MEASURING PERFORMANCE OF UNITED STATES COMMERCIAL AND DOMESTIC BANKS AND ITS IMPACT ON 2007-2009 FINANCIAL CRISIS

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Kekoura Sakouvogui

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Statistics

January 2019

Fargo, North Dakota

NORTH DAKOTA STATE UNIVERSITY

Graduate School

Title

MEASURING PERFORMANCE OF UNITED STATES COMMERCIAL AND DOMESTIC BANKS AND ITS IMPACT ON 2007-2009 FINANCIAL CRISIS

By

Kekoura Sakouvogui

The supervisory committee certifies that this dissertation complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Rhonda Magel
Chair

Dr. Saleem Shaik
Co-Chair

Sir. Curt Doetkott

Dr. Megan Orr

Dr. William Nganje

Approved:

February 8 2019
Date

Dr. Rhonda Magel
Department Chair

ABSTRACT

In the analysis of efficiency measures, the statistical Stochastic Frontier Analysis (SFA) and linear programming Data Envelopment Analysis (DEA) estimators have been widely applied. This dissertation is centered around two main goals. First, this dissertation addresses respectively the individual limitations of SFA and DEA models in chapters 2 and 3 using Monte Carlo (MC) simulations. Motivated by the lack of justification for the choice of inefficiency distributions in MC simulations, chapter 2 develops the statistical parameters, i.e., mean and standard deviation of the inefficiency distributions - truncated normal, half normal, and exponential. MC simulations results show that within the conventional and proposed approaches, misspecification of the inefficiency distribution matters. More precisely, within the proposed approach, the misspecified truncated normal SFA model provides the smallest mean absolute deviation and mean square error when the inefficiency distribution is a half normal. Chapter 3 examines several misspecifications of the DEA efficiency measures while accounting for the stochastic inefficiency distributions of truncated normal, half normal, and exponential derived in chapter 2. MC simulations were conducted to examine the performance of the DEA model under two different data generating processes - logarithm and level, and across five different scenarios - inefficiency distributions, sample sizes, production functions, input distributions, and curse of dimensionality. The results caution DEA practitioners concerning the accuracy of their estimates and the implications within proposed and conventional approaches of the inefficiency distributions. Second, this dissertation presents in chapter 4 an empirical assessment of the liquidity and solvency financial factors on the cost efficiency measures of U.S banks while accounting for regulatory, macroeconomic, and bank internal factors. The results suggest that the liquidity and solvency financial factors negatively impacted the cost efficiency measures of U.S banks from 2005 to 2017. Moreover, during the financial crisis, U.S banks were inefficient in comparison to the tranquil period, and the solvency financial factor insignificantly impacted the cost efficiency measures. In addition, U.S banks' liquidity financial factor negatively collapsed due to contagion during the financial crisis.

ACKNOWLEDGEMENTS

First and foremost, special praises and thanks to Jesus Christ, the Almighty, for His showers of blessings, protection, and guidance throughout this fabulous journey. Without Him, I would not be here Today. I know very well that the future would be bright with Him on my side.

I would like to express my special gratitude and appreciation to my mentor, Dr. Saleem Shaik for his continuous support and immense knowledge. His guidance has helped me in all the time of my research and writing of this dissertation. Additionally, I would like to thank my advisors, Dr. Rhonda Magel and Dr. Saleem Shaik and the rest of the committees: Sir. Curt Doetkott, Dr. Megan Orr, and Dr. William Nganje for their time and supports. Their guidance given to me are the major driving forces in the success of my dissertation. Hence, this dissertation could not have been completed without their invaluable insights. Moreover, I am extremely grateful to Sir. Curt Doetkott for his comments, suggestions, insight, and willingness to always help.

I would like to specially thank my father, Diarra Sakouvogui; my mother, Sogoni Koivogui; my siblings, Kaissa and Victor; and my girlfriend, Genevieve Guilavogui for their supports during this journey. Their unconditional supports and encouragement are making a huge difference all these years. To that, I say "Sincerement Merci Beaucoup du fond du Coeur."

I would also like to thank my instructors and classmates here at North Dakota State University for providing me with excellent advices and supports, and special thanks to my good friends Deborah Diabo, Feifei Huang, and Boubacar Diallo.

DEDICATION

This dissertation is dedicated to:

1. My Savior Jesus Christ;

2. My parents, Diarra Sakouvogui and Sogni Koivogui;

3. My lovely siblings, Kaissa Sakouvogui and Victor Yoko Sakouvogui;

4. My fabulous girlfriend, Genevieve Guilavogui;

5. My good friends, Deborah Diabo, Feifei Huang, and Boubacar Diallo;

6. My uncle Antoine Sakouvogui and his family in Guinea.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

1. INTRODUCTION

1.1. Overview of the 2007-2009 financial crisis

While it has frequently been determined that there is no universally agreed upon definition of a financial crisis, in general, a common view over the recent studies is that "Disruptions in financial markets rise to the level of a crisis when the flow of credit to households and businesses is constrained and the real economy of goods and services is adversely affected (Jickling, 2008)." This was evident during the 2007-2009 financial crisis, which has been characterized as the greatest surprise to the finance world (Jickling, 2009). During the crisis, many countries reported negative impact including the following: increase in unemployment, poverty, and hunger; deceleration of growth and economic contraction; growing budget deficits; falling tax revenues and reduction of fiscal space; contraction of world trade; declining remittances to developing countries; and reduced access to credit and to public confidence in financial institutions (United Nations Department of Economics). These factors impacted the overall bank efficiency measures by the increased number of defaults of mortgages, banks failures, and new policies.

1.2. Impact of the 2007-2009 financial crisis in the United States

In the United States (U.S), the overall impact of the financial crisis was reflected in the real gross domestic products (GDP). The Federal Reserve Bank of Dallas estimated that 14.7 million jobs were lost over the course of the crisis with unemployment peaking at 10 percent in October 2009 and between $6 trillion and $14 trillion loss of national output. Moreover, the U.S. Bureau of Economic Analysis concluded that in comparison to the recent crises, the percentage fall (magnitude) in the real GDP during the financial crisis was twice that of the fall of the 1981 financial crisis. As the effects of the financial crisis spread throughout the U.S. economy specifically in the banking sector, smaller institutions, including about 7,000 community banks (with under $1 billion in assets measured at the individual bank level rather than the bank holding company level) were affected (Hays et al., 2011) during the latter part of 2008. Overall profitability in the banking sector plunged from near record highs in 2006 to an industry loss of $32.1 billion in the fourth quarter of 2008, a -0.94 % quarterly return on average assets (FDIC Quarterly Banking Report, Fourth Quarter 2008; Hays et al., 2011).

In response to the financial crisis, different policies exemplified by: Support to Goldman Sachs and Morgan Stanley, Asset Backed Loan Facility, and Commercial Paper Funding Facility were implemented to provide capital support to significant financial institutions (Barth et al., 2009). The impact of the financial crisis and the effect of policies precisely on the U.S banking sector is illustrated in Figure 1. The financial crisis response in Figure 1 helped to restart the markets by providing financing for auto, credit card, mortgage, and business loans (U.S. Department of Treasury). From Figure 1, at the beginning of the financial crisis, banks had a net percentage easing lending standards of -80%.

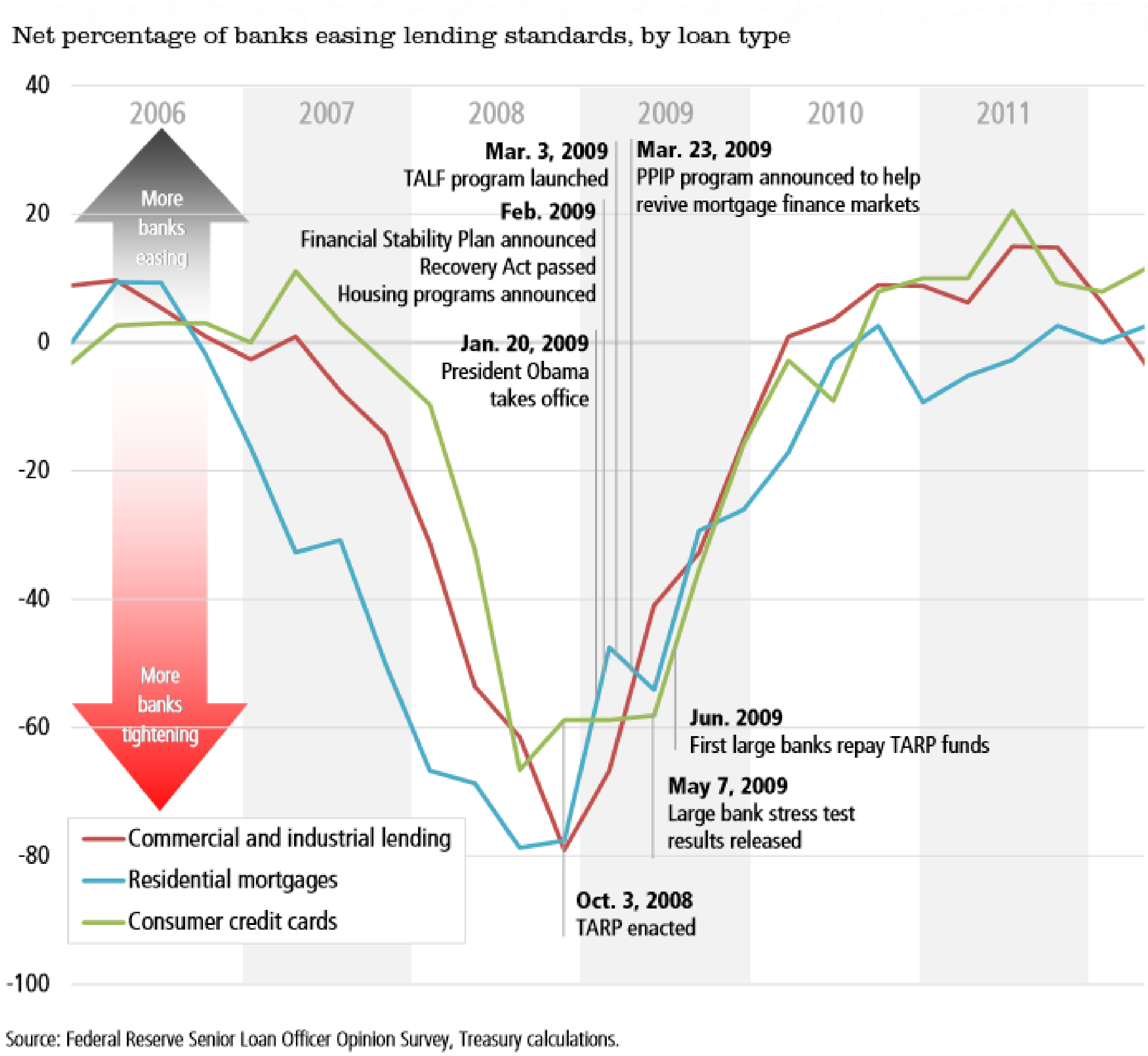


Figure 1.1. Impact of 2007-2009 financial crisis on the banking sector

Furthermore, the most important revision of the financial crisis was the implementation of the Dodd-Frank Act. However, Hank Paulson, Former Treasury Secretary stated that, " To fully prevent the financial crisis of 2007-2009, the Dodd-Frank Act would have needed to have been in place not just before September 2008, but years earlier " (Russo and Katzel, 2011). However, despite these efforts, Barth et al., (2009) concluded that neither the Federal Reserve, the U.S Department of Treasury, nor the Congress proposed an appropriate consistent regulatory structure to prevent a similar financial crisis in the future. Additionally, Mcaleer et al., (2013) and Gandrud and O'Keeffe (2016) concluded that federal policies alone are not sufficient to prevent a financial crisis. Therefore, due to stiff competition characterizing the banking sector and the failure of real appropriate policies to prevent a future financial crisis, benchmarking analysis has become a popular tool to evaluate the performance of financial banks. Hence, we are evaluating and developing efficiency statistical tools for the U.S commercial and domestic banks.

Traditional performance of benchmark is measured by income statements and balance sheets through specific indicators exemplified by the univariate assumption of financial ratios and loan processing, which are appropriate when banks utilize a single input to generate a single output. Alternative methods of efficiency measures assess the performance of banks with multiple inputs/outputs using production, cost, and profit functions based on the production, intermediation, and profitability approaches. The two widely used techniques to evaluate banks efficiency measures, statistical Stochastic Frontier Analysis (SFA) and linear programming Data Envelopment Analysis (DEA) have gained popularity based on their respective methodological advances (Aly et al., 1990; Neffet al., 1994; Berger and Humphrey, 1997; Chen, 2002; Wu and Zhou, 2011; Fen and Zhang, 2012; Lensink and Meesters, 2014; and Moradi-Motlagh and Babacan, 2015). The necessity of a continuous evaluation of banks efficiency measures was shown during the financial crisis. Under such conditions, this dissertation first addresses respectively the individual limitations of SFA and DEA models in chapters 2 and 3 using Monte Carlo (MC) simulations.

A major criticism related to SFA of Aigner et al., (1977) and Meeusen and van den Broeck (1977) is the lack of justifications for the choice of inefficiency distributions. In empirical applications, each inefficiency distribution results in different absolute technical efficiency estimates. Motivated by this drawback, chapter 2 develops a robust model specification based on statistical assumptions and validate with MC simulations in three stages. First, we develop the statistical parameters under which the inefficiency distributions yield comparable sample mean and sample standard deviation. Second, we provide a systematic comparison within the proposed and conventional approaches of the inefficiency distributions. The aim is to compare the performance of truncated normal, half normal, and exponential inefficiency distributions using alternative specifications of functional forms and input distributions within the proposed and conventional approaches. Finally, we compare the sensitivity of the estimation toward misspecification. The application of the proposed approach based on the statistical parameters of the inefficiency distributions is essential and critical since it deals with the dilemma of deciding the best SFA models among the truncated normal, half normal, and exponential inefficiency distributions.

With the limitations of SFA models, Charnes et al., (1978) reformed the piecewise linear convex approach of Farrell (1957) into a mathematical linear programming DEA method which requires no a priori information on the production shape and the distribution of input and output variables. However, motivated by the drawback of the random noise distribution and to mimic the SFA framework in chapter 2, chapter 3 examines several misspecifications of Data Envelopment Analysis (DEA) efficiency measures using MC simulations. MC simulations were conducted to examine the performance of DEA model under two different data generating processes, stochastic and deterministic, and across five different scenarios, inefficiency distributions (traditional and proposed approaches), sample sizes, production functions, input distributions, and curse of dimensionality.[1]

Second, this dissertation in chapter 4 evaluates the impact of liquidity and solvency financial factors while accounting for regulatory, macroeconimic, and bank internal factors on the cost efficiency measures using a two-step approach of SFA and DEA estimators. Hence, chapter 4 contributes to the literature of financial crisis in two ways: (1) by helping policymakers and bank regulators initiate policy measures designed to ensure efficient bank supervision and responses to regulatory changes and (2) by examining the financial factors that may have resulted from the financial crisis.

1.3. Objectives

The present dissertation contributes to the existing literature of the U.S banking efficiency measures in three ways:

1. The first major contribution deals with the statistical SFA model. The objectives include:

[1] Curse of dimensionality results by increasing the number of input variables than necessary.

(a) Developing the statistical conditions under which the inefficiency distributions yield comparable sample mean and sample standard deviation.

(b) Systematic comparison of the inefficiency distributions.

(c) Sensitivity analysis of the estimation within the proposed and conventional apporaches.

(d) Misspecification analysis of the inefficiency distributions.

2. The second major contribution deals with the linear programming DEA model. The objectives include:

    (a) Two schools of thought for the MC simulations.

        i. Logarihm.

        ii. Level.

    (b) Systematic comparison of the proposed and conventional approaches.

    (c) Misspecification of the proposed and conventional approaches:

        i. Production functions.

        ii. Input distributions.

        iii. Curse of dimensionality.

3. The final contribution evaluates and assesses the importance of liquidity and solvency financial factors on the cost efficiency measures. The emphasis will be based on regulatory policies, macroeconomic, and bank internal factors. Figure 1.2 presents the general outline of the dissertation.

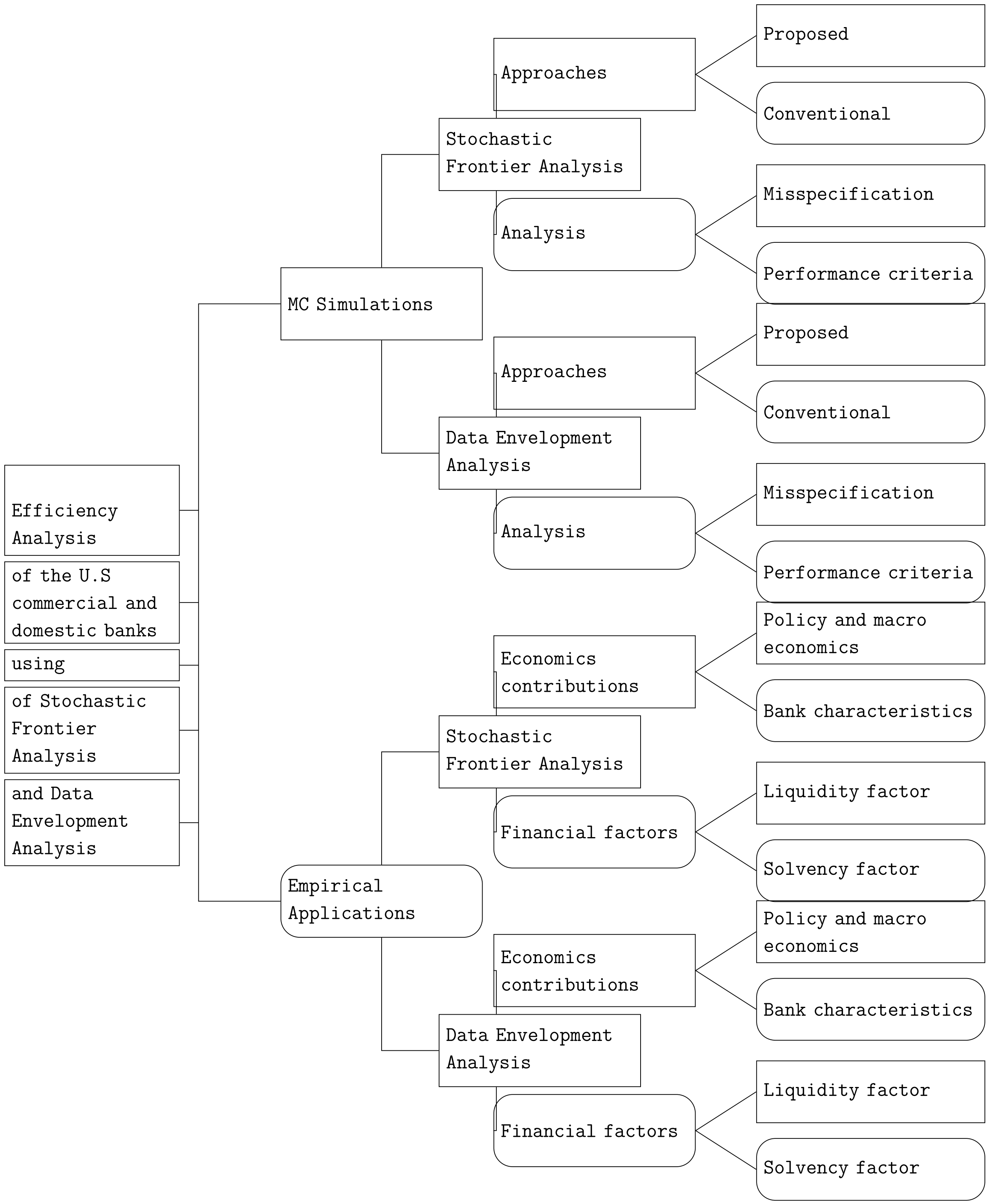


Figure 1.2. Outline of the dissertation

2. SENSITIVITY ANALYSIS OF STOCHASTIC FRONTIER ANALYSIS INEFFICIENCY DISTRIBUTIONS

2.1. Abstract

The efficiency measures estimated by the Stochastic Frontier Analysis (SFA) models are dependent on the distributional assumptions of one-sided error term or inefficiency. Monte Carlo (MC) simulations are conducted to evaluate the statistical properties and robustness of the truncated normal, half normal, and exponential distributions of SFA models on efficiency measures. The use of MC simulations requires the derivation of the population mean and standard deviation from the underlying statistical distributional assumptions. Given the intent of earlier researchers to evaluate a single inefficiency distribution, attention has not been paid to the derivation of population mean and standard deviation for comparative analysis of SFA models. Therefore, this chapter develops the correct statistical parameters of the inefficiency distributions with the objective of having an "apples-to-apples" comparison. MC simulations results show that within the proposed and conventional approaches, misspecification of the inefficiency distribution matters. Moreover, within the proposed approach, the misspecified truncated normal SFA model provides the smallest mean absolute deviation and mean square error when the true data generating process is a half normal.

2.2. Introduction

Conventional Stochastic Frontier Analysis (SFA) model of Aigner et al., (1977), Battese and Corra (1977), and Meeusen and van den Broeck (1977),[1] used to estimate the relative technical efficiency (TE) measures of the decision-making units (DMUs), exemplified by banks (financial and non-financial banks), assumes a stochastic relationship between the input and output variables using parametric distributions. Since its introduction, SFA has appeared to be a promising field of study for researchers in different context. SFA allows deviations from the optimal to be due to the random error, v_i, following a normal distribution and a one-sided inefficiency distribution, u_i, assumed to be a half normal (Aigner et al., 1977 and Meeusen and Van den Broeck, 1977),

[1]During the last 35 years, the efficiency and productivity analysis of the banking industry have been extensively investigated.

an exponential (Aigner et al., 1977), a truncated normal (Stevenson, 1980), or a gamma (Green, 1990). Recently, Stone (2002) and Tsionas (2017) pointed out issues in connection to the efficiency estimation including the statistical distributions of the random error and the one-sided inefficiency distributions.

Contrary to the work of Horrace and Parmeter (2015), the normality assumption of the random error distribution has been universally accepted in both applied and theoretical works (Christopher and Kumbhakar, 2014). However, little is known about how to choose among competing inefficiency distributions, or their implications on the estimation results. In empirical applications, Baccouche and Kouki (2003) concluded that the use of different inefficiency distributions will result in different absolute TE estimates and ranking of DMUs. In addition, Bravo-Ureta and Pinheiro (1993), Coelli (1995), and Simar and Wilson (2009) concluded that most SFA model's applications use cross-sectional data with a one-sided inefficiency distribution following a half normal except for panel SFA models (Shaik, 2015) and Ondrich and Ruggiero (2001). While we agree with the recent literature that the half normal inefficiency distribution plays a pivotal role in the choice of SFA model, it is not in any case a substitute for the exponential and truncated normal inefficiency distributions. In this chapter, our interest typically lies with the underlying choice of the inefficiency distributions in a cross-sectional framework.

The determination of the inefficiency distributions should be dictated by the data and yet SFA has no a priori information for its selection (Coelli, 1995) and this important issue still has not been fully addressed (Stone, 2002 and Tsionas, 2017). Since SFA models yield different efficiency estimates based on an inefficiency distribution, researchers and policy makers always face the problem of determining the true efficiency estimates (Andor and Hesse, 2011). The distributional assumption of the inefficiency problems seems to be resolved only in the context of misspecification during Monte Carlo (MC) simulations. Ruggiero (1999), Ondrich and Ruggiero (2001), Jensen (2005), Liu et al., (2008), and Andor and Hesse (2011 and 2013) have conducted extensive MC simulation studies to examine the impact of misspecifying an inefficiency distribution. Others have used two stages semiparametric SFA models (Badunenko et al., 2012 and Tsionas, 2017). The use of misspecification partly solves the issue of inefficiency distribution and its impact on TE measures and rankings although not in a comprehensive manner.

The current MC simulations, based on pre-defining an inefficiency distribution usually a half normal (Simar and Wilson, 2009) and examining its impact on an alternative exponential inefficiency distribution, creates even greater problems due to four reasons. First, the comparison among the basics inefficiency distributions of half normal, exponential, and truncated normal is almost non-existent. Second, the failure to develop and identify the statistical parameters across the underlying inefficiency distributions with the goal of having an “apples-to-apples" comparison in a comprehensive manner still exist. Third, the existence of the large discrepancies among the sample mean and standard deviation of the different inefficiency distributions still provides bias and inconsistent results. Fourth, the current MC simulations is not robust because for different statistical parameters of the inefficiency distributions, the researchers might get to a different set of conclusions. Therefore, it becomes nearly impossible to truly evaluate and select the most appropriate model among competing SFA models without correctly identifying the statistical parameters of the inefficiency distributions. Henceforth, researchers have been using alternative inefficiency distributions (SFA models) that best suits their issue or goal (Baccouche and Kouki, 2003; Andor and Hesse, 2011; and Huang and Lai, 2012). We refer to this as “ The conventional approach”.

Motivated by the current failure to determine the most efficient inefficiency distributions, this chapter develops a robust model specification of the statistical parameters and validate with MC simulations with three main objectives. First, this chapter identifies the conditions under which the inefficiency distributions yield comparable statistical parameters; we refer to this as “The proposed approach". This procedure has two advantages over the conventional MC simulations. The first advantage includes identification of the statistical properties of half normal, exponential, and truncated normal inefficiency distributions. The second advantage includes an apples-to-apples comparison of the inefficiency distributions. Second, we compare the performance of the inefficiency distributions using alternative specifications of functional forms and input distributions. Third, we compare the sensitivity of both the proposed and conventional approaches toward misspecification.[2] Because the true inefficiency distribution of the data is unknown in practice, misspecification in SFA models can become intractable in the conventional approach MC simulations by providing different efficiency measures based on the inefficiency distribution.

[2]Misspecification arises when the researcher incorrectly applies an SFA model to the data generated under a different SFA model – in other words an incorrect inefficiency distribution was chosen.

The remainder of this chapter is organized as follows. Following the introduction, Section 2.3 provides the theoretical and statistical framework. Under this section, we present the primal SFA model with half normal, exponential, and truncated normal inefficiency distributions. In Section 2.4, we present the standard design of MC simulations. Section 2.5 deals with the simulation results. Finally, the summary of our findings are discussed in Section 2.6.

2.3. Statistical theory of SFA models

In the typical cross-sectional setting with i DMU, the production function estimated using SFA model is expressed as:

$$y_i = f(x_i, \beta) + v_i - u_i, \tag{2.1}$$

where y_i represents an endogenous dependent variable in natural logarithm. x_i is a vector of exogenous independent inputs in natural logarithm used in the production function. β_i is a vector (column vector) of unknown parameters associated with the inputs. In this research, the random error, v_i, which cannot be controlled by the DMU, is assumed to be independent and identically distributed following a normal distribution. A negatively skewed one-sided inefficiency, u_i, associated with the technical inefficiency, which is explained by the shortfall of actual output from its maximum possible value, is represented by alternative distributions including half normal, exponential, and truncated normal. When u_i is zero, equation (2.1) is estimated using the ordinary least square method. In addition, the theory suggests that the inefficiency distributions should be left (right) skewed in production (cost) function (Greene, 1990). Our model in equation (2.1) is comprised of the following assumptions:

Assumption 1 *The set of N inputs, x_{ni} = 1.,.,.,N and i= 1.,.,.,I used to produce a single output, y_i= 1.,.,.,I are available for each of I DMU.*

Assumption 2 *We require that $0 \leq TE_i = e^{-u_i} \leq 1$.*

Assumption 3 *We assume v_i and u_i are independent of each other and of the regressor, y_i.*

From assumptions (1) and (3), Jondrow et. al. (1982) proposed a method to distinguish between the two random components of v_i and u_i by imposing distributional assumptions. The conditional distribution of $(u_i \mid \varepsilon_i)$ can be derived either using $E[u_i \mid \varepsilon_i]$ as a central measure instead of u_i or $E[e^{-u_i} \mid \varepsilon_i]$ as a central measure instead of e^{-u_i}.

