Essays on building and evaluating two-stage DEA models of efficiency and effectiveness

by

Abbas Attarwala

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Management Sciences

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner:

Professor Joe Zhu,
Business School,
Worcester Polytechnic Institute

Supervisor:

Professor Stanko Dimitrov
Dept. of Management Sciences
University of Waterloo

Internal Member:

Professor Brian Cozzarin
Dept. of Management Sciences
University of Waterloo

Professor Fatih Safa Erenay
Dept. of Management Sciences
University of Waterloo

Internal-External Member:

Professor Liping Fu
Dept. of Civil and Environmental Engineering
University of Waterloo

Author's Declaration

Abstract

Researchers are not consistent in their choice of input and output variables when using two-stage data envelopment analysis (DEA) models to measure efficiency and effectiveness. This inconsistency has resulted in the development of many different two-stage DEA models of efficiency and effectiveness for the financial industry.

In this dissertation, I improved the statistical method from the MASc dissertation (Attarwala, 2016) by adding more features. These features are documented in Chapter 2 on page 4 and page 5. This statistical method evaluates efficiency and effectiveness models in the banking industry. It relies on the semi-strong version of the efficient market hypothesis (EMH). The EMH is motivated by the wisdom of the crowds, discussed in Section 2.2.2.

Previously (Attarwala, 2016), I found that the two-stage DEA model of Kumar and Gulati (2010) is not consistent with the semi-strong EMH for Indian and American banks. In this dissertation, using my improved statistical method, I show that the two-stage DEA model of Kumar and Gulati (2010) is not consistent with the semi-strong EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA from 2000-2017. I address the question of whether a universal two-stage DEA model of efficiency and effectiveness exists by building a variable selection framework.

This variable selection framework automatically generates two-stage DEA models of efficiency and effectiveness. To do this, it uses the improved statistical method and a genetic search (GS) algorithm. The variable selection framework finds the best, universal, two-stage DEA model of efficiency and effectiveness consistent with the semi-strong definition of EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and from 2000-2017. I investigated the causal relationship between (a) the quantitative measures of efficiency and effectiveness from the best two-stage DEA model generated by the variable selection framework and (b) Tobin's Q ratio, a financial market-based measure of bank performance. Not only do I provide bank managers with a reasonable proxy for measuring efficiency and effectiveness, but I also address the question of whether acting on these input and output variables improves the performance of banks in the financial market.

Finally, I set up an optimization problem and find an optimal path from the two-stage DEA model of Kumar and Gulati (2010) to the best two-stage DEA model found by the variable selection framework. This optimal path provides a set of actionable items for converting a two-stage DEA model that is not consistent with the semi-strong EMH to one that is.

Acknowledgements

I joined Dr. Stanko Dimitrov's research team in the fall of 2014 as a MASc student. I am very fortunate and humbled to have such a wonderful and inspiring mentor. I appreciate all the patience and the care that he has provided me during these seven years. I have learned so much from Dr. Dimitrov over our weekly meetings. For this, I will always be grateful and indebted. I hope to sow and cultivate the same care and passion that Dr. Dimitrov bestowed on me with my students.

I am very grateful for my committee members' valuable feedback and time: Dr. Joe Zhu, Dr. Fatih Erenay, Dr. Brian Cozzarin and Dr. Liping Fu. Thank you.

My wife, Jenab, put her ambitions on hold so that she could help me nourish mine. I will always be grateful and indebted for this. Thank you.

However difficult it may get, my parents and sister always believed in me and supported me unconditionally. Thank you.

Dedication

To my son, Abraham. And my dada and dadi.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Researchers build many models of efficiency and effectiveness for the financial industry. These models differ in their choice of input and output variables for measuring efficiency and effectiveness. How does one select the appropriate model to use?

In this dissertation, I present a new way to validate whether models of efficiency and effectiveness are consistent with the semi-strong efficient market hypothesis (EMH). In Chapter 2 and Section 2.2.4, I talk more about the semi-strong EMH. The first version of this statistical method (Attarwala, 2016) was developed as part of my MASc dissertation at the University of Waterloo. In this dissertation, I develop a second version of the statistical method that improves the previous statistical method. This new version of the statistical method contains newer features. These newer features are documented in Chapter 2 on pages 4 and 5. In this dissertation, using my improved statistical method, I show that the two-stage DEA model of Kumar and Gulati (2010) is not consistent with the semi-strong EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA from 2000-2017. This begs the question of whether a universal two-stage DEA model of efficiency and effectiveness which meets these requirements exists. This dissertation investigates three research questions.

The first research question addressed in this Ph.D. dissertation is as follows: **how does one build a variable selection framework for finding a universal two-stage DEA model of efficiency and effectiveness consistent with the semi-strong definition of EMH in the financial industry for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA from 2000-2017?** To answer this question, I build a variable selection framework using three search algorithms: (a) a surrogate search optimization (SSO), (b) a genetic search (GS) algorithm, and (c) a multi-armed bandit algorithm (MABA). I evaluate each of the three search algorithms and select the best for traversing the search space of efficiency and effectiveness for two-stage DEA models. I found that the GS performed the best of the three algorithms. More details on why the GS was preferred over the SSO and the MABA are presented in Chapter 3 and Section 3.5 of this dissertation. The search space is characterized by different combinations of input and output variables of efficiency and effectiveness. Using the GS algorithm, the variable selection framework traverses the search space of two-stage DEA efficiency and effectiveness models and finds the best one according to the developed statistical method.

The second research question that this dissertation addresses is **whether there is a consistent (independent of geographic location and time-period) choice of input and output variables for measuring efficiency and effectiveness** in the financial industry using a two-stage DEA model of efficiency and effectiveness. I executed the variable selection framework on bank financial data from eight countries (Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA) from the year 2000 to 2017. I present the resulting universal model in Section 4.4 of Chapter 4 of this dissertation. I also present the input and output dimensions of efficiency and effectiveness frequently consistent with the semi-strong form of EMH in Section 4.4 of Chapter 4 of this dissertation.

Finally, my third research question is **how does one iteratively improve a specific two-stage DEA model of efficiency and effectiveness which is not consistent with the semi-strong form of EMH to a model which is consistent with the semi-strong form of EMH?** I address this by generating an optimal path from the two-stage DEA model of Kumar and Gulati (2010) that is not consistent with the semi-strong EMH to the best two-stage DEA model as found from my variable selection framework. This is accomplished by applying a sequence of elementary operation on the model of Kumar and Gulati (2010) that then transforms it to the best model found from my variable selection framework. An elementary operation includes three options: (1) adding a new variable to a two-stage DEA model, (2) swapping two variables of a two-stage DEA model, and (3) removing an existing variable of a two-stage DEA model. The optimal path contains the least number of steps not consistent with the semi-strong version of the EMH. For example, consider two paths. Each path consists of 7 elementary operations where the first path has 5 of the 7 elementary operations consistent with the semi-strong EMH; the second path has 3 of the 7 elementary operations consistent with the semi-strong EMH. The first path is preferred over the second path in my path-finding algorithm because of its lower number of steps not consistent with the semi-strong EMH-consistent. Chapter 5 and Section 5.3.3 provides more details on the path generation algorithm. In Section 5.4 of Chapter 5, I report the optimal path from Kumar and Gulati (2010) to the best two-stage DEA model of efficiency and effectiveness as found using my variable selection framework.

Chapter 2

A statistical method of goodness on quantitative models for `efficiency` and `effectiveness`

2.1 Introduction

Ever since the financial crisis of 2008, financial institutions all over the world have been required to improve their offerings. For instance, in Canada (McFarland, 2014) in 2014, the big five banks had a flat trend in productivity over the duration of year, and there was no improvement in their efficiency compared to the previous year. These big five banks must become more effective and efficient in their offerings. When banks examine their current state, they need to make difficult choices about how to allocate their limited resources (Stanley, 2015). Researchers agree that a bank should be effective ("do the right things") and efficient ("do the things right"); however, there is no consensus or agreement on what these "right things" are. It is very difficult, if not impossible, to compare studies of efficiency and effectiveness (Cameron, 1978; Steers, 1975), since few studies use the same definition. This inconsistent definition of efficiency and effectiveness has led to many different (Ho and Zhu, 2004; Paradi and Schaffnit, 2004) approaches to measuring efficiency and effectiveness quantitatively. Inconsistent quantitative models of efficiency and effectiveness factor into evaluating publicly traded banks (Kumar and Gulati, 2010).

Given the current state of model generation, a natural overarching research question emerges: *How can a model of banks' efficiency and effectiveness be quantitatively validated?*

Fortunately, for publicly traded banks, financial markets may be viewed as a measure of the wisdom of crowds (Surowiecki, 2005); where the measure is the stock price or another shareholder value creation metric (SHVCM), such as the Tobin's Q ratio. I acknowledge that other factors may impact a firm's financial performance. However, if a particular firm is deemed efficient and effective by market traders, this firm will financially outperform another firm that is less efficient or effective. In this dissertation, I propose a method to validate models of efficiency and effectiveness using financial market data. In particular, the contributions of this chapter are as follows:

1. I propose a statistical method to evaluate quantitative models of efficiency and effectiveness.

   Most financial institutions are publicly traded firms. Therefore, they disclose their operating parameters in quarterly or annual financial statements. Furthermore, as these firms are publicly traded, their market values are known instantaneously at any point in time. Using the semi-strong Efficient Market Hypothesis (EMH) (discussed in Section 2.2.4), the financial markets identify firms that are effective, "doing the right things," and efficient, "doing things right." With these high-level definitions of effectiveness and efficiency, a firm that is "doing the right things right" is both efficient and effective—the market values such firms higher than firms which are not efficient and effective according to the EMH. The proposed statistical method in this research reports the model that can best explain the correlation between SHVCM and the efficiency and effectiveness scores. The statistical method finds the best model from a family of models (I describe this procedure in detail in Section 2.3). The statistical method estimates the quality of each model relative to the others using an information criterion. The method infers whether the correlation of efficiency and effectiveness in the selected model is consistent with the semi-strong EMH by determining if it contains statistically significant parameters and a positive correlation between the selected SHVCM and the determined effectiveness and efficiency measures. I refer to this selected model as the *best* model.

   I developed this method to validate models of efficiency and effectiveness as part of my MASc work at the University of Waterloo. However, in my Ph.D., I improved this method by adding new features such as: (1) Standard errors can now also be computed as robust standard errors assuming that the error terms in the panel data are heteroskedastic (I define heteroskedastic in the Appendix A of this dissertation), (2) Standard errors can now also be computed as cluster standard errors assuming that the error terms in the panel data are heteroskedastic and autocorreleated (I define heteroskedastic and autocorreleated in the Appendix A of this dissertation) within groups, (3) Introduction of random effects model, (4) the Hausman test (I define Hausman test in the Appendix A of this dissertation), (5) the Mundlak test (I define Mundlak in the Appendix A of this dissertation) and (6) the Levin et al. (2002) and both Maddala and Wu (1999) and Choi (2006) of the Fisher panel unit root techniques are employed in this dissertation. The Hausman test is used in the statistical method to determine whether to use a fixed-effects model or random-effects model when the error terms are homoskedastic (I define homoskedastic in the Appendix A of this dissertation). The Mundlak test is used in the statistical method to determine whether to use fixed effects or random effects when the error terms are heteroskedastic and autocorrelated within the cross-sectional units of the panel data.

2. I determine whether the quantitative model of efficiency and effectiveness, proposed for the banking industry by Kumar and Gulati (2010), is consistent with the semi-strong EMH using my statistical method.

    To control for firm size and equities, I use the Tobin's Q ratio (Wolfe, 2003) as discussed in Section 2.4.1 of this dissertation. In my MASc thesis (Attarwala, 2016) I executed the statistical method separately on a subset of all banks in India for 2009 to 2013 and on a subset of all banks in the USA for 2007 to 2015. I determined that the two-stage DEA model of efficiency and effectiveness proposed for the banking industry by Kumar and Gulati (2010), is not consistent with the semi-strong EMH according to my statistical method.

    In this dissertation, for Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA for 2000-2017 (18 time periods), I first averaged the input and output dimensions of efficiency and effectiveness proposed by Kumar and Gulati (2010) across all the banks based on the banks' market capital. In Section 2.4.1 of this chapter I describe this step in more detail. Instead of running the statistical method separately for each of the countries as I did previously in my MASc, I now execute my general statistical method once for all the countries. More details about this are mentioned in Section 2.4 of this chapter.

The intended audience for the work done in this chapter are researchers validating existing or making new models of efficiency and effectiveness. Also, banks that like to assess whether their current model of efficiency and effectiveness is consistent with the semi-strong version of the EMH can use the statistical method of this chapter. In the remainder of this chapter, I first discuss related work in Section 2.2. In Section 2.3, I present my statistical method of goodness based on the semi-strong EMH. In Section 2.4, I present my case study and finally in Section 2.5, I present the result of my case study followed by managerial insight.

2.2 Related work

I highlight in Section 2.2.1 that there are many quantitative models of efficiency and effectiveness. Due to many quantitative models of efficiency and effectiveness how does one compare one model against an another?

Section 2.2.2 highlights work by researchers that consider financial data as crowd-sourced. In Section 2.2.3, I mention related work to my case study, and finally, in Section 2.2.4, I highlight related work in support of the semi-strong version of the EMH.

2.2.1 The need for validating quantitative models of efficiency and effectiveness

Steers (1975) mentions that firms and individuals agree that they would like to be highly effective and efficient. However, there is no agreement or consensus (Mahoney and Weitzel, 1969; Steers, 1975) on how to qualitatively define efficiency and effectiveness. There are many models of efficiency and effectiveness due to the lack of consensus on the qualitative definition of efficiency and effectiveness. Steers (1975) provides a taxonomy of different types of qualitative models (univariate, multivariate, normative, deductive) of efficiency and effectiveness. There is minimal overlap between these models. Sharma and Singh (2019) mentions that measuring organizational effectiveness is very much researcher-dependent. Researchers may select few variables to measure organizational effectiveness. On other occasions, when researchers select many variables to measure effectiveness, they may miss out on the overall context, i.e., researchers may reuse variables from one industry sector in another sector without understanding the overall context. Some researchers have further divided organizational effectiveness into financial measures (Amah and Ahiauzu, 2013) such as using profitability or operational measures (Amah and Ahiauzu, 2013; Budhiraja and Malhotra, 2013) such as using productivity or structural measures (Abd Rahman et al., 2013; Budhiraja and Malhotra, 2013) such as using innovation when measuring organizational effectiveness. In literature, I find the lack of consensus on the qualitative definition of organizational efficiency and effectiveness echoed by early researchers such as (Mahoney and Weitzel, 1969) and also recent researchers such as Mishra and Misra (2017).

Due to the lack of agreement on a qualitative definition, it is not surprising that there is no consensus on quantitative models (Ho and Zhu, 2004; Kumar and Gulati, 2010) of efficiency and effectiveness. One drawback of the lack of consensus amongst quantitative models is that one firm may be characterized as efficient and effective in one model but not efficient or not effective in another.

Steers (1975) mentioned that achieving high efficiency and effectiveness is a global objective among firms. If a manager is interested in improving his firm, each model may lead to different measures of efficiency or effectiveness, therefore suggesting different corrective actions. I partially address the lack of consistency by proposing a statistical method that determines if such models are consistent with the semi-strong EMH. When evaluating and comparing two different quantitative models, I noticed that the selection of input and output variables are subjective depending on the stakeholders. For instance, Morita and Avkiran (2009) mention that deciding whether a variable is considered input or output in a quantitative model of efficiency is based on the "desirability" of the variable. According to the authors, desirable variables are considered outputs, and those deemed undesirable are considered inputs. For example, a firm owner is more interested in running their organization with a smaller number of employees (a less desirable, input variable when measuring efficiency). In contrast, a branch manager may be interested in having more employees (a more desirable, output variable when measuring efficiency). Both the stakeholders may claim their models are correct when measuring performance.

In some cases, such as production (Metzger, 1992), the input and the output variables for measuring efficiency are well defined. However, in the financial industry, there are many competing models (Chu and Lim, 1998; Ho and Zhu, 2004; Kumar and Gulati, 2010; Li et al., 2019) which each have their own definition of efficiency. For instance, Boussemart et al. (2019) propose two different definitions of efficiency: (1) economic efficiency and (2) risk efficiency. Each type includes a different choice of input and output variables. They mention that the loans provided by banks can be classified as good loans that can generate income for the bank or bad loans that generate no income for the bank. For example the authors mention in their paper that economic efficiency comes from the production of good outputs, namely interest and non-interest income, while credit risk management efficiency is related to the minimization of the non-performing loans that can be considered as an unintended or bad output. Tamatam et al. (2019) mentions that their study is limited by the fact that they chose only one definition (as defined by choice of input and output variables) of efficiency when measuring the performance of Indian banks. Instead, additional insights can be drawn by employing different models based on other input and output variables. Wijesiri et al. (2019) proposes that researchers do not define efficiency correctly when measuring the performance of Indian banks. For instance, Wijesiri et al. (2019) defines social efficiency which is a measure of social goodness (empowering women, for example) and proposes using the number of accounts owned by women as an output variable when measuring the efficiency of Indian banks.

With the plethora of model types, inputs, and outputs, evaluating and comparing these models is key. I propose a new method for comparing models of efficiency and effectiveness. Different quantitative models of efficiency and effectiveness lead to different performance measures, leaving a firm manager at a loss as to which model to use. My research helps determine which model of efficiency and effectiveness is consistent with the semi-strong EMH in a financial setting.

2.2.2 Considering financial data as crowdsourced

My statistical method uses correlation to determine how well the efficiency and the effectiveness measures computed from quantitative models matches financial data. Financial data of securities such as: stock price, market capitalization, and Tobin's Q ratio, is viewed as crowdsourced measures because the price of a security is determined by the traders that are trading at that point in time.

This idea of comparing quantitative models using correlation against crowdsourced data that I am using in my statistical method is not new. For example, Barbier et al. (2012) talks about validating unsupervised clustering algorithms using crowdsourced data. The learning task (classification, clustering, or semi-supervised) is first performed using an automated technique, akin to the quantitative models of efficiency and effectiveness in Attarwala (2016). The same task, is then executed on a smaller-scale, by humans (crowdsourcing). The results are compared (using correlation) and the accuracy (using root mean square error) is calculated between both the approaches.

In Agarwal et al. (2008), authors propose a method for identifying influential bloggers in the blogosphere. They evaluate their automatic technique by comparing their results to the crowdsourced data generated on Digg. They assume that the number of 'diggs', humans have assigned to the posts submitted by influential bloggers should be higher and based on this assumption validate their findings. Diasio (2012) mentions that Epagogix (software to predict box-office results of Hollywood movies before release) is correlated with the Hollywood Stock Exchange (HSX). Where HSX is an information market to predict box-office results which provides crowdsourced data.

2.2.3 Work related to our case study

Guenster et al. (2011); Konar and Cohen (2001) use the Tobin's Q ratio as the dependent variable in a regression-based study to determine the financial performance of a firm, like my work. However, unlike my work, the authors do not use their studies to validate models. Also, Konar and Cohen (2001) use a known functional relation in their regression while I do not know the form of that relation; therefore, I use stepwise regression to find one. Guenster et al. (2011) use a database of eco-efficiency values, while I use two-stage data envelopment analysis (DEA) to determine firms' efficiency and effectiveness scores as suggested by Kumar and Gulati (2010). I define and discuss the two-stage DEA in Section 2.4.2 of this dissertation.

Efficiency in the banking industry is a topic already considered in the literature. Adenso-Díaz (1997); Beccalli et al. (2006); Chu and Lim (1998) study the influence of a bank's efficiency scores on its stock price. Chu and Lim (1998) consider Singaporean banks, Beccalli et al. (2006) consider European banks from five countries and Adenso-Díaz (1997) consider Spanish banks. None of the these papers consider effectiveness; they only consider efficiency. In this chapter, I am concerned with both efficiency and effectiveness.

Additionally, linear regression models are also used for predicting SHVCM. Adenso-Díaz (1997); Chu and Lim (1998); Fiordelisi and Molyneux (2010) use a linear model in which a SHVCM is a function of various bank-specific, industry-specific, and other macroeconomic variables and DEA scores. However, the statistical method proposed in my dissertation searches for the best model among a family of models. Using the work of Adenso-Díaz (1997); Chu and Lim (1998); Fiordelisi and Molyneux (2010) as motivation for my case study, I use a family of linear models and find the best model using stepwise regression and the Akaike Information Criterion (AIC) (Bozdogan, 1987).

The case study is an application of my proposed statistical method. The case study shows that the quantitative model proposed by Kumar and Gulati (2010) is not consistent with the semi-strong EMH using market data. One may argue that the EMH should not be used as a benchmark for quantitative models since it is a hypothesis. In the next Section, I show that even though the EMH is a hypothesis, it is a valid benchmark for models of efficiency and effectiveness.

2.2.4 Arguments in support of the semi-strong EMH

There are three versions of the EMH: a) weak, b) semi-strong, and c) strong form. The weak form of the EMH claims that financial asset prices reflect all past publicly available information of the financial asset. The semi-strong form of the EMH claims that prices reflect all publicly available information and that prices instantly change to reflect new public information. In a semi-strong efficient market one would expect that when new and relevant information is made available then this new information is instantly absorbed in the determination of the SHVCM. One would also expect that any relevant information in the financial statements on day of disclosure will be correlated with the the SHCVM. A bank that is efficient and effective, i.e., doing the right things right would have its efficiency and effectiveness scores correlated with its Tobin's Q ratio. The strong form of the EMH additionally claims that prices instantly reflect even private, "insider," information.

I use the semi-strong version of EMH primarily because of data availability in the public domain and because I have no access to private or insider information. I describe the data that I use in the case study of this paper in greater detail in Section 2.4.1. The efficient market hypothesis is associated with the idea of a "random walk" (Fama, 1995). The idea behind the random walk is that changes in future stock prices react to future news and are entirely independent of present price changes. The news is unpredictable (Malkiel, 2003; Sull and Escobari, 2004). Therefore, the resulting price changes are unpredictable and random. Fama (1995) mentions that the random walk theory implies that a series of stock price changes have no memory. The history of the series cannot be used to predict the future. This argument is supported in the correlation study by Moore (1962) in which coefficients computed for successive price changes were nearly zero. Malkiel (2005) argues that if prices are irrational and if market returns are predictable, actively managed investment funds should easily outperform a passive index funds that buy and hold a market portfolio. The most persuasive evidence suggesting that markets are efficient is that professional investors do not beat the market (Malkiel, 2005). There is extensive literature (Jensen, 1968) that shows professional investment managers do not outperform index funds. Recently, Warren Buffett won a $1 million dollar bet by showing a market index fund performed better than a basket of hedge funds (Carrig, 2018). Malkiel (2003) mentions that, due to market efficiency, information in the public domain gets quickly consumed, resulting in no arbitrage opportunities.

Current EMH theory is valid because it has not yet been proven false. To my knowledge, there is no empirical case study or observation that has disproven the EMH (Popper, 1957). Given the volume of empirical and analytical work that suggests the EMH is true or mostly true (i.e., markets correct any inefficiencies within a short period), I would venture to say that benchmarking quantitative efficiency and effectiveness models against the EMH is better than simply accepting the models at face value.

2.3 General statistical method

I now present the main contribution of this chapter: the general statistical method of goodness for quantitative models of efficiency and effectiveness based on the semi-strong EMH. I use the word *general* because the statistical method is independent of any of the following: 1) a specific information criterion, 2) a specific family of models, and 3) specific parameter estimation techniques. The three points mentioned above are grouped inside the "Others" variable in the algorithm that follows. The "Others" variable is one of the preconditions of the general statistical method. I discuss the preconditions of the general statistical method in detail in the next subsection. The statistical method is also agnostic to how the efficiency and the effectiveness scores are computed. In my case study in Section 2.4, I compute the efficiency and the effectiveness scores using the two-stage DEA model of efficiency and effectiveness, however, the efficiency and the effectiveness scores can be computed from any other models as well.

2.3.1 Inputs

In this section, I describe the preconditions of my statistical method in greater detail. As seen in the Algorithm 1 of this chapter, my input is categorized into three separate groups. These three groups are: 1) Data from quantitative models, 2) Data from SHVCM, and 3) 'Others.' In the subsections that follow, I categorize 1) and 2) as Data and 3) as auxiliary inputs for running the general statistical method.

Data

I highlight the fact that the proposed general statistical method is agnostic to how the data is generated. The quantitative measures of efficiency and effectiveness can be generated via DEA or Stochastic Frontier Analysis (SFA) or some other quantitative models.

The data from the quantitative models are made up of: $\mathbf{X}_{\textbf{efficiency}} \in \mathbb{R}^{N\times T}$ which refers to the efficiency scores of N DMUs across T time periods; $\mathbf{X}_{\textbf{effectiveness}} \in \mathbb{R}^{N\times T}$ which refers to the effectiveness scores of N DMUs across T time periods. In the case study that follows in Section 2.4, N is 8 and refers to the 8 countries, i.e., Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and T is 18 and refers to the 18 years in 2000-2017. Regarding the timeline of events, my general statistical method is executed only after determining the quantitative scores of efficiency and effectiveness.

These efficiency and effectiveness scores with their lags serve as independent variables when regressed against the SHVCM, denoted by $\mathbf{Y_{SHVCM}}$ in the Algorithm 1.

$\mathbf{Y_{SHVCM}} \in \mathbb{R}^{\mathbf{N} \times \mathbf{T}}$ consists of market-derived estimates, i.e., the value of the firm. Some examples of SHVCM are the stock price, the market capital, or the Tobin's Q ratio. In my general statistical method, there is no restriction or limit on what is the selected SHVCM. For example, in my case study, see Section 2.4, I execute my general statistical method once by averaging out the input and output variables of efficiency and effectiveness for banks in each of the 18 time periods for all the countries based on their market capital. I use the Tobin's Q ratio as my SHVCM in the case study.

Auxiliary inputs for running the general statistical method

The general statistical method requires other auxiliary inputs as part of its preconditions. In the Algorithm 1 of this chapter, I define these auxiliary inputs as 'Others.' I now provide a brief description of some of these auxiliary inputs:

Calculating standard errors: I provide three options for calculating standard errors: (1) Assuming that the idiosyncratic errors in the panel data regression are homoskedastic, then the homoscedasticity-only standard errors are computed. (2) Assuming that the idiosyncratic errors in the panel data regression are heteroskedastic, then robust standard errors are calculated. (3) Assuming that the idiosyncratic errors in the panel data regression is heteroskedastic and potentially correlated over time within a country, I compute the heteroskedasticity-and-autocorrelation-robust (HAR) standard errors, also referred to as clustered standard errors. For instance, in my case study, see Section 2.4, I use (3) because, like HAR in regression with cross-sectional data, clustered standard errors are valid whether or not there is heteroskedasticity, autocorrelation, or both (Stock and Watson, 2020).

Family of models: Is denoted by $\mathcal{F}$, and is the set of models supplied by the user. There is no restriction on this family of models. This could be a linear or a non-linear family of models. On this family of models supplied by the user, I find the best functional model using some criterion. For example, in the case study, I use a family of linear models.

Criterion: My general statistical method also takes in as input the criterion, i.e., a way to determine the best model from a family of models. The criterion is an estimator of the relative quality of statistical models for a given set of data. Given a collection of models for the data, the criterion estimates the quality of each model, relative to each of the other models. My statistical method is independent and not fixed to any criterion. Some commonly used criterions are the AIC, Bayesian Information Criterion (BIC), the sum of square errors, R^2, or the adjusted R^2. For example, in the case study, I use the AIC criterion.

Number of lags: Each measure of efficiency and effectiveness may, in turn, contain lagged values, meaning that the current market valuation of a firm may be a function of the previous period's efficiency and effectiveness values. As input, I take in as input a number that denotes the number of lags to be used for efficiency and effectiveness measures. Line 2 of the Algorithm 1 of this chapter uses this number of lags in the creation of the panel data. For example, in the case study, I use two lags.

Algorithm 1 Algorithm of General Statistical Method (GSM)

Preconditions:

- Data from quantitative model: $\mathbf{X_{efficiency}} \in \mathbb{R}^{N\times T}, \mathbf{X_{effectiveness}} \in \mathbb{R}^{N\times T}$
  N is the number of DMU and T is the total number of time periods

- Data from SHVCM: $\mathbf{Y_{SHVCM}} \in \mathbb{R}^{N\times T}$
  N is the number of DMU and T is the total number of time periods

- Others: Standard error assumptions, Criterion, Number of lags, Family of Models ($\mathcal{F}$), Parameter estimation technique

Postconditions:

- Best Fitted Model: Coefficients of best fitted model as judged by the criterion

procedure GSP($\mathbf{X_{efficiency}}$,$\mathbf{X_{effectiveness}}$,$\mathbf{Y_{SHVCM}}$,$Others$)
  $PanelData \leftarrow reshape(\mathbf{X_{efficiency}}, \mathbf{X_{effectiveness}}, \mathbf{Y_{SHVCM}}, Others)$
  $BestFittedModel \leftarrow \emptyset$
  for each $f \in \mathcal{F}$ **do**
    $\widehat{f_{fe}} \leftarrow$ fit f on PanelData using fixed-effect and $Others$
    $\widehat{f_{re}} \leftarrow$ fit f on PanelData using random-effect and $Others$
    $Statistic \leftarrow$ HausmanTestOrMundlakTest($\widehat{f_{re}}$,$\widehat{f_{fe}}$)
    if isNullHypothesisTrue(Statistic) **then**
      $\widehat{f} \leftarrow \widehat{f}_{re}$
    else
      $\widehat{f} \leftarrow \widehat{f}_{fe}$
    if (isFittedModelBetterUsingCriterion($\widehat{f}$,$BestFittedModel$,$Criterion$) **then**
      $BestFittedModel \leftarrow \widehat{f}$
  return $BestFittedModel$

2.3.2 Algorithm

I use the stepwise linear regression algorithm (Derksen and Keselman, 1992) to find the best model. The data, as defined in Section 2.3.1, of this paper is first reshaped into panel data. This is shown in line 2 of the Algorithm 1. The panel data is in $\mathbb{R}^{NT\times(3+K)}$. Again, N refers to the number of decision making units (DMUs, countries in my case), and T is the total number of periods/stages. The 3 indicates the three columns in the panel data of 1) 'Time Periods' (refers to each time from $1,\ldots,T$), 2) Decision-Making Unit (DMU) id, and 3) the dependent SHVCM. K is the number of columns that make up the efficiency and the effectiveness and their lags. The user specifies the family of models $\mathcal{F}$. I search only within $\mathcal{F}$ and find the best model. For each model, f, in the family of models, I first fit the model on the panel data using fixed-effect, $\widehat{f}_{fe}$, and then using random-effect, $\widehat{f}_{re}$. Fixed-effect accounts for unobserved heterogeneity. The random-effect model is a special case of fixed-effect and allows for individual effects. If the idiosyncratic errors in the panel data regression are homoskedastic, I perform the Hausman test to discriminate between the fixed and random-effects models. The null hypothesis of the Hausman test is that both fixed-effect and random-effect are similar and will yield similar coefficients. The alternative hypothesis is that the unobserved heterogeneity is fixed and not random, and the fixed-effects model is preferred over the random-effects model on rejection of the null hypothesis. On the other hand, if the idiosyncratic errors in the panel are HAR, I use the Mundlak test (Mundlak, 1978) when deciding between fixed or random effects. The Mundlak approach suggests estimating the following regression: $y_{it} = \alpha + \beta X_{it} + \gamma \bar{X}_i + \mu_i + \eta_{it}$ where $\bar{X}_i$ are country specific means. A Wald joint significance test (I define Wald joint significance test in the Appendix A of this dissertation) on γ is performed. The null hypothesis is set to $H_0 : \gamma = 0$ assuming the random effects model (Álvarez et al., 2017).

Stepwise regression (Draper and Smith, 1981; stepwisefit in Matlab, 2014) is a systematic method for adding and removing terms from a multilinear model based on their statistical significance in a regression. The algorithm can be executed either using forward selection or backward elimination techniques. The algorithm begins with an initial model and then compares the explanatory power of incrementally larger (when performed using forward selection) or smaller (when performed using backward elimination) models. At each step, the p-value of the criterion is computed to test models with and without a potential term. If a term is not currently in the model, the null hypothesis is that the term would have a zero coefficient if added to the model. If there is sufficient evidence to reject the null hypothesis, the term is added to the model. Conversely, if a term is currently in the model, the null hypothesis is that the term has a zero coefficient. If there is insufficient evidence to reject the null hypothesis, the term is removed from the model.

2.3.3 Outputs and relation with semi-strong EMH

The output of the general statistical model is the best model defined as $BestFittedModel$ in the algorithm. This best model is chosen from a family of models, $\mathcal{F}$. The coefficients of the best-fitted model are checked for statistical significance at 5% and whether the inferred correlations are consistent with the semi-strong EMH.

A firm that is "doing the right things right" is both efficient and effective, and according to the EMH, the market values such firms higher than firms that are "not doing the right things right." Given a SHVCM, quantitative measures of efficiency and effectiveness, and associated auxiliary inputs; my proposed method determines a statistical correlation between the quantitative measure of efficiency and effectiveness and the SHVCM. According to the EMH, the correlation must be positive and statistically significant for the quantitative measures of efficiency and effectiveness. If not, then the measures are inconsistent with the EMH.

2.4 Case study

I now present a case study in which I test the validity of the model of efficiency and effectiveness proposed by Kumar and Gulati (2010). When proposing their model, Kumar and Gulati (2010), only consider banks in India. As the authors propose the model for India, I in my MASc (Attarwala, 2016) work tested Kumar and Gulati (2010)'s two-stage DEA model with Indian banks and fount it to be not consistent with the semi-strong version of the EMH. As a sanity check, I also tested the model for eight other countries separately (Brazil, Canada, China, Japan, Mexico, Nigeria, South Korea, USA). I found that the model is not consistent for any of the above countries. In this dissertation, I further investigate Kumar and Gulati (2010)'s two-stage DEA model by combining the banks of Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA for time period 2000-2017. For each country and for each time period, I combine all the countries' banks by using the available market capital data. Before discussing our case, I first acknowledge that the actual selection of an SHVCM and the parameters to the general statistical method should be done by an industry expert and not necessarily by me. The domain experts can use their innate knowledge and select the appropriate values for the general statistical method. As the sector changes, only the inputs of the statistical method need to change and not the method itself. In Section 2.4.3, I provide our rationale and choice of values to the general statistical method when presenting the results of my case study. I generate the efficiency and effectiveness measures using the two-stage DEA (Charnes et al., 1978; Cooper et al., 2004) model. I use the exact definition of efficiency and effectiveness proposed by Kumar and Gulati (2010) where two DEA models are run, one for the first stage (efficiency is determined), and another for the second stage (effectiveness is determined). The outputs of the first stage are inputs of the second stage. Note that some authors use alternative definitions of a two-stage DEA model (Raheli et al., 2017; Simar and Wilson, 2011) and I highlight the differences in the two definitions of two-stage DEA in Section 2.4.2 of this paper. By using the two-stage DEA model of efficiency and effectiveness, I am also consistent with earlier work of other researchers that have used the two-stage DEA model of efficiency and effectiveness (García-Sánchez, 2007; Kumar and Gulati, 2010; Margari et al., 2007) in generating quantitative results of efficiency and effectiveness.

2.4.1 Data

I collected data for this dissertation from Eikon. Eikon is a financial information service from Reuters (2017) that provides company data, financial market data, news, country, and economic data, analytics, and trading tools. From Eikon, I collected the financial banking data of eight countries for the time period of 2000 to 2017. I used the Global Industry Classification System Codes (GICS) when searching for banks across the eight countries. "GICS is a collaboration between Standard & Poor's and Morgan Stanley Capital International. GICS codes are 8-digit codes that correspond to various business or industrial activities, such as Oil & Gas Drilling or Wireless Telecommunication Services. GICS is based upon a classification of economic sectors, which can be further subdivided into a hierarchy of industry groups, industries, and sub-industries. In total, there are 10 economic sectors, 23 industry groups, 59 industries, and 123 sub-industries categories, to date".[1] The GCIS code for banks is 4010. For the time periods that I consider from 2000 to 2017, Brazil contains 22 banks, India contains 28 banks, China contains 10 banks, USA contains 71 banks, Canada contains 13 banks, Mexico contains 8 banks, South Korea contains 6 banks and Japan contains 83 banks. For each country and for each bank and for each time period, I collected 55 dimensions of financial data. 55 dimensions are the dimensions that are common across the banks of all 8 countries and for all the time periods that I consider. This data is used in Chapter 3 and Chapter 4, however, for this chapter, I do not use all

[1] http://fin.ncue.edu.tw/compustat/manual/globdata/Part3d.pdf

the 55 dimensions and instead use only 7 dimensions recommended by Kumar and Gulati (2010).

Eikon provides the option to convert all the local currency data into U.S. dollars and adjust for inflation. The firm's market value and its asset value used in calculating the Tobin's Q ratio are both collected from Eikon. In my case study, I choose the Tobin's Q ratio as the SHVCM. Tobins' Q ratio is used by Deng and Elyasiani (2008) to proxy for a firm value. The Tobin's Q ratio is defined as a ratio of the market value of a firm over the value of its assets. The Tobin's Q ratio allows us to compare firms of various sizes and asset values (Reguera-Alvarado et al., 2017), something share price and market value alone do not. A high value of Tobin's Q ratio is associated with higher efficiency of the firm and which must result in higher financial performance of firms (Reguera-Alvarado et al., 2017). Other studies (Coles et al., 2008; Fich and Shivdasani, 2006; López Iturriaga and Morrós Rodríguez, 2014) also use Tobin's Q ratio as a financial measurement for firms' performance. Some other studies (Ball and Kothari, 1994) have used other SHVCM and have found that the stock market assigns prices to shares that include all the relevant publicly available information. Due to this, Chu and Lim (1998) have considered the stock price as a good proxy for a bank's efficiency in an efficient financial market. Ball and Kothari (1994); Beccalli et al. (2006) mention that efficiency is derived from information available in the public domain. As per the semi-strong definition of the EMH, the same information in the public domain is also used for determining the stock price and other financial metrics contingent on the stock price, such as the Tobin's Q ratio. Hence I hypothesize that there must be some correlation between the SHVCM and the banking performance in an efficient financial market. The studies mentioned above only look at efficiency in their regression model, and I also incorporate effectiveness. The firm's market value and its asset value used in calculating the Tobin's Q ratio are both collected from Eikon.

