	0.8908 (0.013)	0.8915 (0.014)	0.8917 (0.013)	0.8926 (0.014)
DEA1	0.7774 (0.043)	0.7432 (0.049)	0.6064 (0.047)	0.3816 (0.069)	0.2510 (0.061)	0.1864 (0.063)	0.0573 (0.083)
DEA2	0.7801 (0.043)	0.7610 (0.051)	0.6762 (0.058)	0.4239 (0.105)	0.2135 (0.105)	0.1723 (0.087)	0.0298 (0.076)
3. $\sigma_{\hat{y}}=0.33$							
Determ	0.7967 (0.023)	0.7959 (0.023)	0.7966 (0.023)	0.7969 (0.024)	0.7969 (0.022)	0.7975 (0.023)	0.7966 (0.023)
Stoch	0.8432 (0.021)	0.8436 (0.021)	0.8437 (0.020)	0.8438 (0.020)	0.8435 (0.021)	0.8439 (0.018)	0.8436 (0.018)
DEA1	0.7975 (0.039)	0.7655 (0.045)	0.6229 (0.043)	0.3999 (0.061)	0.2608 (0.061)	0.1899 (0.064)	0.0607 (0.083)
DEA2	0.8004 (0.039)	0.7837 (0.046)	0.6999 (0.056)	0.4470 (0.097)	0.2204 (0.112)	0.1685 (0.085)	0.0282 (0.075)

Note that the standard deviations are in the parenthesis.

DEA1 -- Gross DEA efficiency measure

DEA2 -- Residual efficiency measure from DEA-TOBIT

Table 6.4.5
The Effects of Exogenous Variables
Rank Correlation Between The True and Estimated Efficiency

	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
1. $\sigma_{\hat{\eta}}=3.03$							
Determ	0.8693 (0.022)	0.8688 (0.022)	0.8695 (0.023)	0.8693 (0.022)	0.8697 (0.023)	0.8694 (0.022)	0.8684 (0.024)
Stoch	0.8751 (0.018)	0.8742 (0.019)	0.8747 (0.018)	0.8753 (0.017)	0.8750 (0.019)	0.8745 (0.017)	0.8753 (0.018)
DEA1	0.7808 (0.042)	0.7524 (0.047)	0.6320 (0.042)	0.4383 (0.065)	0.3271 (0.047)	0.2735 (0.064)	0.1507 (0.067)
DEA2	0.7860 (0.042)	0.7775 (0.046)	0.7489 (0.044)	0.6653 (0.073)	0.6012 (0.083)	0.4478 (0.085)	0.2057 (0.078)
2. $\sigma_{\hat{\eta}}=1.33$							
Determ	0.8724 (0.021)	0.8717 (0.022)	0.8724 (0.022)	0.8723 (0.022)	0.8724 (0.022)	0.8720 (0.022)	0.8714 (0.023)
Stoch	0.8776 (0.019)	0.8773 (0.019)	0.8776 (0.019)	0.8772 (0.017)	0.8783 (0.019)	0.8784 (0.017)	0.8795 (0.018)
DEA1	0.7776 (0.041)	0.7495 (0.045)	0.6320 (0.039)	0.4420 (0.063)	0.3295 (0.047)	0.2757 (0.064)	0.1515 (0.066)
DEA2	0.7821 (0.041)	0.7739 (0.045)	0.7477 (0.040)	0.6708 (0.067)	0.6111 (0.082)	0.4606 (0.082)	0.2057 (0.075)
3. $\sigma_{\hat{\eta}}=0.33$							
Determ	0.8270 (0.027)	0.8266 (0.027)	0.8269 (0.027)	0.8267 (0.026)	0.8271 (0.027)	0.8275 (0.025)	0.8260 (0.027)
Stoch	0.8266 (0.024)	0.8268 (0.024)	0.8267 (0.023)	0.8272 (0.024)	0.8271 (0.025)	0.8270 (0.022)	0.8268 (0.023)
DEA1	0.7932 (0.039)	0.7661 (0.043)	0.6418 (0.039)	0.4551 (0.059)	0.3367 (0.048)	0.2827 (0.064)	0.1543 (0.064)
DEA2	0.7982 (0.038)	0.7923 (0.042)	0.7652 (0.040)	0.7099 (0.063)	0.6659 (0.087)	0.4949 (0.071)	0.2137 (0.075)

Note that the standard deviations are in the parenthesis.

DEA1 -- Gross DEA efficiency measure

DEA2 -- Residual efficiency measure from DEA-TOBIT

procedure performs poorly. One explanation for this is that when the effects of exogenous variables are significant, the DEA gross efficiency scores are distorted so seriously¹ that the second stage Tobit analysis could not adequately account for the effects of the exogenous variables. Therefore, if it is necessary to use the DEA-TOBIT procedure in situations where there might be large variations in exogenous factors, one must be very cautious in making any inferences from the results.

The one-step parametric methods clearly have a dominating advantage over the two-step DEA-TOBIT procedure for obvious reasons. The two-step procedure has also been used with the parametric methods. With the two-step parametric procedure, the variations in efficiency measures from the first stage are attributed to variations in the exogenous variables in the second stage (see Kalirajan, 1990 for an example). This two-step parametric method first estimates a frontier production function considering only output and input variables², and then uses the TOBIT regression to identify the effects of exogenous variables and to compute the residual efficiencies as in the second stage of the DEA-TOBIT procedure. This two-step procedure may be more appealing to policy and decision makers who are interested in improving efficiency since it relates the exogenous factors (and other explanatory variables) directly to the efficiency performance. It would be interesting to see how the two-step procedure, when applied to the parametric methods, performs in comparison with the one-step parametric procedure discussed above.

¹ This can be confirmed by the extremely low average DEA scores from the DEA estimation.

² As specified by equation (6.1) after removing the exogenous variables.

For the sake of simplicity, the performance of the two-step parametric methods is examined only under the second production function in previous tables where $\sigma_{ij}=1.33$ ($i \neq j$). This will not bias the comparative results since the results in Table 6.4.2 through Table 6.4.5 indicate that the performance of all three methods in relation to the magnitude of exogenous variables does not depend much on the underlying production structures, at least not in the parameter range considered here.

Table 6.4.6 lists the mean efficiency estimates and Table 6.4.8 lists the MAD values from the two-step methods. In both tables, *DET1* denotes the first stage of the deterministic frontier method, and *DET2* refers to the second stage. Similarly, *STOCH1* (*Stoch2*) denotes the first stage (second stage) of the stochastic frontier method. *DEA1* and *DEA2* are the same as those in Table 6.4.2 and Table 6.4.3. Again, the stochastic frontier method is able to filter out the effects of exogenous variables in the first stage fairly well and yield efficiency estimates very close to their true values as indicated by its small MAD values. The residual efficiency estimates from the second stage of the stochastic frontier method tend to overestimate the mean efficiency level, especially when the effects of exogenous variables are becoming larger. On the other hand, both the DEA and the deterministic frontier method underestimate the mean efficiency level in the first stage, and their abilities to reveal the real picture of efficiencies in the first stage fall substantially as α increases. The incorporation of exogenous variables in the second stage helps improve the efficiency estimates by the deterministic frontier method but not those by the DEA. In terms of the MADs, the DEA method performs better than the deterministic frontier method in the first stage, and also in the second stage when α is small. However, when the magnitude of exogenous variables

Table 6.4.6
The Effects of Exogenous Variables
Mean Efficiency Estimates by Two Stage Procedure

	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
$2.\sigma_{ij}=1.33$ True Mean	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)	0.6561 (0.016)
Det1	0.4989 (0.025)	0.4776 (0.030)	0.3897 (0.039)	0.2432 (0.041)	0.1430 (0.038)	0.1018 (0.037)	0.0251 (0.011)
Det2	0.5030 (0.028)	0.5040 (0.030)	0.4923 (0.037)	0.4679 (0.033)	0.4304 (0.068)	0.3958 (0.070)	0.2719 (0.094)
Stoch1	0.6553 (0.024)	0.6594 (0.027)	0.6674 (0.050)	0.7001 (0.123)	0.6785 (0.144)	0.6783 (0.155)	0.5758 (0.227)
Stoch2	0.6786 (0.036)	0.6804 (0.045)	0.7016 (0.081)	0.8005 (0.113)	0.8341 (0.109)	0.8640 (0.114)	0.8247 (0.179)
DEA1	0.6449 (0.020)	0.6268 (0.022)	0.5560 (0.029)	0.4028 (0.034)	0.2977 (0.034)	0.2243 (0.041)	0.1033 (0.027)
DEA2	0.5971 (0.040)	0.5389 (0.044)	0.3882 (0.054)	0.1841 (0.046)	0.0926 (0.035)	0.0962 (0.021)	0.0495 (0.010)

Note that standard deviations are in the parenthesis.

DET1 -- efficiency estimates from first stage by deterministic frontier method

DET2 -- residual efficiency estimates from second stage by deterministic method

Stoch1 -- efficiency estimates from first stage by stochastic frontier method

Stoch2 -- residual efficiency estimates from second stage by stochastic frontier

DEA1 -- gross DEA efficiency estimates

DEA2 -- residual efficiency estimates from DEA-TOBIT

Table 6.4.7
The Effects of Exogenous Variables
The Mean Absolute Deviations by Two Stage Procedure

	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
$2.\sigma_{\hat{y}}=1.33$							
Det1	0.1627 (0.025)	0.1842 (0.025)	0.2716 (0.042)	0.4175 (0.041)	0.5163 (0.038)	0.5579 (0.038)	0.6344 (0.019)
Det2	0.1587 (0.026)	0.1580 (0.030)	0.1694 (0.037)	0.1930 (0.033)	0.2302 (0.068)	0.2639 (0.068)	0.3875 (0.096)
Stoch1	0.0765 (0.043)	0.0848 (0.077)	0.1166 (0.021)	0.1754 (0.059)	0.1922 (0.050)	0.2056 (0.048)	0.2584 (0.058)
Stoch2	0.0799 (0.010)	0.0843 (0.017)	0.1100 (0.042)	0.1831 (0.087)	0.2094 (0.083)	0.2387 (0.084)	0.2468 (0.083)
DEA1	0.1068 (0.009)	0.1212 (0.011)	0.1823 (0.020)	0.3175 (0.029)	0.4154 (0.031)	0.4824 (0.034)	0.5930 (0.027)
DEA2	0.1182 (0.018)	0.1536 (0.032)	0.2813 (0.056)	0.4777 (0.043)	0.5697 (0.041)	0.6100 (0.034)	0.6502 (0.014)

Note that standard deviations are in the parenthesis.

DET1 -- efficiency estimates from first stage by deterministic frontier method

DET2 -- residual efficiency estimates from second stage by deterministic method

Stoch1 -- efficiency estimates from first stage by stochastic frontier method

Stoch2 -- residual efficiency estimates from second stage by stochastic frontier

DEA1 -- gross DEA efficiency estimates

DEA2 -- residual efficiency estimates from DEA-TOBIT

becomes large, the two-step deterministic frontier method appears to have a slight advantage over the DEA. Comparing the results in Table 6.4.2 and Table 6.4.3 with those in Table 6.4.6 and Table 6.4.7, it is clear that technically it would be desirable to employ the one-step parametric methods since they produce more accurate efficiency estimates.

The performance of the two-step models is also examined in terms of correlation and rank correlation (Table 6.4.8 and Table 6.4.9). The results confirm that the efficiency estimates would not reflect the true picture of the relative efficiency performance of the sample firms if the effects of exogenous variables are ignored. This is indicated by the observation that as α increases the correlations and the rank correlations from the first stage fall dramatically for all three methods. The residual efficiency estimates from the second stage, on the other hand, approximate the true efficiency levels fairly well with the two parametric methods as indicated by the fairly high correlation and rank correlation coefficients. Furthermore, it is noted that although the stochastic frontier method performs better than the deterministic frontier method in the first stage, the opposite is true for the residual efficiency estimates from the second stage when α is large. One explanation for this is that the stochastic frontier method filters out some of the effects of exogenous variables as noise in the first stage so that the first stage performs better. However, this leaves it with less explanatory power for the exogenous variables in the second stage regression, thus results in the less satisfactory performance in the second stage.

The DEA-TOBIT procedure does not do as well as the parametric methods over the values of α considered. However, it performs reasonably well in terms of rank correlation when the magnitude of exogenous variables is small, i.e. $\alpha \leq 0.25$.

Table 6.4.8 The Effects of Exogenous Variables Correlation Coefficients by Two Stage Procedure							
	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
$2.\sigma_{\hat{y}}=1.33$							
Det1	0.8556 (0.019)	0.8268 (0.022)	0.6867 (0.038)	0.4546 (0.037)	0.3026 (0.045)	0.2302 (0.052)	0.0560 (0.073)
Det2	0.8620 (0.018)	0.8616 (0.019)	0.8519 (0.023)	0.8202 (0.034)	0.7880 (0.049)	0.7537 (0.068)	0.5606 (0.127)
Stoch1	0.8867 (0.014)	0.8620 (0.020)	0.7440 (0.032)	0.5418 (0.046)	0.4014 (0.053)	0.3325 (0.062)	0.1744 (0.071)
Stoch2	0.8909 (0.014)	0.8848 (0.018)	0.8599 (0.016)	0.8090 (0.038)	0.7656 (0.042)	0.7351 (0.066)	0.5524 (0.119)
DEA1	0.7774 (0.043)	0.7432 (0.049)	0.6064 (0.047)	0.3816 (0.069)	0.2510 (0.061)	0.1864 (0.063)	0.0573 (0.083)
DEA2	0.7801 (0.043)	0.7610 (0.051)	0.6762 (0.058)	0.4239 (0.105)	0.2135 (0.105)	0.1723 (0.087)	0.0298 (0.076)

Note that standard deviations are in the parenthesis.

DET1 -- efficiency estimates from first stage by deterministic frontier method

DET2 -- residual efficiency estimates from second stage by deterministic method

Stoch1 -- efficiency estimates from first stage by stochastic frontier method

Stoch2 -- residual efficiency estimates from second stage by stochastic frontier

DEA1 -- gross DEA efficiency estimates

DEA2 -- residual efficiency estimates from DEA-TOBIT

Table 6.4.9
The Effects of Exogenous Variables
Rank Correlation by Two Stage Procedure

	$\alpha=0.05$	$\alpha=0.10$	$\alpha=0.25$	$\alpha=0.50$	$\alpha=0.75$	$\alpha=1$	$\alpha=2$
$2.\sigma_{ij}=1.33$							
Det1	0.8669 (0.022)	0.8414 (0.028)	0.7175 (0.037)	0.5072 (0.047)	0.3735 (0.050)	0.3139 (0.064)	0.1728 (0.063)
Det2	0.8721 (0.021)	0.8724 (0.022)	0.8650 (0.027)	0.8362 (0.034)	0.8151 (0.031)	0.7851 (0.062)	0.6518 (0.068)
Stoch1	0.8728 (0.018)	0.8450 (0.024)	0.7177 (0.035)	0.5076 (0.047)	0.3738 (0.049)	0.3138 (0.063)	0.1732 (0.063)
Stoch2	0.8774 (0.017)	0.8716 (0.020)	0.8507 (0.020)	0.8074 (0.040)	0.7690 (0.034)	0.7321 (0.070)	0.5644 (0.105)
DEA1	0.7776 (0.041)	0.7495 (0.045)	0.6320 (0.039)	0.4420 (0.063)	0.3295 (0.047)	0.2757 (0.064)	0.1515 (0.066)
DEA2	0.7821 (0.041)	0.7739 (0.045)	0.7477 (0.040)	0.6708 (0.067)	0.6111 (0.082)	0.4606 (0.082)	0.2057 (0.075)

Note that standard deviations are in the parenthesis.

DET1 -- efficiency estimates from first stage by deterministic frontier method

DET2 -- residual efficiency estimates from second stage by deterministic method

Stoch1 -- efficiency estimates from first stage by stochastic frontier method

Stoch2 -- residual efficiency estimates from second stage by stochastic frontier

DEA1 -- gross DEA efficiency estimates

DEA2 -- residual efficiency estimates from DEA-TOBIT

Comparing the residual efficiency estimates from the second stage with those from the one-step procedure reported in Table 6.4.2 to Table 6.4.5, it is clear that the two-step procedure is inferior to the one-step procedure except for the cases where α is small. Since the two-step procedure relates the exogenous variables directly to the observed efficiency performance of the sample firms, it sometimes might be more helpful to policy analyst and decision makers in identifying the effects of exogenous variables on productive performance. If the effects of exogenous variables are not very strong, it may be desirable to employ the two-step procedure because of its intuitive appeal.

In summary, the parametric methods have an obvious advantage over the DEA method in terms of correlations and rank correlations. However, the mean efficiency estimates from the DEA are very close to the "true" efficiency level when the effects of exogenous variables are not very strong. The abilities of the one-step parametric methods do not appear to be affected by the presence of exogenous variables as long as these variables can be correctly identified and accounted for. Omission and misspecification of the exogenous variables, however, would seriously affect the accuracy of the efficiency estimates. The two-step procedure is able to yield reliable efficiency estimates when the effects of exogenous variables are modest, but it is not reliable for cases where the magnitude of exogenous variables is high. However, they are logically and intuitively more appealing for policy analysis and decision making. Before closing this section, one must note that some of the scenarios considered here are not realistic as mentioned in the beginning of this section. Therefore, one should be careful in referring to the poor performance of certain method when α is large which is intended to illustrate the significance

of the effects of exogenous variables.

6.5 The Effects of Outliers

This section examines the effects of outliers on the efficiency estimates. As stated in Chapter 5, only one outlier is considered to minimize computational requirements. The "efficient" outlier is generated by increasing the output level of an efficient observation³. The efficiency levels are then reestimated in order to see to what degree the outliers make the efficiency estimates deviate from the true efficiency level. In particular, the output level of an efficient observation is increased by 5%, 25%, 50%, 75%, and 100%.

The results from earlier sets of experiments indicate that the relative performance of the three methods exhibits similar pattern with the three underlying production functions

Table 6.5.1 The Effects of Outliers The Mean Efficiency Estimates					
	5%	25%	50%	75%	100%
True Mean	0.6514 (0.0157)	0.6514 (0.0157)	0.6514 (0.0157)	0.6514 (0.0157)	0.6514 (0.0157)
Deterministic	0.4815 (0.0259)	0.4086 (0.0215)	0.3444 (0.0188)	0.2981 (0.0176)	0.2632 (0.0171)
Stochastic	0.6582 (0.0237)	0.6686 (0.0225)	0.6790 (0.0230)	0.6881 (0.0242)	0.6964 (0.0256)
DEA	0.6494 (0.0187)	0.6441 (0.0184)	0.6344 (0.0220)	0.6225 (0.0294)	0.6101 (0.0375)

Note that the standard deviations are in the parenthesis.

³ The "efficient" outlier could in fact be inefficient. However, this should not affect the results dramatically.

considered. Therefore, this section considers only the case where $\rho = -0.25$, that is, $\sigma_{12} = \sigma_{13} = \sigma_{23} = 1.33$ (the second production function).

Table 6.5.2 The Effects of Outliers The Mean Absolute Deviations					
	5%	25%	50%	75%	100%
Deterministic	0.1584 (0.0288)	0.1628 (0.0362)	0.1725 (0.0470)	0.1886 (0.0608)	0.2103 (0.0749)
Stochastic	0.0744 (0.0048)	0.0753 (0.0069)	0.0790 (0.0093)	0.0840 (0.0115)	0.0897 (0.0137)
DEA	0.1027 (0.0083)	0.1046 (0.0093)	0.1099 (0.0141)	0.1179 (0.0214)	0.1265 (0.0291)

Note that the standard deviations are in the parenthesis.

The mean efficiency estimates are reported in Table 6.5.1 and the MAD values are reported in Table 6.5.2. The results show that the efficiency estimates by the DEA are relatively stable, changing from 0.649 at 5% to 0.610 at 100%. However, the mean efficiency estimates by the deterministic frontier method fall sharply from 0.482 at 5% to 0.263 at 100% as the scale of outliers rises. This seems to be contradictory to the presumption that the DEA is more sensitive to outliers as stated by many authors in reviewing the various methodologies for efficiency measurements. However, these results are justifiable when we examine how each of the three methods compute efficiency measures. The deterministic frontier method evaluates the efficiency levels based on ONE single "efficient" observation. If this observation is an (efficient) outlier or there is data error, e.g. inflated output level or deflated input level, associated with this observation, the

efficiency estimates for all other observations would be reduced and the mean efficiency estimates would be lower. The more inflated the output level, the lower the mean efficiency estimates. On the other hand, the DEA method evaluates the efficiencies of the firms or observations based on more than one "best practice" points. If one of the "best practice" points happens to be an outlier or has data error, only those observations which are evaluated based on this particular point will be affected. It would have little effects on the efficiency estimates of other observations. Therefore, the DEA method is more flexible, in terms of the mean efficiency estimates, with regards to outliers and data errors than the deterministic frontier method. This result confirms Mensah and Li (1993)'s conclusion that the DEA is relatively more impervious to deletion of outliers than a deterministic frontier translog model. The efficiency estimates by stochastic frontier method are, as expected, rather stable because of its "built in" ability to buffer exogenous shocks.

Both the DEA and the deterministic frontier method underestimate the efficiency levels because the outliers inflate the output level of the reference point with which the productive performance of other observations is compared. The stochastic frontier method appears to overestimate the efficiencies, especially when the scale of outliers rises. This is because the estimated stochastic production frontier would have a wider "band" when the scale of outliers is higher, thus the average efficiency estimate would be higher.

The correlation coefficients are presented in Table 6.5.3. The stochastic frontier method is again shown to be rather robust to the outliers. In terms of correlations, the outliers have similar effects on the performances of the DEA and the deterministic frontier method, as indicated by the similar changing patterns of their correlation coefficients.

Table 6.5.3 The Effects of Outliers Correlation between True and Estimated Efficiencies					
	5%	25%	50%	75%	100%
Deterministic	0.8631 (0.0194)	0.8559 (0.0201)	0.8424 (0.0212)	0.8249 (0.0227)	0.8046 (0.0245)
Stochastic	0.8933 (0.0145)	0.8943 (0.0155)	0.8933 (0.0160)	0.8913 (0.0163)	0.8889 (0.0165)
DEA	0.7881 (0.0412)	0.7843 (0.0425)	0.7727 (0.0505)	0.7568 (0.0626)	0.7395 (0.0755)

Table 6.5.4 The Effects of Outliers Rank Correlation between True and Estimated Efficiencies					
	5%	25%	50%	75%	100%
Deterministic	0.8742 (0.0219)	0.8745 (0.0219)	0.8745 (0.0217)	0.8747 (0.0217)	0.8748 (0.0217)
Stochastic	0.8795 (0.0188)	0.8800 (0.0194)	0.8799 (0.0197)	0.8796 (0.0200)	0.8793 (0.0202)
DEA	0.7856 (0.0396)	0.7840 (0.0411)	0.7763 (0.0483)	0.7649 (0.0573)	0.7514 (0.0669)

Note that the standard deviations are in the parenthesis.

The rank correlations are presented in Table 6.5.4. In terms of the rank correlations, the two parametric methods appear to be rather flexible to the scale of outliers, while the performance of the DEA tends to deteriorate as the scale of outliers increases. The explanation for the stochastic frontier method is obvious. However, some discussions on the deterministic frontier method and DEA are warranted. With the deterministic frontier

method, the scale of the outliers would affect the actual levels of efficiency estimates of other observations (relative to the efficient outlier) but would have less effects on their relative rankings since these observations are evaluated on the basis of the same reference observation - the efficient outlier in this case. This partly explains the higher rank correlations enjoyed by the deterministic frontier method. On the other hand, the efficient outlier would only affect the efficiency estimates of some of the observations with the DEA method, thus relative efficiency rankings of the observations would change. The larger the scale of outliers, the more significant the changes would be, thus the rank correlation for DEA fall noticeably as the scale of outliers rises.

In summary, the stochastic frontier method is rather robust to the problem of outliers in terms of both the mean efficiency estimates (and the MADs) and the rank correlations (and the correlation coefficients). The presence of outliers does not have much effects on the deterministic frontier method in terms of correlations and rank correlations, but does have significant effects in terms of the mean efficiency estimates. On the contrary, outliers have less effects on the DEA in terms of the mean efficiency estimates, but have considerable effects in terms of the relative efficiency rankings.

6.6 The Effects of the Underlying Production Technology

In the previous five sections, the constant returns to scale production technology is assumed in order not to distort the results of experiments by different treatment of returns to scale. In this section, non-constant returns to scale and input complementarity are allowed into the underlying production technology in order to examine how well the three alternative

methods deal with possible misspecification of functional forms of the production technology.

Since the BCC model takes into account of non-constant returns to scale in the DEA context, it is decided to include the BCC model in the comparison together with the CCR model. Ten production functional forms with different degrees of returns to scale and input complementarity are specified, consequently ten sample data sets are generated based on these functions. The functions are labelled 4 to 13 in sequence to the three production functions assumed earlier. The parameter values of these functions are listed in Appendix B. Table 6.6.1 presents the mean efficiency estimates and Table 6.6.2 presents the MAD values from the four alternative methods. The stochastic frontier method is able to yield stable and accurate efficiency estimates across all the functions considered. Therefore, the returns to scale and input complementarity do not appear to have any significant effects on the performance of the stochastic frontier method in terms of the mean efficiency estimates. The efficiency estimates by the deterministic frontier method underestimate the true efficiency level in all the scenarios. The CCR model produces lower average efficiency estimates than the BCC model. This is expected since the CCR ratios reflect the combined effects of technical efficiency and scale effects, and the BCC efficiency ratios are estimated after removing the effects of scale efficiency. Both models tend to underestimate the true efficiency when input complementarity is present. When input complementarity is absent, the BCC model appears to overestimate the true efficiency level. The mean efficiency estimates by both CCR ratio model and BCC model fall dramatically as the returns to scale rise. The results suggest that the BCC is not able to account adequately for the effects of returns to scale.