Definition 1 *Aigner et al., (1977) proposed the normal and half normal SFA model.*

For the normal (N) and half normal (HN) distributions assumed by Aigner et al., (1977), the joint density function of the random error, $v \overset{iid}{\sim} N(0, \sigma_v^2)$ and inefficiency distribution, $u \overset{iid}{\sim} HN(0, \sigma_u^2)$ is expressed as:

$$f(u, v) = \frac{2}{2\pi\sigma_u\sigma_v} \exp\left\{-\frac{u^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{2.2}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating u out, the marginal density function of ε can be written as:

$$f(\varepsilon) = \int_0^\infty f(u, \varepsilon)\, du = \frac{2}{\sigma}\phi\left(\frac{\varepsilon}{\sigma}\right)\Phi\left(-\frac{\varepsilon\lambda}{\sigma}\right), \tag{2.3}$$

where ϕ is the standard normal density, Φ is the standard normal cumulative distribution function (CDF), $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, and $\lambda = \sigma_u/\sigma_v$. Using equation (2.3), the loglikelihood function of $f(\varepsilon)$ can be written as:

$$\ln L(\alpha, \beta, \sigma, \lambda) = constant - n \ln \sigma + \sum_{i=1}^{n} \ln \Phi\left(\frac{-\varepsilon_i\lambda}{\sigma}\right) - \frac{1}{2}\left(\frac{\varepsilon_i}{\sigma}\right)^2, \tag{2.4}$$

where $\varepsilon_i = y_i - f(x_i, \beta)$. By Jondrow et al., (1982), the conditional distribution of $(u|\varepsilon)$ is:

$$f(u|\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_*} e^{-\frac{(u-\mu_*)^2}{2\sigma_*^2}} \left[1 - \Phi(-\frac{\mu_*}{\sigma_*})\right]^{-1}, \tag{2.5}$$

where $\mu_* = \frac{\varepsilon\sigma_u^2}{\sigma^2}$ and $\sigma_*^2 = \frac{\sigma_\nu^2\sigma_u^2}{\sigma^2}$. Clearly, the distribution of μ_* conditional on ε correspond to the point estimator of μ_*. The estimate of the technical efficiency is:

$$TE_i = e^{-\left[\mu_{*i} + \sigma_*\left[\frac{\phi(\frac{-\mu_{*i}}{\sigma_*})}{(1-\Phi(\frac{-\mu_{*i}}{\sigma_*}))}\right]\right]}. \tag{2.6}$$

Definition 2 *Stevenson (1980) proposed the normal and truncated normal SFA model.*

For the normal (N) and truncated normal (TN) distributions assumed by Stevenson (1980), the joint density of the random error, $v \overset{iid}{\sim} N(0, \sigma_v^2)$ and inefficiency distribution, $u \overset{iid}{\sim} TN(\mu, \sigma_u^2)$

can be written as:

$$f(u,v) = \frac{1}{2\pi\sigma_u\sigma_v\Phi(\mu/\sigma_u)} \exp\left\{-\frac{(u-\mu)^2}{2\sigma_u^2} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{2.7}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating u out, the marginal density function of ε can be written as:

$$f(\varepsilon) = \int_0^\infty f(u,\varepsilon)\,du = \frac{1}{\sigma}\phi\left(\frac{\varepsilon+\mu}{\sigma}\right)\Phi\left(\frac{\mu}{\sigma\lambda} - \frac{\varepsilon\lambda}{\sigma}\right)\left\{\Phi\left(\frac{\mu}{\sigma_u}\right)\right\}^{-1}, \tag{2.8}$$

where ϕ is the standard normal density, Φ is the standard normal cumulative distribution function (CDF), $\sigma = \sqrt{\sigma_u^2 + \sigma_v^2}$, and $\lambda = \sigma_u/\sigma_v$. Using equation (2.8), the loglikelihood function of $f(\varepsilon)$ can be written as:

$$\begin{aligned} \ln L(\alpha,\beta,\sigma,\lambda) = \mathit{constant} - n\ln\sigma - n\ln\Phi\left(\tfrac{\mu}{\sigma_u}\right) + \\ \sum_{i=1}^{n} \ln\Phi\left(\tfrac{\mu}{\sigma\lambda} + \tfrac{\varepsilon_i\lambda}{\sigma}\right) - \sum_{i=1}^{n}\left(\tfrac{\varepsilon_i+\mu}{\sigma}\right)^2, \end{aligned} \tag{2.9}$$

where $\varepsilon_i = y_i - f(x_i, \beta)$. Once the maximum likelihood estimates are obtained, the conditional distribution of $(u|\varepsilon)$ is:

$$f(u|\varepsilon) = \frac{1}{\sigma_a\sqrt{2\pi}[1 - \Phi(\frac{-\mu^+}{\sigma_a})]} e^{-\frac{(u-\mu^+)^2}{2\sigma_a^2}}, \tag{2.10}$$

where $\mu^+ = \frac{-\sigma_u^2\varepsilon_i + \mu\sigma_v^2}{\sigma^2}$ and $\sigma_a^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma^2}$. The estimate of the technical efficiency is expressed as:

$$TE_i = e^{-\sigma_a\left[\frac{\mu_i^+}{\sigma_a} + \frac{\phi(\frac{\mu_i^+}{\sigma_a})}{1-\Phi(\frac{-\mu_i^+}{\sigma_a})}\right]}. \tag{2.11}$$

Definition 3 *Meeusen and van den Broeck (1977) and Aigner et al., (1977) proposed the normal and exponential SFA model.*

For the normal (N) and exponential (EXP) distributions assumed by Meeusen and van den Broeck (1977) and Aigner et al., (1977), the joint density of the random error, $v \overset{iid}{\sim} N(0, \sigma_v^2)$ and inefficiency distribution, $u \overset{iid}{\sim} EXP\left(\sigma_u\right)$ can be written as:

$$f\left(u,v\right) = \frac{1}{\sqrt{2\pi}\sigma_u\sigma_v} \exp\left\{-\frac{u}{\sigma_u} - \frac{v^2}{2\sigma_v^2}\right\}. \tag{2.12}$$

For convenient parameterization, substituting $v = \varepsilon + u$ and integrating u out, the marginal density function of ε can be written as:

$$f\left(\varepsilon\right) = \int_0^\infty f\left(u,\varepsilon\right) du = \frac{1}{\sigma_u}\Phi\left(-\frac{\varepsilon}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\exp\left(\frac{\varepsilon}{\sigma_u} + \frac{\sigma_v^2}{2\sigma_u^2}\right). \tag{2.13}$$

Using equation (2.13), the loglikelihood function of $f(\varepsilon)$ can be written as:

$$\ln\, L\left(\alpha,\beta,\sigma,\lambda\right) = constant - n\, \ln\, \sigma + n\left(\frac{\sigma_v^2}{2\sigma_u^2}\right) + \sum_{i=1}^{n}\left(\frac{\varepsilon_i}{\sigma_u}\right) + \sum_{i=1}^{n}\left\{\ln\, \Phi\left(\frac{\varepsilon_i}{\sigma_v} - \frac{\sigma_v}{\sigma_u}\right)\right\}. \tag{2.14}$$

where $\varepsilon_i = y_i - f(x_i,\beta)$. Once the maximum likelihood estimates are obtained, the conditional distribution of $(u|\varepsilon)$ is:

$$f(u|\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma_v\Phi(-\frac{\mu^+}{\sigma_v})} e^{-\frac{(u-\mu^+)^2}{2\sigma^2}}, \tag{2.15}$$

where $\mu^+ = \frac{-\sigma_u^2\varepsilon_i + \mu\sigma_v^2}{\sigma^2}$ and $\sigma_a^2 = \frac{\sigma_u^2\sigma_v^2}{\sigma^2}$. The estimate of the technical efficiency is expressed as:

$$TE_i = e^{-\left[\mu_i^+ + \sigma_v\left[\frac{\phi(-\frac{\mu_i^+}{\sigma_v})}{\Phi\frac{(\mu_i)^+}{\sigma_v}}\right]\right]}. \tag{2.16}$$

A major criticism of the current MC simulations is the presence of the large discrepancies in the sample mean and sample standard deviation of the one-sided inefficiency distributions. Most applied papers do not rigorously check differences in estimates and inference across different distributional assumptions (Christopher and Kumbhakar, 2014). A few papers that have engaged in MC simulations only checked the impact of misspecification imposed on the inefficiency distribution. Hence, this research is proposing the correct methodology assumptions to be used in MC simulations by specifically deriving the correct and true parameters of the inefficiency distributions of truncated normal, half normal, and exponential.

To concisely evaluate the performance and account for the sensitive estimations of truncated normal, half normal, and exponential inefficiency distributions, we must resolve the following problems: (i) Find the correct statistical parameters of the truncated normal and half normal inefficiency distribution; (ii) Find the correct statistical parameters of the exponential distribution; and (iii) With the goal of having apples-to-apples comparison of the inefficiency distribution, we need to find a scaling parameter of the exponential distribution.

Theorem 2.3.1 *Deriving the true and correct population mean and standard deviation of truncated normal distribution.*

A continuous random, y, is said to have a normal distribution with mean, μ, and variance, σ^2, if its probability density function (pdf) is given by:

$$p(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \tag{2.17}$$

where $-\infty < y < \infty$, $-\infty < \mu < \infty$, and $\sigma > 0$. Moreover, a continuous random, y, is said to have a truncated normal distribution with mean, μ, standard deviation, σ, lower bound, a, and upper bound, b, if its pdf is given by:

$$f(y) = \frac{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}}{\Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})}, \tag{2.18}$$

where $\Phi(.)$ is the standard normal cumulative distribution function. Following Absanullah et al., (2014), the population mean, $E(y)$, and variance, $V(y)$, of the truncated normal distribution are:

$$\begin{aligned} E(y) &= E(y|y \in [a,b]) \\ &= \mu + \sigma\left[\frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)}\right] \end{aligned} \tag{2.19}$$

$$V(y) = \sigma^2\left[1 + \frac{Z_a\phi(Z_a) - Z_b\phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)}\right] - \sigma^2\left[\frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)}\right]^2, \tag{2.20}$$

where $Z_a = \frac{a-\mu}{\sigma}$ and $Z_b = \frac{b-\mu}{\sigma}$. $\phi(.)$ is the probability density function of a standard normal distribution. $\Phi(.)$ is the standard normal cumulative distribution function normal distribution.

A special case of the truncated normal distribution is the standard normal distribution. For simplicity, we assume that $\mu = 0$ and σ^2=1. Given assumptions (1) and (2), the maximum and minimum TE measures attainable by any DMU in the production function are respectively 1 and 0. Hence, the maximum and minimum TE measures respectively correspond to the upper and lower bound of the truncated normal distribution. Conclusively, the population mean, $E(y)$ and variance, $V(y)$ of the truncated normal distribution with a lower bound 0 and an upper bound 1 are respectively:

$$
\begin{aligned}
E(y) &= 0 + 1\left[\frac{\phi(0) - \phi(1)}{\Phi(1) - \Phi(0)}\right] \\
&= \frac{0.157}{0.3413} = 0.46.
\end{aligned} \tag{2.21}
$$

$$
\begin{aligned}
V(y) &= 1^2\left[1 + \frac{0\phi(0) - 1\phi(1)}{\Phi(1) - \Phi(0)}\right] - 1^2\left[\frac{\phi(0) - \phi(1)}{\Phi(1) - \Phi(0)}\right]^2 \\
&= 0.2909 - 0.2116 = 0.0793.
\end{aligned} \tag{2.22}
$$

Theorem 2.3.2 *Deriving the correct population mean and standard deviation of the half normal and exponential distributions.*

A continuous random, y, is said to have a half normal distribution with mean, μ, and variance, σ^2, if its pdf is given by:

$$
P(y) = \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2}, \tag{2.23}
$$

where $y > 0$ and $\sigma > 0$. The population mean, $E(y)$, of the half normal inefficiency distribution is:

$$
\begin{aligned}
E(y) &= \int_0^\infty y \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy \\
&= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty (-\sigma^2) \frac{d}{dy} exp(-\frac{(y-\mu)^2}{2\sigma^2} dy \\
&= \mu + \sqrt{\frac{2}{\pi}}\sigma.
\end{aligned} \tag{2.24}
$$

The $E(y^2)$ of the half normal inefficiency distribution is:

$$E(y^2) = \int_0^\infty y^2 \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{y-\mu}{\sigma})^2} dy \tag{2.25}$$
$$= \frac{2}{\sigma\sqrt{2\pi}} \int_0^\infty (-\sigma^2) \frac{d}{dy} exp(-\frac{(y-\mu)^2}{2\sigma^2} dy$$
$$= \sigma^2 + \mu^2.$$

Then the population variance, $V(y)$, of the half normal inefficiency distribution is:

$$V(y) = E(y^2) - (E(y))^2 \tag{2.26}$$
$$= \sigma^2 \left[1 - \frac{2}{\pi}\right].$$

In order to have MC simulations representing comparable sample statistics, the population mean and standard deviation of the half normal inefficiency distribution should correspond to the population mean and standard deviation of the truncated normal inefficiency distribution.

A continuous random, y, is said to have an exponential distribution with mean, β, if its pdf is given by :

$$p(y) = \frac{1}{\beta} e^{-\frac{y}{\beta}}, y > 0. \tag{2.27}$$

The population mean, $E(y)$, of the exponential inefficiency distribution is:

$$E(y) = \int_0^\infty y \frac{1}{\beta} e^{-\frac{y}{\beta}} dy \tag{2.28}$$
$$= \beta$$

The expectation of y^2 of exponential distribution is:

$$E(y^2) = \int_0^\infty y^2 \frac{1}{\beta} . e^{-\frac{y}{\beta}} dy \tag{2.29}$$
$$= \frac{\beta^3}{\beta} \int_0^\infty z^2 exp(-z) dz$$
$$= 2\beta^2$$

Then the population variance, $V(y)$, of the exponential inefficiency distribution is:

$$V(y) = \beta^2 \tag{2.30}$$

However, a 2 parameter (shifted) exponential would be used in this research. The pdf of a two-parameter exponential distribution can be written as:

$$p(y) = \frac{1}{\beta} e^{-\frac{y-\theta}{\beta}}, y > 0 \tag{2.31}$$

For the MC simulations to represent an “apples-to-apples" comparison across truncated normal, half normal, and exponential inefficiency distributions, the population variance of the exponential distribution should correspond to the square root of the variance of the truncated normal distribution. Therefore, the population mean of the exponential distribution is 0.2816. Since the difference between the population mean of the truncated normal or half normal distribution and the population mean of the exponential distribution is 0.1784, then the exponential distribution is shifted positively by 0.1784 units. Using the population parameters, the generation of the inefficiency distributions for MC simulations was as follows:

1. Truncated normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal (0.46, 0.2186).
   (b) Step 2: If $0 < u_i < 1$ then accept u_i. Otherwise re-generate u_i.
2. Half normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal (0.46, 0.2186).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i. This process is similar to taking the absolute of u_i.
3. Exponential:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Exponential (0.2186).
   (b) Step 2: If $u_i > 0$ then accept u_i and add 0.1784. Otherwise re-generate u_i.

For the conventional approach, we follow the works of Behr and Tente (2008) for the half normal, and applied the statistical parameter of Andor and Hesse (2011) and Hafner et al., (2016) respectively to the truncated normal and exponential inefficiency distributions. The algorithms for generating the inefficiency distributions were as follows:

1. Truncated normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal(0, 0.20).
   (b) Step 2: If $0 < u_i < 1$ then accept u_i. Otherwise re-generate u_i.
2. Half normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal(0, 0.587).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i. This process is similar to taking the absolute of u_i.
3. Exponential:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Exponential (0.35).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i.

A well-known problem across the truncated normal and half normal inefficiency distributions of the proposed approach is that the derived parameters are the same (mean and standard deviation).[3] To address this issue, we perform an acceptance rejection method of MC simulations in which we only accept positive inefficiency values of the half normal distribution which follows the theory of SFA production function (Aigner et al., 1977). For the truncated normal distribution, the inefficiency values are bounded between 0 and 1. This will, however, result into different sample standard deviation across the inefficiency distributions due to the screening constraint. Table 2.1 presents the sample statistics of the realized inefficiency distributions.

As already mentioned, since the inefficiency distributions are based on different screening criteria, the relationship between the standard deviation of the inefficiency distributions, σ_u, and the random error distribution, σ_v are presented. As the random error, σ_ε, composed of σ_u and

[3]During MC simulations in exception of the exponential inefficiency distribution, the truncated normal and half normal inefficiency distributions of the proposed and conventional approaches were from a normal distribution with their respective screening constraints.

σ_v, an interesting aspect of all the models is related to the ratio of σ_u $/\sigma_v = \lambda$. Aigner et al., (1977) found that the smallest the ratio of σ_u $/\sigma_v$, the better the models become near-identifiable.[4] Therefore, from Table 2.1, for a priori comparison of the inefficiency distributions, we would expect the truncated normal inefficiency distribution to provide the smallest MAD and MSE values within each approach.

Table 2.1. Sample statistics of the inefficiency distributions

n	Parameters	Standard production function		
		Exponential	Truncated normal	Half normal
		Proposed approach		
50	$\bar{u}$	0.46	0.47	0.47
	σ_u	0.269	0.228	0.256
	σ_v	0.147	0.147	0.147
	$\frac{\sigma_u}{\sigma_v}$	1.829	1.551	1.742
		Conventional approach		
50	$\bar{u}$	0.348	0.159	0.468
	σ_u	0.330	0.117	0.344
	σ_v	0.147	0.147	0.147
	$\frac{\sigma_u}{\sigma_v}$	2.225	0.796	2.340

50: Sample size of 50 with 5000 replications. $\bar{u}$: Sample mean of the inefficiency distributions. σ_u: Sample standard deviation of the inefficiency distributions. σ_v: standard deviation of the random noise distribution and $v_i \overset{iid}{\sim}$ Normal (0, 0.15).

2.4. MC simulations of SFA models

The overall aim of the MC simulations is two-fold. First, this chapter examines the performance of SFA models within the proposed and conventional approaches. Second, this chapter evaluates the impact of misspecification of the inefficiency distributions within the proposed and conventional approaches. Accordingly, different factors affecting the performance of SFA models were considered for the data generation process (DGP) of the MC simulations.

[4]From Aigner et al (1977), the standard SFA models tend to be near-identifiable as the λ tends ∞. As the λ tends 0, there is no inefficiency in the disturbance, and the model can be efficiently estimated by ordinary least squares regression.

2.4.1. Sample size

Studies have shown that the SFA models are impacted by the sample sizes. For small sample sizes, the performance of SFA model might provide small chance of reliable results (Andor and Hesse, 2011). Following Andor and Hesse (2011), we considered a range of sample sizes of 50, 100, and 200 with 5000 replications.

2.4.2. Input distributions

The design of the MC simulations represents the case where the inputs are uncorrelated with the TE measures (Andor and Hesse, 2011, 2013). Two inputs, x_{1i} and x_{2i} were randomly and independently generated from a uniform distribution over the interval [5,15] to obtain one output, y_i.

2.4.3. Production function

The performance of SFA models can be dictated by the choice of the functional form. In the literature of SFA, Cobb Douglas production function has been widely used (Ruggiero, 1999 and Andor and Hesse, 2011 and 2013). Hence for the baseline scenario of MC simulations, we considered the Cobb-Douglas production function with the property of constant return to scale (CRS) and defined as:

$$y_i = ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i. \tag{2.32}$$

2.4.4. Noise distribution

Following Ruggiero (1999), the random noise, v_i, is drawn independently from a normal distribution with mean 0 and standard deviation 0.15.[5]

2.4.5. Misspecification of SFA models

To assess the quality and robustness of the proposed and conventional approaches, we conducted five misspecification scenarios.

1. In the first scenario, we falsely applied the truncated normal and half normal SFA models to the data generated with the exponential inefficiency distribution.

2. In the second scenario, we falsely applied the truncated normal and exponential SFA models to the data generated with the half normal inefficiency distribution.

[5] Additionally, two normal distributions were generated to account for the different levels of statistical noise, assuming values of 0.05, and 0.10. Doing so, truncated normal still performs better in terms of misspecification.

3. In the third scenario, we falsely applied the exponential and half normal SFA models to the data generated with the truncated normal inefficiency distribution.

4. In the fourth scenario, we assumed that the baseline production function was wrongly specified and considered three additional production functions: Cobb-Douglas production function with the properties of increasing return to scale (IRS) and decreasing return to scale (DRS) and the translog production function, defined as:

    (a) IRS: $y_i = \ln(2) + 0.6 \ln x_{1i} + 0.6 \ln x_{2i} - u_i + v_i$.

    (b) DRS: $y_i = \ln(2) + 0.4 \ln x_{1i} + 0.4 \ln x_{2i} - u_i + v_i$.

    (c) Translog: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} + (0.5 \ln x_{1i})^2 + (0.5 \ln x_{2i})^2 + 0.1 \ln x_{1i} \ln x_{2i} - u_i + v_i$.

    The parameters of the translog production function used for DGP are from Andor and Hesse (2013).

5. Finally, in the fifth scenario, we assumed that the DGP of the input distribution was misspecified. Accordingly, based on the financial operations of the institutions composing the United States Farm Credit System obtained from the Financial Credit Administration (FCA), two input distributions were used:

    (a) $x_1 \overset{iid}{\sim} Normal$ (8.4, 1.51) and $x_2 \overset{iid}{\sim} Normal$ (7.92, 1.18).

    (b) $x_1 \overset{iid}{\sim} Weibull$ (17.4, 20.5) and $x_2 \overset{iid}{\sim} Gamma$ (285, 0.07).

2.4.6. Performance criteria

To assess the performance of the inefficiency distributions within the proposed and conventional approaches, we calculated Mean Squared Error (MSE) and Mean Absolute Deviation (MAD) between the estimated and actual TE measures. MSE and MAD are defined respectively as:

$$MSE = \frac{1}{rK} \sum_{r=1}^{R} \sum_{k=1}^{K} (\widehat{TE_{i,r}} - TE_{i,r})^2 \tag{2.33}$$

$$MAD = \frac{1}{rK} \sum_{r=1}^{R} \sum_{k=1}^{K} |(\widehat{TE_{i,r}} - TE_{i,r})|, \tag{2.34}$$

where $TE_{i,r} = exp(-u_{i,r})$ and $\widehat{TE_{i,r}} = exp(-E(u_{i,r} \mid \varepsilon_i])_{i,r})$. r is the number of replications. $\widehat{TE_{i,r}}$ denotes the estimated TE measures from SFA model and $TE_{i,r}$ is the true (actual) TE measures for each inefficiency distribution.

In addition, since the performance measures of MAD and MSE are based on the mean value, we additionally considered an alternative method to mean based performance, Spearman rank-order correlation of TE (Rank TE). Ondrich and Ruggiero (2001) showed that the efficiency measures of Jondrow et al., (1982) used to compute TE measures of SFA models are absolute only if the researcher has a prior knowledge of the true SFA model and relative whenever such assumption cannot be made. This is typically seen in the linear programming data envelopment analysis because of its relative frontier[6]. Therefore, the Rank TE instead of the absolute TE was used as an additional performance measure (Ondrich and Ruggiero, 2001).

The absolute mean TE as a performance creteria provides inconsistent results. The mean predicted TE of Jondrow et al., (1982) is absolute because it can either increase or decrease depending on the input or output variables of the banks. For instance, if the researcher has n available banks with TE_n then for additional number of banks k with TE_{n+k}, the overall mean TE can only increase if and only if mean $TE_k >$ mean TE_n and decrease if $TE_k < TE_n$. Therefore, the increase in mean TE_k is dependent on the input and output values of the k banks. However, the ranking of the n DMUs will be consistent even with $n+k$ banks. As, a result, the Spearman rank-order correlation using the average ranking correlation between the true and estimated TE measures is defined as:

$$Rank\ TE = \frac{1}{R}\sum_{r=1}^{R} \frac{\sum_{i=1}^{n}(\hat{t}e_{i,r} - \hat{\bar{t}}e_{i,r})(te_{i,r} - \bar{t}e_{i,r})}{\sqrt{\sum_{i=1}^{n}(\hat{t}e_{i,r} - \hat{\bar{t}}e_{i,r})^2 \sum_{i=1}^{n}(te_{i,r} - \bar{t}e_{i,r})^2}} \tag{2.35}$$

where $\hat{t}e_{i,r}$ and $\hat{\bar{t}}e_{i,r}$ are respectively the rank and mean rank of n predicted TE measures. $te_{i,r}$ and $\bar{t}e_{i,r}$ are respectively the rank and mean rank of the actual TE measures. All models and performance criteria of MSE, MAD, and Rank TE were estimated as primary using SAS. Figure 2.1 displays the scheme of the MC simulations.

[6] Hence, one could assume that the technical efficiency measures of SFA could either increase or decrease for an increased in the sample size.

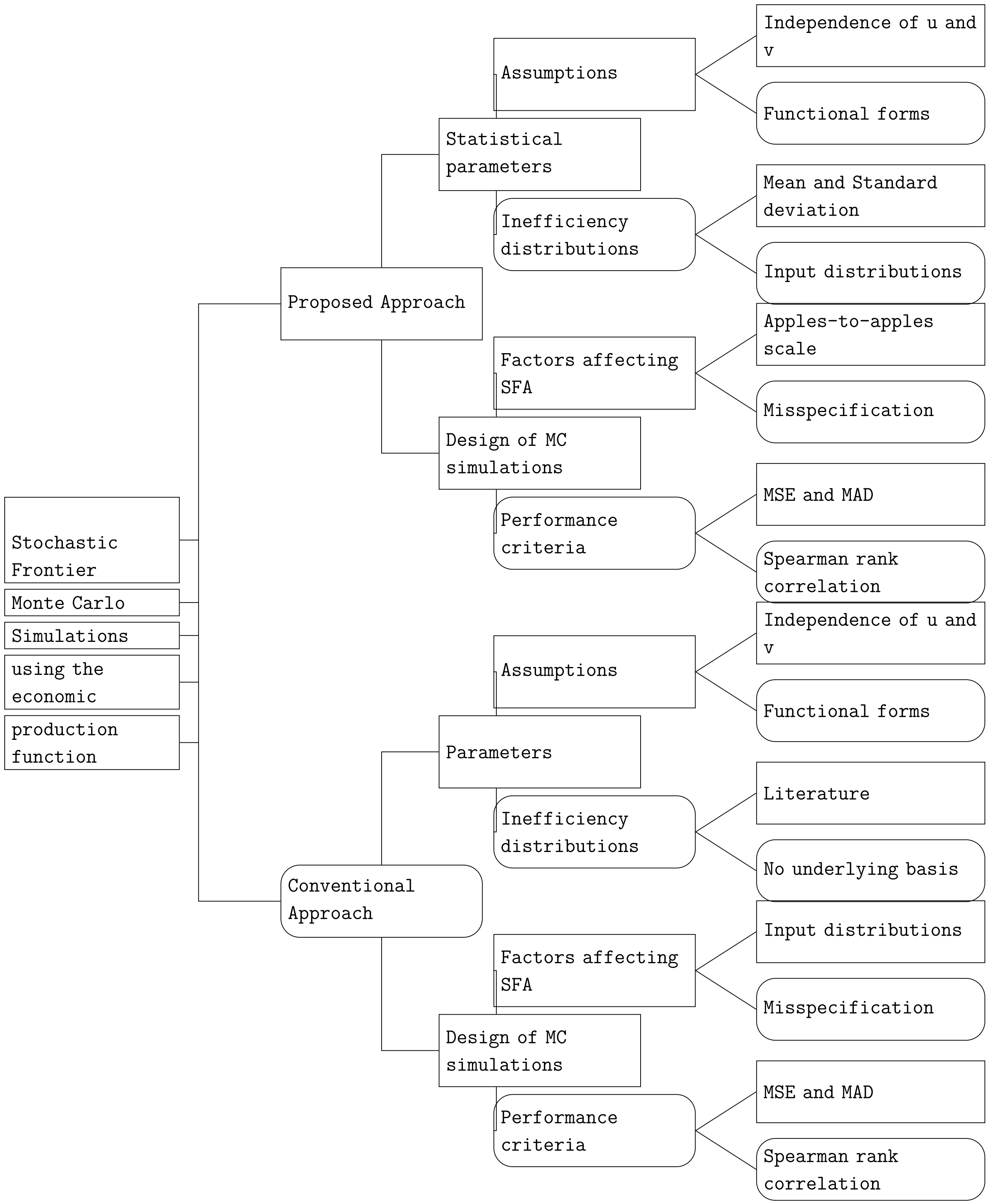


Figure 2.1. Design of the MC simulations

2.5. Results of MC simulations

In this chapter, our aim is to compare the performance of half normal, truncated normal, and exponential SFA models using the performance criteria of Rank TE, MAD, and MSE within the proposed and conventional approaches. That is because in the conventional approach, the results of the SFA models are not robust since for different statistical parameters of the inefficiency distributions, the researchers will derive to different conclusions.[7] The situations examined include: 1) The estimation of TE measures; 2) The impact of the sample size; 3) The impact of the misspecified inefficiency distributions; 4) The impact of the production functions; and finally, 5) The impact of the input distributions.