Traditionally, Tobin's Q ratio has been used to determine whether a firm is overvalued or undervalued (Chappell Jr and Cheng, 1984). For instance, if Tobin's Q ratio is greater than 1, then the bank is overvalued. If Tobin's Q ratio of a bank is less than 1 then, the bank is undervalued. In this dissertation, I do not use Tobin's Q ratio to identify overvalued or undervalued banks. I use Tobin's Q ratio to compare large and small banks alike. Tobin's Q ratio may be the most appropriate SHVCM to use in my study because it controls for bank size, while other SHVCMs may depend on a bank's size. Dezsö and Ross (2012) mention that better (efficient and effective) firms create more economic value from a given quantity of assets, i.e., better firms have a higher Tobin's Q ratio compared to firms that are less efficient and less effective. Dezsö and Ross (2012) also mention that Tobin's Q ratio is a forward-looking measure that includes the value of a firm's future cash flows, which are capitalized in the market value of the firm. Table C.1 suggests that Tobin's Q ratios in the panel data are non-stationary (I define stationary and non-stationary time series in the Appendix A of this dissertation) due to the presence of the unit root. In a weak form of the EMH, any shock to stock price or Tobin's Q ratio due to news or political events must be transitory, resulting in a random walk process. A random walk process is usually a non-stationary process, and the presence of unit roots establishes the weak form of the market efficiency (Juliana et al., 2011). Now that the weak form of the market efficiency has been established, in this chapter, I validate whether a two-stage DEA model of efficiency and effectiveness is consistent with the semi-strong version of the EMH. Later in Chapter 3, I build models of efficiency and effectiveness consistent with the semi-strong version of the EMH. The prices are end-of-day closing prices at the end of the fiscal year when calculating the Tobin's Q ratio. I fill any missing entries using a moving window mean of length 3.

For every country, I then perform a weighted average based on the market capital for that year. I repeat the process also for the Tobin's Q ratio. I do this because the financial market is a collection of firms that is not necessarily representative of the economy as a whole.[2] Firms with a higher market capital dictate the performance of the whole financial market. This is one reason why we see a divide and a disconnect between the economy and the performance of the financial markets. Table 2.1 describe the dimensions of Kumar and Gulati (2010) from Eikon along with the mean and standard deviation of the input and output dimensions of efficiency and effectiveness. Table 2.1 also lists the mean and standard deviation of Tobin's Q ratio.

In the Appendix B, Table B.1 and Table B.2 summarizes the mean and standard deviation of all the 55 dimensions from Eikon and for Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017. Data from these tables is used in Chapter 4 when I consider all the 55 dimensions from Eikon in finding a universal two-stage DEA model of efficiency and effectiveness for the above countries and for timeperiod 2000-2017.

[2] https://knowledge.wharton.upenn.edu/article/why-is-the-stock-market-so-strong-when-the-economy-is-weak/

Table 2.1 Mean and standard deviation of input and output variables used in Kumar and Gulati (2010) across Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017

	Brazil	**India**	**China**	**USA**	**Canada**	**Mexico**	**SouthKorea**	**Japan**
Total Assets, Reported (mean)	15505474539.06	62012169859.04	1214148198097.82	1146487076763.28	427797504191.02	38177690344.25	195205366056.09	722660173677.39
Total Assets, Reported (std dev)	39402282173.75	45697940627.89	897955856811.74	309184644754.08	195826731259.31	13976732058.00	101648697034.17	355656579447.02
Full-Time Employees (mean)	12497.31	39569.07	230843.83	163295.07	49269.37	20052.42	10610.93	27949.66
Full-Time Employees (std dev)	16608.82	15728.59	107803.02	25715.49	11923.02	7372.15	8122.11	13005.50
Net Loans (mean)	6093079849.32	39092061228.24	618984979926.54	468402075538.06	206119599977.13	17594101744.68	134042911939.01	310033510429.56
Net Loans (std dev)	9710112037.08	33092344482.36	468037838296.31	128771791565.37	101336773177.33	5869632706.74	68786919973.49	129724106093.70
Other Earning Assets, Total (mean)	9600252776.41	19903479396.11	326839162227.61	512163524583.44	191991912819.32	14886641786.97	39798792447.62	309393708110.21
Other Earning Assets, Total (std dev)	26343958189.11	18043786142.34	230638567340.68	144707854265.46	84268313435.42	7991447261.65	23425654837.08	151172893978.31
Long Term Investments (mean)	1023912732.04	545006250.21	1084638186.01	6103757354.64	1607687247.50	255840402.07	406014828.14	3343965999.38
Long Term Investments (std dev)	561328155.28	860891368.27	711387543.82	4202849351.19	964234526.92	223532462.17	229434423.04	1979301047.10
Non-Interest Income, Bank (mean)	986392973.70	1274569466.89	9125309026.85	25204158687.13	7154864912.90	1013800753.59	8150989504.85	8349764721.25
Non-Interest Income, Bank (std dev)	2342834193.42	1490084247.83	8297582284.10	6398710242.17	2661315841.54	435705165.25	5673852135.58	3947149427.03
Interest Income, Bank (mean)	2006406781.49	4721677224.51	46344214146.62	40338242283.90	12535104947.25	3237312935.84	8431821025.29	9228791349.14
Interest Income, Bank (std dev)	4269107681.16	3865316154.22	34461120688.87	10702768879.96	4044809615.63	745751893.88	4580522308.79	3786322112.71
Tobin's Q ratio (mean)	0.222678	0.203511	0.254289	0.209894	0.387389	0.402661	0.421956	0.431717
Tobin's Q ratio (std dev)	0.155121	0.147646	0.247316	0.211861	0.312029	0.325002	0.354690	0.388474

Variable Description:

- Total Assets, Reported (measured as value of fixed asset in USA dollars)
- Full-Time Employees (measured as number of employees)
- Net Loans (measured as sum of deposits and borrowings in USA dollars)
- Other Earning Assets, Total (measured as advances/loans to general public in USA dollars)
- Non-Interest Income, Bank (measured as income from commission and brokerage etc in USA dollars)
- Interest Income, Bank (measured as difference between interest income and interest expense in USA dollars)

Finally, this dissertation builds models of efficiency and effectiveness and validates these models against the semi-strong version of the EMH. For this reason, I do not use the raw data from Eikon and regress it against Tobin's Q ratio.

2.4.2 Two-stage DEA model

DEA (Charnes et al., 1978) is a nonparametric approach to evaluating the performance of a set of peers. These peers are commonly referred to as Decision Making Units (DMUs). The performance of a single DMU is measured across multiple input and output dimensions.

For each DMU, DEA infers an efficiency score defined by the ratio of a weighted sum of its output to a weighted sum of its input subject to the constraint that the ratio does not exceed 1 for any DMU.

Some researchers (Chen et al., 2009; Chu and Zhu, 2021; Cook et al., 2010a; Ho and Zhu, 2004; Kumar and Gulati, 2010; Lim and Zhu, 2016; Wang et al., 2014, 2010) extend the DEA model just described to the two-stage DEA model. The first stage of the DEA model is used in calculating the efficiency scores, and the second stage of the DEA model is used in calculating the effectiveness scores. In a two-stage DEA model of efficiency and effectiveness, the output dimensions of efficiency are the same as the input dimensions of effectiveness. Figure 2.1 shows the two-stage DEA evaluation model I use in my case study. This kind of two-stage DEA model is also commonly referred to as a simple two-stage DEA (Cook et al., 2010b). I will henceforth continue to refer to the simple two-stage process DEA as two-stage DEA model of efficiency and effectiveness in this dissertation.

Cook et al. (2010b); Yang et al. (2017) differentiates a simple two-stage DEA from a two-stage network DEA. In a two-stage network DEA, additional inputs could be fed into the second stage. These additional inputs are not the output variables of the first stage. Also, the first stage of the two-stage network DEA can produce outputs that are not fed into the second stage. The mathematical program for such network DEA models is presented in the research of Cook et al. (2010b); Yang et al. (2017). In the variable selection framework presented in Chapter 3, I build models of efficiency and effectiveness using the simple two-stage DEA model. In order to use two-stage network DEA in the variable selection framework, I will first replace the mathematical program presented in Equations (2.1)–(2.4) with the mathematical program of two-stage network DEA presented in Cook et al. (2010b); Yang et al. (2017). Subsequently, in Chapter 4 where I find the best two-stage simple DEA model for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA for time period of 2000-2017, I will then find a best two-stage network DEA model. Finally, in Chapter 5, I will use the pathfinding algorithm and find an optimal path of transforming a certain two-stage network DEA[3] model that is not consistent with the semi-strong version of the EMH to the best two-stage network DEA found from the variable selection framework.

My reasons for using Kumar and Gulati (2010)'s two-stage DEA model for validation against the semi-strong version of the EMH and then later in Chapter 3 use the two-stage

[3]In my dissertation, I use Kumar and Gulati (2010)'s two-stage simple DEA model as an example of a model that is not consistent with the semi-strong version of the EMH. If I use two-stage network DEA in my variable selection framework, I would either need to find some existing two-stage network DEA model that is not consistent with the semi-strong version of the EMH or build one myself.

DEA model for building models of efficiency and effectiveness are (1) popularity of two-stage DEA models among researchers. For instance, Seiford and Zhu (1999) used a two-stage DEA model to measure the profitability and marketability of US banks. Zhu (2000) presented a financial performance model based on a two-stage DEA model for measuring the performance of Fortune Global 500 companies. Seiford and Zhu (1999); Zhu (2000) did not refer to the first stage as efficiency and second stage as effectiveness but instead refer to them as measuring profitability and measuring marketability. Yang (2006) used a two-stage DEA model for measuring the performance of the Canadian insurance industry and Kao and Hwang (2008) used a two-stage DEA model for estimating the efficiencies of Taiwanese non-life insurance companies. (2) Kumar and Gulati (2010) explicitly refer to their first stage in the two-stage DEA model as efficiency and the second stage as effectiveness. Other researchers such as Hafsal et al. (2020) also use the same definition and variables of Kumar and Gulati (2010)'s in their two-stage DEA model of efficiency and effectiveness. This suggests that Kumar and Gulati (2010)'s choice of input and output variables for measuring efficiency and effectiveness has some appeal among other researchers when measuring the performance of Indian banks. In my MASc (Attarwala, 2016), I already tested Kumar and Gulati (2010)'s two-stage DEA model of efficiency and effectiveness separately for Indian and US banks and found it to be not consistent with the semi-strong version of the EMH for both the countries.

For completeness, I acknowledge that there is another interpretation of the two-stage DEA model in the literature. Other researchers (Paleckova, 2019; Raheli et al., 2017; Simar and Wilson, 2011; Wanke, 2012; Yu et al., 2020) define the two-stage DEA model as one in which, in the first stage, efficiencies are calculated. In the second stage, to examine the effect of factors that influence the efficiency of DMUs, a regression model is estimated with the efficiency scores (computed from the first stage of DEA) as the dependent variable. There is no effectiveness calculated in the variant of the two-stage DEA model used by Raheli et al. (2017); Simar and Wilson (2011). However, in this variant of the two-stage DEA model, as identified by Simar and Wilson (2011), the regression used in the second stage of the DEA with efficiency scores as dependent variables will result in an incorrect data-generating model. Their concern is that linear regression in such a model will predict efficiency scores outside the bounds of 0 and 1. The efficiency scores are limited between the range of 0 and 1 by the DEA model. They argue that using a Tobit model is also incorrect because a Tobit model enforces a censoring that limits the dependent variables between 0 and 1. Meanwhile, efficiency and effectiveness scores between 0 and 1 have no such censoring applied when calculated using DEA. However, in my work, I am not using the DEA scores as dependent variables; instead, I use them as independent variables. The dependent variable is some SHVCM (Tobin's Q ratio), and hence my claim that the statistical model is agnostic to the underlying efficiency and effectiveness holds in the variant of the two-stage DEA model we consider. My statistical model uses an existing model of efficiency and effectiveness and regresses the SHVCM on the efficiency and effectiveness scores, and infers whether they are consistent with the semi-strong definition of the EMH.

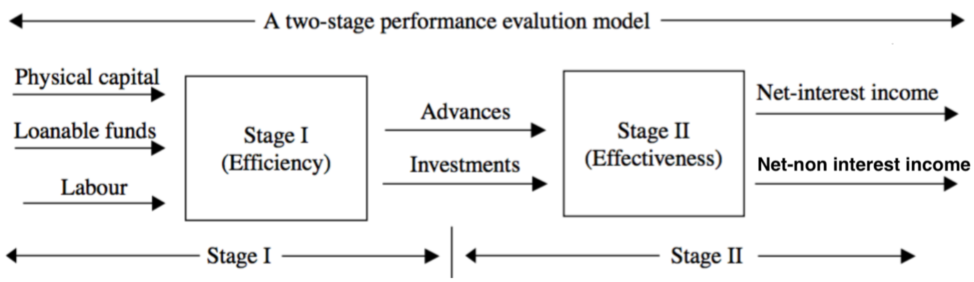

Figure 2.1: Two-stage DEA model as defined in Kumar and Gulati (2010) with their choice of input and output dimensions for efficiency and effectiveness. I use the same input and output dimensions in my case study.

1. In *stage I* (for efficiency), the selected inputs used for computing the efficiency scores are:
   (a) physical capital
   (b) labour
   (c) loanable funds
2. In *stage I* (for efficiency), the selected output variables are:
   (a) advances
   (b) investments
3. In *stage II* (for effectiveness), the selected input variables are:
   (a) advances
   (b) investments
4. In *stage II* (for effectiveness), the selected output variables are:

(a) net interest income

(b) net non interest income

In *stage II* of effectiveness scores calculation, I use the specific effectiveness criterion as outlined by Kumar and Gulati (2010). The effectiveness criterion is a measure of the extent to which policy objectives of an organization are achieved. In the years after 1992, i.e., the post-reform period in India, strong competition in Indian banking forced banks to minimize all non-essential costs while maximizing income from traditional and non-traditional activities. Mohan and Ray (2004) point out that Indian banks are maximizing incomes from all possible sources in the post-reforms period. This is the reason for Kumar and Gulati (2010) selecting: net interest income and non-interest income as the output variables in *stage II* of the performance evaluation model. Based on the above set of variables, the DEA model on each stage is as follows:

$$Max \quad \phi_k \tag{2.1}$$
$$s.t \quad \phi_k \mathbf{w_k} - \mathbf{W}\boldsymbol{\mu} \leq 0 \tag{2.2}$$
$$\mathbf{z_k} - \mathbf{Z}\boldsymbol{\mu} \geq 0 \tag{2.3}$$
$$\boldsymbol{\mu} \geq 0 \tag{2.4}$$

In mathematical program (2.1)–(2.4), subscript k refers to the k^{th} DMU under evaluation. The ϕ_k refers to the efficiency of DMU k when referring to stage one (see Figure 2.1) of the two-stage DEA model. The ϕ_k refers to the effectiveness of DMU k when referring to stage two (see Figure 2.1) of the two-stage DEA model. The vectors $\mathbf{z_k} \in \mathbb{R}^l$ and $\mathbf{w_k} \in \mathbb{R}^m$ refer to the input and output for bank k. m is the number of outputs and l is the number of inputs. The matrix $\mathbf{Z} \in \mathbb{R}^{l \times n}$ contains all the input dimensions of all the DMUs. The matrix $\mathbf{W} \in \mathbb{R}^{m \times n}$ contains all the output dimensions of all the DMUs. The inequalities on Equation (2.2) and (2.3) ensure that the input vector and output vector of DMU k is within or on the production frontier. I execute this linear program twice for each DMU, i.e. 1) for efficiency and then again 2) for effectiveness. The vector $\boldsymbol{\mu} \in \mathbb{R}^n$ is a semipositive vector.

2.4.3 Model setup and statistical method parameter selection

I first determine the independent ($\mathbf{X_{efficiency}}$ and $\mathbf{X_{effectiveness}}$) and dependent variables, $\mathbf{Y_{SHVC}}$ for my statistical method. I use AIC to determine the appropriate lag length when selecting a model for my panel data. AIC requires that the number of data points in my panel data cannot change when selecting a model. I first preprocessed the panel data by including 3 lags. By including 3 lags, I lose 3 data points for each of the 8 cross sectional units. I lose a total of $3 \times 8 = 24$ data points. The new panel data now consists of $8 \times 15 = 120$ data points. I notice that after preprocessing the panel data to include three lags, AIC determines the best statistical model of 1 lag. I repeat this process again by preprocessing the panel data to include 2 lags instead of 3 lags. By including 2 lags, I lose 2 data points for each of the 8 cross sectional units. I lose a total of $2 \times 8 = 16$ data points. The new panel data now consists of $8 \times 16 = 128$ data points. I notice that after preprocessing my panel data to include 2 lags, AIC determines the best statistical model of 1 lag. I therefore use up to 2 lags when executing the statistical method, similar to the study by Jung (1986), and because I notice no significant difference between the 2 and 3 lag models. Liew (2004) presents guidelines to use when selecting lag lengths. Liew (2004) concludes that the AIC is superior to most other information criteria in determining the correct lag length.

In my case study, I also use the AIC for the model selection criterion of model selection. AIC is a measure of the relative quality of statistical models for a given observed set of data. Given a family of models, AIC provides a means for model selection and balances the goodness of fit while discouraging overfitting of data (Stone, 1979). In the case study, I do not use R^2 because, by definition, R^2 is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. One can easily increase R^2 by adding more independent variables and overfitting the model to the data. The model that solely depends on R^2 will perhaps not explain or predict well for unobserved data.

With data and information criterion selected, I now turn my attention to choosing the best model from a family of models. I use the forward selection technique of stepwise linear regression. The forward selection technique involves starting with no variables in the model and tests the addition of each variable using our chosen information criterion. The starting model of the forward selection technique is set to model with no variables. The ending model is set to linear.

Model specification is a process of determining which independent variables to include and exclude from a statistical model. A model may be over-specified when there are too many variables, and a model may be under-specified when there are too few variables. In my statistical method, I use the AIC in the stepwise linear regression when balancing goodness of fit against the complexity of the model. In Section 4.4 of Chapter 4, I test for multicollinearity (I define multicollinearity in the Appendix A of this dissertation) amongst the independent variables of the regression model of the best two-stage DEA model of efficiency and effectiveness using the variance inflation factor. I address any unobserved heterogeneity in the regression model using the fixed effects model. If the covariance between the independent variables in the regression model is zero, then the standard error of the random effects estimator is smaller than the fixed-effects model. I use the Mundlak or the Hausman test to determine whether to use the random effects or the fixed effects model. The Mundlak test is used when the error terms in the panel regression model are heteroskedastic with autocorrelation. The Hausman test is used when the error terms in the panel regression are homoscedastic. Finally, in Section 4.4, I use the Pesaran cross-sectional dependency test to check for any cross-sectional dependence among the cross-sectional units in the panel data. When running the statistical method, I further assume that the error terms in the regression model are heteroskedastic with autocorrelation rather than assume homoscedasticity. Using a regression model with the assumption that the error terms are heteroskedastic and autocorrelated is valid because homoskedastic errors are a special case of heteroskedastic and autocorrelated errors (Stock and Watson, 2020). In the case study, I restrict the search to linear models. I restrict myself to linear models so that I am consistent with the work of Beccalli et al. (2006); Chu and Lim (1998); Ioannidis et al. (2008); Kirkwood and Nahm (2006). In all of these papers, the authors use either DEA or stochastic frontier scores to measure efficiency and regress computed scores with stock price using a linear model for different world banking sectors. I also use a linear model, but in addition to efficiency, I also compute effectiveness.

I use a t-test when inferring whether the coefficients of efficiency and effectiveness estimated are statistically significant or not. In Table 2.2, the model parameters are presented with a detailed breakdown of their cluster error along with the p-value of their estimates.

2.5 Output of statistical method

In Section 2.5.1, I talk about the statistical tests that I use for checking unit roots in panel data. In Table 2.2 the best linear model as determined by the statistical method for efficiency and effectiveness scores computed from Kumar and Gulati (2010)'s two-stage DEA model on Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017 is shown. In Section 2.5.2, I present an overview of my results and in Section 2.5.3, I provide managerial insights.

2.5.1 Unit root in panel data

I have used Choi (2006); Levin et al. (2002); Maddala and Wu (1999) statistical tests for checking unit root in panel data. Maddala and Wu (1999) and Choi (2006) are also commonly referred to as the Fisher panel unit root techniques. The Fischer panel unit root techniques are of two kind, i.e. (1) Fischer Augmented Dicky Fuller (ADF) test and (2) Fischer Philips-Perron (PP) test. Unlike the Fischer ADF test, the Fischer PP test is non-parametric.

Other researchers such as Ansari et al. (2020, 2019); Charfeddine and Mrabet (2017) have also used similar unit root finding tests on panel data in their studies. If the panel data has unit root, then spurious regression (I define spurious regression in the Appendix A of this dissertation) will occur due to non-stationary (Kao, 1999). When interpreting the coefficients from the regression, spurious regression leads to nonsensical results.

Levin et al. (2002) test assume that there is a common unit root process so that ρ_i is identical across all the cross sections i.e., $\rho_i = \rho$. The null hypothesis in Levin et al. (2002) assumes that there is a common unit root in the panel data while the alternative hypothesis is that there is no unit root in the panel data. Levin et al. (2002) uses the following model when determining whether y_i is stationary or not: $\Delta y_{i,t} = \alpha y_{i,t-1} + \sum_{j=1}^{p_i} \beta_{i,j} \Delta y_{i,t-j} + \delta X_{i,t} + \eta_{i,t}$. $X_{i,t}$ is the set of exogenous variables. $\alpha = \rho - 1$ and p_i is the lag order for the difference terms that is allowed to vary across the cross sections.

The Fischer ADF and Fischer PP tests allow for individual unit root processes, i.e., $\rho_i \neq \rho$. Maddala and Wu (1999) mention that the Fischer ADF and Fischer PP tests combine the evidence on the unit-root hypothesis from the N unit root tests performed on the N cross section units. In relation to finding whether unit root exists in the panel data; these tests perform a unit-root test on each of the cross section separately, and then combine the p-values to obtain an overall test of whether the panel data contains a unit root. The null hypothesis is that all cross sectional unit contain a unit root. The alternative is that at least one cross sectional unit is stationary.

I report the panel unit root test results for Tobin's Q ratio, efficiency and effectiveness in the Appendix C of this dissertation in Table C.1, Table C.2 and Table C.3 for Kumar and Gulati (2010)'s two-stage DEA model of efficiency and effectiveness. Table C.2 and Table C.3 indicate that the efficiency and effectiveness scores calculated from Kumar and Gulati (2010) are non-stationary. Table C.1 shows that Tobin's Q ratio is non-stationary. I take the log transformation of the efficiency and the effectiveness scores and then rerun these statistical tests and find that the $\log(efficiency)$ is stationary as per the three statistical tests, see Table C.5.

In Table C.6, as per Levin et al. (2002), $\log(effectiveness)$ is non-stationary. However, $\log(effectiveness)$ is stationary as per the Fischer ADF and Fischer PP test. However because the majority of the three test detect $\log(effectiveness)$ to be stationary, I conclude that $\log(effectiveness)$ is stationary. Table C.4 shows that $\log(TobinQ)$ is stationary across all the tests.

Table 2.2 lists the output from the statistical method after validating Kumar and Gulati (2010)'s two-stage DEA model of efficiency and effectiveness against the semi-strong EMH.

Table 2.2 Output of the best model from the statistical method on Kumar and Gulati (2010)'s two-stage DEA model

$N = 128$	
n=8	$T = 16$
$R^2 = .0229$	$AdjR^2 = -.070402$
Wald F(4,7)=.596511	p-value = .6770

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$\log(efficiency_t)$	-.855792	.707110	-1.2103	.265
$\log(effectiveness_t)$	-.105970	.202489	-.5233	.617
$\log(efficiency_{t-1})$	1.099682	.851496	1.2915	.238
$\log(effectiveness_{t-1})$	.050239	.257883	.1948	.851

Standard errors robust to heteroskedasticity adjusted for 8 clusters .

2.5.2 Hypothesis statement and results overview

I state my null hypothesis H_0 as follows: *There is no correlation of efficiency or effectiveness on Tobin's Q ratio.* I check the validity of the H_0 at significance level (α), of 0.05. If the p-value is less than (or equal to) α, then the H_0 is rejected for the alternative hypothesis. If the p-value is greater than α, then the H_0 is not rejected. I infer that the quantitative model of efficiency and effectiveness of Kumar and Gulati (2010) is not consistent with the semi-strong definition of the EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period of 2000-2017 due to the coefficient on the efficiency and effectiveness value at time period t not being positive and statistically significant.

For the quantitative model of efficiency and effectiveness to be consistent with the semi-strong definition of the EMH, the efficiency and effectiveness estimate at the time period t must be positive and statistically significant. In another paper of us, I infer similar results, i.e., the quantitative model of efficiency and effectiveness of Kumar and Gulati (2010) is also not consistent with the semi-strong definition of the EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017 when I run the the two-stage DEA model of Kumar and Gulati (2010) separately for each of the 8 countries.

2.5.3 Relating results to EMH and managerial insights

I only consider linear relationships between the measure of efficiency and effectiveness and the Tobin's Q ratio. Using our proposed statistical method, I find the best linear model, using AIC, and provide the coefficients on the efficiency and effectiveness scores. Suppose the coefficients are either not statistically significant or are negative, in that case, I say the quantitative models used to estimate the efficiency and effectiveness are not consistent with the semi-strong definition of the EMH.

I find the model of Kumar and Gulati (2010) not consistent with the semi-strong definition of EMH for Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017. In the next chapter, Chapter 3, I build a variable selection framework that can be used to address the research question of whether there is a universal quantitative model of efficiency and effectiveness that works for all banks in all geographic regions? This question is important for any firm looking to expand its services into new geographic markets. A firm can employ the same definition of efficiency and effectiveness in the new geographic market when such a universal definition of efficiency and effectiveness exist.

2.6 Conclusion

The financial market identifies firms that are effective ("doing the right things") and efficient ("doing things right"). Firms that "do the right things right", will be valued much higher than other firms as per EMH. This chapter presents a statistical method of goodness that finds the best model using the correlation of financial measures against the efficiency and effectiveness scores. The statistical method may demonstrate its practicality as a tool for analyzing financial measures and the relationship among the operational performance. My statistical method can also be applied to other industries where firms belonging to these industries have financial securities traded in financial markets.

In ongoing research of mine, I utilized our general statistical method to charity organizations in Canada, with the collaboration of an industry partner. The industry partner was interested in knowing whether the current methodology of evaluating efficiency and effectiveness for charity organizations is correlated with improvement in charity ranking. In the charity research project, I am using the charity ranking by MoneySense[4] as SHVCM and then using the general statistical method, I find whether the efficiency and effectiveness scores of charity organizations are correlated and statistically significant with the SHVCM. I am using my general statistical method in the Canadian charity industry to inform the industry partner of the appropriate dimensions for measuring efficiency and effectiveness and whether their current methodology of measuring efficiency and effectiveness has a statistically significant improvement on rankings.

Due to the generality of my statistical method, this approach of validating can be applied to other markets. The exact metric to use as an SHVCM is a function of the sector or the market. It changes from the banking industry where I used the Tobin's Q ratio to charity organizations where, as suggested by our industry partner, I use charity rankings. As mentioned previously in this chapter, the selection of SHVCM is performed by an expert of a certain domain (such as charity or banking or service industry). I see my general statistical method used by this expert in determining which organizations in that domain are efficient and effective. This domain expert can then select and choose the appropriate input values for the general statistical method. As a sector matures and changes, only the inputs of the statistical method need to change and not the method itself.

[4]A sample of the 2018 ranking of charity organizations by MoneySense may be found here: https://www.moneysense.ca/save/financial-planning/2018-charity-100-full-list/.

Chapter 3

Variable selection framework for two-stage DEA models

3.1 Introduction

Incorrect selection of input and output variables when measuring efficiency and effectiveness will lead to inaccurate performance estimates. Muller (2018) provides numerous case studies from medicine, education, and the military where incorrect input and output variable choices skews performance measurement. Muller (2018) suggested that the World Health Organization (WHO) is using the wrong choice of variables when measuring the health-care performance of countries; the United States is ranked 1st in healthcare spending per capita, but ranked 43rd for adult female mortality, 42nd for adult male mortality, and 36th for life expectancy. Muller (2018) has argued that the WHO uses variables for ranking that are in large part outside the medical system. He believes these variables are more influenced by lifestyle and culture choices. For instance, Americans are, on average, more obese than citizens of other nations. America also has more smokers. Americans are also, on average, more likely to be wounded by a gunshot. In short, many of America's health problems are a function of social and cultural factors beyond the medical system.

In education and at universities, is using student course evaluations with predefined questions at the end of the term appropriate for measuring an instructor's performance? Probably not. Many students form a new appreciation for the value of a course and an instructor when they are at a job, internship, or co-op. This new-found appreciation for the course is not measured when course evaluations are conducted at the end of the university semester. I am not aware of any university that asks students to re-evaluate their instructor or their course months after having completed the course. However, in many places (Anderson et al., 2005; McCallum, 1984), student course evaluations are used to decide whether a certain assistant professor is offered tenure or not.

In addition to Muller (2018), Beckett (2009); Economist (2009) have considered the efficiency and effectiveness of railway transport in different countries. Karlaftis (2004) mentions that the need to measure the various aspects of the performance of a transit system has led to the development of many quantitative performance indicators. However, depending upon the specific indicator examined, different conclusions can often be reached regarding performance. Karlaftis (2004) also mentions that, before 1970, urban transit in the U.S. was privately owned; the efficiency of these urban transit systems was measured via economic profits and ridership. However, after 1970, most of the U.S. transit systems were controlled by the government and heavily subsidized. Karlaftis (2004) points out that the variables of efficiency and effectiveness used to measure urban transit systems prior to 1970 cannot be used to measure the efficiency and effectiveness of a modern transit system that the government heavily subsidizes.

An article in the *Harvard Business Review Magazine* (Chew, 1988) mentioned a firm that hired a gifted mathematician to evaluate the efficiency of its various divisions. Much to the firm's dismay, the quantitative model developed by the mathematician suggested that none of the business units were performing efficiently. How was it possible that the firm was generating large profits but had low efficiency? Perhaps the quantitative model of efficiency used inappropriate input and output variables. The underlying issue with these different fields is that there is no consensus on which input and output variables to choose when measuring efficiency and effectiveness.

The lack of consistency in input and output variables for measuring efficiency and effectiveness is also prevalent in the financial industry. For instance, Kumar and Gulati (2010) measured the efficiency and effectiveness of Indian banks. Additionally, Chu and Lim (1998) measured the efficiency but not the effectiveness of Singaporean banks. However, both papers selected different input and output variables when measuring bank efficiency. How does one decide what input and output variables to use when measuring efficiency and effectiveness in the financial industry? The current state of model generation in the financial industry lends itself to this research question: *How can a model of banks' efficiency and effectiveness be quantitatively validated?* Fortunately, for publicly traded banks, financial markets may be viewed as a measure of the wisdom of crowds (Surowiecki, 2005). Financial metrics such as (1) stock price, (2) market capitalization, and (3) Tobin's Q ratio are considered measures of the performance of a firm in the financial markets (Bharadwaj et al., 1999; Lang and Stulz, 1994; Muller, 2018; Wernerfelt and Montgomery, 1988; Yermack, 1997). Tobin's Q is the ratio of the market value of a company's assets as measured by the market value of its outstanding stock and debt divided by the replacement cost of the company's assets (book value) (Jose et al., 1996; Lexicon, 2009). Tobin's Q ratio also controls for firm sizes and equities. While other factors may impact a firm's financial performance, firms deemed efficient and effective by market traders will have better financial performance than a firm that is either inefficient or ineffective. I explored this idea of validating quantitative models of efficiency and effectiveness in the previous chapter. Driven by the need for the banking industry to be more effective and efficient as well as the lack of consistency in input and output variables in two-stage data envelopment analysis (DEA) models of effectiveness and efficiency, this chapter addresses the following research question: **How does one build a variable selection framework for finding two-stage DEA models of efficiency and effectiveness that are consistent with the semi-strong definition of the efficient market hypothesis (EMH) in the financial industry?**

The variable selection framework consists of two major components: (a) The statistical method from Chapter 2 and (b) a genetic search (GS) algorithm. The variable selection framework uses (a) and (b) for finding two-stage DEA models of efficiency and effectiveness that are consistent with the semi-strong definition of the EMH.

I evaluated the performance of three search algorithms: (a) a GS, (b) a surrogate search optimization algorithm (SSO), and (c) a multi-armed bandit algorithm (MABA) when traversing the search space of efficiency and effectiveness for the two-stage DEA models. The search space is characterized by different combinations of input and output variables for efficiency and effectiveness. The GS performed the best among all the search algorithms; the results are presented in Section 3.5 of this chapter. The variable selection framework uses the GS to traverse the search space by continuously generating a neighborhood of potential two-stage DEA models of efficiency and effectiveness. The variable selection framework finds the model deemed best by the statistical method from Chapter 2. The variable selection framework for the two-stage DEA model developed as part of my Ph.D. is a significant contribution to the literature on building two-stage DEA models of efficiency and effectiveness as well as validating whether these models are consistent with the semi-strong EMH. My contribution and how it differs from other work is explained further in Section 3.2 of this dissertation. This framework is not restricted to the banking industry; it can also be used for other industries such as services and retail.

The intended audience for the work done in this chapter are researchers building new models of efficiency and effectiveness. Researchers from other industry sectors can also use the variable selection framework to build models of efficiency and effectiveness. For instance in another ongoing project of mine, I use the variable selection framework to build models of efficiency and effectiveness for non-profit organizations. I use charity rankings as the SHVCM in the non-profit organization. In the remainder of this chapter, I first discuss related work in Section 3.2. In Section 3.4, I present an overview of the three search algorithms. In Section 3.5, I evaluate the three search algorithms and find that the GSA performs the best.

3.2 Related work

Cinca and Molinero (2004); Nunamaker (1985) suggested that adding more variables to a DEA model results in less discrimination power between decision-making units (DMUs). An asymptotic analysis suggests that as the number of variables increases, the efficiency scores for all DMUs approaches 1 and it becomes harder to discriminate between DMUs. The recommended approach is to select as few variables as possible in a DEA model. On the other hand, a guideline commonly applied to variable selection is that there should be at least three times as many DMUs as variables (Friedman and Sinuany-Stern, 1998). These recommendations have been incorporated in the search algorithms which are further discussed in Section 3.3.1 of this dissertation.

Ruggiero (2005) talks about the reliable measurement of efficiency. He does not mention effectiveness. Results can be biased when the wrong input variables are used in a DEA model. Careful selection of an appropriate set of variables is necessary to reliably measure efficiency. He used a simulation analysis to develop a statistical procedure that provides guidelines for input selection for two-stage DEA models. His recommendations incorporate the guidelines of Friedman and Sinuany-Stern (1998). Additionally, Sharma and Yu (2015) regressed efficiency score against the input and output variables of a DEA model. Using this approach, they dropped the input and output variables that were not statistically significant in their impact on the efficiency score. In this dissertation, I use the statistical method from Chapter 2 which allows me to validate any quantitative model of efficiency and effectiveness against the semi-strong definition of the EMH. My statistical method serves as a fitness score in the variable selection framework built in this dissertation. The search algorithms in the variable selection framework traverse the search space of two-stage DEA models of efficiency and effectiveness to find a set of models that are consistent with the semi-strong definition of the EMH. In contrast, Ruggiero (2005)'s statistical procedure finds those input variables in a DEA model that maximize the discrimination of DMUs. It will be interesting to see as future work whether a model that maximizes the discrimination of DMUs is any different from a model that is consistent with the semi-strong definition of the EMH.

Adler and Yazhemsky (2010); Nataraja and Johnson (2011); Sirvent et al. (2005) run

Monte Carlo simulations that compare the two methods. They run two simulations: (1) principal component analysis (PCA) and (2) variable reduction based on partial covariance. However, for this analysis, they assume some true and correct efficiency score. This true and correct efficiency score is a DEA model of efficiency that is most often used by firm managers; there is a general consensus on what the input and output variables must be. Based on this true efficiency score, the researchers perform the discrimination-improving methods and record the lowest number of variables via their variable selection framework for which the discrimination starts to worsen. The new parsimonious model with the lowest number of variables produced by applying the discrimination-improving methods is then an improvement over the original model. This true and correct efficiency score is not available in the banking sector; as mentioned previously in this dissertation, researchers use different two-stage DEA models of efficiency and effectiveness in the banking sector. My contribution in this Ph.D. work is to find whether such a true model exists. This true model is found using the variable selection framework built in this dissertation. A detailed breakdown of these results for different countries and time periods is presented in Chapter 4 and Section 4.4.