Table 6.6.1 The Effects of Underlying Production Technology Mean Efficiency Estimates				
	DET	STOCH	CCR	BCC
True	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
4. $\gamma=0.927$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.5048 (0.0279)	0.6557 (0.0232)	0.5894 (0.0251)	0.6465 (0.0216)
5. $\gamma=0.927$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.5045 (0.0280)	0.6556 (0.0229)	0.5968 (0.0223)	0.6515 (0.0190)
6. $\gamma=0.929$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.4611 (0.0437)	0.6558 (0.0286)	0.6478 (0.0235)	0.7221 (0.0190)
7. $\gamma=0.934$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.5046 (0.0280)	0.6557 (0.0231)	0.5892 (0.0252)	0.6460 (0.0216)
8. $\gamma=0.934$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.5047 (0.0279)	0.6556 (0.0231)	0.5886 (0.0252)	0.6452 (0.0216)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.1(Cont.) The Effects of Underlying Production Technology Mean Efficiency Estimates				
	DET	STOCH	CCR	BCC
True	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)	0.6561 (0.0157)
9. $\gamma=1.209$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.5045 (0.0280)	0.6556 (0.0229)	0.4723 (0.0390)	0.5452 (0.0423)
10. $\gamma=1.227$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.5047 (0.0280)	0.6557 (0.0231)	0.4379 (0.0519)	0.5152 (0.0509)
11. $\gamma=1.239$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.4285 (0.0594)	0.6562 (0.0232)	0.6070 (0.0263)	0.6924 (0.0225)
12. $\gamma=1.286$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.5050 (0.0280)	0.6557 (0.0232)	0.3951 (0.0555)	0.4736 (0.0557)
13. $\gamma=1.573$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.5048 (0.0277)	0.6556 (0.0230)	0.2438 (0.0549)	0.3188 (0.0587)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.2 The Effects of Underlying Production Technology The Mean Absolute Deviations				
	DET	STOCH	CCR	BCC
4. $\gamma=0.927$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.1570 (0.0277)	0.0744 (0.0044)	0.1239 (0.0153)	0.1278 (0.0110)
5. $\gamma=0.927$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.1573 (0.0278)	0.0743 (0.0043)	0.1222 (0.0137)	0.1276 (0.0110)
6. $\gamma=0.929$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.1995 (0.0444)	0.0826 (0.0067)	0.1083 (0.0087)	0.1227 (0.0075)
7. $\gamma=0.934$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.1572 (0.0279)	0.0744 (0.0043)	0.1237 (0.0154)	0.1280 (0.0111)
8. $\gamma=0.934$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.1571 (0.0276)	0.0744 (0.0044)	0.1240 (0.0156)	0.1283 (0.0112)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.2(Cont.) The Effects of Underlying Production Technology The Mean Absolute Deviations				
	DET	STOCH	CCR	BCC
9. $\gamma=1.209$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.1573 (0.0278)	0.0743 (0.0043)	0.2239 (0.0360)	0.2016 (0.0328)
10. $\gamma=1.227$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.1572 (0.0278)	0.0744 (0.0043)	0.2544 (0.0506)	0.2278 (0.0476)
11. $\gamma=1.239$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.2315 (0.0565)	0.0855 (0.0057)	0.1304 (0.0142)	0.1248 (0.0092)
12. $\gamma=1.286$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.1569 (0.0275)	0.0744 (0.0044)	0.2959 (0.0545)	0.2666 (0.0526)
13. $\gamma=1.573$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.1570 (0.0273)	0.0744 (0.0044)	0.4443 (0.0549)	0.4154 (0.0567)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

In summary, when inputs are all substitutes, and when there are decreasing returns to scale, the deterministic frontier method produces the most conservative efficiency estimates which underestimate the true efficiency level. However, when the underlying production technology shows increasing returns to scale and there exists complementarity among inputs, the DEA models would be more likely to produce the lowest efficiency estimates. The poor performance of the DEA models in the cases where there are increasing returns to scales is due to the fact that DEA assumes a convex production possibility set, and increasing returns to scale technology is inconsistent with this convexity assumption (Petersen, 1990). Therefore, discretion is necessary in applying the DEA models where there is evidence suggesting increasing returns to scale.

Regarding the correlations in Table 6.6.3 and the rank correlations in Table 6.6.4, we find that the stochastic frontier method dominates the competition among the four models with its high correlation and rank correlation coefficients. The performance of the deterministic frontier method is very close to that of the stochastic frontier especially in terms of rank correlations. The performance, in terms of correlations, of the two parametric methods do not appear to be affected by the returns to scale. However, it is noticed that the correlation and rank correlation coefficients are slightly lower for cases where all inputs are substitutes (function 6 and 11). As for the two DEA models, the CCR model appears to perform better than the BCC model in all cases examined. The performance of the two DEA models deteriorates as the returns to scale increases (for the reasons given in the previous paragraph). Unlike the parametric methods, the two DEA models perform slightly better when input complementarity is absent. In general, the parametric approach clearly has an

Table 6.6.3 The Effects of Underlying Production Technology Correlation Between True and Estimated Efficiencies				
	DET	STOCH	CCR	BCC
4. $\gamma=0.927$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.8644 (0.0192)	0.8927 (0.0145)	0.7951 (0.0478)	0.6785 (0.0427)
5. $\gamma=0.927$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.8644 (0.0192)	0.8928 (0.0147)	0.7621 (0.0498)	0.6802 (0.0454)
6. $\gamma=0.929$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.8327 (0.0175)	0.8672 (0.0177)	0.7734 (0.0399)	0.7227 (0.0359)
7. $\gamma=0.934$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.8643 (0.0192)	0.8928 (0.0146)	0.7662 (0.0479)	0.6780 (0.0429)
8. $\gamma=0.934$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.8644 (0.0192)	0.8927 (0.0146)	0.7660 (0.0480)	0.6774 (0.0431)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.3(Cont.) The Effects of Underlying Production Technology Correlation between True and Estimated Efficiencies				
	DET	STOCH	CCR	BCC
9. $\gamma=1.209$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.8644 (0.0192)	0.8928 (0.0147)	0.6238 (0.0647)	0.5551 (0.0576)
10. $\gamma=1.227$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.8644 (0.0192)	0.8928 (0.0146)	0.5908 (0.0790)	0.5120 (0.0801)
11. $\gamma=1.239$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.8096 (0.0192)	0.8520 (0.0208)	0.7192 (0.0465)	0.6940 (0.0456)
12. $\gamma=1.286$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.8643 (0.0192)	0.8927 (0.0146)	0.5258 (0.0832)	0.4511 (0.0850)
13. $\gamma=1.573$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.8643 (0.0192)	0.8927 (0.0147)	0.2919 (0.0837)	0.2336 (0.0838)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.4 The Effects of Underlying Production Technology Rank Correlation Between True and Estimated Efficiencies				
	DET	STOCH	CCR	BCC
4. $\gamma=0.927$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.8742 (0.0219)	0.8790 (0.0188)	0.7926 (0.0420)	0.6837 (0.0447)
5. $\gamma=0.927$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.8743 (0.0219)	0.8792 (0.0188)	0.7837 (0.0447)	0.6804 (0.0474)
6. $\gamma=0.929$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.8502 (0.0204)	0.8537 (0.0194)	0.7693 (0.0385)	0.6832 (0.0382)
7. $\gamma=0.934$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.8742 (0.0218)	0.8791 (0.0187)	0.7939 (0.0421)	0.6836 (0.0448)
8. $\gamma=0.934$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.8743 (0.0219)	0.8791 (0.0188)	0.7939 (0.0422)	0.6837 (0.0447)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

Table 6.6.4(Cont.) The Effects of Underlying Production Technology Rank Correlation between True and Estimated Efficiencies				
	DET	STOCH	CCR	BCC
9. $\gamma=1.209$ $\sigma_{12}=0.399$ $\sigma_{13}=0.454$ $\sigma_{12}=-0.182$	0.8742 (0.0218)	0.8792 (0.0187)	0.7141 (0.0535)	0.6467 (0.0540)
10. $\gamma=1.227$ $\sigma_{12}=-0.209$ $\sigma_{13}=0.399$ $\sigma_{12}=0.454$	0.8742 (0.0218)	0.8791 (0.0187)	0.7059 (0.0502)	0.6367 (0.0535)
11. $\gamma=1.239$ $\sigma_{12}=0.356$ $\sigma_{13}=0.447$ $\sigma_{12}=0.508$	0.8339 (0.0204)	0.8349 (0.0236)	0.7284 (0.0414)	0.6673 (0.0416)
12. $\gamma=1.286$ $\sigma_{12}=-0.277$ $\sigma_{13}=0.540$ $\sigma_{12}=0.512$	0.8743 (0.0219)	0.8789 (0.0188)	0.6637 (0.0532)	0.6109 (0.0557)
13. $\gamma=1.573$ $\sigma_{12}=-0.091$ $\sigma_{13}=0.399$ $\sigma_{12}=0.453$	0.8742 (0.0219)	0.8790 (0.0188)	0.4747 (0.0597)	0.4650 (0.0606)

- (a) γ and σ_{ij} are the returns to scale and elasticities of substitution as defined by (5.2) and (5.3)
- (b) DET - Deterministic Frontier Method
- (c) STOCH - Stochastic Frontier Method
- (d) BCC - the BCC Model
- (e) CCR - the CCR ratio
- (f) The standard deviations are in the parenthesis

advantage over the DEA method in distinguishing the effects of efficiency from the effects of returns to scale. Both the parametric and the DEA methods are affected by the presence of input complementarity, but the potential effects are in the opposite directions.

Chapter 7

A Summary of Part I

In Part I of this dissertation, three alternative methods for efficiency measurements, namely, the deterministic frontier method, the stochastic frontier method, and the data envelopment analysis method, are first reviewed in relation to the basic definition of productive efficiency. A Monte Carlo study is then carried out to evaluate the relative merits of the three methods in measuring efficiency under certain known conditions. The results are summarized as follows:

- The performance of all three methods, in terms of correlations and rank correlations between the true and the simulated efficiency estimates, improves as the sample size increases. However, the mean efficiency estimates fall when the sample size increases.
- Non-homogeneity in inputs does not appear to have much effect on the performance of the three methods in terms of correlations and rank correlations in cases where the elasticity of input substitution is greater than one. However, in the case of weak input substitution the performance of the two parametric methods, in terms of correlations and rank correlations, fall considerably as the degree of "non-homogeneity" in inputs increases. The mean efficiency estimates by the DEA appear to increase with the variations in input values, but the opposite is true for the deterministic frontier method. The mean efficiency estimates by the stochastic frontier method meanwhile remain relatively stable regardless the variations in input

variables.

- Higher noise level reduces the efficiency estimates by the deterministic frontier method and the DEA, but tends to increase those by the stochastic frontier method. In terms of correlations and rank correlations, the performance of all three alternatives deteriorate sharply as the noise level rises.
- The one-step parametric methods, both deterministic and stochastic, can adequately account for the effects of exogenous variables as long as the exogenous variables are correctly identified. On the other hand, the two-step procedure performs reasonably well when the effects of exogenous variables are modest. However, when the magnitude of exogenous variables is large, the two-step procedure performs poorly with all three methods.
- The experiment results confirm that the stochastic frontier method is rather robust with respects to outliers. In terms of mean efficiency estimates, the DEA appears to be more flexible to outliers than the deterministic frontier method. However, their performances are quite similar in terms of correlations and rank correlations. Although both the DEA method and the deterministic frontier method are affected by the presence of outliers, the extent of these effects is modest.
- The performance of the stochastic frontier method does not depend much on the structure of the underlying production technology over the parameter range considered. The deterministic frontier method appears to perform slightly better in the presence of input complementarity, but its performance does not seem to be affected by the returns to scale. The DEA method performs poorly when there are

increasing returns to scale, and when input complementarity is present.

In general, the stochastic frontier method dominates the competition among the three alternative methods in all aspects examined. The deterministic frontier method performs better than the DEA method in most cases in terms of correlations and rank correlations, while the DEA method often outperforms the deterministic frontier method in terms of mean efficiency estimates and the MAD values. In drawing any conclusions from the Monte Carlo results, one must not ignore the fact that the production data are generated from a parametric production function which is expected to give the DEA method a comparative disadvantage. However, the DEA performs reasonably well in most situations where potential disturbance is not very serious.

It should be noted that the results summarized above, like results from other experimental studies, are based on a specific experimental design. Although it is believed that the experimental design, as described in Chapter 5, is based on reasonable assumptions about possible empirical situations, at least for the transportation industry, there still exists the possibility that the findings might be changed if the experimental design is altered. However, it is clear that the stochastic frontier method has an inherent advantage in measuring efficiency over the DEA and the deterministic frontier method in most cases. Therefore, it would probably be preferred to the other methods if data permit. The estimation of stochastic frontier production functions might encounter difficulty in situations involving multiple outputs, and the estimation of stochastic frontier cost functions would require multi-lateral price index for inputs which may be difficult, if not impossible, to obtain. In such cases, the DEA method would have a competitive advantage since it is

designed to deal with multiple inputs and outputs, and it does not require information on prices. In light of the Monte Carlo results, however, discretion would be necessary in applying the DEA method in situations where there are large variations in the firms' operating environments and in firms' production characteristics, and where there is evidence suggesting possible increasing returns to scale. The following lists some specific guidelines that may be useful to empirical researchers in selecting an appropriate method for the particular case on hand.

- In situations where there is evidence indicating weak input substitution, and where large variations in sample firms' input variables are observed, the DEA method may be a good choice. It is shown to perform well in terms of both rank correlations and the actual level of efficiency estimates. Further, it requires less data than the two parametric methods. However, if the sample size is very large, the computational costs of DEA might be higher compared to the other two methods.
- When there are large differences in the operating environments of the sample firms, the stochastic frontier method is obviously the best choice. However, if the situation involves multiple outputs, and there is no consistent price data available, such as in studies involving cross-country comparisons, the application of the stochastic frontier method may not be practical.
- The deterministic frontier method has the drawbacks of both the DEA method and the stochastic frontier method: it is deterministic and parametric. However, it is the easiest one to apply. If one is mainly interested in the relative efficiency rankings rather than the actual efficiency levels, the deterministic frontier method is probably

sufficient.

- One should avoid the use of the DEA methods as much as possible in situations where there are increasing returns to scale in the production. Although the BCC model is developed to account for the effects of returns to scale, the Monte Carlo results show that the BCC model can not effectively deal with the increasing returns to scale condition. Further, in situations where input complementarity is present, one should also try to use the parametric methods if data permit.
- In dealing with the identifiable exogenous variables, technically it is desirable to use the one-step stochastic frontier method. If the two-step procedure is preferred for its intuitive appeal, the deterministic frontier method is a better choice if one is mainly interested in the relative efficiency rankings, while the stochastic frontier method is more appropriate if the main concern is in the actual level of efficiency estimates. The DEA-TOBIT method is a reasonably good alternative in situations where the effects of exogenous variables are not too strong.

In the Monte Carlo experiments, the efficiency estimates are compared to the known efficiency profile. In empirical situations, it is not possible to have an accurate picture about the true efficiency performance of the sample firms. In Part II of this dissertation, the three alternative methods are applied to two real world data sets, namely, a railways data set, and an international airlines data set. These two data sets depict two fairly different situations. With the railways data, services are mostly provided by highly regulated, nationalized firms thus the effects of exogenous variables are expected to be significant, and it is very likely to have outliers in the sample. With the airline data set, firms operate in a fairly competitive

environment, and they have access to essentially the same technology available even though there is a high degree of diversity in size. The results from the empirical case studies are examined to see if they are consistent with the observations from the Monte Carlo experiments. Moreover, the empirical results would also be useful in providing policy implications for the industries examined.

Part II

Applications in the Transportation Industry

Chapter 8

The Efficiency of Railway Systems in OECD Countries

In this chapter, the three alternative methods are applied to a sample of 19 passenger railways in OECD countries to examine the productive efficiency of the sample railways over the period of 1978-1989. The first section provides an introduction to the problem and the motivation for the study. The second section describes the data and the model specification. The empirical results are reported and discussed in section 3. Section 4 compares the relative merits of the alternative methods in this application, and finally summary and concluding remarks are given in Section 5.

8.1 Introduction

In recent years, the financial and economic performance of the passenger railway system has attracted a great deal of attention due primarily to the mounting subsidies and alleged inefficiencies imbedded in the system. The economic efficiency of railways is believed to be influenced heavily by the degree of government intervention and subsidization (or taxation), and the institutional and regulatory setting within which the railways operate. The productive efficiency measured from observable data is also heavily influenced by the market and operating environments to which the railways are subjected. These include factors largely beyond managerial control such as topography and climate of the region, the extent of development of other transport modes, traffic density, average load, average distance of haul, the economic development stage of the nation, etc.

The differences in policy adopted by different countries provide an excellent

opportunity to investigate the effects of policy choices on the economic efficiency of the industry. Nash (1981) seeks to discover how much of the variation in the performance of Western European railways, measured by market share, traffic trends and support (subsidy) requirements, may be accounted for by government policies. He finds that, in the passenger sector, the significant factors determining both market share and support requirements are the prices charged and the mix of services offered. These decisions regarding prices and services are almost entirely determined by government policy. Schwier, Jones and Pignal (1990) compares the performance and policy environment of regional rail passenger services operated by VIA Rail Canada, British Railways (BR), French National Railways (SNCF) and Amtrak, with respect to the way they organize regional services, frequency and utilization of services, and financial performance of the railways. Perelman and Pestieau (1988) and Compagnie, Gathon and Pestieau (1991) have analyzed the efficiency of railways, with respect to their differences in the operating environments. Gathon and Perelman (1990) and Compagnie, Gathon and Pestieau (1991) introduce an index of regulatory and institutional autonomy to correct for inefficiency caused by a lack of managerial autonomy.

Different methodologies have been employed in studies on the performance of railways. Nash (1981) and Schwier, Jones and Pignal (1990) base their analysis mainly on simple financial and performance indicators. Nash (1985), Jackson (1991), Thompson, Wood and Lures (1991), Jackson (1992), and Thompson and Fraser (1993) examine the performance of railways through the use of some simple productivity measures, such as labour, fuel, and rolling stock productivities. Caves, Christensen and Swanson (1980, 1981) and McGeehan (1993) measure railways' productivities by estimating cost functions.

Freeman, Oum, Tretheway and Waters (1987) uses the multilateral TFP index to examine the performance of the two largest Canadian railways during the period of 1956-1981. Perelman and Pestieau (1988), Deprins and Simar (1989), and Grabowski and Mehdian (1990) apply the deterministic frontier method to measure the efficiency of railways¹. Gathon and Perelman (1990) and Compagnie, Gathon and Pestieau (1991) estimate stochastic frontier functions to evaluate the productive performance of the European railways, and Jha and Singh (1994) estimates zone specific technical efficiency in the Indian railways using a stochastic frontier model. Bookbinder and Qu (1993) compares the performance of seven Class I North-American railroads using the data envelopment analysis method.

Some of these studies have considered the effects of exogenous variables, e.g. average trip length, average load, etc., on measuring the railways' efficiency, such as Perelman and Pestieau (1988), Deprins and Simar (1989), and Gathon and Perelman (1990). However, none of the studies has examined the effects of public subsidies on the efficiency performance of the railways. There has been no study comparing the results from applications of the DEA and the parametric methods to the railway industry. Simar (1992) compares the deterministic frontier method, the stochastic frontier models with panel data and a non-parametric method, the so-called Free Disposal Hull (FDH) method (Deprins, Simar, and Tulkens, 1984), in estimating railways' efficiencies. However, he did not consider the effects of the output attribute variables in applying the non-parametric approach which rated 50 percent of the observations as efficient.

¹ Grabowski and Mehdian (1990) is concerned with revenue efficiency where railroad revenue is used as the output measure.

This chapter attempts to measure productive efficiency of the railway systems of nineteen OECD countries and analyze it in order to identify effects on efficiency of the public subsidy and degree of managerial autonomy. In the process, the study also attempts to compare and reconcile the results of productive efficiency obtained by using the three alternative methods.

8.2 Data and Model Specification

In this study, each observation in the data set, that is, combination of railway and year, is regarded as an individual "decision-making unit" (DMU) in the terminology of the data envelopment analysis. In total, there are 208 DMUs. The sample firms and their characteristics as well as the data sources are presented in Appendix C. More discussions can be found in Oum and Yu (1991) and Oum and Yu (1994).

For the DEA model, each DMU is assumed to produce two outputs: passenger services and freight services. Two alternative sets of output measures are considered: (1) revenue output measures, as measured by passenger kilometres and freight tonne-kilometres, and (2) available output measures, as measured by passenger train kilometres and freight train kilometres. The available output measures indicate the level of capacity supplied, while the revenue output measures indicate the level of output consumed by users and the value they derive from them.

For the parametric methods, there are problems in estimating multiple output production functions due to the presence of more than one dependent variables (outputs). Estimation of the dual cost functions would require data on the railways' input prices, for

which there is no reliable data which is consistent across firms in different countries and over time. Further, information on revenue shares is not available to aggregate passenger kilometres and freight kilometres, or to aggregate the train kilometres in a more sensible way. Therefore, total train kilometres² is used as the output measure with the parametric methods. This variable is certainly a crude measure of the production of a railway, but it offers an aggregate measure of a railway's activity. An attempt is made to account for the potential effects of different service orientations by incorporating the percentage of passenger train kilometres in total train kilometres, and to account for the degree of train utilization by incorporating the average load per train as additional explanatory variables.

For both the DEA model and the parametric methods, seven inputs are considered: (1) labour, (2) energy consumption, (3) ways and structures, (4) materials, (5) the number of passenger cars, (6) the number of freight wagons, and (7) the number of locomotives (see Appendix C for more details).

The observed productive performance of the railways are influenced by variations in the market, operating, institutional and regulatory policy environment, on which railways have very limited control. The effects of these variables should be accounted for in making efficiency comparison among the railways. Table 8.1 lists the definitions of such variables examined in this study.

² Deprins and Simar (1989) also uses total train kilometres as the output measure.

Table 8.1
Definition of Policy and Uncontrollable Variables

Variables	Description
	Uncontrollable Factors
PDENSITY(KM)	Passenger-km per route-km
FDENSITY(KM)	Freight Tonne-km per route-km
PDENSITY(TR)	Passenger Train Density: Passenger train-km per route-km
FDENSITY(TR)	Freight Train Density: Freight train-km per route-km
%PASSENGER	Percentage of passenger train-km in total train-km
TRIP	Average length of passenger trip (km)
HAUL	Average length of Haul of freight traffic (km)
PLOAD	Average passenger load per train
FLOAD	Average freight load per train
TIME	Time Trend
	Controllable Factors
SUBSIDY/C	Percentage of subsidy to total operating costs
%ELECTRIC	Percentage of Electrified Route Miles
AUTONOMY	Degree of managerial autonomy measured by Compagnie, Gathon and Pestieau (1991)*

* Since the information on the degree of management autonomy for the European railways was available only at a single point in time (measured by Compagnie, Gathon and Pestieau, 1991), it was necessary to assume that the railways' autonomy ratings remain unchanged during the study period. In addition, JNR is given the lowest autonomy rating.

The CCR model³ is applied to the data set, using the formulation (4.2.2), and the second stage regression analysis is accomplished using equations (4.3.1) and (4.3.2). Both the deterministic frontier and the deterministic core of the stochastic frontier are specified as follows:

$$\ln y_j = \alpha_0 + \sum_i \alpha_i \ln x_{ij} + \sum_k \beta_k \ln z_{kj} \quad (8.1)$$

where y_j is the output in train kilometres for the j -th DMU, x_{ij} is the i -th input of DMU j , z_{kj} are the variables listed in Table 8.1, and α , β are the coefficients to be estimated. The deterministic frontier function is estimated using the COLS, described by equations (4.1.5) to (4.1.7), and the stochastic frontier function is estimated using the Battese and Coelli procedure as described in section 4.1.2 of Chapter 4.

8.3 The Results

The study proceeds first with the DEA-TOBIT analysis. Then the deterministic frontier method is used to measure the residual efficiencies of the railways, and to identify the effects of the exogenous variables on the efficiency. Finally, the stochastic frontier method is applied to the sample data. The results from the deterministic frontier method and the stochastic frontier method are briefly discussed in comparison with the DEA-TOBIT results.

³ The BCC model is not used here, since the results in Chapter 6 indicate that it does not perform better than the CCR ratio even in cases of non-constant returns to scale.

8.3.1 The DEA Results

With the DEA-TOBIT procedure, we first estimate a gross efficiency index using the CCR ratio model, then use the TOBIT analysis to identify the effects of the exogenous variables on the gross efficiency and to compute a residual efficiency index.

8.3.1.1 The Computation and Comparison of Gross DEA Efficiency Index

The DEA gross efficiency estimates are computed using a computer code slightly modified from the one listed in Appendix A, and executed on a UNIX computer. As mentioned in previous chapters, the LP problem in equation (4.2.2) has to be solved once for each DMU (observation) to compute the DEA gross efficiency ratings.

Table 8.2 presents the gross efficiency estimates from the CCR model. These DEA indices reflect the combined outcome of true managerial and operational efficiency and the effects of constraints imposed by the institutional, regulatory, market and operating environment. Therefore, one cannot make inferences about true managerial and operational efficiency from these indices without accounting for the variations caused by the variables beyond a firm's control.

There are some significant differences between the two sets of DEA gross efficiency indices: the case of using revenue outputs versus the case of using available outputs. NSB (Norway) has one of the lowest DEA ratings when passenger-kilometres and tonne-kilometres are used as outputs, while it has one of the highest DEA ratings when train-kilometres are used as outputs. This may be partly explained by the fact that NSB's (Norway) average loads per train are on the lower end for both passenger and freight trains.