2.5.1. Impact of efficiency measures

Table 2.2 presents the performance of fitting the correct inefficiency distributions across the sample sizes of 50, 100, and 200 with 5000 replications. The results of Table 2.2 are subdivided into two blocks. The first block indicates the results of the proposed approach. The second block presents the results of the conventional approach. Within each block, the estimated Rank TE, MAD, and MSE values are presented.

Three interesting results emerge from Table 2.2. First, the Rank TE increases as the sample size increases within the proposed and conventional approaches of the inefficiency distributions. As the number of observations increase, the linear relationship between the actual and predicted TE measures become stronger with a strictly monotonic relationship. This is consistent with the studies of Ruggiero (1999), Ondrich and Ruggiero (2001), and Andor and Hesse (2011, 2013). Second, across the sample size, the magnitude of Rank TE of the truncated normal SFA model within the proposed approach is higher in comparison to the conventional approach. This suggests that within the proposed approach, the actual and predicted TE provides a better positive linear relationship. Third, within the proposed approach, the truncated normal SFA model provides smaller variation in comparison to the exponential SFA model of the conventional approach. The results within each approach entail that even though the truncated normal of the proposed and conventional approaches provide the lowest ratio of $\frac{\sigma_u}{\sigma_v}$ (Table 2.1), it does not always lead to the smallest MAD and MSE values.

[7] The proposed approach provides robust estimation of TE measures because the statistical parameters of the inefficiency distributions were set equal.

Table 2.2. Summary of technical efficiency measures

n	Parameters	Standard production function (PF)		
		Exponential	Truncated normal	Half normal
		Proposed approach		
50	Rank TE	0.621	0.305	0.558
	MAD	0.148	0.138	0.156
	MSE	0.031	0.031	0.037
100	Rank TE	0.746	0.349	0.657
	MAD	0.136	0.134	0.150
	MSE	0.025	0.029	0.035
200	Rank TE	0.770	0.413	0.697
	MAD	0.133	0.127	0.142
	MSE	0.023	0.027	0.030
		Conventional approach		
50	Rank TE	0.668	0.212	0.573
	MAD	0.100	0.143	0.129
	MSE	0.021	0.032	0.033
100	Rank TE	0.791	0.248	0.784
	MAD	0.077	0.135	0.090
	MSE	0.011	0.029	0.016
200	Rank TE	0.825	0.318	0.882
	MAD	0.070	0.132	0.086
	MSE	0.008	0.027	0.014

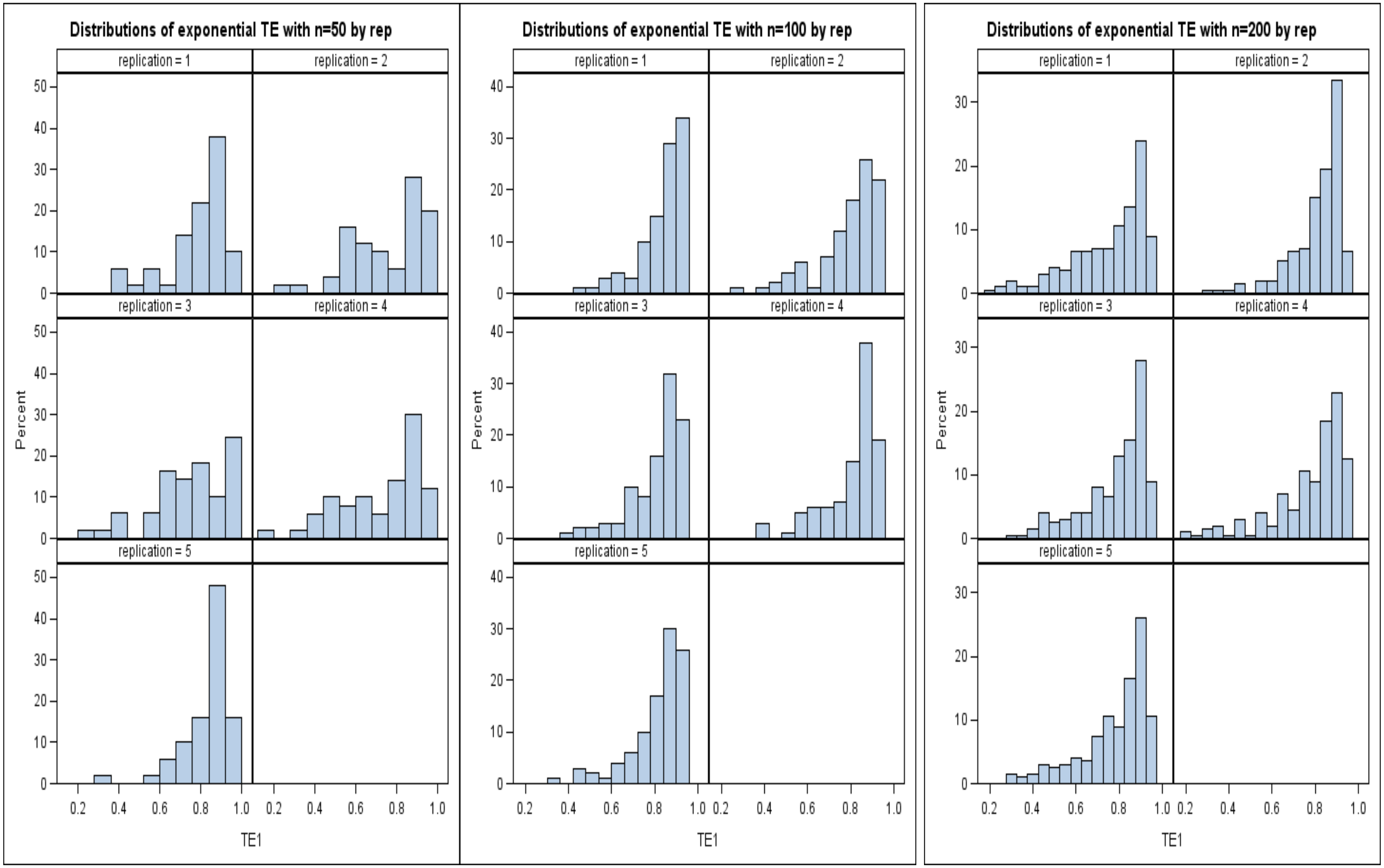
n: Sample size. Rank TE: Mean of Spearman rank correlation of TE. MSE: Mean Square Error. MAD: Mean Absolute Deviation. PF: $y_i = ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. $v_i \overset{iid}{\sim}$ Normal (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

2.5.2. Impact of sample size

The sample size is identified as one important factor influencing the performance of SFA efficiency measures (Olson et al., 1980; Ruggiero, 1999; and Andor and Hesse (2011, 2013)). To assess the impact of the sample size on the SFA models, Ruggiero (1999) and Andor and Hesse (2011) have combined the sample size and the number of replication. In contrast, the effect of the sample size in this chapter is studied by examining the impact of the replication variability on TE measures. There are two reasons for doing this. First, as reported by Andor and Hesse (2011, 2013), the effect of the sample size is minimal once the sample size is greater than 50. Second, from our point of view, we are more interested in the replication to replication variability rather than combining the number of replication and the sample size.

From Table 2.2, regarding the comparison within the proposed and conventional approaches, the sample size has an influence on the performance of the Rank TE of SFA models. Furthermore, using the performance criteria of MAD and MSE, the truncated normal SFA model of the proposed approach yields a better performance in comparison to the exponential SFA model of the conventional approach. Furthermore, since the exponential SFA model of the proposed approach provides a higher absolute change in terms of Rank TE, Figure 2.2 displays the distributions of its estimated TE measures for the first five replications.

From Figure 2.2, the distribution of TE measures for the exponential SFA model is skewed to the left. The results suggest that not only the sample size affects the performance creteria of SFA model (Table 2.2) but also the number of replications (Figure 2.2). That is for different sample sizes, Figure 2.2 presents a changed in scale. Our results coincide with the works of Banker et al., (1993) and Andor and Hesse (2011) but provide more information about the variability of the sample size and the impact of number of replication on TE measures.



(a) n=50 (b) n=100 (c) n=200

Figure 2.2. TE measures of exponential SFA Model for the first five replications

Next, we discuss additional factors known to influence SFA models. The first factor is misspecification - that is when the researcher applies an inefficiency distribution of SFA model that differs from the correct inefficiency distribution of the generated data or empirical data. We thus analyze the influence of the inefficiency distribution by comparing the results of the correct inefficiency distribution with misspecified inefficiency distributions. This is particularly interesting because we can analyze the effect of models' specification error.

2.5.3. Impact of the misspecified inefficiency distributions

A major question concerning SFA models is whether the choice of the underlying inefficiency distribution matters. To truly evaluate the performance of SFA models using different underlying assumptions, we need to get the residuals (that is the difference between the actual TE and the predicted TE). Accordingly, for different magnitudes of the sample size, we simulated each inefficiency distribution with misspecified SFA models. Tables 2.3-2.5 present the results of the misspecification within each approach.

The first column of Tables 2.3-2.5 represents respectively the estimation of the correct SFA models, exponential, half normal, and truncated normal. The next two SFA models are the misspecified models. Furthermore, a priori, based on the underlying statistical properties of the inefficiency distributions, the correct SFA models should provide the smallest MAD and MSE values. Hence, we would expect the misspecified SFA models to perform poorly relative to the correctly specified SFA models. Three interesting results emerge from Tables 2.3-2.5.

First, in Table 2.3 of the proposed approach, we simulated data from a shifted exponential inefficiency distribution and applied the correct exponential SFA model and two misspecified SFA models, truncated normal and half normal. For the conventional approach, we simulated from an exponential inefficiency distribution and applied the correct exponential SFA model and two misspecified SFA models, truncated normal and half normal. The results of the proposed approach suggest that the misspecified half normal SFA model provides the smallest variability in terms of MAD. This could be attributed to the tighter clustering of the TE values near zero for the correct exponential SFA model. However, in the conventional approach, the exponential SFA model provides the lowest variability in terms of MAD. These results are expected since the correct inefficiency distribution is exponential and both half normal and truncated normal are misspecified SFA models.

Second, in Table 2.4 and within each approach, we simulated data from a half normal inefficiency distribution, then applied the correct half normal SFA model and two misspecified SFA models, truncated normal and exponential. The results of the proposed approach suggest that the misspecified truncated normal SFA model provides the smallest variability in terms of MAD. This could be first attributed to the flexibility in modeling truncated SFA model. Another attribution could be due to the ratio of $\sigma_u\ /\sigma_v$ (Table 2.1). However, within the conventional approach, the correct half normal SFA model provides better results in comparison to the misspecified exponential and truncated normal SFA models.

Third, Table 2.5 displays the results of the misspecification by incorrectly applying an exponential and a half normal SFA models to the data generated with a truncated normal inefficiency distribution. Within the proposed approach, the truncated normal SFA model provides smaller MAD values in comparison to the misspecified SFA models, exponential and half normal. However, for the conventional approach, the misspecified half normal and exponential SFA models provide the smallest variability in terms of the MAD and MSE values.

Table 2.3. Misspecification of the exponential SFA model

n	Parameters	Standard production function		
		Exponential (correct model)	Truncated normal	Half normal
		Proposed approach		
50	Rank TE	0.621	0.507	0.463
	MAD	0.148	0.139	0.126
	MSE	0.031	0.031	0.027
100	Rank TE	0.746	0.673	0.641
	MAD	0.136	0.129	0.103
	MSE	0.025	0.024	0.017
200	Rank TE	0.770	0.743	0.750
	MAD	0.132	0.125	0.093
	MSE	0.023	0.021	0.014
		Conventional approach		
50	Rank TE	0.668	0.491	0.485
	MAD	0.101	0.142	0.146
	MSE	0.021	0.040	0.041
100	Rank TE	0.791	0.673	0.616
	MAD	0.077	0.102	0.111
	MSE	0.011	0.022	0.024
200	Rank TE	0.825	0.779	0.803
	MAD	0.070	0.075	0.096
	MSE	0.008	0.010	0.017

n: Sample size. Rank TE: Mean of Spearman rank correlation of TE. MSE: Mean Square Error. MAD: Mean Absolute Deviation. Production function: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. u_i is Exponential. Incorrect or misspecified SFA models include: Truncated normal and Half normal SFA models. $v_i \overset{iid}{\sim}$ Normal (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

Table 2.4. Misspecification of the half normal SFA model

n	Parameters	Standard production function		
		Half normal (correct model)	Truncated normal	Exponential
		Proposed approach		
50	Rank TE	0.558	0.474	0.346
	MAD	0.156	0.147	0.222
	MSE	0.037	0.035	0.067
100	Rank TE	0.657	0.530	0.406
	MAD	0.150	0.141	0.234
	MSE	0.035	0.033	0.072
200	Rank TE	0.697	0.625	0.447
	MAD	0.142	0.130	0.242
	MSE	0.030	0.029	0.075
		Conventional approach		
50	Rank TE	0.573	0.469	0.350
	MAD	0.089	0.152	0.135
	MSE	0.016	0.043	0.032
100	Rank TE	0.784	0.557	0.423
	MAD	0.089	0.121	0.116
	MSE	0.010	0.029	0.022
200	Rank TE	0.882	0.65	0.547
	MAD	0.090	0.091	0.108
	MSE	0.007	0.016	0.018

n: Sample size. Rank TE: Mean of Spearman rank correlation of TE. MSE: Mean Square Error. MAD: Mean Absolute Deviation. Production function: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. u_i is half normal. Incorrect or misspecified SFA models include: Truncated normal and Exponential. $v_i \overset{iid}{\sim}$ Normal (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

Table 2.5. Misspecification of the truncated normal SFA model

n	Parameters	Standard production function		
		Truncated normal (correct model)	Half normal	Exponential
		Proposed approach		
50	Rank TE	0.305	0.298	0.245
	MAD	0.138	0.175	0.248
	MSE	0.031	0.044	0.078
100	Rank TE	0.349	0.314	0.259
	MAD	0.134	0.192	0.268
	MSE	0.029	0.051	0.089
200	Rank TE	0.413	0.378	0.301
	MAD	0.128	0.209	0.281
	MSE	0.027	0.058	0.097
		Conventional approach		
50	Rank TE	0.212	0.200	0.167
	MAD	0.134	0.083	0.084
	MSE	0.029	0.012	0.013
100	Rank TE	0.248	0.234	0.205
	MAD	0.133	0.071	0.082
	MSE	0.020	0.008	0.012
200	Rank TE	0.318	0.268	0.245
	MAD	0.134	0.067	0.078
	MSE	0.012	0.007	0.010

n: Sample size. Rank TE: Mean of Spearman rank correlation of TE. MSE: Mean Square Error. MAD: Mean Absolute Deviation. Production function: $y_i = \ln(2)+0.5\ln x_{1i}+0.5\ln x_{2i}-u_i+v_i$. u_i is truncated normal. Incorrect or misspecified SFA models include: Half normal and Exponential. $v_i \overset{iid}{\sim}$ Normal (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

Figures 2.3 - 2.5 display the histograms of the residual[8] associated with Tables 2.3 - 2.5 of the proposed approach. From top to bottom, the first plot of Figure 2.3 represents the distribution of the correct exponential SFA model, the second and third plots represent the misspecified half normal and truncated normal SFA models. From Figure 2.4, the first and third plots represent the misspecified truncated normal and exponential SFA models respectively and the second plot represents the correct half normal SFA model. From Figure 2.5, the first and second plots represent the misspecified exponential and half normal SFA models and the third plot represents the correct truncated normal SFA model.

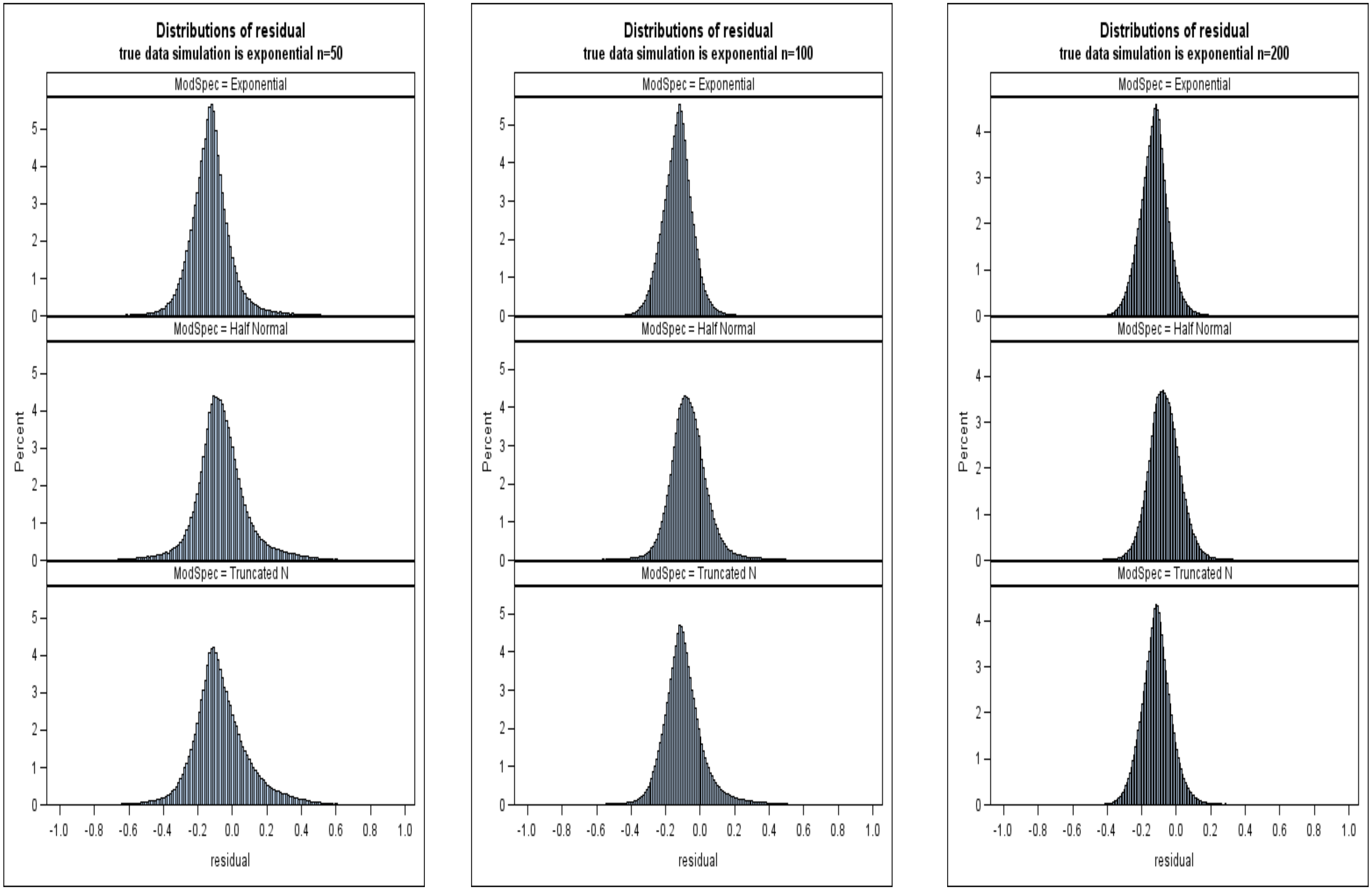
Figures 2.6 - 2.8 display the histograms of the residual associated with Tables 2.3 - 2.5 of the conventional approach. From top to bottom, the first plot of Figure 2.6 represents the correct exponential SFA model and the second and third plots represent the misspecified half normal and truncated normal SFA models. From Figure 2.7, the first and third plots represent respectively the misspecified exponential and truncated normal SFA models and the second plot represents the correct half normal SFA model. From Figure 2.8, the first and second plots represent respectively the distribution of the misspecified exponential and half normal SFA models and the third plot represents the correct truncated normal SFA model.

Overall, Figures 2.3 - 2.8 suggest that within each approach, some of the residual's distributions of the correct and misspecified SFA models are quite different. For each misspecified SFA model, the researcher is either over or under estimating TE measures based on the MAD values. The magnitude of this estimation is based on correctly identifying the inefficiency distributions (SFA models).

In conclusion, the results suggest that the TE measures may be more sensitive to the measurement technique within the conventional approach in comparison to the proposed approach. For consistent MC simulations, each distribution is a combination of all 5000 samples of different sizes. Hence, the important features to compare are the shapes within a column. Furthermore, within the proposed approach (Figures 2.3-2.5), the residuals are not centered at zero. This suggests the presence of significant inefficiency distributions. In contrast, within the conventional approach, the results in Figures 2.6-2.8 show that the residuals are in some cases centered at zero. This

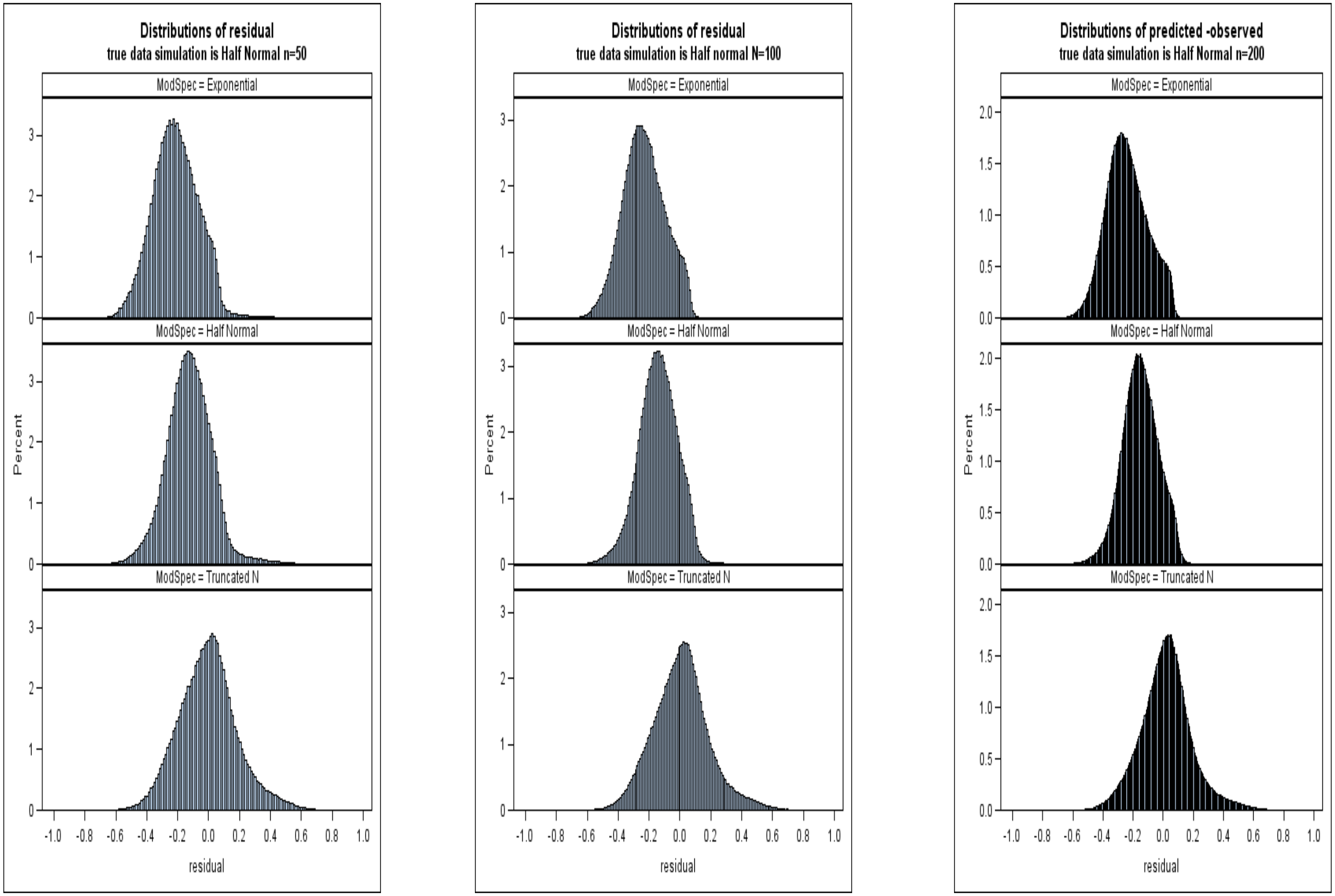
[8] The residuals are the differences between the actual TE and the predicted TE. The results show that the estimated TE are statistically different which makes sense in practice (Baccouche and Kouki, 2003).

might indicate that the inefficiency distribution was insignificant or the lack of inefficiency term during misspecification. Moreover, comparing the exponential SFA model from the proposed approach (Figure 2.3) to the conventional approach (Figure 2.6), the results show that the residuals in Figure 2.3 are clearly shifted negatively whereas in Figure 2.6, the residuals are centered at zero. The contrast results in both approaches is due to the choice of the parameters associated with the inefficiency distributions of the conventional approach.



(a) n=50 $\times$ 5000 (b) n=100 $\times$ 5000 (c) n=200 $\times$ 5000

Figure 2.3. Proposed approach - Residual of misspecification using exponential data.