Luo et al. (2012) discusses another approach for selecting variables. They mention some of the common problems in variable selection such as the selection tools, correlation analysis, and the misclassification of variables as input or output. Their work is based on cash value added. They use statistical tests to decide which input and output variables to select. To evaluate efficiency, the authors consider the cash flow of DMUs in their variable selection framework. In their method, variables are selected based on their influence on the cash flow of the DMUs. A variable is taken as an output if it has a positive influence on a DMU's cash flow. Otherwise, it is used as an input. Likewise, Nataraja and Johnson (2011) mention three approaches to variable selection in DEA via Monte Carlo simulations. They measure efficiency using PCA, a regression-based test, and bootstrapping. They determine the advantages and disadvantages of each approach. Similarly, Amirteimoori et al. (2014) mention that researchers doing DEA modeling often find that the closeness of the number of operational units and the number of inputs and outputs is problematic. In performance measurement using DEA, the closeness of these two numbers could yield many efficient units. The authors aggregate some input and output variables and iteratively reduce the number of variables. They provide numerical examples and show that, in comparison to the single DEA method, their approach produces the fewest efficient units. According to them, this implies that their approach can better discriminate the performance of DMUs. Luo et al. (2012) uses the cash flow of DMUs as an objective criterion when evaluating DEA models of efficiency; however, there is no mention of what objective criteria to use when evaluating DEA models of effectiveness. In Amirteimoori et al. (2014); Luo et al.

(2012); Nataraja and Johnson (2011), there is an absence of objective criteria that can be used to validate quantitative models of efficiency and effectiveness built using their variable selection framework. My statistical method provides such an objective criterion. It tests any quantitative model of efficiency and effectiveness against the semi-strong definition of the EMH using the financial performance metric of a firm such as stock price, market capitalization or Tobin's Q ratio. In this dissertation, I built a variable selection framework for two-stage DEA models of efficiency and effectiveness that finds input and output variables consistent with the semi-strong definition of the EMH. My contribution will give bank managers a set of input and output variables that they can act on to improve their bank's efficiency and effectiveness. Bank managers will also be able to compare the performance of their bank relative to their competitors. My approach recognizes those who have excelled and helps those who have fallen behind. Also, a bank's shareholders can benefit from these input and output variables because, as the bank's efficiency and effectiveness increases, the bank's financial market performance will improve.

Wagner and Shimshak (2007) developed a formal procedure for a stepwise approach to variable selection that involves sequentially maximizing or minimizing the average change in efficiency as variables are added or dropped from the DEA model. In addition, they discuss how their method yields useful managerial results as well as identifying DEA models that include variables with the largest impact on the results. In my dissertation, my objective is not to find input and output variables of two-stage DEA models that yield the largest change in efficiency and effectiveness scores of banks, but rather to find a two-stage DEA model of efficiency and effectiveness that is consistent with the semi-strong version of the EMH. If the variables that yield the largest efficiency and effectiveness scores are not consistent with the semi-strong EMH, then how would the financial market reward firms (via a higher Tobin's Q ratio) that are doing the right things (effectiveness) and are doing things right (efficiency)?

Daraio and Simar (2007); Jenkins and Anderson (2003); Lee and Choi (2010); Lewin et al. (1982); Sengupta (1990) suggested removing the highly correlated variables that appear within the input or the output selection of DEA to reduce the total number of variables. Others (Adler and Golany, 2001; Serrano-Cinca et al., 2005; Ueda and Hoshiai, 1997) suggested using PCA to reduce the number of variables. These researchers found the eigenvectors that explain 90% of the variance of the original dataset. They use these eigenvectors as a new set of variables for measuring efficiency using the DEA model. I also use PCA to find the eigenvectors that explain 90% of the variance in the banking dataset; however, I find the original variables that lie on the eigenspace or are close to the eigenspace. I use these original variables as the initial starting points of the GS algorithm in the variable selection framework. A detailed breakdown of these results are presented in Chapter 4 and Section 4.3.3. I use this approach to assist my search algorithms by providing a promising starting solution (i.e., it is close to those input and output variables that generate eigenvectors which explain 90% of the variance). This approach of selecting starting points using PCA is performed in addition to using the random starting point generator that I developed which is described in Section 3.3.1 of this dissertation. The random starting point generator provides up to 5,000 different starting points selected randomly from a uniform distribution. However, there are two challenges in using PCA to measure efficiency and effectiveness that have not been addressed by Adler and Golany (2001); Serrano-Cinca et al. (2005); Ueda and Hoshiai (1997): (1) Comparing the eigenspace with the original space, how accurate is the measurement of efficiency and effectiveness and what is the error component? (2) When measuring efficiency and effectiveness in the eigenspace, one loses the original set of variables. How does one use the efficiency and the effectiveness scores computed from the eigenspace to explain this in terms of the original set of variables?

One of the goals of this Ph.D. dissertation is to provide managerial insight. This insight, discussed in Chapter 4 and Section 4.4.6, is provided using the input and output variables that are in the original space and not the eigenspace. I also provide insight to shareholders of banks; this insight also uses variables in the original space rather than the eigenspace. For these reasons, I use the variable selection framework on the original banking data variables and restrict the use of PCA to only providing promising starting points to the search algorithms in the variable selection framework.

Paradi et al. (2011) mention that measuring bank branch performance is a difficult task because bank branches operate in different economic regions and serve different customers. Further, Paradi et al. (2011) mention that performance evaluation of bank branches, both within a country and globally, remains an important research area. Does this mean that there is no universal quantitative model that can measure efficiency and effectiveness? One of the major contributions of this dissertation is that it provides an answer to whether the quantitative models (i.e, two-stage DEA models) of efficiency and effectiveness are independent of geographic location and time period. In my Ph.D. work, I answer this question by traversing the search space of two-stage efficiency and effectiveness DEA models on a subset of banks from Brazil, Canada, China, India, Japan, Mexico, South Korea, and the U.S. and during period of 2000-2017. The detailed breakdown of these results is presented in Chapter 4 and Section 4.4 of this dissertation. How does one choose the input and output variables for measuring efficiency and effectiveness? What are the input and output variables for measuring efficiency and effectiveness? Are the variables independent of space and time? How is one quantitative model of efficiency and effectiveness evaluated against some other quantitative model? These questions highlight the need for the following:

1. A standard that can be used to compare and evaluate different models.

2. A variable selection framework that finds two-stage DEA models of efficiency and effectiveness that are consistent with that standard.

3. A search for a universal two-stage DEA model of efficiency and effectiveness that is independent of space (geographic location) and time (time period).

Different quantitative models of efficiency and effectiveness lead to different performance measures, leaving a firm manager at a loss as to which to use. Stockholders of the firm are also at a loss, as they do not know whether the model of efficiency and effectiveness used by the firm manager improves the firm's performance in the financial markets.

4. Is there any *causal* relationship between (a) efficiency and effectiveness measures from two-stage DEA models of banks and (b) the Tobin's Q ratio of banks?

In the previous chapter, Chapter 2, I addressed the first point from above by building a statistical method that validates the quantitative models of efficiency and effectiveness with the semi-strong definition of EMH. In this chapter, I address the next two points from above that are also presented in detail in Section 3.4. Finally in Chapter 4, I address the last point.

3.3 Variable Selection Framework

I used the GS, SSO and the MABA when traversing the search space of efficiency and effectiveness for two-stage DEA models. The search space is characterized by different combinations of input and output variables of efficiency and effectiveness. The search algorithm traverses the search space of potential two-stage DEA models of efficiency and effectiveness and finds the two-stage DEA model of efficiency and effectiveness deemed best by the statistical method of Chapter 2. In Section 3.3.1, I first describe the formulation of a two-stage DEA model of efficiency and effectiveness as a matrix in the search algorithm. Later in the same section I mention some of the constraints that must be satisfied when creating a two-stage DEA model of efficiency and effectiveness. Finally in Section 3.4, I describe the three search algorithms of (1) GS algorithm, (2) SSO and (3) MABA in greater detail.

3.3.1 Constrained satisfaction problem (CSP)

The search algorithms require an initial population of two-stage DEA models of efficiency and effectiveness on which the search can begin. A two-stage DEA model of efficiency and effectiveness, $Model$, is represented as a matrix of $\mathbb{B}^{3\times 55}$. The three rows of the matrix represents: (1) efficiency input dimensions, (2) efficiency output dimensions and effectiveness input dimensions and (3) effectiveness output dimensions. Each column in the matrix represents a dimension from the Eikon dataset. There are 55 dimensions from Eikon.

In the two-stage DEA model of efficiency and effectiveness that I consider in this dissertation the efficiency output dimensions are identical to the effectiveness input dimensions, see Figure 2.1 for more detail. For instance if row 1 and column j is 1 then that indicates that variable j belongs to the set of efficiency input vectors, if row 2 and column j is 1 then that indicates that variable j belongs to the set of efficiency output vectors and if row 3 and column j is 1 then that indicates that variable j belongs to the set of effectiveness output vectors. Likewise a 0 in any row i and column j indicates that the dimension j is not present in the row i of the two-stage DEA model of efficiency and effectiveness. However here are the linear constraints that must be met:

1. A variable j if it is present in the two-stage DEA model of efficiency and effectiveness then it must occur only in one of three positions of the two-stage DEA model. The three positions are: (1) efficiency input variables, (2) efficiency output variables and (3) effectiveness output variables. Each column in the matrix $Model$ represents a variable. The sum of the entries in each of the column of $Model$ must be less than or equal to 1, i.e., $\sum_{k=1}^{3} Model_{k,i} \leq 1 \; \forall i$.

2. The sum of the entries in each row of $Model$ must be greater than or equal to 2. Mathematically this is represented as $\sum_{i=1}^{55} Model_{k,i} \geq 2 \; \forall k$. I do not wish to reduce the production frontier of calculating efficiency and effectiveness to simple ratios of one input and one output variable. Therefore I require that the minimum number of efficiency input variables or efficiency output variables or effectiveness output variables must be at least 2. In the DEA model of Daraio and Simar (2007) the number of input variable in DEA is exactly one, and the number of output variable in DEA is also exactly one. I suspect that using one input variable and one output variable will yield a single ratio when performing DEA analysis. Yang and Pollitt (2009) has documented that single ratios do not provide reliable results due to banks' complex operational environment.

3. As the number of variables in a single stage DEA increases, the discriminating power across DMU decreases. Golany and Roll (1989) mentions that the number of DMU must be at least two times the number of input and output variables. The two-stage DEA model of efficiency and effectiveness can be viewed as combining two single stage DEA model where the output of the first single stage DEA model are the inputs to the second single stage DEA model. Mathematically to have the number of DMUs atleast two times the sum of efficiency input variables and efficiency output variables is $2 \times \sum_{k=1}^{2} \sum_{i=1}^{55} Model_{k,i} \leq 8$. Note that k goes from 1 to 2 where the entries in row 1 represents the variables present as input of efficiency and row 2 represents the variables present as output of effectiveness. 8 is the number of DMU or the 8 countries that we consider in this dissertation. Repeating this once more, but now for effectiveness, $2 \times \sum_{k=2}^{3} \sum_{i=1}^{55} Model_{k,i} \leq 8$. Note that k goes from 2 to 3 where the entries in row 2 represents the variables present as input of effectiveness and row 3 represents the variables present as output of effectiveness.

4. I generate 5000 two-stage DEA model of efficiency and effectiveness using the dimensions from my principal component analysis (PCA) of Section 4.3.3 when executing the variable selection framework for Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for timeperiod of 2000-2017. These 5000 two-stage DEA model of efficiency and effectiveness serve as initial population size for the GS algorithm to begin the search on. Mathematically this is represented as $\sum_{k=1}^{3} \sum_{j=1}^{J} Model_{k,j} = 0 \ \ \forall J \notin Set_{pca}$ where Set_{pca} is the set of dimensions recommended by the PCA algorithm in Section 4.3.3.

   By summing this to 0, I essentially turn off all those variables that are not recommended by the PCA algorithm. If I wish to not discriminate among any of the 55 dimensions in Eikon, I can simply drop this constraint of (4).

I finally take the four constraints above and run it on IBM's CPlex[1] and generate the desired population size.

[1] https://www.ibm.com/analytics/cplex-optimizer

3.4 Search algorithms

In this section, I describe each of the three search algorithms in greater detail. In Section 3.4.1, I describe the objective function that the three search algorithms minimize. I present the genetic search algorithm in Section 3.4.2. In Section 3.4.3, I present the surrogate search optimization. Finally, in Section 3.4.4, I present the multi-armed bandit search algorithm.

3.4.1 Objective function

For each two-stage DEA model of efficiency and effectiveness that the search algorithm of the variable selection framework visits, the statistical method reports the AIC value if the efficiency and the effectiveness of the model are consistent with the semi-strong version of the EMH. AIC is calculated as $2k - 2\ln(\hat{L})$ where $\hat{L}$ is the likelihood of the model, and k is the number of parameters. Lower AIC values indicate better fits, (i.e., higher likelihood with fewer parameters). AIC is the objective function that the search algorithm minimizes. In Chapter 2, Section 2.4.3, I explain the preference of using AIC over R^2.

The objective function is non-linear. Mixed integer non-linear programming (MINLP) solvers are appropriate to minimize this non-linear function given the constraints of Section 3.3.1. MINLP solvers involving extended cutting plane (Westerlund and Pettersson, 1995), Bonmin (Bonami et al., 2019) solver which uses branch and bound, DICOPT (Grossmann et al., 2002), KNITRO (Waltz and Nocedal, 2004) solver which also uses branch and bound, and SCP (Achterberg, 2009), CPlex (Bliek1ú et al., 2014) either requires the function that is getting optimized is convex or the function is twice differentiable. In a convex problem, any local solution from these solvers is guaranteed to be an optimum solution.

In a non convex problem such as involving the AIC objective function there may be multiple local solutions neither of which are optimum. What is required then are algorithms that are derivative free and their termination criteria is not based on some gradient or stationary points. Also what is desired then is an algorithm that favor global optimum solution rather than local optimum solution for a non convex problem. The genetic search, surrogate search optimization and multi-armed bandit on the other hand are derivative free algorithms that do not impose any constraints on the objective function.

To use AIC correctly, the number of data points in the panel cannot change. However, the inclusion or exclusion of independent variables, including lags of efficiency and effectiveness, changes the number of data points. For instance, adding a lag variable will result in the loss of an entire time period across all the cross-sectional units. Comparing two linear models in which one contains lags and the other does not contain any lags will result in incorrect comparison.

In this dissertation, before using AIC for any comparison, I first preprocess the panel data by including two lags of efficiency and effectiveness. More detail on this is presented in Chapter 4 and Section 4.4.1 of this dissertation. Including two lags as part of preprocessing, results in panel data with 16 time periods instead of the original 18 time periods. It is now correct to use AIC when comparing linear models using the preprocessed panel data of 16 time periods that include up to a maximum of two lags of efficiency and effectiveness. As the search algorithm traverses different two-stage DEA models of efficiency and effectiveness, the efficiency and the effectiveness scores generated by these two-stage DEA models are preprocessed to contain a maximum of two lags. The statistical method uses the efficiency and the effectiveness scores of these different models, including their lags and regresses, against a common SHVCM of Tobin's Q ratio. The AIC score from the statistical method is the objective function minimized by the search algorithms.

The Bayesian information criterion (BIC), like AIC, is a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. BIC is defined as $-2ln(\hat{L}) + kln(n)$ where $\hat{L}$ is the maximized value of the likelihood function, and k is the number of parameters estimated by the model, and n is the size of the data. Unlike AIC, BIC heavily penalizes models that contain more parameters. The drawback of BIC is that it may result in serious under-fitting when compared to AIC. Asymptotically, BIC is consistent in that it will select the true model if the true model is among the set of candidate models. The data collected in my dissertation is non-experimental observational data where the underlying process that connects SHVCM with the efficiency and the effective measures is not well understood. Therefore, I do not know whether the true model is present in the family of models that I consider.[2] Berger et al. (2003); Vrieze (2012) mention that when the true model is not in the candidate models, then AIC is preferred. For these reasons, I use the AIC as the objective function and not the BIC.

3.4.2 Genetic search (GS) algorithm

GS algorithms are adaptive metaheuristic search algorithms classified as an evolutionary computing algorithm. They use techniques inspired by natural evolution (Hassanat et al., 2019). Holland developed the first GS in 1975 to solve optimization problems based on

[2]As mentioned previously in Chapter 1, I restrict myself to linear models to be consistent with the work of other researchers.

biological, genetic, and evolutionary ideas (Holland, 1975). In the algorithm, Algorithm 2, the preconditions are as follows: (1) financial data of banks across 8 countries and 17 time periods as per Section 2.4.1, (2) Tobin's Q ratio data, (3) the CSP of Section 3.3.1 and (4) the length of time to run the algorithm for. The best two-stage DEA model of efficiency and effectiveness is returned in the postcondition of the algorithm.

Each individual (commonly referred to as a chromosome in the jargon of GS algorithms) in the population is a two-stage DEA model of efficiency and effectiveness. These models are generated by the CSP of Section 3.3.1 (see line 10 of Algorithm 2). The initial size of the population depends on the nature of the problem. For instance in Section 3.5.1 when working with 6 dimensions from Eikon, I experimented with different population number. Population size of 30 resulted in not only the optimal AIC but by the 14th generation the entire population had converged to the optimal solution. A larger population size will result in increased diversity among the population but also require significantly more computation resources per generation. With a larger population size, I observed[3] that it takes more number of generations for the entire population in GS to converge to the optimal solution. I used a similar approach later in Chapter 4 and Section 4.4 by trying different population sizes and then selecting the initial population to 5000 when running the GS algorithm on all 55 dimensions from Eikon.

The fitness function on line 19, calculates the population's fitness by generating the efficiency and the effectiveness scores of each two-stage DEA model in the population, then checking whether the log transformation of efficiency and effectiveness scores are stationary or not. I use the same approach as mentioned in Section 2.5.1 of Chapter 2 when checking for unit root in panel data. Suppose the log transformation of efficiency and effectiveness scores are stationary. In this case, the statistical method of Chapter 2 is executed. Then, it checks whether the two-stage DEA model of efficiency and effectiveness is consistent with the semi-strong definition of the EMH and reports the AIC value.

On line 14, the selection function selects the "parent" models from which the next generation of "offspring" models will be created. In Section 3.4.2, I use the stochastic uniform option[4] for selection of parents in the GS. On line 15, crossover produces new offspring. Crossover[5] selects a random point on the chromosome and exchanges parts

[3] I noticed this behavior when selecting first 7 or first 8 or first 9 dimensions from Eikon

[4] Stochastic uniform lays out a line in which each parent corresponds to a section of the line of length proportional to its normalized fitness value. The algorithm moves along the line in steps of equal size. At each step, the algorithm allocates a parent from the section it lands on. The first step is a uniform random number less than the step size.

[5] I use the crossoverintermediate function in MATLAB to generate cross over. The usage of crossover-intermediate is recommended when there are linear constraints.

of the parents between them to create an offspring. The crossover fraction controls the fraction of the population that the crossover function uses to create the next generation, not including elite children.

DeJong (1975) recommends a mutation rate of 0.001 for a population with 50 to 100 individuals. Abolghasemi Dehaghani (2020) on the other hand, sets the mutation rate to 0.2 in his dissertation when running GS. Abolghasemi Dehaghani (2020) mentions that a high mutation rate ensures diversity among the population, preventing the algorithm from getting stuck in a local optimum. Instead of fixing the mutation rate to some value that remains constant across generations, I control the rate at which the average amount of mutation decreases from the initial generation to the final generation. I set the mutation rate to 0.2 in the first generation and 0.001 in the final generation. The mutation rate decreases linearly from the first generation to the last generation. By setting a higher mutation rate in the earlier generations, I encourage diversity in the population so that the search does not get stuck in any local optimum. The lower mutation rate in the later generations allows the search algorithm to converge, producing a solution.

On line 17, the top 5% of the population based on fitness value survive until the next generation. I set the elite value to the top 5% of the population. Mishra and Shukla (2017) mention that using high values for the elite count will prevent diversity, causing the algorithm to get trapped in local minima. On the other hand, with low values for the elite count such as the top 5% of the entire population, there is a higher likelihood of obtaining solutions near the global optimum.

Finally as part of the post condition, the GS algorithm reports the two-stage DEA model of efficiency and effectiveness with the lowest AIC and the most consistent with the semi-strong version of the EMH.

Algorithm 2 Genetic Algorithm

Preconditions:

- Bank financial data: $Data \in \mathbb{R}^{N \times D \times T}$
  N is the number of countries and T is the total number of time periods and D is the number of data dimensions.

- Data from Tobin's Q ratio: $\mathbf{Y_{TobinQRatio}} \in \mathbb{R}^{N \times T}$
  N is the number of countries and T is the total number of time periods.

- CSP of Chapter 3 and Section 3.3.1: Linear equality and linear inequality constraints of the form $\mathbf{A_{eq}x} = \mathbf{b_{eq}}$ and $\mathbf{A_{ineq}x} \leq \mathbf{b_{ineq}}$ where $\mathbf{x} \in \mathbb{B}^{3 \times D}$
  and $A_{eq} \in \mathbb{B}^{m_1 \times (3 \times D)}$ and $b_{eq} \in \mathbb{I}^{m_1 \times 1}$ and $A_{ineq} \in \mathbb{B}^{m_2 \times (3 \times D)}$ and $b_{ineq} \in \mathbb{I}^{m_2 \times 1}$ where m_1 is the number of linear equality constraints and m_2 is the number of linear inequality constraints.

- Time: The length of time to run the algorithm for.

Postconditions:

- Best Solution: Best solution found from the statistical method of Chapter 2

```
procedure GeneticAlgorithm(Data, Y_TobinQRatio, CSP, time)
    startTime ←getCurrentTime()
    bestSolution ← empty
    population ← empty
    for each generation in ∞ do
        endTime ←getCurrentTime()
        if endTime − startTime > time OR noChangeInBestSolutionAcrossNgenerations(N) then
            return bestSolution, functionValues, population
        if generation=0 then
            startingPoints ← GenerateRandomStartingPoints(CSP)
            population ← startingPoints
    ⋮
```

Genetic Algorithm cont...

12: $functionValues, population$ ←FitnessFunction($Data, Y_{TobinQRatio}, population$)
13: $bestSolution$ ←checkWhetherBestSolutionRequiresUpdate($functionValues$)
14: $newpopulation$ ←performSelectionOnPopulation($population$)
15: $newpopulation$ ←updatePopulationWithCrossOver($newpopulation, CSP$)
16: $newpopulation$ ←updatePopulationWithMutation($newpopulation, CSP$)
17: $newpopulation$ ←updatePopulationWithEliteism($newpopulation, CSP$)
18: $population \leftarrow newpopulation$

Genetic Algorithm cont...

19: **procedure** FitnessFunction($Data, Y_{TobinQRatio}, population$)
20: **for each** *individual in population* **do**
21: $(X_{efficiency}, X_{effectiveness}) \leftarrow$ generateEfficiencyAndEffectiveness($individual, Data$)
22: **if** isNotStationary($\log(X_{efficiency}), \log(X_{effectiveness})$) **then** continue
23: $aicValue \leftarrow$ statisticalMethod($\log(X_{efficiency}), \log(X_{effectiveness}), \log(Y_{TobinQRatio})$)
24: UpdateBestSolutionInCurrentGeneration($aicValue, individual$)
25: **return** $aicValues, population$

3.4.3 Surrogate search optimization (SSO)

The "surrogate" in surrogate search optimization is a function that approximates the fitness function (also referred to as the objective function). The fitness function is shown in Algorithm 3 as the function `FitnessFunction`. The surrogate is computationally less expensive. For example, to search for a point that minimizes the fitness function, evaluate the surrogate on thousands of points and take the best value as an approximation to the minimizer of the fitness function (MATLAB, 2020). The preconditions and postcondition of the algorithm are almost identical to the preconditions and postcondition of the GS of Section 3.4.2 except that in pre conditions I do not set the time for how long to run the SSO.

The SSO alternates between two phases: (1) the construct surrogate phase, and (2) the search for the minimum. In the construct surrogate phase, as seen on line 3 of the algorithm, an initial population of two-stage DEA models of efficiency and effectiveness is generated that meets the CSP of Section 3.3.1. The fitness function is evaluated for the entire population. The fitness function of surrogate search optimization is very similar to the fitness function of the GS of Section 3.4.2. The fitness function on line 7 of the algorithm is computationally more expensive to evaluate when compared to evaluating the surrogate function on line 12. After computing the fitness function from the initial population, a surrogate function is created (see lines 9 and 10 of Algorithm 3) by interpolating a radial basis function through the fitness value of the initial population.

To search for the minimum of the objective function, several thousand random points are generated around the minimum of the surrogate function (see line 11 of Algorithm 3). These points are then evaluated on the surrogate function; the lowest among them is called the adaptive point (see line 12 of Algorithm 3). Then, the adaptive point is evaluated on the fitness function. The adaptive point is checked to see whether it results in a better solution than previously found from the population (see line 14 of Algorithm 3). When the surrogate optimization resets (i.e., the loop in line 6 of Algorithm 3 runs again) a newer random population is generated (see line 15). The randomness prevents the surrogate search optimization from getting stuck in any local optimum. In Section 3.5.2, I note that the surrogate search optimization performs significantly worse than the genetic search algorithm. This is because the CSP of Section 3.3.1 is introduced in the fitness function of the surrogate search optimization. This causes many individuals to be rejected; in the GS, all individuals in the population are feasible because they are generated using the CSP in every generation.

Finally as part of the post condition, the SSO algorithm reports the two-stage DEA model of efficiency and effectiveness with the lowest AIC and the most consistent with the semi-strong version of the EMH.

Algorithm 3 Surrogate Search Optimization

Preconditions:

- Bank financial data: $Data \in \mathbb{R}^{N \times D \times T}$
  N is the number of countries and T is the total number of time periods and D is the number of data dimensions.

- Data from Tobin's Q ratio: $\mathbf{Y_{TobinQRatio}} \in \mathbb{R}^{N \times T}$
  N is the number of countries and T is the total number of time periods.

- CSP of Chapter 3 and Section 3.3.1: Linear equality and linear inequality constraints of the form $\mathbf{A_{eq}x} = \mathbf{b_{eq}}$ and $\mathbf{A_{ineq}x} \leq \mathbf{b_{ineq}}$ where $\mathbf{x} \in \mathbb{B}^{3 \times D}$ and $A_{eq} \in \mathbb{B}^{m_1 \times (3 \times D)}$ and $b_{eq} \in \mathbb{I}^{m_1 \times 1}$ and $A_{ineq} \in \mathbb{B}^{m_2 \times (3 \times D)}$ and $b_{ineq} \in \mathbb{I}^{m_2 \times 1}$ where m_1 is the number of linear equality constraints and m_2 is the number of linear inequality constraints.

Postconditions:

- Best Solution: Best solution found from the statistical method of Chapter 2

```
procedure SurrogateSearchOptimization(Data, CSP, Y_TobinQRatio)
    bestSolution ← empty
    population ← GenerateRandomStartingPoints(CSP)
    randomPoints ←GenerateRandomStartingPoints(minimumNumberOfSurrogatePoints-
size(population))
    population ←population+randomPoints
    for each resetNumber in totalNumberOfResets do
        functionValues ←FitnessFunction(Data, Y_TobinQRatio, population)
        bestSolution ←checkWhetherBestSolutionRequiresUpdate(functionValues)
        surrogateFunction ←createSurrogateFunction(functionValues, population)
        surrogateFunction ←updateSurrogateFunctionWithRadialBiasFunction(
            surrogateFunction)
        samplePoints ←generateManySampleAroundMinimumUsingMeritFunction(bestSolution)
    ⋮
```

Surrogate Search Optimization cont...

$dea_{adaptive}, f_{adaptive}$ ←getMinimumOfSamplePoints($samplePoints$,$surrogateFunction$)
$functionValueOfAdaptivePoint$ ←FitnessFunction($Data, Y_{TobinQRatio}, dea_{adaptive}$)
surrogateFunction←updateSurrogateFunction($dea_{adaptive}, functionValueOfAdaptivePoint$)
$bestSolution$ ←checkWhetherBestSolutionRequiresUpdate($functionValueOfAdaptivePoint$
$population$ ←GenerateRandomStartingPoints(minimumNumberOfSurrogatePoints)
return $bestSolution$

procedure FITNESSFUNCTION($CSP, Data, Y_{TobinQRatio}, population$)
for each *individual in population* **do**
if individualNOTMeetsCSP($individual, CSP$) **then**
continue
$(X_{efficiency}, X_{effectiveness}) \leftarrow$ generateEfficiencyAndEffectiveness($individual, Data$)
if isNotStationary($\log(X_{efficiency}), \log(X_{effectiveness})$) **then**
continue
$fitnessFunctionValue \leftarrow$ statisticalMethod($\log(X_{efficiency}), \log(X_{effectiveness}), \log(Y_{TobinQRat}$
UpdateBestSolution($fitnessFunctionValue, individual$)
return $functionValues, population$

3.4.4 Multi-Armed Bandit algorithm (MABA)

The MABA has two distinct phases: *exploration* and *exploitation*. I summarize both these stages of the algorithm below. The purpose of the exploration stage is to generate a belief system that will help find variables of the two-stage DEA model that will likely be consistent with the semi-strong definition of the EMH. Specifically, MABA's belief system keeps count of the two-stage DEA model variables that generated a consistent model with the semi-strong definition of the EMH. Conversely, the algorithm also keeps count of the variables that generated a model that was not consistent with the semi-strong definition of the EMH. The belief system from the exploration stage is later used by the exploitation stage in generating a two-stage DEA model. Using the belief system, the exploitation stage generates two-stage DEA models such that top k frequently occurring dimensions that results in the statistical significant model (as per the semi-strong version of the EMH) are always present in the two-stage DEA models generated by the exploitation stage. In addition, top z frequently occurring dimensions that results in the non-significant statistical model (as per the semi-strong version of the EMH) are always absent in the two-stage DEA models generated by the exploitation stage. By setting a probability, ϵ, the MABA can be in the exploration stage ϵ probability and with $1 - \epsilon$ probability in the exploitation stage.

Exploration

The preconditions of the exploration stage are identical to GS and SSO with the added inclusion of a starting point on which the algorithm will initially begun. The financial bank data and the Tobin's Q ratio are further explained in Section 2.4.1 of this paper. The exploration stage first generates a neighborhood of two-stage DEA models for the given starting solution of a certain two-stage DEA model. See line 4 of the Algorithm 4. For each of the two-stage DEA model in the neighborhood, the efficiency and the effectiveness measures are computed. The statistical method is then executed on the efficiency and the effectiveness measures along with the Tobin's Q data. The statistical method reports back whether the efficiency and the effectiveness scores are consistent with the semi-strong definition of the EMH along with the AIC value.

The belief system comprises of two matrices. The first matrix is a matrix $belief_{SS}$ of size $\mathbb{Z}^{3\times 55}$. This matrix keeps track of how many of the 55 dimensions and in what of the three orientations did they occur that caused them to be statistically significant. The second matrix is a matrix $belief_{NSS}$ of size $\mathbb{Z}^{3\times 55}$. This matrix keeps track of how many of the 55 dimensions and in what of the three orientations did they occur that caused them to not be statistically significant.

Algorithm 4 Algorithm of exploration in multi-armed bandit algorithm

Preconditions:

- Bank financial data: $Data \in \mathbb{R}^{N \times D \times T}$
  N is the number of countries and T is the total number of time periods and D is the number of data dimensions.

- Data from Tobin's Q ratio: $\mathbf{Y_{TobinQRatio}} \in \mathbb{R}^{N \times T}$
  N is the number of countries and T is the total number of time periods.

- CSP of Chapter 3 and Section 3.3.1: Linear equality and linear inequality constraints of the form $\mathbf{A_{eq}x} = \mathbf{b_{eq}}$ and $\mathbf{A_{ineq}x} \leq \mathbf{b_{ineq}}$ where $\mathbf{x} \in \mathbb{B}^{3 \times D}$
  and $A_{eq} \in \mathbb{B}^{m_1 \times (3 \times D)}$ and $b_{eq} \in \mathbb{I}^{m_1 \times 1}$ and $A_{ineq} \in \mathbb{B}^{m_2 \times (3 \times D)}$ and $b_{ineq} \in \mathbb{I}^{m_2 \times 1}$ where m_1 is the number of linear equality constraints and m_2 is the number of linear inequality constraints.

- Starting Solution: $StartingSolution \in \mathbb{B}^{3 \times D}$

Postconditions:

- Belief System: Belief System that can be used later for exploitation

```
procedure EXPLORATION(Data, CSP, StartingSolution, Y_TobinQRatio)
    Queue.enqueue(StartingSolution)
    tabuList←Nil
    for each potentialStartingPointToExplore in Queue do
        neighborhood ← GenerateNeighborhood(potentialStartingPointToExplore,CSP)
    ⋮
```

Exploration cont...

for each *neighbor in neighborhood* $\wedge$ *neighbor not in tabuList* **do**
    $(X_{efficiency}, X_{effectiveness}) \leftarrow$ generateEfficiencyAndEffectiveness(neighbor,Data)
    if isNotStationary($\log(X_{efficiency}), \log(X_{effectiveness})$) **then** continue
    $Model \leftarrow$ statisticalMethod($(X_{efficiency}, X_{effectiveness}), Y_{TobinQRatio}$)
    UpdateBeliefSystem($Model$)
    Queue.enqueue(neighbor)
    tabuList.enqueue(StartingSolution)

return $getCurrentBeliefSystem()$

Exploitation

The preconditions of the exploitation stage are the bank financial data, data from the Tobin's Q ratio and the belief system from the exploration stage of the multi-armed bandit algorithm.

The belief system comprises of two matrices. The first matrix is a matrix $belief_{SS}$ of size $\mathbb{Z}^{3\times55}$. This matrix keeps track of how many of the 55 dimensions and in what of the three orientations did they occur that caused them to be statistically significant. The second matrix is a matrix $belief_{NSS}$ of size $\mathbb{Z}^{3\times55}$. This matrix keeps track of how many of the 55 dimensions and in what of the three orientations did they occur that caused them to be not statistically significant.

The exploitation stage uses the belief system and generates two-stage DEA models such that the top k variables from $belief_{SS}$ are always present in the two-stage DEA model. I also use z such that top z variables from $belief_{NSS}$ are never in the two-stage DEA model. I then use this two-stage DEA model and generate an entire neighborhood that the exploitation stage of the multi-armed bandit algorithm can exploit. This makes the exploitation happen in the vicinity of those top k variables always present and z variables not present. In case of any dimension j that is present in the top k and the top z for the same orientation in $belief_{NSS}$ and $belief_{SS}$, I take the maximum of how many times does j occur in $belief_{SS}$ and $belief_{NSS}$ and then decide whether to use j or drop j in exploitation. The exploitation returns back the best model if any that exists from the belief system.

Algorithm of exploitation in multi-armed bandit algorithm

Preconditions:

- Bank financial data: $Data \in \mathbb{R}^{N \times D \times T}$
  N is the number of countries and T is the total number of time periods and D is the number of data dimensions.

- Data from Tobin's Q ratio: $\mathbf{Y_{TobinQRatio}} \in \mathbb{R}^{N \times T}$
  N is the number of countries and T is the total number of time periods.

- CSP of Chapter 3 and Section 3.3.1: Linear equality and linear inequality constraints of the form $\mathbf{A_{eq}x} = \mathbf{b_{eq}}$ and $\mathbf{A_{ineq}x} \leq \mathbf{b_{ineq}}$ where $\mathbf{x} \in \mathbb{B}^{3 \times D}$
  and $A_{eq} \in \mathbb{B}^{m_1 \times (3 \times D)}$ and $b_{eq} \in \mathbb{I}^{m_1 \times 1}$ and $A_{ineq} \in \mathbb{B}^{m_2 \times (3 \times D)}$ and $b_{ineq} \in \mathbb{I}^{m_2 \times 1}$ where m_1 is the number of linear equality constraints and m_2 is the number of linear inequality constraints.

- Current Belief System: $BeliefSystem$

- k: $k \in \mathbb{Z}^+$
  Top K dimensions for exploitation to exploit by always turning them on when generating a solution.

- z: $z \in \mathbb{Z}^+$
  Top Z dimensions for exploitation to exploit by always turning them off when generating a solution.

Postconditions:

- Best Solution: Best solution found from the statistical method of Chapter 2

```
procedure EXPLOITATION(Data, Y_TobinQRatio, BeliefSystem, k, z, CSP)
    solution ← FromBeliefSystemSetTopKAndTurnOffWorstZ(BeliefSystem, k, z)
    neighborhood ← GenerateNeighborhood(solution,CSP)
```

Exploitation cont...

for each *neighbor in neighborhood* **do**
    $(X_{efficiency}, X_{effectiveness}) \leftarrow$ generateEfficiencyAndEffectiveness(neighbor,Data)
    if isNotStationary$(\log(X_{efficiency}), \log(X_{effectiveness}))$ **then** continue
    $Model \leftarrow$ statisticalMethod$((X_{efficiency}, X_{effectiveness}), Y_{TobinQRatio})$
    UpdateBeliefSystem($Model$)
    UpdateBestSolution($Model$)
return $getBestSolution()$

3.5 Results

In this section, I present the results of executing the GS and the SSO algorithm on the Eikon dataset from Chapter 2 and Section 2.4.1. The GS and the SSO algorithms are preferred over the MABA because MATLAB offers support for vectorized functions and parallel processing. The GS and the SSO usually run faster when the fitness function is vectorized. Function vectorization implies that the search algorithm only calls the fitness function once. However, the fitness function computes the fitness for all individuals in the current population at once. The parallel computing toolbox in MATLAB offers support for running the GS and the SSO in parallel. This addition significantly speeds up the number of two-stage DEA models of efficiency and effectiveness evaluated per unit of time. My implementation of a MABA evaluates a lower number of two-stage DEA models of efficiency and effectiveness per unit of time compared to GS and SSO. On average, it is about five to eight times slower. For this reason, I discard the MABA in favor of the GS and the SSO.