Table 8.2					
DEA Gross Efficiency Index					
		Output Measure			
Railways	Countries	Revenue Outputs (I) Passenger.-Km & Tonne-Km		Available Outputs (II) Passenger Train-Km & Freight Train-Km	
		1978	1989	1978	1989
BR	U.K.	0.89	1.00	1.00	1.00
CFF-SBB	Switzerland	0.69	0.73	0.94	0.89
CFL	Luxembourg	0.93	0.70	0.95	0.76
CH	Greece	0.52	0.50a	0.62	0.44a
CIE	Ireland	-	1.00	-	1.00
CP	Portugal	0.89	1.00	0.72	0.85
DB	Germany	0.63	0.65	0.81	0.91
DSB	Denmark	0.54	0.75	0.66	0.82
FS	Italy	0.82	0.82	0.71	0.76
JNR	Japan	1.00	1.00b	0.96	1.00b
NS	Netherlands	0.88	0.94	1.00	1.00
NSB	Norway	0.74	0.67	1.00	0.94
ÖBB	Austria	0.60	0.62	0.84	0.85
RENFE	Spain	0.76	0.77	0.97	1.00
SJ	Sweden	0.90	1.00	1.00	1.00
SNCB	Belgium	0.61	0.71	0.66	0.73
SNCF	France	0.77	0.84	0.95	0.99
TCDD	Turkey	0.88	0.94	0.91	0.64
VR	Finland	0.79	1.00	0.80	0.96

a. for 1987;

b. for 1986.

Similarly, TCDD (Turkey) has a quite high rating in terms of passenger-kilometres and tonne-kilometres, while it is rated low in terms of train-kilometres. This may be partly explained by the fact that TCDD (Turkey) has relatively higher average loads per train for both freight and passenger services.

Both sets of the DEA gross efficiency indices show that by 1989 BR (U.K.), SJ (Sweden), JNR (Japan), CIE (Ireland), NS (Netherlands) and VR (Finland) have attained a position close to the efficient production frontier. DSB (Denmark) and VR (Finland) have achieved significant improvement during the period: 0.54 and 0.66 in 1978 to 0.75 and 0.82 in 1989 for DSB (Denmark), and 0.79 and 0.80 in 1978 to 1.00 and 0.96 in 1989 for VR (Finland). CH (Greece) and SNCB (Belgium) are indicated as the least efficient railways by both sets of the DEA gross efficiency indices. CFL (Luxembourg) and NSB (Norway) are the two railways which exhibit a noticeable decline in their DEA gross efficiency index during the sample period.

Overall, the two sets of the DEA gross efficiency indices are comparable. The correlation coefficient between the two DEA gross efficiency indices is 0.624 while their Spearman's rank correlation coefficient is 0.615. These DEA gross efficiency results are generally consistent with the findings of Jackson's surveys of European railway performance (Jackson, 1991 and 1992), the world bank surveys of railway performance (Thompson, Wood, and Lures, 1991, and Thompson and Fraser, 1993), and Compagnie, Gathon and Pestieau, 1991.

8.3.1.2 The Effects of Exogenous Variables

As mentioned earlier, the DEA gross efficiency indices are influenced by variations in the market, operating, institutional and regulatory policy environment, variations over which railways have limited control. The effects of these variables are identified by analyzing the DEA gross efficiency index. Subsequently, after controlling for the effects of these variables, the residual efficiency indices are computed and analyzed. This residual efficiency index is a closer indicator of managerial and technical efficiency than the DEA gross efficiency index.

TOBIT regression as specified by equation (4.3.1) is used to identify the effects of policy and other variables, and to measure the "residual" efficiency⁴ of the railways. The variables listed in Table 8.1 are examined in the TOBIT analysis. Some of these variables are considered controllable by government and/or regulatory agencies while others are not. Note that two policy variables are considered: SUBSIDY and AUTONOMY. SUBSIDY is measured by the ratio of subsidy to operating costs. Subsidy policy should ideally be examined according to the types of subsidies and the ways in which they are provided (e.g. loss/balancing subsidy vs. a fixed sum subsidy, unconditional subsidy vs. payment conditional on meeting a certain performance standards, etc.), which are likely to have substantial impacts on a firm's efficiency. However, due to limited information, this study

⁴ We use the term "residual efficiency" because there are some important factors left out in our analysis, such as weather and climate, topography of the land, the general economic condition of the country, the extent of development of other transport modes, and quality of services, etc. The presence of these factors (which are left out mainly due to the lack of data) makes it difficult to interpret the residual efficiency index as an indicator for true managerial and operational efficiency.

examines the effects of subsidization with respect to the level of aggregate subsidy only. The AUTONOMY variable is an index of regulatory and institutional environments, based on 1988-1989 figures, which was constructed by Perelman and Gathon (1990) using the information collected through a survey of railways' management. Its values range between 40 and 100. The more autonomous management is, the higher the value of the AUTONOMY index. The degree of managerial autonomy is affected by a large number of factors including ownership form and managerial mandate. It is therefore very difficult to quantify managerial autonomy consistently across railways even with the best of efforts. Another problem with this variable is that this variable was based on only one year's observations. In using this variable it is therefore assumed that the institutional environment of the railways had varied only minimally over our sample period, an assumption which is not realistic for most railways. Although it is realized that the AUTONOMY variable is not ideally defined, this variable is used in the analysis as it is the only information of its kind which had been collected systematically.

Table 8.3 reports the two best log-linear TOBIT regressions: one with the passenger-kilometres and freight tonne-kilometres as outputs (henceforth referred to as Model I), and the other with the train-kilometres as outputs (henceforth referred to as Model II). The results can be interpreted as follows:

- (a) **Effects of Uncontrollable Variables:** Four of the variables used in the TOBIT model (for both freight and passenger services) are largely beyond managerial or governmental control, i.e. traffic density (DENSITY: passenger and freight), average distance of the trip (TRIP or HAUL), load per train (LOAD: passenger and freight)

Table 8.3
Tobit Regression Results

Dependent Variable				
Log (DEA Efficiency Indices)				
	Model (I) Passenger-Km & Tonne-Km as Outputs		Model (II) Passenger and Freight Train-Kms as Outputs	
Independent Variables	Coef.	(t-stat.)	Coef.	(t-stat.)
DENSITY-Freight	-	-	-	-
DENSITY-Pax	-	-	-	-
HAUL	-	-	-	-
TRIP	-	-	-	-
%PASS	-	-	- 0.567	(6.04)
LOAD-Pax	0.26	(7.30)	- 0.062	(1.96)
LOAD-Freight	-	-	- 0.418	(8.21)
%ELECTRIC	0.018	(2.54)	0.043	(7.28)
SUBSIDY	-0.088	(6.27)	-0.052	(5.20)
AUTONOMY	0.452	(6.78)	0.163	(3.55)
TIME	0.008	(2.04)	0.007	(2.71)
CONSTANT	-3.49	(9.81)	1.524	(4.82)
No.of Observ.	208		208	
LOG-Likelihood	27.8568		86.3713	
R ²	0.4196		0.6067	

Note that all variables are in their logarithms, except for TIME. In this way, the coefficient for the variable TIME indicates the constant growth rate directly.

and percentage of passenger train-kilometres in the total train-kilometres (%PASSENGER).

The DENSITY and TRIP (HAUL) variables turn out to be statistically insignificant in both models. The insignificance of the DENSITY variables may be explained by the fact that: (i) railways with higher passenger traffic density do not necessarily perform well in their freight services, and vice versa; and (ii) substantial correlations exist between traffic density and load per train (0.33 for passenger services and 0.58 for freight services).

Both LOAD variables, Passenger LOAD and Freight LOAD, are statistically significant in the Model (II) while only Passenger LOAD is significant in Model (I). These results indicate that railways with higher average LOADs tend to attain a higher DEA gross efficiency index when passenger-kilometres and freight tonne-kilometres, rather than train-kilometres, are used as the output. These results confirmed our hypotheses that: (i) the cost per train-kilometer increases as the load per train increases; and (ii) the cost per passenger-kilometer (or tonne-kilometer) decreases as the load per train increases.

The %PASSENGER variable has a significant negative coefficient in Model (II) while the same variable is statistically insignificant in Model (I). The negative coefficient in Model (II) can be interpreted as meaning that railways with heavy concentration in passenger services are at a disadvantage when their output is measured in terms of train-kilometres. Gathon and Perelman (1990) draws a similar conclusion.

- (b) **Effects of Electrification:** The %ELECTRIC variable has a statistically significant, positive coefficient in both Models (I) and (II). This indicates that electrification improves the performance of the railways. This may be due to the positive effect electrification has on reducing energy consumption and the amount of labour required. The positive effect on labour productivity is also shown by Gathon and Perelman (1990).
- (c) **Effects of Subsidy:** The ratio of subsidy to the total operating expenses has a statistically significant negative coefficient in both models. This implies that heavily subsidized railways tend to be less efficient than other railways⁵.

The subsidization of a particular firm (or mode) puts its rivals in a weaker competitive position, that is, the subsidized firm is protected from the pressure of competition by the subsidy. As a result, these firms have little incentive to improve their productivities or minimize costs, thus becoming less innovative and efficient. The extent of those potential subsidy effects depends on how the subsidies are provided. However, as stated earlier, our analysis does not distinguish amongst the types of subsidies and how they are provided due to the lack of information required

⁵ At first glance, the small coefficients for the SUBSIDY variable appear to indicate that the negative effect of subsidy on the efficiency is not that high. However, in terms of the magnitude of the effect on cost efficiency relative to the change in subsidy amount can be relatively large. For example, the Model (1) in Table 8.3 shows that doubling of the subsidy would reduce efficiency of the subsidized railway system (the freight and passenger services together) by 8.8%. Suppose that the total size of a company of our interest is \$2 billion and the current subsidy level is \$100 million. Then, doubling the subsidy to \$200 million would cost the firm about \$176 million (8.8% of \$2 billion) due to the increased inefficiency.

for such an analysis⁶.

- (d) **Effects of Managerial Autonomy:** The managerial autonomy index constructed by Gathon and Perelman (1990) has a statistically significant positive coefficient in both models. This shows that efficiency is expected to improve if management is given more autonomy in making strategic and operational decisions. Higher degrees of managerial autonomy enable management to response quickly to new opportunities and circumstances, thus keeping the firm competitive in the ever changing market. Managerial autonomy on the choice of the markets to serve (or abandon) and frequency of services would heavily influence the productive efficiency of the railway. Moreover, higher degrees of managerial autonomy would require management to be more accountable for the firm's performance, thus giving them greater incentive to improve productive efficiency.
- (e) **Effects of TIME:** The TIME trend variable is intended to assess the overall improvement of the industry's technological progress and managerial efficiency (i.e. the efficiency improvement common across all of the railways in the sample). This variable has a statistically significant positive coefficient, indicating that the industry experienced technological progress at a rate of about 0.7% ~ 0.8% each year during the study period from 1978 to 1989.

Although there are some differences between the two TOBIT models, the overall

⁶ It should be noted that subsidization and inefficiency are in fact inter-dependent. Subsidization causes inefficiency, and inefficiency is very likely to result in larger requirements for subsidy. Since the main interest of this study is the comparative performance of the alternative methods, subsidy is assumed as an exogenous variable.

results are consistent with our expectations. Certainly the differences in the signs of regression coefficients are reconcilable. The regression results indicate that a railway's efficiency performance may be significantly enhanced by an institutional and regulatory framework which would provide railway management with greater freedom in decision making. This is evident from the statistically significant positive coefficients for the AUTONOMY variable and the negative coefficients for SUBSIDY. These results are also consistent with the general belief that a high degree of direct government intervention and high subsidy levels interfere with market mechanisms and encourage inefficiency to remain in railways' operation.

8.3.1.3 The Residual Efficiencies

The statistical significance of some of the variables beyond managerial control implies that these variables do influence the DEA gross efficiency index. This confirms our earlier statement that the DEA gross efficiency index does not reflect true managerial and operational efficiency. The results will be closer to the real level of efficiency only after the effects of these variables are removed from the DEA gross efficiency index.

Residual efficiencies are computed from the residuals of the Tobit regressions. However, the reader is cautioned once again that these residual efficiency indicators may not reflect the true picture of the railways' efficiency performance. The lack of adequate data makes it impossible to control for such factors as quality of service, weather and climate, topography of the land, etc. However, the residual efficiency index is much closer to a railway's true efficiency than the DEA gross efficiency index.

Table 8.4					
Residual Efficiency Index					
		Output Measure			
Railways	Countries	Revenue Outputs (I) Passenger-Km & Tonne-Km		Available Outputs (II) Passenger Train-Km & Freight Train-Km	
		1978	1989	1978	1989
BR	U.K.	0.76	0.79	0.80	0.80
CFF-SBB	Switzerland	0.53	0.55	0.73	0.72
CFL	Luxembourg	1.00	0.72	1.00	0.71
CH	Greece	0.63	0.56a	0.82	0.61a
CIE	Ireland	-	0.84	-	0.77
CP	Portugal	0.73	0.75	0.61	0.76
DB	Germany	0.59	0.55	0.69	0.76
DSB	Denmark	0.63	0.79	0.67	0.77
FS	Italy	0.65	0.64	0.66	0.67
JNR	Japan	0.74	0.71b	0.90	0.89b
NS	Netherlands	0.83	0.80	0.86	0.89
NSB	Norway	0.80	0.67	0.82	0.80
ÖBB	Austria	0.59	0.62	0.70	0.73
RENFE	Spain	0.69	0.69	0.83	0.85
SJ	Sweden	0.73	0.77	0.81	0.83
SNCB	Belgium	0.59	0.64	0.66	0.74
SNCF	France	0.60	0.61	0.80	0.82
TCDD	Turkey	0.68	0.68	0.84	0.64
VR	Finland	0.54	0.78	0.62	0.82

a. for 1987;

b. for 1986.

The residual efficiency indices are listed in Table 8.4. The variation of the residual efficiency index values are smaller than the DEA gross efficiency index values. This is because some of the variations in the DEA gross efficiency indices have been explained by the variables included in the TOBIT regression model. The residual indices from Models (I) and (II) show that in 1989, BR (UK), NS (Netherlands), SJ (Sweden) and VR (Finland) were among the most efficient performers; CH (Greece) and ÖBB (Austria) were among the least efficient performers.

It is noteworthy to observe that the performance of CP (Portugal) and JNR (Japan), which are among the top performers in terms of the DEA gross efficiency index, are reduced to that of the "mid-range" performers when using the residual efficiency. On the other hand, the efficiency rating of DSB (Denmark), using the residual efficiency, is improved as compared to its rating in terms of the DEA gross efficiency index. Furthermore, a greater number of railways join the 'mid-range' performers group when the residual efficiency index is used. Finally, the performance of DSB (Denmark) and SNCB (Belgium) improves over the years, in terms of residual efficiency, while the ratings of CFL (Luxembourg), CH (Greece), NSB (Norway), and TCDD (Turkey) noticeably deteriorate.

8.3.2 The Deterministic Frontier Method Results

In this subsection, the results from the deterministic frontier method is examined. As stated in section 2 of this chapter, the total train-kilometer is used as the output measure. This is because it is difficult to deal with multiple outputs with the parametric production function methods, and there is no consistent price data for estimating the alternative cost

function. The input variables are the same as those used in the DEA analysis.

The estimated "best practice" production function and resulting efficiency estimates from the one-step deterministic frontier method are presented in Table 8.5 and Table 8.6, respectively. Recall from Chapter 6 that the one-step procedure includes the exogenous variables directly in the estimation of the production frontier. Among the policy and uncontrollable variables, TRIP has a statistically significant positive coefficient, implying that the railways with longer average trip distance are expected to enjoy consistently higher efficiency ratings than other railways. As expected, %ELECTRIC has a statistically significant positive coefficient reflecting its positive effects on labour and energy productivities. The two LOAD variables have statistically significant negative coefficients. This is the case because the output is measured in train-kilometres, thus cost per train-kilometer increase as the average load per train goes up. %PASSENGER has a statistically significant negative coefficient. This indicates that railways with larger portion of their business in passenger traffic are likely to have a lower efficiency rating when train kilometers is used as the output measure. SUBSIDY has a significantly negative coefficient implying that government financial support tends to decrease the railways' efficiency. The degree of managerial freedom appears to have significant positive effects on the railways' performance as indicated by its statistically significant positive coefficient. As expected, the TIME trend variable is also statistically significant, indicating that technological progress has improved the overall efficiency performance of the industry. The results in Table 8.5 essentially confirm those from the DEA-TOBIT analysis.

Table 8.6 lists the efficiency estimates from the deterministic frontier method. There

Table 8.5
Deterministic Production Frontier Function

Dependent Variable		
Log (Train-Kilometres)		
Independent Variables	Coef.	(t-stat.)
LABOUR	0.062	(1.11)
ENERGY	0.590	(17.63)
WAYS AND STRUCTURAL	0.186	(12.91)
FREIGHT CAR	0.131	(3.21)
DENSITY-Freight	-	-
DENSITY-Pax	-	-
HAUL	-	-
TRIP	0.258	(11.38)
%PASS	-0.791	(7.33)
LOAD-Pax	-0.232	(6.78)
LOAD-Freight	-0.603	(12.17)
%ELECTRIC	0.116	(15.08)
SUBSIDY	-0.076	(8.55)
AUTONOMY	0.299	(5.51)
TIME	0.001	(3.45)
CONSTANT	3.049	(7.86)
No.of Observ.	208	
LOG-Likelihood	153.904	
R ²	0.9919	

Note that all variables are in their logarithms, except for TIME.

Table 8.6
Efficiency Estimates by Deterministic Frontier

Railways	Countries	Train-Kilometres as Output	
		1978	1989
BR	U.K.	0.71	0.80
CFF-SBB	Switzerland	0.62	0.66
CFL	Luxembourg	0.89	0.64
CH	Greece	0.83	0.74a
CIE	Ireland	-	0.76
CP	Portugal	0.65	0.83
DB	Germany	0.74	0.74
DSB	Denmark	0.73	0.73
FS	Italy	0.63	0.60
JNR	Japan	0.99	0.81b
NS	Netherlands	0.77	0.76
NSB	Norway	0.77	0.74
ÖBB	Austria	0.73	0.70
RENFE	Spain	0.74	0.70
SJ	Sweden	0.86	0.90
SNCB	Belgium	0.74	0.72
SNCF	France	0.86	0.78
TCDD	Turkey	0.52	0.67
VR	Finland	0.54	0.82

a. for 1987;

b. for 1986.

are some noticeable differences between the results in Table 8.6 and those in Table 8.4 from the DEA-TOBIT analysis. For example, the results by deterministic frontier method indicate that the performance of CP (Portugal) is among the top performers in 1989 while the DEA-TOBIT⁷ results put it in the group of "mid-range" performers. In addition, the deterministic frontier method also shows that there is significant improvement in BR's performance in 1989 from 1978, but the DEA-TOBIT results indicate that there is essentially no change in BR's performance. Also, the deterministic frontier method indicates that JNR's performance deteriorates over time while the DEA-TOBIT results show that there is little change in JNR's performance between 1978 and 1986. One possible explanation for the difference is that the deterministic frontier method follows an one-step procedure, while the DEA-TOBIT analysis is a two-step procedure, and the Monte Carlo results show that the two procedures yield different efficiency estimates. Further, the DEA-TOBIT method considers two separate outputs, but the deterministic frontier method consider only one aggregate output. However, both methods indicate that BR, JNR, SJ, and VR are among the most efficient performers, while CFL, FS and TCDD are among the least efficient performers in 1989.

8.3.3 The Stochastic Frontier Results

The results from the application of the stochastic frontier method are reported and examined in this section. As with the deterministic frontier method, total train-kilometer is

⁷ In comparing the results of the parametric methods and DEA, all references to the residual efficiencies are directed to the second set of efficiency index in Table 8.4 using the available output measures so that they are consistent with the output measure used in the parametric methods. .

used as the output measure. Again, seven input variables are considered.

The estimated deterministic core of the stochastic frontier production function, using the one-step procedure, is listed in Table 8.7. As one can see, the estimated frontier function is essentially the same as the one-step deterministic frontier production function presented in Table 8.5 except for the intercept term. This is not unexpected since the estimation procedure employed in the FRONTIER program starts with OLS estimates of the coefficients which are all unbiased except for the intercept term (Coelli, 1991), the procedure then proceeds with a two-phase grid search and an iterative process using a Quasi-Newton method to obtain the final (approximate) maximum-likelihood estimates.

The efficiency estimates by the one-step stochastic frontier method are presented in Table 8.8. The stochastic frontier efficiency estimates are considerably higher than both the residual efficiencies from the DEA-TOBIT procedure (Table 8.4) and the deterministic frontier efficiency estimates (Table 8.6). It is noteworthy to observe that some results in Table 8.8 are noticeably different from the earlier results: CH appears to be among the "upper-middle" group in terms of the stochastic frontier efficiency estimates, while it appears to be among the "mid-range" group in terms of the deterministic frontier efficiency estimates, and among the least efficient performers in terms of the DEA residual efficiency estimates. JNR is among the efficient performers in terms of both the DEA residual efficiencies and the deterministic frontier efficiency estimates, however, it is shown to be among the "mid-range" performer in terms of the stochastic frontier efficiency estimates. CP is ranked among the top performers by the two parametric methods but it is placed among the mid-range group in terms of the DEA residual efficiency index in Table 8.4.

<p>Table 8.7 Stochastic Production Frontier Function</p>		
Dependent Variable		
Log (Train-Kilometres)		
Independent Variables	Coef.	(t-stat.)
LABOUR	0.062	(1.27)
ENERGY	0.602	(17.55)
WAYS AND STRUCTURAL	0.183	(14.60)
FREIGHT CAR	0.111	(3.60)
DENSITY-Freight	-	-
DENSITY-Pax	-	-
HAUL	-	-
TRIP	0.238	(11.82)
%PASS	-0.751	(7.93)
LOAD-Pax	-0.139	(4.63)
LOAD-Freight	-0.508	(11.43)
%ELECTRIC	0.112	(17.33)
SUBSIDY	-0.094	(9.18)
AUTONOMY	0.409	(9.22)
TIME	0.003	(1.39)
CONSTANT	1.911	(5.23)
No.of Observ.	208	
LOG-Likelihood	167.99	

Note that all variables are in their logarithms, except for TIME.

Table 8.8
Efficiency Estimates by Stochastic Frontier

Railways	Countries	Train-Kilometers as Output	
		1978	1989
BR	U.K.	0.84	0.97
CFF-SBB	Switzerland	0.71	0.79
CFL	Luxembourg	0.97	0.78
CH	Greece	0.96	0.91a
CIE	Ireland	-	0.94
CP	Portugal	0.74	0.94
DB	Germany	0.90	0.94
DSB	Denmark	0.88	0.92
FS	Italy	0.73	0.75
JNR	Japan	0.99	0.88b
NS	Netherlands	0.91	0.92
NSB	Norway	0.95	0.95
ÖBB	Austria	0.88	0.92
RENFE	Spain	0.87	0.90
SJ	Sweden	0.97	0.99
SNCB	Belgium	0.87	0.89
SNCF	France	0.96	0.92
TCDD	Turkey	0.55	0.73
VR	Finland	0.56	0.96

a. for 1987;

b. for 1986.

Similar to the results by the other methods, BR, SJ, and VR are again ranked among the most efficient railways in terms of the stochastic frontier efficiency estimates, while CFL, FS, and TCDD are among the least efficient performers in 1989.

8.4 Comparison of the Alternative Efficiency Estimates

In the previous section, comparison of the efficiency estimates from the three methods is focused on the performance of some specific railways. In this section, however, the focus is on comparison of overall pattern of the alternative efficiency estimates.

Table 8.9 Means of the Railways' Efficiency Estimates			
	Deterministic	Stochastic	DEA-TOBIT ¹
Mean	0.7366	0.8729	0.7731
Standard Deviation.	0.0835	0.0943	0.0759
Minimum	0.5247	0.5529	0.5521
Maximum	1.0000	0.9875	1.0000

1. With passenger and freight train kilometres as the output measures.

Table 8.9 lists the mean efficiency estimates by the three alternative methods. The deterministic frontier method produces the most conservative average efficiency estimate, while the stochastic frontier method yields the highest average efficiency estimate. This result is consistent with the results from the Monte Carlo experiments in Chapter 6. Furthermore, the difference between the stochastic frontier mean efficiency estimate and the

mean of the DEA-TOBIT residual efficiencies is rather significant. The Monte Carlo results indicate that the gaps between the alternative efficiency estimates become larger when the effects of the exogenous variables are significant. Therefore, the differences among the three sets of efficiency estimates confirm the initial hypothesis that exogenous factors have substantial effects on the observed productive performance of the railways. However, the ranges of the efficiency estimates by the three methods are quite similar, with the railways

Table 8.10 Correlation and Rank Correlation Coefficients Between the Efficiency Estimates			
Correlation Coefficient			
	Deterministic	Stochastic	DEA-TOBIT
Deterministic	1		
Stochastic	0.8859	1	
DEA-TOBIT	0.5204	0.4150	1
Rank Correlation Coefficient			
Deterministic	1		
Stochastic	0.8996	1	
DEA-TOBIT	0.5486	0.5152	1

ranging from efficient performers to about 45 percent less efficient.

Table 8.10 lists the correlation and rank correlation coefficients among the alternative efficiency estimates. It is noted that the efficiency estimates from the two parametric methods are highly correlated in terms of both the correlation coefficient and the rank

correlation coefficient. However, the residual efficiencies from the DEA-TOBIT are only marginally correlated with the residual efficiency estimates from the other two methods. This happens partly because the parametric methods employ an one-step procedure while the DEA-TOBIT is a two-step procedure, and the two procedures are shown by the Monte Carlo experiments to produce different efficiency estimates. In addition, the parametric methods consider only one output variable, while DEA considers two outputs.