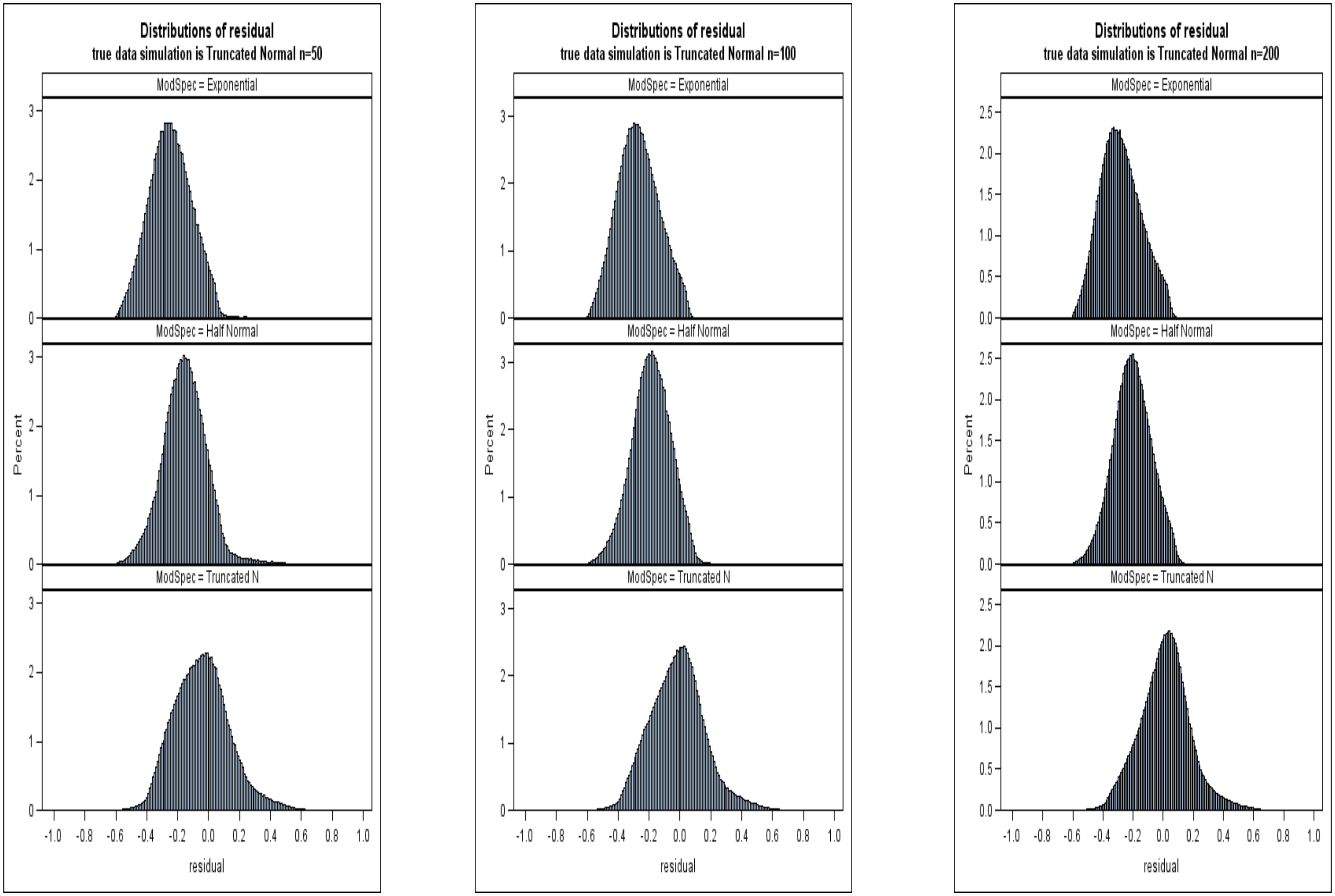


(a) n=50 × 5000

(b) n=100 × 5000

(c) n=200 × 5000

Figure 2.4. Proposed approach - Residual of misspecification using half normal data.

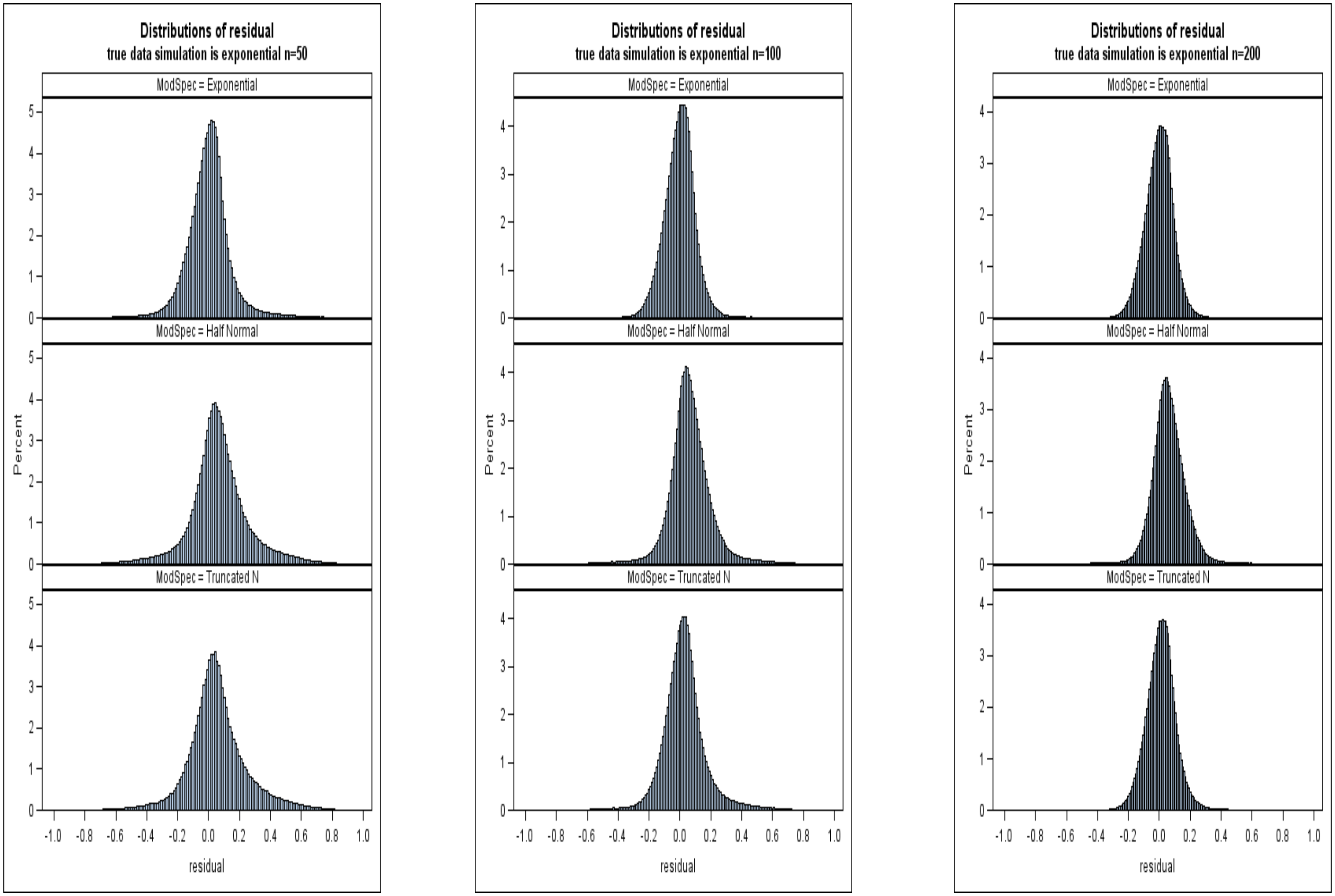


(a) n=50 × 5000

(b) n=100 × 5000

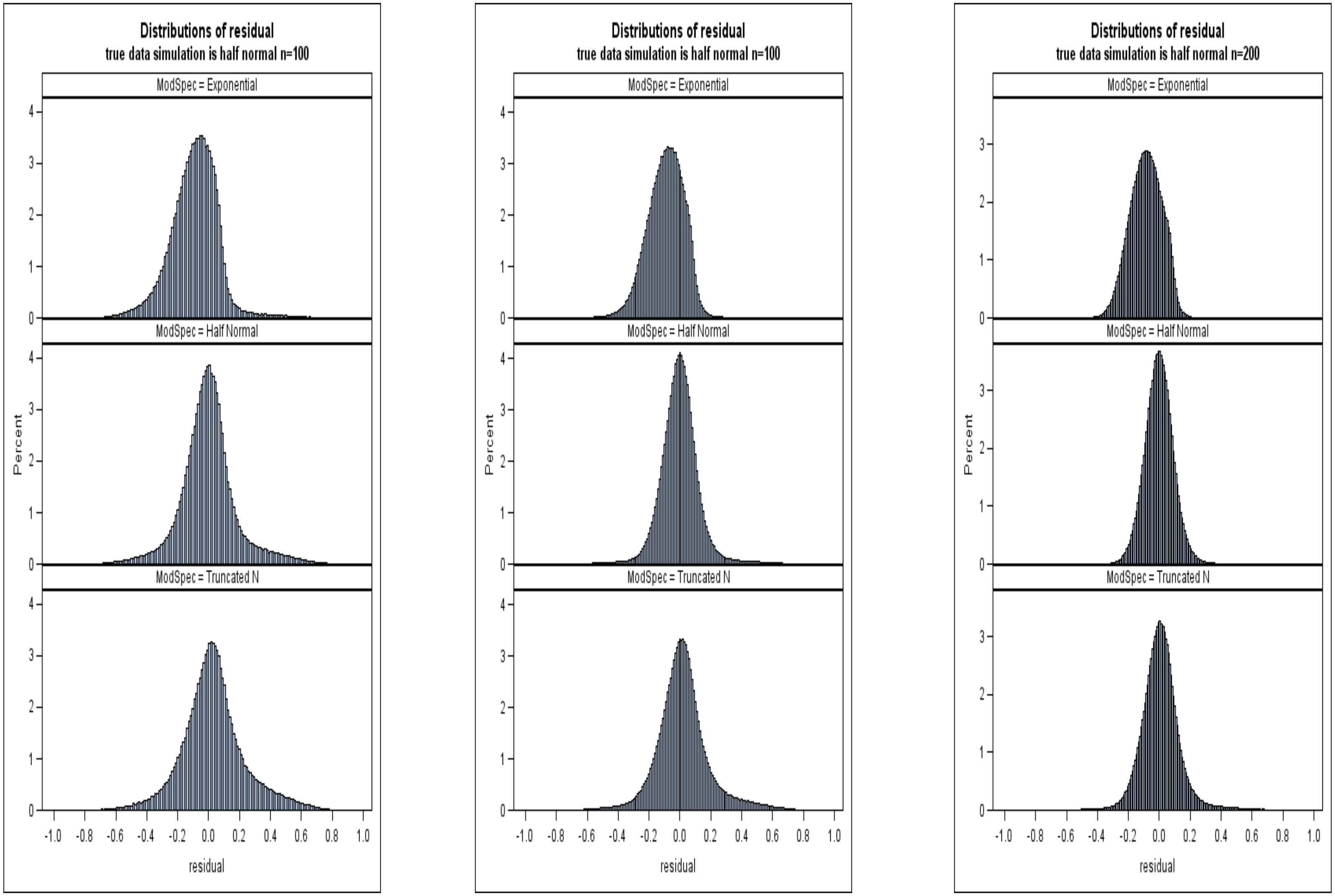
(c) n=200 × 5000

Figure 2.5. Proposed approach - Residual of misspecification using truncated normal data.



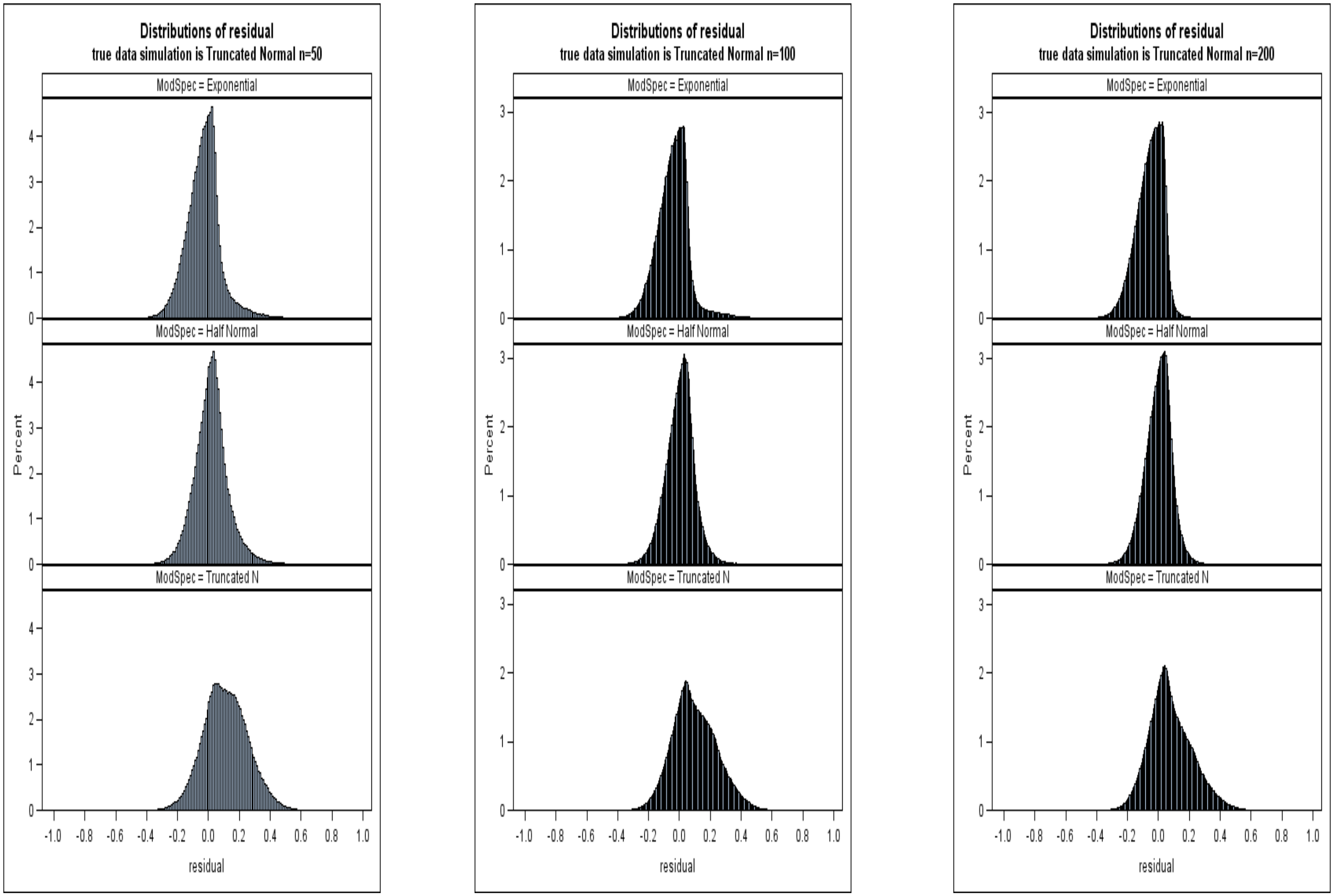
(a) n=50 × 5000 (b) n=100 × 5000 (c) n=200 × 5000

Figure 2.6. Conventional approach- Residual of misspecification using exponential data.



(a) n=50 × 5000 (b) n=100 × 5000 (c) n=200 × 5000

Figure 2.7. Conventional approach - Residual of misspecification using half normal data.



(a) n=50 × 5000 (b) n=100 × 5000 (c) n=200 × 5000

Figure 2.8. Conventional approach - Residual of misspecification using truncated normal data.

Overall, the results of the misspecification suggest that within the proposed approach, the truncated normal SFA model performs better in terms of MAD and MSE when the correct inefficiency distributions are half normal and truncated normal. Hence, we further assess the MC simulations results within the proposed approach. Since the SFA models are fitted to the same data within a sample during misspecification, we can look at the performance keeping that connection. Therefore, the idea of identifying which model is the best during misspecification could also be applied to the performance creteria of MAD and MSE since in both cases we are presumably interested in the model that minimizes these criteria. Of course, the 'best' model based on the MAD values might be only a tiny bit better than the next best model, so this approach is blind to this possibility. The results in Tables 2.6-2.8 summarize the main conclusions of the proposed approach during misspecification.

From Tables 2.6-2.8, we simulated the correct inefficiency distributions (actual SFA models) then fitted misspecified SFA models assuming exponential, truncated normal, and half normal while keeping track of the best MAD values- that is the minimum MAD values across observation. As result, Tables 2.6-2.8 respectively present the percentage of the best MAD value for sample sizes of 50, 100, and 200 with 5000 replications. In addition, the diagonals of Tables 2.6-2.8 represent the correct SFA models (with correct inefficiency distributions) and the off-diagonals are the misspecified SFA models.

From Table 2.6, a correct exponential SFA model was fitted in addition to two misspecified SFA models, truncated normal and half normal. The results suggest that: 1) The correct exponential SFA model is identified as the 'best' model 19% of the time 2) The misspecified truncated normal SFA model is identified as the 'best' model 46% of the time. 3) The misspecified half normal SFA model is identified as the 'best' model 35% of the time. Concerning the correct truncated normal inefficiency distribution (truncated normal SFA model), the results suggest the following: 1) The misspecified exponential SFA model is identified as the 'best' model 11% of the time. 2) The correct truncated normal SFA model is identified as the 'best' model 70% of the time. 3) The misspecified half normal SFA model is identified as the 'best' model 19% of the time. For the correct half normal distribution inefficiency distribution (half normal SFA model), the results suggest that: 1) The misspecified exponential SFA model is identified as the 'best' model 9.31% of the time. 2) The misspecified truncated normal SFA model is identified as the 'best' model 66.36%

of the time. 3) The correct half normal SFA model is identified as the 'best' model 24.34% of the time. In conclusion, the results in Table 2.6 suggest that the truncated normal has the highest percentage of providing the minimum MAD value.

In addition, Tables 2.7 and 2.8 respectively present the percentage of the best MAD values for sample sizes of 100 and 200 with 5000 replications. The results of Tables 2.7 and 2.8 suggest that the truncated normal SFA model provides the highest percentage of the best MAD value. In Table 2.7, when the correct inefficiency distribution is exponential (exponential SFA model), the misspecified truncated normal and half normal SFA models are respectively identified as the 'best' models 20.77% and 59.03% of the time. In addition, when the sample size increases to 200 (Table 2.8), for the correct exponential inefficiency distribution (exponential SFA model), the misspecified truncated normal and half normal SFA models are respectively identified as the 'best' models 19.93% and 81.27% of the time. Hence, for smaller sample size, the truncated is a good choice, but the half normal is better when the true distribution is exponential for large sample size (n=200). Moreover, the misspecified truncated normal provides the highest percentage of the best MAD only when the correct SFA models are truncated normal and half normal. Next, the second factor that can influence the performance of SFA models is the input distributions.

Table 2.6. Confusion matrix of the proposed approach using MAD_{50}

Actual SFA models	Misspecified SFA models		
	Exponential	Truncated normal	Half normal
Exponential	18.99	46.47	34.54
Truncated normal	10.70	70.27	19.03
Half normal	9.31	66.36	24.34

Production function: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. Sample size of 50. $v_i \overset{iid}{\sim}$ N (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15]. This table provides the percentage of the frequency count.

Table 2.7. Confusion matrix of the proposed approach using MAD_{100}

Actual SFA models	Misspecified SFA models		
	Exponential	Truncated normal	Half normal
Exponential	17.46	20.77	59.03
Truncated normal	8.96	72.77	18.52
Half normal	3.39	61.68	34.92

Production function: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. Sample size of 100. $v_i \overset{iid}{\sim}$ N (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15]. This table provides the percentage of the frequency count.

Table 2.8. Confusion matrix of the proposed approach using MAD_{200}

Actual SFA models	Misspecified SFA models		
	Exponential	Truncated normal	Half normal
Exponential	7.80	10.93	81.27
Truncated normal	10.06	77.98	11.97
Half normal	10.41	55.73	33.86

Production function: $y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} - u_i + v_i$. Sample size of 200. $v_i \overset{iid}{\sim}$ N (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15]. This table provides the percentage of the frequency count.

2.5.4. Impact of input distributions

The performance of SFA models has been dependent on the input distributions (Andor and Hesse, 2011 and 2013). Research has shown that not only the production functions impact TE measures, but the input distributions play an important role. Table 2.9 summarizes the performance measures of the Cobb-Douglas production function with the property of CRS using normal-normal and weibull-gamma input distributions based on FCA financial data. The results are subdivided into two blocks, proposed and conventional approaches of the inefficiency distributions. Within each approach, the results suggest that the SFA models are not impacted by the input distributions in terms of the MAD and MSE values. Similar conclusions were reported by Andor and Hesse (2011).

Table 2.9. Impact of the input distributions

n	Paramters	Normal and Normal			Gamma and Weibull		
		EXP	TN	HN	EXP	TN	HN
		Proposed approach					
50	Rank TE	0.621	0.323	0.572	0.633	0.335	0.560
	MAD	0.148	0.136	0.156	0.147	0.134	0.155
	MSE	0.031	0.030	0.036	0.029	0.029	0.039
100	Rank TE	0.745	0.359	0.658	0.745	0.368	0.656
	MAD	0.137	0.133	0.152	0.137	0.132	0.152
	MSE	0.025	0.029	0.034	0.024	0.028	0.036
200	Rank TE	0.770	0.418	0.700	0.772	0.425	0.698
	MAD	0.133	0.128	0.149	0.133	0.127	0.148
	MSE	0.023	0.027	0.031	0.023	0.026	0.032
		Conventional approach					
50	Rank TE	0.669	0.218	0.579	0.675	0.228	0.576
	MAD	0.100	0.166	0.127	0.099	0.172	0.127
	MSE	0.020	0.040	0.032	0.018	0.042	0.034
100	Rank TE	0.793	0.227	0.785	0.794	0.238	0.788
	MAD	0.077	0.158	0.089	0.074	0.164	0.085
	MSE	0.011	0.038	0.016	0.010	0.040	0.016
200	Rank TE	0.825	0.259	0.871	0.825	0.256	0.871
	MAD	0.070	0.139	0.073	0.703	0.148	0.072
	MSE	0.008	0.032	0.009	0.007	0.035	0.008

Rank TE: Mean Spearman rank-order correlation efficiency. MSE: Mean Square Error. MAD: Mean Absolute Deviation. n: Sample size. Inpu distributions: $x_1 \sim Normal$ (8.4, 1.51) and $x_2 \sim Normal$ (7.92, 1.18) and $x_1 \sim Weibull$ (17.4, 20.5) and $x_2 \sim Gamma$ (285, 0.07). TN: Truncated normal SFA model. HN: Half normal SFA model. EXP: Exponential SFA model.

The third factor that can influence the performance of SFA models is the functional form of the production functions. While applying SFA models, it is impossible to know beforehand the type of production functions that can provide a better performance. Different production functions result into different performance measures. Hence, we considered three classes of production functions, the Cobb-Douglas production functions with the properties of IRS and DRS and translog production functions. Misspecification of the production functions arises due to wrongly assuming IRS, DRS, and translog production functions. That is when the researcher assumes a production function that is different from the correct production function of the data.

2.5.5. Impact of production functions

The influence of the production functions on the performance of SFA models is well known in the literature (Gong and Sickles, 1992; Banker et al., 1993; Jensen, 2005; Perelman and Santin, 2009; and Andor and Hesse, 2011). Throughout this chapter, the Cobb-Douglas production function with the property of CRS has been used to analyze the impact of the sample size, the estimation of TE measures and the impact of misspecification.

Accordingly, we generated data using the restrictive Cobb-Douglas production functions with the properties of IRS and DRS and a more flexible and homothetic, translog production function. Table 2.10 presents the results of the MC simulation studies involving 5000 replications across IRS of 1.2, DRS of 0.8 and translog production functions. The results of Table 2.10 are partitioned within the proposed and conventional approaches of the inefficiency distributions. Within each approach, the performance criteria of Rank TE, MAD, and MSE are presented. Concerning both approaches, three important results emerge from Table 2.10.

First, the performance of the SFA models is impacted by the sample size across the different production functions. As the sample size increases, the average Rank TE increases and MAD and MSE values decrease. Similar results were summarized in Ondrich and Ruggiero (2001) and Andor and Hesse (2013).

Second, for the exponential and truncated normal SFA models, MAD and MSE values provide quite similar performances. In addition, for the half normal SFA model, the performance of MAD and MSE values are different across the production functions. Moreover, the performance of MAD and MSE values was consistent with the a prior expectations formulated by Andor and Hesse (2011). For each SFA model, the average Rank TE provides different rate of trend within the proposed and conventional approaches. For example, for the truncated normal SFA model of the proposed approach, the magnitude of Rank TE is higher in comparison to the conventional approach.

Third, within production function, the results suggest that there is a significant influence on the rank TE when the researchers use the Cobb-Douglas production functions with the properties of IRS and DRS, or translog production function within each approach of the inefficiency distributions. This agrees with the work of Andor and Hesse (2013). In summary, regarding the comparison of the inefficiency distributions for the proposed and conventional approaches of the production functions, the results suggest that SFA models are affected.

Table 2.10. Impact of the production functions

n	Parameters	IRS			DRS			TL		
		EXP	TN	HN	EXP	TN	HN	EXP	TN	HN
		Proposed Approach								
50	Rank TE	0.622	0.309	0.557	0.616	0.306	0.506	0.617	0.302	0.564
	MAD	0.148	0.138	0.156	0.148	0.139	0.155	0.147	0.139	0.155
	MSE	0.031	0.032	0.037	0.031	0.032	0.037	0.031	0.031	0.037
100	Rank TE	0.746	0.343	0.661	0.742	0.352	0.658	0.744	0.342	0.656
	MAD	0.138	0.134	0.154	0.137	0.133	0.153	0.138	0.134	0.151
	MSE	0.025	0.030	0.033	0.025	0.030	0.034	0.025	0.029	0.035
200	Rank TE	0.770	0.409	0.709	0.768	0.410	0.698	0.769	0.407	0.701
	MAD	0.133	0.128	0.148	0.132	0.129	0.150	0.133	0.129	0.143
	MSE	0.023	0.028	0.029	0.023	0.027	0.030	0.023	0.027	0.032
		Conventional Approach								
50	Rank TE	0.668	0.216	0.566	0.674	0.217	0.506	0.667	0.214	0.571
	MAD	0.100	0.143	0.130	0.099	0.144	0.131	0.100	0.143	0.129
	MSE	0.021	0.032	0.033	0.020	0.032	0.034	0.020	0.032	0.033
100	Rank TE	0.793	0.251	0.786	0.792	0.254	0.781	0.794	0.252	0.781
	MAD	0.077	0.133	0.089	0.077	0.134	0.090	0.076	0.134	0.090
	MSE	0.011	0.029	0.016	0.010	0.029	0.017	0.010	0.029	0.016
200	Rank TE	0.824	0.292	0.872	0.825	0.295	0.871	0.824	0.297	0.872
	MAD	0.070	0.120	0.072	0.070	0.122	0.072	0.072	0.121	0.070
	MSE	0.008	0.025	0.009	0.008	0.025	0.009	0.009	0.025	0.011

n: Sample size. Rank TE: Mean of Spearman rank correlation of TE. MSE: Mean Square Error; MAD: Mean Absolute Deviation. IRS: increasing return to scale. DRS: decreasing return to scale. TL: Translog. Input distributions: $x_{1i} \overset{iid}{\sim}$ Uniform [5,15] and $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

2.6. Conclusion

Stochastic Frontier Analysis (SFA), an econometric approach, requires the researcher to make an assumption about functional form of the inefficiency distributions which could bias the results of the technical efficiency measures. Hence, the contribution of this chapter is two-fold.

First, to assess the prediction accuracy of TE measures and minimize its degree of bias, we present an approach that correctly derives and identifies the expected population mean and standard deviation of half normal, exponential and truncated normal inefficiency distributions. To the best of our knowledge, this is the first research that provides a comprehensive comparison of the three stochastic frontier models while deriving a consistent estimate of their respective statistical parameters.