3.5.1 Results of GS algorithm

I first execute the GS and the SSO by selecting the first 6 dimensions of the 55 dimensions from Eikon. The selection of 6 dimensions yields a total of $\frac{6!}{2^3} = 90$ two-stage DEA models of efficiency and effectiveness. I enumerated across all these 90 models and found the best two-stage DEA model of efficiency and effectiveness from these 90 models. I executed the statistical method once each across the 90 models in order to know which is the best model. Since there are only 90 models, it is not computationally expensive to enumerate each of them and find the best two-stage DEA model of efficiency and effectiveness. I find the best two-stage DEA model of efficiency and effectiveness with an AIC of -46.5799 after enumerating across all the 90 models. Knowing upfront what the best two-stage DEA model of efficiency and effectiveness is, I can now determine whether the GS converges to this model or not. I can also determine how long convergence takes if it does converge.

In Section 3.4.2 I have mentioned how the model parameters of GS algorithm are set. Figure 3.1 shows the result of running the GS algorithm on 6 dimensions. On the same figure, the bottom plot shows that the best AIC of -46.5799 is achieved in the first generation. The second plot in the same figure shows that by the 7th generation, most of the population with 30 individuals has converged to the best two-stage DEA model of efficiency and effectiveness. The range between the best and the worst AIC in the population in the 7th generation is small. The first plot in the same figure shows the average distance in the population. In the earlier generations the average distance is high among individuals indicating high diversity. As the algorithm converges to the optimal solution, the average distance reduces to 0 by the 14th generation. This suggests that all the individuals in the population (i.e., all the two-stage DEA models of efficiency and effectiveness in the population) have converged to a single model. The average value of the AIC in the 14th generation is 46.5799; this is the same as the best AIC value. The GS algorithm executed for 15 generations with a population size of 30. 411 two-stage DEA models were evaluated by the GS algorithm in about 85.089 seconds.

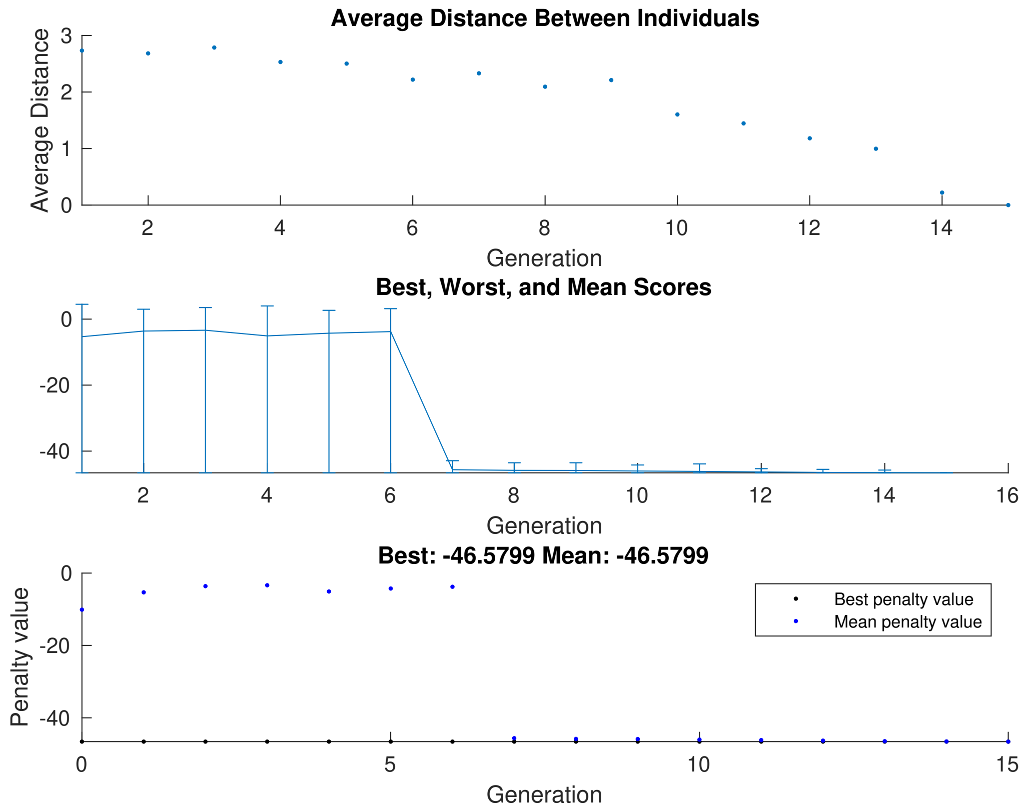

Figure 3.1: Genetic Algorithm run on 6 variables with population size set to 30

I then execute the GS algorithm by selecting the first 7 dimensions of the 55 dimensions from Eikon. The selection of 7 dimensions yields a total of $\frac{7!}{2^3} = 630$ two-stage DEA models of efficiency and effectiveness. Again, it is trivially easy to enumerate all the 630 two-stage DEA models of efficiency and effectiveness and find the best model among these 630 two-stage DEA models. The best two-stage DEA model has an AIC of -55.7257. Figure 3.2 shows the result of running the GS algorithm on 7 variables. On the same figure, the third plot at the bottom shows that the best AIC of -55.7257 is achieved in the first generation. The second plot in the same figure shows that by the 20th generation, the average AIC of the entire population has decreased to -50.7297. The first plot in the same figure shows the average distance in the population. The average distance is high in the earlier generation (close to 3) indicating high diversity. However, the average distance decreases. By the 20th generation, the average distance in the population is close to zero, suggesting that most two-stage DEA models of efficiency and effectiveness in the population have converged to a single model. The GS algorithm executed for 20 generations, with a population size of 200. 4106 two-stage DEA models were evaluated by the GS algorithm in about 728.357 seconds.

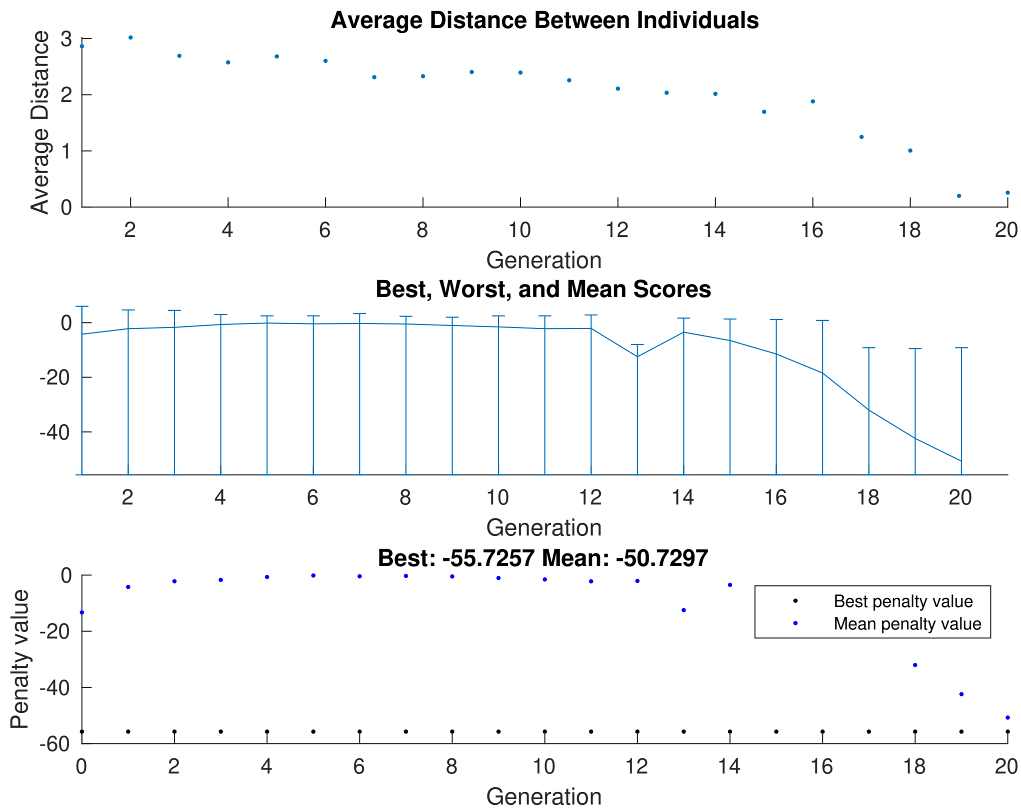

Figure 3.2: Genetic Algorithm run on 7 variables with population size set to 200

I finally execute the GS algorithm by selecting the first 8 dimensions of the 55 dimensions from Eikon. The selection of 8 dimensions yields a total of $\frac{8!/2}{2^3} = 2520$ two-stage DEA models of efficiency and effectiveness. Again, it is trivially easy to enumerate all the 2520 two-stage DEA models of efficiency and effectiveness and find the best model. I find that the best two-stage DEA model contains an AIC of -55.7257. Figure 3.3 shows the result of running the GS algorithm on 8 variables. On the same figure, the third plot at the bottom shows that the best AIC of −55.7257 is achieved in the first generation. The second plot in the same figure shows that by the 39th generation, the average AIC of the entire population has decreased to −41.9562. The first plot in the same figure shows the average distance in the population. The average distance is high in the earlier generation (close to 3) indicating high diversity. However, the average distance decreases. By the 39th generation, the average distance in the population is zero suggesting that most two-stage DEA models of efficiency and effectiveness in the population have converged to a single model. The GS algorithm executed for 40 generations, with a population size of 600. 24406 two-stage DEA models were evaluated by the GS algorithm in about 3874.144 seconds.

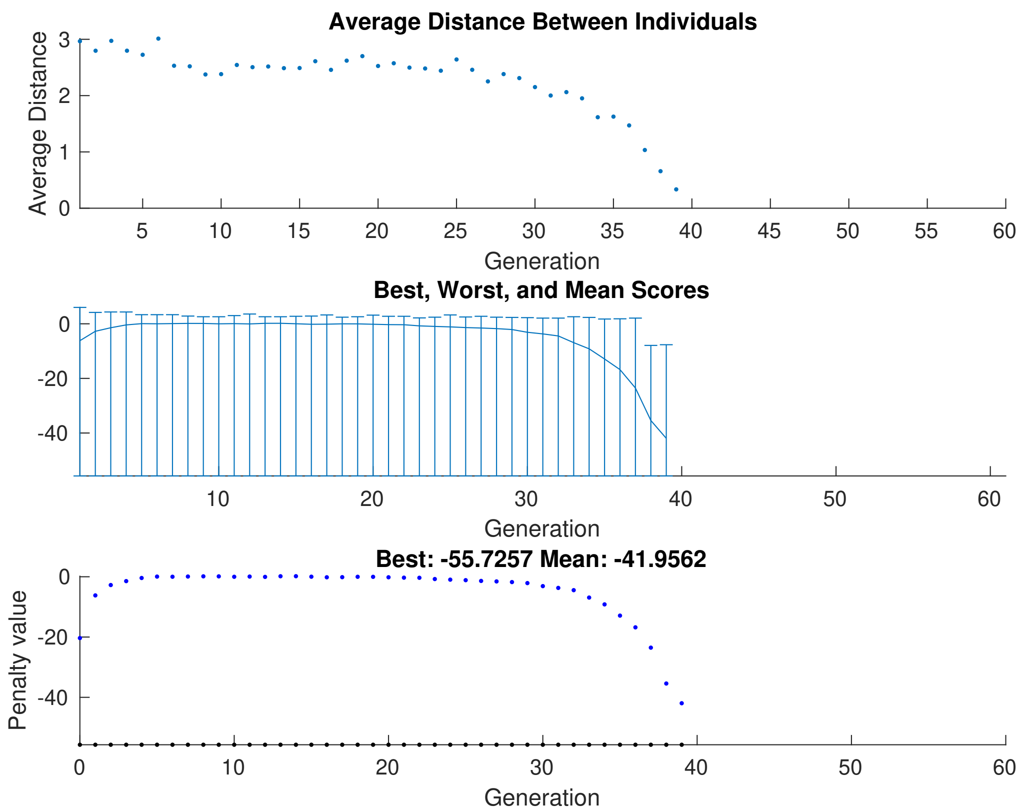

Figure 3.3: Genetic Algorithm run on 8 variables with population size set to 600

However, with every increase in the number of dimensions, the number of two-stage DEA models of efficiency and effectiveness increases combinatorially. I use the empirical evidence collected from Figures 3.1, 3.2 and 3.3 and compare the performance of the GS algorithm with the SSO in the next section.

3.5.2 Results of SSO algorithm

I executed the GS algorithm on 6, 7, and 8 dimensions from the Eikon dataset of Chapter 2 and Section 2.4.1. I know from Section 3.5.1 that the best two-stage DEA model has an AIC of -46.5799 when selecting the first 6 of the 55 dimensions. I repeat this process now for SSO. I set the minimum surrogate points to 110 and the maximum number of function evaluations to 250. The population is set to 30. The result of running the surrogate search optimization on 6 dimensions from Eikon is shown in Figure 3.4. The best AIC value found from running the surrogate search optimization is -46.5319, which is not the optimal value of -46.5799.

Unlike the GS algorithm, I implemented the CSP of Section 3.3.1 into the fitness function of the surrogate optimization. MATLAB[6] does not allow for any user specified constraints in the setup of the problem for SSO. I overrode this by including the user specific constraints in the fitness function of SSO. For any unfeasible point generated by the algorithm, the fitness function penalizes it with a penalty of 70. The infeasible points are shown in red in Figure 3.4.

Each two-stage DEA model of efficiency and effectiveness is represented as a matrix of $\mathbb{B}^{3\times 6}$. Where 6 again is the number of dimensions selected from Eikon. There are a total of $2^6 \times 2^6 \times 2^6 = 262144$ two-stage DEA model of efficiency and effectiveness. However, with the CSP of Section 3.3.1, only 90 of these models are feasible. The rest of the models are infeasible. The initial population of size 30 models for SSO is generated from the CSP of Section 3.3.1. These initial points are represented as pink points in the Figure 3.4. Due to the penalization imposed inside the fitness function of the SSO, all the models after the initial population are deemed as infeasible and rejected. In the worst case there is a probability of only $\frac{90}{262144}$ for SSO to generate a feasible point. The performance of SSO greatly suffers when compared to the GS algorithm. An inordinate amount of time may pass before SSO evaluates a feasible point.

[6] MATLAB R2020A

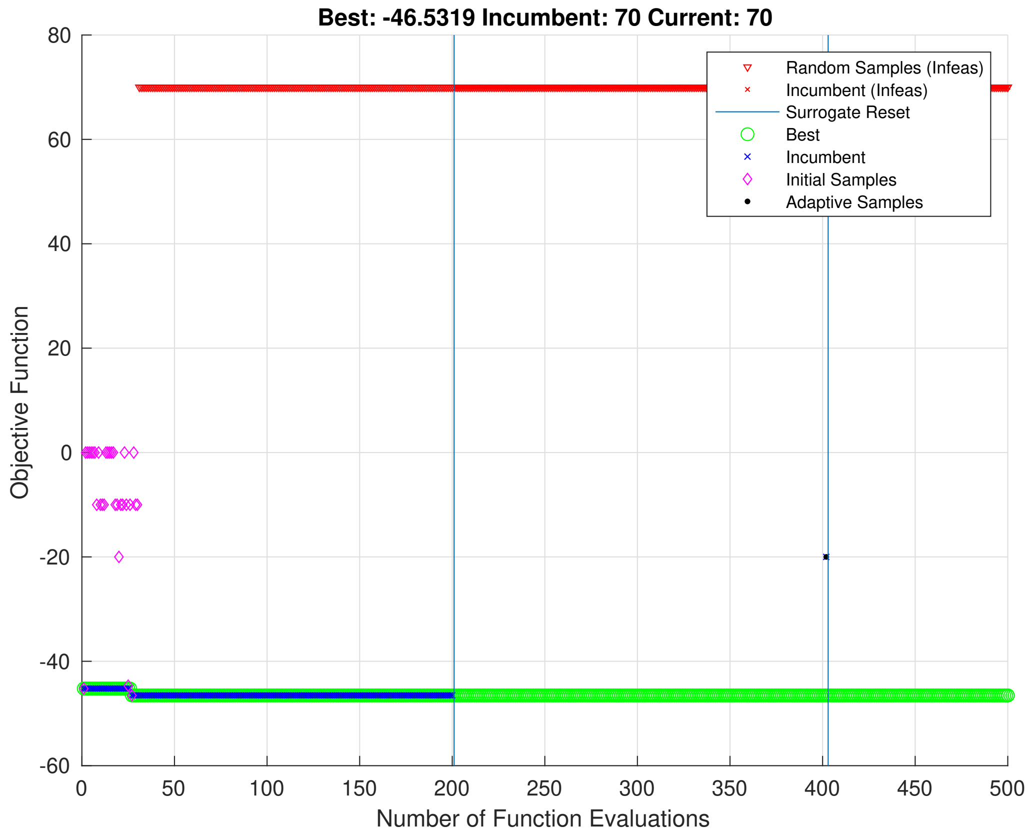

Figure 3.4: Surrogate Optimization run on 6 variables with population size set to 30, minimum surrogate points set to 200 and maximum number function evaluation set to 500.

I now increase my initial population size from 30 models to 90 models. All the 90 models are generated from the CSP of Section 3.3.1 and the result is shown in Figure 3.5. As expected, I see the best AIC value of -46.5799. By increasing the size of the population to 90, which now includes all the feasible models, I am not utilizing any benefits of SSO. The algorithm has degraded into an enumeration of all the 90 feasible points.

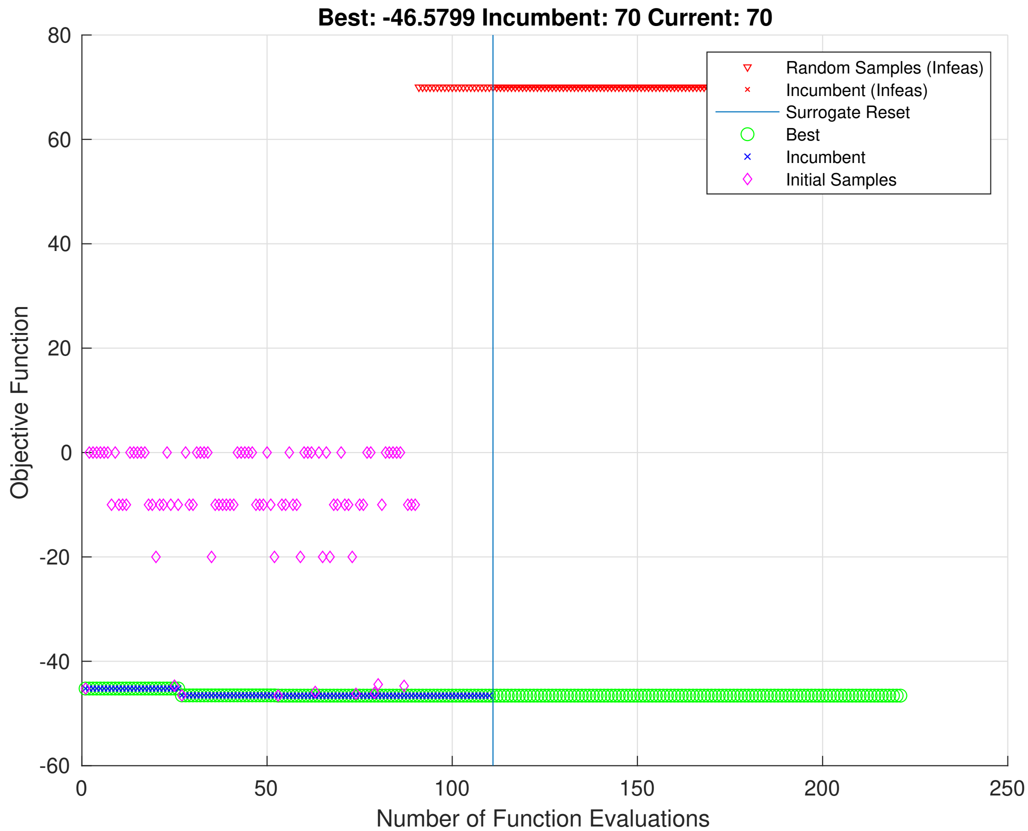

Figure 3.5: Surrogate Optimization run on 6 variables with population size set to 90, minimum surrogate points set to 110 and maximum number function evaluation set to 250.

From Section 3.5.1, the best two-stage DEA model has an AIC of -55.7257 when selecting the first 7 variables from Eikon. I repeat this process now for SSO. I set the minimum surrogate points to 700 and the maximum number of function evaluations to 1200. The population is set to 200. As mentioned previously, each two-stage DEA model of efficiency and effectiveness is represented as a matrix of $\mathbb{B}^{3\times 7}$. There are a total of $2^7 \times 2^7 \times 2^7 = 2097152$ two-stage DEA model of efficiency and effectiveness. However, with the CSP of Section 3.3.1, only 630 of these models are feasible. The rest of the models are infeasible. The initial population of 200 models is generated from the CSP of Section 3.3.1. The result of running the SSO on 7 dimensions from Eikon is shown in Figure 3.6. The best

AIC value found from running the surrogate search optimization is also -55.7257. This best model was not part of the initial population of size 200. The best model was generated by the adaptive point represented in black color at about the 700 function evaluation in the figure. As mentioned previously in Section 3, the adaptive point is the minimum of the surrogate function and gets evaluated against the fitness function. As I select more dimensions from Ekion, the algorithm will not scale well. In Figure 3.6, I am dealing with only 7 dimensions. And with 7 dimensions, only two adaptive models were generated where one of them resulted in the minimum AIC. However, after the reset (represented as a vertical blue line) in the figure, all subsequent models were deemed infeasible. Due to the penalization imposed on a model that does not satisfy the CSP by the fitness function; the performance of SSO greatly suffers when compared to the GS algorithm. In the worst case the probability of SSO generating a feasible point is $\frac{630}{2097152}$.

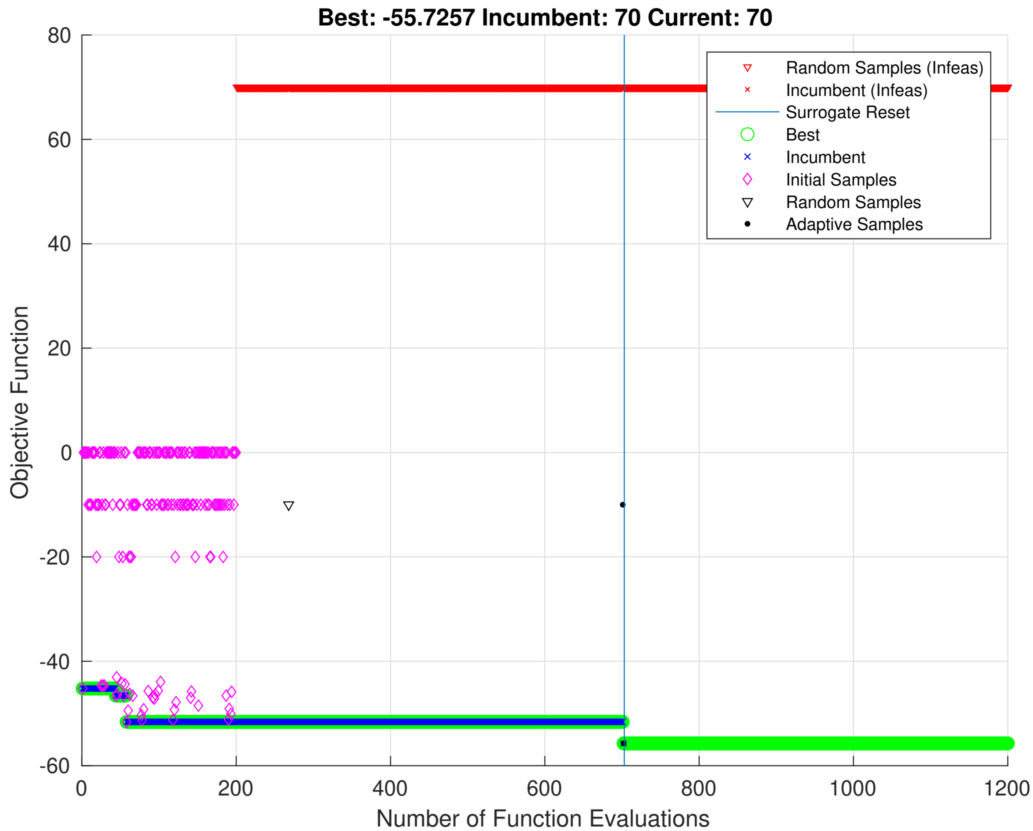

Figure 3.6: Surrogate Optimization run on 7 variables with population size set to 200, minimum surrogate points set to 700 and maximum number function evaluation set to 1200.

Now SSO may work out well if the CSP of Section 3.3.1 is built into the problem setup similar to GS algorithm. However, until then, the GS algorithm works well. I select the GS algorithm over SSO and MABA. I use the GS algorithm as part of the variable selection framework. In Chapter 3, I execute the variable selection framework on my case study. In the case study, I address the next research question **whether there is a consistent (independent of geographic location and time-period) choice of input and output variables for measuring efficiency and effectiveness** in the financial industry using a two-stage DEA model of efficiency and effectiveness.

3.5.3 Results of MABA

I executed the GS algorithm and SSO on 6, 7, and 8 dimensions from the Eikon dataset of Chapter 2 and Section 2.4.1. I repeat this process now for MABA. MABA did not converge to the optimal solution when selecting 7 and 8 dimensions from Eikon. It did converge to the optimal solution for 6 dimensions. Unlike the GS algorithm that converged to the optimal solution in all three cases (i.e., For 6 variables from Eikon, 7 variables from Eikon, and 8 variables from Eikon).

The reason for the non-convergence property of MABA is the current implementation of the belief system. The belief system maintains a frequency counter of all dimensions and their orientation in the two-stage DEA model that resulted in a model consistent with the semi-strong version of the EMH. The belief system also maintains a frequency counter of all dimensions and their orientation in the two-stage DEA model that did not result in a model consistent with the semi-strong version of the EMH. The exploitation stage uses the belief system and selects the top dimensions that resulted in the most two-stage DEA models consistent with the semi-strong version of the EMH. However, when selected together, the top dimensions don't need to result in a better two-stage DEA model of efficiency and effectiveness than previously found from the exploration stage. It is very much possible that the two-stage DEA model generated by combining the top dimensions may result in a worse model and a model that is not consistent with the semi-strong version of the EMH.

Instead of just maintaining a counter as I currently do, what may be required for improved performance of MABA is also to keep some context around each dimension. For example, a context may include a set of dimensions that must be present in the two-stage DEA model when including any of the top dimensions. In other words, conditioned on the fact that top dimensions are selected, what other variables must be included that will increase the likelihood of generating two-stage DEA models consistent with the semi-strong version of the EMH? By including a context, the exploitation stage can exploit the top dimensions that include other dimensions that increase the likelihood of generating models consistent with the semi-strong version of the EMH. In the current implementation of MABA, I use the top k dimensions in the exploitation stage, and because the top k dimensions, when combined together, do not result in an improved model, the exploitation stage does not find a better solution.

In Table 3.1, I see this observation for 7 and 8 variables from Eikon where the top dimensions, when placed together, do not result in an improved model. And even if the exploitation stage results in a two-stage DEA model consistent with the semi-strong version of the EMH, it is not the optimal solution. This suggests that if I were to use MABA (at least the implementation that I have) across all the 55 dimensions, it is unlikely that the algorithm will converge to the optimal solution.

Another issue is collisions of variables in the belief system that compete for the same orientation in the two-stage DEA model. Due to collisions, I consider each of them, which further forces the exploitation to enumerate across many more two-stage DEA models, increasing the time it takes to run the algorithm and the number of two-stage DEA models it must execute. A high value of k and z will result in a very rigid configuration of a two-stage DEA model that may miss out on an optimal solution. In Table 3.1, I select low values for k and z so that the two-stage DEA model from the exploitation stage is not rigid and many more two-stage DEA models in the neighborhood can be explored by the exploitation.

Table 3.1 Result of the best AIC function value from MABA

Number of Variables from Eikon	**k,z**	**AIC value**
6	2,0	-46.5799
7	4,1	No solution consistent with the semi-strong version of EMH found by exploitation stage
8	4,2	No solution consistent with the semi-strong version of EMH found by exploitation stage

A better implementation of MABA will probably make the k, z, and the epsilon adaptive rather than making them constant. If the exploitation stage does not result in an improved model, the adaptive nature of k and z will allow the exploitation stage to broaden the size of the neighborhood and exploit more models. The adaptive nature of the epsilon will allow the algorithm to place more emphasis on exploration if the exploitation stage does not result in an improved model. In the current implementation of MABA, a single starting point is accepted into the exploration stage as opposed to the GS algorithm that can take in an entire population.

For instance, as seen in Figure 3.7,[7] the GS algorithm evaluated 411 two-stage DEA models in 85 seconds when 6 variables were selected from Eikon. MABA evaluated 63 two-stage DEA models in the same amount of time. The GS algorithm evaluated 4106 two-stage DEA models in 728 seconds when 7 variables were selected from Eikon. MABA evaluated 380 two-stage DEA models in the same amount of time. The GS algorithm evaluated 24,406 two-stage DEA models in 3874 seconds when 8 variables were selected from Eikon. MABA evaluated 2198 two-stage DEA models in the same amount of time. The GS algorithm converged in all three cases to the optimal solution. The MABA algorithm did not converge to the optimal solution when selecting 7 and 8 variables from Eikon.

[7] I do not plot the SSO because of the large number of infeasible points that are generated. In other words, an inordinate amount of time passes before SSO encounters a feasible point. In this figure, the Y-axis measures the number of feasible two-stage DEA models evaluated by the search algorithms.

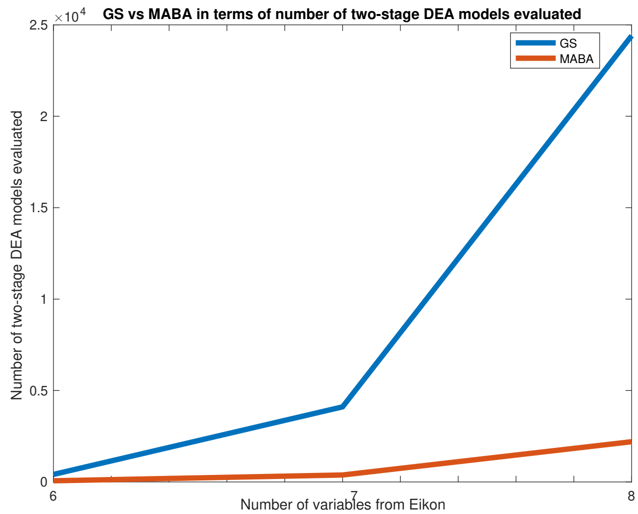

Figure 3.7: Comparing the number of two-stage DEA models evaluated in MABA and GS algorithm, for the same fixed wall-clock time (higher better).

Due to vectorized and parallel computing support in the GS, I select GS over MABA. Another improvement could be to have a matrix of starting points accepted into the exploration stage. This could then take advantage of the code not only being vectorized but also can execute in parallel.

3.6 Conclusions

The purpose of this chapter was to evaluate the three different search algorithms: (1) the GS, (2) the SSO, and (3) the MABA. Based on the results of Section 3.5, the GS performed the best. The GS converged to the best solution in all three cases. The three cases are (1) selecting six dimensions from Eikon, (2) selecting seven dimensions from Eikon, and (3) selecting eight dimensions from Eikon. The best solution was pre-determined by enumerating all the two-stage DEA models of efficiency and effectiveness on the above three cases and running them against the statistical method in Chapter 2. The GS algorithm did not only converge to the optimal solution but most individuals in the population also converged to the optimal solution. This can be observed in the middle plot of best, worst and mean scores of Figures 3.1, 3.2 and 3.3. This convergence of every individual in the population is not observable in the SSO algorithm and the MABA. This convergence behavior of every individual in the population is also complemented with the average distance that decreases across generations suggesting that the two-stage DEA models are converging to a single two-stage DEA model. For these reasons, I select the GS algorithm and use the GS algorithm in the next Chapter 4 when finding the best universal two-stage DEA model across 8 countries and across 18 time periods. SSO is best suited to time-consuming objective functions; the objective function need not be smooth, but the algorithm works best when it is continuous. However, in this dissertation, the objective function is not smooth. In this chapter, I modified the SSO by introducing the CSP from Section 3.3.1 into the objective function (or the fitness function of the SSO algorithm as shown in Algorithm 3). This modification allows SSO to work with discrete input. However, it leads to many individuals in the population being classified as infeasible. In Chapter 4, I use the variable selection framework[8] from this chapter and find a universal two-stage DEA model of efficiency and effectiveness for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for 2000-2017.

[8]Comprises of the GS algorithm and the statistical method

Chapter 4

Building Models of efficiency and effectiveness for the financial industry

4.1 Introduction

The statistical method in Chapter 2 validates whether a certain model of efficiency and effectiveness is consistent with the semi-strong definition of the EMH. The variable selection framework in Chapter 3 uses the GS algorithm and the statistical method to find the best model of efficiency and effectiveness consistent with the semi-strong version of the EMH in a search space of two-stage DEA models. In this chapter, I investigate the following research question: **Is there a universal two-stage DEA model of efficiency and effectiveness that is independent of space (geographic location) and time (time period)?**

In Section 4.4, I present the universal two-stage DEA model of efficiency and effectiveness that is consistent with the semi-strong definition of the EMH along with its input and output variables for banks from eight countries, including Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA for 2000-2017. The universal two-stage DEA model of efficiency and effectiveness reported in Section 4.4 is the best-two stage DEA model according to my variable selection framework. I also address causality, i.e., whether changes to the efficiency and effectiveness measures from the best universal two-stage DEA model directly cause changes to a bank's Tobin's Q ratio and vice versa.

In an efficient market, efficiency and effectiveness cannot cause the Tobin's Q ratio. Otherwise traders can exploit the cause and effect relationship for arbitrage opportunities.

I investigate whether there is any causality between the efficiency and the effectiveness scores computed from the best two-stage DEA model of efficiency and effectiveness and the Tobin's Q ratio using Toda Yamamoto's test (TYT) (Toda and Yamamoto, 1995). I present the TYT test in Section 4.3.1 of this chapter. In Section 4.4.5, using the TYT, I check for the following causalities: (1) Does the efficiency and effectiveness Granger cause the Tobin's Q ratio? (2) Does the Tobin's Q ratio Granger cause efficiency and effectiveness? (3) Does the efficiency Granger cause the Tobin's Q ratio when controlling for effectiveness? (4) Does the effectiveness Granger cause Tobin's Q ratio when controlling for efficiency? (5) Does the Tobin's Q ratio Granger cause efficiency when controlling for effectiveness? (6) Lastly, does the Tobin's Q ratio Granger cause effectiveness when controlling for efficiency? In Section 4.4, I find that efficiency **and** effectiveness from the two-stage DEA model considered best by the variable selection framework does not Granger cause Tobin's Q ratio for all countries except for India.

I find that in India, efficiency Granger causes Tobin's Q ratio when controlling for effectiveness. I also find in India, effectiveness Granger causes Tobin's Q ratio when controlling for efficiency. A bank in India, may not control all the input and output variables of efficiency and effectiveness that I report in Section 4.4. For instance, a bank may only control the input variables of efficiency or the output variables of effectiveness. My recommendation from this chapter for banks in India is that a bank that only controls the input variables of efficiency can increase its efficiency by lowering (or optimizing) its consumption of input variables of efficiency. This will cause an effect in the Tobin's Q ratio. My other recommendation from this chapter for banks in India is that a bank that only controls the output variables of effectiveness can increase its effectiveness by increasing (or optimizing) its output variables of effectiveness. This will cause an effect in the Tobin's Q ratio. More details of my recommendations are provided in Section 4.4.6 of this chapter.

The input and output variables of efficiency and effectiveness that I present in Figure 4.3 of this chapter are consistent with the semi-strong definition of the EMH for banks in countries of Brazil, Canada, China, Japan, Mexico, South Korea and the USA. A bank in any one of these countries that is considered efficient and effective can use the same input and output variables of efficiency and effectiveness when expanding to any one of the above with a high degree of confidence. Besides banks opening newer locations in other places, equity market traders can also benefit from the results of this chapter. In an inefficient market (such as Indian banks as presented in Section 4.4.6 of this chapter), traders can benefit the most by using the cause and effect relationship between efficiency and effectiveness and Tobin's Q ratio. The cause and effect relation can be used to predict the firm's performance in the financial market, and traders can profit from such information.

4.2 Related work

In Section 4.2.1, I highlight some of the differences between TYT test and the Granger causality test and my reasons for choosing the TYT test. In Section 4.2.2, I present some related work in testing of the semi-strong EMH.

4.2.1 Causality

Causality is a central concept in the discussion of economic laws, yet economists disagree about its definition (Hsiao, 1979). The confusion about the use of causality occurs because of its different interpretation in science and economics. Scientists can conduct controlled experiments which generate data; however, economists cannot usually do this outside of the laboratory setting. The data available to an economist are not generated by a controlled experiment. In my dissertation, I use the word "causality" in the context proposed by Granger (1980), which is used frequently in the econometric literature. Granger causality for two time series X and Y is defined as "X is said to Granger cause Y if Y can be better predicted using the histories of both X and Y than it can by using the history of Y alone." Causality findings have important decision implications. Understanding the direction of causality between two variables is crucial for banks when formulating policies that help them not only do well in financial markets but also improve the allocation of their offerings. An efficient and effective bank will allocate resources to projects that have strong growth prospects and pay handsomely for the bank's investment. In this section, I review some of the recent work on Granger causality and the TYT, I also present my reasons for choosing the TYT when inferring causality. I present TYT in Section 4.3.1.