An examination of the efficiency estimates from the two-step parametric methods⁸ shows that there are higher correlations between corresponding efficiency estimates: the rank correlation coefficient is 0.81 between the gross efficiency estimates by the stochastic frontier method and the gross DEA efficiency estimates, and 0.64 between the corresponding residual efficiency estimates; and is 0.83 between the gross efficiency estimates from the deterministic frontier method and the gross DEA efficiency estimates, and 0.71 between the corresponding residual efficiency estimates. This is also consistent with the simulation results in Chapter 6 where the efficiency estimates from the two-step parametric methods are closer to the DEA-TOBIT estimates than those using the one-step procedure.

In general, the basic pattern of the efficiency estimates from the two parametric methods are very similar, although there are noticeable differences in the actual levels of estimated efficiency between the two sets of efficiency estimates. On the other hand, the efficiency estimates from the DEA-TOBIT analysis exhibit some significant differences from those obtained from the parametric methods. The main reason for this appears to be due to

⁸ The results from the two-step parametric methods are not reported in this chapter, they are available from the author upon request.

the uses of one-step vs two-step procedure, and the use of different output measures. However, the policy implications from all three methods are consistent.

8.5 Summary and Concluding Remarks

This chapter attempts to identify the effects of government intervention and subsidization on productive efficiency of those railways which derive a high proportion of their business from passenger services. In particular, the productive efficiency of the railway systems in 19 OECD countries are measured and analyzed in order to identify the effects of both public subsidies and degree of managerial autonomy on efficiency. A two-step DEA-TOBIT procedure is used along with two one-step parametric methods: the deterministic frontier method and the stochastic frontier method. With the DEA-TOBIT procedure, the Data Envelopment Analysis (DEA) method is first used to measure the gross efficiency index from the panel data of the 19 railways over the 1978-89 period, and then the Tobit regression is used to identify the effects of the public subsidies and the extent of managerial autonomy after controlling for the effects of various operating characteristics and market environments such as traffic density, average load per train, average distance hauled and percentage of electrified route network, which are largely beyond managerial control. With the parametric methods, the policy and uncontrollable variables are incorporated directly in estimating the frontier production functions.

The empirical results from all three methods show that railway systems with high dependence on public subsidies are significantly less efficient than similar railways with less dependence on subsidies, and railways with high degree of managerial autonomy from

regulatory authority tend to achieve higher efficiency. These two results together imply that productive efficiency of railway systems may be significantly enhanced by an institutional and regulatory framework which provides a greater freedom for managerial decision making. Therefore, the institutional and regulatory framework for railway industry must squarely address the question of railways' managerial autonomy. Subsidy policies must encourage railways to use normal market mechanisms to improve their cost recovery and to use the subsidies only for improving services.

In addition, the empirical results indicate that efficiency measures may not be meaningfully compared across railways without controlling for the differences in operating and market environments.

Due to limited information and knowledge, this study examines the effects of government intervention and subsidization in a broad sense only. There are many alternative and complementary means by which governments intervene in railway business. These include a partial or full ownership by a government, operation of railways by a government branch, regulation of prices and service frequencies, and use of taxation and subsidy. Each of these means of intervention has a distinct effect on the firm's efficiency. Also, not only the amount of subsidy but also the method of subsidization affect management's incentive to improve efficiency. For example, a predetermined amount of subsidy is considered to be more effective than subsidizing 100% of the financial loss. Therefore, an informed policy making would require detailed studies for measuring the effects of various means of government intervention and alternative methods of subsidization on economic efficiency.

In comparing the efficiency estimates from the three alternative methods, it is found

that the mean efficiency estimate by the deterministic frontier method is the lowest among the three sets of efficiency estimates while the stochastic frontier method produces the highest average efficiency estimate. This is consistent with the Monte Carlo results in Chapter 6. The study also shows that the efficiency estimates from the two parametric methods exhibit essentially the same pattern. However, there exist noticeable differences between the DEA-TOBIT efficiency estimates and those of the parametric methods although the policy implications from all three methods are consistent. The main reason for the difference between the DEA and the parametric methods is that the DEA-TOBIT is a two-step procedure while the parametric methods follow a one-step procedure which produces different estimates as shown by the simulation results in Chapter 6. Further, the DEA-TOBIT considers two output variables whereas the parametric methods consider only one aggregate output variable. The results from Monte Carlo experiments indicate that the efficiency estimates from the stochastic frontier method are closer to the "true" efficiency than the DEA method if the inputs and exogenous variables are correctly identified and accounted for. However, in this case study, we encounter the problem of collinearity among inputs and exogenous variables in estimating the stochastic frontier production function (and the deterministic frontier function). This is a rather common problem associated with estimating production functions. This may have caused misspecification of the "best practice" production function, consequently may have affected the accuracy of the efficiency estimates. Therefore, the differences between the DEA-TOBIT estimates and the parametric estimates may be partly attributable to the potential mis-specification of the model as well.

Chapter 9

The Efficiency of International Airlines

In this chapter, the three alternative methods are applied to the international airline industry to examine the productive efficiency of 36 major international airlines over the period of 1980-1992. Section 1 gives a brief introduction and reviews selected literature on airline productivity and efficiency. Section 2 describes the variables and the model specifications. The efficiency estimates are reported and discussed in Section 3, and the relative merits of the three methods are compared in Section 4. Finally, a summary and some concluding remarks are given in Section 5. Detailed descriptions of the sample airlines can be found in Appendix D.

9.1 Introduction

The airline industry is characterized by falling unit costs and a growing demand, however, the profitability of the airline industry world wide has been marginal. Since the 1978 deregulation of the U.S. domestic market, major institutional changes have been introduced in the world's commercial air transport industry resulting in increased competition in international aviation. As the market becomes more competitive, the ultimate ability of a carrier to survive and prosper depends greatly on improvement in its efficiency and productivity. The issue of efficiency and productivity will become increasingly important for airline industry since the differences in labour costs are likely to diminish over time as more and more airlines practice global sourcing of their flight crews and maintenance work.

There have been numerous studies on airline productivity and efficiency. Caves, Christensen and Tretheway (1981) compares 11 U.S. trunk air carriers on the basis of levels and rates of growth of outputs, inputs, and total factor productivity for the period of 1972-1977. They also examine the relationships between productivity and differences in outputs, average stage length, and load factor. Caves, Christensen, Tretheway and Windle (1987) compares the productive performance of a sample of U.S. and non-U.S. airlines over the 1970-1983 period in terms of the growth rate of total factor productivity, focusing on the effects of the U.S. deregulation. Gillen, Oum and Tretheway (1985, 1990) measure and compare the productive performance of seven Canadian airlines in terms of total factor productivity and unit cost. The study is conducted for the period of 1964 to 1981. Windle (1991) measures the productivity and unit costs of a set of U.S. and non-U.S. airlines and attempts to identify the factors which are most influential in explaining the observed differences in the airlines productivity. The study is based mainly on the 1983 data. Encaoua (1991) examines cost and productivity differences among European carriers, and finds that the gap in productivity measures between the carriers are shrinking during the 1981-1986 period.

Barla and Perelman (1989) adopts the stochastic frontier method to measure the efficiency performance of 26 airlines from OECD countries during 1976 to 1986. They estimate the production frontier function using the one-step procedure, and use available tonne-kilometre as output measure and consider two inputs, labour and capital, and four exogenous variables. Their study compares the efficiency performance of American carriers operating in deregulated markets with those of European and other airlines still subject to

regulatory control, and finds that the deregulated airlines do not perform better than regulated airlines. Bruning (1991) applies the two-step stochastic frontier (cost) function method to a sample of 73 international airlines in 1987 to assess the relationship between operating efficiency and measures of market competition. The study does not find any positive effects of competition on airlines' efficiency. Jha and Sahni (1992) applies the panel data approach proposed by Cornwell, Schmidt and Sickles (1989) to the case of six Canadian airlines to measure these airlines' technical inefficiency for the period of 1970 to 1986. They estimate two separate frontier production functions with available passenger kilometres (seat kilometres) and available tonne-kilometres as the output measures, respectively, using the same set of input variables in both estimations. Good, Nadiri, Röller, and Sickles (1993) compares the technical efficiency and productivity growth among the four largest European carriers and eight U.S carriers using the stochastic frontier method. They estimate the potential efficiency gains of the European aviation liberalization by comparing efficiency differences between the two carrier groups. The study is carried out for the period of 1976 to 1986. Loeb, Bruning and Hu (1994) estimates a Cobb-Douglas total cost frontier function to measure airlines' relative inefficiency levels for the years 1982, 1985, 1988, and 1990. The inefficiency estimates are then compared across market groups and years using ANOVA analysis. The study finds that the Pacific Rim carriers group is the least inefficient while the Northern European carriers are the most inefficient.

Good and Rhodes (1991) uses the data envelopment analysis method to examine the productive performance of 37 airlines from the Pacific region over the period of 1976 to 1986, and finds that there is a strong correlation between the productive efficiency of a

country's national carrier and the restrictiveness of the country's bilateral agreement with the U.S. Cooper and Gallegos (1991) employs a combination of the DEA method, the conventional regression analysis, and the frontier methods to examine the performance of Latin American airlines, in particular, the effects of ownership and international competition on airlines' performance. Based on the same data set as in Cooper and Gallegos (1991), Ray and Hu (1993) examines the annual productivity growth rate for the airlines during 1981 to 1988 using the data envelopment analysis method, and finds that the U.S. carriers experience faster growth in productivity than the Latin American carriers mainly due to technical progress. Distexhe and Perelman (1993) uses the data envelopment analysis method to measure the technical efficiency and productivity gains for 33 airlines during 1977-1988. They also find technological progress as a major source of productivity growth.

Most of the previous studies are based on the data prior to mid-1980s. Significant changes have occurred in the industry during the last few years such as consolidation of the airlines in the U.S. and Europe, regulatory liberalization of the industry in Europe and other countries, continued liberalization of international air transportation, and privatization of many airlines. It would be interesting to examine if there are any significant changes in the airlines' productive performance following the regulatory and institutional development in the international aviation industry.

In examining the efficiency performance of the airlines, most of the studies have taken into consideration of the effects of some output attribute variables, mostly load factor and stage length. Although being considered as important in most studies the effects of government ownership are examined analytically only in Bruning (1991) and Loeb (1994).

Their results appear to indicate, in contrast to common arguments, possible positive effects of government ownership on airlines's efficiency. Further examination of this issue is necessary.

People usually distinguish airlines' outputs as passenger, cargo and charter services. However, a closer examination of airlines's revenue shares (see Table D.2 in Appendix D) shows that the so-called incidental services, including catering services, ground handling services and maintenance services performed for other airlines, account for up to 30 percent of total operating revenues for some airlines with an average of about 8 percent for the sample airlines in this study. It therefore is an important component of the airlines' business. However, this component of the airlines' business has been largely ignored in previous studies on airline performance. This study considers the incidental services as one of the airlines' output in order to properly reflect the total output of airlines.

This chapter attempts to measure and compare the efficiency of 36 airlines over the 1980-1992 period and analyze it in order to identify the effects on efficiency of government ownership and technical progress. The study also attempts to control for the effects of other factors such as load factor, stage length, and output composition. The efficiency performance of the airlines is compared between airlines in different continents. In addition, the efficiency estimates from the three methods are compared.

9.2 The Data and Model Specification

Each observation in the sample is treated as an individual "decision-making-unit" (DMU). There are 447 DMUs in total. The sample carriers and data sources are described in Appendix D.

Airlines produce several distinct outputs, including scheduled passenger, scheduled freight, mail, non-scheduled, and incidental services. The DEA method is designed to deal with multiple output problems. However, there are problems in estimating parametric frontier production function¹ with multiple outputs. In order to have consistent results comparable across the three alternative methods, the outputs are aggregated into output index using the translog multilateral index procedure proposed by Caves, Christensen and Diewert (1982). Revenue shares are used as the weights in computing the output index since there is empirical evidence indicating the existence of constant returns to scale in the airline industry (Caves, Christensen, and Tretheway, 1984, Gillen, Oum and Tretheway, 1990, Cornwell, Schmidt, and Sickles, 1990).

Five categories of inputs are considered: (1) labour, (2) fuel, (3) materials, (4) flight capital, and (5) ground property and equipment. Appendix D provides details on definition and construction of these input variables.

Airlines operate over routes of varying size and length and in markets with different traffic densities. Therefore, a realistic characterization of the industry's productive

¹ The Monte Carlo study in Part I considers only frontier production functions. Therefore, production frontier functions are estimated here instead of the frontier cost functions in order to be consistent with Part I.

Table 9.1 Definition of Exogenous Variables	
Variables	Description
LOAD	Weight Load Factor for all traffic
STAGE	Stage Length: average flight distance
%PASS	Percentage of Passenger RTK in total RTK
%INTL	Percentage of International RTK in total RTK
MAJ	A Dummy for majority government ownership
MIN	A Dummy for minority government ownership
TIME	Time trend

performance requires not only the incorporation of multiple outputs and multiple inputs, but also the special characteristics of the route networks, output attributes, and technological changes over time. Table 9.1 lists a number of variables which reflect some special characteristics of airlines' operation. Note that the variable MAJ is a dummy variable which equals to 1 if government owns 50 percent or more of the airline. Similarly, the variable MIN is a dummy variable which equals to 1 if government's ownership of the airline is less than 50 percent.

The CCR ratio model of the DEA method, specified by equation 4.2.2, is used here. It is appropriate since earlier studies have found that there are roughly constant returns to scale for airlines over a rather broad ranges of network sizes. The CCR ratio model estimates a gross efficiency index using the output index and the five input variables. The TOBIT analysis is then used to examine the effects of the exogenous variables on the DEA

gross efficiency index and compute a residual efficiency index using equations (4.3.1) and (4.3.2).

The estimation of the parametric frontier functions assumes a Cobb-Douglas production technology with five inputs producing a single output². The deterministic frontier production function is estimated by COLS (see Chapter 4), and the stochastic frontier production function is estimated using the Battese and Coelli (1988) procedure. In particular, the deterministic core of the production frontiers is specified as:

$$\ln y_j = \alpha_0 + \sum_i \alpha_i \ln x_{ij} + \sum_k \beta_k \ln z_{kj} \quad (9.1)$$

where y_j is the output index for observation j , x_{ij} are the input variables, z_{kj} are the exogenous variables listed in Table 9.1, and α , β are the coefficients to be estimated.

9.3 The Results

This section reports and discusses the empirical results. The study proceeds first with the DEA-TOBIT analysis. Then the one-step deterministic frontier method is used to examine the airlines' productive performance. Finally, the one-step stochastic frontier method is applied to the sample data.

² Madala (1979) noted that measurement of technological change and efficiency are quite insensitive to the choice of functional form of production since these measures are related to shifts of the isoquants rather than their shapes.

9.3.1 The DEA Results

We first discuss the DEA gross efficiency estimates, and then examine the results from the TOBIT regression. A residual efficiency index is computed after removing the effects of the exogenous variables, and then compared with the DEA gross efficiency index.

9.3.1.1 The Gross Efficiency Estimates

The DEA gross efficiency estimates are computed using the CCR ratio model on a UNIX computer. The linear programming problem has to be solved once for each DMU (observation) in the sample, that is, a total of 447 linear programming solutions are required in order to compute the DEA gross efficiency index.

The DEA efficiency estimates are given in Table 9.2³. As pointed out in the earlier chapters, these DEA indices measure the "observed" productive performance, thus they reflect the combined effects of true managerial and operational efficiency and of the exogenous factors which are beyond the airlines' managerial control. One can only make meaningful inferences about the true efficiency performance of the airlines after controlling for the effects of the potentially influential factors.

In terms of the DEA gross efficiency index, Finnair, KLM, Singapore Airlines, Continental, TWA, JAL, and Cathay Pacific have achieved a position close to the production frontier in 1989-1991. Sabena, Finnair, Singapore Airlines, and Continental have experienced the most significant improvement in their performance during the sample period:

³ The 1992 data are not very reliable, therefore, the discussions on the results are focused on earlier years.

Table 9.2
Gross DEA Efficiency Index¹

	1980-82	1983-85	1986-88	1989-91	1992
Qantas	0.808	0.818	0.820	0.812	1.000
AUA	0.364	0.385	0.419	0.456	0.463
Sabena	0.776	0.882	0.806	0.927	0.913
Air Canada	0.759	0.743	0.738	0.729	0.691
Canadian	0.959	0.946	0.781	0.779	0.734
Finnair	0.734	0.741	0.831	0.944	1.000
Air France	0.712	0.757	0.728	0.691	0.676
UTA	0.932	0.984	0.990	0.964	-
Lufthansa	0.730	0.773	0.804	0.772	0.815
Air India	0.567	0.633	0.581	0.579	-
Alitalia	0.640	0.776	0.732	0.759	0.832
JAL	0.811	0.865	0.914	0.963	0.894
MAS	0.634	0.618	0.619	0.465	-
Mexicana	0.890	0.796	0.596	0.575	0.482
KLM	0.916	0.886	0.907	0.985	0.994
PIA	0.483	0.476	0.455	0.540	0.517
PAL	0.490	0.754	0.688	0.639	0.562
Tap Air	0.760	0.849	0.869	0.807	0.687
KAL	0.767	0.744	0.722	0.733	0.726
Saudia	0.768	0.896	0.838	0.827	-
SAS	0.659	0.691	0.610	0.682	0.690
SIA	0.783	0.856	0.947	0.986	1.000
Iberia	0.584	0.669	0.774	0.744	0.630
Swiss Air	0.811	0.818	0.874	0.872	1.000
British Airways	0.636	0.602	0.561	0.635	0.728
Cathay	-	-	1.000 ³	0.944	1.000
American	0.848	0.842	0.854	0.844	0.898
US Air	0.573	0.606	0.656	0.633	0.683
Continental	0.745	0.806	0.927	0.910	0.978
Delta	0.776	0.744	0.801	0.807	0.794
Northwest	0.955	0.992	0.812	0.867	0.914
TWA	0.844	0.870	0.820	0.948	1.000
United	0.967	0.912	0.889	0.873	0.893
American W	-	0.833 ³	0.795	0.918	1.000
Eastern	0.739	0.746	0.804	0.810	-
Pan Am	0.824	0.894	0.852	0.967	-

1. Table entries are all three-year average unless otherwise noted.

2. for 1988.

3 for 1985.

0.776 in 1980-82 to 0.927 in 1989-91 for Sabena, 0.734 in 1980-82 to 0.944 in 1989-91 for Finnair, 0.783 in 1980-82 to 0.986 in 1989-91 for Singapore Airlines, and 0.745 in 1980-82 to 0.910 in 1989-91 for Continental. The DEA gross efficiency index of British Airways improved consistently after its privatization in December 1986. On the other hand, Mexicana has experienced noticeable decrease in its performance during the same period. The DEA gross efficiency index of Canadian Airlines International (formally Canadian Pacific) has also declined considerably after 1986, the year in which PWA purchased CP Air and created CAI. In terms of these gross efficiency index, Austrian Airlines, Air India, and Pakistan Airlines appear to be among the least efficient carriers in this sample.

Table 9.3 compares mean values of the DEA gross efficiency estimates by continents of the airlines' registry. Major Asian carriers, including JAL, KAL, Singapore Airlines, and Cathay Pacific, are rated as the most efficient. Other Asian carriers, namely, Air India, Pakistan International Airlines, Philippines Airlines, and Malaysia Airlines, are rated as the least efficient ones. The U.S. carriers are very close to the major Asian Carriers in terms of the DEA gross efficiency index while the Canadian carriers are rated slightly better than the European Carriers. There is apparently no significant difference between the major European carriers and other European carriers in the sample.

Table 9.3 The Means of DEA Gross Efficiency Index						
	Europe Major	Europe other	Asia Major	Asia other	U.S.	Canada
Mean	0.756	0.752	0.860	0.581	0.834	0.797
St. Dev	0.116	0.193	0.100	0.099	0.110	0.093
Ob. No	91	77	44	50	121	26

9.3.1.2 The Effects of Exogenous Variables

The DEA gross efficiency estimates are affected by the differences in the airlines' operating environments. A number of factors have been identified as potentially having influences on the airlines' observed performance. They are listed in Table 9.1 and are discussed earlier. In the next step, we attempt to attribute the differences among the DEA gross efficiency estimates to the variations in these exogenous variables. A residual efficiency index is consequently computed after removing the effects of these variables.

The Tobit analysis is used to examine the effects on the DEA gross efficiency of these exogenous variables. Table 9.4 reports the TOBIT regression results. The main results are discussed below:

- (a) **The Effects of Output Characteristics:** The weight load factor (LOAD) is a measure of output performance (or service quality). It has a marginally significant positive coefficient, indicating that airlines with high load factor tend to achieve high DEA gross efficiency index. This occurs because when the load factor is high, more outputs are produced for a given level of input. That is, cost per unit output falls as

the load factor increases.

It should be noted that in the deregulated domestic markets, such as in the U.S., Canada, Australia, New Zealand, etc., airlines can manage load factor by choosing flight frequency and aircraft size for specific routes. In these markets, therefore, the load factor is not an "exogenous variable", but rather a policy variable. However, in regulated domestic markets, such as in Japan, China, India, etc, and many of the international markets governed by restrictive bilateral agreements between countries, airlines do not have direct control of the "load factor". Since most of the airlines in

Table 9.4 Tobit Regression Results		
Dependent Variable: LOG (DEA Index)		
	Coefficient	T-Value
LOAD	0.137	1.551
STAGE	0.196	9.271
%PASS	-0.333	5.740
%INTL	-0.069	5.824
MAJ	-0.127	6.305
T	0.003	1.323
CONSTANT	-2.370	6.147
No. of Observations	447	
Log-Likelihood	94.863	
R ²	0.381	

Note that the first four variables are in logarithm form.

our sample operate in both markets, a part of the effect of load factor is attributable to good management while the rest of the effect are beyond managerial control.

- (b) **The Effects of Network Characteristics:** The stage length (STAGE) measures the average flight distance, and is an important dimension of an airline's network. STAGE has a statistically significant positive coefficient, meaning that airlines flying longer routes are expected to achieve higher DEA gross efficiency index. This is expected because unit cost per passenger kilometer or per revenue tonne-kilometer decrease with average stage length. Since airport charges, station costs and other ground expenses do not vary with stage length, the costs spread over more passenger kilometers (or revenue tonne-kilometers) the longer the stage length is.
- (c) **The Effects of Output Composition:** Two variables describing the composition of airline output, %PASS and %INTL, are included in the Tobit regression in order to determine the direction and magnitude of their effects on the observed efficiency performance of the airlines. The variable %PASS is the percentage of passenger RTK in total RTK, and has a statistically significant negative coefficient. The negative coefficient indicates that, *ceteris paribus*, the airlines with larger share of their traffic in passenger services are likely to have a lower DEA efficiency rating. This is expected because passenger services are more input-intensive than freight services. Therefore, an airline with smaller %PASS would be able to produce more outputs with a given level of inputs than an airline with higher %PASS. The variable %INTL is the share of international traffic in total traffic. It has a significant negative coefficient indicating that an airline with a higher proportion of international

services tends to have a lower DEA gross efficiency index. For a given load factor and stage length, international services may incur additional costs such as crew stopovers due to lower frequencies, costs associated with sales, ticketing and handling of passengers, etc. Another reason for the negative coefficient of %INTL is that restrictions on airlines' operations imposed by bilateral air services agreements cause inefficiency.

- (d) **The Effects of Government Ownership:** Two dummy variables are used to indicate the extent of government ownership: MAJority and MINority. Since variable MINority is not statistically significant, it is not included in the final regression. This implies that a minority government ownership would not have any significant positive or negative effect on the productive performance of the airlines. On the other hand, variable MAJority has a statistically significant negative coefficient indicating airlines with government majority ownership are likely to be less efficient. These airlines often have less managerial freedom for making strategic and operational changes in order to improve efficiency. In addition, existence of subsidized services required by government makes airlines less efficient by reducing incentives to improve efficiency.
- (e) **The Effects of Time:** TIME variable is used to estimate the residual technical (or managerial) efficiency over time. TIME has a (marginally) statistically significant positive coefficient, indicating that the airline industry experienced technological progress at an annual rate of about 0.3% during the sample period, 1980-1992.

9.3.1.3 The Residual Efficiency

As seen from the forgoing discussions, the exogenous factors do have a significant effect on the DEA gross efficiency index. Therefore, it is necessary to purge the effects of these factors from the airlines' observed efficiency levels in order to have a better picture about the "true" relative efficiency performance of the airlines.

Table 9.5 presents the airlines' residual efficiencies computed from the TOBIT regression after removing the effects of the exogenous variables. It should be noted that these residual efficiencies may still be distorted by factors which are left out of the regression because of lack of information. However, it is overall a better indicator of the airlines' productive performance than the DEA gross index.

From Table 9.5, we can see that the differences in the residual efficiencies among the airlines are much smaller than the DEA gross efficiency index. Some of the variations in the observed efficiency performance as measured by the DEA gross efficiency index are caused by variations in the exogenous variables which are controlled for in the TOBIT regression.