Second, we built on this theoretical framework by shedding more light on a systematic comparison between the technical efficiencies of the inefficiency distributions through MC simulations. To generalize our simulations, we compare the inefficiency distributions of the proposed and the conventional approaches toward misspecification. The design of MC simulations was conceptually done in a manner consistent with the statistical theory of comparing different distributions. To that effect, the exponential inefficiency distribution was changed to a two-parameter distribution. Additionally, this research extended other recent papers to evaluate the performance of SFA models in MC simulations. Therefore, the impact of misspecified SFA models was evaluated relative to the correctly specified SFA model.

In conclusion, the truncated normal SFA model within the proposed approach provided smaller variation when the correct inefficiency distribution generated was from a half normal inefficiency distribution. However, within the conventional approach, the correct inefficiency distributions of half normal and exponential provided the smaller variation during misspecification. The results of our MC simulations suggest that SFA may be sensitive to misspecification..

SFA works well when the researcher applies the truncated normal model, but for that to be useful information, we must know in advance that the statistical parameters of the inefficiency distributions are comparable in terms of sample mean and sample standard deviation. Absent from that, the results of the MC simulations suggest that SFA can give very misleading results, especially as far as the least efficient firms (the ones in which we are, presumably, most interested) are concerned. Furthermore, the results suggest that TE measures of the inefficiency distributions of the truncated normal, half normal and exponential matter.

3. MISSPECIFICATION OF DEA EFFICIENCY BASED ON INEFFICIENCY DISTRIBUTIONS

3.1. Abstract

This chapter examines several misspecifications of Data Envelopment Analysis (DEA) efficiency measures while accounting for the stochastic inefficiency distributions of truncated normal, half normal, and exponential. Within the proposed and conventional approaches of the inefficiency distributions, the technical efficiency measures of the input oriented variable return to scale DEA model are estimated. Monte Carlo simulations were conducted to study the impact of the inefficiency distributions by examining the performance of DEA under two different data generating processes, logarithm and level, and across five different scenarios, inefficiency distributions, sample sizes, production functions, input distributions, and curse of dimensionality.

3.2. Introduction

In the application of Stochastic Frontier Analysis (SFA) models to evaluate the efficiency measures of decisions making units (DMUs) such as banks, each inefficiency distribution results in different performance of technical efficiency (TE) measures (Baccouche and Kouki, 2003). Hence, with the inefficiency distribution limitations of SFA models, Charnes et al., (1978) reformed the piecewise linear convex approach of Farrell (1957) and Aigner and Chu (1968) into a mathematical linear programming method known as Data Envelopment Analysis (DEA). Contrary to SFA, Ruggiero (2004) states that, "Since DEA compares a given DMU's observed outputs and inputs, it has been criticized because measurement error and other statistical noise are not accounted for".

In the literature, Monte Carlo (MC) simulation studies have concluded that the random error can cause insignificant and inconsistent results in DEA estimation. See Grosskopf and Valdmanis (1987), Gong and Sickles (1992), Banker et al., (1993), Wilson (1993), Andersen and Petersen (1993), Dusansky and Wilson (1994, 1995), Kittelsen (1995), Pastor et al., (1999), Lee and Holland (2000), Holland and Lee (2002), Fried et al., (2002), Ondrich and Ruggiero (2002), Banker and Chang (2006), and Simar and Zelenyuk (2011). In this chapter, we explore the measurement error associated with efficient production. In the analysis of DMUs, the measurement error of the input and output variables is not only subject to random error but to inefficiency error.[1] The distributional assumptions of the inefficiency error are important to both policy makers and managers because the performance of DMUs operating under optimal capacity depend on the inefficiency distributions. Therefore, the researcher must determine the values for the parameters of the respective inefficiency distributions such that the distributions are comparable (promoting an "apples-to-apples" type of comparison) in order to decide the most efficient inefficient distributions.

However, the impact of the statistically driven inefficiency distribution assumptions on DEA efficiency has received little attention. Analyses of Cooper and Tone (1997), Ruggiero (1999), Ondrich and Ruggiero (2001), Ruggiero (2004), and Olessen and Petersen (2016) provided MC simulations based on inefficiency distributions that fail to account for: 1) The correct statistical parameters of the inefficiency distributions and 2) The lack of a comparative statistical approach of the inefficiency distributions of half normal (Meeusen and Van den Broeck, 1977), exponential (Aigner et al., 1977) or truncated normal (Stevenson, 1980) in a comprehensive manner. We refer to this type of simulation as "The conventional approach." In addition, the conventional approach focuses on specifying a random reference inefficiency distribution on the observed inputs in the generation of the observed outputs with supports in the input-output space. Therefore, to provide accurate and comparative results in the estimation of DEA TE measures, the statistical parameters of the inefficiency distributions should be consistent and on similar and comparable scales.

More specifically, in this chapter, we assess the impact of the statistically driven inefficiency distributions on DEA efficiency measures following the process outlined in chapter 2. Chapter 2 derived the correct values for the distributional parameters in SFA setting to facilitate an "apples-to-apples" comparison across the inefficiency distributions; we refer to this as "The proposed approach". This approach, thus, minimizes the bias in DEA estimation and the specification error across the inefficiency distributions.

To provide a comparison within the proposed and conventional approaches, one research objective formulates a consistent simulation by generating data under two schools of thought. The first school is oriented toward the traditional SFA cross-sectional production function in logarithm

[1] The deviations of the observed choices from the optimal ones are due to the failure to optimize - that is the inefficiency due to random shocks.

scale (Aigner et al., 1977 and Holland and Lee, 2002). The second school is oriented toward the traditional DEA of Charnes et al., (1978) without including a random noise distribution. A second research objective examines the impact of inefficiency distributions on the performance of DEA efficiency measures.[2] Finally, a third research objective examines the impact of the efficiency measures on a misspecified DEA model. The problem of misspecification in DEA is not new (Smith 1997). The key issue in applying the misspecification technique is to study the robustness within each approach. Misspecification in DEA model can arise due to the several factors including: inefficiency distributions, sample sizes, production functions, input distributions, and curse of dimensionality.

The rest of this chapter is organized into four sections: Section 3.3 presents the theoretical framework. We present the input distance DEA model and the statistical parameters of the proposed and conventional approaches. Section 3.4 discusses the standard design of MC simulations. Section 3.5 deals with the simulation results. Finally, the summary of our findings are presented in Section 3.6.

3.3. DEA theoretical framework

The technology that transforms i inputs $x = (x_1, x_2, ..., x_i) \in \mathbb{R}^I_+$ into outputs, j, $y = (y_1, x_2, ..., y_j) \in \mathbb{R}^J_+$ is represented by the input set, $L(y)$. The input set satisfying constant returns to scale (CRS) and strong disposability of outputs and inputs is defined as:

$$L(y) = \{x : \text{y can produce x}; \quad x \in \mathbb{R}^I_+ \quad and \quad y \in \mathbb{R}^J_+\} \tag{3.1}$$

The input set denotes the collection of input vector that yield output vector. Non-parametric output, input, and graph efficiency measures are based on the distance functions from output, input, and graph sets, respectively. Input distance function is defined in terms of scalar shrinkage of observed inputs with output held fixed. An input distance function evaluated for any bank, t, using reference production possibilities set, T, is represented as:

$$D^T_i(y^t, x^t)^{-1} = \min\{\lambda : (\lambda x^t \in L^T(y^t)\} \tag{3.2}$$

[2] The inefficiency distributions across both approaches are generated using an acceptance-rejection method of MC Simulation.

or

$$
\begin{aligned}
& \min_{\theta z} \\
\text{subject to} \quad y^t &\leq Yz,\ Y = y_1, \ldots, y_T. \\
\lambda x^t &\geq Xz,\ X = x_1, \ldots, x_T. \\
z &\geq 0
\end{aligned}
\tag{3.3}
$$

Here, equation (3.3) identifies the linear program that is used to calculate the distance function, with the z's being a Dx1 vector of intensity variables that identify CRS boundaries of the reference set. Under the variable return to scale (VRS) DEA model, the intensity variable z=1. Thus, a stochastic aspect of the efficiency measures is added to the production function and examined for characterizing the empirical distribution. Following the process outlined in chapter 2, to compare two or more inefficiency distributions, we must identify the statistical parameters of the inefficiency distributions such that they are on a similar scale.

Within the proposed approach, the inefficiency distributions are generated as follows:

1. Truncated normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal (0.46, 0.2186).
   (b) Step 2: If $0 < u_i < 1$ then accept u_i. Otherwise re-generate u_i.
2. Half normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal (0.46, 0.2186).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i. This process is similar to taking the absolute of u_i.
3. Exponential:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Exponential (0.2186).
   (b) Step 2: If $u_i > 0$ then accept u_i and add 0.1784. Otherwise re-generate u_i.

To further analyze the extensions of our MC simulations, we followed the works of Behr and Tente (2008), Andor and Hesse (2011), and Hafner et al., (2016) for the conventional approach. The generation of the inefficiency distributions is as follows:

1. Truncated normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal(0, 0.20).
   (b) Step 2: If $0 < u_i < 1$ then accept u_i. Otherwise re-generate u_i.
2. Half normal:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Normal(0, 0.587).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i. This process is similar to taking the absolute of u_i.
3. Exponential:
   (a) Step 1: Generate $u_i \overset{iid}{\sim}$ Exponential (0.35).
   (b) Step 2: If $u_i > 0$ then accept u_i. Otherwise re-generate u_i.

3.4. Standard MC simulations

The standard VRS DEA model assumes that the input and output variables are given by fixed values. To study the impact of the inefficiency distributions on the performance of VRS DEA efficiency measures within each approach, we performed MC simulations with the probabilistic inefficiency distributions of truncated normal, half normal, and exponential. The design of MC simulations was as follow:

3.4.1. Variation of sample size

One aim of this chapter was to establish the relationship between mean TE measures and sample sizes. Zhang and Bartels (1998) showed that as the sample size increases the estimated mean TE measures decrease. Hence, three sets of sample sizes of 50, 100, and 200 were considered with 5000 replications (Banker, et al., 1993).

3.4.2. Input distributions

Following Lee and Holland (2000), we assumed two inputs x_{1i} and x_{2i} independently[3] drawn from a uniform distribution in the interval [5, 15].

[3]Zhang and Bartels (1998) concluded that the mean technical efficiency does not seem to depend on the values of the correlation coefficient of the inputs. Additional input distributions based on the Farm Credit Administration data sets are applied in the results section.

3.4.3. Production functions

During DGP, two experimental frameworks were conducted. The first framework, F_1, included the intercept and the random noise distribution, v_i, independently drawn from a normal distribution with mean 0 and standard deviation of 0.15. In F_1, the input distributions were transformed in logarithm. Accordingly, for each DMU, i, a single output, y_i, the Cobb-Douglas production function with the property of CRS was used to generate the data dependent on x_{1i} and x_{2i}. The production function of F_1 was:

$$lny_i = ln(2) + 0.5lnx_{1i} + 0.5lnx_{2i} - u_i + v_i. \tag{3.4}$$

F_1 addresses the importance of the random noise distribution of DEA performance. By the introduction of v_i, F_1 induces bias in DEA efficiency measures which is of an importance in the analysis of DMUs. The second framework, F_2, was in level and the input and output variables were not affected by v_i. Moreover, F_2 examined the impact of the inefficiency distributions on DEA TE measures. For each i^{th}DMU, a single output, y_i, dependent on x_{1i} and x_{2i}, the Cobb-Douglas production function with the property of CRS used to generate the data was:

$$y_i = 0.5x_{1i} + 0.5x_{2i} - u_i. \tag{3.5}$$

3.4.4. Misspecification

The robustness analysis of the VRS DEA model within the proposed and conventional approaches and across both F_1 and F_2 was examined in three scenarios. In the first scenario, we assumed that the baseline production function of CRS was misspecified and considered three additional production functions: increasing return to scale (IRS) and decreasing return to scale (DRS) Cobb-Douglas and translog production functions. Doing so, we examined the impact of the homogenous Cobb-Douglas production function and the non-homogenous translog production function on the DEA TE measures. In the second scenario, we assumed that DGP of the input distributions was wrongly misspecified, and hence DEA model was sensitive to input data which is a fundamental step in DEA analysis (Santos et al., 2011; Wu et al., 2016; and Mullarkey et al., 2015). The effects of different levels of input distributions were investigated using the Financial

Credit Administration data. The secondary input distributions considered were:

1. $x_1 \overset{iid}{\sim} Normal$ (8.4, 1.51) and $x_2 \overset{iid}{\sim} Normal$ (7.92, 1.18).

2. $x_1 \overset{iid}{\sim} Weibull$ (17.4, 20.5) and $x_2 \overset{iid}{\sim} Gamma$ (285, 0.07).

Finally, the third scenario dealt with the curse of dimensionality. The performance of DEA TE measures can be biased without a proportional increase of the number of observations due to the curse of dimensionality (Daraio, and Simar 2007; Pastor, et al., 2002; and Jenkins and Anderson, 2003). We studied the impact of the curse of dimensionality by increasing the number of inputs from two to three, four, and five.

3.4.5. Performance criteria

To compare the performance of DEA TE measures based on different inefficiency distributions, we calculated the mean squared error (MSE) and mean absolute deviation (MAD) between the true and estimated TE measures. MSE and MAD are defined as:

$$MSE = \frac{1}{rK} \sum_{r=1}^{R} \sum_{k=1}^{K} (\widehat{TE_{i,r}} - TE_{i,r})^2 \tag{3.6}$$

$$MAD = \frac{1}{rK} \sum_{r=1}^{R} \sum_{k=1}^{K} |(\widehat{TE_{i,r}} - TE_{i,r})| \tag{3.7}$$

where $TE_{i,r} = exp(-u_{i,r})$ and $\widehat{TE_{i,r}} = exp(-E(u_{i,r} \mid \varepsilon_i])_{i,r})$. r is the number of replications. $\widehat{TE_{i,r}}$ and $TE_{i,r}$ denotes respectively the predicted and actual DEA TE measures.

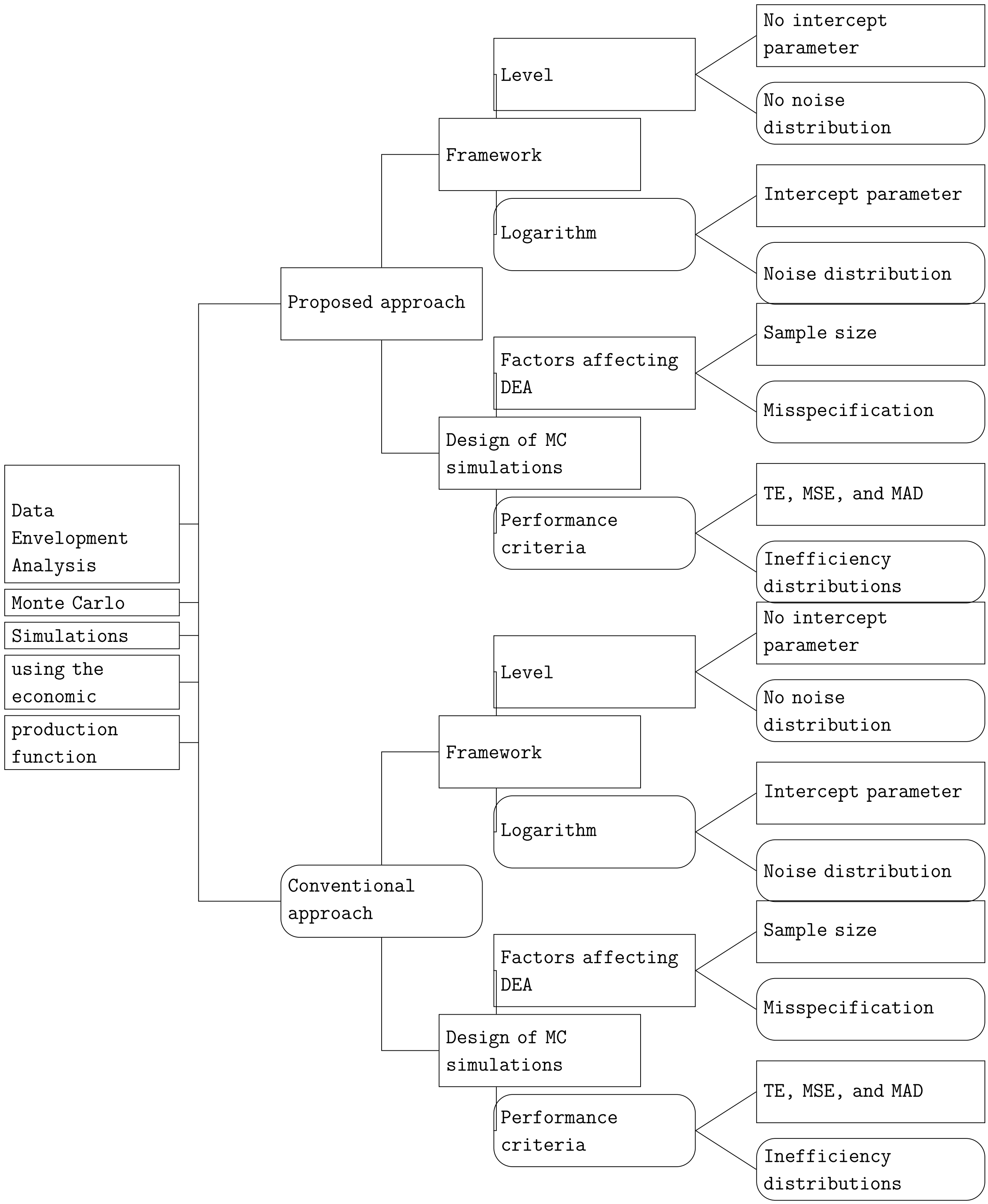


Figure 3.1. Design of the DEA MC simulations

3.5. Simulation results

This chapter compares the performance of two DGPs present in the DEA literature within the proposed and conventional approaches of the inefficiency distributions. An acceptance-rejection method of MC simulations was used to generalize the comparison as to which framework was preferable across half normal, exponential, and truncated normal inefficiency distributions. The situations examined include: 1) The impact of TE measures; 2) The impact of production functions; 3) The impact of input distributions; and 4) The impact of curse of dimensionality.

3.5.1. Impact of the sample size

The estimates of mean TE measures yielded by the VRS DEA model for F_1 and F_2 are respectively in Tables 3.1 and 3.2. The results in Tables 3.1 and 3.2 are partitioned within the proposed and conventional approaches of the inefficiency distributions. Additionally, within each approach, the performance of mean TE, MAD, and MSE are provided across the sample sizes of 50, 100, and 200.

Within the proposed and conventional approaches of the inefficiency distributions, the results suggest that the sample size matters when estimating the structural DEA efficiency measures. As the sample size increases, the mean TE decreases accordingly. These results are supported by Zhang and Bartels (1998) and Podnovski and Thanassoulis (2007). Another important finding is the rate in the mean TE measures. The rate of decrease in the mean TE measures from Tables 3.1 and 3.2 is dependent on the presence of the random noise distribution in Table 3.1.

The results in Tables 3.1 and 3.2 suggest that in the presence of the random noise distribution, DEA efficiency measures cause cautions. Thus, the DEA efficiency measures remain poor in terms of mean TE. This stands to reason because the random noise distribution more likely impacts the performance of DEA model.

Concerning MAD, within each approach of F_1 and F_2, the results of Tables 3.1 and 3.2 provide declining values as the sample size increases. A low MAD value implies that on an average the estimated TE measures is close to the true TE measures. Therefore, it is preferable to obtain small MAD values. Moreover, within the proposed approach, the exponential and half normal inefficiency distributions provide smaller MAD values (Tables 3.1 and 3.2) in comparison to the truncated normal inefficiency distribution of the conventional approach (Tables 3.1 and 3.2).

Overall, as the sample size increases, the rate of change in MAD values is affected by the presence of the random noise distribution (Tables 3.1 and 3.2). However, contrary to Table 3.1, Table 3.2 provides considerably higher MAD values.

Table 3.1. Average technical efficiency measures using F_1

n	Parameters	Proposed approach			Conventional approach		
		EXP	TN	HN	EXP	TN	HN
		Performance criteria					
50	TE	0.908	0.899	0.897	0.904	0.919	0.897
	MAD	0.255	0.260	0.254	0.176	0.086	0.240
	MSE	0.082	0.084	0.084	0.056	0.013	0.086
100	TE	0.891	0.881	0.879	0.887	0.902	0.878
	MAD	0.239	0.242	0.238	0.166	0.081	0.225
	MSE	0.074	0.075	0.075	0.052	0.011	0.078
200	TE	0.876	0.865	0.863	0.872	0.887	0.864
	MAD	0.224	0.228	0.224	0.159	0.078	0.214
	MSE	0.067	0.068	0.069	0.049	0.010	0.073

n: Sample size. EXP: Exponential. TN: Truncated normal. HN: Half normal. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. $F_1 : y_i = ln(2) + 0.5lnx_{1i} + 0.5lnx_{2i} - u_i + v_i$.. $v_i \overset{iid}{\sim}$ N (0, 0.15). $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

Table 3.2. Average technical efficiency measures using F_2

n	Parameters	Proposed approach			Conventional approach		
		EXP	TN	HN	EXP	TN	HN
		Performance criteria					
50	TE	0.981	0.975	0.973	0.976	0.990	0.970
	MAD	0.328	0.334	0.328	0.235	0.131	0.308
	MSE	0.124	0.129	0.129	0.086	0.025	0.126
100	TE	0.978	0.969	0.967	0.972	0.988	0.964
	MAD	0.325	0.329	0.322	0.231	0.129	0.303
	MSE	0.121	0.125	0.124	0.083	0.024	0.122
200	TE	0.975	0.964	0.962	0.970	0.986	0.961
	MAD	0.323	0.324	0.318	0.229	0.128	0.299
	MSE	0.119	0.121	0.121	0.081	0.024	0.119

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribtuion. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. $F_2 : y_i = 0.5x_{1i} + 0.5x_{2i} - u_i$. $x_{1i} \overset{iid}{\sim}$ Uniform [5,15]. $x_{2i} \overset{iid}{\sim}$ Uniform [5,15].

3.5.2. Impact of production functions

DEA has emerged as an important nonparametric method for evaluating the performance of DMUs through benchmarking. A possible misspecification error in DEA model arises from assuming an inaccurate production function (Gong and Sickles, 1992 and Banker et al., 1993). We generated data using three production functions, IRS, DRS, and Translog. Table 3.3 provides the parameters associated with the three production functions summarized by Andor and Hesse (2013).

Table 3.3. Variation of the production functions

PF	Framework 1	Framework 2
IRS	$y_i = \ln(2) + 0.6 \ln x_{1i} + 0.6 \ln x_{2i} - u_i + v_i$	$y_i = 0.6x_{1i} + 0.6x_{2i} - u_i$
DRS	$y_i = \ln(2) + 0.4 \ln x_{1i} + 0.4 \ln x_{2i} - u_i + v_i$	$y_i = 0.4x_{1i} + 0.4x_{2i} - u_i$
TL	$y_i = \ln(2) + 0.5 \ln x_{1i} + 0.5 \ln x_{2i} + (0.5 \ln x_{1i})^2 + (0.5 \ln x_{2i})^2 + 0.1 \ln x_{1i} \ln x_{2i} - u_i + v_i$	$y_i = 0.5x_{1i} + 0.5x_{2i} + (0.5x_{1i})^2 + (0.5x_{2i})^2 + 0.1x_{1i}x_{2i} - u_i$

PF: Production functions. IRS: Increasing Return to Scale. DRS: Decreasing Return to Scale. x_{1i} and $x_{2i} \overset{iid}{\sim}$ U[5,15]. $v_i \overset{iid}{\sim}$ N (0, 0.15). u_i: Inefficiency distributions. TL: Translog

Concerning F_1, Tables 3.4 and 3.5 present respectively the results for the impact of the production function within the proposed and conventional approaches of the inefficiency distributions. In addition, from Tables 3.4 and 3.5, the performance of mean TE, MAD, and MSE are provided across the sample sizes of 50, 100, and 200. Two important results emerge from Tables 3.4 and 3.5.

First, within the proposed approach of Table 3.4, the mean TE decreases accordingly. However, the rate of decreased is proportional to the sample size and the type of prooduction function. In comparison to IRS and DRS, TL provides higher magnitude in terms of mean TE. This is consistent across the inefficiency distributions. Second, concerning MAD values, the results suggest that the DRS production function provides the best performance in comparison to IRS and TL production functions. More precisely, the exponential inefficiency distribution of DRS provides the smallest MAD values. Additionally, within production function, MAD and MSE decrease as as the sample size increases.

Within the conventional approach in Table 3.5, the DEA efficiency measures associated with the truncated normal inefficiency distribution provides smaller MAD values across DRS, IRS, and translog production functions. Additionally, in comparison to the proposed approach, the results of the conventional approach in Table 3.5 provide smaller MAD values within production function and inefficiency distribution. Furthermore, in Table 3.5, the results suggest that independent of the approaches, the production function does impact the performance of DEA efficiency measures.

Table 3.4. Impact of the production functions using F_1 of the proposed approach

n	Parameters	IRS			DRS			TL		
		EXP	TN	HN	EXP	TN	HN	EXP	TN	HN
		Performance criteria								
50	TE	0.917	0.908	0.906	0.898	0.889	0.889	0.938	0.931	0.929
	MAD	0.264	0.269	0.263	0.246	0.251	0.247	0.285	0.291	0.285
	MSE	0.085	0.088	0.088	0.078	0.081	0.081	0.096	0.100	0.099
100	TE	0.901	0.890	0.888	0.880	0.870	0.869	0.925	0.916	0.914
	MAD	0.248	0.251	0.246	0.228	0.233	0.229	0.272	0.277	0.271
	MSE	0.077	0.079	0.079	0.070	0.071	0.072	0.089	0.092	0.091
200	TE	0.887	0.875	0.874	0.876	0.865	0.863	0.914	0.904	0.902
	MAD	0.235	0.237	0.233	0.224	0.228	0.224	0.262	0.265	0.259
	MSE	0.071	0.072	0.072	0.067	0.068	0.069	0.083	0.085	0.085

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribtuion. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. IRS: Increasing Return to scale. DRS: Decreasing Return to scale. TL: Translog production function.

Table 3.5. Impact of the production functions using F_1 of the conventional approach

n	Parameters	IRS			DRS			TL		
		EXP	TN	HN	EXP	TN	HN	EXP	TN	HN
					Performance criteria					
50	TE	0.912	0.928	0.904	0.895	0.907	0.888	0.933	0.949	0.926
	MAD	0.182	0.088	0.246	0.173	0.086	0.234	0.196	0.099	0.265
	MSE	0.059	0.013	0.088	0.056	0.012	0.083	0.065	0.016	0.098
100	TE	0.896	0.913	0.887	0.876	0.889	0.869	0.920	0.937	0.911
	MAD	0.170	0.082	0.231	0.163	0.082	0.219	0.185	0.091	0.252
	MSE	0.053	0.012	0.080	0.051	0.011	0.076	0.060	0.014	0.090
200	TE	0.882	0.899	0.873	0.861	0.873	0.854	0.909	0.927	0.900
	MAD	0.162	0.078	0.220	0.157	0.082	0.209	0.177	0.085	0.241
	MSE	0.050	0.010	0.075	0.048	0.011	0.071	0.056	0.012	0.084

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribtuion. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE:Mean Square Error. IRS: Increasing Return to scale. DRS: Decreasing Return to scale. TL: Translog production function.