Aziz et al. (2000) documented the shortcomings of the Granger causality test. They used the TYT instead to infer the causal relationships between government spending and tax revenues from 1960 to 1997 in Malaysia. They pointed out that the TYT can infer causality for a nonstationary time series and for any co-integration of time series data. The Granger causality test can be used only on stationary time series data. Similarly, in my dissertation, I use the TYT due to the nonstationarity of Tobin's Q ratio time series data and efficiency and effectiveness time series scores calculated using two-stage DEA models.

Emirmahmutoglu and Kose (2011) documented the shortcomings of the Granger causality test. First, a two-variable Granger causality test that does not consider the effect of other variables is subject to possible specification bias. Second, a Granger causality test cannot handle nonstationary time series data. The recommended approach to test for causality on nonstationary time series data is to use the TYT. Gujarati (2006) mentioned that a causality test is sensitive to model specification as well as the number of lags. Omitted variables and the absence of other relevant variables in the causality model will lead to incorrect results. Second, time series data are often nonstationary (Maddala and Lahiri, 1992). This can amplify the problem of spurious regression which occurs when one nonstationary time series is regressed against another nonstationary time series. The results of such a regression are not only incorrect, but also nonsensical (Entorf, 1997; Roberts, 2000). Spurious regression can be addressed by converting the nonstationary time series data into a stationary time series.

Di Iorio and Triacca (2011) mention that incorrect causal relationships can be derived from a Granger causality test if one does not know whether the time series data are stationary or nonstationary. They recommend using a Granger causality test only when the time series data are $I(0)$ (i.e., stationary). When the data are not $I(0)$, but $I(1)$, the authors recommend performing a first-order of differences to convert the nonstationary data into stationary data, then use a Granger causality test. Finally, if the time series data are co-integrated, then the authors recommend using a vector error correction before performing the Granger causality test. I apply the TYT test on the efficiency and the effectiveness scores from the best model that was found from my variable selection framework. The variable selection framework applies the log transformation on the three time series, Tobin's Q ratio, efficiency, and effectiveness, then checks whether any unit root exists on the panel data by running the statistical test of Section 2.5.1. The variable selection framework rejects any model that continues to have nonstationarity after applying the log transformation. Hence, the best-two stage DEA model as found from my variable selection framework already contains the log transformation of Tobin's Q ratio, efficiency, and effectiveness as a stationary time series. The results of checking for unit root in panel data suggests that, after the log transformation, the time series of Tobin's Q ratio, efficiency, and effectiveness are stationary. The results are mentioned in the Appendix C in Tables C.7, C.10, C.8, C.11, C.9 and C.12.

Shan et al. (2001) talk about the drawbacks of the f-test for Granger causality, especially when trying to find causality between two integrated time series. Instead, they recommend using a modified Wald test statistic and TYT. This is because the asymptotic distribution of the Wald test statistic converges to chi-square independent of whether the time series data are stationary or nonstationary (i.e., $I(0)$, $I(1)$, or $I(2)$) or even if the time series data are co-integrated. Chowdhury and Mavrotas (2006) use the TYT for causality in determining if direct foreign investment in a country affects its growth. Due to their small sample size, TYT test may suffer from size distortion and low power (Mavrotas and Kelly, 2001). To address this, the authors check for the robustness of the causality test results by recalculating the p-values obtained in the initial Wald test using a bootstrap test with 1,000 replications. I also bootstrap using the block bootstrap techniques proposed by Andriansyah and Messinis (2019). I describe the block bootstrap technique in Section 4.3.2.

Andriansyah and Messinis (2019) classify the Dumitrescu–Hurlin (DH) test as a standard Granger causality test. They noted that, in a bivariate setting (i.e., a VAR model with two variables X and Y), when one of the variables is nonstationary, a standard Granger causality test such as DH is not valid due to the problems described in Chowdhury and Mavrotas (2006); Maddala and Lahiri (1992). However, the work of Andriansyah and Messinis (2019) extends the DH test to a trivariate setting by integrating the TYT approach with Granger causality. This work is important to my research because I have a trivariate setting of efficiency, effectiveness, and Tobin's Q ratio. I use the same approach recommended by Andriansyah and Messinis (2019) when running the TYT test.

Hiemstra and Jones (1994) use a linear and a non-linear form of the Granger causality test to examine the relation between Dow Jones stock returns and the percentage change in the New York Stock Exchange. They found evidence of statistically significant bidirectional nonlinear causality between returns and volume. Other papers, such as Hurlin and Venet (2008); Kar et al. (2011); Xie and Chen (2014), include other approaches to converting nonstationary time series data into stationary data. These authors looked for panel data causality and studied financial development and growth. My Ph.D. dissertation uses panel data causality to see whether changing the efficiency and effectiveness from the best two-stage DEA model of efficiency and effectiveness as found from our variable selection framework causes a change in its Tobin's Q ratio in financial markets.

4.2.2 Work related to testing of the semi-strong EMH

Groenewold and Kang (1993) studied whether the Australian share market is consistent with the semi-strong version of the EMH using the data from the 1980s. The authors set up a linear regression where the dependent variable is the rate of return on the stock price. The independent variables are a set of endogenous variables that may affect the rate of return on the stock price. The authors also include the lag variables of these endogenous variables. However, they do not include any endogenous variable at the current time period t. On the other hand, I include efficiency and effectiveness scores computed at time period t along with its lags when setting up the regression model in the statistical method. The semi-strong version of the EMH states that the market instantly reacts and absorbs any new and relevant information. I hypothesize as mentioned in Chapter 2 and Section 2.5.3 that the efficiency and effectiveness scores computed using relevant input and output variables on the day when financial statements are made public will be positive and statistically significant when regressed against the Tobin's Q ratio. The idea is that the same relevant information that was absorbed in the making of Tobin's Q ratio in time period t must also be present in the efficiency and the effectiveness scores for time period t. Others such as Nan and Kaizoji (2019) studied market efficiency of the bitcoin exchange rate. The authors tested whether the bitcoin exchange rate behaves like a random-walk assuming the semi-strong version of the EMH holds. Similar to what I do, the authors include independent variables at time period t along with their lags when testing for semi-strong version of the EMH. Groenewold and Kang (1993) provide no rationale on their choice of selecting the appropriate lag length. On the other hand, I use the AIC information criterion when determining the appropriate lag length when running the correlation on panel data and then later also use AIC when determining the lag length before running the TYT test. The availability of data limits the author's choice of endogenous variables. On the other hand, the variable selection framework uses many different combinations of input and output variable of efficiency and effectiveness from a set of 55 dimensions and finds the two-stage DEA model of efficiency and effectiveness that is consistent with the semi-strong version of EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017.

Ayodele and Maxwell (2017) test whether the Nigerian stock market is consistent with the semi-strong definition of the EMH. The authors conduct a linear regression where the dependent variable is the market index at time period t and the independent variables include the computed index of selected securities with n lags. If any of the n lags include a statistically significant coefficient, the authors reject the semi-strong version of the EMH. Based on this analysis, the authors claim that the Nigerian Stock Exchange is not efficient. Correlation in the regression output does not imply causality. It is important to run a causality check such as the TYT test to determine whether the semi-strong version of the EMH holds. My dissertation's statistical method finds a model that is consistent with the semi-strong version of the EMH. The best model that is found contains lags, see Section 4.4. However, I do not claim that the semi-strong version of the EMH can be rejected. Instead, I run the TYT. Only after running the TYT, I find that efficiency does Granger cause the Tobin's Q ratio in India. Similar result also holds when I check for whether effectiveness Granger cause the Tobin's Q ratio. Therefore, I conclude that at the time, the financial market of India is inefficient. However, all other country financial markets may be said to be efficient.

Alexakis et al. (2010) examine the predictability of stock returns in the Athens Stock Exchange from 1993 to 2006 using accounting information. The authors use panel data analysis to conclude that certain financial ratios contain significant information for predicting the cross-section of stock returns. The authors set up a linear regression where the dependent variable is the stock price for a firm i at time period t. The independent variables include a set of endogenous variables that have no lags, and the stock price for the firm at time period $t-1$. In my dissertation, instead of fixing the lag length to 1, the statistical method and the TYT test use AIC to determine the appropriate lag length. The statistical method traverses a family of models, finding the best model consistent with the semi-strong version of the EMH rather than fixing some linear model as Alexakis et al. (2010) have done. However, the authors check for unit roots in their panel data but do not check for any cross-sectional dependency between firms trading in the Athens Stock Exchange. As a sanity check, I also run the cross-sectional dependency check (see Section 4.4) on the best two-stage DEA model of efficiency and effectiveness found from the variable selection framework.

Torun and Serdar (2008) use co-integration testing and causality testing when inferring whether stock exchanges in European Monetary Union countries are consistent with the semi-strong definition of the EMH. The authors use the fact that if two time series, X and Y, are cointegrated, there must exist Granger causality either from X to Y, or from Y to X, or both. Like Torun and Serdar (2008), Mabakeng and Sheefeni (2014) use co-integration analysis and causality when checking whether Namibia's Foreign Exchange Market is consistent with the semi-strong version of the EMH. Wickremasinghe (2004) uses co-integration testing to determine whether Sri Lanka's foreign exchange market is consistent with the semi-strong version of the EMH. In my dissertation, the variable selection framework performs a log transformation on the data. Then, it checks whether efficiency and effectiveness are stationary. If the log transformation of efficiency and effectiveness is non-stationary, then the variable selection framework rejects that two-stage DEA model of efficiency and effectiveness. This implies that the best two-stage DEA model of efficiency and effectiveness from my variable selection framework always contains a stationary series. Therefore, I perform no co-integration. Instead, I use the TYT test to check whether either efficiency or effectiveness Granger causes the Tobin's Q ratio when controlling for the other. Mabakeng and Sheefeni (2014); Torun and Serdar (2008); Wickremasinghe (2004) check for unit roots in their panel data by using the same statistical test that I use in Chapter 2 and Section 2.5.1. Based on the results of their co-integration analysis, Torun and Serdar (2008) claim that the stock exchanges in European Monetary Union countries are not efficient. Wickremasinghe (2004) finds that the Sri Lanka's foreign exchange market is not consistent with the semi-strong version of the EMH. However, Mabakeng and Sheefeni (2014) claim there is evidence of semi-strong form market efficiency in Namibia's foreign exchange market.

4.3 Methodology

In Section 4.3.1, I present the TYT in a tri-variate setting. In Section 4.3.1, I also present the algorithm for executing TYT of tri-variate setting in a panel data. In Section 4.3.2, I describe the block bootstrapping technique. In Section 4.3.3, I demonstrate how principal component analysis (PCA) is used in generating the initial population size for running the GS algorithm from the previous chapter.

4.3.1 TYT test in tri-variate setting

After the variable selection framework from Chapter 3 finds the best two-stage DEA model of efficiency and effectiveness, I then run the TYT test. I run the TYT test using the efficiency and the effectiveness scores computed from the best two-stage DEA model and Tobin's Q ratio in determining any cause and effect relationship among them. The statistical method of Chapter 2 finds the linear model that best describes the correlation between the dependent variable of Tobin's Q ratio and the independent variables of efficiency, effectiveness, and their lags.

The result from the statistical method determines whether the efficiency and effectiveness measures of a model are consistent with the semi-strong version of the EMH. The correlation result of the statistical method does not imply causality. The TYT test can determine whether there is any causal relationship involving Tobin's Q ratio and the efficiency and the effectiveness. In an efficient market, one does not expect to find any causal relationship between efficiency and effectiveness, causing Tobin's Q ratio. However, in an inefficient market, arbitrage opportunities may exist where traders can use the causal relationship for financial gains. In Section 4.4, I present the results of the TYT test and identify the market of India as inefficient.

Using TYT, I perform a causal analysis in panel data on the following six research questions:

1. Does efficiency *and* effectiveness Granger cause Tobin's Q ratio?
2. Does efficiency Granger cause Tobin's Q ratio when controlling for effectiveness?
3. Does effectiveness Granger cause Tobin's Q ratio when controlling for efficiency?
4. Does Tobin's Q ratio Granger cause efficiency when controlling for effectiveness?

5. Does Tobin's Q ratio Granger cause effectiveness when controlling for efficiency?

6. Does Tobin's Q ratio Granger cause efficiency *and* effectiveness?

The regression setup of Equation (4.1) is used for determining whether efficiency and effectiveness Granger cause Tobin's Q ratio. The subscript i refers to country i and t refers to time period t. This regression setup is executed N times where N is the total number of countries.

$$\log(TobinQ_{i,t}) = \alpha_i + \sum_{j=1}^{m+k} \beta_{1,i,j} \log(TobinQ_{i,t-j}) + \sum_{j=1}^{m+k} \beta_{2,i,j} \log(Efficiency_{i,t-j}) + \sum_{j=1}^{m+k} \beta_{3,i,j} \log(Effectiveness_{i,t-j}) + \mu_{i,t} \quad (4.1)$$

The null hypothesis, H_0 is stated as *efficiency and effectiveness does not Granger cause Tobin's Q ratio*. Equivalently, $\beta_{2,i,1} = \beta_{2,i,2} = \beta_{2,i,3} = \ldots = \beta_{2,i,m} = 0$ and $\beta_{3,i,1} = \beta_{3,i,2} = \beta_{3,i,3} = \ldots = \beta_{3,i,m} = 0$ for $i = 1, \ldots, N$. The alternative hypothesis, H_1 then is stated as *efficiency or effectiveness Granger cause Tobin's Q ratio*. Equivalently, $\beta_{2,i,1} = \beta_{2,i,2} = \beta_{2,i,3} = \ldots = \beta_{2,i,m} = 0$ for $i = 1, \ldots, N_2$ and $\beta_{2,i,1} \neq 0$ or $\ldots$ or $\beta_{2,i,m} \neq 0$ for $i = N_2 + 1, \ldots, N$. N_2 is unknown where $N_2 \in [0, N-1]$, and $\beta_{3,i,1} = \beta_{3,i,2} = \beta_{3,i,3} = \ldots = \beta_{3,i,m} = 0$ for $i = 1, \ldots, N_3$ and $\beta_{3,i,1} \neq 0$ or $\ldots$ or $\beta_{3,i,m} \neq 0$ for $i = N_3 + 1, \ldots, N$. N_3 is unknown and $N_3 \in [0, N-1]$. If $N_2 = 0$ and $N_3 = 0$ there is causality for all countries in the panel. $N_3 < N$ and $N_2 < N$. The Equation (4.1) above has a total of $m + k$ lag variables. m is calculated using an information criteria. I use AIC (Snipes and Taylor, 2014) when calculating m. k is calculated using the maximal order of integration (Amiri and Ventelou, 2012).

In Algorithm 5, from line 6 to 12, I describe how k is calculated. k is the maximal order of integration which refers to the maximum number of differencing that must be applied for converting non stationary series in panel data to stationary. m on line 18 and 19 is calculated as the maximum of $lag_{country}$. $lag_{country}$ is the optimal lag length of the VAR model from line 15 for each of the country. $lag_{country}$ is determined using AIC. If any serial correlation exist on $lag_{country}$ (see line 17) then $lag_{country}$ is increased until the serial correlation is removed. In Section 4.4.5, I present the results of $lag_{country}$, k and m. On line 22, I setup the VAR model of Equation (4.1) with its corresponding null hypothesis when addressing research question (1) as mentioned in the top of this section. Later in this section, I develop the VAR model for research question (2) and its corresponding hypothesis. I then estimate on line 25 to 29 the average Wald statistic ($\bar{W}$), $\bar{Z}$, $\tilde{Z}$ and the corresponding $p$values of $\bar{Z}$ and $\tilde{Z}$.

Dumitrescu and Hurlin (2012) mention that $\bar{W}$, where $\bar{W} = \frac{\sum_{i=1}^{N} W_i}{N}$ is asymptotically well behaved and can be used to detect causality at panel level. W_i is the ith country's Wald statistic calculated from line 23 of the algorithm. Using $\bar{W}$, two test statistics of $\bar{Z}$ and $\tilde{Z}$ are computed (see line 26 and 27 of the algorithm). $\bar{Z}$ follows the standard normal distribution when $T \to \infty$ and then $N \to \infty$. $\tilde{Z}$ works well for a fixed T dimension where $T > 5 + 3(k + m)$ and also follows the standard normal distribution when $N \to \infty$. Dumitrescu and Hurlin (2012) presents how to calculate $\bar{Z}$ and $\tilde{Z}$ in a bi-variate setting and Andriansyah and Messinis (2019) presents how to calculate $\bar{Z}$ and $\tilde{Z}$ in a tri-variate setting. If $\bar{Z}$ and $\tilde{Z}$ are greater than the critical values, then the null hypothesis is rejected in favor of the alternative hypothesis.

I show the hypothesis setup for research question (2) next, however a similar setup can be performed for research questions (3), (4) and (5). In Section 4.4.5, I use the results of (4) and (5) to estimate the result of (6). The regression setup, as shown in Equation (4.1) is also used for whether efficiency Granger causes Tobin's Q ratio when controlling for effectiveness. The null hypothesis, H_0 is *efficiency does not Granger cause Tobin's Q ratio when controlling for effectiveness*. Equivalently, $\beta_{2,i,1} = \beta_{2,i,2} = \beta_{2,i,3} = \ldots = \beta_{2,i,m} = 0$ for $i = 1, \ldots, N$. The alternative hypothesis, H_1 then is *efficiency does Granger cause Tobin's Q ratio when controlling for effectiveness*. Equivalently, $\beta_{2,i,1} = \beta_{2,i,2} = \beta_{2,i,3} = \ldots = \beta_{2,i,m} = 0$ for $i = 1, \ldots, N_1$ and $\beta_{2,i,1} \neq 0$ or $\ldots$ or $\beta_{2,i,m} \neq 0$ for $i = N_1 + 1, \ldots, N$. $N_1 \in [0, N - 1]$ is unknown.

On line 22 of the Algorithm 5, I setup the VAR model of Equation (4.1) with the corresponding null hypothesis when addressing research question (2).

Algorithm 5 Algorithm for Toda-Yamamoto Causality Test Procedure in Panel Data

Preconditions:

- Panel data: Panel data of $log(TobinQ_{i,t})$, $log(Efficiency_{i,t})$ and $log(Effectiveness_{i,t})$
  where $1 \leq i \leq 8$ and $1 \leq t \leq 18$.
  8 represents the number of countries and 18 represents the number of time periods.

Postconditions:

- Wald score, Test statistic and pValues: Wald, $\bar{Z}$, $\tilde{Z}$, $pValue_{\bar{Z}}$, $pValue_{\tilde{Z}}$

```
procedure PerformTYT(panelData)
    X,X_copy ← getTobinQ(panelData)
    Y,Y_copy ← getEfficiency(panelData)
    Z,Z_copy ← getEffectiveness(panelData)
    k← 0
    while true do
        if isStationary(X_copy) ∧ isStationary(Y_copy) ∧ isStationary(Z_copy) then
            break
        k← k + 1
        X_copy ←performDifferencingIfNonStationary(X_copy)
        Y_copy ←performDifferencingIfNonStationary(Y_copy)
        Z_copy ←performDifferencingIfNonStationary(Z_copy)
    m← 0
    for each country in {Brazil, Canada, China, India, Japan, Mexico, SouthKorea, USA} do
        v_country ← setupVarModelForCountryWithLags(X, Y, Z, country, 4)
        lag_country ← runVarAndFindBestLagLength(lag_country, AIC)
        lag_country ← checkForSerialCorrelationAndIncreaseLagIfNeccessary(lag_country)
        if m < lag_country then
            m ← lag_country
    Wald← 0
    ⋮
```

PERFORMTYT CONT...

for each $country\ in\ \{Brazil, Canada, China, India, Japan, Mexico, SouthKorea, USA\}$ **do**

  $v_{country} \leftarrow$ setupVarModelForCountryWithLags$(X, Y, Z, country, m + k)$

  $wald_{country} \leftarrow$ runVarAndEstimateWald$(v_{country})$

  $Wald \leftarrow Wald + wald_{country}$

$Wald \leftarrow \frac{Wald}{numberOfCountries}$

$\bar{Z} \leftarrow$ estimateZBarFromWald(Wald)

$\tilde{Z} \leftarrow$ estimateZTildeFromWald(Wald)

$pValue_{\bar{Z}} \leftarrow$ estimatePValue$(\bar{Z})$

$pValue_{\tilde{Z}\leftarrow}$ estimatePValue$(\tilde{Z})$

Return Wald, $\bar{Z}$, $\tilde{Z}$, $pValue_{\bar{Z}}$, $pValue_{\tilde{Z}}$

4.3.2 Block bootstrap

Due to cross-sectional dependency that may exist in panel data, Andriansyah and Messinis (2019); Dumitrescu and Hurlin (2012) recommends using a block bootstrap approach. I will present the block bootstrapped approach for **whether efficiency Granger cause Tobin's Q ratio when controlling for effectiveness**. A similar setup can be created for other scenarios.

1. Run the TYT test and compute the $\bar{Z}$ and $\tilde{Z}$ for Equation (4.1).

2. Under the null hypothesis that *efficiency does not Granger cause Tobin's Q ratio when controlling for effectiveness*, for Equation (4.1), $\beta_{2,i,1} = \beta_{2,i,2} = \beta_{2,i,3} = \ldots = \beta_{2,i,k+m} = 0$ for all i. The following model is then estimated:

$$\log(TobinQ_{i,t}) = \hat{\alpha}_i + \sum_{j=1}^{k+m} \hat{\beta}_{1,i,j} \log(TobinQ_{i,t-j}) + \sum_{j=1}^{k+m} \hat{\beta}_{3,i,j} \log(Effectiveness_{i,t-j}) + \mu_{i,t} \tag{4.2}$$

   and the residuals are collected in the matrix $\hat{\mu}$ where $\hat{\mu} \in \mathbb{R}^{(T-(k+m))\times N}$.

3. A matrix $\mu^\star$ is built where $\mu^\star \in \mathbb{R}^{(T-(k+m))\times N}$ by resampling blocks of rows of matrix $\hat{\mu}$. Each row j in the matrix $\mu^\star$ contains the residual of all countries for time period j. Resampling blocks of row will preserve any temporal dependency that may be present

among the countries. In a non block bootstrap approach, temporal dependency or correlation among countries is not accounted for.

4. Generate a random draw of $(\log(\mathbf{TobinQ_1})^\star, \ldots, \log(\mathbf{TobinQ_{k+m}})^\star)$ where $(\log(\mathbf{TobinQ_t})^\star = (\log(TobinQ_{1,t})^\star, \log(TobinQ_{2,t})^\star, \ldots, \log(TobinQ_{N,t})^\star)$ by randomly selecting a block of $k+m$ consecutive time periods with replacement (Berkowitz and Kilian, 2000; Stine, 1987).

5. Construct the resampled series of $\log(TobinQ^\star_{i,t}) = \hat{\alpha}_i + \sum_{j=1}^{k+m} \hat{\beta}_{1,i,j} \log(TobinQ^\star_{i,t-j}) + \sum_{j=1}^{k+m} \hat{\beta}_{3,i,j} \log(Effectiveness_{i,t-j}) + \mu^\star_{i,t}$. The $\hat{\alpha}$'s and $\hat{\beta}_1$'s are used from Equation (4.2).

6. Fit the model of $\log(TobinQ^\star_{i,t}) = \alpha_i + \sum_{j=1}^{k+m} \beta_{1,i,j} \log(TobinQ^\star_{i,t-j}) + \sum_{j=1}^{k+m} \beta_{2,i,j} \log(Efficiency_{i,t-j})$ $+ \sum_{j=1}^{k+m} \beta_{3,i,j} \log(Effectiveness_{i,t-j}) + \mu_{i,t}$ and compute $\bar{Z}^b$ and $\tilde{Z}^b$.

7. Repeat steps 3 to 6, $10,000$ times.

8. Compute the $p$values and critical values of $\bar{Z}$ and $\tilde{Z}$ of step (1) based on the distribution of $\bar{Z}^b$ and $\tilde{Z}^b$ where $b = 1, \ldots, 10,000$.

4.3.3 Principal Component Analysis (PCA)

The data that we collected from Eikon as mentioned previously in Chapter 2 and Section 2.4.1 is of size $\mathbb{R}^{N \times D \times T}$ where N is the total number of banks across the eight countries, D is the number of dimensions, and T is the number of time periods. N is 241, D is 55 and T is 18. As mentioned in section 2.4.1, I average out the data based on the market capitalization in each country for all banks and for each time period. The result is a new dataset of size $\mathbb{R}^{8 \times 55 \times 18}$. The purpose of using PCA is to find promising starting points for executing the GS algorithm in the variable selection framework. A starting point referred to as sp where $sp \in \mathbb{B}^{3 \times 55}$, in the GS algorithm is a two-stage DEA model of efficiency and effectiveness. The 3 indicates the three different sets of variables in the two-stage DEA model, i.e., (1) input variable of efficiency, (2) output variable of efficiency, and (3) output variable of effectiveness. Note that the output efficiency variables are identical to the input effectiveness variables in the two-stage DEA model. The $\mathbb{B}$ indicates that the matrix of size 3×55 is a matrix of boolean values. A boolean value of 1 in row 1 and column j

indicates that the variable j is part of the efficiency input variables. A boolean value of 1 in row 3 and column k indicates that the variable k is part of the effectiveness output variables.

A promising starting point is a point in a region of dimensions where the dimensions are closely aligned with the eigenvectors of the dataset's covariance. The eigenvectors of the covariance of the dataset are calculated on line 6 of the algorithm as shown on page 81. I then filter out (see line 7) those eigenvectors that cumulatively capture 90% of the total variance of the dataset. Here is the algorithm:

PCA algorithm

Postconditions:

- Promising Dimensions $\in \mathbb{Z}^{+p}$
  where p is the number of promising dimensions.

```
1: procedure performPCA
2:     bankData← LoadEikonDataForAllBanksAndForAllCountries()
3:     marketCapitalDataForBanks← LoadMarketCapitalDataForAllBanksAndForAll-Countries()
4:     countryData← AverageOutData(bankData,marketCapitalDataForBanks)
5:     cov← FindCovariance(countryData)
6:     (eigenvectors,eigenvalue)← GetEigenvectorsAndEigenValues(cov)
7:     (eigenvectors,eigenvalue)=Filter(eigenvectors,eigenvalue)
8:     dotProductResult=Identity(55,55)^T× eigenvectors
9:     topDimensionsAcrossEachEigenvector=GetMax(dotProductResult,7)
10:    Promising Dimensions ← PerformUnionOnColumns(topDimensionsAcrossEachEigenvector)
```

Line 7, filters out 7 eigenvectors that captures 90% of the total variance of the Eikon dataset.

Table 4.1 % of variance captured by each of the 7 eigenvectors

Eigenvector	**% of total variance captured**
$\mathbf{e_7}$	38%
$\mathbf{e_6}$	16.93%
$\mathbf{e_5}$	11.85%
$\mathbf{e_4}$	8.38%
$\mathbf{e_3}$	6.21%
$\mathbf{e_2}$	5.58%
$\mathbf{e_1}$	3.06%
Total Variance Captured:	90%

Each of the 7 eigenvectors is an orthonormal vector and is in the coordinate system of the basis vectors from the Eikon dataset. The basis vectors of the Eikon dataset is the standard basis. Each eigenvector ($\mathbf{e_i}$) contains coordinates in the standard basis i.e, $\mathbf{e_i} \in \mathbb{R}^{55}$ and $i \in \{1, 2, 3, 4, 5, 6, 7\}$. The standard basis is the set of 55 dimensions from the Eikon dataset. Due to this, the dot product (see line 8 of the algorithm) of each of the basis vectors from the Eikon dataset against the eigenvectors will result in a dot product between -1 to 1. In Appendix E, Table E.1 lists the results of the 7 largest dot product of the dimensions from Eikon across each of the 7 eigenvectors of the covariance of the dataset. In Appendix E, Table E.2 lists the actual dimension number corresponding with each of the dot product results of Table E.1. Each of the dimension numbers varies from 1 to 55, representing one of the 55 dimensions from the Eikon dataset.

Table E.2 highlights the dimensions from Eikon that are similar to each of the eigenvectors or dimensions that are pointing in the direction of the largest variance. I now take the union of the columns of Table E.2 i.e.,

$PD = \{$1,2,3,4,5,6,7,9,10,16,17,18,19,20,21,22,25,26,27,28,29,30,31,32,33,34,35,37,38,39, 41,42,43,46,47,48,49,50,52,53,54,55$\}$

where PD is the set of promising dimensions. Line 10 of the algorithm generates PD. PD is now used in the constrained satisfaction problem of Chapter 3 and Section 3.3.1 in generating starting points for the GS algorithm of the variable selection framework. The PCA algorithm used in the panel data cannot distinguish between time and cross-section-specific variances in the variables.

4.4 Results

In Section 4.4.1, I mention the lag length selection criteria in the statistical method for panel data regression. In Section 4.4.2, I present the results of the variable selection framework on banks in Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA from 2000 to 2017. In Section 4.4.6, I provide discussion and managerial insights about my results.

4.4.1 Lag length selection criteria in the statistical method

As mentioned in Chapter 3 and Section 3.4.1, to use AIC, the number of data points in the panel data must remain constant when comparing a statistical model containing lags against a model containing no lags. The panel data is first preprocessed by removing K_{max} time periods. K_{max} is a positive integer representing the number of maximum lags that the panel data contains (Lopez et al., 2017).

First, I set $K_{max} = 2$, resulting in a panel data of size $8 \times (18 - 2)$. 8 refers to the number of countries, and 18 refers to the total number of time periods that I consider in my panel data. I run the statistical method for $10,000$ randomly chosen two-stage DEA models of efficiency and effectiveness. I find 310 that are statistically significant, meaning they are consistent with the semi-strong EMH. Out of these, 53 have no lags, 75 have lags of 1, and 182 have lags of 2. Out of the 310 models, the model with the lowest AIC score contains 1 lag. When $K_{max} = 3$, i.e., the new panel data is now of size $8 \times (18 - 3)$, I run the statistical method for the same $10,000$ two-stage DEA models of efficiency and effectiveness. I find that 242 are statistically significant. Out of these, 32 have no lags, 126 have lags of 1, and 84 have lags of 2. I find no statistically significant models with lags of 3. Out of these 242 models, the model with the lowest AIC score contain no lag. Based on these results and on losing an extra 8 data points when setting to $K_{max} = 3$, I decide to set $K_{max} = 2$ in the statistical method when running the GS algorithm of the variable selection framework. This also makes it consistent with the 2 lags that was set in Chapter 2 when validating Kumar and Gulati (2010)'s two-stage DEA model of efficiency and effectiveness against the semi-strong version of the EMH.

4.4.2 Results from the variable selection framework

I executed the GS algorithm of the variable selection framework 30 different times. On each of these 30 runs, an initial population of 5000 points was set. Each individual in the population is a certain two-stage DEA model of efficiency and effectiveness. From these 5000 starting points, about 50% were generated in the neighborhood of the dimensions suggested from PCA (see section 4.3.3) using the constrained satisfaction problem from Chapter 3 and section 3.3.1. The other 50% were randomly generated using a uniform distribution. Each of the 30 runs, ran for a total of 10 days or when the number of consecutive generations without any improvement reached 500 which ever came first. On average, each run explored about $15,200,000$ two-stage DEA models. The statistical method from Chapter 2 reported the best AIC score of -61.9107 across all the 30 runs. The output of the statistical method is shown in table 4.2.

Table 4.2 Output from the statistical method on the best two-stage DEA model of efficiency and effectiveness from variable selection framework

$N = 128$	AIC=-61.9107
n=8	$T = 16$
$R^2 = .11160$	$AdjR^2 = .02736$
Wald F(4,7)=9.296765	p-value = .0062

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$\log(efficiency_t)$	.356226	.090470	3.9375	.006
$\log(effectiveness_t)$	.388642	.082897	4.6882	.002
$\log(efficiency_{t-1})$	-.437795	.133423	-3.2813	.013
$\log(efficiency_{t-2})$	.228851	.121640	1.8814	.102

Standard errors robust to heteroskedasticity adjusted for 8 clusters .

The corresponding two-stage DEA model is shown in figure 4.1.

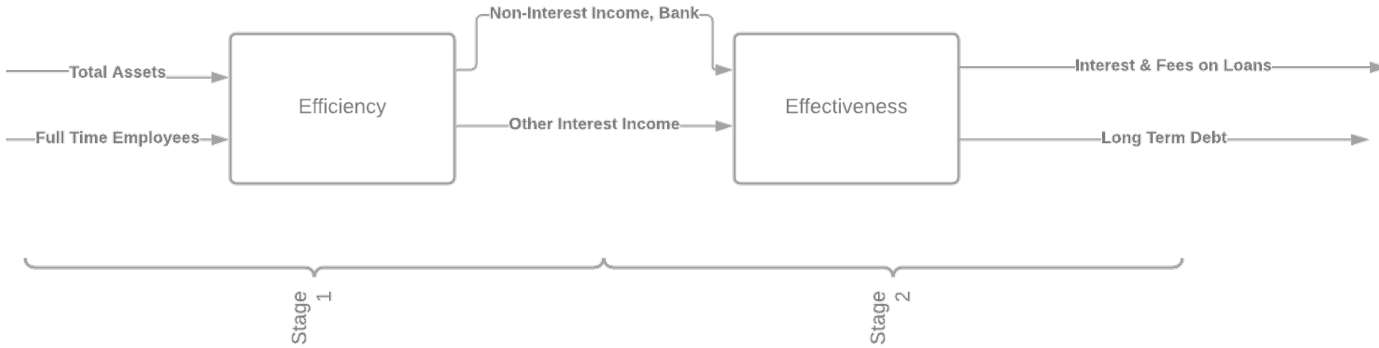

Figure 4.1: Best two-stage DEA model from the variable selection framework.

The average AIC from all the 30 different runs is -47.4718. The best two-stage DEA model on 14 of the 30 runs always contains the variables of 'total assets,' 'number of employees' and 'non-interest income.' 8 of the 14 runs always contain the variables of 'total assets' and 'number of employees' as input variables of efficiency and 'non-interest income' as output variables of effectiveness.

I decided to enumerate all the two-stage DEA models such that 'total assets' and 'number of employees' always occur together either as input variables of efficiency or output variables of efficiency, or output variables of effectiveness. In addition, the variable of 'Non-interest income' is always present in the two-stage DEA model. The reason for doing this is to check if there is a better two-stage DEA model by fixing the above variables in the two-stage DEA model than what the variable selection framework found. One may think of the enumeration as a process of performing a localized search around the best-two-stage DEA model of efficiency and effectiveness found from the variable selection framework. There are a total of 55 dimensions in the Eikon dataset. The enumeration yields a total of $6 \times \binom{52}{1} \times \binom{51}{2} = 397800$ two-stage DEA models of efficiency and effectiveness. There are 6 different ways (also referred to as permutations in Figure 4.2) to arrange 'total assets' and 'number of employees' always occurring together and 'non-interest income' in the two-stage DEA model. The result from our enumeration is summarized in Figure 4.2.

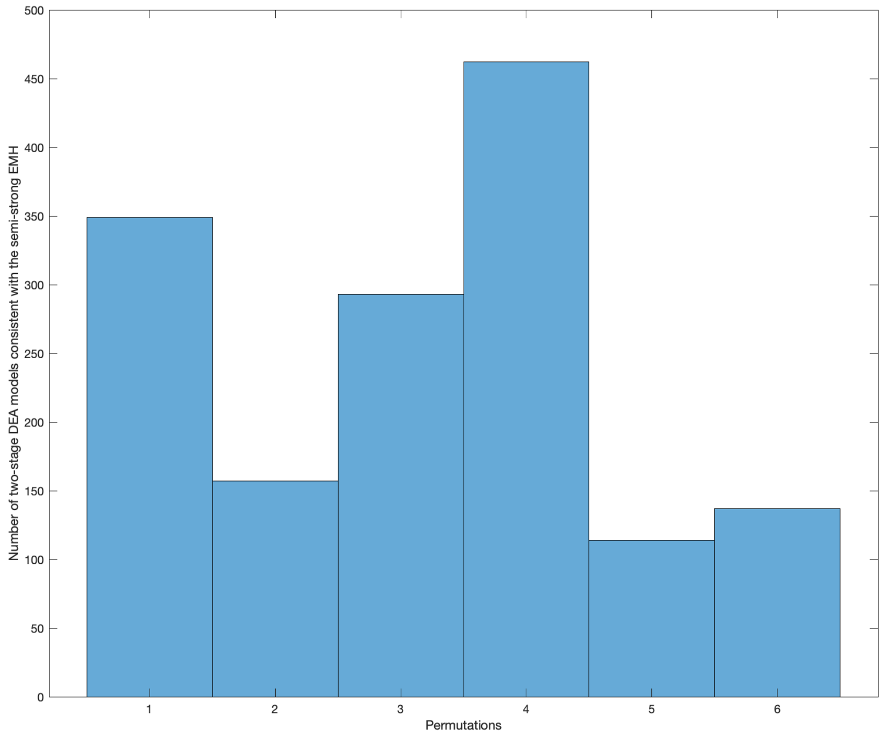

Figure 4.2: Distribution of two-stage DEA model consistent with semi-strong EMH after performing enumeration across all the 6 different permutations

A total of 1592 two-stage DEA models of efficiency and effectiveness were consistent with the semi-strong EMH from the $397,800$ models enumerated. Figure 4.2 displays the distribution of these 1592 models across the 6 different permutations. In Appendix F, in Figure F.1, Figure F.2, Figure F.3, Figure F.4, Figure F.5 and Figure F.6; I list the orientation of the variables 'total assets', 'full time employees' and 'non-interest income' in the two-stage DEA model of efficiency and effectiveness across the 6 different permutations. After enumerating $397,800$ two-stage DEA models of efficiency and effectiveness, a better two-stage DEA model of efficiency and effectiveness resulted than the model found by the variable selection framework. This new two-stage DEA model is presented in Figure 4.3.