Finnair, KLM, SIA, Continental, and TWA are still rated among the most efficient carriers in 1989-91 in terms of the residual efficiency index, while the performance of Cathay Pacific and JAL have been reduced from the top performers in terms of the DEA gross efficiency index to that of slightly above average performers in terms of residual efficiency. The relative efficiency rankings of UTA, KAL, and American Airlines are also reduced considerably in terms of the residual efficiency index. On the other hand, Malaysia Airlines, Saudia, SAS, and Iberia are given relatively higher ratings by the residual

Table 9.5
Residual Efficiency Index¹

	1980-82	1983-85	1986-88	1989-91	1992
Qantas	0.665	0.647	0.636	0.620	0.756
AUA	0.432	0.452	0.488	0.501	0.493
Sabena	0.611	0.674	0.613	0.718	0.760
Air Canada	0.764	0.713	0.691	0.586	0.555
Canadian	0.792	0.771	0.661	0.651	0.599
Finnair	0.784	0.749	0.778	0.837	0.887
Air France	0.644	0.648	0.607	0.572	0.556
UTA	0.713	0.731	0.729	0.704	-
Lufthansa	0.689	0.692	0.693	0.662	0.718
Air India	0.471	0.529	0.470	0.456	-
Alitalia	0.638	0.741	0.683	0.685	0.763
JAL	0.603	0.625	0.648	0.684	0.644
MAS	0.753	0.720	0.679	0.483	-
Mexicana	0.854	0.779	0.564	0.531	0.448
KLM	0.716	0.665	0.670	0.715	0.724
PIA	0.490	0.479	0.454	0.533	0.513
PAL	0.470	0.706	0.650	0.590	0.505
Tap Air	0.742	0.803	0.825	0.765	0.656
KAL	0.565	0.540	0.520	0.551	0.518
Saudia	0.819	0.992	0.856	0.813	-
SAS	0.712	0.742	0.656	0.727	0.738
SIA	0.683	0.716	0.754	0.745	0.742
Iberia	0.625	0.689	0.785	0.734	0.623
Swiss Air	0.731	0.703	0.734	0.718	0.801
British Airways	0.641	0.590	0.498	0.527	0.593
Cathay	-	-	0.727 ²	0.679	0.711
American	0.664	0.677	0.717	0.701	0.739
US Air	0.521	0.522	0.532	0.536	0.596
Continental	0.607	0.672	0.783	0.770	0.819
Delta	0.683	0.646	0.688	0.688	0.683
Northwest	0.814	0.806	0.680	0.714	0.745
TWA	0.729	0.740	0.786	0.799	0.836
United	0.688	0.693	0.721	0.716	0.732
American W	-	0.699 ³	0.658	0.718	0.774
Eastern	0.635	0.639	0.684	0.699	-
Pan Am	0.644	0.730	0.709	0.789	-

1. Table entries are all three-year average unless otherwise noted.

2. for 1988.

3. for 1985.

efficiency index. However, Air India, Pakistan International Airlines, and Austrian Airlines are still rated as the least efficient carriers among the sample airlines.

Continental and Sabena have had the most significant improvement in their performance over the 13-year time period, while the performance of Air Canada, Canadian Airlines International, Air France, Malaysia Airlines, and Mexicana have experienced noticeable declines. Another observation from the results is related to British Airways: the efficiency ratings for British Airways declined consistently during the earlier 1980s, then improved consistently to the end of 1980s and earlier 1990s. This reflects the sluggish performance of BA prior to 1986, and its ambitious restructuring process following its privatization in 1986.

The mean DEA residual efficiency estimates by regions are shown in Table 9.6. As one can see, there are very small differences in the mean residual efficiency estimates between regions. Recall from Table 9.3 the regional mean DEA gross efficiency estimates range from 0.581 to 0.860 which is considerably reduced in terms of the DEA residual efficiency index to ranging from 0.565 to 0.719. The *Major Asian* airlines have the highest mean DEA gross efficiency rating, however, in terms of the DEA residual efficiency, their performance is reduced to that close to the lower end. The *Other Asian* carriers are still the least efficient carrier group. It is also noted that the relative performance of European Majors and Other European Carriers, and the relative performance of Canadian carriers and U.S. carriers are reversed as compared to the DEA gross efficiency estimates.

Table 9.6
Mean DEA Residual Efficiency Estimates by Region

	Europe Major	Europe other	Asia major	Asia other	U.S.	Canada
Mean	0.669	0.688	0.643	0.565	0.692	0.694
St. Dev.	0.075	0.122	0.081	0.110	0.083	0.079
Obs No.	91	77	44	50	121	26

9.3.2 The Deterministic Frontier Method

The deterministic frontier method is applied to the airline data set with the same set of output and input variables as used in the DEA-TOBIT analysis. The estimation of the frontier production function follows the one-step procedure which includes the exogenous variables directly in the estimation.

The estimated "best practice" production function is reported in Table 9.7. The four input variables are all statistically significant and, as expected, have the positive coefficients. Except for dummy variable MIN, all the exogenous variables appear to be statistically significant⁴. LOAD factor has a positive coefficient implying that airlines with high load factor achieve high efficiency ratings when their output is measured in terms of revenue generating output such as RPK, RTK, etc. STAGE length also has a significant positive coefficient implying that airlines with longer average stage length are expected to achieve

⁴ Actually %INTL is not significant at the traditional 5%. However, it is included since it serves to confirm the results from the DEA-TOBIT procedure.

Table 9.7 Deterministic Frontier Production Function		
Dependent Variable: LOG (OUTPUT INDEX)		
Variables	Coefficient	T-Value
LABOUR	0.268	8.195
FUEL	0.309	9.369
MATERIALS	0.048	1.714
FLIGHT EQUIPMENT	0.386	12.63
LOAD	0.682	8.100
STAGE LENGTH	0.218	11.18
%PASS	-0.239	4.834
%INTL	-0.014	1.292
MAJORITY	-0.118	6.475
TIME	0.004	1.796
CONSTANT	-1.064	2.506
R ²	0.9719	
Log-likelihood Function	216.526	
No. of Observations	447	

Note that all variables are in logarithm form except for MAJ and TIME.

higher observed efficiency. On the other hand, a carrier with a large proportion of passenger traffic or a large proportion of international traffic is expected to achieve a lower efficiency rating. The ownership variable MAJ has a negative coefficient of -0.12 which is very similar to the DEA-TOBIT results. This indicates that airlines with government majority ownership are about 11 to 12 % less efficient. The coefficient for the TIME is statistically significant, and again similar to the DEA-TOBIT results, indicating that the annual technological progress in the international airline industry is about 0.3 to 0.4 %. In summary, the results in Table 9.7 essentially confirm the DEA-TOBIT results.

By following the procedure described in equations (4.1.5) to (4.1.7), the observation specific efficiency estimates are estimated using the production function shown in Table 9.7. Table 9.8 lists the efficiency estimates from the deterministic frontier method. Since the exogenous variables are included in the estimation of the frontier production function, and their effects are removed in computing the efficiency estimates, these efficiency estimates in Table 9.8 are closer to the DEA residual efficiency index than the DEA gross efficiency index. SIA, Continental, and TWA are still among the most efficient carriers in 1989-91, however, the performance of Finnair is reduced to that of mid-range. Air India, Pakistan International Airlines, and Austrian Airlines are still among the least efficient carriers as indicated before by the DEA residual efficiency, but the performance of Philippines Airlines is given a better rating. Sabena and Continental are shown to have experienced considerably improvement during the sample period, while the performance of Air Canada and Canadian Airlines International have declined during the same period.

The mean efficiency estimates from the deterministic frontier method by regions are

Table 9.8
Efficiency Estimates by Deterministic Frontier Method¹

	1980-82	1983-85	1986-88	1989-91	1992
Qantas	0.624	0.628	0.643	0.609	0.686
AUA	0.568	0.518	0.528	0.525	0.579
Sabena	0.682	0.689	0.670	0.830	0.853
Air Canada	0.776	0.725	0.735	0.599	0.528
Canadian	0.767	0.803	0.692	0.689	0.623
Finnair	0.687	0.727	0.784	0.753	0.663
Air France	0.621	0.596	0.589	0.556	0.717
UTA	0.829	0.846	0.895	0.814	-
Lufthansa	0.705	0.774	0.768	0.698	0.734
Air India	0.551	0.532	0.489	0.475	-
Alitalia	0.701	0.747	0.583	0.721	0.847
JAL	0.590	0.586	0.552	0.585	0.569
MAS	0.790	0.731	0.713	0.535	-
Mexicana	0.773	0.842	0.651	0.647	0.654
KLM	0.722	0.648	0.682	0.658	0.649
PIA	0.527	0.495	0.518	0.582	0.565
PAL	0.583	0.755	0.706	0.697	0.620
Tap Air	0.803	0.773	0.873	0.837	0.730
KAL	0.624	0.651	0.626	0.660	0.578
Saudia	0.883	0.875	0.762	0.783	-
SAS	0.616	0.787	0.722	0.724	0.737
SIA	0.702	0.831	0.877	0.837	0.794
Iberia	0.626	0.746	0.816	0.730	0.659
Swiss Air	0.639	0.612	0.749	0.704	0.816
British Airways	0.664	0.639	0.558	0.581	0.637
Cathay	-	-	0.813 ²	0.706	0.691
American	0.717	0.732	0.747	0.702	0.689
US Air	0.586	0.629	0.600	0.570	0.612
Continental	0.668	0.706	0.802	0.799	0.817
Delta	0.698	0.751	0.727	0.691	0.692
Northwest	0.692	0.741	0.651	0.718	0.726
TWA	0.718	0.753	0.803	0.805	0.856
United	0.704	0.706	0.738	0.708	0.700
American W	-	0.685 ³	0.617	0.689	0.682
Eastern	0.644	0.655	0.717	0.592	-
Pan Am	0.616	0.702	0.732	0.757	-

1. Table entries are all three-year average unless otherwise noted. 2. for 1988. 3 for 1985.

given in Table 9.9. The average efficiency level is slightly higher than those of the DEA residual efficiency index. Carriers in the "other Asia" group are again the least efficient carriers. The average performance of major Asian carriers and major European carriers are similar at group level. Canadian carriers appear to perform slightly better than the US carriers on average. The relative performance between the regions are similar to that in terms of the DEA residual efficiency index. The efficiency gaps among the sample airlines are fairly small as in the case of DEA residual efficiency estimates.

Table 9.9 Mean Efficiency Estimates by Deterministic Frontier						
	Europe Major	Europe other	Asia Major	Asia other	US	Canada
Mean	0.677	0.727	0.680	0.612	0.701	0.712
St. Dev	0.078	0.117	0.110	0.113	0.072	0.077
Obs.	91	77	44	50	121	26

9.3.3 The Stochastic Frontier Method

This section reports and discusses the results from the stochastic frontier method. Again the same set of input and output variables are considered, and the estimation of the frontier production function follows the one step procedure which includes the exogenous variables directly in the estimation.

Table 9.10 presents the deterministic core of the stochastic frontier function. This function is very close to the deterministic frontier production function (Table 9.7) except that

Table 9.10 Stochastic Frontier Production Function		
Dependent Variable: LOG (OUTPUT INDEX)		
Variables	Coefficient	T-Value
LABOUR	0.263	9.043
FUEL	0.253	10.79
MATERIALS	0.113	4.056
FLIGHT EQUIPMENT	0.342	13.35
GROUND EQUIPMENT	0.019	1.647
LOAD	0.574	8.539
STAGE LENGTH	0.264	16.12
%PASS	-0.187	4.777
%INTL	-0.011	1.012
MAJORITY	-0.117	5.077
MINORITY	-0.042	1.511
TIME	0.007	3.645
CONSTANT	-0.773	1.987
Log-likelihood Function	239.966	
No. of Observations	447	

Note that all variables are in logarithm form except for MAJ and TIME.

the GPE and MINORITY variables here are statistically significant while they are not in the deterministic model, and the variable %INTL is not statistically significant here while it is marginally significant in the deterministic model.

Similar to the results from previous production functions, the stochastic frontier results indicate that airlines with higher load factor and longer stage length are expected to have higher gross efficiency ratings, while airlines which have a larger percentage of passenger traffic and/or international traffic are expected to have lower gross efficiency ratings. Government ownership in airlines will have negative effects on the productive efficiency of the airlines. Technological progress has improved the overall performance of the industry during the sample time period. This is particularly the case for some of the major Asian carriers. They exhibit considerable improvement in terms of the gross efficiency measures⁵, but after removing the effects of the exogenous variables their "net" efficiency improvements over time appear to be reduced substantially. However, it is noted that the annual rate of technological progress indicated by the stochastic frontier method is higher than that from the deterministic frontier method or the DEA-TOBIT analysis. This may be partly due to the fact that the stochastic frontier method attempts to filter out the statistical noise, thus attribute some of the "left over" positive variations in the efficiency estimates to technological progress.

The efficiency estimates from the stochastic frontier model are given in Table 9.11. Overall, these efficiency estimates are higher than those from the other two methods. This

⁵ The efficiency estimates from first stage of the two-step parametric (both deterministic and stochastic) procedure also show similar pattern.

Table 9.11
Efficiency Estimates by Stochastic Frontier Method¹

	1980-82	1983-85	1986-88	1989-91	1992
Qantas	0.730	0.734	0.740	0.707	0.802
AUA	0.626	0.584	0.588	0.593	0.651
Sabena	0.851	0.854	0.820	0.954	0.967
Air Canada	0.944	0.884	0.891	0.722	0.643
Canadian	0.934	0.941	0.834	0.829	0.748
Finnair	0.874	0.907	0.850	0.926	0.796
Air France	0.785	0.762	0.750	0.707	0.846
UTA	0.944	0.948	0.963	0.903	-
Lufthansa	0.890	0.951	0.944	0.873	0.913
Air India	0.572	0.595	0.549	0.524	-
Alitalia	0.848	0.910	0.939	0.876	0.970
JAL	0.773	0.770	0.712	0.732	0.699
MAS	0.936	0.875	0.856	0.678	-
Mexicana	0.916	0.920	0.751	0.734	0.737
KLM	0.922	0.834	0.856	0.833	0.826
PIA	0.599	0.572	0.607	0.661	0.659
PAL	0.715	0.914	0.837	0.836	0.734
Tap Air	0.887	0.873	0.932	0.889	0.776
KAL	0.761	0.784	0.750	0.797	0.711
Saudia	0.965	0.955	0.895	0.910	-
SAS	0.796	0.945	0.871	0.873	0.882
SIA	0.845	0.953	0.964	0.937	0.909
Iberia	0.772	0.906	0.955	0.863	0.770
Swiss Air	0.830	0.796	0.913	0.885	0.966
British Airways	0.820	0.788	0.685	0.706	0.768
Cathay	-	-	0.962 ²	0.868	0.855
American	0.891	0.910	0.931	0.875	0.860
US Air	0.760	0.805	0.776	0.734	0.777
Continental	0.836	0.872	0.964	0.959	0.966
Delta	0.907	0.937	0.918	0.877	0.870
Northwest	0.907	0.941	0.841	0.908	0.912
TWA	0.896	0.921	0.963	0.961	0.977
United	0.913	0.904	0.931	0.891	0.880
American W	-	0.755 ³	0.786	0.867	0.853
Eastern	0.841	0.846	0.903	0.774	-
Pan Am	0.778	0.873	0.892	0.922	-

1. Table entries are all three-year average unless otherwise noted

2. for 1988. 3 for 1985.

observation is consistent with the results from Monte Carlo study in Chapter 6.

Sabena, Finnair, SIA, Continental, and TWA are among the most efficient carriers in 1989-91 according to these efficiency estimates. Air India, Austrian Airlines and Pakistan International Airlines are still among the least efficient carriers. Sabena and Continental are shown to have experienced substantial improvement in efficiency performance over the 13-year time period. Singapore Airlines, Iberia, and American Airlines had exhibited considerable efficiency improvement during the first half of the 1980s, however, their performance show a downward trend since the mid-1980. The performance of Air Canada, Canadian Airlines International, and Malaysia Airlines appear to deteriorate throughout the sample period.

Table 9.12 Mean Efficiency Estimates Using Stochastic Frontier						
	Europe Major	Europe other	Asia major	Asia other	US	Canada
Mean	0.876	0.846	0.820	0.718	0.877	0.859
St. Dev	0.083	0.124	0.092	0.144	0.074	0.091
Obs	91	77	44	50	121	26

Table 9.12 lists the mean efficiency estimates by region. According to these results the US carriers are the most efficient airlines on average followed by the Canadian carriers and the carriers in the "other European" group. The "other Asian" carriers are the least efficient carriers.

9.4 Comparison of Efficiency Estimates by Alternative Methods

In the previous section, the three alternative methods are applied to the same airline data set to examine the efficiency performance of the airlines. Earlier discussions about the results are focused on the general policy implications and the performance of individual airlines. In this section, the three sets of efficiency estimates are compared in terms of the means and correlations.

Table 9.13 Means of Airlines' Efficiency Estimates			
	Deterministic	Stochastic	DEA-TOBIT
Mean	0.690	0.834	0.669
St. Dev	0.100	0.112	0.108
Maximum	1.000	0.987	1.000
Minimum	0.434	0.484	0.415

Table 9.13 lists the means of the residual efficiency estimates by the three methods. The mean of the efficiency estimates by the stochastic frontier method is considerably higher than those by the deterministic frontier method and the data envelopment analysis method. This observation is expected in view of the Monte Carlo results in Part I, which consistently show higher mean efficiency estimates by the stochastic frontier method than the other two methods. The lower means of the deterministic frontier method and the DEA are due to the fact that both methods attribute deviations from the frontier as inefficiency. Some of the inefficiencies indicated by these two methods are in fact due to the effects of some unidentifiable exogenous variables which are accounted for by the stochastic frontier method.

Table 9.14 Correlation and Rank Correlation Coefficients Between the Alternative Efficiency Estimates of Airlines			
Correlation Coefficients			
	Deterministic	Stochastic	DEA-TOBIT
Deterministic	1		
Stochastic	0.9262	1	
DEA-TOBIT	0.7945	0.8282	1
Rank Correlations			
Deterministic	1		
Stochastic	0.9562	1	
DEA-TOBIT	0.7936	0.8277	1

The correlations among the three sets of residual efficiency estimates are investigated using both Pearson correlation coefficients and Spearman's rank correlation coefficients. The results are reported in Table 9.14. Although there are some significant differences in terms of individual airlines' efficiency estimates as mentioned in the previous section, there is a very high correlation between the residual efficiency estimates from the two parametric methods, and the correlation between the DEA results and those from the two parametric methods are also reasonably high. This indicates that for the present airline case the results are not very sensitive to the choice between the stochastic frontier method and the deterministic frontier method, and the choice between the parametric methods and the DEA does not make any dramatic differences in the results either. The high rank correlations indicate that the choice of different methods does not impinge much on the ranking of firm

(observation) specific efficiency estimates among the airlines. Thus, the results in Table 9.14 suggest that the three different methods yield broadly similar results in measuring airlines' efficiency, at least for the functional form considered in this chapter. The Monte Carlo results in Part I indicate that when the variations in the firms' environments are not very large, the three alternative methods are expected to yield similar efficiency estimates. The high correlations and rank correlations in Table 9.14 are essentially consistent with the Monte Carlo results.

9.5 Summary and Concluding Remarks

This chapter provides some empirical evidence on the comparative performance of the three alternative efficiency measurement methods, namely, the deterministic frontier method, the stochastic frontier method, and the data envelopment analysis method. The three methods are applied to a sample of 36 international airlines during the period of 1980 to 1992 to measure the observation specific efficiency of the carriers and to identify the effects on the efficiency performance of government ownership, operating characteristics, and technological progress.

The two parametric frontier methods are estimated following an one-step procedure which incorporates the exogenous variables directly in the estimation of the frontier functions. The data envelopment analysis method utilizes a two-step procedure. At the first stage, the CCR ratio model is used to estimate a gross efficiency index of the airlines. At the second stage, the gross efficiency estimates are analyzed using the TOBIT regression to identify the effects of exogenous variables and to compute a residual efficiency index. The

empirical results from all three models indicate that: (1) airlines with high load factor and/or long stage length are expected to achieve higher gross efficiency ratings, (2) airlines with a large proportion of passenger services and/or international traffic are likely to have lower efficiency ratings, (3) airlines with majority government ownership are significantly less efficient, implying that a greater degree of managerial freedom may enhance an airline's productive efficiency substantially, and (4) technological progress has significantly improved the overall performance of the industry during the sample period, especially for some of the major Asian carriers. In addition, the discrepancy between the "gross" efficiency index and the "residual" efficiency index indicates that the exogenous variables do have considerable effects on the efficiency performance of the airlines.

The comparison of the results from the three methods illustrates that although there are noticeable differences in the actual levels of estimated efficiency, the overall pattern of the efficiency estimates from the two parametric methods are essentially the same, and the results by the DEA-TOBIT are broadly similar to those from the parametric methods. The policy implications from all three methods are consistent.

By comparing the results from Chapter 8 and Chapter 9, it is found that the three alternative methods give more consistent results in the airline case than in the railway case. This is because that the airline data set is considered as a 'better' data set for application of the three methods for the following reasons: (1) the data are fairly consistent across the airlines, (2) there are more observations, and (3) there is less degree of variations in the carriers' operating environments. This is consistent with the Monte Carlo results in Part I which indicate that the three alternative methods are expected to yield similar efficiency

estimates when the variations in firms' environments are not very large. The empirical lesson from this is that in situations where there are large variations in the DMUs' operating environments and production characteristics the results would be sensitive to the choice of method, therefore, careful examination of the situation and the estimation results is necessary.

Chapter 10

Summary and Conclusions

Productive efficiency is a performance measure to evaluate production units, and is an indicator of success. Measuring efficiency and identifying the sources of efficiency differentials are essential for designing public and corporate policies to improve performance. During the last three decades, various methods have been developed to measure productive efficiency. Different methods often yield different efficiency rankings among the firms being considered, and may lead to different policy implications on how to improve the efficiency of a particular firm and of the overall industry. Each method has its strengths and weaknesses. Knowledge of these strengths and weaknesses will help researchers and policy analysts to choose the most "suitable" method for a particular situation, and thus to make accurate measurements of efficiency. These in turn will help policy makers to make appropriate policy decisions. Therefore, it is important to study the relative merits of different methods in terms of their abilities to reveal the structure of production technology and the nature and extent of inefficiency under different conditions. This study compares three alternative methods in measuring firm specific efficiency, namely the deterministic frontier method, the stochastic frontier method, and the data envelopment analysis method.

In Part I, Monte Carlo experiments are carried out to examine the relative merits of the three methods where the underlying production technology and efficiency profile are known. The Monte Carlo results and their methodological implications are summarized as follows:

- The performance of all three methods improve as the sample size increases. Therefore, it is desirable to have a large sample whenever possible. However, it should be noted that when the sample size is too large computational cost of the DEA method becomes high.
- The variations in inputs do not appear to have much effect on the performance of the three methods in cases where the elasticity of input substitution is greater than one. However, in the case of weak input substitution, the performance of the deterministic frontier method and the stochastic frontier method fall noticeably with the variations in input variables. Therefore, the DEA method would be a good choice in situations where there are evidences indicating weak input substitution, and where large variations in sample firms' input variables are observed.
- The performance of all three methods deteriorates sharply as the level of noise rises, especially the DEA and the deterministic frontier method. The stochastic frontier method is expected to produce better estimates than the other two methods when the noise level is high.
- The magnitude of exogenous variables does not appear to have any significant effects on the performance of the one-step parametric methods as long as the exogenous variables can be correctly identified and accounted for. On the other hand, the performance of the two-step procedure, especially for the DEA and the deterministic frontier method, is very sensitive to the magnitude of the exogenous variables. Technically, it is desirable to use the one-step stochastic frontier method. However, the two-step procedure relates the exogenous variables directly to efficiency

performance, and thus may be appealing to the policy and decision makers. The DEA-Tobit method is a reasonably good alternative in situations where variations in the firms' operating environments are not very large.

- The stochastic frontier method is rather robust with respects to outliers. Both the DEA method and the deterministic frontier method are affected by the presence of outliers, but the extent of the effects are not as significant as expected. If it is possible, one should try to identify and investigate the causes of any potentially influential outliers before applying the DEA method and the deterministic frontier method to a particular data set.
- The performance of the stochastic frontier method is not heavily influenced by the structure of the underlying production technology, while that of the deterministic frontier method is affected by the presence of input complementarity, but not by the returns to scale. The performance of the DEA method deteriorates as the returns to scale increases, and in the presence of input complementarity. The use of DEA methods should be avoided as much as possible in situations where there are increasing returns to scale, and/or high complementarity among inputs. Although the BCC model is developed to account for the effects of returns to scale, our Monte Carlo results show that it can not effectively deal with the increasing returns to scale condition.

The simulation results in Part I provide some general guidelines regarding the performance of the three alternative methods under certain known conditions. Since in practice the "true" underlying production technology is not likely to be known, the choice

of methods would depend on careful examination of available information and a good understanding of the production situation. The two case studies in Part II serve as examples of how the three methods can be applied to a real world problem. Further, researchers may encounter problems which are not considered in simulation studies. The two case studies also illustrate some of the problems one may encounter in empirical situations where there is no prior knowledge of either the production technology nor the efficiency profile.

Chapter 8 reports on a case study of efficiency performance of railways from 19 OECD countries during the period of 1978-89. The three alternative methods are applied to the sample data to measure the productive efficiency of the railways and to identify the effects of government intervention and subsidization on the productive efficiency. The results show that: (1) railway systems with high dependence on public subsidies are significantly less efficient than similar railways with less dependence on subsidies, and (2) railways with high degree of managerial autonomy from regulatory authority tend to achieve higher efficiency. The empirical results also confirm that efficiency measures may not be meaningfully compared across railways without controlling for the variations in railways' operating and market environments. In addition, comparison of the efficiency estimates from the three alternative methods confirms the Monte Carlo result in Part I that the stochastic frontier method yields higher efficiency estimates, on average, than the other two methods. The efficiency estimates by the two parametric methods are highly correlated although their mean values are different. There are substantial differences between the efficiency estimates obtained from the DEA-TOBIT analysis and those obtained by the parametric methods. However, the policy implications from all three methods are consistent. The main reason

for the differences is that the DEA-TOBIT analysis is a two-step procedure while the parametric frontier production functions are estimated using an one-step procedure. The Monte Carlo study in Part I has shown that these two procedures often produce different efficiency estimates. Further, the DEA-TOBIT analysis considers two output variables, while the parametric methods consider only one output variable.