Next, we discuss the results of F_2 in Tables 3.6 and 3.7 respectively for the proposed and conventional approaches. Within the proposed approach of the IRS production function (Table 3.6), the DEA efficiency measures associated with the half normal inefficiency distribution provides a consistent decrease of MAD and MSE values as the sample size increases. Moreover, the overall performance of DRS is consistently better in comparison to IRS and translog production functions across the three inefficiency distributions. Within the conventional approach in Table 3.7, IRS provides inconsistent results. As the sample size increases for the exponential and truncated normal inefficiency distributions in the IRS production function, MAD values increase then decrease. This could be due to the sampling error or MC simulations errors.

Table 3.6. Impact of the production functions using F_2 of the proposed approach

n	Parameters	IRS			DRS			TL		
		EXP	TN	HN	EXP	TN	HN	EXP	TN	HN
		Performance criteria								
50	TE	0.984	0.979	0.977	0.976	0.968	0.966	0.972	0.972	0.972
	MAD	0.331	0.339	0.333	0.323	0.328	0.322	0.319	0.332	0.328
	MSE	0.126	0.133	0.132	0.119	0.124	0.123	0.122	0.131	0.133
100	TE	0.981	0.974	0.972	0.972	0.961	0.959	0.965	0.965	0.963
	MAD	0.327	0.333	0.327	0.320	0.321	0.314	0.312	0.325	0.321
	MSE	0.124	0.128	0.128	0.116	0.119	0.118	0.118	0.127	0.128
200	TE	0.979	0.970	0.969	0.970	0.955	0.953	0.959	0.960	0.960
	MAD	0.327	0.329	0.324	0.317	0.315	0.308	0.307	0.320	0.316
	MSE	0.124	0.126	0.125	0.114	0.114	0.113	0.115	0.123	0.124

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. IRS: Increasing Return to scale. DRS: Decreasing Return to scale. TL: Translog production function.

Table 3.7. Impact of the production functions using F_2 of the conventional approach

n	Parameters	IRS			DRS			TL		
		EXP	TN	HN	EXP	TN	HN	EXP	TN	HN
					Performance criteria					
50	TE	0.970	0.987	0.963	0.969	0.987	0.965	0.971	0.972	0.971
	MAD	0.230	0.129	0.301	0.232	0.137	0.303	0.233	0.118	0.310
	MSE	0.081	0.024	0.120	0.088	0.027	0.112	0.091	0.023	0.134
100	TE	0.977	0.990	0.970	0.966	0.978	0.956	0.964	0.966	0.964
	MAD	0.236	0.131	0.308	0.225	0.126	0.294	0.228	0.113	0.304
	MSE	0.087	0.025	0.127	0.078	0.023	0.115	0.088	0.021	0.130
200	TE	0.975	0.989	0.967	0.963	0.961	0.959	0.959	0.961	0.958
	MAD	0.234	0.130	0.305	0.222	0.321	0.314	0.223	0.110	0.299
	MSE	0.085	0.025	0.124	0.076	0.119	0.118	0.085	0.020	0.127

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. IRS: Increasing Return to scale. DRS: Decreasing Return scale. TL: Translog production function.

Overall, examining the results of F_1 (Tables 3.4 and 3.5) and F_2 (Tables 3.6 and 3.7), it is evident that there exists a considerable variation in the bias of the estimates between the two DGPs. The mean bias of MAD values (Tables 3.4 and 3.5) shows that in the presence of random noise distribution, MAD values decrease in comparison to the DGP in level (Table 3.5). In fact, the level of the mean bias in Tables 3.4 and 3.5 is smaller than the mean bias in Tables 3.6 and 3.7. However, it is important to note that in terms of mean TE, Tables 3.6 and 3.7 provide better results.

Conclusively, from Tables 3.4 to 3.7, the results suggest the production functions influence the performance of TE and MAD contrary to the results of Andor and Hesse (2011). Moreover, the percentage magnitude of the influence is dependent on DGP and the approach used. These results caution the researcher to pay attention to the inefficiency distribution of the output variables. Next, we discuss an additional factor that impact the performance of DEA efficiency measures.

3.5.3. Impact of input distributions

A third misspecification that can affect the performance of DEA TE measures is the input distributions (Andor and Hesse, 2011 and 2013). Hence, for this analysis, the normal-normal and gamma-weibull distributions are applied. These results are presented in Tables 3.8 and 3.9 respectively across F_1 and F_2. The results of Tables 3.8 and 3.9 are partitioned within the proposed and conventional approaches of the inefficiency distributions. Additionally, the performance creteria of mean TE, MAD, and MSE are presented.

Concerning the proposed approach of the inefficiency distributions, two important results emerge from Tables 3.8 and 3.9. First, all methods perform consistently better. As the sample size increases, MAD values decrease across the DEA efficiency measures associated with the exponential, truncated normal, and half normal inefficiency distributions of normal-normal and gamma-weibull input distributions. Second, the normal-normal input distributions provide the lowest MAD values consistently across the three inefficiency distributions (Tables 3.8 and 3.9). Additionally, across each sample size and input distributions, $MAD_{EXP} < MAD_{HN} < MAD_{TN}$ (Tables 3.8 and 3.9).

Furthermore, comparing the proposed approach of Table 3.8 to Table 3.9, it is worth noting that the mean TE, MAD, and MSE values in Table 3.9 have quite increase. Hence, DEA confounded the random noise distribution with the inefficiency distribution (Table 3.8). The results in the presence of the random noise distribution are inconsistent in terms of mean TE. Moreover, the MAD values in Table 3.8 are smaller in comparison to Table 3.9. In fact, the effect of the log transformation of the input distributions and the presence of the random noise distribution perform better in terms of MAD values in comparison to the DEA's model without log transformation and the random noise distribution. Overall, the input distributions impact the proposed approach of the inefficiency distributions.

We now begin the analysis of the conventional approach of the inefficiency distributions by examining the mean TE, MAD and MSE values across F_1 (Table 3.8) and F_2 (Table 3.9). The results in Table 3.9 indicate a small drop in MAD values across the inefficiency distributions of the gamma-weibull input distributions as the sample size increases. With the normal-normal's input distribution of Table 3.9, it is worth noting as the sample size increases, the MAD values of the DEA TE associated with the exponential inefficiency distribution increases. The results in

Tables 3.8 and 3.9 suggest that the input distributions matter. Moreover, the input distributions of normal-normal and gamma-weibull perform better in terms of MAD in comparison to the uniform distribution of Tables 3.1 and 3.2 across identical inefficiency distributions. Overall, the analysis of the input distribution supports the supposition of Resti (2000) that the input distributions can have an influence on the performance of TE measures (Tables 3.8 and 3.9).

Table 3.8. Impact of the input distributions using F_1

n	Parameters	Normal and Normal			Gamma and Weibull		
		EXP	TN	HN	EXP	TN	HN
		Proposed approach					
50	TE	0.923	0.917	0.917	0.976	0.975	0.975
	MAD	0.270	0.278	0.274	0.323	0.335	0.331
	MSE	0.091	0.094	0.096	0.125	0.132	0.134
100	TE	0.905	0.898	0.898	0.970	0.972	0.970
	MAD	0.252	0.260	0.256	0.318	0.329	0.325
	MSE	0.081	0.084	0.085	0.122	0.129	0.130
200	TE	0.887	0.880	0.879	0.965	0.964	0.965
	MAD	0.234	0.243	0.239	0.312	0.324	0.320
	MSE	0.072	0.074	0.076	0.118	0.126	0.127
		Conventional approach					
50	TE	0.921	0.927	0.918	0.981	0.976	0.975
	MAD	0.191	0.089	0.259	0.242	0.119	0.314
	MSE	0.066	0.014	0.098	0.097	0.023	0.137
100	TE	0.902	0.910	0.898	0.976	0.970	0.970
	MAD	0.178	0.081	0.242	0.236	0.115	0.309
	MSE	0.059	0.011	0.088	0.093	0.022	0.134
200	TE	0.884	0.893	0.879	0.970	0.965	0.965
	MAD	0.167	0.076	0.227	0.232	0.112	0.304
	MSE	0.053	0.010	0.079	0.091	0.021	0.130

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. $F_1 : y_i = ln(2) + 0.5lnx_{1i} + 0.5lnx_{2i} - u_i + v_i$. $v_i \overset{iid}{\sim}$ N (0, 0.15). $x_{1i} \overset{iid}{\sim}$ N (8.4, 1.51) and $x_2 \overset{iid}{\sim}$ N (7.92, 1.18). $x_1 \overset{iid}{\sim}$ Weibull (17.4, 20.5) and $x_2 \overset{iid}{\sim}$ Gamma (285, 0.07).

Table 3.9. Impact of the input distributions using F_2

n	Parameters	Normal and Normal			Gamma and Weibull		
		EXP	TN	HN	EXP	TN	HN
		Proposed approach					
50	TE	0.975	0.967	0.965	0.990	0.986	0.985
	MAD	0.322	0.327	0.320	0.337	0.346	0.341
	MSE	0.119	0.123	0.122	0.132	0.139	0.140
100	TE	0.972	0.960	0.958	0.988	0.983	0.983
	MAD	0.319	0.320	0.313	0.336	0.343	0.338
	MSE	0.116	0.118	0.117	0.131	0.137	0.137
200	TE	0.804	0.794	0.792	0.987	0.981	0.980
	MAD	0.162	0.169	0.169	0.335	0.341	0.336
	MSE	0.042	0.042	0.043	0.130	0.135	0.136
		Conventional approach					
50	TE	0.976	0.987	0.962	0.990	0.995	0.984
	MAD	0.235	0.128	0.299	0.249	0.136	0.323
	MSE	0.085	0.024	0.119	0.097	0.027	0.140
100	TE	0.970	0.984	0.955	0.987	0.994	0.982
	MAD	0.230	0.126	0.293	0.247	0.135	0.319
	MSE	0.081	0.023	0.114	0.095	0.027	0.137
200	TE	0.965	0.983	0.951	0.986	0.993	0.980
	MAD	0.244	0.124	0.289	0.245	0.135	0.319
	MSE	0.077	0.023	0.110	0.094	0.027	0.137

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average technical efficiency. MAD: Mean Absolute Deviation. MSE: Mean Square Error. $F_2 : y_i = 0.5x_{1i} + 0.5x_{2i} - u_i$. $x_{1i} \overset{iid}{\sim}$ N (8.4, 1.51) and $x_2 \overset{iid}{\sim}$ N (7.92, 1.18). $x_1 \overset{iid}{\sim}$ Weibull (17.4, 20.5) and x_2 $\overset{iid}{\sim}$ Gamma (285, 0.07).

3.5.4. Impact of curse of dimensionality

Finally, a third factor that impacts the performance of DEA TE measures is the curse of dimensionality. Statistical variable selection aims to keep the significant inputs and outputs. By removing the unimportant variables, the DEA model can estimate the production frontiers more precisely and avoid the problem of curse of dimensionality (Golany and Roll, 1989 and Dyson et al., 2001). Studies have shown that the number of observations required to obtain meaningful estimates of inefficiency increases dramatically with the number of production inputs and outputs (Wheelock and Wilson, 2007).

With an increasing number of inputs, the estimation of the production function becomes more challenging. Our focus was to vary the number of inputs of the constant return Cobb-Douglas production function and keep the scale elasticity constant (elasticity=1). The additional variables are not correlated with the included inputs. Hence, the additional variables will result in poor measurement error. It is expected that the correct specification of the number of input variables will outperform all the additional input variables used in the production function. Tables 3.10 and 3.11 present the results of mean TE and MAD values across, three, four, and five independent inputs within sample sizes of 50, 100, and 200. The results are partitioned into the proposed and conventional approaches across F_1 and F_2.

The results reported are in line with the issue of curse of dimensionality in DEA model found in Smith (1997). As the number of inputs increases, the increased in mean TE is proportional to the sample size. The results in Tables 3.10 and 3.11 show that the performance of DEA across both frameworks of the inefficiency distributions are influenced by the variations in the number of inputs. The results further show that the MAD values decrease linearly while the number of observations increases. Therefore, the minimal requirements scholars promoted are not enough and the data set should be increased linearly while the dimension become higher.

Table 3.10. Curse of dimensionality for the proposed approach

n	Parameters	Framework 1			Framework 2		
		EXP	TN	HN	EXP	TN	HN
		Three inputs					
50	TE	0.939	0.934	0.934	0.985	0.979	0.977
	MAD	0.286	0.295	0.290	0.333	0.339	0.332
100	TE	0.923	0.917	0.916	0.972	0.969	0.968
	MAD	0.270	0.277	0.272	0.319	0.329	0.323
200	TE	0.907	0.900	0.899	0.978	0.961	0.959
	MAD	0.254	0.262	0.258	0.325	0.321	0.314
		Four inputs					
50	TE	0.976	0.975	0.975	0.992	0.991	0.990
	MAD	0.323	0.334	0.330	0.340	0.351	0.346
100	TE	0.967	0.966	0.966	0.989	0.986	0.985
	MAD	0.315	0.326	0.322	0.336	0.345	0.340
200	TE	0.958	0.957	0.957	0.986	0.981	0.980
	MAD	0.306	0.317	0.313	0.333	0.341	0.335
		Five inputs					
50	TE	0.988	0.987	0.987	0.996	0.995	0.995
	MAD	0.335	0.347	0.343	0.342	0.356	0.350
100	TE	0.983	0.983	0.983	0.993	0.992	0.991
	MAD	0.331	0.343	0.338	0.340	0.352	0.346
200	TE	0.978	0.978	0.978	0.990	0.988	0.987
	MAD	0.326	0.338	0.333	0.338	0.348	0.342

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average Technical efficiency measures. MAD: Mean Absolute Deviation.

Table 3.11. Curse of dimensionality for the conventional approach

n	Parameters	Framework 1			Framework 2		
		EXP	TN	HN	EXP	TN	HN
		Three inputs					
50	TE	0.937	0.944	0.934	0.982	0.993	0.978
	MAD	0.203	0.100	0.274	0.241	0.134	0.315
100	TE	0.921	0.928	0.916	0.977	0.990	0.971
	MAD	0.191	0.091	0.258	0.236	0.132	0.309
200	TE	0.905	0.913	0.899	0.973	0.988	0.965
	MAD	0.180	0.085	0.245	0.232	0.129	0.303
		Four inputs					
50	TE	0.976	0.977	0.975	0.990	0.996	0.988
	MAD	0.236	0.121	0.313	0.249	0.138	0.326
100	TE	0.967	0.969	0.966	0.986	0.994	0.983
	MAD	0.229	0.114	0.305	0.245	0.136	0.320
200	TE	0.958	0.960	0.957	0.982	0.993	0.977
	MAD	0.221	0.108	0.296	0.241	0.134	0.316
		Five inputs					
50	TE	0.988	0.988	0.988	0.994	0.998	0.993
	MAD	0.247	0.131	0.326	0.254	0.140	0.331
100	TE	0.983	0.984	0.983	0.991	0.997	0.989
	MAD	0.224	0.127	0.321	0.250	0.138	0.327
200	TE	0.978	0.979	0.978	0.988	0.995	0.985
	MAD	0.239	0.122	0.316	0.247	0.137	0.323

n: Sample size. EXP: Exponential distribution. TN: Truncated normal distribution. HN: Half normal distribution. TE: Average Technical efficiency measures. MAD: Mean Absolute Deviation.

3.6. Conclusion and future research

In this research, MC simulations were conducted to examine the performance of the DEA model under two different data generating processes, logarithm and level, and across five different scenarios, inefficiency distributions, sample size, production functions, input distributions, and curse of dimensionality. In the consistent evaluation of the DEA model with inefficiency distributions, two approaches, proposed and conventional, were applied. In the proposed approach, the sample mean and sample standard deviation of the inefficiency distributions were on a similar scale and promoting an “apples-to-apples" comparison. Therefore, it is statistically appropriate in the estimation of DEA model. In contrast, for the traditional approach, the sample statistics were not consistent. Hence, it is merely impossible to compare the performance of DEA model associated with half normal, truncated normal, and exponential inefficiency distributions.

Conclusively, this research cautions the DEA practitioners concerning the accuracy of their estimates. Within the proposed and traditional approaches of the logarithm simulation, our results follow the literature and suggest that the DEA TE measures are inconsistent in the presence of the random noise distribution. Furthermore, in the proposed approach, the accuracy of the DEA model to minimize bias in its estimation was sufficiently done and the comparison across the inefficiency distributions could be carried out. Additionally, the results of our simulations suggest that the input distributions of normal-normal and gamma-weibull perform better in comparison to the uniform distribution used in the literature of DEA efficiency measures. In addition, when the DEA model includes more variables than necessary, DEA overestimates the efficiency measures. Hence, we believe that the results of our simulations conducted would be useful to the DEA model builders when choosing the input variables.

Past researchers have calculated technical efficiency based on the DEA method, but they have not really considered the values of the parameters or the inclusions of various inefficiency distributions. The assessment of how this issue is applied in empirical frontier analysis is beyond the scope of this research and should be tested through real−world problems. We think that more research is still needed, the sources of endogeneity affecting the estimation of production frontiers, and the sensitivity of the DEA model to incorporate multiple outputs.

4. IMPACT OF LIQUIDITY AND SOLVENCY FINANCIAL FACTORS ON BANKS' COST EFFICIENCY MEASURES

4.1. Abstract

This chapter evaluates the impact of liquidity and solvency financial factors while accounting for additional set of explanatory variables, i.e. regulatory, macroeconomic, and bank characteristics on the cost efficiency measures using the Stochastic Frontier Analysis and Data Envelopment Analysis. The results show that the liquidity and solvency financial factors negatively impacted the cost efficiency measures of U.S banks from 2005 to 2017. Moreover, during the financial crisis, U.S banks were inefficient in comparison to the tranquil period, and the solvency financial factor impacted inignificantly the cost efficiency measures. However, during the financial crisis, U.S banks' liquidity financial factor collapsed due to contagion.

4.2. Introduction

The efficiency of banks is an important element of analysis in the world and the importance of this efficiency on the 2007-2009 financial crisis has been addressed (Pasiouras et al., 2009; Chortareas, et al., 2012; Pessarossi and Weill, 2014; Barth et al., 2013; Aspachs et al., 2005; Distinguin et al., 2013; and Bonner et al., 2015). Existing research has shown that the financial crisis provides disruptive effects on the real economy (Hoggarth and Saporta, 2001 and Aspachs et al., 2005) while taking its toll on the banking efficiency measures (Shaik, 2015).[1] In recent analysis of the banking sector, opinions have differed on the two main factors that led to the 2007-2009 financial crisis, liquidity (Gorton and Metrick, 2010; Lucas and Stokey, 2011; and Cochrane, 2013) and solvency (Johnson and Kwak, 2011; and Volcker[2]).

In this research, the primary objective is to evaluate and assess the role or importance of liquidity and solvency financial factors on the cost efficiency measures of the United States' commercial and domestic banks. A liquidity crisis occurs when banks have their liabilities greater than their assets and are unable to provide cash in the short run. Studies have emphasized that the

[1] Starting in the late 2007 and early 2008, when financial crisis hit the economy, 30 banks were closed. This number jumped to 148 bank failures in 2009 and 150 bank failures in 2010. However, in 2011, the number of failed banks dropped to 92 and there were 41 failures through the third quarter of 2012.

[2] Paul Volker is the former Chair of the Board of Governors of the Federal Reserve System.

creation of liquidity financing illiquid assets makes banks susceptible to liquidity risks (Kashyap et al., 2002; Berger and Bouwman, 2009; and Ratnovski, 2013). On a contrary, a solvency crisis arises when the value of banks assets falls short of the value of their liabilities (Demirguc-Kunt et al., 2008) or when increased interest rates reduce the demand for liquidity (Diamond and Rajan, 2005). The debate of identifying the impact of both liquidity and solvency financial factors is an important debate because the efficiency of the banking sector influences the stability of the financial system.

In the evaluation of liquidity and solvency financial factors, this chapter additionally accounts for exogenous variables, regulatory, macroeconomic, and bank characteristics. This is particularly important from a practical-stand point of survivability of banks and ability to identify the primary factors that caused the financial crisis. The measurement and identification of the exogenous variables that impacted the cost efficiency is of paramount importance to the banking sector. This is exemplified by the identification of the best performance among the competing banks, the influence of the policy intervention, the strength of the banking sector, and the prevention of bank failures.

Recent studies have suggested that banks operate in heavily regulated sectors (Pasiouras et al., 2009 and Chortareas, et al., 2012) with conflicting views about which regulations are the most appropriate to use (Barth et al., 2004; Demirguc-Kunt et al., 2008; Pasiouras et al., 2009; and Barth et al., 2013). Because policymakers are concerned with preventing future bank failures; therefore, our research accounts for the two popular regulatory factors, Basel Accord III and Dodd Frank Act. These regulatory factors are considered while thinking about the impact of the structural shift of the regulatory framework. In addition, this chapter considers internal and external factors as exogenous variables. Internal factors are bank characteristics, bank type, bank size and asset classification (Aspachs et al., 2005; Distinguin et al., 2013; and Bonner et al., 2015). The external factor, annual state gross domestic product (GDP), is used as a proxy for the state economic development (Distinguin et al., 2013).

To investigate the impact of liquidity and solvency financial factors on the cost efficiency measures, two approaches have caught the attention of researchers, Data Envelopment Analysis (DEA) (Charnes et al., 1978) and Stochastic Frontier Analysis (SFA) (Aigner et al., 1977). This research fundamentally contributes to the existing literature in three different aspects: two-step approach, comparative analysis of DEA and SFA, and properties of estimations.

First, we employ the two-step approach for panel data. In the first step, we estimate the dual cost function of SFA and DEA economic efficiency measures. Then in a second step, for consistency estimation, we use a Tobit regression model (Tobin, 1958) and examine the impact of exogenous variables on the banks' cost efficiency measures by regressing the economic efficiency measures on a series of proxies for liquidity and solvency financial factors, regulatory factors, macroeconomic variables, and bank characteristics. Second, we look at the comparative performance of both DEA and SFA to identify whether the cost efficiency measures can influence the 2007-2009 financial crisis. Third, using the properties of the estimation techniques, we infer which method is likely to be outperforming the other and identify the main sources of inefficiency using the exogenous variables. Moreover, in doing so we are mainly interested in the sign and significance level of the exogenous variables.

The remainder of this chapter is structured as follows. Section 4.3 presents the theoretical SFA and DEA models. Section 4.4 presents the empirical data and the construction of the dependent and independent variables. Section 4.5 presents the empirical results of SFA and DEA estimators. Section 4.6 summarizes the research and provides additional discussion.

4.3. Theoretical framework

To concisely address the theoretical framework of this research, let's assume that there are K input prices and J output quantities available on each of the N banks. For the i^{th} bank, we have vectors of input price, w_i, and output quantity, y_i, respectively representing the K input prices and J output quantities. Therefore, we have $K \times N$ input prices matrix, w, and $J \times N$ output quantities matrix, $\mathbf{y}$, representing the U.S banks.

Dual cost theory assuming the relationship between J producing output quantities, $y = (y_1, x_2, ..., y_j) \in \mathbb{R}_+^J$ and K input prices, $w = (w_1, w_2, ..., w_k) \in \mathbb{R}_+^K$ is reflected by the concept of cost function[3]. In general, the cost function of a bank i at time t can be written as follow:

$$TC_{it} = F(y_{it}, w_{it}; \beta) \quad i = 1, 2, .., N; \quad t = 1, 2, .., T, \tag{4.1}$$

where TC_{it} is the total cost of bank i at time t. y_{it} is the vector of output quantities of a bank i at time t. w_{it} is the vector of input prices of a bank i at time t. F is the associated functional form of

[3]In the theoretical framework of the cost function, it is important that it satisfies a number of properties including: non-decreasing in outputs and input prices, homogeneous of degree one in input prices and concave in input prices.

the cost function. β is the vector of unknown parameters that will be estimated. The cost function framework in equation (4.1) forms the bases in the estimation of the banks' cost efficiency measures using the statistical SFA and the mathematical programming DEA models under variable return to scale.

4.3.1. Efficiency measures

Equation (4.1) estimated using the statistical SFA model that decomposes the traditional error, ε_{it}, into a symmetrical random error, v_{it}, and a one-sided error or inefficiency, u_{it}, is represented by:

$$\ln TC_{it} = F(y_{it}, w_{it}; \beta) + v_{it} + u_{it} \quad i = 1, 2, .., N; \quad t = 1, 2, .., T, \tag{4.2}$$

where $\ln TC_{it}$ is the logarithm of the observed total cost of bank i at time t. $F(y_{it}, w_{it}; \beta)$ is the cost function frontier common to bank. u_{it} is the one-sided inefficiency represented with alternative distributions including a half normal distribution (Meeusen and Van den Broeck, 1977), an exponential distribution (Aigner et al., 1977), a truncated normal distribution (Stevenson, 1980), and a gamma (Green, 1990). In this chapter, the half normal SFA model is used.

The cost function in equation (4.1) can also be estimated using the linear programming DEA model. As any nonparametric method, DEA is a mathematical programming that provides the construction of efficiency limit obtained using the available banks. Each bank will only maximize its efficiency score under the constraint that the sum of banks efficiency score is not allowed to exceed one. Following Färe et al., (1985), the cost minimization DEA model is defined as:

$$\begin{array}{ll} \min\limits_{\lambda x_i} & w_i x_i^* \\ \text{subject to} & y_i + \mathbf{y}\lambda \geq 0 \\ & \lambda \mathbf{x} - x_i^* \leq 0, \\ & \lambda \geq 0 \end{array} \tag{4.3}$$

where x_i^* is the cost minimizing vector of input quantities for the i^{th} bank given input price, w_i and output quantities, y_i. λ is $N \times 1$ vector of constants. $\mathbf{x}$ is a matrix of input quantities associated with the input prices. This constrained optimal minimization is obtained from a linear combination of banks that produces at least as much of each of the outputs using the same or less amount of inputs as calculated for each bank in the data.

4.3.2. Determinants of efficiency

In the evaluation of the exogenous variables using SFA and DEA estimators, this chapter employs a two-step approach.[4] In the first step, one estimates the cost efficiency measures of U.S banks. In the second step, one regresses the cost efficiency measures on the exogenous variables using the discrete choice Tobit regression model.

After the estimation of cost efficiency measures in equations (4.2) and (4.3), the second step analysis consists in the specification of the Tobit regression model to analyze the cost inefficiency measures of banks. The Tobit regression is an appropriate tool to be used, because the cost inefficiency measures obtained from equations (4.2) and (4.3) are censored, and cannot exceed 1 nor be below 0. The Tobit regression of the cost inefficiency measures can be expressed as:

$$\begin{aligned} u_{it}^{*} &= \delta\theta_{it} + \varepsilon_{it} \\ \text{where} \quad u_{it} &= u_{it}^{*} \quad if \quad u_{it}^{*} > 0 \\ u_{it} &= 0 \quad if \quad u_{it}^{*} \leq 0 \end{aligned} \tag{4.4}$$

where u_{it} is the dependent variable-that is the cost inefficiency measures obtained from SFA and DEA which is defined by a latent variable u_{it}^{*} for positive values of the inefficiency measures and censored otherwise. δ is a vector of the estimated parameters. θ_{it} is a vector of explanatory variables associated with the economic inefficiency of bank i over time t. ε_{it} is the random variable that capture the effect of the unobserved factors of bank i over time t and distributed with zero mean and constant variance, σ^2.