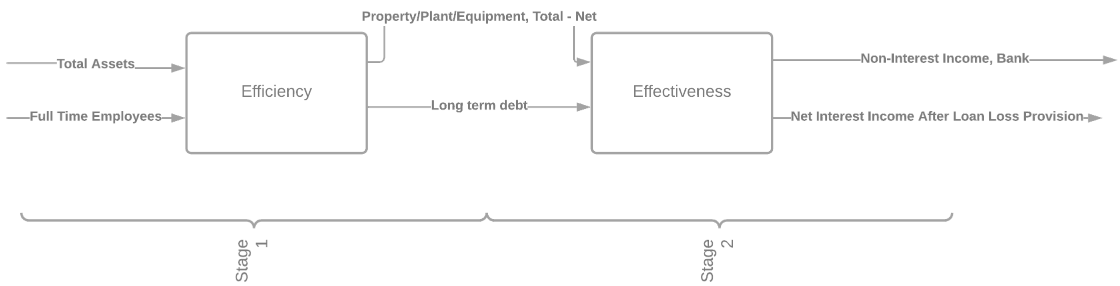

Figure 4.3: Best two-stage DEA model after enumeration.

The statistical test of finding unit root in panel data infers that the log transformation of Tobin Q ratio; and the log transformation of the efficiency and the log transformation of effectiveness computed from the two-stage DEA model of Figure 4.3 are stationary. In Table C.7, Table C.8 and Table C.9 of Appendix C, I report the results of checking for unit roots without the log transformation. After applying the log transformation I report the results in Table C.10, Table C.11 and Table C.12 of Appendix C. I talk more about the statistical test of finding unit root in panel data in Chapter 2 and Section 2.5.1 of this dissertation. Table 4.3 presents the result of the statistical method on the two-stage DEA model of efficiency and effectiveness of Figure 4.3.

Table 4.3 Output from the statistical method on the best two-stage DEA model of efficiency and effectiveness from running the enumeration.

$N = 128$	AIC=-63.1407
n=8	$T = 16$
$R^2 = .11594$	$AdjR^2 = .03211$
Wald F(4,7)=18.146207	p-value = .0008

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$\log(efficiency_t)$	.588653	.242731	2.4251	.046
$\log(effectiveness_t)$	.465596	.169885	2.7406	.029
$\log(effectiveness_{t-1})$	-.727668	.111479	-6.5274	.000
$\log(efficiency_{t-2})$	-.744820	.426930	-1.7446	.125

Standard errors robust to heteroskedasticity adjusted for 8 clusters .

4.4.3 Variance inflation factor (VIF)

I calculate VIF for each of the independent variable and check whether any multicollinearity exist among the independent variables listed in Table 4.3. A VIF of 1 represents no multicollinearity. A VIF greater than 1 represents multicollinearity. VIFs represent the factor by which the correlations amongst the predictors inflate the variance. For example, a VIF of 5 indicates that multicollinearity inflates the variance by a factor of 5 compared to a model with no multicollinearity. Because no formal cutoff value or method exists to determine when a VIF is too large, typical suggestions for acceptable range of VIF is less than 10 (Garg and Tai, 2013). Values greater than 10 suggest severe multicollinearity. Others have taken conservative approach and suggested values must be less than 5 to be said to have no multicollinearity in practice (Craney and Surles, 2002; Stine, 1995). The VIF for the independent variables of Table 4.4 are less than 5.

Table 4.4 VIF for independent variables of Table 4.3

Variable	**VIF**
$\log(efficiency_t)$	2.4587
$\log(effectiveness_t)$	2.6441
$\log(effectiveness_{t-1})$	3.3665
$\log(efficiency_{t-2})$	3.6521

4.4.4 Cross section dependency

Due to banks' inter-connectedness among the eight countries, we believe that there may be a high degree of cross-section dependency among the error terms. Cross-sectional dependence in the errors may arise because of common shocks or when the estimated models present spatial dependence in the disturbances (Álvarez et al., 2017). Not accounting for cross-sectional dependence will result in inefficient estimators, and we may draw incorrect inferences from it.

I use Pesaran (2004)'s cross-sectional dependence test and check for any cross-sectional dependence among the error terms. This test is executed on the statistical model of Table 4.3. The null hypothesis of the Pesaran (2004) assumes that there is no cross-sectional dependence. The Pesaran test statistic calculated is -.1616 for a p-value of .4358. The null hypothesis that there is no cross-sectional dependency is not rejected.

4.4.5 Toda Yamamoto causality test (TYT) results

I now check whether the best two-stage DEA model of Figure 4.3 exhibits any causality. I use the TYT and check for causality in the following scenarios:

1. Does efficiency and effectiveness Granger cause Tobin's Q ratio?
2. Does efficiency Granger cause Tobin's Q ratio while controlling for effectiveness?
3. Does effectiveness Granger cause Tobin's Q ratio while controlling for efficiency?
4. Does Tobin's Q ratio Granger cause efficiency while controlling for effectiveness?
5. Does Tobin's Q ratio Granger cause effectiveness while controlling for efficiency?
6. Does Tobin's Q ratio Granger cause efficiency and effectiveness?

From Table C.10, Table C.11 and Table C.12 of Appendix C, I know that the log transformation of Tobin's Q ratio, log transformation of efficiency and the log transformation of effectiveness are stationary series of $I(0)$. To determine the appropriate lag length for our panel data, I use the same approach as suggested by Andriansyah and Messinis (2019). I first create a VAR model of lag length 4 on $log(efficiency)$, $log(effectiveness)$ and $log(tobinQ)$ for each of the 8 countries. A VAR model of lag length 4 contains a maximum of 4 lags. For each of the VAR models, I use the AIC criterion to determine the ideal lag length, where the ideal lag length can vary from 1 to 4.

I then check whether the lag length selected by AIC has any serial correlation on it. I use the Rao F-test version of the Lagrange multiplier (LM) statistic when testing for serial correlation. The Rao F-test version of the LM statistic augments the Edgeworth Likelihood Ratio (LR) form of the test, and Edgerton and Shukur (1999) mentioned that it performs best among the many variants they consider. The null hypothesis for the Rao F-test is that there is no serial correlation at the lag length k where k is the optimal lag length found from AIC. If the null hypothesis is rejected in favor of the alternative hypothesis, I increase the lag length until there is no serial correlation.

Tables D.1, D.2, D.3, D.4, D.5, D.6, D.7 and D.8 in Appendix D presents the results of the optimal lag length from AIC for the 8 countries along with the serial correlation test on the optimal lag length. For Brazil, from Table D.1 the optimal lag length with no serial correlation is 1. For India, from Table D.2 the optimal lag length with no serial correlation is 1. For China, from Table D.3 the optimal lag length with no serial correlation is 3. For the USA from Table D.4 the optimal lag length with no serial correlation is 2. Notice that

I increased the lag length as determined from AIC by 1. This is because there is a serial correlation on lag length of 1 and no serial correlation on lag length of 2. For Canada from Table D.5 the optimal lag length with no serial correlation is 2. For Mexico, from Table D.6 the optimal lag length with no serial correlation is 2. For South Korea, from Table D.7 the optimal lag length with no serial correlation is 3. For Japan, from Table D.8 the optimal lag length with no serial correlation is 2. I select the lag length of 3, which is the maximum across all 8 cross-sectional time series. Similar to Andriansyah and Messinis (2019) I then select m to 3 in Equation (4.1) when running the TYT on panel data. k is set to 0 in Equation (4.1) because after the log transformations, all the three-time series are stationary and of $I(0)$. Before I present the TYT tests for all the six scenarios and their results, I formally state the hypothesis for each of the six scenarios.

1. Does efficiency and effectiveness Granger cause Tobin's Q ratio?

   H_0 is stated as, *In the given panel data, efficiency and effectiveness does not Granger cause Tobin's Q ratio for all countries.*

   H_1 is stated as, *In the given panel data, efficiency or effectiveness Granger cause Tobin's Q ratio for atleast one country.*

2. Does efficiency Granger cause Tobin's Q ratio when controlling for effectiveness?

   H_0 is stated as, *In the given panel data, the efficiency does not Granger cause Tobin's Q ratio for all countries when controlling for effectiveness.*

   H_1 is stated as, *In the given panel data, the efficiency Granger causes Tobin's Q ratio for at least one country when controlling for effectiveness.*

3. Does effectiveness Granger cause Tobin's Q ratio when controlling for efficiency?

   H_0 is stated as, *In the given panel data, the effectiveness does not Granger cause Tobin's Q ratio for all countries when controlling for efficiency.*

   H_1 is stated as, *In the given panel data, the effectiveness Granger causes Tobin's Q ratio for at least one country when controlling for efficiency.*

4. Does Tobin's Q ratio Granger cause efficiency when controlling for effectiveness?

   H_0 is stated as, *In the given panel data, Tobin's Q ratio does not Granger cause efficiency for all countries when controlling for effectiveness.*

   H_1 is stated as, *In the given panel data, Tobin's Q ratio Granger causes efficiency for at least one country when controlling for effectiveness.*

5. Does Tobin's Q ratio Granger cause effectiveness when controlling for efficiency?

   H_0 is stated as, *In the given panel data, Tobin's Q ratio does not Granger cause effectiveness for all countries when controlling for efficiency.*

   H_1 is stated as, *In the given panel data, Tobin's Q ratio Granger causes effectiveness for at least one country when controlling for efficiency.*

6. Does Tobin's Q ratio Granger cause efficiency and effectiveness?

   H_0 is stated as, *In the given panel data, Tobin's Q ratio does not Granger cause efficiency and effectiveness for all countries.*

   H_1 is stated as, *In the given panel data, Tobin's Q ratio Granger causes efficiency or effectiveness for atleast one country.*

For each of the above four cases; I calculate two different test statistic of $\bar{Z}$ and $\tilde{Z}$. Dumitrescu and Hurlin (2012); Lopez et al. (2017) have mentioned that $\bar{Z}$ referred to as standardized test statistic follows a normal distribution when $T \to \infty$ and then $N \to \infty$. $\tilde{Z}$ is referred to as the approximate standardized statistic and is used for a fixed time T dimension such that $T > 5 + 3(k + m)$. Dumitrescu and Hurlin (2012) used Monte Carlo simulations and found that even for a small panel where N is small, and T is small, $\tilde{Z}$ exhibits good finite sample properties. I calculate the two p values, one using $\bar{Z}$ and another using $\tilde{Z}$.

Econometric models are used to formulate policy recommendations, and inaccurate conclusions may be harmful (Lopez et al., 2017). Due to large trading among banks across different countries, I suspect that there may be cross-dependence among the countries. Dumitrescu and Hurlin (2012) recommends a block bootstrap procedure to compute bootstrapped critical values and then to use these bootstrapped critical values for determining whether $\bar{Z}$ and $\tilde{Z}$ are statistically significant. Hence, in addition to calculating p values based on asymptotic critical values of standard normal distribution, I also calculate the bootstrapped critical values for determining whether $\bar{Z}$ and $\tilde{Z}$ are statistically significant in the bootstrapped critical values. I use the block bootstrapping approach for (3), (4), (5) and (6). I bootstrap $10,000$ samples and then calculate critical values at 1%, 5%, and 10% level. I use the same approach to bootstrapping as implemented and suggested by Andriansyah and Messinis (2019).

Some researchers such as Gorus and Aydin (2019); Kim et al. (2018); Paramati et al. (2016) report only $\bar{Z}$ for relatively small panel data. To be consistent with them, I also use $\bar{Z}$ when providing managerial insights and discussion in Section 4.4.6.

Does efficiency and effectiveness Granger cause Tobin's Q ratio?

Using the hypothesis development of Equation (4.1) from Section 4.3.1 and stated again in the first of the six scenarios on top of this section, I first find the p value and decide whether to reject the null hypothesis H_0 (i.e., p value is less than 0.05) in favour of the alternative hypothesis H_1. My result is summarized in Table 4.5. Based on asymptotic critical values of standard normal distribution, $\bar{Z}$ is statistically significant at 1%. $\tilde{Z}$ is not statistically significant at 10%. Using $\bar{Z}$, I can reject the null hypothesis in favor of the alternative hypothesis, i.e, that efficiency or effectiveness does Granger cause Tobin's Q ratio for at least 1 country.

However I cannot reject the null hypothesis when checking $\tilde{Z}$ against the asymptotic critical values and against the bootstrapped critical values at 10%. I also cannot reject the null hypothesis when checking $\bar{Z}$ against the bootstrapped critical values at 10%.

Table 4.5 Does efficiency and effectiveness Granger cause Tobin's Q ratio.

Extra lag k = 0
Lag order m = 3
Size T = 18 Size N = 8
$\bar{W}$ statistic = 3.8965
$\bar{Z}$ statistic = 3.1057 $p$value = 0.0019
$\tilde{Z}$ statistic (standardized for fixed T value) = -0.0732 $p$value = 0.9417

	1 %	**5 %**	**10 %**
Bootstrapped critical values for $\bar{Z}$	79.0739	49.8877	39.8806
Bootstrapped critical values for $\tilde{Z}$	-0.0046	0.7509	1.2072

Cross Unit Identifier	**Wald statistics**	**pvalue**
Brazil	3.7780	0.2865
India	13.5731	0.0035
China	3.5549	0.3137
USA	0.1680	0.9826
Canada	3.2435	0.3556
Mexico	3.2344	0.3569
South Korea	3.0356	0.3862
Japan	0.5848	0.8999

At the country level, India exhibits causality at $\alpha = 1\%$.

Does efficiency Granger cause Tobin's Q ratio when controlling for effectiveness?

Using the hypothesis development from Section 4.3.1 and stated again in the second of the six scenarios on top of this section, I first find the p value and decide whether to reject the null hypothesis H_0 (i.e., p value is less than 0.05) in favour of the alternative hypothesis H_1. My result is summarized in Table 4.6. Based on asymptotic critical values of standard normal distribution, $\bar{Z}$ is statistically significant at 1%. Using $\bar{Z}$, I can reject the null hypothesis in favor of the alternative hypothesis i.e, that efficiency does Granger cause Tobin's Q ratio when controlling for effectiveness for at least 1 country.

However I cannot reject the null hypothesis when checking $\tilde{Z}$ against the asymptotic critical values and against the bootstrapped critical values at 10%. I also cannot reject the null hypothesis when checking $\bar{Z}$ against the bootstrapped critical values at 10%.

Table 4.6 Does efficiency Granger cause Tobin's Q ratio when controlling for effectiveness?

Extra lag k = 0
Lag order m = 3
Size T = 18 Size N = 8
$\bar{W}$ statistic = 4.1688
$\bar{Z}$ statistic = 4.0488 $p$value = 0.0001
$\tilde{Z}$ statistic (standardized for fixed T value) = 0.1193 $p$value = 0.9050

	1 %	**5 %**	**10 %**
Bootstrapped critical values for $\bar{Z}$	46.8200	27.8605	21.8951
Bootstrapped critical values for $\tilde{Z}$	8.8500	4.9799	3.7622

Cross Unit Identifier	**Wald statistics**	**pvalue**
Brazil	4.5555	0.2074
India	7.9170	0.0478
China	6.6980	0.0822
USA	0.2641	0.9666
Canada	5.8946	0.1169
Mexico	3.2660	0.3524
South Korea	0.9476	0.8139
Japan	3.8074	0.2830

At the country level, India and exhibits causality at $\alpha = 5\%$

Does effectiveness Granger cause Tobin's Q ratio when controlling for efficiency?

Using the hypothesis development from Section 4.3.1 and stated again in the third of the six scenarios on top of this section, I first find the p value and decide whether to reject the null hypothesis H_0 (i.e., p value is less than 0.05) in favour of the alternative hypothesis H_1. My result is summarized in Table 4.7. Based on asymptotic critical values of standard normal distribution, $\bar{Z}$ is statistically significant at 1%. Using $\bar{Z}$, I can reject the null hypothesis in favor of the alternative hypothesis i.e, that effectiveness does Granger cause Tobin's Q ratio when controlling for efficiency for at least 1 country.

However I cannot reject the null hypothesis when checking $\tilde{Z}$ against the asymptotic critical values and against the bootstrapped critical values at 10%. I also cannot reject the null hypothesis when checking $\bar{Z}$ against the bootstrapped critical values at 10%.

Table 4.7 Does effectiveness Granger cause Tobin's Q ratio when controlling for efficiency?

Extra lag k = 0
Lag order m = 3
Size T = 18   Size N = 8
$\bar{W}$ statistic = 3.9645
$\bar{Z}$ statistic = 3.3413 $p$value = 0.0008
$\tilde{Z}$ statistic (standardized for fixed T value) = -0.0251 $p$value = 0.9800

	1 %	**5 %**	**10 %**
Bootstrapped critical values $\bar{Z}$	123.5686	74.8559	59.7948
Bootstrapped critical values $\tilde{Z}$	1.4718	2.4488	3.0215

Cross Unit Identifier	**Wald statistics**	**pvalue**
Brazil	2.2671	0.5189
India	13.6892	0.0034
China	6.9761	0.0727
USA	0.1798	0.9808
Canada	1.4733	0.6885
Mexico	2.4509	0.4842
South Korea	3.1061	0.3756
Japan	1.5739	0.6653

At the country level India exhibits causality at $\alpha = 5\%$.

Does Tobin's Q ratio Granger cause efficiency when controlling for effectiveness?

Using the hypothesis development from Section 4.3.1 and stated again in the fourth of the six scenarios on top of this section, I first find the p value and decide whether to reject the null hypothesis H_0 (i.e., p value is less than 0.05) in favour of the alternative hypothesis H_1. My result is summarized in Table 4.8. Based on asymptotic critical values of standard normal distribution, $\bar{Z}$ and $\tilde{Z}$ are statistically significant at 1%. I can reject the null hypothesis in favor of the alternative hypothesis i.e, that Tobin's Q ratio does Granger cause efficiency when controlling for effectiveness at least for 1 country. Also using the bootstrapped critical value of $\bar{Z}$ and $\tilde{Z}$ at the 10% level, the null hypothesis can be rejected.

Table 4.8 Does Tobin's Q ratio Granger cause efficiency when controlling for effectiveness?

Extra lag k = 0
Lag order m = 3
Size T = 18 Size N = 8
$\bar{W}$ statistic = 10.8896
$\bar{Z}$ statistic = 27.3304 $p$value = 0.0000
$\tilde{Z}$ (standardized for fixed T value) = 4.8717 $p$value = 0.0000

	1 %	**5 %**	**10 %**
Bootstrapped critical values $\bar{Z}$	49.1897	28.6704	21.8668
Bootstrapped critical values $\tilde{Z}$	9.3337	5.1452	3.7564

Cross Unit Identifier	**Wald statistics**	**pvalue**
Brazil	2.9288	0.4027
India	4.1245	0.2483
China	5.8622	0.1185
USA	4.9292	0.1771
Canada	40.6029	0.0000
Mexico	25.8129	0.0000
South Korea	2.0569	0.5607
Japan	0.7995	0.8496

At the country level, Canada and Mexico exhibit causality at $\alpha = 1\%$.

Does Tobin's Q ratio Granger cause effectiveness when controlling for efficiency?

Using the hypothesis development from Section 4.3.1 and stated again in fifth of the six scenarios on top of this section, I first find the p value and decide whether to reject the null hypothesis H_0 (i.e., p value is less than 0.05) in favour of the alternative hypothesis H_1. My result is summarized in Table 4.9. Based on asymptotic critical values of standard normal distribution, $\bar{Z}$ is statistically significant at 1%. Using $\bar{Z}$, I can reject the null hypothesis in favor of the alternative hypothesis i.e, that Tobin's Q ratio does Granger cause effectiveness when controlling for efficiency for at least 1 country.

However I cannot reject the null hypothesis when checking $\tilde{Z}$ against the asymptotic critical values and against the bootstrapped critical values at 10%. I also cannot reject the null hypothesis when checking $\bar{Z}$ against the bootstrapped critical values at 10%.

Table 4.9 Does Tobin's Q ratio Granger cause effectiveness when controlling for efficiency?

Extra lag k = 0
Lag order m = 3
Size T = 18 Size N = 8
$\bar{W}$ statistic = 3.9399
$\bar{Z}$ statistic = 3.2558 $p$value = 0.0011
$\tilde{Z}$ statistic (standardized for fixed T value) = -0.0425 $p$value = 0.9661

	1%	**5%**	**10%**
Bootstrapped critical values $\bar{Z}$	46.6626	26.1181	19.1528
Bootstrapped critical values $\tilde{Z}$	-1.4392	-1.0202	-0.7772

Cross Unit Identifier	**Wald statistics**	**pvalue**
Brazil	1.1978	0.7535
India	3.1957	0.3624
China	10.3911	0.0155
USA	6.1891	0.1028
Canada	0.7364	0.8646
Mexico	4.8688	0.1817
South Korea	2.6653	0.4462
Japan	2.2750	0.5173

At the country level China exhibits causality at $\alpha = 5\%$.

Does Tobin's Q ratio Granger cause efficiency and effectiveness?

I can use the results of *Does Tobin's Q ratio Granger cause efficiency when controlling for effectiveness* on page 95 and the results of *Does Tobin's Q ratio Granger cause effectiveness when controlling for efficiency* on page 96 for answering whether Tobin's Q ratio Granger cause efficiency and effectiveness.

The null hypothesis is stated as, *Tobin's Q ratio does not Granger cause efficiency or effectiveness*. Alternatively, this is the same as *Tobin's Q ratio does not Granger cause efficiency when controlling for effectiveness* or *Tobin's Q ratio does not Granger cause effectiveness when controlling for efficiency*. From page 95, atleast based on $\bar{Z}$'s asymptotic value, the null hypothesis of *Tobin's Q ratio does not Granger cause efficiency when controlling for effectiveness* was rejected. From page 96, atleast based on $\bar{Z}$'s asymptotic value, the null hypothesis of *Tobin's Q ratio does not Granger cause effectiveness when controlling for efficiency* was rejected. Therefore based on the $\bar{Z}$'s asymptotic value the null hypothesis of *Tobin's Q ratio does not Granger cause efficiency or effectiveness* is rejected.

4.4.6 Discussion of results

The results of the statistical method shown in Table 4.3 suggest that the market values banks that exhibit continuous improvement in efficiency and effectiveness from year to year. For instance, in Table 4.3, the coefficient for $\log(efficiency_t)$ is .588653, and $\log(effectiveness_t)$ is .465596. Both of these are statistically significant. A 1% increase in the efficiency at time t results in an increase of .588653% of Tobin's Q ratio. A 1% increase in the effectiveness at time period t results in an increase of .465596% of Tobin's Q ratio. However, the coefficient for $\log(effectiveness_{t-1})$ is .727668 and is statistically significant. The market values those banks that compensate for the decrease in the Tobin's Q ratio caused by the negative coefficient on the lagged variable, efficiency. The net effect on the Tobin's Q ratio is $(.588653 + .465596) - .727668 = .326581$.

In Figure 4.3, the input and output variables of efficiency and effectiveness indicate what is best aligned with the semi-strong version of the EMH. However, knowing what these variables of efficiency and effectiveness are not enough, it is necessary for banks to use this new information to optimize their input and output variables of efficiency and effectiveness so their performance exhibits continuous improvement in efficiency and effectiveness from year to year.

Cetorelli and Peretto (2000); Huangfu et al. (2017); Khemraj (2010) mention that only a handful of banks in a country capture a large market value by capital. For instance, 6 major Canadian banks comprise over 70% of the market value, by capital, of the Canadian banking sector (Allen et al., 2007). In the U.S., there are about 20 banks that capture over 90% of the market value by capital. In the data I used from Eikon, this same trend applies for banks in Brazil, China, India, South Korea, Japan, and Mexico, where only a few banks control over 70% of the market value by capital. Martin (2020) explains that a handful of banks controlling the entire banking market is known as a Pareto distribution.

In this dissertation, as mentioned previously, I average the data collected from Eikon for each country's bank across all time periods based on the market value by capital of the bank. The input and output dimensions of efficiency and effectiveness in Figure 4.3 are representative of this small group of banks that dominate the banking market. The rewards arising from efficiency and effectiveness increase as their performance improves, creating a high degree of specialization and conferring an ever-growing market power on the most efficient competitors (Martin, 2019). Due to economies of scale, the larger banks continue to get more efficient and effective while the smaller banks may suffer. For instance, in Figure 4.3, one of the output variables of efficiency or the input variables of effectiveness is 'Property/Plant/Equipment, Total' (PPE).

PPE includes any of a company's long-term, fixed assets. PPE assets are tangible, identifiable, and expected to generate an economic return for the company for a period of more than one year.[1] PPE includes machinery, equipment, vehicles, buildings, land, office equipment, and furnishings, among other things. A larger bank can increase its effectiveness manifold with every 1 unit increase in PPE due to economies of scale; a smaller bank struggles to accomplish this. Axos Bank,[2] in the U.S., had the smallest PPE in the data I collected from Eikon for 2017 at $21,454,000. JPMorgan Chase[3] had the largest PPE in that data at $1.4934e+10. This is about 696 times larger. In 2017, JP Morgan captured about 21.39% of the market by capital and Axos Bank captured 0.0631% of the market by capital.

Also, due to economies of scale, some of the bigger banks benefit from selling products at a lower price than those offered by the smaller firms, allowing them to attract and serve more customers. This increases their "non-interest income." Non-interest income is derived primarily from fees, including deposit and transaction fees, insufficient funds fees, annual fees, monthly account service charges, inactivity fees, check and deposit slip fees, and so on. Credit card issuers also charge penalty fees such as late fees and over-limit fees.

[1] https://corporatefinanceinstitute.com/resources/knowledge/accounting/ppe-property-plant-equipment/

[2] https://www.axosbank.com

[3] https://www.jpmorganchase.com

Banks charge fees that generate non-interest income as a way of increasing revenue and to ensure liquidity in the event of increased default rates. 'Non-interest income' is one of the output variables of effectiveness shown in Figure 4.3.

How do smaller banks fare in this environment? Martin (2020) presents a case study of the U.S. waste management industry. There were once many smaller waste management companies (garbage collectors) across the country; each had anything from one to several trucks serving customers across a particular route. The profitability of these smaller waste management companies was found to be normally distributed. However, Wayne Huizenga, the founder of Waste Management (WM), realized that if he acquired several routes in a region, then he would have greater purchasing power with truck manufacturers, thus resulting in cheaper trucks. Additionally, a single maintenance facility located in a city hub was far more efficient than several smaller maintenance facilities around a city. This pursuit of greater efficiency by Wayne Huizenga resulted in WM acquiring many smaller waste management companies. Martin (2020) illustrates that this resulted in a Pareto distribution; WM generates more profit in comparison to other waste management companies across the country. A similar Pareto distribution exists in the banking market. As larger banks pursue greater efficiency and effectiveness, they acquire or merge will smaller banks, strengthening the Pareto distribution.

A study by George Mason University (Peirce and Miller, 2015) found that from 2000 to 2014, the number of small banks in the U.S. decreased by 28%. During the financial crisis of 2008, JPMorgan, Bank of America, Wells Fargo, and Citigroup all acquired weaker competitors that had been overexposed to subprime mortgages. Washington Mutual, Bear Stearns, Countrywide Financial, Merrill Lynch, and Wachovia were all acquired by these larger banks during this time (Desjardins, 2019).

Kumar and Gulati (2010) used 'net interest income' as an output dimension of effectiveness in their two-stage DEA model of efficiency and effectiveness. Figure 4.3, illustrates how I use 'net interest income after loan loss provision' as an output dimension of effectiveness. 'Net interest income after loan loss provision' is calculated as 'net interest income' minus 'loan loss provisions,' where 'loan loss provisions' is a buffer that the bank sets aside for any loans that may default. The financial crisis of 2007-2008 resulted in many banks worldwide setting aside such a buffer for loans that may default. Once again, during the COVID-19 pandemic, banks increased their loan loss provisions (Armstrong, 2020; Economist, 2020). Net income that already subtracts loan loss provision is a better indicator of income because it accounts for loans that may default. A model of efficiency and effectiveness that optimizes income without loan loss provisions does not hold well in a financial crisis.

Any bank would like to know whether efficiency and effectiveness affects its Tobin's Q ratio. However, the correlation results from my statistical method do not imply causation. Table 4.3 cannot be used to develop any causal interpretation of efficiency or effectiveness on the Tobin's Q ratio. Dong et al. (2013) mentioned that, in an efficient market, efficiency and effectiveness or its lags cannot affect the SHVCM. Otherwise, arbitrage opportunities exist that will cause traders to use historical information to predict the future Tobin's Q ratio. I use the TYT causal test for inferring cause and effect relationships. In Table 4.5, the null hypothesis that efficiency and effectiveness does not Granger cause the Tobin's Q ratio is rejected for only India. This suggest that the banking market in India is not efficient. As India is the only country where efficiency or effectiveness Granger causes Tobin's Q ratio, this further strengthens my claim that the two-stage DEA model shown in Figure 4.3 is consistent with the semi-strong definition of the EMH for other countries.

Efficiency and the effectiveness are computed using the two-stage DEA model of constant return to scale (CRS). This is consistent with Kumar and Gulati (2010)'s model. Efficiency and effectiveness computed using the input minimization or the output maximization technique of DEA with CRS will result in identical results (Ray, 2008). Therefore, a bank can increase its efficiency by lowering its consumption of efficiency input variables. Additionally, a bank can increase its effectiveness by increasing its generation of effectiveness output variables. Some banks may control both the input variables of efficiency and the output variables of effectiveness. Others may only have control over the input variables of efficiency or only the output variables of effectiveness. For this reason, I investigate (1) whether efficiency Granger causes the Tobin's Q ratio when controlling for effectiveness and (2) whether effectiveness Granger causes the Tobin's Q ratio when controlling for efficiency.

In Table 4.6, using the $\bar{Z}$, efficiency Granger causes the Tobin's Q ratio when controlling for effectiveness. At the country level, Table 4.6 indicates that banks in India can expect efficiency to Granger cause the Tobin's Q ratio when controlling for effectiveness. If banks in India control the input variables of efficiency, i.e., 'total assets' and 'full time employees,' then they can optimize their efficiency; they can expect a corresponding effect on their Tobin's Q ratio.

In Table 4.7, based on the asymptotic critical values of $\bar{Z}$, the null hypothesis is rejected in favor of the alternative hypothesis, i.e., effectiveness Granger causes the Tobin's Q ratio when controlling for efficiency. At the country level, banks in India exhibit causality and can expect effectiveness to Granger cause the Tobin's Q ratio when controlling for efficiency. If banks in India control the output variables of effectiveness, i.e., 'non-interest income' and 'net interest income after loan loss provisions,' then they can optimize their effectiveness and expect an effect on their Tobin's Q ratio. In the next paragraph, I provide supporting evidence of other researchers finding the equity market in India to be inefficient.

Awasthi and Malafeyev (2015) study the efficiency of Bombay's Stock Exchange (BSE). They analyze five popular stock indices, finding that the BSE is inefficient, allowing traders to make excessive returns. The authors provide strong evidence in favor of the inefficient form of the BSE for the years 2005 to 2015. They further suggest that the inefficiency of the Indian stock market could be because the stock market is not well regulated, or may be corrupt. For instance, there may be insider trading, fraud, false statements, or trading abuses occurring within this market.

In Table 4.8 and Table 4.9, I perform reverse causality. In Table 4.8, I check whether the Tobin's Q ratio Granger causes efficiency when controlling for effectiveness. In Table 4.9, I check whether the Tobin's Q ratio Granger causes effectiveness when controlling for efficiency. In Table 4.8, the null hypothesis is rejected in favor of the alternative hypothesis based on the asymptotic critical value of $\bar{Z}$ and $\tilde{Z}$; the Tobin's Q ratio Granger causes efficiency when controlling for effectiveness. At the country level, Canada and Mexico exhibit causality. In Table 4.9, the null hypothesis is rejected in favor of the alternative hypothesis based on the asymptotic critical value of $\bar{Z}$; the Tobin's Q ratio Granger causes effectiveness when controlling for efficiency. At the country level, China exhibits causality. Niederhoffer (1971) mentions that good headlines tend to be followed by more good headlines in the news. Also, bad headlines tend to be followed by further bad headlines. The stock price, the price of an asset, or the Tobin's Q ratio in an equity market factors in news or information in its price determination. For instance, if the Tobin's Q ratio for banks in these countries for t is higher than the previous time $t - x$, then the anticipation of future increase in the efficiency or effectiveness is priced into the current Tobin's Q ratio. This suggests that the Tobin's Q ratio Granger causes effectiveness when controlling for efficiency and that the Tobin's Q ratio Granger causes efficiency when controlling for effectiveness.

4.5 Future work

My statistical method uses a linear family of models to find a relationship between the dependent variable, the Tobin's Q ratio, and the independent variables, efficiency, effectiveness, and their lags. Beccalli et al. (2006); Chu and Lim (1998) have also used linear models when regressing efficiency on the stock price. Before including non-linear models, one must understand how to interpret the economic significance of the coefficients from the model. Martin (2019) mentions that exploring non-linear models is crucial for studying the economy's input and output relationships. However, he does not provide any further insight into which non-linear models to consider and how to interpret their economic significance.

In another paper of mine,[4] I am studying the cause and effect of Glassdoor ratings on stock price, market capital, and the Tobin Q ratio of U.S. banks. I find that the intangible assets like employee satisfaction (I proxy this via Glassdoor rating of the firm) significantly affect stock price, market capital, and the Tobin's Q ratio of some banks. These intangible assets are not captured in banking statements and therefore not considered when building two-stage DEA models. In the future, I would like to explore these additional intangible assets when finding models of efficiency and effectiveness. In this dissertation, I considered banks in Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA from 2000 to 2017. However, some of these countries only have annual rather than quarterly data in Eikon. The motivation for this research is finding a single two-stage DEA model of efficiency and effectiveness that works for banks in most countries and for most time periods. However, Eikon only contains annual data for banks in countries outside North America. In the future, I would like to rerun my variable selection framework on banks from the USA and Canada only because quarterly data is available for these banks from CompuStat.[5]

In this dissertation, I use the CRS model of two-stage DEA of efficiency and effectiveness. This is the same model used by Kumar and Gulati (2010). I use this same model because in this dissertation, I first validate Kumar and Gulati (2010)'s model as not consistent with the semi-strong version of the EMH. I then use the variable selection framework and find a better two-stage DEA model consistent with the semi-strong version of the EMH for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017. Finally in Chapter 5, I provide an optimal path of transforming Kumar and Gulati (2010)'s model into the model recommended by my variable selection framework. Others such as Kao and Hwang (2008) have also used the CRS model in the two-stage DEA model for measuring efficiency to non-life insurance companies in Taiwan. As future work, what are the input and output variables of efficiency and effectiveness with the variable return to scale model? How are these input and output variables different from the CRS model?

[4] I am collaborating with Dr. Dimitrov and Dr. Duimering. We plan to submit this paper to some journal by June 2021

[5] https://wrds-www.wharton.upenn.edu

4.6 Conclusions

In this chapter, I used the variable selection framework of the previous chapter to find a universal two-stage DEA model of efficiency and effectiveness that works for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA for 2000 to 2017. I find that the best two-stage DEA model of efficiency and effectiveness contains (1) total assets and (2) full-time employees as efficiency input variables. Additionally, it contains (1) property/plant/equipment and (2) long-term debt as output variables of efficiency or input variables of effectiveness as ell as (1) non-interest income and (2) net interest income after loan loss provisions as output variables of effectiveness.

As mentioned in Section 4.4.6, the banking market is a Pareto distribution in which a handful of banks in all countries control over 90% of the market value by capital. The result of Table 4.3 suggests that the market values of those banks exhibit continuous improvement in efficiency and effectiveness from year to year. This is because the net cumulative effect of a 1% increase in the efficiency and effectiveness at time period t and effectiveness at time period $t-1$ leads to a .3265% increase in the Tobin's Q ratio. The input and output variables of efficiency and effectiveness from the variable selection framework are more appropriate for these larger banks.

However, correlation does not imply causation. After finding the best two-stage DEA model, I performed the TYT test for causation on six different questions: (1) Does efficiency and effectiveness Granger cause the Tobin's Q ratio? (2) Does the Tobin's Q ratio Granger cause efficiency and effectiveness? (3) Does efficiency Granger cause the Tobin's Q ratio when controlling for effectiveness? (4) Does effectiveness Granger cause the Tobin's Q ratio when controlling for efficiency? (5) Does the Tobin's Q ratio cause efficiency when controlling for effectiveness? (6) Does the Tobin's Q ratio Granger cause effectiveness when controlling for efficiency? I find that in India, efficiency Granger causes Tobin's Q ratio when controlling for effectiveness. I also find in India, effectiveness Granger causes Tobin's Q ratio when controlling for efficiency. A bank in India may not control all the input and output variables of efficiency and effectiveness that I report in Section 4.4. For instance a bank may only control the input variables of efficiency or the output variables of effectiveness. My recommendation from this chapter for banks in India is that a bank that only controls the input variables of efficiency can increase its efficiency by lowering (or optimizing) its consumption of input variables of efficiency. This will cause an effect in the Tobin's Q ratio. My other recommendation from this chapter for banks in India is that a bank that only controls the output variables of effectiveness can increase its effectiveness by increasing (or optimizing) its output variables of effectiveness. This will cause an effect in the Tobin's Q ratio. More details of my recommendations are provided in Section 4.4.6 of this chapter.