The second case study is reported in Chapter 9. It measures and compares the efficiency of 36 international airlines during the period of 1980-1992, and identifies the effects on efficiency of government ownership and technical progress. The empirical results show that technological progress has improved the productive efficiency of the airline industry over time, especially for some of the major Asian carriers. The airlines with majority government ownership are shown to be less efficient than other airlines with similar operating characteristics. The results also indicate that the effects of network and market environments should be controlled for in order to measure productive efficiency meaningfully comparable across airlines.

These two cases represent two rather different situations. In the railway case, the services are mostly provided by highly regulated, nationalized firms. The firms operate in very different environments. The noise level and the effects of exogenous variables are expected to be high, and it is very likely to have outliers in the sample. In the case of international airlines, firms operate in a fairly competitive environment. Although there is a high degree of diversity in size, the firms have access to essentially the same technologies. They acquire basic capital inputs (aircrafts), fuel and other inputs from the international market. The noise level and the magnitude of the effects of exogenous variables are likely

to be lower than those in the railway case. Also, the sample size in the airline case is larger than that in the railway case. Therefore, in view of the Monte Carlo results in Part I we expect more consistent results from the three alternative methods in the airline case. There is much less discrepancy and higher correlations among the alternative efficiency estimates in the airline case than in the railway case. This is consistent with the results of the Monte Carlo study in Part I.

Both cases are affected by statistical noise and identifiable exogenous variables. According to the Monte Carlo results, the one-step stochastic frontier method would produce best estimates if one has a clear picture of the production situation and the environment such that the model can be correctly specified. However, there are some practical limitations to the parametric methods. First, data on input and output prices are often required to apply the parametric methods, especially for the cases involving multiple outputs. However, consistent price data across sample firms may not be available. Further, the intrinsic collinearities among the explanatory variables pose another problem for estimation of the parametric frontier functions. When these problems cannot be overcome, the Data Envelopment Analysis method may be the only choice available.

It is worth noting some of the limitations of this study. (1) The results from the Monte Carlo experiments in Part I are based on a specific experimental design. Although it is believed that the experimental design as described in Chapter 5 is based on reasonable assumptions about possible empirical situations, at least for the transportation industry, there still exists a possibility that the findings may change if the experimental design, such as the ranges of exogenous variables and outliers, is altered. (2) Total train kilometres is used as

the output measure in estimating the parametric frontier productions in the railway case because data required to aggregate the multiple outputs into a more sensible single output measure are not available. (3) The number of employees is used as the labour input in both applications, because of the lack of better data such as number of labour hours by job categories. (4) For convenience, all exogenous variables are assumed to affect only the observed production output level, thus are treated in the same manner. Further study can be conducted to address this issue under the assumption that some of these variables would affect only the observed production level (input consumption and output level), while others would affect the "true" efficiency level and consequently affect the observed production level. Furthermore, exogenous variables are assumed to be independent of the efficiency performance. In practice, however, there may exist certain causal relationship between some "exogenous variables" and efficiency, such as efficiency and subsidization. How to treat such causal relationship properly in making efficiency measurement requires further research.

This thesis is concerned only with technical efficiency. Allocative efficiency and profitability are not considered in the thesis. Technical efficiency measures only the relation between output and input quantities, it does not consider the costs of the inputs nor the (monetary) values of the outputs. That is, there is no obvious relationship between efficiency and profitability. This is confirmed by examining the efficiency performance of the railways and airlines and their financial performance in terms of operating ratios. For example, British Airways is considered as a successful example by the international aviation industry, it has made profits when most airlines have been losing money. However, its efficiency rating is relatively low among the sample airlines. Although high technical efficiency is

desirable in any circumstances, the survival of a business depends on its profitability as well. Therefore, one area for further research is the evaluation of the overall performance of a firm or organization in terms of technical efficiency, allocative efficiency and profitability. Goal programming techniques may be useful in this aspect. Goal programming is often used in situations where a firm wishes to simultaneously achieve a number of objectives. Goal programming refers to these potential objectives as goals. The goal programming procedure places all the goals into the constraint set, and imposes a new objective function equal to a weighted sum of the deviations (from the goals). The solution will achieve a compromise among the goals based on the weights attached to the deviations.

To apply the goal programming techniques to evaluate the overall performance of a firm or organization, we could treat technical efficiency, allocative efficiency, and profitability as three separate but related goals. The questions then are how to determine the appropriate weights to each of these goals, and how to relate these goals together.

Bibliography

- Afriat, S. N. (1972), "Efficiency Estimation of Production Functions", **International Economic Review** 13(3), pp.568-598
- Aigner, D.J. and S.F. Chu (1968), "On Estimating the Industry Production Function", **American Economic Review**, 58, pp.826-839
- Aigner, D., C.A.K. Lovell and P. Schmidt (1977), "Formulation and Estimation of Stochastic Frontier Production Function Models", **Journal of Econometrics** 6, pp.21-37
- Ali, A.I. and L.M. Seiford (1993), "The Mathematical Programming Approach to Efficiency Analysis", in H.O. Fried, C.A. Knox Lovell, and S.S. Schmidt eds. **The Measurement of Productive Efficiency: Techniques and Applications**, Oxford University Press, 120-159
- Antoniou, A. (1992), "The Factors Determining the Profitability of International Airlines: Some Econometric Results", **Managerial and Decision Economics**, Vol.13, pp503-514
- Arrow, K.J., H.B. Chenery, B. Minhas and R.M. Solow (1961), "Capital-Labour Substitution and Economic Efficiency", **Review of Economics and Statistics**, 43, 225-250.
- Banker, R.D. (1984), "Estimating Most Productive Scale Size Using Data Envelopment Analysis", **European Journal of Operational Research**, 17, 35-44
- Banker, R.D., A. Charnes and W.W. Cooper (1984), "Some Models for Estimating Technical and Scal Inefficiencies in Data Envelopment Analysis", **Management Science** 30, pp.1078-1092
- Banker, R.D., R.F. Conrad and R.P. Strauss (1986), "A Comparative Application of Data Envelopment Analysis and Translog Methods: an Illustrative Study of Hospital Production", **Management Science**, 32, No.1, pp.30-44.
- Banker, R. D., A. Charnes, W.W. Cooper and A. Maindiratta (1988), "A Comparison of DEA and Translog Estimates of Production Frontiers using Simulated Observations from a Known Technology", in A. Dogramaci and R. Färe (eds) **Applications of Modern Production Theory: Efficiency and Productivity**, Boston: Kluwer Academic Publishers, pp.33-55.
- Banker, R., A. Charnes, W.W. Cooper, J.Swarts, and D.A. Thomas (1989), "An Introduction to Data Envelopment Analysis with Some of Its Models and Their Uses",

Research in Governmental and Nonprofit Accounting, Vol.5, 125-163

- Banker, R.D., V.M. Gadh, and W. L. Gorr (1993), "A Monte Carlo Comparison of Two Production Frontier Estimation Methods: Corrected Ordinary Least Squares and Data Envelopment Analysis", **European Journal of Operational Research**, 67, 332-343
- Barla, P. and S. Perelman (1989), "Technical Efficiency in Airlines under Regulated and Deregulated Environments", **Annals and Public and Cooperative Economics**, 60, No.1, pp103-124
- Battese, G.E. and T.J. Coelli (1988), "Prediction of Firm-level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data", **Journal of Econometrics**, 38, 387-399
- Battese, G.E. and T.J. Coelli (1992), "Frontier Production Functions, Technical Efficiency and Panel Data: With Application to Paddy Farmers in India", **The Journal of Productivity Analysis**, 3, 153-169
- Bauer, P.W. (1990), "Recent Developments in the Econometric Estimation of Frontiers", **Journal of Econometrics**, 46(1/2), 39-56
- Bjurek, H., L. Hjalmarsson and F.R. Forsund (1990), "Deterministic Parametric and Nonparametric Estimation of Efficiency in Service Production: a Comparison", **Journal of Econometrics**, 46, pp.213-227
- Bookbinder, James H. and Wendy W. Qu (1993), "Comparing the Performance of Major American Railroads", **Journal of Transportation Research Forum**, Vol.33, No.1, 1993
- Borger, B.D. (1991), "Hedonic versus Homogeneous Output Specifications of Railroad Technology: Belgian Railroads 1950-1986", **Transportation Research**, Vol 25A, No.4, 227-238
- Bowlin, W.F., A. Charnes, W.W. Cooper and H.D. Sherman (1985), "Data Envelopment Analysis and Regression Approaches to Efficiency Estimation and Evaluation", **Annals of Operations Research**, 2, pp.113-138.
- Bruning, E.R. (1991), "Market Liberalization and Operating Efficiency in the International Aviation Industry", **International Journal of Transport Economics**, Vol.XVIII, No.3, October, pp259-274
- Caves, D.W., L.R. Christensen, and W.E. Diewert (1982), "Multilateral Comparisons of Output, Input, and Productivity Using Superlative Index Numbers", **Economic**

Journal, 92, March, pp73-86

Caves, D.W., L.R. Christensen, and J.A. Swanson (1980), "Productivity in U.S. Railroads, 1951-1974", **Bell Journal of Economics**, Vol.11, Spring 1980, pp166-181.

Caves, D.W., L.R. Christensen, and J.A. Swanson (1981), "Productivity Growth, Scale Economics, and Capacity Utilization in U.S. Railroads, 1955-1974", **American Economic Review**, Vol.71, December 1981, pp994-1002.

Caves, D.W., L.R. Christensen, and M.W. Tretheway (1981), "U.S. Truck Air Carriers, 1972-1977: A Multilateral Comparison of Total Factor Productivity" in T.G. Cowing and R.E. Stevenson eds. **Productivity Measurement in Regulated Industries**, New York: Academic Press, pp47-77

Caves, D.W., L.R. Christensen, and M.W. Tretheway (1984), "Economies of Density versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ", **Rand Journal of Economics**, Vol.15, No.4, Winter, pp471-489

Caves, D.W., L.R. Christensen, M.W. Tretheway, and R.J. Windle (1987), "An Assessment of the Efficiency Effects of U.S. Airline Deregulation via an International Comparison" in E.E. Bailey ed. **Public Regulation: New Perspectives on Institutions and Policies**, Cambridge, Mass: MIT Press, pp285-320

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

Charnes, A. and W.W. Cooper (1985), "Preface to Topics in Data Envelopment Analysis", **Annals of Operations Research**, 2, pp.59-94

Christensen, L.R. and D.W. Jorgensen (1969), "The Measurement of U.S. Real Capital Input, 1929-1967", **The Review of Income and Wealth**, Series 15, No.1, pp293-320

Christensen, L.R. and D.W. Jorgensen (1970), "U.S. Real Product and Real Factor Input, 1929-1967", **The Review of Income and Wealth**, Series 16, No. 1, pp19-50

Christensen, L.R., D.W. Jorgensen and L.J. Lau (1971), "Conjugate Duality and the Transcendental Production Function", **Econometrica**, July 1971, 255-256.

Christensen, L.R., D.W. Jorgensen and L.J. Lau (1973), "Transcendental Logarithmic Production Frontiers", **Review of Economics and Statistics**, February, 28-45

Coelli, T.J. (1991), "Maximum-Likelihood Estimation of Stochastic Frontier Production Functions with Time-Varying Technical Efficiency Using the Computer Program,

- FRONTIER Version 2.0", working paper No.57, Department of Econometrics, University of New England. Armidale, NSW, Australia
- Compagnie, I., H. Gathon and P. Pestieau (1991), "Autonomy and Performance in Public Enterprises: the case of Railways and Postal Services", a paper presented at CIRIEC Seminar **Public Versus Private Enterprises: In Search of the Real Issues**, Liège, April 4-5, 1991.
- Cooper. W.W. and A. Gallegos (1991), "A Combined DEA-Stochastic Frontier Approach to Latin American Efficiency Evaluation", The University of Texas at Austin, May 1991
- Corbo, V. and J. de Melo (1986), "Measuring Technical Efficiency: A Comparison of Alternative Methodologies with Census Data", in A. Dogramaci (ed) **Measurement Issues and Behaviour of Productivity Variables**, Boston: Kluwer Nijhoff Publishing, pp.1-39.
- Cornwell, C., P. Schmidt, and R.C. Sickles (1990), "Production Frontiers with Cross-Sectional and Time-Series Variation in Efficiency Levels", **Journal of Econometrics**, 46(1/2), 185-200
- Deprins, D., L. Simar, and H. Tulken (1984), "Measuring Labour Inefficiency in Post Offices", M. Marchand, P. Pestieau, and H. Tulken (eds), **The Performance of Public Enterprises: Concepts and Measurements**, Amsterdam: North-Holland, pp243-267
- Deprins, D. and L. Simar (1989), "Estimating Technical Inefficiencies with Correction for Environmental Conditions with an Application to railway companies", **Annals of Public and Cooperative Economics**, 1989, pp81-101.
- Diewart, W.E. and C. Parkan (1983), "Linear Programming Tests of Regularity Conditions for Production Functions", in W. Eichhorn, R. Henn, K. Neumann and R.W. Shephard eds. **Quantitative Studies on Production and Prices**, Physica Verlag, pp131-158
- Distexhe, V. and S. Perelman (1993), "Technical Efficiency and Productivity Growth in an Era of Deregulation: the Case of Airlines", a paper presented at the **Third European Workshop on Efficiency and Productivity Measurement**, CORE, Belgium, October
- Doganis, R. (1991), **Flying Off Course: the Economics of International Airlines**, London, UK: HarperCollinsAcademic
- ECMT (1993), **Privatisations of Railways**, the Report of the Ninetieth Round Table on

Transport Economics, February, Paris, France

Encaoua, D. (1991), "Liberalizing European Airlines: Cost and Factor Productivity Evidence", International Journal of Industrial Organization, 9, pp109-124

European Conference of Ministers of Transport (1989), Statistical Trends in Transport 1965-1988, Paris, France, 1993

Färe, R., S. Grosskopf, and C.A. K. Lovell (1985), The Measurement of Efficiency of Production, Kluwer-Nijhoff, 1985

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994), "Productivity Growth, Technical Progress, and Efficiency Change in Industrialized Countries", The American Economic Review, Vol. 84, No.1, March, pp67-83

Farrell, M.J. (1957), "The Measurement of Productive Efficiency", Journal of the Royal Statistical Society, Series A (General), 120, Part III, 253-281

Farrell, M.J. and M. Fieldhouse (1962), "Estimating Efficient Production under Increasing Returns to Scale", Journal of the Royal Statistical Society, Series A (General), 125, part II, 252-267

Fecher, F., D. Kessler, S. Perelman and P. Pestieau (1993) "Productive Performance of the French Insurance Industry", The Journal of Productivity Analysis, 4, 77-93

Ferrier, G.D. and C.A.K. Lovell (1990), "Measuring Cost Efficiency in Banking: Econometric and Linear Programming Evidence", Journal of Econometrics 46 (1990), pp.229-245

Försund, F.R. and L. Hjalmarsson (1979), "Frontier Production Functions and Technical Progress: A Study of General Milk Processing in Swedish Dairy Plants", Econometrica 47 (4), pp.893-900

Försund, F.R., C.A.K. Lovell and P. Schmidt (1980), "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement", Journal of Econometrics, 13, 5-25

Försund, F.R. (1992), "A Comparison of Parametric and Non-parametric Efficiency Measures: the Case of Norwegian Ferries", The Journal of Productivity Analysis, Vol.4, No. 1/2, June 1992, pp.25-43

Forsyth, P.J., R.D. Hill, and C.D. Trengove (1986), "Measuring Airline Efficiency", Fiscal Studies, Vol.7, No. 1, February, pp61-81

- Freeman, K.D., T.H. Oum, M.W. Tretheway, W.G. Waters II (1987), **The Growth and Performance of the Canadian Transcontinental Railways 1956-1981**, Centre for Transportation Studies, University of British Columbia, Vancouver, B.C, Canada.
- Gathon, H. and S. Perelman (1990), "Measuring Technical Efficiency in European Railways: a Panel Data Approach", an unpublished paper, University de Liege.
- Gillen, D.W., T.H. Oum and M.W. Tretheway (1985), **Airline Cost and Performance: Implications for Public and Industry Policies**, Centre for Transportation Studies, University of British Columbia, Vancouver, B.C., Canada
- Gillen, D.W., T.H. Oum and M.W. Tretheway (1990), "Airline Cost Structure and Policy Implications", **Journal of Transport Economics and Policy**, Vol. XXIV, No.2, May, pp9-34
- Gong, B.-H. (1987), **Finite Sample Evidence on the Performance of Stochastic Frontier and Data Envelopment Models in the Estimation of Firm-Specific Technical Efficiency Using Panel Data**, unpublished Ph.D dissertation, Rice University, Houston, Texas
- Gong, B.-H. and R.C. Sickles (1989), "Finite Sample Evidence on the Performance of Stochastic Frontier Models Using Panel Data", **The Journal of Productivity Analysis**, 1, pp.229-261
- Gong, B.-H. and R.C. Sickles (1992), "Finite Sample Evidence on the Performance of Stochastic Frontiers and Data Envelopment Analysis Using Panel Data", **Journal of Econometrics**, 51, pp259-284
- Good, D.H. and E.L. Rhodes (1991), "Productive Efficiency, Technological Change and the Competitiveness of U.S. Airlines in the Pacific Rim", **Journal of the Transportation Research Forum**, Vol.31, No.2, pp347-358
- Good, D.H., M.I. Nadiri, L.H. Röller, and R.C. Sickles (1993), "Efficiency and Productivity Growth Comparisons of European and U.S. Air Carriers: A First Look at the Data", **The Journal of Productivity Analysis**, 4, pp115-125
- Grabowski, R. and S. Mehdian (1990), "Efficiency of the Railroad Industry: A Frontier Production Function Approach", **Quarterly Journal of Business and Economics**, Vol.29, Issue 2, Spring, pp26-42
- Greene, W.H. (1980a), "Maximum Likelihood Estimation of Econometric Frontier Functions", **Journal of Econometrics** 13, pp.27-56

- Greene, W.H. (1980b), "On the Estimation of a Flexible Frontier Production Model", **Journal of Econometrics** 13, pp.101-115
- Greene, W.H. (1990), "A Gamma Distributed Stochastic Frontier Model", **Journal of Econometrics**, 46,
- Greene, W.H. (1993), "The Econometric Approach to Efficiency Analysis", in H.O. Fried, C.A. Knox Lovell, and S.S. Schmidt eds. **The Measurement of Productive Efficiency: Techniques and Applications**, Oxford University Press, 68-119
- Guilkey, D. K. and C.A. K. Lovell (1980), "On the Flexibility of the Translog Approximation", **International Economic Review**, Vol. 21, No.1 , February, 1980, 137-147
- Guilkey, D. K., C.A. K. Lovell and R. C. Sickles (1983), "A Comparison of the Performance of Three Flexible Functional Forms", **International Economic Review**, Vol. 24, No.3, October 1983, 591-616
- Hanoch, G. (1971), "CRESH Production Functions" ,**Econometrica**, Vol.39, No.5, September 1971, 695-712
- IATA, **World Air Transport Statistics**, various issues, London, UK
- International Union of Railways, **International Railways Statistics**, various issues
- Jackson, C. (1991), "European Railway Performance 1990", **Railway Gazette International**, July 1991, pp483-484.
- Jackson, C. (1992), "Swedes Ahead in European Rail Productivity", **Railway Gazette International**, July 1992, pp479-481.
- Jha, R. and B.S. Sahni (1992), "Towards Measuring Airline Technical Inefficiency: the Case of Canadian Airlines Industry", **International Journal of Transport Economics**, Vol. XIX, No. 1, February, pp45-59
- Jha, R. and S.P. Singh (1994), "Inter-temporal and Cross-section Variations in Technical Efficiency in the Indian Railways", **International Journal of Transport Economics**, Vol.XXI, No.1, February, pp57-73
- Jondrow, J., C.A.K. Lovell, I. Materov and P. Schmidt (1982), "On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function", **Journal of Econometrics** 19, pp.233-238

- Kalirajan, K. (1990), "On Measuring Economic Efficiency", **Journal of Applied Econometrics**, 5(1), 75-86
- Keeler, J.P. and J.P. Formby (1994), "Cost Economies and Consolidation in the U.S. Airline Industry", **International Journal of Transport Economics**, Vol.XXI, No.1, February 1994, pp21-45
- Kopp, R. (1981), "The Measurement of Productive Efficiency: A Reconsideration", **The Quarterly Journal of Economics**, August 1981, 477-503
- Kopp, R. and W.E. Diewert (1982), "The Decomposition of Frontier Cost Function Deviations into Measures of Technical and Allocative Efficiency", **Journal of Econometrics**, 19, pp.319-331
- Kumbhakar, S. (1987), "The Specification of Technical and Allocative Inefficiency in Stochastic Production and Profit Frontiers", **Journal of Econometrics**, 34(3), 335-348
- Kumbhakar, S. (1989), "Modelling Technical and Allocative Inefficiency in Translog Production Function", **Economics Letters**, 31(2), 119-124
- Lee, Y. H., and P. Schmidt (1993), "A Production Frontier Model with Flexible Temporal Variation in Technical Inefficiency", in H.O. Fried, C.A. Knox Lovell, and S.S. Schmidt eds. **The Measurement of Productive Efficiency: Techniques and Applications**, Oxford University Press, 237-255
- Li, D. (1991), **Alternative Approaches to Technical Efficiency Estimation: A Comparative Study**, unpublished Ph.D dissertation, Department of Economics, McMaster University, Hamilton, Ont. ,Canada
- Loeb, S. G. (1994), **An Analysis of the Influence of Economic Regulation on International Air Carrier Efficiency**, unpublished Ph.D. dissertation, Graduate School of Management, Kent State University, May
- Loeb, S., E. Bruning, and M. Hu (1994), "Government's Role in a Post-Deregulated Environment: the Case of International Aviation", **Proceedings of the 29th Canadian Transportation Research Forum**, Victoria, B.C., May
- Lovell, C.A. K. and P. Schmidt (1988), "A Comparison of Alternative Approaches to the Measurement of Productive Efficiency", in A. Dogramaci and R. Färe (eds) **Applications of Modern Production Theory: Efficiency and Productivity**, Boston: Kluwer Academic Publishers, pp.3-32.