4.4. Empirical data and construction of the input prices and outputs quantities

This research employs annual data of the Federal Financial Institutions Examination Council's based on the Council Form 041 Report of Condition and Income of U.S. commercial and domestic banks that report to the Federal Reserve banks. In the literature of the cost efficiency measures, three approaches are pertinent to the selection of input and output variables, production, profitability, and intermediate respectively based on the theory of production, profit, and cost functions.

[4] Studies have concluded that the two-step approach leads to omitted variables bias (Pessarossi and Weill, 2014 and Shaik 2015) and heteroskedasticity (Greene, 2004) and it is only applicable to SFA estimator.

4.4.1. Input prices and output variables selection

The production approach of Sherman and Gold (1985) is the most popular approach for bank branch performance analysis and has been used to identify any operational inefficiencies (Parkan, 1987). Profitability approach of Drake et al., (2006) is quite similar to the production approach, but the outputs of profitability approach are more profit-oriented (Thagunna and Poudel, 2013). In the intermediation approach, banks serve as intermediate between depositors and borrowers and regard loans and other assets as bank outputs while deposits and other liability funds as bank inputs (Sealey and Lindley, 1977).[5]

Concerning the selection of the input and output variables, the intermediation approach of Berger and Humphrey (1997) is used. Following Pessarossi and Weill (2014), two outputs and three inputs prices are identified. The outputs selected are, total loans (y_1), and other earning assets, (y_2). The three input prices selected are, price of labor, (w_1), calculated as the ratio of personnel expenses to total assets; price of physical capital, (w_2), calculated as the ratio of other operating expenses to premises and fixed assets, and price of borrowed funds, (w_3), computed as the ratio of the interest expenses to total deposits. Finally, to construct the cost function, total cost, (TC) is calculated as the sum of interest expenses, personnel expenses, and other operating expenses. We additionally impose homogeneity conditions by normalizing respectively TC, w_1, and w_2 by w_3 (see Table 4.2 for further definitions). For the estimation of SFA model in equation (4.2), a Cobb-Douglas functional form of the stochastic frontier model of bank i at time t with two outputs and three input prices is specified as:

$$\ln(\frac{TC_{it}}{w_{3it}}) = \alpha_0 + \beta_1 \ln(\frac{w_{1it}}{w_{3it}}) + \beta_2 \ln(\frac{w_{1it}}{w_{3it}}) + \gamma_1 \ln y_{1it} + \gamma_2 \ln y_{2it} + \eta t + v_{it} + u_{it} \qquad (4.5)$$

where α_0 is the intercept. β_1 and β_2 are the parameters estimate for the input prices of labor and physical capital. γ_1 and γ_2 are the parameters estimate associated with respectively the output quantities of total loans and other earning assets. η is the parameter associated with the time trend t. In addition, since in the cost function, total cost, input prices, and output quantities are in logarithms, then the dual cost function coefficients represent the elasticities.

[5] Overall, the production approach may be more appropriate for evaluating branch-level efficiency, while the intermediation approach is better for measuring the efficiency of banks as a whole (Berger and Humphrey, 1997).

Table 4.1 presents the summary statistics of the total cost, input prices, and output quantities in thousand based on 48 states from 2001 to 2017 with a total of 119,146 observations. The results in Table 4.1 are subdivided in three groups, the original data, the post-hoc cleaning data, and the post-hoc unique identifier. First, in the original data ranging from 2001 to 2017, the number of missing observations ranged from 8,547 to 96,716. Second, for the post-hoc data cleaning ranging from 2005 to 2017, we excluded 1) banks from 2001 to 2004 due to the volume of missing values and 2) rows that had five at least percent missing. As a result, we had 13,950 complete observations. Finally, in the third stage of the post-hoc data cleaning, and to provide a systematic and consistent result, we excluded banks with at least four unique identifiers. Overall, the final data used for the analysis is an unbalanced panel with 11,044 observations ranging from 2005 to 2017.[6]

[6] Berrospide (2013) evaluated Bank liquidity hoarding and the financial crisis using 106,817 bank-quarter observations for approximately 6,750 institutions.

Table 4.1. Descriptive statistics of the variables

Variable	N	Missing	Mean	Standard deviation	Maximum
	Original data from 2001-2017 (in thousand)				
Total cost (TC)	119, 146	8547	12349.82	60374.87	4960846.00
Price of borrowed funds (w_1)	119, 146	18119	0.174	12.730	1973.33
Price of labor (w_2)	119, 146	17,243	0.021	0.061	6.198
Price of physical capital (w_3)	119, 146	96,716	0.625	15.190	1113.53
Total loans (y_1)	119, 146	0	6514.70	121797.76	16642207.00
Other earning assets (y_2)	119, 146	0	16822.78	161773.15	22301000.00
	post-hoc data cleaning from 2005-2017 (in thousand)				
Total cost (TC)	13,950	0	42069.09	133709.61	4960846
Price of borrowed funds (w_1)	13,950	0	0.018	0.1828	21.554
Price of labor (w_2)	13,950	0	0.016	0.01	0.424
Price of physical capital (w_3)	13,950	0	0.415	13.87	1113.53
Total loans (y_1)	13,950	0	20524.27	320484.8	16642207
Other earning assets (y_2)	13,950	0	62088.92	308173.03	22301000
	post-hoc data cleaning accounting for unique identifier (in thousand)				
Total cost (TC)	11,044	0	43367.57	128165.46	4960846
Price of borrowed funds (w_1)	11,044	0	0.018	0.012	0.142
Price of labor (w_2)	11,044	0	0.016	0.009	0.282
Price of physical capital (w_3)	11,044	0	0.452	15.398	1113.53
Total loans (y_1)	11,044	0	21716.89	324202.59	16642207
Other earning assets (y_2)	11,044	0	62086.71	219869.21	7803155

After the estimation of SFA and DEA cost efficiency measures respectively from equations (4.5) and (4.3), a Tobit regression model (equation 4.4) was used to investigate the influence of a broader set of determinants. The potential exogenous factors affecting the cost efficiency measures include liquidity (Liq) and solvency (Sol) financial factors, financial crisis (crisis), bank size (size), state GDP (GDP), Basel Accord III (Basel3) , Dodd Frank Act (Dodd), bank headquarter (BMHO), and agricultural asset specialization (Asset) (see Table 4.2). Equation (4.4) can be rewritten into equation (4.6) to respectively account for the explanatory variables. Equation (4.6) is expressed for a bank i at time t as:

$$\begin{aligned} u_{it}^* = \delta_0 + \delta_1 \ln(Liq_{it}) + \delta_2 \ln(Sol_{it}) + \delta_3 crisis + \delta_4 \ln(size_{it}) + \delta_5 \ln(GDP_{it}) \\ + \delta_6 Basel3 + \delta_7 Dodd + \delta_8 BMHO + \delta_9 Asset + \epsilon_{it} \end{aligned} \tag{4.6}$$

where δ_0 is the intercept of the Tobit regression model. δ_1 and δ_2 are the estimated parameters of liquidity (Liq_{it}) and solvency (Sol_{it}) financial factors. δ_3 is the estimated parameter of the financial crisis indicator (crisis). δ_4 and δ_5 are the estimated parameters of size and GDP. δ_6 and δ_7 are the estimated parameters of the indicators, Basel3 and Dodd. δ_8 and δ_9 are the estimated parameters of the indicators variables, BMHO and Asset.

Table 4.2. Variable descriptions

Description of the cost variables and financial factors			
Variables	Formula	Definitions	FFEIC database (Corresponding variables in parenthesis)
Price of labor (w_1)	$\frac{Personnel\ expenses}{Total\ assets}$	Price of labor is the price associated with the sum of all wages paid to employees, as well as the price of employee benefits.	Personnel expenses (RIAD4135) include salaries and employee benefits. Total asset (RCON2170) is the sum of total loans and leases, total securities (HTM), total securities (AFS), trading assets, total intangible assets, other real estate owned, all other assets minus Allowance for loan and lease losses
Price of physical capital (w_2)	$\frac{Other\ operating\ expenses}{Fixed\ assets}$	Price of physical capital is the price of maintaining building.	Other operating expenses (RIADC216 +RIADC232) is the sum of Goodwill impairment losses, amortization expenses and impairment losses for other intangible assets. Fixed assets (RCON2145) are assets which are purchased for long-term use and unlikely to be quickly converted into cash.
Price of borrowed funds (w_3)	$\frac{Interest\ expenses}{Total\ deposits}$	Price of borrowed funds is the price of associated with borrowing money.	Total interest expense (RIAD4073) is the sum of the interest expense. Total deposit (RCON2200) is the sum of all domestic deposits including demand, saving and fixed deposits minus noninterest bearing and interest bearing.
Total loans (y_1)		Sum of all type of loans	Sum of loans including: loans secured by real estate (RCON6560); loans to finance agricultural production and other loans to farmers (RCON1583); loans to finance commercial real estate, construction and land development activities (RCON6560); loans to individuals for household, family, and other personal expenditures (RIADB486); loans to individuals for households, family, and other personal expenditures: credit cards (RCONB577); Other construction Loans (RCONF177).
Other earning assets (y_2)			Other earning assets (RCON2160) consists of balances due from the bank, inter-bank loans, investments, and securities.
Total cost (TC)		Sum of interest expenses, personnel expenses, and other operating expenses	Interest expenses (RIAD4073), personnel expenses (RIAD4135), and other operating expenses (RIADC216+RIADC232).
Liquidity factor (Liq)	$\frac{Liquid\ assets}{Total\ deposits}$	Ability to quickly rise cash	Liquid asset is sum of saving deposits (RIAD0093), federal funds sold and securities purchased under agreements to resell in domestic offices of the bank and its edge and agreement subsidiaries and in IBFS (RCON1350) and the total trading assets (RCON3545).
Solvency risk (Sol)	$\frac{Total\ equity}{Total\ assets}$	Capacity to face difficulties during the downturn	Total equity (RCON3210) is the total holding company or bank equity capital, including paid-up capital, share premiums and reserves.
Financial crisis (crisis)	dummy	1 if 2007 $\leq$ year $\leq$ 2009 and 0 otherwise.	

Table 4.3. Exogenous factors

Description of the exogenous variables		
Variables	Formula	Definitions
Banking characteristics		
Bank size ($Size$)	Log of total assets (RCON2170).	Measured by the natural logarithm of total amount of assets owned by the bank.
Bank headquarter (BMKO)	1 if headquarters and 0 if branch	`https://www5.fdic.gov/sod/sodInstBranchRpt.asp?rCert=10057&rYear=2018&barItem=1`
primary asset specialization ($Asset$)	1 if the asset specialization is agricultural and 0 otherwise.	`https://www5.fdic.gov/sod/sodInstBranchRpt.asp?rCert=10057&rYear=2018&barItem=1`
Regulation factors		
Basel Accord III (basel3)	1 if during the Basel III accord and 0 otherwise	`https://www.federalreserve.gov/supervisionreg/basel/USImplementation.htm#Basel_III_Tools`. Basel III is the last of the banking regulation agreements proposed in 2010 and implemented from 1st of January of 2013 till 1st of January of 2019. Basel III is a dummy characterized by 1 since implementation =2013 and 0 otherwise.
Dodd Frank Act (Dodd)	1 if during the Dodd Frank Act accord and 0 otherwise	
Macroeconomic factors		
State GDP ($logGDP$)	Log of GDP	GDP is the state gross domestic products. `https://www.bea.gov/`
Interaction terms		
$Liq \times BKMO$		Interaction term between liquidity risk factor and headquarter of the banks.
$Sol \times BKMO$		Interaction term between solvency financial factor and headquarter of the banks.
$Liq \times agricultural$		Interaction term between liquidity financial factor and agricultural primary asset classification.
$Sol \times agricultural$		Interaction term between solvency financial factor and agricultural primary asset classification.
$Liq \times basel3$		Interaction term between liquidity financial factor and Basel Accord III.
$Liq \times dodd$		Interaction term between liquidity financial factor and Dodd Frank Act.
$Sol \times basel3$		Interaction term between solvency financial factor and Basel Accord III.
$Sol \times dodd$		Interaction term between solvency financial factor and Dodd Frank Act.

4.5. Empirical results

To evaluate the impact of liquidity and solvency financial factors on the cost efficiency measures while accounting for regulatory, macroeconomic, and bank characteristics, this section presents the results of SFA and DEA cost efficiency measures models. All estimations were performed by maximum likelihood function incorporated into statistical software analysis SAS and R. Alternative specifications and assumptions about the distribution of the one-sided error term of SFA model were also tried. Specifically, a half normal distribution of the inefficiency distribution was estimated.

4.5.1. Efficiency distributions

In the evaluation of exogenous factors on the cost efficiency measures, it is important to first present the distributional properties of the estimated cost efficiency measures of DEA and SFA. These results are presented in Table 4.4. It is important to note that the average cost efficiency measures over time range from 0 (fully inefficient) to 1 (fully efficient). From Table 4.4, the mean efficiency measures fluctuate over time. The average cost efficiency measures over time range between 0.362 to 0.961 for SFA model and from 0.681 to 1.000 for DEA model. Regarding the overall mean values of cost efficiency scores over time, the results indicate that the average bank could reduce its cost efficiency measure from 0.039 to 0.638 for SFA model and from 0 to 0.319 for DEA model to match the bank's performance with the best possible bank practices.

Table 4.4 further shows that the SFA cost mean efficiency (0.853) is higher than the DEA cost mean (0.839). The difference in both mean efficiency measures is 1.4 percent. Such difference is not surprising given that SFA model is characterized by its ability to incorporate measurement errors while DEA model does not capture random noise where any deviation from the estimated frontier is interpreted as being due to inefficiency. The inconsistency between both methods is further illustrated by the standard deviation of efficiency estimates. There exists a large variation between the minimum efficiency measure for a given year and the pooled efficient measures for both methods. However, it is important to accentuate that these results are consistent with Dong et al., (2014). In the evaluation of the cost efficiency measures over time, the results suggest that the financial crisis of 2007-2009 took its toll on banks during the latter part of the financial crisis.

In Table 4.4, two interesting additional results emerge. First, the minimum and maximum cost efficiency measures of SFA are always less than the minimum and maximum efficiency measures of DEA. This is expected due to the random noise present in SFA model and the choice of the inefficiency distribution. Second, the mean efficiency of DEA is always less than the mean efficiency of SFA even though the minimum and maximum efficiency measures of SFA are always less than the minimum and maximum efficiency measures of DEA. The reason for this result lies in the standard deviation.

In the case of the pooled efficiency measures, the mean efficiency measure of DEA is 0.839 with a standard deviation 0.049, indicating that most of the banks fall between 0.79 and 0.888 efficiency scores. However, in SFA, with a pooled efficiency measure of 0.853 and a standard deviation of 0.058, most of the banks have efficiency measures falling 0.795 and 0.911. Furthermore, to illustrate this discrepancy, consider the year 2005. In 2005, the mean of DEA efficiency measure is 0.837 with a standard deviation of 0.046, indicating that most of the efficiency measures of the banks fall between 0.791 and 0.883 with only one bank having the minimum efficiency measure of 0.681. However, in SFA, with a mean of 0.854 and a standard deviation of 0.053, most of the banks efficiency measures fall between 0.801 and 0.907 with only one bank having an efficiency measure of 0.603.

Table 4.4. Summary of the cost efficiency measures

Year	Mean	Std.dev	Maximum	Minimum	Year	Mean	Std.dev	Maximum	Minimum
	SFA cost efficiency measures								
Pooled	0.853	0.058	0.362	0.961	2011	0.833	0.072	0.362	0.951
2005	0.854	0.053	0.603	0.947	2012	0.843	0.064	0.407	0.952
2006	0.867	0.048	0.645	0.953	2013	0.85	0.057	0.453	0.954
2007	0.869	0.047	0.608	0.948	2014	0.86	0.053	0.55	0.954
2008	0.858	0.054	0.568	0.957	2015	0.865	0.05	0.518	0.953
2009	0.836	0.059	0.371	0.948	2016	0.866	0.047	0.654	0.953
2010	0.829	0.064	0.464	0.961	2017	0.876	0.04	0.685	0.955
	DEA cost efficiency measures								
Pooled	0.839	0.049	0.681	1.000	2011	0.826	0.047	0.681	1.000
2005	0.837	0.046	0.702	1.000	2012	0.834	0.048	0.71	1.000
2006	0.847	0.049	0.712	1.000	2013	0.84	0.048	0.711	1.000
2007	0.851	0.051	0.698	1.000	2014	0.848	0.049	0.731	1.000
2008	0.842	0.05	0.705	1.000	2015	0.852	0.049	0.731	1.000
2009	0.822	0.042	0.688	0.99	2016	0.852	0.048	0.745	1.000
2010	0.82	0.044	0.684	1.000	2017	0.859	0.045	0.764	1.000

Year: time of the efficiency measures. Pooled: overall mean and standard deviation of the cost efficiency measures. Minimum: minimum cost efficiency measures over time. Maximum: maximum cost efficiency measures over time.

Figure 4.1 presents the comparison of the efficiency measures using DEA and SFA models. Since most of the points lie below the reference line at 0.7 of the y-axis, this tells us that most of the DEA efficiency measures yield higher economic cost efficiency than the SFA does. This is further quantified in Table 4.4. Table 4.5 presents the correlation coefficient between the cost efficiency measures and the financial factors of liquidity and solvency and the state GDP.

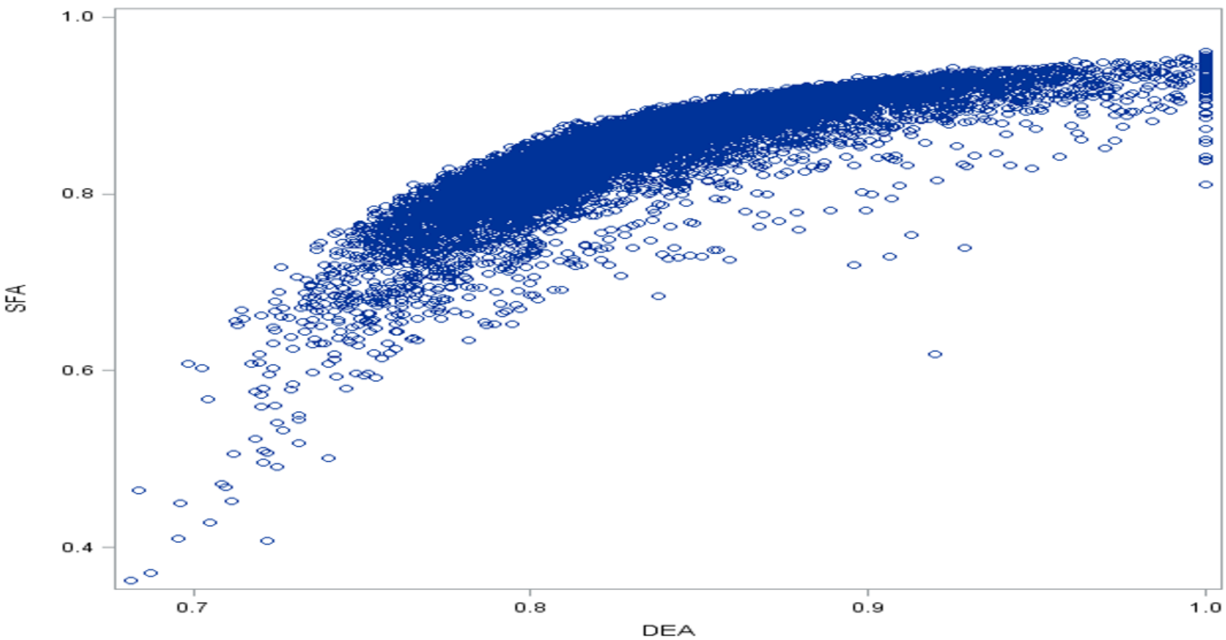


Figure 4.1. Comparing economic efficiency measures of SFA and DEA

Table 4.5. Strength of relationship between the efficiency measures and exogenous variables

Tobit Variables	Economic efficiency	Liquidity factor	Solvency factor	Bank size
	Stochastic Frontier Analysis Correlation			
Economic efficiency	1.000	0.056	-0.020	0.264
Liquidity factor	0.056	1.000	-0.144	-0.105
Solvency factor	-0.020	-0.144	1.000	0.026
Bank size	0.264	-0.105	0.026	1.000
	Data Envelopment Analysis Correlation			
Economic efficiency	1.000	0.043	0.0107	0.181
Liquidity factor	0.043	1.000	-0.144	-0.105
Solvency factor	0.0107	-0.144	1.000	0.026
Bank size	0.181	-0.105	0.026	1.000

Strong correlation: -1.0 to -0.5 or 1.0 to 0.5. Moderate correlation: -0.5 to -0.3 or 0.3 to 0.5 Weak correlation: -0.3 to -0.1 or 0.1 to 0.3. Very weak correlation: -0.1 to 0.1.

4.5.2. Impact of liquidity and solvency risks factors

To examine the determinants of cost efficiency, Table 4.6 presents the results of liquidity and solvency financial factors. The dependent variable is the inefficiency measures of either DEA and SFA and the exogenous variables of liquidity, solvency, bank size, and state GDP are in logarithms, and a dummy variable, crisis, is used to represent the financial crisis. The effect of the exogenous variables on the cost inefficiency measures were estimated using Tobit regression model (equation 4.6).[7] Our results reveal some interesting findings.

First, the estimated coefficient of liquidity financial factor, *Liq*, across both SFA and DEA models is negative and significant at a 1 percent significance level suggesting that liquidity has a negative impact on the cost efficiency measures. The results suggest that increasing the ratio of liquid assets to that of deposits of U.S banks does negatively impact the cost efficiency scores. This is consistent with the finding of Gorton and Metrick (2010), Lucas and Stokey (2011), and Cochrane (2013). In the assessment of liquidity during the financial crisis, the interaction term, $Liq \times crisis$, is negative and significant at a 5 percent significance level across SFA and DEA models. This suggests that banks' liquidity negatively collapses due to contagion. This result follows the subsequent paper of Berger and Bouwman (2008) who found that for the 2007-2009 financial crisis, there was an indication of a fall after the start of the crisis. The results further suggest that during the crisis, U.S banks with high exposure to liquidity demand suffered a further erosion in that advantage. Second, the solvency financial factor, *Sol*, is significantly negatively related to the economic cost efficiency respectively at a 10 percent and a 5 percent significance levels for SFA and DEA estimators. This result suggests that high capital requirements decrease the cost efficiency of banks (Berger and Bonaccorsi 2006; Pasiouras et al., 2009; and Pessarossi and Weill, 2014). The significance of solvency factor in explaining efficiency implies that banks with higher solvency risk (capital adequacy ratio) were less efficient since they were risk-averse. However, when controlling for the financial crisis, $Sol \times crisis$, solvency factor is positively insignificant at 10 percent significance level in DEA and SFA.

[7] The cost function of input prices and output quantities of the stochastic frontier in equation (4.5) indicate that the price of labor, the price of physical capital, total loans and other earning assets are positive and statistically significant at 1 percent. This is consistent with the theory associated with the cost function that an increase in the use of the input prices or output quantities will increase the total cost of production, all things being equal. These results are available upon request.

Third, bank size, *Size*, has significantly a negative effect on the cost efficiency measures of SFA and DEA estimators at a 1 percent significance level. Bank size is negatively related to the cost inefficiency measures, indicating larger bank performed better. This result further suggests that the amount of total assets of U.S banks does matter in the improvement of the cost efficiency measures. Similar results are found in Naceur and Roulet (2017), Pessarossi and Weill (2014), and Berger et al., (2009). We further interpret the significance of bank size as an indication of higher efficiency of large banks. Also, more profitable banks achieved higher economic efficiency. Fourth, the financial crisis, *crisis*, is negative and significant respectively at a 10 percent in SFA and a 1 percent in DEA signifance levels. The negative coefficient of crisis shows that during the financial crisis, banks were inefficient in comparison to the tranquil period.

Additionally, Table 4.6 presents the spearman rank correlation between the predicted and actual inefficiency measures of banks, the normality tests of the residual of the inefficiency measures using Kolmogorov-Smirnov test and Anderson-Darling test for normality of the respective SFA and DEA models, the log likelihood estimation value of SFA and DEA, and the variance equality test between the residual of the inefficiency measures of SFA and DEA methods. Figure 4.2 presents the comparison of the residual inefficiency measures of DEA and SFA models. Since most of the points lie below 0.1 of the y-axis the diagonal, this tells us that most of the SFA residuals are higher compared to the DEA inefficiency residuals.

Table 4.6. Impact of liquidity and solvency risks across SFA and DEA models

Parameter	SFA		DEA	
	Estimate	P-value	Estimate	P-value
Constant	0.213	<.0001	0.175	<.0001
Liq	-0.006	0.0054	-0.005	<.0001
Sol	-0.012	0.0609	-0.012	<.0001
$crisis$	-0.004	0.0875	-0.018	0.001
$Liq \times crisis$	-0.006	0.0234	-0.005	<.0001
$Sol \times crisis$	0.009	0.2216	0.003	0.419
$Dodd$	0.116	<.0001	0.07	<.0001
$Liq \times Dodd$	0.005	0.0517	0.004	<.0001
$Sol \times Dodd$	0.026	0.0003	0.014	<.0001
$basel3$	-0.049	0.0057	-0.045	<.0001
$Liq \times basel3$	-0.004	0.0322	-0.005	<.0001
$Sol \times basel3$	0.003	0.0409	0.006	<.0001
$Asset$	0.026	0.1297	0.021	<.0001
$Liq \times Asset$	0.007	<.0001	0.002	<.0001
$Sol \times Asset$	-0.007	0.2261	0.001	0.7744
$BKMO$	0.016	0.3137	-0.007	0.0559
$Liq \times BKMO$	0.002	0.1351	-0.002	<.0001
$Sol \times BKMO$	0.002	0.6716	0.006	<.0001
Size	-0.012	<.0001	-0.006	<.0001
GDP	0.002	0.0014	0.001	<.0001
Performance criteria				
σ	0.055	<.0001	0.047	<.0001
Spearman Rank	0.349		0.269	
KS		<0.010		<0.010
AD		<0.005		<0.005
Log likelihood	16351		22958	
Equality of variance		<.0001		

KS: Kolmogorov-Smirnov test for normality. AD: Anderson-Darling test for normality. Equal variance test between SFA and DEA inefficiency residual. Correlation between the predicted and actual inefficiency.