Chapter 5

Building an optimal path for transforming a two-stage DEA model not consistent with semi-strong EMH to one that is

5.1 Introduction

Due to intense competition for scarce resources, many domestic and international firms are searching for more efficient and effective ways of managing their operations. As measured by efficiency and effectiveness, every firm wants to achieve high performance. Firms strive to be highly efficient (i.e., doing things right) and highly effective (i.e., doing the right things). What is also desirable is that any improvement in efficiency or effectiveness leads to higher shareholder equity. This is because financial markets can be viewed as a measure of the wisdom of crowds (Surowiecki, 2005). The measure could be the stock price or some other shareholder value creation metric (SHVCM), such as Tobin's Q ratio. Although other factors may impact a firm's SHVCM, if a firm is deemed efficient and effective by market traders, it tends to perform financially better than firms that are either inefficient or ineffective, or both.

For most firms, though, an improvement in performance does not lead to an increase in shareholder equity. For instance, Kaplan et al. (2005), in their article in the Harvard Business Review, report that from 1987 to 1990, a certain NYSE electronics company improved its quality and on-time delivery. Its defect rate dropped from 500 parts per million to 50, and on-time delivery improved from 70% to 96%. However, these performance improvements did not result in an increase in shareholder value, since its stock price fell to a third of its July 1987 value. The operational improvements the company had made in performance were real. However, these did not correlate with any change in its stock price.

Associating a firm's performance with its stock price or some other SHVCM has other benefits as well. In the Economist (Economist, 2012), Hermann Stern, the CEO of Obermatt, said that earnings growth and shareholder return should determine how much a CEO should be paid. Remunerating a CEO and managers based solely on their firm's performance is flawed and incorrect. In another article in the Economist (Economist, 2003), Alan Dunn, an instructor at Caltech, describes a 2-day seminar titled "Measuring Business Performance: Aligning Strategy, Metrics, and Rewards." In this seminar, Mr. Dunn gives recommendations and advice on measuring performance using dimensions such as cash flow instead of profit, because these are better correlated with shareholder equity.

Perhaps, then, the NYSE firm mentioned above was using the wrong dimensions for measuring its performance. Often, firms use dimensions for measuring performance that are biased or constrained by managers' prior views of what drives performance (Silvestro, 2016). Managers make assumptions about the relation between performance, customer loyalty, and profitability, even when these presumed links have not been tested. In a study (Ittner and Larcker, 2003), as also noted by Silvestro (2016), only 21% of managers had tested the dimensions of performance, and many of those who had tested them found that their assumed dimensions of performance were incorrect. Such incorrect dimensions of performance lead to very little or no significant impact on SHVCM. Instead, perhaps, the NYSE firm should first learn how to change its existing model of measuring performance using newer measures that are consistent with its stock price.

Gallivan et al. (1994) identified two dimensions of management change. The first is the scope of the change, i.e., whether the change is radical or incremental. The second dimension is the pace of the change measured across time, i.e., whether the change is rapid or gradual. Gallivan et al. (1994) defines radical change as replacing the status quo with a new order, which may result in serious disruption to structures, processes, operations, knowledge, and morale. Jobs are altered or eliminated, skills are gained or lost, the information flow is redefined and rerouted, processes are transformed and created, responsibilities are transferred, and power bases are undermined. On the other hand, in incremental change, established structures, processes, and knowledge are extended and augmented. Incremental change is not as disruptive as radical change. Tushman et al. (1986) note that resistance to change is natural. An organization can introduce radical change gradually or rapidly. If done gradually, the radical change will be accepted more readily with little to no resistance.

For example, Orlikowski (1993) discusses an organization that introduced radical change by implementing new software. The radical change consisted of using new programming languages and new software development methodologies.

Hage and Aiken (1970) recommend implementing radical change gradually, as this allows more time to be allocated for trial and error, and thus, the intended objectives are more likely to be fulfilled. Rogers (2003) mentions that if rapid change is introduced quickly, the results can be disastrous because the organization, especially the employees, may not have enough time to adapt to the changes.

I hypothesize that a set of incremental changes that transform the two-stage DEA model of Kumar and Gulati (2010) into a model consistent with the semi-strong version of EMH will be more readily adopted. Instead of telling a bank about the best two-stage DEA model of efficiency and effectiveness, it would be more meaningful for the bank to know what path of incremental changes will transform its current model of measuring efficiency and effectiveness into the best model. Implementing the changes incrementally would allow a bank to optimize and act on the dimensions one at a time, leading to little resistance, unlike a rapid and large-scale overhaul of its model in which it would have to change multiple dimensions, thus leading to widespread disruption. The intended audience for the work done in this chapter is a bank hesitant to adopt a big radical change. A bank can instead adopt incremental changes as suggested in the optimal path. The incremental change promises to yield immediate improvement, incentivizing the bank to adopt further changes on the path. I address the following research questions in this chapter:

1. How should I define incremental change in the context of a two-stage DEA model of efficiency and effectiveness?

   Incremental change in this chapter is defined as applying a single elementary operation on a two-stage DEA model of efficiency and effectiveness. I define three elementary operations: (1) the addition of a dimension, (2) the removal of a dimension, and (3) swapping two dimensions. Given a candidate two-stage DEA model of efficiency and effectiveness, one can generate a neighborhood of two-stage DEA models of efficiency and effectiveness in which each model in the neighborhood is separated by one incremental change from the candidate model. Not all elementary operations are consistent with the semi-strong EMH.

2. What path or actionable items must a firm take to transform from an existing model of efficiency and effectiveness that is inconsistent with the semi-strong definition of the EMH to a model that is consistent with the semi-strong definition of the EMH?

A path is defined as a set of elementary operations between the existing model, which is not consistent with the semi-strong EMH, to a model that is. Kumar and Gulati (2010)'s two-stage DEA model of efficiency and effectiveness is not consistent with the semi-strong version of the EMH, as already seen in the results of Chapter 2 and Section 2.5.2. On the other hand, the two-stage DEA model of efficiency and effectiveness, from my variable selection framework in Chapter 4, is. In Section 5.3.4 of this chapter, I present an algorithm for generating an optimal path from Kumar and Gulati (2010)'s model to the model in my variable selection framework. Later in Section 5.4, I present the result of the optimal path. The optimal path is defined as the shortest path with the least number of elementary operations that are not consistent with the semi-strong EMH. For example, consider two shortest paths with a length of 7. Each path has 7 elementary operations. In the first path, 5 of the 7 elementary operations are consistent with the semi-strong EMH. In the second path, 3 of the 7 elementary operations are consistent with the semi-strong EMH. Thus, the first path is preferred over the second path in our path-finding algorithm because it has more elementary operations consistent with the semi-strong EMH.

In the remainder of this chapter, I first discuss related work in Section 5.2. In Section 5.3, I present the algorithms of (1) for generating a neighborhood using elementary operations and (2) for finding the optimal path. In Section 5.4, I present the results of this chapter, and finally, in Section 5.5, I state my conclusions.

5.2 Related work

Like Gallivan et al. (1994), Del Val and Fuentes (2003) distinguish between radical change and incremental change. Del Val and Fuentes (2003) refer to radical change as strategic change and incremental change as evolutionary change. The latter introduces small changes that improve the present situation while keeping the general framework more or less consistent. This idea of incremental or evolutionary change is also echoed by others (Blumenthal and Haspeslagh, 1994; Goodstein and Burke, 1991; Mezias and Glynn, 1993). In contrast, strategic change is revolutionary and more transformational (Blumenthal and Haspeslagh, 1994; Ghoshal and Bartlett, 1996). Del Val and Fuentes (2003) further mention that resistance to change is generally higher for strategic than for evolutionary change.

Researchers agree that radical change is vastly different from incremental change. However, how does one quantitatively separate radical change from incremental change? In this dissertation, I define incremental change as applying a single elementary operation to a two-stage DEA model of efficiency and effectiveness. Radical change consists of applying more than one such instance of incremental change to a two-stage DEA model of efficiency and effectiveness. The best two-stage DEA model of efficiency and effectiveness as shown in Figure 4.3 is 7 incremental steps away from Kumar and Gulati (2010)'s two-stage DEA model as seen in Figure 2.1. 3 addition of new variables and 4 removal of old variables are required for changing Kumar and Gulati (2010)'s two-stage DEA model into the best two-stage DEA model. This change is then classified as a radical change. The radical change or the optimal path that I present in Section 5.4 is given as an optimal path of 7 incremental steps. My search algorithm finds an optimal path of 7 elementary operations, out of which 5 are consistent with the semi-strong EMH. Other paths of length 7 have less than 5 elementary operations consistent with the semi-strong EMH. As a sanity check, I also checked all two-stage DEA models that are at most two elementary steps away from Kumar and Gulati (2010)'s model, and none of these are consistent with semi-strong EMH. This suggests no path of length greater than seven contains at most two incremental steps not consistent with the semi-strong EMH.

Del Val and Fuentes (2003) mention that when an organization undergoes incremental change, there is little to no resistance. By breaking radical change into a series of incremental changes, I hypothesize that each incremental change in the optimal path is well received by banks that want to gravitate towards a model of efficiency and effectiveness consistent with the semi-strong definition of EMH. As mentioned previously, I define three elementary operations: (1) the addition of a dimension, (2) the removal of a dimension, and (3) swapping two dimensions. The motivation for defining three elementary operations is that Simar and Wilson (2008) mention that a parsimonious model is desirable since the discriminating power of DEA decreases as the number of dimensions increases. I do not explore other operations, such as adding more than one variable or removing more than one variable at a time. A parsimonious DEA model is desirable since it has as many input and output variables as needed but as few as possible (Jenkins and Anderson, 2003; Wagner and Shimshak, 2007).

Others (Adler and Golany, 2001; Jenkins and Anderson, 2003; Pastor et al., 2002; Ruggiero, 2005; Simar and Wilson, 2001; Ueda and Hoshiai, 1997) in the literature have talked about variable selection in DEA. The variable selection framework proposed by these authors is designed to improve efficiency or introduce more discrimination. For instance, Pastor et al. (2002) developed an efficiency contribution measure (ECM) for a single dimension. The approach recommended by the authors is first to compute the efficiency by including the dimension of interest and then comparing the overall efficiency with another DEA for which the dimension of interest is absent. A statistical test is used to infer whether to include or reject the dimension of interest. Ruggiero (2005) set up a regression where the dependent variable is the efficiency computed from DEA, and the independent variables are candidate variables for input into or output from DEA. The coefficients of the regressors from the regression, i.e., whether the coefficient is positive for an input variable and negative for an output variable, are used in deciding which of the candidate variables to select for two-stage DEA. Simar and Wilson (2001) use bootstrapping and outline a statistical procedure on how best to find relevant input and output variables. Jenkins and Anderson (2003) use a variable-reduction approach that removes variables containing minimum information, which is defined via a partial correlation. The removal of correlated variables can have a significant impact on the efficiency computed from DEA. Wagner and Shimshak (2007) present a stepwise approach to variable selection that involves sequentially maximizing (or minimizing) the average change in the efficiency as variables are added to or dropped from the DEA model. Titko et al. (2014) built 14 DEA models for measuring bank efficiency in Latvia. These 14 DEA models use different combinations of input and output variables. The authors performed the two-sample Kolmogorov–Smirnov test to identify which of these 14 models are substantially different from the others. The authors then provide general recommendations on variable selection for DEA in the Latvian banking sector. Sigala et al. (2004) observed a paradox in the tourism industry. Due to increased competition in the industry, Sigala et al. (2004) expected that productivity would have increased. However, this was not the case. Most likely, the lack of productivity was because incorrect variables were used to measure performance. The authors present a stepwise approach for selecting variables for measuring performance consistent with the increase in competition in the tourism industry. Others, such as Aydın et al. (2020); Chen et al. (2017); Li and Cui (2017), present different techniques for evaluating performance in the airline industry using DEA. These authors also present guidelines on how to remove undesirable output variables from DEA. For instance, Aydın et al. (2020) first select inputs and outputs based on a literature survey. For each output, the most significant inputs are selected using a stepwise backward variable regression analysis. The authors then use a social network analysis to identify which output variables for which efficient airline companies are leaders and role models in terms of input variables selected by the stepwise regression analysis. The authors provide a methodology for finding input and output variables best suited to the airline industry.

In this dissertation, I use my variable selection framework of Chapter 3 to find a two-stage DEA model of efficiency and effectiveness for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA for 2000-2017. Furthermore, each dimension and its orientation (whether as an input of efficiency, an output of efficiency, or an output of effectiveness) within the two-stage DEA model is validated against the semi-strong version of EMH using the statistical method presented in Chapter 2. The above researchers do not recommend what incremental changes an airline or a bank must make to transform from its current model of measuring efficiency to the model suggested by them. This is what I do in this chapter. I present an optimal path, which consists of incremental changes that transform a model not consistent with semi-strong EMH to one that is.

5.3 Methodology

In Section 5.3.1, I describe the elementary operations in more detail. In Section 5.3.2, I outline the algorithm which generates a neighborhood of two-stage DEA models of efficiency and effectiveness using these elementary operations. In Section 5.3.3, I present the algorithm for finding the optimal path from a two-stage DEA model of efficiency and effectiveness not consistent with the semi-strong EMH to one that is.

5.3.1 Elementary operations

Three elementary operations can be performed on a certain two-stage DEA model (D_{old}) of efficiency and effectiveness:

1. Add variable

   The *add variable* operation adds a new variable, N_v to D_{old} resulting in three new two-stage DEA models of efficiency and effectiveness. N_v is chosen from the Eikon dataset such that $N_v \in E - D_{old,vs}$ where E is the set of all variables from the Eikon dataset and $D_{old,vs}$ is the set of all variables from the two-stage DEA model of efficiency and effectiveness D_{old}.

   In the *add variable* operation, three new two-stage DEA models of efficiency and effectiveness are created by (1) adding N_v either as an input variable of efficiency, (2) as an output variable of efficiency, or (3) as output variable of effectiveness.

2. Remove variable

   The *remove variable* operation removes an existing variable from D_{old}. The variable selected for removal (R_v) is chosen from $D_{old,vs}$. $D_{old,vs}$ is the set of all variables from the two-stage DEA model of efficiency and effectiveness, D_{old}. The removal of R_v from $D_{old,vs}$ will result in one new two-stage DEA model of efficiency and effectiveness.

3. Swap variable

   The *swap variable* operation does not add any new variables from E to $D_{old,vs}$ or removes any variables from $D_{old,vs}$. However, the swap variable operation selects two variables ($S_i \in D_{old,i}$ and $S_j \in D_{old,j}$) where $i \in \{1, 2, 3\}$ and $j \in \{1, 2, 3\}$ and $i \neq j$. i and j indicates whether the variable, S belongs in the set of input variables of efficiency, or the output variables of efficiency or the output variables of effectiveness. $D_{old,i}$ is the set of all variables from the two-stage DEA model of efficiency and effectiveness that belongs in orientation i. A two-stage DEA model of efficiency and effectiveness has three orientations. When $i = 1$, $D_{old,i}$ refers to the set of efficiency input variables. When $i = 2$, $D_{old,i}$ refers to the efficiency output or effectiveness input variables. When $i = 3$, $D_{old,i}$ refers to the effectiveness output variables. A swap operation results in a new two-stage DEA model of efficiency and effectiveness where variables $S_{1,i}$ and $S_{1,j}$ swap their variables.

5.3.2 Neighborhood generation algorithm

Using the elementary operations from Section 5.3.1, I will now describe how to generate a neighborhood of two-stage DEA models of efficiency and effectiveness around Kumar and Gulati (2010)'s model. I refer to Kumar and Gulati (2010)'s model in this section as an example. However, the algorithm can generate the neighborhood around any two-stage DEA model of efficiency and effectiveness.

The two-stage DEA model from Kumar and Gulati (2010) is called $Model_{K\&G}$ where $Model_{K\&G} \in \mathbb{B}^{3\times 55}$. 55 is the number of columns which is the total number of dimensions from Eikon. 3 is the number of rows, representing the three different orientations of the two-stage DEA model. Row one refers to the input efficiency variables, row two refers to output efficiency variables or the effectiveness input variables, and row three represents output effectiveness variables. Each entry in the matrix is a binary number, i.e., $\mathbb{B}$. For example, the value 1 in row 3, column j indicates that the dimension j from Eikon is present as an output variable of effectiveness in the two-stage DEA model. Any zero-vector column j in $Model_{K\&G}$ indicates the absence of dimension j in the two-stage DEA model.

The neighborhood generation algorithm returns a neighborhood of two-stage DEA models around Kumar and Gulati (2010)'s model using the elementary operations. In the post-condition of Algorithm 6, $Neighborhood$ is returned where $Neighborhood \in \mathbb{B}^{3\times 55\times N}$ and N is the size of the neighborhood. The algorithm has three distinct stages; there is one stage for each of the three different elementary operations.

1. In the first stage, from line 3 to line 7 of Algorithm 6, I build the linear equality constraints of the form $Ax = b$ for the addition elementary operation. The linear equality constraints are set such that the existing variables in $Model_{K\&G}$ remain fixed. However, the sum of variables selected from the set, $E - Model_{K\&G,vs}$, where E is the set of all dimensions from Eikon, and $Model_{K\&G,vs}$ is the set of all variables from Kumar and Gulati (2010)'s two-stage DEA model must be equal to 1 to reflect the addition of a new variable in $Model_{K\&G}$. I also set the inequality constraints for addition such that all the rows of column j in $Model_{K\&G}$ must sum to at most 1 where $1 \leq j \leq 55$. Suppose variable j is present in the model. In that case, it can occur in only one of the three orientations of the model. In other words, a variable cannot occur as both input and as output. The three different orientations in the two-stage DEA are the (1) the input variable of efficiency, (2) the output variable of efficiency and (3) the output variable of effectiveness.

2. In the second stage, from line 8 to line 12 of the algorithm, I build the linear equality constraints of the form $Ax = b$ for the removal elementary operation. The sum of variables selected from the set, $E - Model_{K\&G,vs}$, where E is the set of all dimensions from Eikon and $Model_{K\&G,vs}$ is the set of all variables from Kumar and Gulati (2010)'s model is equal to 0. This prevents any dimension not already present in $Model_{K\&G,vs}$ from being considered. In addition, I also set the equality constraints such that any one variable can be removed from all the variables in the set $Model_{K\&G,vs}$.

3. In the third stage, from line 14 to line 20 of the algorithm, I build the linear equality constraints of the form $Ax = b$ for the swap elementary operation. The sum of variables selected from the set, $E - Model_{K\&G,vs}$, where E is the set of all dimensions from Eikon, and $Model_{K\&G,vs}$ is the set of all variables from Kumar and Gulati (2010)'s model is equal to 0. I also set the equality and the inequality constraints such that for every two variables in $Model_{K\&G,vs}$, a swap is permissible as long as the two variables are in two different orientations in $Model_{K\&G}$.

For each stage in the algorithm, I use IBM CPlex[1] to generate the neighborhood based on the equality and the linear inequality constraints.

[1] https://www.ibm.com/analytics/cplex-optimizer

Algorithm 6 Neighborhood Generation Algorithm

Preconditions:

- Two-Stage DEA model of Kumar and Gulati (2010): $Model_{K\&G} \in \mathbb{B}^{3\times55}$
  where the number of columns, 55, is the number of dimensions and the number of rows, 3, represents the three different positions of the two-stage DEA model, i.e., row one represents input efficiency variables, row two represents output efficiency variables, and row three represents output effectiveness variables.

Postconditions:

- Solution: $Neighborhood \in \mathbb{B}^{3\times55\times N}$
  where N is the size of the Neighborhood.

procedure GENERATENEIGHBORHOOD($Model_{K\&G}$)
    Neighborhood← empty
    $A_{eq,addition} \leftarrow$ setAEqualityLinearConstraintsOfAddition($Model_{K\&G}$)
    $b_{eq,addition} \leftarrow$ setbEqualityLinearConstraintsOfAddition($Model_{K\&G}$)
    $A_{ineq,addition} \leftarrow$ setAInEqualityLinearConstraintsOfAddition($Model_{K\&G}$)
    $b_{ineq,addition} \leftarrow$ setbInEqualityLinearConstraintsOfAddition($Model_{K\&G}$)
    $neighborhoodFromAddition$ ←runCplex($A_{eq,addition}$,$b_{eq,addition}$,$A_{ineq,addition}$,$b_{ineq,addition}$)

Neighborhood Generation of removal of one variable

$A_{eq,removal} \leftarrow$ setAEqualityLinearConstraintsOfRemoval($Model_{K\&G}$)
$b_{eq,removal} \leftarrow$ setbEqualityLinearConstraintsOfRemoval($Model_{K\&G}$)
$A_{ineq,removal} \leftarrow$ setAInEqualityLinearConstraintsOfRemoval($Model_{K\&G}$)
$b_{ineq,removal} \leftarrow$ setbInEqualityLinearConstraintsOfRemoval($Model_{K\&G}$)
$neighborhoodFromRemoval$ ←runCplex($A_{eq,removal}$,$b_{eq,removal}$,$A_{ineq,removal}$,$b_{ineq,removal}$)
$\vdots$

Neighborhood Generation of swap of one variable

$A_{eq,swap} \leftarrow$ setAEqualityLinearConstraintsOfSwap($Model_{K\&G}$)
$b_{eq,swap} \leftarrow$ setbEqualityLinearConstraintsOfSwap($Model_{K\&G}$)
$A_{ineq,swap} \leftarrow$ setAInEqualityLinearConstraintsOfSwap($Model_{K\&G}$)
$b_{ineq,swap} \leftarrow$ setbInEqualityLinearConstraintsOfSwap($Model_{K\&G}$)
$neighborhoodFromSwap \leftarrow$runCplex($A_{eq,swap}$,$b_{eq,swap}$,$A_{ineq,swap}$,$b_{ineq,swap}$)
Neighborhood $\leftarrow$ $neighborhoodFromAddition \cup neighborhoodFromRemoval \cup neighborhoodFromSwap$
return Neighborhood

5.3.3 Path finding algorithm

The path finding algorithm finds all paths from Kumar and Gulati (2010) to the target two-stage DEA model. The target two-stage DEA model in Algorithm 7 is the best two-stage DEA model from Chapter 4; it was found by executing the variable selection framework on banks from Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA from 2000-2017. *Steps*, another precondition in the algorithm, refers to the maximum number of elementary operations each path from Kumar and Gulati (2010)'s model to the target two-stage DEA model can contain. I set the *Steps* to seven[2] when running the algorithm. 7 is the least number of steps required to move from $Model_{K\&G}$ to $Model_{target}$, i.e., there are 4 removal of old variables and 3 addition of new variables required on $Model_{K\&G}$ for transforming it to $Model_{target}$.

The path finding algorithm returns all paths referred to as $Paths \in \mathbb{I}^{Steps\times N}$ and $DEAModels \in \mathbb{B}^{3\times 55\times Steps\times N}$ in the algorithm. When the algorithm is first called, $Model_{start}$ on line 1 is set to $Model_{K\&G}$. On line 8, the path finding algorithm generates a neighborhood of two-stage DEA models of efficiency and effectiveness using Algorithm 6 for $Model_{start}$. For each of the neighbors in the neighborhood, if an intermediate path from the neighbor to $Model_{target}$ exists, then an attempt is made to determine what elementary operation connects $Model_{K\&G}$ to the neighbor which then completes the path from $Model_{K\&G}$ to $Model_{target}$ (see line 14 to 17 of the algorithm). This is a recursive algorithm. For each step along the path from $Model_{K\&G}$ to $Model_{target}$, the algorithm tracks

[2]because there are no two-stage DEA model that is consistent with semi-strong that is at most two steps away from Kumar and Gulati (2010)'s model. This implies that any path from Kumar and Gulati (2010)'s model to the best model will have atleast two steps that are not consistent with the semi-strong EMH.

the two-stage DEA models of efficiency and effectiveness (referred to as $DEAModels$ in Algorithm 7).

Algorithm 7 Path Finding Algorithm

Preconditions:

- Two-Stage DEA model of Kumar and Gulati (2010): $Model_{K\&G} \in \mathbb{B}^{3\times55}$
- Two-Stage DEA target model $Model_{target} \in \mathbb{B}^{3\times55}$
- Steps: Maximum number of steps for each path.

Postconditions:

- Solution: $Paths \in \mathbb{I}^{Steps\times N}$ and $DEAModels \in \mathbb{B}^{3\times55\times Steps\times N}$ where N is the number of paths found.

procedure GENERATEPATHS($Model_{start}$,$Model_{target}$,$steps$)
  if $Model_{start} = Model_{target}$ **then**
    paths=padZerosOfLength(steps)
    deaModelsOnPaths=padZerosOfLength(steps)
  else if steps=0
     return NO_PATH,NO_DEA
  else
  neighborhood ← GenerateNeighborhood($Model_{start}$)
  completePaths← Nil
  DEAModels← Nil
  for each *neighbor in neighborhood* **do**
    intermediatePaths,intermediateDEAModels ← GeneratePaths(neighbor,$Model_{target}$,steps-1)
    for each *partialValidPath in intermediatePaths* **do**
      completePath ←generateCompletePathByAddingValidElementaryOperation(partialValidPath,neighbor)
      DEAModel ←generateCompleteDEAInPaths(neighbor,getDEA(partialValidPath,intermediateDEAModels))
      completePaths←completePaths+completePath
      DEAModels←DEAModels+DEAModel
  return completePaths,DEAModels

5.3.4 Optimization and finding the best path

Given all the paths from $Model_{K\&G}$ to $Model_{target}$ from the path finding algorithm, Algorithm 7, I now find the optimal path from all the available paths using the Algorithm 8. To find the optimal path, all paths are first generated on line 2 of the Algorithm 8. Each path consists of a set of elementary operations. For each elementary operation, the resultant two-stage DEA model of efficiency and effectiveness is represented as *dea* on line 6 of Algorithm 8. The efficiency and effectiveness are generated on line 7. Next, the statistical method is run on line 8. On line 9, if the resultant two-stage DEA model of efficiency and effectiveness is consistent with the semi-strong version of the EMH, then that elementary operation is saved. On line 12 and line 13, the algorithm gets all paths with the least number of steps not consistent with the EMH. From these paths, the algorithm returns the set of path with the least number of steps. If the total number of such paths is one (see line 14) then the path is an optimal path. However, if there are multiple paths, then I compute the average AIC along all the elementary operations; the optimal path is the one with the lowest average AIC.

Algorithm 8 Find Best Path Algorithm

Preconditions:

- Two-Stage DEA model of Kumar and Gulati (2010): $Model_{K\&G} \in \mathbb{B}^{3\times 55}$
- Two-Stage DEA target model $Model_{target} \in \mathbb{B}^{3\times 55}$
- Steps: Maximum number of steps to run for each path before termination.

Postconditions:

- Solution: $Path \in \mathbb{I}^{Steps}$ and $DEAModel \in \mathbb{B}^{3\times 55\times Steps}$
  where N is the number of paths found.

```
procedure findBestPath
    paths,deas ← GeneratePaths(Model_{K&G},Model_{target},steps=7)
    pathNumber ← 0
    for each path in paths do
        stepNumber ← 0
        for each dea in deas do
            efficiency,effectiveness ← calculateDEAScores(data,dea)
            aic,isConsistentWithEMH ← statisticalMethod(
              efficiency,effectiveness,tobinQ)
            results ← populate(stepNumber,pathNumber,aic,isConsistentWithEMH)
            stepNumber ← stepNumber+1
        pathNumber ← pathNumber+1
    solutions ← getPathsWithLeastNumberOfStepsNotConsistentWithEMH(results)
    solutions ← getShortestPath(solutions)
    if len(solutions)=1 then
        best_path ← solutions[0]
        return solutions[0]

    best_path=compute_average_aic_for_paths_and_return_best(solutions)

    return best_path
```

5.4 Results and discussions

Using Algorithm 6, I found that there is no two-stage DEA model of efficiency and effectiveness that is two elementary operations away from $Model_{K\&G}$. This suggests that any path from $Model_{K\&G}$ to $Model_{target}$ will have at least two steps that are not consistent with the semi-strong definition of the EMH. $Model_{target}$ refers to the two-stage DEA model shown in Figure 4.3. $Model_{target}$ is 7 elementary operations away from $Model_{K\&G}$; three variables are added ($set_{target} - set_{K\&G}$) and four variables variables are removed ($set_{K\&G} - set_{target}$) where $set_{K\&G}$ refers to the set of variables present in $Model_{K\&G}$ and set_{target} refers to the set of variables present in $Model_{target}$.

The goal is to find the shortest path from $Model_{K\&G}$ to $Model_{target}$ that contains the greatest number of steps that are consistent with the semi-strong definition of the EMH. I use Algorithm 7 and find all paths from $Model_{K\&G}$ to $Model_{target}$ that consists of 7 steps. In the generation of intermediate two-stage DEA in a single path, I modify the constraint number (3) in Chapter 3 and Section 3.3.1 to allow the inclusion of an extra variable. I also modify the constraint (2) in the same section for the sum of the entries in each row of the matrix *Model* in Section 3.3.1 can be greater than or equal to 1. These couple of modifications are required in order to generate the intermediate two-stage DEA. From Algorithm 8, the optimal path contains 5 elementary operations consistent with the semi-strong version of the EMH. This path recommends the following elementary operations from $Model_{K\&G}$ to $Model_{target}$:

1. Perform *removal* of variable 'net loans' from efficiency input.
2. Perform *addition* of variable 'long term debt' to efficiency output.
3. Perform *removal* of variable 'long term investments' from efficiency output.
4. Perform *addition* of variable 'net interest income after loan loss provision' to effectiveness output.
5. Perform *removal* of variable 'net interest income' from effectiveness output.
6. Perform *addition* of variable 'property/plant/equipment' to efficiency output.
7. Perform *removal* of variable 'other earning assets' from efficiency output.

Steps (3), (4), (5), (6) and (7) result in two-stage DEA models of efficiency and effectiveness that are consistent with the semi-strong version of the EMH. The output from the statistical method for each of these steps is presented in Table 5.1, Table 5.2, Table 5.3, Table 5.4, and Table 5.5.

Table 5.1 Output from the statistical method after performing step 3

N=128	$AIC = -47.0056$
n=8	T=16
$R^2 = .04000$	Adj $R^2 = -.05104$
Wald F(4,7)=7.907589	p-value = .0098

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$log(efficiency_t)$	.906199	.388465	2.3328	0.052
$log(effectiveness_t)$	.413602	.120883	3.4215	.011
$log(effectiveness_{t-1})$	-1.044003	.216526	-4.8216	.002
$log(efficiency_{t-2})$	-1.058600	.736042	-1.4382	.194

Standard errors robust to heteroskedasticity adjusted for 8 clusters

Table 5.2 Output from the statistical method after performing step 4

N=128	$AIC = -48.2813$
n=8	T=16
$R^2 = .05501$	$AdjR^2 = -.03461$
Wald F(4,7)=6.585157	p-value = .0160

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$log(efficiency_t)$	.899877	.299534	3.0043	.020
$log(effectiveness_t)$	.375271	.093054	4.0328	.005
$log(effectiveness_{t-1})$	-.585199	.174111	-3.3611	.012
$log(efficiency_{t-2})$	-1.131958	.728598	-1.5536	.164

Standard errors robust to heteroskedasticity adjusted for 8 clusters

Table 5.3 Output from the statistical method after performing step 5

N=128	AIC=-50.8363
n=8	T=16
$R^2 = .06438$	$AdjR^2 = -.02435$
Wald F(4,7)=5.783737	p-value = .0223

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$log(efficiency_t)$	1.198015	.398183	3.0087	.020
$log(effectiveness_t)$	.388275	.095883	4.0495	.005
$log(effectiveness_{t-1})$	-.598528	.184370	-3.2463	.014
$log(efficiency_{t-2})$	-1.331313	.654118	-2.0353	.081

Standard errors robust to heteroskedasticity adjusted for 8 clusters

Table 5.4 Output from the statistical method after performing step 6

N=128	$AIC = -55.5372$
n=8	T=16
$R^2 = .08287$	$AdjR^2 = -.00410$
Wald F(4,7)=9.245219	p-value = .0063

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$log(efficiency_t)$	.905135	.247911	3.6511	.008
$log(effectiveness_t)$	.490411	.094993	5.1626	.001
$log(effectiveness_{t-1})$	-.684544	.162889	-4.2025	.004
$log(efficiency_{t-2})$	-.579604	.398737	-1.4536	.189

Standard errors robust to heteroskedasticity adjusted for 8 clusters

Table 5.5 Output from the statistical method after performing step 7, resulting in the same two-stage DEA model of Table 4.3

$N = 128$	AIC=-63.1407
n=8	$T = 16$
$R^2 = .11594$	$Adj R^2 = .03211$
Wald F(4,7)=18.146207	p-value = .0008

Variable	**Coefficient**	**Cluster Standard Error**	**t-stat**	**p-value**
$\log(efficiency_t)$	.588653	.242731	2.4251	.046
$\log(effectiveness_t)$	.465596	.169885	2.7406	.029
$\log(effectiveness_{t-1})$	-.727668	.111479	-6.5274	.000
$\log(efficiency_{t-2})$	-.744820	.426930	-1.7446	.125

Standard errors robust to heteroskedasticity adjusted for 8 clusters

One can view $Model_{target}$ as a radical change from $Model_{K\&G}$ due to the 7 elementary operations that separate the two models. Rose and Lawton (1999) observe that changes in organizations arise from the need to become more efficient and effective. Recommending that banks use $Model_{target}$ without a clear path to move from their current model to $Model_{target}$ may be met with resistance. In this section, I provide an optimum set of recommendations or incremental changes that transform $Model_{K\&G}$ to $Model_{target}$.

HBR (2018) interviewed managers of banks in 14 countries. They learned that, for any change to be effective, there must be consistent measurement and monitoring of whether the new change is improving efficiency and effectiveness. By providing an optimal path of incremental changes, banks can measure whether adding a new variable of efficiency or effectiveness leads to improvement. Each incremental change in the optimal path is a single elementary operation, giving banks the option to monitor and measure each incremental change before moving onto the next change in the optimal path. Bank management requires time to not only understand why these new metrics are required but also time to understand how these metrics lead to improvement in the Tobin's Q ratio.

HBS (2002) mentions that when a radical change is broken down into small incremental changes, it is easier to manage. Small changes have a greater probability of success; any disruptions are more easily addressed. Rousseau (1989) notes that when a change is introduced in a step-by-step fashion, stakeholders have more confidence about adopting the change.

Some banks, for instance, may only control the input variables of efficiency or the output variables of effectiveness. Banks can use their innate knowledge of what variables they control best, perhaps combining two or more incremental steps in the optimal path into a single step. The goal is to reach $Model_{target}$.

5.5 Conclusions

In conclusion, my optimal path of recommendation from the previous section aligns with Raffaelli (2015)'s roadmap to change. Raffaelli (2015) outlines the following steps for managing change successfully:

1. Diagnosis: Why is change needed?

   $Model_{K\&G}$ is not consistent with the semi-strong definition of the EMH. On the other hand, $Model_{target}$ is consistent with the semi-strong definition of the EMH. Table 5.5 suggests that a 1% increase in efficiency and a 1% increase in effectiveness are correlated with an increase in the Tobin's Q ratio.

2. Design: What sort of change is called for?

   Change management literature has noted resistance to change which is perceived as radical. However, making changes incremental reduces this resistance. I define radical change and incremental change in Section 5.2 of this chapter. I provide a path of incremental change from $Model_{K\&G}$ to $Model_{target}$ containing the least number of elementary operations. Furthermore, the path that I recommend contains the greatest number of steps that are consistent with the semi-strong definition of the EMH.

3. Delivery: How can change best be implemented?

   The change can best be implemented by performing the elementary operations mentioned in Section 5.4 that will transform $Model_{K\&G}$ to $Model_{target}$.

4. Evaluation: How can the impact of the change be assessed and measured?

   Five of the seven steps in the path that I recommend are consistent with the semi-strong definition of the EMH. Each of these five steps result in incremental changes to input and output variables of efficiency and effectiveness, resulting in improved models. The efficiency and effectiveness measures in these newer models are correlated with Tobin's Q ratio. Each step is an improvement, resulting in a model that better captures the performance of the market. Finally, the last step of the path results in $Model_{target}$. As per Chapter 4 and Section 4.4 of this dissertation, $Model_{target}$ is the best model according to my statistical method.

Chapter 6

Conclusion

In this thesis, I enhanced my statistical method from my M.ASc by adding new features to it. These new features are documented in Chapter 2. The statistical method validates whether a quantitative model of efficiency and effectiveness is consistent with the semi-strong definition of the EMH. I find that the two-stage DEA model of efficiency and effectiveness proposed by Kumar and Gulati (2010) is not consistent with the semi-strong definition of the EMH.

In Chapter 3, I study how one can find a quantitative model of efficiency and effectiveness consistent with the semi-strong version of the EMH. I use the same definition of a two-stage DEA model of efficiency and effectiveness as Kumar and Gulati (2010). The search space of the two-stage DEA model of efficiency and effectiveness is characterized by different combinations of input and output variables of efficiency and effectiveness. Using the GS algorithm, the variable selection framework traverses the search space of models and finds the best one according to my statistical method. I evaluate three search algorithms: GS, SSO, and the MABA. I find that the GS algorithm performs best.