- Lovell, C.A.K (1993), "Production Frontiers and Productive Efficiency", in H.O. Fried, C.A. Knox Lovell, and S.S. Schmidt eds. **The Measurement of Productive Efficiency: Techniques and Applications**, Oxford University Press, 3-67
- Madala, G.S. (1979) " A Note on the Form of the Production Function and Productivity", in **Measurement and Interpretation of Productivity**, National Research Council, Washington, D.C., pp309-317
- Marsten, R. E. (1981), "The Design of the XMP Linear Programming Library", **Transactions on Mathematical Software**, Vol.7, No.4, December 1981
- McGeehan, H. (1993), "Railway Costs and Productivity Growth", **Journal of Transport Economics and Policy**, Vol. XXVII, No.1, January 1993, 19-32
- McMullen, B. S. and H. Tanaka (19**), "Structural Differences Between Large and Small U.S. Motor Carriers Following Deregulation: Implications for Market Structure",
- Meeusen, W. and J. van den Broeck (1977), "Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error", **International Economic Review**, 8, 435-444
- Mendoza, M.N. (1989), **Essays in Production Theory: Efficiency Measurement and Comparative Statics**, unpublished Ph.D dissertation, Department of Economics, The University of British Columbia, Vancouver, B.C. Canada
- Mensah, Y.M. and S.H. Li (1993), "Measuring Production Efficiency in a Not-for-Profit Setting: An Extension", **The Accounting Review**, Vol .68, No. 2 January 1993, pp66-88
- Nash, C. (1981), "Government Policy and Rail Transport in Western Europe", **Transport Review** 1981, Vol.1, No.3, pp225-250.
- Nash, C. (1985), European Railway Comparisons - What can we learn', **International Railway Economics**, ed. by K.J. Button and D.E. Pitfield, Gower Publishing, 1985, pp237-269.
- Nerlove, M. (1965), **Estimation and Identification of Cobb-Douglas Production Functions**, Chicago: Rand McNally & Company
- Nerlove, M. (1971), "Further Evidence on the Estimation of Dynamic Economic Relations from A Time Series of Cross Sections", **Econometrica**, Vol.39, No.2, 359-382
- OECD (1992), **National Accounts: Main Aggregates**, Vol.I, Department of Economics and

Statistics, Paris, France

- Olson, J.A., P. Schmidt and D. W. Waldman (1980), "A Monte Carlo Study of Estimators of the Stochastic Frontier Production Function", **Journal of Econometrics**, 13, pp.67-82
- Oum, T.H. and C. Yu (1991), **An International Comparison of the Economic Efficiency of Passenger Railway Systems**, a report prepared for the Research Division, Royal Commission on National Passenger Transportation, RES 91-21, December 1991
- Oum, T.H. and C. Yu (1994), "Economic Efficiency of Railways and Implications for Public Policy: A Comparative Study of the OECD Countries' Railways", **The Journal of Transport Economics and Policy**, Vol.XXVIII, No.2, May, pp121-138.
- Perelman, S. and P. Pestieau (1988), "Technical Performance in Public Enterprises: A Comparative Study of Railway and Postal Services", **European Economic Review**, 32 (1988), pp432-441.
- Petersen, N.C. (1990), "Data Envelopment Analysis on a Relaxed Set of Assumptions", **Management Science**, Vol. 36, No.3, March, pp305-314
- Phillips, D., A. Ravindran, and J. Solberg (1987), **Operations Research: Principles and Practice**, John Wiley & Sons (New York), 1987.
- Ray, S. C. (1991), "Resource-Use Efficiency in Public Schools: A Study of Connecticut Data", **Management Science**, 37(12), 1620-1628
- Ray, S. and X. Hu (1993), "A Nonparametric Decomposition of the Malmquist Productivity Index: A Study of Airlines' Data", a paper presented at the **Third European Workshop on Productivity**, CORE, Belgium, October
- Reifschneider, D. and R. Stevenson (1989), "Systematic Departures from Minimum Cost Performance: A Framework for the Analysis of Firm Inefficiency", Discussion paper No. 178, The Institute of Social and Economic Research, Osaka University, Osaka, Japan
- Richmond, J. (1974), "Estimating the Efficiency of Production", **International Economic Review**, 15(2), pp.515-521
- Röller, L.-H. and R. Sickles (1994), "Competition, Market Niches, and Efficiency: A Structural Model of the European Airline Industry", Discussion Paper No. 926, March 1994, Center for Economic Policy Research, London, England

- Rubinstein, R.Y. (1981), Simulation and the Monte Carlo Method, John Wiley & Sons (New York), 1981
- Schmidt, P. (1976), "On the Statistical Estimation of Parametric Frontier Production Functions", Review of Economics and Statistics, 58 (2), pp.238-239
- Schmidt, P. and C.A.K. Lovell (1979), "Estimating Technical and Allocative Inefficiency Relative to Stochastic Production and Cost Frontiers", Journal of Econometrics, 9(3), 343-366
- Schmidt, P. (1985-86), "Frontier Production Functions", Econometric Reviews, 4(2), 289-328
- Schmidt, P and R. Sickles (1984), "Production Frontiers and Panel Data", Journal of Business and Economic Statistics, 2, pp.367-374
- Schwier, C., J. Jones and A. Pignal (1990), "Canadian Regional Railway Passenger Services: Was There Ever Any Potential for Financial Improvement", Proceedings of 25th Canadian Transportation Research Forum, Saskatoon, June 6-8, 1990, pp254-267.
- Seiford, L.M. and R.M. Thrall (1990), "Recent Developments in DEA, the Mathematical Programming Approach to Frontier Analysis", Journal of Econometrics 46, pp.7-38
- Seitz, W.D. (1971), "Productive Efficiency in the Steam-electric Generating Industry", Journal of Political Economy, 79, No.4, July/August, 878-886
- Sexton, T.R., R. H. Silkman, and A.J. Hogan (1986), "Data Envelopment Analysis: Critique and Extensions", in R.H. Silkman ed. Measuring Efficiency: An Assessment of Data Envelopment Analysis, New Directions for Program Evaluation, No. 32, San Francisco: Jossey-Bass, 1986, pp73-105
- Sickles, R.C. and M.L. Streitwieser (1992), "Technical Inefficiency and Productivity Decline in the U.S. Interstate Natural Gas Pipeline Industry under the Natural Gas Policy Act", The Journal of Productivity Analysis, Vol.3, No.1/2, June 1992, pp.119-133
- Simar, L. (1992), "Estimating Efficiencies from Frontier Models with Panel Data: A Comparison of Parametric, Nonparametric and Semi-parametric Methods with Bootstrapping", The Journal of Productivity Analysis, Vol.3, No.1/2, June 1992, pp.171-203
- Stevenson, R.E. (1980) "Likelihood Functions for Generalized Stochastic Frontier Estimation", Journal of Econometrics, 13, pp.57-66

- Summers, R. and A. Heston (1991), "The Penn World Table (Mark 5): An Expanded Set of International Comparisons, 1950-1988", Quarterly Journal of Economics, Vol. 106, Issue 2, May, pp327-368
- Thanassoulis, E. (1993), "A Comparison of Regression Analysis and Data Envelopment Analysis as Alternative Methods for Performance Assessments", Journal of the Operational Research Society, Vol. 44, No. 11, 1129-1144
- Timmer, C.P. (1971), "Using a Probabilistic Frontier Production Function to Measure Technical Efficiency", Journal of Political Economy, 79, 775-794
- Thompson, L.S., W.G. Wood, and F. Lures (1991), "World Railway Performance Update", Developing Railways, 1991, pp8-16.
- Thompson, L.S. and J.M. Fraser (1993), "World Bank Survey of Railway Performance", Developing Railways, 1993, pp5-13.
- Thursby, J.G. and C.A.K. Lovell (1978), "An Investigation of the Kmenta Approximation to the CES Function", International Economic Review, Vol.19, No.2, June
- van de Broek, J., F.R. Forsund, L. Hjalmarsson and W. Meeusen (1980), " On the Estimation of Deterministic and Stochastic Frontier Production Functions: a Comparison", Journal of Econometrics, 13, pp.117-138
- Wales, T.J. (1977), "On the Flexibility of Flexible Functional Forms", Journal of Econometrics, 5, 183-193
- White, K.J., S. D. Wong, D. Whistler and S.A. Haun (1990), SHAZAM User's Reference Manual Version 6.2, McGraw-Hill, 1990
- Windle, R.J. (1991), "The World's Airlines: A Cost and Productivity Comparison", Journal of Transport Economics and Policy, Vol.XXV, No.1, January, pp31-49
- Zieschang, K.D. (1983), "A Note on the Decomposition of Cost Efficiency into Technical and Allocative Components", Journal of Econometrics, 23, pp.401-405

Appendix A

Description of Computer Programs

A.1 Computer Program for DEA

In actual computation, the dual formulation of the DEA method is more tractable than the primal. In the primal formulation, the constraints are indexed on all DMUs, while in the dual formulation the constraints are indexed on inputs and outputs and sum over DMUs. For the problems in this thesis, the number of inputs and outputs is far less than the number of DMUs. Phillips, Ravindran and Solberg (1987) have shown that the computational efficiency of the simplex method falls with increases in the size of the constraint set. Hence the dual program with only $(m+s)$ constraints on inputs and outputs is computed in preference to its primal with $n+1$ constraints.

Implementation of the DEA method in the Monte Carlo experiments is illustrated by the following sample FORTRAN program¹ which follows the one step non-Archimedean model procedure where ϵ , the non-Archimedean scalar, is given the value of 10^{-6} . The DEA method requires the solution of linear programming problems. The user-callable subroutines from the XMP Mathematical Programming Library (Marsten, 1981) are used. Note that the linear programming problem is solved for each of the n DMUs.

¹ The computer codes for BCC model and for railway application requires only minor changes, so they are not listed here.

```

C
C   THIS IS THE SAMPLE FORTRAN CODE FOR THE DEA METHOD
C   USING THE DUAL FORMULATION OF THE CCR MODEL(1978)
C       For use on a P.C.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      CHARACTER*24 FIN, FOUT, INX

      WRITE(*,999)
      READ(*,998) FIN
      WRITE(*,997)
      READ(*,998) FOUT

      WRITE(*,996)
      READ(*,998) INX

999   FORMAT(10X,' INPUT DATA FILE NAME=')
997   FORMAT(10X,' OUTPUT FILE NAME=')
996   FORMAT(10X,' INDEX FILE NAME=')
998   FORMAT(A24)

      OPEN(5, FILE=FIN, STATUS='OLD')
      OPEN(6, FILE=FOUT)
      OPEN(7, FILE=INX)

C
      CALL MATRX
      STOP
      END

C
C   SUBROUTINE MATRX
C
C   THE PURPOSE OF THIS SUBROUTINE IS TO GENERATE XMP MATRIX
C   FROM THE ORIGINAL DATA FILE WHICH IS ORGANIZED VARIABLE BY
C   VARIABLE,
C
C       DEFINITION OF VARIABLES
C           Y:      THE OBSERVED OUTPUT LEVEL
C           X1:     INPUT 1
C           X2:     INPUT 2
C           X3:     INPUT 3
C
C
C   DECLARATIONS FOR THE PROBLEM DATA.
C
      REAL DATA(4,250)
      REAL RATIO(250)

```

```

      REAL WEIGHT(4,250)
C
C
C   DECLARATIONS FOR THE XMP VARIABLES AND ARRAYS.
C
      DOUBLE PRECISION B(4), BASCB(4), BASLB(4), BASUB(4),
X   BOUND, CAND(4,6), CANDCJ(6), CJ, COLA(4),
X   LJ, MEMORY(20000), UJ, UZERO(4), XBZERO(4), YQ(4),
X   Z, ZTEMP, UTEMP
      DOUBLE PRECISION BLOW(4)
      INTEGER BNDTYP, COLLEN, COLMAX,
X   ERROR, FACTOR, IOERR, IOLOG, ITER,
X   ITER1, ITER2, LENMA, LENMI, LENMY, LOOK, M, MAPA(20),
X   MAPI(20), MAXA, MAXM, MAXN, N, NTYPE2, PICK, PRINT, TERMIN,
X   UNBDDQ
C THE NEXT STATEMENT SHOULD SPECIFY HALF-WORDS IF POSSIBLE.
      INTEGER*2 CAND(6), CANDI(4,6), CANDL(4), BS(4),
X   COLI(4), ROWTYP(4), STATUS(259), MINMAX, BASIS(4)
C
C   MINMAX = -1 MEANS MINIMIZATION
C             +1 MEANS MAXIMIZATION
C
C       Note that the convention in XMP is maximization
C
C
C   INTEGER IDNO,K,I,J
C
C   SET VALUES FOR I/O UNITS
C
      IOIN=5
      IOLOG=6
      IOERR=6
C
C
C   INITIALIZE XMP PARAMETERS.
C
      MINMAX=1
      MAXA=1011
      MAXM=4
      MAXN=259
      COLMAX=4
      PICK=6
      LOOK=50
      FACTOR=50
      LENMY=20000
C
C   DO 310 I=1,N
310 STATUS(I)=0
C

```

```

C   UP TO HERE WE FINISH DEFINING THE DATA STRUCTURE
C
C
C   THE FOLLOWING SUBROUTINE IS TO SET UP A STARTING BASIS
C
C       CALL XSLACK(B,BASCB,BASIS,BASLB,BASUB,BLOW,
X       BNDTYP,BOUND,COLA,COLI,COLMAX,IOERR,
X       LENMA,LENMI,LENMY,M,MAPA,MAPI,MAXM,MAXN,MEMORY,
X       N,ROWTYP,STATUS,UZERO,XBZERO,Z)
C
C   THE FOLLOWING SUBROUTINE IS TO SOLVE THE LP BY THE
C   PRIMAL SIMPLEX METHOD
C
C       CALL XPRIML(B,BASCB,BASIS,BASLB,BASUB,BNDTYP,BOUND,
X       CAND,CANDA,CANDCJ,CANDI,CANDL,COLA,COLI,COLMAX,
X       FACTOR,IOERR,IOLOG,ITER1,ITER2,LENMA,LENMI,LENMY,
X       LOOK,M,MAPA,MAPI,MAXM,MAXN,MEMORY,N,NTYPE2,PICK,
X       PRINT,STATUS,TERMIN,UNBDDQ,UZERO,XBZERO,YQ,Z)
C       WRITE(6,650)
650  FORMAT(5X,'XPRIMAL')
C
C
C   "Z" IS THE VALUE OF THE OBJECTIVE FUNCTION.
C
C       ZTEMP=Z
C       IF (MINMAX .EQ. -1) ZTEMP=-ZTEMP
C       RATIO(K)=1/ZTEMP
C
C   PRINT OUT SOLUTIONS: IT INCLUDES THE CURRENT BASIC
C   SOLUTION AND THE OBJECTIVE FUNCTION VALUE
C
C       CALL XPRINT(BASIS,BNDTYP,BOUND,IOERR,IOLOG,
X       LENMA,LENMY,M,MAPA,MAXM,MAXN,MEMORY,N,NTYPE2,
X       STATUS,XBZERO,ZTEMP)
C
C   PRINT OUT THE VALUES OF THE DUAL VARIABLES
C
C       WRITE(6,910)
910  FORMAT(1H0///5X,28HVALUES OF THE DUAL VARIABLES)
C       WRITE(6,911)
911  FORMAT(1H0,5X,10HCONSTRAINT,12X,14H DUAL VARIABLE)
C
C   FLIP THE SIGN IF WE ARE MINIMIZING
C
C       WEIGHT(1,K)=-UZERO(1)
C       DO 300 I=2,M
C       UTEMP=UZERO(I)
C       WEIGHT(I,K)=UTEMP
C       IF (MINMAX.EQ.-1) WEIGHT(I,K)=-UTEMP

```

```

300 CONTINUE
      WRITE(7,799) K, RATIO(K),WEIGHT(1,K), WEIGHT(2,K),
      X WEIGHT(3,K), WEIGHT(4,K)
799 FORMAT(2X,I8,5F12.6)
C
C
C   END OF ONE ROUND
C
800 CONTINUE
C
      WRITE(*,803) KN
803 FORMAT(2X," NUMBER OF SAMPLE", F8.0)
999 CONTINUE
C
C   END OF THE SUBROUTINE
C
      RETURN
      END

```

A.2 The Program for Stochastic Frontier Model

The stochastic frontier models are estimated using the program *FRONTIER* version 2.0 developed by Coelli (1991). The program provides maximum likelihood estimation of a wide variety of stochastic frontier production function model formulations. The program follows a three-step procedure in estimating the maximum likelihood estimates of the parameters of a stochastic frontier production function. The first step obtains the OLS estimates of the production function; then the second step conducts a two-phase grid search of the parameters in the log likelihood function with the OLS estimators excepting the intercept; finally the values selected in the grid search are used as starting values in an iterative procedure using Davidon-Fletcher-Powell Quasi-Newton method to obtain the final (approximate) maximum-likelihood estimates.

Appendix B

Data in Monte Carlo Experiments

B.1 Data Generation

The inputs, the exogenous variable and the error terms are generated using the random number generators in SHAZAM version 6.2.

The inputs X are drawn from lognormal distributions with the following p.d.f.:

$$f_x(x) = \begin{cases} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\tau^2}\right], & 0 \leq x \leq \infty \\ 0 & \text{otherwise} \end{cases} \quad \text{B.1}$$

$$\mu = 3, \quad \tau = 1$$

Since SHAZAM version 6.2 has only uniform random number generator and normal random number generator, the lognormal random numbers are generated using the following algorithm²:

1. Generate x_0 from $N(0,1)$
2. $W \leftarrow \mu + \tau x_0$
3. $X \leftarrow e^w$
4. Deliver X .

The error term for inefficiency u is from a half-normal distribution which takes absolute values from $N(0, 0.36)$.

² Rubinstein (1981) gives a very good description about the methods for generating random variables and random vectors from different probability distributions.

The values for inputs and inefficiency term are fixed throughout the experimentation. Table B.1 lists the summary statistics for X and e^u . Note that the values listed here are the averages from 25 replications.

Table B.1 Summary Statistics for Variables in Monte Carlo Experiments				
	Means	St.dev.	Minimum	Maximum
X1	34.04	42.675	1.2851	358.48
X2	33.45	44.948	1.1938	410.64
X3	33.02	42.172	1.3541	379.39
e^u	0.6561	0.2034	0.1641	0.9979

The values for exogenous variables and the noise term vary from experiment to experiment, and given in Chapter 6.

B. 2 Determination of the Number of Replications

As mentioned in Chapter 5, the more replications, the better the results. However, computation costs increase significantly with the number of replications. Therefore, it is necessary to examine whether or not it makes any significant differences in the simulation results to use different number of replications. In particular, we test the hypothesis that

Appendix B Monte Carlo Experiments

there are no significant differences among the results with 25, 50, and 100 replications, for each of the three alternative methods, in terms of mean efficiency levels and correlation coefficients. This is accomplished by doing three sets of experiments, with 25, 50 and 100 replications respectively, and comparing the results between the different sets of experiments. Here the underlying production technology is specified by functions (5.1), (5.4) and (5.5) with $\theta=0$, $\gamma=1$, $\delta_1=\delta_2=0.3$, $\delta_3=0.4$, $\alpha=0$, and $\sigma_v = 0.15$. With each experiment, the sample size is set at 100, and ρ is set at -0.25.

In comparing the results from different sets of the experiments, we consider each set of experiments as simulating a sample representing a particular population. Consequently, hypotheses tests are conducted about the differences among the means of three populations. Each replication is considered as a sample observation from that particular population, thus we have three separate samples with 25, 50 and 100 observations respectively, representing three different populations.

B.2.1 Stochastic Frontier Method

First we look at the average efficiency estimates. The following lists the results from three sets of experiments with 25, 50 and 100 replications, respectively.

$$\begin{array}{ll} \hat{\mu}_{25} = 0.53614 & s_{25} = 0.020749 \\ \hat{\mu}_{50} = 0.54150 & s_{50} = 0.025803 \\ \hat{\mu}_{100} = 0.54231 & s_{100} = 0.034441 \end{array} \quad \text{B.2}$$

where μ_n is the mean of average efficiency estimates with n replications, s_n is the corresponding standard deviations. To determine whether the number of replications makes

any significant differences in the simulation results, we test the following hypothesis:

$$\begin{aligned} H1_0: \mu_{25} &= \mu_{50} \\ H2_0: \mu_{25} &= \mu_{100} \end{aligned} \quad B.3$$

using the test statistic z :

$$z = \frac{(\hat{\mu}_1 - \hat{\mu}_2) - (\mu_1 - \mu_2)}{\sqrt{s^2/n_1 + s^2/n_2}} \quad B.4$$

$$z_{sto\ 1} = \frac{\hat{\mu}_{25} - \hat{\mu}_{50}}{\sqrt{s^2_{25}/25 + s^2_{50}/50}} = \frac{0.53614 - 0.54150}{\sqrt{0.020749^2/25 + 0.025803^2/50}} = -0.97$$

$$z_{sto\ 2} = \frac{\hat{\mu}_{25} - \hat{\mu}_{100}}{\sqrt{s^2_{25}/25 + s^2_{100}/100}} = \frac{0.53614 - 0.54231}{\sqrt{0.020749^2/25 + 0.034441^2/100}} = -1.144$$

since $-1.96 < z_{sto\ 1} < 1.96$, and $-1.96 < z_{sto\ 2} < 1.96$, we can not reject either $H1_0$ or $H2_0$. That is, there is no significant differences in mean efficiency estimates between 25 replications and 50 replications, and between 25 replications and 100 replications.

Next we look at correlation coefficients and rank correlation coefficients. The means of average correlation coefficients and rank correlation coefficients (between the estimated efficiency levels and the true efficiency levels) and the corresponding standard deviations are as follows:

$$\begin{array}{ll}
\hat{C}_{25} = 0.76653 & s_{c25} = 0.34063 \\
\hat{C}_{50} = 0.77842 & s_{c50} = 0.31610 \\
\hat{C}_{100} = 0.79202 & s_{c100} = 0.30153
\end{array}$$

$$\begin{array}{ll}
\hat{R}_{25} = 0.76761 & s_{R25} = 0.33988 \\
\hat{R}_{50} = 0.77763 & s_{R50} = 0.31725 \\
\hat{R}_{100} = 0.79273 & s_{R100} = 0.30202
\end{array}$$

Following the same procedure used for the mean efficiency estimates, the following hypotheses are tested:

$$\begin{array}{ll}
H1_0: C_{25} = C_{50} & R_{25} = R_{50} \\
H2_0: C_{25} = C_{100} & R_{25} = R_{100}
\end{array} \quad \text{B.5}$$

Using B.4 , the test statistics are computed as:

$$\begin{array}{ll}
z_{C1} = -0.1459, & z_{C2} = -0.342 \\
z_{R1} = -0.123, & z_{R2} = -0.3377
\end{array}$$

All of the four test statistics fall within the interval of (-1.96, +1.96), thus the hypotheses stated in B.5 can not be rejected. There is no significant differences, in terms of average correlation coefficients and rank coefficients, between 25 replications and 50 replications, and between 25 replications and 100 replications.

From the above test results, we can conclude that it is sufficient to use 25 replications for the stochastic frontier models.

Similar hypothesis tests are conducted for the deterministic frontier method and the DEA. These tests are summarized in the following tables.

B.2.1 Deterministic Frontier Method

Table B.2 Determination of Number of Replication Experiment Results by Deterministic Frontier Method			
	No. of Replications	Sample Mean	Standard Deviation
Average Efficiency	25	0.42031	0.041606
	50	0.41362	0.063023
	100	0.40822	0.058574
Correlation	25	0.87340	0.032179
	50	0.87312	0.034146
	100	0.87182	0.039585
Rank Correlation	25	0.89287	0.033548
	50	0.89845	0.028357
	100	0.89767	0.032981

Table B.3 Hypothesis Test for Deterministic Frontier Method		
	Null Hypothesis	Test Statistic
Average Efficiency	$\mu_{25} = \mu_{50}$	$z_1 = 0.5487$
	$\mu_{25} = \mu_{100}$	$z_2 = 1.1881$
Correlation	$C_{25} = C_{50}$	$z_{C1} = 0.0348$
	$C_{25} = C_{100}$	$z_{C2} = 0.2091$
Rank Correlation	$R_{25} = R_{50}$	$z_{R1} = -0.7139$
	$R_{25} = R_{100}$	$z_{R2} = -0.6420$

As one can see from Table B.3, all test statistics are within the interval of $(-1.96, +1.96)$, that is the null hypotheses can not be rejected. Therefore, the use of 25 replications for the deterministic frontier methods will not give different results from using 50 replications, or 100 replications.

B.2.1 Data Envelopment Analysis Method

Table B.4 Determination of Number of Replications Experiment Results by Data Envelopment Analysis			
	No. of Replications	Sample Mean	Standard Deviation
Average Efficiency	25	0.61645	0.028649
	50	0.61954	0.032030
	100	0.62000	0.030646
Correlation	25	0.79608	0.051961
	50	0.79139	0.067507
	100	0.79254	0.060102
Rank Correlation	25	0.77979	0.059231
	50	0.77669	0.073567
	100	0.77907	0.065774

Table B.5
Hypothesis Test for Data Envelopment Analysis

	Null Hypothesis	Test Statistic
Average Efficiency	$\mu_{25} = \mu_{50}$	$z_1 = -0.4231$
	$\mu_{25} = \mu_{100}$	$z_2 = -0.5463$
Correlation	$C_{25} = C_{50}$	$z_{C1} = 0.3323$
	$C_{25} = C_{100}$	$z_{C2} = 0.2949$
Rank Correlation	$R_{25} = R_{50}$	$z_{R1} = -0.1966$
	$R_{25} = R_{100}$	$z_{R2} = -0.0531$

Again all test statistics fall within the interval of $(-1.96, +1.96)$. Therefore, it is sufficient to use 25 replications for the DEA method.

In summary, the sample results indicate that there is no significant difference between the results using 25 replications and 50 or 100 replications. Therefore, it is sufficient to conduct 25 replications for all three alternative methods.

B. 3 Specification of Functional Forms of Underlying Production Technology

In the first five sets of Monte Carlo experiments, the underlying production technology is assumed to be constant returns to scale. In the last set of Monte Carlo experiments, non-constant returns to scale and input complementarity are allowed for the underlying production function. Table B.7 lists the related parameter values of the production function. The basic functional form is the CRESH function as specified by equation (5.1). The first three functions are the constant returns to scale technologies used

in the first five sets of experiments. Function 4 to 13 are the functions considered in the last set of experiments.

Table B.6 Parameter Values for Underlying Production Function								
	γ	σ_{12}	σ_{13}	σ_{23}	ρ	ρ_1	ρ_2	ρ_3
1	1	3.03	3.03	3.03	-0.67	-0.67	-0.67	-0.67
2	1	1.33	1.33	1.33	-0.25	-0.25	-0.25	-0.25
3	1	0.33	0.33	0.33	+2.00	+2.00	+2.00	+2.00
4	0.927	-0.277	0.540	0.512	-2.15	+0.85	+0.95	-2.00
5	0.927	0.399	0.454	-0.182	-2.15	-2.00	+1.50	1.20
6	0.929	0.356	0.447	0.508	+1.00	+1.50	+1.20	+0.75
7	0.934	-0.209	0.400	0.454	-2.30	+1.50	+1.20	-2.15
8	0.934	-0.091	0.399	0.453	-1.60	+1.50	+1.20	-1.50
9	1.209	0.399	0.454	-0.182	-1.65	-2.00	+1.50	+1.20
10	1.227	-0.209	0.400	0.454	-1.75	+1.50	+1.20	-2.15
11	1.239	0.356	0.447	0.508	-0.75	+1.50	+1.20	+0.75
12	1.286	-0.277	0.540	0.512	-1.55	+0.85	+0.95	-2.00
13	1.573	-0.091	0.399	0.453	-0.95	+1.50	+1.20	-1.50

γ , and σ_{ij} are the returns to scale and Allen-Uzawa partial elasticities of substitution as specified by equation (5.2) and equation (5.3) in Chapter 5.

Appendix C

Railway Sample Data

C. 1 Sample Firms and Their Characteristics

An annual panel of 19 selected railways for the period of 1978-1989 forms the primary data base for the case study of railways. The sample railways are selected because their data are available in relatively consistent form over the sample period. The sample firms represent different institutional settings and operate in a variety of environments. The names of these railways and the years for which data are collected for this study are listed in Table C.1³.

The management of some of the railways such as British Rail (BR), and Netherlands Railways (NS) enjoy substantial freedom for making strategic and operational decisions without government intervention, provided that the predetermined minimum performance criteria are met. On the other hand, some of the railways such as the Finnish State Railways (VR), Austrian Federal Railways (ÖBB), and Norwegian State Railways (NSB) are subject to strict governmental control. Of the sample railways, only German Federal Railways (DB), Italian State Railways (FS prior to 1985 only) and Danish State Railways (DSB prior to 1986) are run by governmental agencies while French National Railways (SNCF prior to 1982 only) is organized as quasi-public firms⁴. The rest of the railways are organized as public firms (crown corporations).

³ More discussions for some of the railways can be found in Oum and Yu (1991).