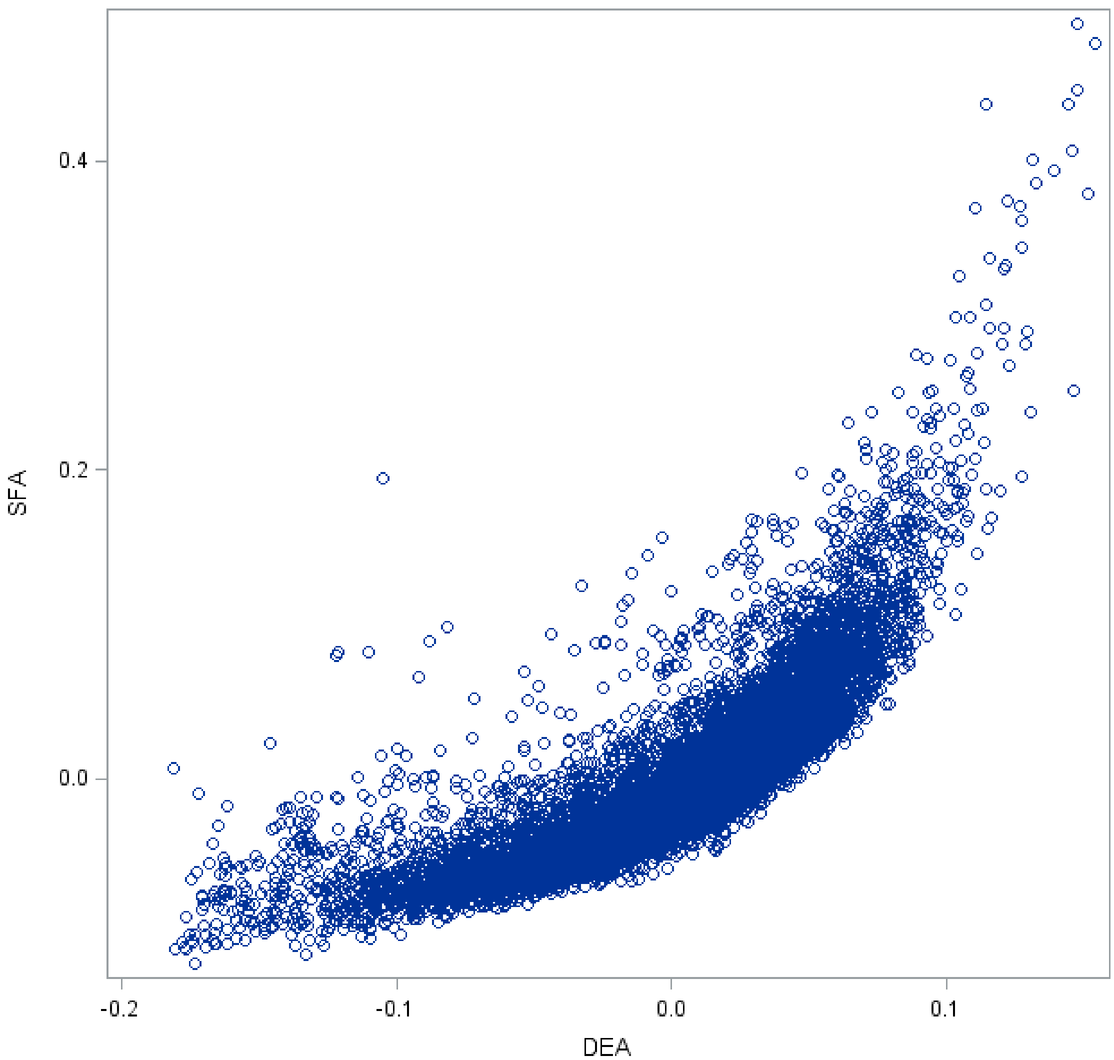


Figure 4.2. Comparing inefficiency residual of SFA and DEA

4.5.3. Impact of regulation changes

In this chapter, we additionally attempted to examine the impact of the regulatory factors on the cost efficiency measures. The regulations of concern were related to Basel Accord III (basel3) and the Dodd Franck Act (Dodd), both defined as indicators. With the on-going debate as to the costs and benefits of the regulatory frameworks (Barth et al., 2004 and Barth et al., 2013), the results of our research might provide empirical answers dealing with the impact of these regulatory approaches on the banking sector. Table 4.6 presents the estimation of the impact of the regulatory factors on the cost efficiency measures of banks. The results reveal two interesting findings.

First, we observe that basel3 is negative and statistically significant at a 1 percent significance level. This is consistent across both SFA and DEA methods. We observe that the estimated coefficients of the interaction term of liquidity financial factor and Basel Accord III, $Liq \times basel3$, is negative and significant at a 5 percent significance level, indicating a downturn in the bank efficiency measures. In addition, the interaction term of solvency financial factor and Basel Accord III, $Sol \times basel3$, is positive and significant at a 5 percent significance level. This indicates an improvement in the cost efficiency measures of banks. This further confirms the Basel Accord III recommendations by increasing the capital adequacy ratio and lowering the leverage ratio.

Second, Dodd is positive and statistically significant at a 1 percent significance level across both SFA and DEA methods. We additionally observe that the estimated coefficients of the interaction term of liquidity financial factor and Dodd Frank Act, $Liq \times Dodd$, and solvency financial factor and Dodd Frank Act, $Sol \times Dodd$, are positive and significant at a 5 percent significance level, indicating an upturn in the banks efficiency measures. Overall, the negative (positive) signs of regulations is consistent with the view that less (more) regulatory control allows banks to engage in less (more) activities and operate under (above) economies of scale.

4.5.4. Impact of asset classification and type of banks

With different goals in the banking sector, the cost efficiency of U.S banks might be influenced by the location of the headquarters and the type of banks (branch or headquarter). An important question that has not received widely the attentions of researchers is whether banks still need headquarters. In addition, since the cost efficiency of the banking industry is influenced by the overall financial markets stability, a better allocation of assets can improve bank performance. In Table 4.6, $BMKO$ is a dummy variable that is equal to one if the U.S bank is a headquarter and zero otherwise. *Asset* is a dummy variable that is equal to one if the U.S bank's primary asset is agricultural and zero otherwise. In the evaluation of $BMKO$ and *Asset* on the cost efficiency measures, the results provide different conclusions. $BMKO$ is negative and significant at a 10 percent significance level. The negative coefficient of $BMKO$ in DEA confirms that the state of headquarters worsens cost efficiency measures. However, the effect is not statistically significant in SFA at a 10 percent significance level. We can speculate on the reasons for this. The expansion of branch networks to direct services to rural as well as to metropolitan areas may have increased costs and thus, affected efficiency negatively in DEA.

Asset has a positive and statistically significant effect on the cost efficiency measures of DEA model at a 5 percent significance level. However, *Asset* is insignificant at a 10 percent significance level in SFA model. The significance of *Asset* implies that a higher share of agricultural asset banks contributes to lower cost on the efficiency measures. Therefore, agricultural assets are expected to face a smaller variation in efficiency, thus they contribute to the economic efficiency.

4.6. Conclusion and future research

The evaluation of the exogenous affecting the economic efficiency measure of U.S commerical and domestic banks is an important concept that can address the issues of maintaining confidence and stability in the banking sector. Hence, a two-step approach analysis estimation technique was conducted. In the first stage, the cost efficiency measures of U.S. banks were estimated using DEA and SFA models. In the second step, a Tobit regression model was used to evaluate the impact of liquidity and solvency risk factors, in addition to regulatory, macroeconomics, and bank characteristics.

Our sample consisted of a panel dataset of 11,044 observations covering the period of 2005-2017. Using the input prices and output quantities, the variable return to scale economic efficiency measures were estimated over time. The yearly variability allows the cost efficiency measures to differ through the technological change while accounting for the cross-sectional banks. The empirical estimates of the cost efficiency measures and different factors influencing the cost inefficiency terms present distinctive conclusions.

First, the negative estimated coefficient of liquidity financial factor suggests that increasing the ratio of liquid assets to that of deposits of the banks does negatively impact the cost efficiency scores. In the assessment of the role of liquidity financial factor during the financial crisis, U.S. banks negatively collapses due to contagion. In addition, the solvency financial factor is significant and negatively related to the cost efficiency, suggesting that high capital requirements decrease the cost efficiency of banks. However, during the financial crisis, solvency financial factor was insiginficant in DEA and SFA models. Second, bank size and financial crisis have significantly negative effects on the cost efficiency measures of U.S banks. These results suggest that the amount of total assets of banks matters in the improvement of the cost efficiency measures. Moreover, during the financial crisis, U.S. banks were inefficient in comparison to the tranquil period. Third, Basel Accord III negatively impacted on the cost efficiency measures of U.S banks, consistently across both SFA and DEA methods. The interaction term, liquidity financial factor and Basel Accord III negatively impacted the cost efficiency measures of U.S banks. In addition, the interaction, solvency risk factor and Basel Accord III positively impacted the cost efficiency measures of U.S banks, indicating that an improvement in the cost efficiency measures.

There are, however, limitations that future researchers could study to evaluate the impact of exogenous variables on the cost efficiency measures. For example, future research could incorporate the variability of the cost efficiency measures by using the variance as a function of explanatory variables. In comparison to the current framework, the results of the efficiency measures could vary regarding the first and second moments of the financial factors. It could be great to further incorporate the financial crisis as a indicator and study its implication on DEA and SFA cost efficiency measures. Research could also focus on applying cluster analysis as an extra exogenous factor for bank total assets and potential banking policy implications.

5. REFERENCE

Ahsanullah, M., Kibria, B., and Shakil, M. (2014). Normal and Students t-Distributions and Their Applications. Atlantis Press.

Aigner, D., and Chu, S. (1968). On estimating the industry production function. The American Economic Review, 58(4): 826-839.

Aigner, J., Lovell, C., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. Journal of Econometrics, 6:21-37.

Aly, G, Pasurka C., and Rangan N. (1990). Technical scale and allocative efficiencies in US banking: An empirical investigation. Review of Economics and Statistics, 72 (2):211-218.

Andersen, P., and Petersen, C. (1993). A procedure for ranking efficient units in data envelopment analysis. Management science, 39(10): 1261-1264.

Andor, M., and Hesse, F. (2011). A Monte Carlo simulation comparing DEA, SFA and two simple approaches to combine efficiency estimates. Center of Applied Economic Research Munster, University of Munster.

Andor, M., and Hesse, F. (2013). The StoNED Age: The Departure into a New Era of Efficiency Analysis? A Monte Carlo Comparison of StoNED and the 'Oldies' (SFA and DEA) Ruhr Economic Paper No. 394.

Aspachs, O., Nier, E., and Tiesset, M. (2005). Liquidity, Banking Regulation and the Macroeconomy. Electronic Journal, 673-883.

Baccouche, R., and Kouki, M. (2003). Stochastic production frontier and technical inefficiency: A sensitivity analysis. Econometric Reviews, 22(1): 79-91.

Badunenko, O., Henderson, J., and Kumbhakar, C. (2012). When, Where and How to Perform Efficiency Estimation. Journal of the Royal Statistical Society, Series A (175): 863-892.

Banker, D., and Chang, H. (2006). The super-efficiency procedure for outlier identification, not for ranking efficient units. European Journal of Operational Research, 175(2): 1311-1320.

Banker, D., Gadh, M., and Gorr, L. (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(3): 332-343.

Barth, J., Lin, C., Ma, Y., Seade, J., and Song, M. (2013). Do bank regulation, supervision and monitoring enhance or impede bank efficiency? Journal of Banking and Finance, 37(8): 2879-2892.

Barth, R., Caprio, G., and Levine, R. (2004). Bank regulation and supervision: what works best? Journal of Financial Intermediation, 13: 205-248.

Barth, R., Li, T., and Lu, W. (2009). Bank regulation in the United States. Economic Studies, 56(1):112-140.

Battese, E., and Corra, S. (1977). Estimation of a Production Function Model with Application to the Pastoral Zone of Eastern Australia. Australian Journal of Agricultural Economics, 21: 169-179.

Behr, A., and Tente, S. (2008). Stochastic frontier analysis by means of maximum likelihood and the method of moments. Discussion Paper Series 2: Banking and Financial Studies 19.

Berger, A., and Humphrey, D. (1997). Efficiency of Financial Institutions: International survey and directions for future research. European Journal of Operations Research, 98: 175-212.

Berger, N. (1993). Distribution-Free Estimates of Efficiency in the U.S. Banking Industry and Tests of the Standard Distributional Assumptions. The Journal of Productivity Analysis, 4: 261-292.

Berger, N., and Bonaccorsi, E. (2006). Capital structure and firm performance: A new approach to testing agency theory and an application to the banking industry. Journal of Banking and Finance, 30: 1065-1102.

Berger, N., and Bouwman, S. (2009). Bank liquidity creation. Review of Financial Studies, 22: 3779-3837.

Berger, N., and Humphrey, B. (1992). Measurement and Efficiency Issues in Commercial Banking, in Measurement Issues in the Service Sector, Z. Griliches (ed.), NBER, Chicago.

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

Berger, N., Hassan, I., and Zhou, M. (2009). Bank ownership and efficiency in China: What will happen in the world's largest nation? Journal of Banking & Finance, 33(1): 113-130.

Berrospide., M. (2013). Bank liquidity hoarding and the financial crisis: an empirical evaluation. Finance and Economics Discussion Series, Board of Governors of the Federal Reserve System.

Bonner, C., Lelyveld, I., and Zymek, V. (2015). Banks' Liquidity Buffers and the Role of Liquidity Regulation. Journal of Financial Services Research, 48(3): 215-234.

Bravo-Ureta, E., and Pinheiro, E. (1993). Efficiency analysis of developing country agriculture: A review of the frontier function literature. Agricultural and Resource Economics Review, 22: 88-101.

Chang, P. (1999). Measuring efficiency with quasi-concave production frontiers. European Journal of Operational Research, 115(3):497-506.

Charnes, A., Cooper, W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. European Journal of Operational Research, 2: 429-444.

Chen, T. (2002). A comparison of chance-constrained DEA and stochastic frontier analysis: Bank efficiency in Taiwan. The Journal of the Operational Research Society, 53(5): 492-500.

Chortareas, E., Girardone, C, Ventouri, A. (2012). Bank supervision, regulation, and efficiency: Evidence from the European Union. Journal of Financial Stability, 8(4): 292-302.

Christopher, P., and Kumbhakar, S. (2014). Efficiency Analysis: A Primer on Recent Advances. Foundations and Trends in Econometrics, 7(34): 191-385.

Cochrane, H. (2013). Stopping Bank Crises Before They Start. The Wall Street Journal: Opinion.

Coelli, J. (1995). Recent developments in frontier modelling and efficiency measurement. Australian Journal of Agricultural and Resource Economics, 39: 219-245.

Coelli, T. (1995). Estimators and hypothesis tests for a stochastic frontier function: A monte Carlo analysis. Journal of Productivity Analysis, 6: 247-268.

Cooper, W., and Tone, K. (1997). Measures of Inefficiency in Data Envelopment Analyis and Stochastic Frontier Estimation. European Journal of Operational Research, 99(1): 72-88.

Daraio, C., and Simar, L. (2007). Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications. New York, NY: Springer.

Demirguc-Kunt, A., Detragiache, E., and Tressel, T. (2008). Banking on the principles: Compliance with Basel core principles and bank soundness. Journal of Financial Intermediation, 17: 511-542.

Diamond, W., and R. G. Rajan. (2005). Liquidity Shortages and Banking Crises. Journal of Finance, 60 (2): 615-647.

Distinguin, R., & Tarazi. T. (2013). Bank regulatory capital and liquidity: Evidence from US and European publicly traded banks. Journal of Banking and Finance, 37(9): 3295-3317.

Dong, Y., Hamilton, R., and Tippett, M. (2014). Cost efficiency of the Chinese banking sector: A comparison of stochastic frontier analysis and data envelopment analysis, Economic Modelling, 36(C), 298-308.

Drake, M., Hall, B., and Simper, R. (2006). The Impact of Macroeconomic and Regulatory Factors on Bank Efficiency: A Non-Parametric Analysis of Hong Kong Banking System. Journal of Banking Finance, 30:1443- 1466.

Dusansky, R., and Wilson, W. (1994). Measuring efficiency in the care of developmentally disabled. Review of Economics and Statistics, 76(27): 340-345.

Dusansky, R., and Wilson, W. (1995). On the relative efficiency of alternative modes of producing public sector output: The case of developmentally disabled. European Journal of Operational Research, 80: 608-628.

Dyson, G., Allen, R., Camanho, S., Podinovski, V., Sarrico, S., and Shale, A. (2001). Pitfalls and Protocols in DEA. European Journal of Operational Research, 132: 245-259.

Farrell, J. (1957). The Measurement of productive efficiency, Journal of Royal Statistical Society, Series A (120): 253-290.

Färe R, Grosskopf S, Lovell., K. (1985). The Measurement of Efficiency of Production. Kluwer-Nijhoff: Boston.

Federal Deposit Insurance Corporation (FDIC).

Fen, G., and Zhang, X. (2012). Productivity and efficiency at large and community banks in the US: A Bayesian true random effects stochastic distance frontier analysis. Journal of Banking and Finance, 36(7), 1883-1895.

Fried, O., Lovell, K., Schmidt, S., and Yaisawarng, S. (2002). Accounting for environmental effects and statistical noise in data envelopment analysis. Journal of Productivity Analysis, 17(1-2): 157-174.

Gandrud, C., and O'Keeffe, M. (2016). Information and financial crisis policy-making. Journal of European Public Policy, 1-20.

Golany, B., and Roll, Y. (1989) An Application Procedure for DEA. Omega, 17: 237-250.

Gong, B., and Sickles, C. (1992). Finite sample evidence on the performance of stochastic frontiers and data envelopment analysis using panel data. Journal of Econometric, 51: 259-284.

Gorton, B., and Metrick, A. (2010). Securitized Banking and the Run on Repo. Yale ICF Working Paper, 09-14.

Greene, W. (1990). A gamma-distributed stochastic frontier model. Journal of Econometrics, 46: 141-164.

Greene,W. (2004). Reconsidering heterogeneity and inefficiency: Stochastic frontier models. Journal of Econometrics, 126(2): 269-303.

Grosskopf, S., and Valdmanis, V. (1987). Measuring hospital performance: A non-parametric approach. Journal of Health Economics, 6 (2): 89-107.

Gstach, D. (1998). Another approach to data envelopment analysis in noisy environments: DEA+. Journal of Productivity Analysis, 9: 161-176.

Hafner, C., Manner, H., and Simar, L. (2016). The Wrong Skewness Problem in Stochastic Frontier Models: A New Approach. Econometric Reviews.

Hays, F., DeLurgio, S., and Gilbert, A. (2011). Efficiency Ratios and Community Bank Performance. Journal of Finance and Accountancy.

Hoggarth, G., and Saporta, V. (2001). 'Costs of Banking System Instability: Some Empirical Evidence. Financial Stability Review, Bank of England.

Holland, D., and Lee, T. (2002). Impacts of random noise and specification on estimates of capacity derived from data envelopment analysis. European Journal of Operational Research, 137(1): 10-21.

Horrace, C. and Schmidt, P. (1996). Confidence Statements for Efficiency Estimates from Stochastic Frontier Models, Journal of Productivity Analysis, 7: 257-282.

Horrace, C., and Christopher, P. (2015). A Laplace Stochastic Frontier Model. Econometric Reviews.

Huang, C., and Lai, J. (2012). Estimation of stochastic frontier models based on multi-model inference. Journal of Productivity Analysis, 38(3): 273-284.

Jenkins, L., and Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. European Journal of Operational Research, 147: 51-61.

Jensen, U. (2005). Misspecification preferred: the sensitivity of inefficiency rankings. Journal of Productivity Analysis, 23: 223-244.

Jickling, M. (2008). Averting financial crisis. Report, Congressional Research Service, Library of Congress.

Jickling, M. (2009). Causes of the financial crisis. Washington, DC.: Congressional Research Service, Library of Congress.

Johnson S., and Kwak, J. (2011). 13 Bankers: The Wall Street Takeover and the Next Financial Meltdown. Vintage; Reprint edition.

Jondrow, J., Materov, I., Lovell, K., and Schmidt, P. (1982). On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. Journal of Econometrics, 19: 233-238.

Kashyap K., Raghuram R., and Stein, J. (2002). Banks as Liquidity Providers: An Explanation for the Co-Existence of Lending and Deposit-Taking. Journal of Finance, 57 (1): 33-73.

Kittelsen, S. (1995). Monte Carlo simulations of DEA Efficiency Measures and Hypothesis Test. University of Oslo doctoral thesis, presented at Georgia productivity workshop.

Kumbhakar, S., and Lovell, K. (2000). Stochastic Frontier Analysis, Cambridge University Press.

Lee, T., and Holland, D. (2000). The impact of noisy catch data on estimates of fishing capacity derived from DEA and stochastic frontier models: A Monte Carlo comparison. In: Proceedings of the 10th Conference of the International Institute for Fisheries Economics and Trade. Corvallis, Oregon.

Lensink R., and Meesters A. (2014). Institutions and Banks Performance: A stochastic Frontier Analysis. Oxford bulletin of economics and statistics 76(1): 0305-9049.

Liu, C., Laporte, A., and Ferguson, S. (2008). The quantile regression approach to efficiency measurement: insights from Monte Carlo simulations. Health Economic, 17: 1073-1087.

Lucas, E., and Stokey, N (2011). Liquidity Crises. Mimeo University of Chicago.

Mcaleer, M., Jimenez-Martin, J., and Perez-Amaral, T. (2013). Has the Basel Accord improved risk management during the global financial crisis? North American Journal of Economics and Finance, 26: 250-265.

Meeusen, W., and van den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. International Economic Review, 18: 435-444.

Moradi-Motlagh, M., and Babacan, A. (2015). The impact of the global financial crisis on the efficiency of Australian banks. Economic Modeling, 46: 397-406.

Mullarkey, S., Caulfield, B., McCormack, S., and Basu, B. (2015). A framework for establishing the technical efficiency of electricity distribution counties (EDCs) using data envelopment analysis. Energy Convers Management, 94: 112-123.

Naceur, S., Almarzoqi, R., and Scopelliti, A. (2015). How Does Bank Competition Affect Solvency, Liquidity and Credit Risk? Evidence from the MENA Countries. IMF Working Papers.

Naceur, N., and Roulet, C. (2017). Basel III and Bank-Lending: Evidence from the United States and Europe, IMF Working Paper No.17/245.

Neff, D., Dixon, L., and Zhu, S. (1994). Measuring the efficiency of agricultural banks. American journal of agricultural economics, 76: 662-668.

Olesen, B., and Petersen, N. (2016). Stochastic Data Envelopment Analysis A review. European Journal of Operational Research, 251(1): 2-21.

Olson, A., Schmidt, P., and Waldman, M. (1980). A monte Carlo study of estimators of stochastic frontier production functions. Journal of Econometrics, 13: 67-82.

Ondrich, J., and Ruggiero, J. (2001). Efficiency Measurement in the Stochastic Frontier Model. European Journal of Operational Research, 129: 434-442.

Ondrich, J., Ruggiero, J. (2002). Outlier detection in data envelopment analysis: an analysis of Jackknifing. Journal of the Operational Research Society, 53 (3): 342-346.

Orea, L. and Kumbhakar, S. (2004). Efficiency Measurement using a Stochastic Frontier Latent Class Model. Empirical Economics, 29: 169-183.

Parkan, C. (1987). Measuring the efficiency of service operations: An application to bank branches. Engineering Costs and Production Economics, 12: 237-242.

Pasiouras, F., Tanna, S., and Zopounidis, C. (2009). The Impact of Banking Regulations on Banks' Cost and Profit Efficiency: Cross-Country Evidence. International Review of Financial Analysis, 18: 294-302.

Pastor, T., Ruiz, L., and Sirvent, I., (1999). A statistical test for detecting influential observations in DEA. European Journal of Operational Research, 115 (3): 542-554.

Perelman, S., and Santin, D. (2009). How to generate regularly behaved production data? A Monte Carlo experimentation on DEA scale efficiency measurement. EJOP, 199(1): 303-310.

Pessarossi, P., and Weill, L. (2014). Do capital requirements affect cost efficiency? Evidence from China. Journal of Financial Stability, 19: 10-16.

Podinovski, V., and Thanassoulis, E. (2007). Improving discrimination in data envelopment analysis: some practical suggestions. Journal of Productivity Analysis, 28: 117.

Ratnovski, L. (2013). Liquidity and transparency in bank risk management. Journal of Financial Intermediation, 22(3): 422-439.

Resti, A. (2000). Efficiency measurement for multi-product industries: A comparison of classic and recent techniques based on simulated data. European Journal of Operation Research, 121(3): 559-578.

Ruggiero, J. (1999). Efficiency estimation and error decomposition in the stochastic frontier model: A Monte Carlo analysis. European Journal of Operational Research.

Ruggiero, J. (2004). Data Envelopment Analysis with Stochastic Data. Journal of the Operational Research Society, 55: 1008-1012.

Russo, A., and Katzel, J. (2011). The 2008 Financial Crisis and Its Aftermath: Addressing the Next Debt Challenge. Occasional Paper 82.

Santos, P., Amado, C., and Rosado, J. (2011). Formative evaluation of electricity distribution utilities using data envelopment analysis. Journal of Operational Research Society, 62(7):1298-1319.

Saunders, A., and Schumacher, L. (2000). The determinants of bank interest rate margins: An international study. Journal of International Money and Finance, 19(6), 813-832.

Sealey, W., and Lindley, T. (1977). Inputs, outputs, and a theory of production and cost at depository financial institutions, The Journal of Finance 32(4): 1251-1266.

Settlage, D., Dixon, B., and Thomsen, M. (2000). A comparison of various frontier estimation methods under differing data generation assumptions. University of Arkansas, Department of Agricultural Economics and Agribusiness, Staff Papers.

Shaik, S. (2015). Impact of liquidity risk on variations in efficiency and productivity: A panel gamma simulated maximum likelihood estimation. European Journal of Operational Research, 245(2): 463-469.

Sherman, D., and Gold, F. (1985). Bank branch operating efficiency: Evaluation with Data Envelopment Analysis. Journal of Banking and Finance, 9: 297-315.

Simar, L. and Wilson, W. (2009). Inferences from cross-sectional, stochastic frontier models. Econometric Reviews, 29: 62-98.

Simar, L., and Zelenyuk, V., (2011). Stochastic FDH/DEA estimators for frontier analysis. Journal of Productivity Analysis, 36(1): 1-20.

Smith, P. (1997). Model misspecification in Data Envelopment Analysis. Annals of Operations Research, 73: 233.

Stevenson, R. (1980). Likelihood Functions for Generalized Stochastic Frontier Estimation. Stochastic Frontier Estimation. European Journal of Operational Research, 99: 72-88.

Stone, M. (2002). How Not to Measure the Efficiency of Public Services (and How one Might). Journal of the Royal Statistical Society, Series A, 165: 405-434.

Thagunna, S., and Poudel, S. (2013). Measuring bank performance of Nepali banks: a data envelopment analysis (DEA) perspective. International Journal of Economics and Financial, 3(1): 54-65.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. Econometrica. 26 (1): 24-36.

Tsionas, G. (2017). When, Where, and How of Efficiency Estimation: Improved Procedures for Stochastic Frontier Modeling. Journal of the American Statistical Association.

United Nations Department of Economic: http://www.un.org/esa/ffd/documents/Outcome-2009.pdf.

United States Department of Treasury: Federal Reserve Senior Loan Officer Survey, Treasury calculations.

Volker. P. (2014). http://noahpinionblog.blogspot.com/2014/02/2008-liquidity-crisis-or-solvency-crisis.html.

Wheelock, C. (1993). Is the Banking Industry in Decline? Recent Trends and future Prospects from a Historical Perspective. Review, 3-22.

Wheelock, C., and Wilson, P. (1995). Explaining Bank Failures: Deposit Insurance, Regulation, and Efficiency. Review, 77.

Wheelock, D., and Wilson, P. (2007). Non-parametric, Unconditional Quantile Estimation for Efficiency Analysis with an Application to Federal Reserve Check Processing Operations. Journal of Econometrics.

Wilson, W. (1993). Detecting Outliers in Deterministic Nonparametric Frontier Models With Multiple Outputs. Journal of Business & Economic Statistics, 11(3); 319-323.

Wu, D., Zhou, D., and Birge, Z. (2011). Estimation of potential gains from mergers in multiple periods: A comparison of stochastic frontier analysis and Data Envelopment Analysis. Annals of Operations Research, 186(1): 357-381.

Wu, J., Yin, P., Sun, J., Chu, J., and Liang, L. (2016). Evaluating the environmental efficiency of a two-stage system with undesired outputs by a DEA approach: an interest preference perspective. European Journal of Operational Research, 254: 1047-1062.

Zhang, H., and Bartels, R. (1998). The effect of sample size on the mean efficiency in DEA with Application to Electricity Distribution in Australia, Sweden and New Zealand. Journal of Productivity Analysis, 9: 187-204.