⁴ The term "quasi-public firm" is used to indicate mixed ownership by public and private interests.

There are large variations in operating environments across the selected railways. Japanese National Railways (JNR, however since 1987 it is known as JR) enjoys very high

Table C.1

List of Sample Railways, Abbreviations, & Years Used

BR :	British Railways (UK)	1978-1989
CFF:	Swiss Federal Railways (Switzerland)	1978-1989
CFL:	Luxembourg National Railway Company	1978-1989
CH:	Hellenic Railways Organization (Greece)	1978-1987
CIE:	Irish Transport Company	1988-1989
CP:	Portuguese Railways	1978-1989
DB:	Deutsche Bundesbahn-German Federal Railways	1978-1989
DSB:	Danish State Railways (Denmark)	1978-1989
FS:	Italian State Railways	1978-1989
JNR:	Japanese National Railways	1978-1986
NS:	Netherlands Railways	1978-1989
NSB:	Norwegian State Railways	1978-1989
ÖBB:	Austrian Federal Railways	1978-1989
RENFE:	National System of Spanish Railways	1978-1989
SJ:	Swedish State Railways	1978-1989
SNCB:	Belgium National Railways	1978-1989
SNCF:	French National Railways	1978-1989
TCDD:	Turkish Republic State Railways	1978-1989
VR:	Finnish State Railways	1978-1989

passenger traffic density⁵ (9.6 million passengers as an average density in 1987) while NSB has much lower passenger traffic density (528,000 in 1989). Hellenic Railways (CH) of Greece serves relatively long distance passenger traffic, while most other railways have high

⁵ Traffic density is measured in passenger-kilometers per route-kilometers.

percentage of their passenger services in short suburban and commuter traffic (see the average passenger trip length in Table C.2). JNR (Japan) and TCDD (Turkey) enjoy much higher average passenger load per train than the other railways. SNCF (France), VR (Finland) and TCDD (Turkey) exhibit a fair balance between passenger and freight services, while NS (Netherlands) and DSB (Danish State Railways) provide primarily passenger services. CFF-SBB (Swiss) is fully electrified while CH (Greece), CIE (Ireland), DSB (Denmark), and TCDD (Turkey) still operate fuel traction over nearly their entire network.

The sizes of the railways as indicated by length of route range from a 270 kilometer (Table C.3) network of the Luxembourg Railway (CFL) to larger firms such as the French National Railway (SNCF) that serves rail lines of over 30,000 kilometers. Traffic volumes as measured by passenger kilometers, freight tonne kilometers and train kilometers also vary over a wide range. JNR produces over 200 billion passenger kilometers each year while CFL produces only a little over 200 million passenger-kilometers (Table C.3).

Cost recovery conditions vary greatly among the sample railways. Table C.4 presents the overall operating cost recovery ratios and the ratio of direct subsidy to total operating cost. It is noted that the subsidy ratio is not necessarily equal to:

$$(1 - \text{operating cost recovery ratio})$$

This is because most countries do not provide balancing subsidies to their railways⁶. Other railways, such as JNR (Japan), before its reorganization, and DB (Germany), finance part of their deficits by raising debt. Among the 19 OECD railways, the Italian State Railways

⁶ Most of the railways receive payments from governments for specific services. Some railways, such as CH, DSB and NSB, get their operating losses subsidized from the state.

Table C. 2
Railway Characteristics
(Mean Value of the variables during the sample period)

Railways	Countries	Load ^a (Pax)	Density ^b (Pax) (000)	Trip ^c (Km)	%PASS ^d (%)	%ELEC ^e (%)	Load ^f (FRE)	Density ^g (FRE) (000)	Haul ^h (Km)	Pden ⁱ (000)	Fden ^j (000)
BR	UK	93	1862	44	78	23	232	1019	119	20	4
CFF	Switzerland	123	3166	41	73	99	257	2358	157	26	9
CFL	Luxembourg	79	868	21	65	59	424	2253	38	11	5
CH	Greece	125	670	152	81	0	254	295	201	5	1
CIE	Ireland	118	528	57	65	1	137	309	178	4	2
CP	Portugal	194	1608	26	79	13	174	352	254	8	2
DB	Germany	103	1441	38	66	40	303	2142	207	14	7
DSB	Denmark	99	1781	34	83	7	201	742	239	18	4
FS	Italy	173	2467	102	77	55	304	1079	336	14	4
JNR	Japan	372	8950	28	81	40	241	1391	318	24	6
NS	Netherlands	91	3201	44	89	64	243	1080	159	35	5
NSB	Norway	98	527	62	67	58	262	651	112	5	2
ÖBB	Austria	116	1272	45	62	53	298	1880	205	11	6
RENFE	Spain	151	1171	81	68	46	264	918	350	8	3
SJ	Sweden	102	552	84	60	62	410	1443	328	5	4
SNCB	Belgium	93	1678	43	76	47	375	2070	118	18	5
SNCF	France	195	1686	79	61	32	312	1709	338	9	5
TCDD	Turkey	263	757	48	57	3	380	812	496	3	2
VR	Finland	129	527	78	57	21	423	1290	266	4	3

a Average passenger load: number of passengers per train

b Passenger traffic density: passenger-kilometers per route-kilometer

c Average trip length of rail passengers

d Percentage of passenger train-kilometers in total train-kilometers

e Percentage of electrified route mileage in total route mileage

f Average freight load: number of tonnes per train

g Freight traffic density: freight tonne-kilometers per route-kilometer

h Average length of haul of freight traffic

i Passenger train density: passenger train-kilometers per route-kilometer

j Freight train density: freight train-kilometers per route-kilometer

Table C. 3 Railways Route Length and Traffic Volume*						
Railway	Country	Route (KM)	Train-Km (Pax) (000)	Train-Km (Freight) (000)	Pax-Km (000000)	Ton-Km (000000)
BR	UK	16588	362430	77232	33323	16742
CFF	Switzerland	2994	91063	27366	11021	8161
CFL	Luxembourg	272	3083	1402	224	669
CH	Greece	2479	13711	2313	2011	657
CIE	Ireland	1944	9541	4136	1220	560
CP	Portugal	3064	29559	6939	5908	1719
DB	Germany	27045	397437	196029	41144	61109
DSB	Denmark	2344	42410	7350	4649	1677
FS	Italy	16030	236451	65660	44443	18650
JNR	Japan	20341	655095	91164	222670	24752
NS	Netherlands	2828	106664	11479	10164	3108
NSB	Norway	4044	21461	9561	2136	2749
ÖBB	Austria	5641	71297	37562	8445	11849
RENFE	Spain	12565	109853	48214	14715	14048
SJ	Sweden	11022	59171	40620	6060	18532
SNCB	Belgium	3513	71285	21041	6400	9275
SNCF	France	34322	312348	167964	64256	52449
TCDD	Turkey	8430	26959	16690	6844	7564
VR	Finland	5884	22027	17239	3208	7958

* 1989 data

Source: International Railway Statistics

(FS) has the lowest revenue-cost ratio (24% in 1989) while the operating revenue of the JR system (Japan) exceeded costs in 1988. It is worth noting that the most heavily subsidized railways including FS (Italy), SNCB (Belgium), and CH (Greece) provide primarily passenger services with passenger train-kilometers accounting for over 75% of their total train-kilometers.

Table C.4 Cost Recovery Indicators					
Railways	Countries	Revenue/Costs		Subsidy/Costs	
		1980	1989	1980	1989
BR	U.K.	0.74	0.90	0.24	0.18
CFF	Switzerland	0.74	0.73a	0.08	0.12a
CFL	Luxembourg	0.31	0.23	0.66	0.75
CH	Greece	0.54	0.28b	0.47	0.72b
CIE	Ireland	0.68	0.76	0.31	0.25
CP	Portugal	0.46	0.46	0.45	0.29
DB	Germany	0.58	0.60	0.30	0.28
DSB	Denmark	0.68	0.62	0.32	0.38
FS	Italy	0.29	0.24	0.45	0.71
JNR	Japan	0.70	1.10	0.07	0.01
NS	Netherlands	0.56	0.57	0.43	0.45
NSB	Norway	0.72	0.60	0.28	0.40
ÖBB	Austria	0.67	0.64	0.14	0.40
RENFE	Spain	0.55	0.47	0.45	0.48
SJ	Sweden	0.83	0.81c	0.13	0.15c
SNCB	Belgium	0.44	0.42	0.55	0.58
SNCF	France	0.66	0.68	0.30	0.29
TCDD	Turkey	0.45	0.73	0.41	0.13
VR	Finland	0.77	0.72	0.02	0.02

a. The Government has been paying the infrastructure costs of CFF since 1987, on the other hand, CFF pay the government a contribution towards the infrastructure expenses. To be consistent with the earlier data, government payment for infrastructure and CFF's contribution towards infrastructure expenditure are excluded in calculating the financial ratios.

b. for 1988.

c. for 1988. SJ was separated into two organizations SJ and BV in 1989.

Sources: UIC: International Railway Statistics; EC: COM(88) 12, Com(89) 364, 564

C. 2 Source of Data and Description of Variables

The annual data are compiled from the **International Railway Statistics**, railways' annual reports, and the (statistical) Yearbooks of the respective countries as well as publications by the Commissions for the European Community.

There are two sets of variables collected for the railways in order to measure the productive performance of the railways: (1) outputs and inputs, and (2) the variables describing the operating environment and production characteristics of the railways.

Two alternate sets of output measures are considered: (i) revenue output measures (passenger-kilometers and freight tonne-kilometers); and (ii) available output measures (passenger train kilometres and freight train-kilometres). The available output measures indicate essentially the level of capacity supplied while the revenue output measures indicate the level of output consumed by users, and the value they derive from them. These output data are collected from **International Railways Statistics**.

Seven input measures are used: (i) labour; (ii) energy consumption; (iii) ways and structures; (iv) materials; (v) the number of passenger cars; (vi) the number of freight wagons; and (vii) the number of locomotives.

The number of employees from **International Railways Statistics** is used as the measure of labour input as uniform statistics on labour hours are not available for all firms in the sample. Energy input is measured by the total BTU consumed, using the following conversion factors:

Diesel oil : 0.1657 million Btu / imperial gallons

Fuel oil: 0.1801 million Btu/ imperial gallons

Electricity: 3412 Btu/KWH

Coal: 26.2 million Btu/ton

1 ton of diesel oil = 1.571 tons of coal

Railways' energy consumption for 1978-1984 are obtained from **International Railways Statistics**. Energy data of the European railways for 1985-1987 are obtained from **Statistical Trends in Transport 1965-1989**⁷. JNR's energy consumption for 1985-1986 are obtained from **Japan Statistical Yearbook**.

The ways and structures input is determined, using the perpetual inventory method (Christensen and Jorgenson, 1969), by the amount of land and infrastructure capital stock reported by **International Railways Statistics**. The land and infrastructure assets are first converted into U.S. currency using the Purchasing Power Parity (PPP) index for GDP (OECD, 1992), and then the following perpetual inventory method is applied to construct the ways and structures capital stock:

$$K_t = I_t + (1-\delta)K_{t-1} \quad \text{C-1}$$

where K_t is the capital stock at year t , I_t is the real value of the net investment in year t . δ is the depreciation rate. The U.S. GDP deflator and 3% annual depreciation rate are used to create the real capital stock series for each railway.

Finally, the railways' expenditures on services provided by third parties and

⁷ An OECD publication prepared by the European Conference of Ministers of Transport (ECMT).

purchased materials are used as a measure of materials input. These expenditures are first converted into U.S. currency using the Purchasing Power Parity (PPP) index for GDP (OECD, 1992) and are then deflated to constant 1985 value using the U.S. GDP deflator.

The environmental and production characteristics variables include: (1) passenger and freight traffic densities, measured by passenger kilometers (freight tonne-kilometers) per route kilometer and train-kilometers per route kilometers (Table C.2), these variables reflects in part the market demand condition of the railways; (2) average load per train, it indicates the level of vehicle utilization; (3) average length of trip and average length of haul, they indicate the type of traffic, e.g. long distance vs short distance commuter traffic; (4) percentage of passenger train kilometers in total train kilometers, this variable is considered as an indicator of the importance of passenger services in the railways' overall operation; (5) electrification rate, as measured by the percentage of electrified lines in total rail lines operated, is intended to reflect the state of technology being employed and the extent of infrastructure investment.

In addition, two policy variables are considered: SUBSIDY and AUTONOMY. SUBSIDY is measured by the ratio of subsidy to operating costs. Subsidy policy should ideally be examined according to the types of subsidies and the ways in which they are provided (e.g. loss/balancing subsidy vs. a fixed sum subsidy, unconditional subsidy vs. payment conditional on meeting a certain performance standards, etc.), which are likely to have substantial impacts on a firm's efficiency. However, due to limited information, this study considers only the aggregate subsidy. The AUTONOMY variable is an index of regulatory and institutional environments, based on 1988-1989 figures, which was

constructed by Perelman and Gathon (1990) using the information collected through a survey of railways' management. Its values range between 40 and 100. The more autonomous management was, the higher the value of the AUTONOMY index. The degree of managerial autonomy is affected by a large number of factors including ownership form and managerial mandate. It is therefore very difficult to quantify managerial autonomy consistently across railways even with the best of efforts. Another problem with this variable is that it was based on only one year's worth of observations. In using this variable, it is assumed that the institutional environment of the railways had varied only minimally over the sample period, an assumption which is not realistic for most railways. Moreover, there is no index available for JNR, it is given the lowest autonomy rating in light of JNR was under strict governmental control for the sample period. Although it is realized that the AUTONOMY variable is not ideally defined, it is used here as it was the only information of its kind which had been collected systematically.

Appendix D

Airline Sample Data

D.1 The Sample Airlines and Data Sources

The airline data base contains 36 airlines observed over the period 1980-1992. The airlines are selected mainly due to data availability, however, they must be international air carriers, and have a significant involvement in scheduled passenger services. That is, carriers that are mainly engaged in cargo or charter services are excluded. The data are mainly collected from **Digest of Statistics** published by the International Civil Aviation Organization (ICAO), in particular, the annual series on **Traffic, Fleet-Personnel and Financial Statistics**. A primary advantage of using the ICAO statistics lies in the fact that, for the most part, they are compiled in a consistent manner across airlines. Additional data are obtained from airlines' annual reports, AVMARK and directly from airlines. The names of the airlines, their countries of origin, and the years for which the data are collected are listed in Table D.1.

The sample includes airlines from Europe, North America and Asia-Pacific. Some of the airlines are 100 percent stated owned, some are private companies, while others have mixed ownership. For example, Air France, Air India, Pakistan International Airlines are 100 percent government owned, while the US carriers are all private companies. There are large variations in size, traffic mix, and other operating characteristics among the sample airlines (Table D.2). The size of the airlines as measured by the length of the route network range from 92 thousand kilometers for Finnair to over 940 thousand kilometers for Lufthansa

Table D.1

List of Airlines & Years Used

Qantas (Australia)	1980-1992
AUA (Austrian Airlines)	1980-1992
SABENA (Belgium)	1980-1992
Air Canada (Canada)	1980-1992
CP (Canadian Airlines International)	1980-1992
Finnair (Finland)	1980-1992
Air France (France)	1980-1991
UTA (France)	1980-1991
Lufthansa (Germany)	1980-1992
Air India (India)	1980-1991
Alitalia (Italy)	1980-1992
JAL (Japan Airlines)	1980-1992
MAS (Malaysia Airlines)	1980-1991
Mexicana (Mexico)	1980-1992
KLM (Netherlands)	1980-1992
PIA (Pakistan International Airlines)	1980-1992
PAL (Philippines Airlines)	1980-1992
Tap Air (Tap Air Portugal)	1980-1992
KAL (Korean Air)	1980-1992
SAS (Scandinavian Airlines Systems)	1980-1992
SAUDIA (Saudi Arabian Airline)	1980-1991
SIA (Singapore Airlines)	1980-1992
Iberia (Spain)	1980-1992
Swiss Air (Switzerland)	1980-1992
British Airways (U.K.)	1980-1992
Cathay Pacific (Hong Kong)	1988-1992
American Airlines (U.S.)	1980-1992
US Air (U.S.)	1980-1992
American West (U.S.)	1985-1992
Continental (U.S.)	1980-1992
Delta (U.S.)	1980-1992
Eastern (U.S.)	1980-1990
Northwest (U.S.)	1980-1992
Pan American (U.S.)	1980-1990
TWA (Trans-World Airlines, U.S.)	1980-1992
United (U.S.)	1980-1992

Table D.2
Airline Characteristics¹

	Route (000'km)	Stage Length (km)	Load (%)	%PASS ² (%)	%INT'L ³ (%)
Qantas	452	4368	65	67	99.95
AUA	95	1051	50	78	100
Sabena	190	988	62	58	100
Air Canada	329	1485	54	73	60
Canadian	268	1514	54	77	68
Finnair	92	1072	57	85	93
Air France	858	1585	61	49	89
UTA	224	3841	68	48	100
Lufthansa	942	1063	65	48	95
Air India	292	2712	58	69	96
Alitalia	486	1173	66	63	94
JAL	492	2476	64	60	85
MAS	-	672	64	67	87
Mexicana	-	1131	48	96	61
KLM	492	1919	70	53	100
PIA	263	978	57	68	80
PAL	117	1045	63	73	88
Tap Air	136	1504	59	80	87
KAL	-	1292	75	41	94
Saudia	310	1103	49	72	75
SAS	212	786	59	77	85
SIA	719	4039	67	58	100
Iberia	326	1191	55	78	74
Swiss Air	309	1243	61	55	99
British Airways	581	1564	66	70	98
Cathay	243	2901	67	58	100
American	798	1493	54	86	30
US Air	-	832	49	93	9
Continental	484	1370	51	86	30
Delta	849	1184	52	87	27
Northwest	735	1397	55	74	49
TWA	304	1338	57	85	35
United	778	1551	57	84	39
American W	140	1015	53	92	2
Eastern	-	1009	51	92	9
Pan Am	289	1735	63	78	77

1. The most recent year's (in the sample) data; 2. Percentage of passenger RTK including non-scheduled service.
3. Percentage of international traffic in terms of RTK (including non-scheduled services)

of Germany⁸. Aside from the US carriers, most of the airlines in the sample provide mainly international services, some do not provide domestic services at all. Qantas, Singapore Airlines, Cathay Pacific, and JAL serve mostly inter-continental traffic as indicated by the long stage length, while other airlines have a large proportion of their business in intra-continental traffic. KLM achieved the highest average (weight) load factor at 70 % , on the other hand, the average loads for Mexicana, Saudia and US Air are below 50%. Air France and Lufthansa have a rather significant proportion of their business in cargo services, while the U.S. carriers, Finnair, and Mexicana provide primarily passenger services.

The size of the airlines in terms of annual total operating revenue ranges from 780 million US dollars for Austrian Airlines to over 13,000 million US dollars for American Airlines (Table D.3). There are also large variations in terms of revenue shares. Passenger revenue accounts for over 90 percent of operating revenues for Austrian Airlines, US Air, and Delta, while accounts for less than 60 percent for Sabena and Korean Air. Air France, Korean Air, and Singapore Airlines' cargo services contribute about 18 percent of their operating revenues. The revenue shares of incidental services range from 0.3 percent for British Airways to 41 percent for UTA and 30 percent for Sabena.

Cost recovery conditions vary greatly among airlines (Table D.3). In 1992, about one third of the airlines were able to recover their operating expense. On the other hand, none of the North American carriers had a revenue cost ratio over 1.00. Iberia incurred the lowest revenue cost ratio of 79 percent compared to 119 percent achieved by Sabena,

⁸ Airlines' route kilometers are obtained from **World Air Transport Statistics** published by International Air Transport Association (IATA).

Table D.3
Airline Financial Performance¹

	Revenue (Mill. US\$)	%Pax Rev ² (%)	%Fre Rev ² (%)	%Inc Rev ² (%)	Rev/ Exp ³
Qantas	2614	77	10	6	1.08
AUA	787	90	6	3	0.94
Sabena	1424	58	11	30	1.19
Air Canada	2226	83	9	3	0.92
Canadian	2011	84	8	3	0.89
Finnair	982	70	5	9	1.19
Air France ⁴	5284	77	18	3	0.98
UTA ⁴	1399	42	16	41	0.99
Lufthansa	9433	67	14	17	0.99
Air India	799	76	14	7	1.09
Alitalia	4105	72	8	19	1.02
JAL	7858	78	14	5	0.95
MAS	1390	76	13	7	0.98
Mexicana	1044	86	6	6	0.87
KLM	3402	67	16	16	0.84
PIA	824	85	11	3	1.08
PAL	1204	84	11	3	1.08
Tap Air	1123	74	7	18	0.83
KAL	2966	58	23	15	0.99
Saudia ⁴	2093	68	7	13	0.98
SAS	3756	74	5	20	1.05
SIA	3153	78	18	3	1.12
Iberia	3298	86	7	6	0.79
Swiss Air	3392	66	10	23	0.97
British Airways	8444	89	7	0.3	1.07
Cathay	2902	77	16	5	1.19
American	13581	88	3	8	0.99
US Air	6236	93	1	4	0.94
Continental	5210	89	3	5	0.96
Delta	11639	92	4	2	0.93
Northwest	7964	87	7	3	0.96
TWA	3570	83	3	12	0.91
United	12725	88	5	5	0.96
American W	1303	92	2	3	0.95
Eastern ⁵	2182	89	1	7	0.80
Pan Am ⁵	3931	86	5	6	0.89

1. all entries are 1992 data except for noted otherwise; 2. revenue shares for passenger, freight and incidental;
 5.3. ratio of operating revenue over operating expenses; 4. 1991 data 1990 data

Finnair, and Cathay Pacific.

D.2 Description of Variables

There are two sets of variables collected for the airlines: (1) inputs and outputs; and (2) the variables describing the operating environments and production characteristics of the airlines.

The outputs are distinguished by the following five categories: scheduled passenger service, scheduled freight service, mail service, non-scheduled (charter) services, and incidental services. Incidental services include, among other things, equipment leasing, maintenance provided to other carriers' equipment, and catering. They have been ignored in most of previous studies⁹, even though they account for upto 30 to 40 percent of revenues for some airlines such as UTA and Sabena (see Table D.3). The first four outputs are measured by revenue tonne-kilometers (RTK), while incidental output is measured by a quantity index which is constructed by deflating the incidental revenue using the Purchasing Power Parity index for consumption obtained from Summers and Heston (1991) and U.S. consumer price index.

Five categories of inputs are used: labour, fuel, material, flight equipment, and ground property and equipment (GPE). Labour input is measured by total number of employees, since labour compensation data required for constructing multilateral index is not available for some airlines in the sample. Fuel input is measured by number of gallons (US

⁹ Good, Nadiri, Röller, and Sickles (1993) includes incidental service output.

gallons) of fuel consumed. ICAO reports fuel expense data, but does not report fuel price or quantity. The data on fuel quantity for some airlines are provided by the airlines directly, those of other airlines are estimated on the basis of fuel expense data and estimated fuel prices.

Flight equipment is a multilateral index of 14 types of aircraft constructed using the following translog multilateral index procedure proposed by Caves, Christensen, and Diewert (1982):

$$LnQPL_{jh} = \frac{1}{2} \sum_k (W_k^j + \bar{W}_k) (LnQ_k^j - Ln\bar{Q}_k) \quad D-1$$

where Q_k is number of type k aircraft, W_k^j is the weight for aircraft of type k for j -the observation, the bar indicates the arithmetic mean over all the observations, and QPL_{jh} is the multilateral index in comparison with a hypothetical representative observation (h) with aircraft number vector $Ln\bar{Q}_k$, and weights \bar{W}_k . Aircrafts' leasing values are used as the weights. The data on leasing prices for the period of 1986-1992 is provided by AVMARK, the leasing prices for earlier years are obtained from Professor Tretheway for some of the aircraft types and rest are estimated based on similar aircrafts.

GPE input is estimated following the Christensen and Jorgenson (1969) procedure. The method assumes that the flow of capital services is proportional to the stock. The net GPE investment series for the airlines are collected from ICAO account data. The 1980's GPE stock is considered as the initial GPE stock. Both the initial GPE stock and net investment series are delated by the PPP investment price index (Summers and Heston, 1991)

in order to convert them into real quantity series comparable across countries and over time. The following perpetual inventory method is used to create the real GPE capital stock series for each firm:

$$K_t = I_t + (1-\delta)K_{t-1} \quad (D-2)$$

where K_t is the GPE capital stock at year t , I_t is the real value of the net investment in year t , and δ is the depreciation which is assumed to be 7%.

The last category of input is materials. The materials input is an catch-all-expense category, it includes all other inputs or material costs which are not included in the other input categories. Material cost is defined as the part of total operating cost which is not attributable to labour, fuel or capital, and is computed as follows: materials cost = carrier's operating cost (reported in ICAO pub.) - labour cost - fuel cost - reported rental and depreciation of flight equipment - depreciation of GPE. The materials input is a quantity index constructed by deflating the materials cost with PPP Consumption Price Index (Summers and Heston, 1991).

The second sets of variables describe the operating environment and production characteristics. They include that: (1) average stage length which is the average distance between takeoffs and landings, and serves as an indicator how wide an airline's network is; (2) average load factor which indicates the level of aircraft utilization; (3) revenue shares of freight, mail, non-scheduled services, and incidental services which are to reflect service mix of a particular airline; (4) government ownership dummy variables, one for majority government ownership where government ownership is over 50 percent, one for minority

government ownership where government ownership is between 20 and 50 percent, and default is private ownership; (5) geographic dummy variables, distinguishing the airlines according to their home countries geographic locations, and their importance in that region, in particular, seven groups are classified: Europe, Europe Major, Asia, Asia Major, Canada, other, and the default is U.S. carriers; (6) technical changes over time: yearly dummy variables are used to reflect any potential technical changes occurring over time.