The proposed variable selection framework in Chapter 4 of this dissertation can be used by managers when defining organizational goals and key performance indicators for the bank. The key performance indicators found from the variable selection framework will be aligned with how the market evaluates the bank. The actions that the bank then takes on the key performance indicators may result in happier shareholders. This is because the market will reward improvement in its efficiency and effectiveness with higher market evaluations. I use the variable selection framework to find the best universal two-stage DEA model of efficiency and effectiveness for banks in Brazil, Canada, China, India, Japan, Mexico, South Korea, and the USA and for 2000-2017. I also run the TYT test to check whether any cause and effect relationship exists between the efficiency and the effectiveness scores computed from the best universal two-stage DEA model and the Tobin's Q ratio. I find that efficiency and effectiveness do not Granger-cause the Tobin's Q ratio for all countries except for India. In an efficient market, efficiency and effectiveness and their lags cannot affect the Tobin's Q ratio. However, for banks in India, I find that (1) efficiency Granger-causes the Tobin's Q ratio when controlling for effectiveness and (2) effectiveness Granger-causes the Tobin's Q ratio when controlling for efficiency. I provide recommendations in Chapter 4, Section 4.4, on how banks can use this to their benefit. The intended audience for the work done in this chapter are banks planning to open in other locations. Suppose the exact definition of efficiency and effectiveness exists between two locations. In that case, banks can safely open new branches in the second location and use the definition of efficiency and effectiveness from the first location. In an inefficient market, traders can benefit the most by using the cause and effect relationship between efficiency and effectiveness and Tobin's Q ratio. The cause and effect relation can be used to predict the firm's performance in the financial market, and traders can profit from such information.

In Chapter 5, I provide an optimal path to transform Kumar and Gulati (2010)'s two-stage DEA model into the best two-stage DEA model found from my variable selection framework. I hypothesize that a set of incremental changes that transform Kumar and Gulati (2010)'s model into a model consistent with the semi-strong version of EMH will be quickly adopted by banks. Instead of telling a bank about the best two-stage DEA model of efficiency and effectiveness, it is more meaningful for the bank to know what path of incremental changes will transform its current model of measuring efficiency and effectiveness into the best model. Implementing changes incrementally allows the bank to optimize one dimension at a time, leading to little resistance to change, unlike a rapid, large-scale overhaul of its model which would lead to widespread disruption within the bank.

References

Abd Rahman, A., Ng, S. I., Sambasivan, M., and Wong, F. (2013). Training and organizational effectiveness: moderating role of knowledge management process. *European Journal of Training and Development*.

Abolghasemi Dehaghani, Y. (2020). Three essays on business analytics: time-series causality, panel data analysis, and design of experiments.

Achterberg, T. (2009). Scip: solving constraint integer programs. *Mathematical Programming Computation*, 1(1):1–41.

Adenso-Díaz, G. (1997). Linking and weighting efficiency estimates with stock performance in banking firms. http://fic.wharton.upenn.edu/fic/papers/97/adenso.pdf. (Accessed on 10/14/2016).

Adler, N. and Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to western europe. *European Journal of Operational Research*, 132(2):260–273.

Adler, N. and Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *Journal of the Operational Research Society*, 53(9):985–991.

Adler, N. and Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: Pca–dea or variable reduction. *European Journal of Operational Research*, 202(1):273 – 284.

Agarwal, N., Liu, H., Tang, L., and Yu, P. S. (2008). Identifying the influential bloggers in a community. In *Proceedings of the 2008 international conference on web search and data mining*, pages 207–218. ACM.

Alexakis, C., Patra, T., and Poshakwale, S. (2010). Predictability of stock returns using financial statement information: evidence on semi-strong efficiency of emerging greek stock market. *Applied Financial Economics*, 20(16):1321–1326.

Allen, J., Engert, W., et al. (2007). Efficiency and competition in canadian banking. *Bank of Canada Review*, 2007(Summer):33–45.

Álvarez, I. C., Barbero, J., Zofío, J. L., et al. (2017). A panel data toolbox for matlab. *Journal of Statistical Software*, 76(6):1–27.

Amah, E. and Ahiauzu, A. (2013). Employee involvement and organizational effectiveness. *Journal of Management Development*.

Amiri, A. and Ventelou, B. (2012). Granger causality between total expenditure on health and gdp in oecd: Evidence from the toda–yamamoto approach. *Economics Letters*, 116(3):541–544.

Amirteimoori, A., Despotis, D. K., and Kordrostami, S. (2014). Variables reduction in data envelopment analysis. *Optimization*, 63(5):735–745.

Anderson, H. M., Cain, J., and Bird, E. (2005). Online student course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education*, 69(1):5.

Andriansyah, A. and Messinis, G. (2019). Stock prices, exchange rates and portfolio equity flows: A toda-yamamoto panel causality test. *Journal of Economic Studies*.

Ansari, M. A., Haider, S., and Khan, N. (2020). Does trade openness affects global carbon dioxide emissions. *Management of Environmental Quality: An International Journal*.

Ansari, M. A., Khan, N. A., and Ganaie, A. A. (2019). Does foreign direct investment impede environmental quality in asian countries? a panel data analysis. *OPEC Energy Review*, 43(2):109–135.

Armstrong, R. (2020). Jpmorgan and citi upbeat on outlook for us economy.

Attarwala, A. (2016). Statistical method of goodness on quantitative models of efficiency and effectiveness. Master's thesis, University of Waterloo.

Awasthi, A. and Malafeyev, O. (2015). Is the indian stock market efficient-a comprehensive study of bombay stock exchange indices. *arXiv preprint arXiv:1510.03704*.

Aydın, U., Karadayi, M. A., and Ülengin, F. (2020). How efficient airways act as role models and in what dimensions? a superefficiency dea model enhanced by social network analysis. *Journal of Air Transport Management*, 82:101725.

Ayodele, A. J. and Maxwell, O. O. (2017). Test of the semi-strong efficiency theory in the nigerian stock market: An empirical analysis. *Journal of Finance and Accounting*, 5(4):139–146.

Aziz, M. A., Habibullah, M. S., Azman-Saini, W., and Azali, M. (2000). Testing for causality between taxation and government spending: An application of toda-yamamoto approach. *Pertanika Journal of Social Science and Humanities*, 8(1):45–50.

Ball, R. and Kothari, S. (1994). *Financial statement analysis*. McGraw-Hill.

Barbier, G., Zafarani, R., Gao, H., Fung, G., and Liu, H. (2012). Maximizing benefits from crowdsourced data. *Computational and Mathematical Organization Theory*, 18(3):257–279.

Beccalli, E., Casu, B., and Girardone, C. (2006). Efficiency and stock performance in European banking. *Journal of Business Finance and Accounting*, 33(1-2):245–262.

Beckett, D. (2009). The problem of measuring software productivity. http://www.qsm.com/blog/2014/problem-measuring-software-productivity. (Accessed on 10/31/2017).

Berger, J. O., Ghosh, J. K., and Mukhopadhyay, N. (2003). Approximations and consistency of bayes factors as model dimension grows. *Journal of Statistical Planning and Inference*, 112(1-2):241–258.

Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews*, 19(1):1–48.

Bharadwaj, A. S., Bharadwaj, S. G., and Konsynski, B. R. (1999). Information technology effects on firm performance as measured by tobin's q. *Management science*, 45(7):1008–1024.

Bliek1ú, C., Bonami, P., and Lodi, A. (2014). Solving mixed-integer quadratic programming problems with ibm-cplex: a progress report. In *Proceedings of the twenty-sixth RAMP symposium*, pages 16–17.

Blumenthal, B. and Haspeslagh, P. (1994). Toward a definition of corporate transformation. *MIT Sloan Management Review*, 35(3):101.

Bonami, P., Belotti, P., Forrest, J., Ladanyi, L., and Laird, C. (2019). Basic open-source nonlinear mixed integer programming. *accessed on Aug*, 1.

Boussemart, J.-P., Leleu, H., Shen, Z., Vardanyan, M., and Zhu, N. (2019). Decomposing banking performance into economic and credit risk efficiencies. *European Journal of Operational Research*, 277(2):719–726.

Bozdogan, H. (1987). Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370.

Budhiraja, S. and Malhotra, M. (2013). Leadership style & organizational effectiveness in indian it & banking industry. *The Indian Journal of Industrial Relations*, pages 270–285.

Cameron, K. (1978). Measuring organizational effectiveness in institutions of higher education. *Administrative Science Quarterly*, pages 604–632.

Carrig, D. (2018). Warren Buffett wins $1M bet against hedge funds and gives it to girls' charity. *USA Today*, pages 1–2.

Cetorelli, N. and Peretto, P. F. (2000). Oligopoly banking and capital accumulation.

Chappell Jr, H. W. and Cheng, D. C. (1984). Firms' acquisition decisions and tobin's q ratio. *Journal of Economics and Business*, 36(1):29–42.

Charfeddine, L. and Mrabet, Z. (2017). The impact of economic development and social-political factors on ecological footprint: A panel data analysis for 15 mena countries. *Renewable and Sustainable Energy Reviews*, 76:138–154.

Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.

Chen, Y., Liang, L., and Zhu, J. (2009). Equivalence in two-stage dea approaches. *European Journal of Operational Research*, 193(2):600–604.

Chen, Z., Wanke, P., Antunes, J. J. M., and Zhang, N. (2017). Chinese airline efficiency under co2 emissions and flight delays: A stochastic network dea model. *Energy Economics*, 68:89–108.

Chew, B. (1988). Harvard business review, no-nonsense guide to measuring productivity. https://hbr.org/1988/01/no-nonsense-guide-to-measuring-productivity. (Accessed on 10/26/2017).

Choi, I. (2006). Combination unit root tests for cross-sectionally correlated panels. *Econometric Theory and Practice: Frontiers of Analysis and Applied Research: Essays in Honor of Peter CB Phillips. Cambridge University Press, Chapt*, 11:311–333.

Chowdhury, A. and Mavrotas, G. (2006). Fdi and growth: What causes what? *The World Economy*, 29(1):9–19.

Chu, J. and Zhu, J. (2021). Production scale-based two-stage network data envelopment analysis. *European Journal of Operational Research*, 294(1):283–294.

Chu, S. F. and Lim, G. H. (1998). Share performance and profit efficiency of banks in an oligopolistic market: evidence from Singapore. *Journal of Multinational Financial Management*, 8(2-3):155–168.

Cinca, C. S. and Molinero, C. M. (2004). Selecting dea specifications and ranking units via pca. *Journal of the Operational Research Society*, 55(5):521–528.

Coles, J. L., Daniel, N. D., and Naveen, L. (2008). Boards: Does one size fit all? *Journal of financial economics*, 87(2):329–356.

Cook, W. D., Liang, L., and Zhu, J. (2010a). Measuring performance of two-stage network structures by dea: A review and future perspective. *Omega*, 38(6):423–430.

Cook, W. D., Zhu, J., Bi, G., and Yang, F. (2010b). Network dea: Additive efficiency decomposition. *European Journal of Operational Research*, 207(2):1122–1129.

Cooper, W. W., Seiford, L. M., and Zhu, J. (2004). Data envelopment analysis. In *Handbook on data envelopment analysis*, pages 1–39. Springer.

Craney, T. A. and Surles, J. G. (2002). Model-dependent variance inflation factor cutoff values. *Quality Engineering*, 14(3):391–403.

Daraio, C. and Simar, L. (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Springer Science & Business Media.

DeJong, K. (1975). Analysis of the behavior of a class of genetic adaptive systems. dept. *Computer and Communication Sciences, University of Michigan, Ann Arbor*.

Del Val, M. P. and Fuentes, C. M. (2003). Resistance to change: a literature review and empirical study. *Management decision*.

Deng, S. E. and Elyasiani, E. (2008). Geographic diversification, bank holding company value, and risk. *Journal of Money, Credit and Banking*, 40(6):1217–1238.

Derksen, S. and Keselman, H. (1992). Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *British Journal of Mathematical and Statistical Psychology*, 45(2):265–282.

Desjardins, J. (2019). The making of the big four banking oligopoly in one chart.

Dezsö, C. L. and Ross, D. G. (2012). Does female representation in top management improve firm performance? a panel data investigation. *Strategic management journal*, 33(9):1072–1089.

Di Iorio, F. and Triacca, U. (2011). Testing for non-causality by using the autoregressive metric.

Diasio, S. (2012). Open models of decision support towards a framework. http://www.tdx.cat/bitstream/handle/10803/82075/DIASIO_PhD%20Thesis_FV_Amb%20revisi?sequence=2. (Accessed on 04/07/2017).

Dong, H., Bowers, H. M., and Latham, W. R. (2013). Evidence on the efficient market hypothesis from 44 global financial market indexes. *Economics Research International*, 2013.

Draper, N. and Smith, H. (1981). *Applied Regression Analysis*. Wiley, New York, NY, second edition.

Dumitrescu, E.-I. and Hurlin, C. (2012). Testing for granger non-causality in heterogeneous panels. *Economic Modelling*, 29(4):1450–1460.

Economist (2003). Going with the flow — unknown — the economist. https://www.economist.com/unknown/2003/09/05/going-with-the-flow. (Accessed on 02/15/2021).

Economist (2012). Focus - executive pay and performance — graphic detail — the economist. https://www.economist.com/graphic-detail/2012/02/07/executive-pay-and-performance. (Accessed on 02/15/2021).

Economist (2020). Wall street says it is braced for losses. now what?

Economist, C. r. g. (2009). Economist, china's rapid growth is due not just to heavy investment, but also to the world's fastest productivity gains. http://www.economist.com/node/14844987. (Accessed on 10/31/2017).

Edgerton, D. and Shukur, G. (1999). Testing autocorrelation in a system perspective testing autocorrelation. *Econometric Reviews*, 18(4):343–386.

Emirmahmutoglu, F. and Kose, N. (2011). Testing for granger causality in heterogeneous mixed panels. *Economic Modelling*, 28(3):870–876.

Entorf, H. (1997). Random walks with drifts: Nonsense regression and spurious fixed-effect estimation. *Journal of Econometrics*, 80(2):287–296.

Fama, E. F. (1995). Random walks in stock market prices. *Financial analysts journal*, 51(1):75–80.

Fich, E. M. and Shivdasani, A. (2006). Are busy boards effective monitors? *The Journal of finance*, 61(2):689–724.

Fiordelisi, F. and Molyneux, P. (2010). The determinants of shareholder value in European banking. *Journal of Banking and Finance*, 34(6):1189–1200.

Fox, J. (1997). *Applied regression analysis, linear models, and related methods.* Sage Publications, Inc.

Friedman, L. and Sinuany-Stern, Z. (1998). Combining ranking scales and selecting variables in the dea context: The case of industrial branches. *Computers & Operations Research*, 25(9):781–791.

Gallivan, M. J., Hofman, J. D., and Orlikowski, W. J. (1994). Implementing radical change: Gradual versus rapid pace.

García-Sánchez, I. (2007). Efficiency and effectiveness of spanish football teams: a three-stage-dea approach. *Central European Journal of Operations Research*, 15(1):21–45.

Garg, A. and Tai, K. (2013). Comparison of statistical and machine learning methods in modelling of data with multicollinearity. *International Journal of Modelling, Identification and Control*, 18(4):295–312.

Ghoshal, S. and Bartlett, C. A. (1996). Rebuilding behavioral context: A blueprint for corporate renewal. *The Strategic Management of Intellectual Capital*, pages 141–161.

Golany, B. and Roll, Y. (1989). An application procedure for dea. *Omega*, 17(3):237–250.

Goodstein, L. D. and Burke, W. W. (1991). Creating successful organization change. *Organizational Dynamics*, 19(4):5–17.

Gorus, M. S. and Aydin, M. (2019). The relationship between energy consumption, economic growth, and co2 emission in mena countries: Causality analysis in the frequency domain. *Energy*, 168:815–822.

Granger, C. W. (1980). Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and control*, 2:329–352.

Groenewold, N. and Kang, K. C. (1993). The semi-strong efficiency of the australian share market. *Economic Record*, 69(4):405–410.

Grossmann, I. E., Viswanathan, J., Vecchietti, A., Raman, R., Kalvelagen, E., et al. (2002). Gams/dicopt: A discrete continuous optimization package. *GAMS Corporation Inc*, 37:55.

Guenster, N., Bauer, R., Derwall, J., and Koedijk, K. (2011). The economic value of corporate eco-efficiency. *European Financial Management*, 17(4):679–704.

Gujarati, D. (2006). Essential of econometrics 3rd edition mcgraw–hill. *New York*.

Hafsal, K., Suvvari, A., and Durai, S. R. S. (2020). Efficiency of indian banks with non-performing assets: evidence from two-stage network dea. *Future Business Journal*, 6(1):1–9.

Hage, J. and Aiken, M. (1970). *Social change in complex organizations (Studies in sociology)*. Random House, hardcover edition.

Hassanat, A., Almohammadi, K., Alkafaween, E., Abunawas, E., Hammouri, A., and Prasath, V. (2019). Choosing mutation and crossover ratios for genetic algorithms—a review with a new dynamic approach. *Information*, 10(12):390.

HBR (2018). Making process improvements stick: early excitement usually leads to backsliding. *Harvard business review*, 96(6):16.

HBS (2002). Toward continuous change: Staying competitive through change.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*, 49(5):1639–1664.

Ho, C.-T. and Zhu, D.-S. (2004). Performance measurement of taiwan's commercial banks. *International Journal of Productivity and Performance Management*, 53(5):425–434.

Holland, J. (1975). Adaptation in natural and artificial systems: an introductory analysis with application to biology. *Control and artificial intelligence*.

Hsiao, C. (1979). Causality tests in econometrics. *Journal of Economic Dynamics and Control*, 1(4):321–346.

Huangfu, S. X., Sun, H., Zhou, C., Dong, M., et al. (2017). A macroeconomic theory of banking oligopoly. In *2017 Meeting Papers*, number 191. Society for Economic Dynamics.

Hurlin, C. and Venet, B. (2008). Financial development and growth: a re-examination using a panel granger causality test.

Ioannidis, C., Molyneux, P., and Pasiouras, F. (2008). The relationship between bank efficiency and stock returns: evidence from asia and latin america.

Ittner, C. D. and Larcker, D. F. (2003). Coming up short on nonfinancial performance measurement. *Harvard business review*, 81(11):88–95.

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(1):51–61.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *The Journal of Finance*, 23(2):389–416.

Jose, M. L., Lancaster, C., Stevens, J. L., and Jennings, J. A. (1996). Stability of excellence: revealed patterns in tobin's q-ratios. *Journal of Applied Business Research*, 12(2):83.

Juliana, I., Hartini, A. G., Safrul Izani, M. S., and Yasmiza, L. (2011). Weak-form efficiency of foreign exchange market in the organisation for economic cooperation and development countries: Unit root test. *International Journal of Business and Management*, 6(6):1833–3850.

Jung, W. S. (1986). Financial development and economic growth: international evidence. *Economic Development and cultural change*, 34(2):333–346.

Kao, C. (1999). Spurious regression and residual-based tests for cointegration in panel data. *Journal of econometrics*, 90(1):1–44.

Kao, C. and Hwang, S.-N. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in taiwan. *European journal of operational research*, 185(1):418–429.

Kaplan, R. S., Norton, D. P., et al. (2005). The balanced scorecard: measures that drive performance. *Harvard business review*, 83(7):172.

Kar, M., Nazlıoğlu, Ş., and Ağır, H. (2011). Financial development and economic growth nexus in the mena countries: Bootstrap panel granger causality analysis. *Economic modelling*, 28(1-2):685–693.

Karlaftis, M. G. (2004). A dea approach for evaluating the efficiency and effectiveness of urban transit systems. *European Journal of Operational Research*, 152(2):354–364.

Khemraj, T. (2010). The simple analytics of oligopoly banking in developing economies.

Kim, D.-W., Yu, J.-S., and Hassan, M. K. (2018). Financial inclusion and economic growth in oic countries. *Research in International Business and Finance*, 43:1–14.

Kirkwood, J. and Nahm, D. (2006). Australian banking efficiency and its relation to stock returns. *Economic Record*, 82(258):253–267.

Konar, S. and Cohen, M. A. (2001). Does the market value environmental performance? *Review of economics and statistics*, 83(2):281–289.

Kronqvist, J., Bernal, D. E., Lundell, A., and Grossmann, I. E. (2019). A review and comparison of solvers for convex minlp. *Optimization and Engineering*, 20(2):397–455.

Kumar, S. and Gulati, R. (2010). Measuring efficiency, effectiveness and performance of indian public sector banks. *International Journal of Productivity and Performance Management*.

Lang, L. H. and Stulz, R. M. (1994). Tobin's q, corporate diversification, and firm performance. *Journal of political economy*, 102(6):1248–1280.

Lee, K. and Choi, K. (2010). Cross redundancy and sensitivity in dea models. *Journal of Productivity Analysis*, 34(2):151–165.

Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of econometrics*, 108(1):1–24.

Lewin, A. Y., Morey, R. C., and Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega*, 10(4):401–411.

Lexicon, F. T. (2009). Financial times lexicon. http://lexicon.ft.com/Term?term=Tobin's-q-ratio. (Accessed on 10/31/2017).

Li, Y., Chiu, Y.-h., Lin, T.-Y., and Huang, Y. Y. (2019). Market share and performance in taiwanese banks: min/max sbm dea. *Top*, 27(2):233–252.

Li, Y. and Cui, Q. (2017). Carbon neutral growth from 2020 strategy and airline environmental inefficiency: A network range adjusted environmental data envelopment analysis. *Applied Energy*, 199:13–24.

Liew, V. K.-S. (2004). Which lag length selection criteria should we employ? *Economics bulletin*, 3(33):1–9.

Lim, S. and Zhu, J. (2016). A note on two-stage network dea model: Frontier projection and duality. *European Journal of Operational Research*, 248(1):342–346.

Lopez, L., Weber, S., et al. (2017). Testing for granger causality in panel data. *Stata Journal*, 17(4):972–984.

López Iturriaga, F. J. and Morrós Rodríguez, I. (2014). Boards of directors and firm performance: the effect of multiple directorships. *Spanish Journal of Finance and Accounting/Revista Espanola de Financiacion y Contabilidad*, 43(2):177–192.

Luo, Y., Bi, G., and Liang, L. (2012). Input/output indicator selection for dea efficiency evaluation: An empirical study of chinese commercial banks. *Expert Systems with Applications*, 39(1):1118–1123.

Mabakeng, M. E. P. and Sheefeni, J. P. S. (2014). Investigating the semi-strong efficiency in namibia's foreign exchange market. *Global Journal of Contemporary Research in Accounting, Auditing and Business Ethics*, 1(3):168–181.

Maddala, G. S. and Lahiri, K. (1992). *Introduction to econometrics*, volume 2. Macmillan New York.

Maddala, G. S. and Wu, S. (1999). A comparative study of unit root tests with panel data and a new simple test. *Oxford Bulletin of Economics and statistics*, 61(S1):631–652.

Mahoney, T. A. and Weitzel, W. (1969). Managerial models of organizational effectiveness. *Administrative Science Quarterly*, pages 357–365.

Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *The Journal of Economic Perspectives*, 17(1):59–82.

Malkiel, B. G. (2005). Reflections on the efficient market hypothesis: 30 years later. *Financial Review*, 40(1):1–9.

Margari, B. B., Erbetta, F., Petraglia, C., and Piacenza, M. (2007). Regulatory and environmental effects on public transit efficiency: a mixed dea-sfa approach. *Journal of Regulatory Economics*, 32(2):131–151.

Martin, R. L. (2019). The high price of efficiency. *Harvard business review*, 97(1):42–55.

Martin, R. L. (2020). *When More Is Not Better: Overcoming America's Obsession with Economic Efficiency*. Harvard Business Review Press, kindle edition edition.

MATLAB (2020). Surrogate optimization algorithm - matlab & simulink.

Mavrotas, G. and Kelly, R. (2001). Old wine in new bottles: Testing causality between savings and growth. *The Manchester School*, 69:97–105.

McCallum, L. (1984). A meta-analysis of course evaluation data and its use in the tenure decision. *Research in Higher Education*, 21(2):150–158.

McFarland, K. (2014). pwc-canadian-banks-2015-en.pdf. https://www.pwc.com/ca/en/banking-capital-markets/publications/pwc-canadian-banks-2015-en.pdf. (Accessed on 10/14/2016).

Metzger, L. M. (1992). Measuring production department efficiency using data envelopment analysis. *Journal of Managerial Issues*, 4(4):494–509.

Mezias, S. J. and Glynn, M. A. (1993). The three faces of corporate renewal: Institution, revolution, and evolution. *Strategic management journal*, 14(2):77–101.

Mirzaei, A., Saad, M., and Emrouznejad, A. (2020). Bank stock performance during the covid-19 crisis: Does efficiency explain why islamic banks fared relatively better? *Available at SSRN 3702116*.

Mishra, A. and Shukla, A. (2017). Analysis of the effect of elite count on the behavior of genetic algorithms: a perspective. In *2017 IEEE 7th International Advance Computing Conference (IACC)*, pages 835–840. IEEE.

Mishra, P. and Misra, R. K. (2017). Entrepreneurial leadership and organizational effectiveness: A comparative study of executives and non-executives. *Procedia computer science*, 122:71–78.

Mohan, T. and Ray, S. C. (2004). Productivity growth and efficiency in indian banking: a comparison of public, private, and foreign banks. *Economics Working Papers*, page 200427.

Moore, A. B. (1962). *A statistical analysis of common stock prices*. PhD thesis, University of Chicago.

Morita, H. and Avkiran, N. K. (2009). Selecting inputs and outputs in data envelopment analysis by designing statistical experiments (special issue, operations research for performance evaluation). *Journal of the Operations Research society of Japan*, 52(2):163–173.

Muller, J. Z. (2018). *The Tyranny of Metrics*. Princeton University Press.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society*, pages 69–85.

Nan, Z. and Kaizoji, T. (2019). Market efficiency of the bitcoin exchange rate: Weak and semi-strong form tests with the spot, futures and forward foreign exchange rates. *International Review of Financial Analysis*, 64:273–281.

Nataraja, N. R. and Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, 215(3):662–669.

Niederhoffer, V. (1971). The analysis of world events and stock prices. *The Journal of Business*, 44(2):193–219.

Nunamaker, T. R. (1985). Using data envelopment analysis to measure the efficiency of non-profit organizations: A critical evaluation. *Managerial and decision Economics*, 6(1):50–58.

Orlikowski, W. J. (1993). Case tools as organizational change: Investigating incremental and radical changes in systems development. *MIS quarterly*, pages 309–340.

Paleckova, I. (2019). Cost efficiency measurement using two-stage data envelopment analysis in the czech and slovak banking sectors. *Acta Oeconomica*, 69(3):445–466.

Paradi, J. C., Rouatt, S., and Zhu, H. (2011). Two-stage evaluation of bank branch efficiency using data envelopment analysis. *Omega*, 39(1):99–109.

Paradi, J. C. and Schaffnit, C. (2004). Commercial branch performance evaluation and results communication in a canadian bank—-a dea application. *European Journal of Operational Research*, 156(3):719–735.

Paramati, S. R., Ummalla, M., and Apergis, N. (2016). The effect of foreign direct investment and stock market growth on clean energy use across a panel of emerging market economies. *Energy Economics*, 56:29–41.

Pastor, J. T., Ruiz, J. L., and Sirvent, I. (2002). A statistical test for nested radial dea models. *Operations Research*, 50(4):728–735.

Peirce, H. and Miller, S. M. (2015). Small banks by the numbers, 2000–2014. *Mercatus Center, March*, 17.

Pesaran, M. H. (2004). General diagnostic tests for cross-sectional dependence in panels. *Empirical Economics*, pages 1–38.

Popper, K. (1957). Philosophy of science. *British Philosophy in the Mid-Century (ed. CA Mace). London: George Allen and Unwin.*

Raffaelli, R. (2015). Leading and managing change.

Raheli, H., Rezaei, R. M., Jadidi, M. R., and Mobtaker, H. G. (2017). A two-stage dea model to evaluate sustainability and energy efficiency of tomato production. *Information Processing in Agriculture*, 4(4):342–350.

Ray, S. C. (2008). Comparing input-and output-oriented measures of technical efficiency to determine local returns to scale in dea models.

Reguera-Alvarado, N., de Fuentes, P., and Laffarga, J. (2017). Does board gender diversity influence financial performance? evidence from spain. *Journal of Business Ethics*, 141(2):337–350.

Reuters, T. (2017). Thomson reuters eikon. *Thomson Reuters*.

Roberts, J. (2000). Spurious regression problems in the determinants of health care expenditure: a comment on hitiris (1997). *Applied Economics Letters*, 7(5):279–283.

Rogers, E. M. (2003). *Diffusion of Innovations, 5th Edition*. Free Press, paperback edition.

Rose, A. and Lawton, A. (1999). *Public services management*. Financial Times/Prentice Hall.

Rousseau, D. M. (1989). Psychological and implied contracts in organizations. *Employee responsibilities and rights journal*, 2(2):121–139.

Ruggiero, J. (2005). Impact assessment of input omission on dea. *International Journal of Information Technology & Decision Making*, 4(03):359–368.

Seiford, L. M. and Zhu, J. (1999). Profitability and marketability of the top 55 us commercial banks. *Management science*, 45(9):1270–1288.

Sengupta, J. K. (1990). Tests of efficiency in data envelopment analysis. *Computers & Operations Research*, 17(2):123–132.

Serrano-Cinca, C., Fuertes-Callén, Y., and Mar-Molinero, C. (2005). Measuring dea efficiency in internet companies. *Decision Support Systems*, 38(4):557–573.

Shan, J. Z., Morris, A. G., and Sun, F. (2001). Financial development and economic growth: An egg-and-chicken problem? *Review of international Economics*, 9(3):443–454.

Sharma, M. J. and Yu, S. J. (2015). Stepwise regression data envelopment analysis for variable reduction. *Applied Mathematics and Computation*, 253:126–134.

Sharma, N. and Singh, R. K. (2019). A unified model of organizational effectiveness. *Journal of Organizational Effectiveness: People and Performance*.

Sigala, M., Airey, D., Jones, P., and Lockwood, A. (2004). Ict paradox lost? a stepwise dea methodology to evaluate technology investments in tourism settings. *Journal of Travel Research*, 43(2):180–192.

Silvestro, R. (2016). Do you know what really drives your business's performance? *MIT Sloan Management Review*, 57(4):28.

Simar, L. and Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models. *Communications in Statistics-Simulation and Computation*, 30(1):159–184.

Simar, L. and Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: recent developments and perspectives. *The measurement of productive efficiency and productivity growth*, pages 421–521.

Simar, L. and Wilson, P. W. (2011). Two-stage dea: caveat emptor. *Journal of Productivity Analysis*, 36(2):205.

Sirvent, I., Ruiz, J. L., Borrás, F., and Pastor, J. T. (2005). A monte carlo evaluation of several tests for the selection of variables in dea models. *International Journal of Information Technology & Decision Making*, 4(03):325–343.

Snipes, M. and Taylor, D. C. (2014). Model selection and akaike information criteria: An example from wine ratings and prices. *Wine Economics and Policy*, 3(1):3–9.

Stanley, M. (2015). Banks need to optimize their resources. http://www.morganstanley.com/ideas/banks-need-to-optimize-their-resources/.

Steers, R. M. (1975). Problems in the measurement of organizational effectiveness. *Administrative Science Quarterly*, pages 546–558.

stepwisefit in Matlab (2014). Stepwise regression matlab stepwisefit. http://www.mathworks.com/help/stats/stepwisefit.html?s_tid=gn_loc_drop. (Accessed on 08/02/2016).

Stine, R. A. (1987). Estimating properties of autoregressive forecasts. *Journal of the American statistical association*, 82(400):1072–1078.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician*, 49(1):53–56.

Stock, J. H. and Watson, M. W. (2020). Introduction to econometrics 4th ed.

Stone, M. (1979). Comments on model selection criteria of akaike and schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 276–278.

Sull, D. N. and Escobari, M. (2004). Creating value in an unpredictable world. *Business strategy review*, 15(3):14–20.

Surowiecki, J. (2005). *The wisdom of crowds*. Anchor Books, reprint edition.

Tamatam, R., Dutta, P., Dutta, G., and Lessmann, S. (2019). Efficiency analysis of indian banking industry over the period 2008–2017 using data envelopment analysis. *Benchmarking: An International Journal*.

Titko, J., Stankevičienė, J., and Lāce, N. (2014). Measuring bank efficiency: Dea application. *Technological and Economic Development of Economy*, 20(4):739–757.

Toda, H. Y. and Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of econometrics*, 66(1-2):225–250.

Torun, M. and Serdar, K. (2008). Testing weak and semi-strong form efficiency of stock exchanges in european monetary union countries: Panel data causality and co-integration analysis. *Uluslararası İktisadi ve İdari İncelemeler Dergisi*, (1).

Tushman, M. L., Newman, W. H., and Romanelli, E. (1986). Convergence and upheaval: Managing the unsteady pace of organizational evolution. *California management review*, 29(1):29–44.

Ueda, T. and Hoshiai, Y. (1997). Application of principal component analysis for parsimonious summarization of dea inputs and/or outputs. *Journal of the Operations Research Society of Japan*, 40(4):466–478.

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228.

Wagner, J. M. and Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European journal of operational research*, 180(1):57–67.

Waltz, R. A. and Nocedal, J. (2004). Knitro 2.0 user's manual. *Ziena Optimization, Inc.[en ligne] disponible sur http://www. ziena. com (September, 2010)*, 7:33–34.

Wang, K., Huang, W., Wu, J., and Liu, Y.-N. (2014). Efficiency measures of the chinese commercial banking system using an additive two-stage dea. *Omega*, 44:5–20.

Wang, R.-T., Ho, C.-T. B., and Oh, K. (2010). Measuring production and marketing efficiency using grey relation analysis and data envelopment analysis. *International Journal of Production Research*, 48(1):183–199.

Wanke, P. F. (2012). Determinants of scale efficiency in the brazilian 3pl industry: A 10-year analysis. *International Journal of Production Research*, 50(9):2423–2438.

Wernerfelt, B. and Montgomery, C. A. (1988). Tobin's q and the importance of focus in firm performance. *The American Economic Review*, pages 246–250.

Westerlund, T. and Pettersson, F. (1995). An extended cutting plane method for solving convex minlp problems. *Computers & Chemical Engineering*, 19:131–136.

Wickremasinghe, G. (2004). Efficiency of foreign exchange markets: A developing country perspective. *Available at SSRN 609285*.

Wijesiri, M., Martínez-Campillo, A., and Wanke, P. (2019). Is there a trade-off between social and financial performance of public commercial banks in india? a multi-activity dea model with shared inputs and undesirable outputs. *Review of Managerial Science*, 13(2):417–442.

Wolfe, J. (2003). The tobinq as a company performance indicator.

Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

Xie, Z. and Chen, S.-W. (2014). Untangling the causal relationship between government budget and current account deficits in oecd countries: Evidence from bootstrap panel granger causality. *International Review of Economics & Finance*, 31:95–104.

Yang, G.-L., Song, Y.-Y., Xu, D.-L., and Yang, J.-B. (2017). Overall efficiency and its decomposition in a two-stage network dea model. *Journal of Management Science and Engineering*, 2(3):161–192.

Yang, H. and Pollitt, M. (2009). Incorporating both undesirable outputs and uncontrollable variables into dea: The performance of chinese coal-fired power plants. *European journal of operational research*, 197(3):1095–1105.

Yang, Z. (2006). A two-stage dea model to evaluate the overall performance of canadian life and health insurance companies. *Mathematical and computer modelling*, 43(7-8):910–919.

Yermack, D. (1997). Good timing: Ceo stock option awards and company news announcements. *The journal of Finance*, 52(2):449–476.

Yu, M.-M., Hsiung, N.-H., and Chen, L.-H. (2020). Determinants of banks' nerlovian economic efficiency: a dea-bootstrap approach. *Applied Economics*, pages 1–19.

Zhu, J. (2000). Multi-factor performance measure model with an application to fortune 500 companies. *European journal of operational research*, 123(1):105–124.

APPENDICES

Appendix A

Glossary of terms and definitions

1. **Hausman test:**

   Is used in panel data to decide between fixed effects estimators and random effects estimator. Assume the following panel data regression model, $y_{i,t} = \alpha + \beta X_{i,t} + \mu_i + \eta_{i,t}$ where $X_{i,t}$ are the independent variables, μ_i is the unobserved heterogeneity. The error terms $\eta_{i,t}$ are assumed to be homoskedastic.

   In the null hypothesis, the covariance between the unobserved heterogeneity and the independent variables is zero. Under the null hypothesis, the Hausman test statistic is calculated as $W = \frac{(\beta_{FE} - \beta_{RE})^2}{Var(\beta_{FE}) - Var(\beta_{RE})}$ where W follows a chi-squared distribution with 1 degree of freedom. β_{FE} is the fixed effects estimator of the panel regression model and β_{RE} is the random effects estimator of the panel regression model (Wooldridge, 2015, Chapter 14).

2. **Heteroskedastic:**

   In time series, heteroskedastic refers to a condition in which the variance of the error term, in a regression model varies with time (Stock and Watson, 2020, Chapter 5).

3. **Heteroskedastic with auto correlation:**

   Heteroskedasticity with autocorrelation consistent (HAC) refers to the condition where the error terms are assumed to be heteroskedastic. There is also a correlation among the error terms of the cross-sectional units in the panel data (Stock and Watson, 2020, Chapter 5).

4. **Homoskedastic:**

   In time series, homoskedastic refers to a condition in which the variance of the error term in a regression model is constant (i.e., not a function of time) (Stock and Watson, 2020, Chapter 5).

5. **Multicollinearity:**

   Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a regression model (Stock and Watson, 2020, Chapter 6).

6. **Mundlak test:**

   Is used in panel data to decide between fixed effects estimators and random effects estimator. The error terms in the panel data are assumed to be heteroskedastic.

   The Mundlak approach suggests estimating the following regression: $y_{it} = \alpha + \beta X_{it} + \gamma \bar{X}_i + \mu_i + \eta_{it}$ where $\bar{X}_i$ are country specific means. A Wald joint significance test on γ is performed where the null hypothesis is set to $H_0 : \gamma = 0$ (i.e., the random-effects model holds under the null hypothesis) (Mundlak, 1978).

7. **Spurious regression:**

   When two non-stationary time series data are regressed against each other, spurious regression may occur (Kao, 1999). For example, consider a village's ice cream sales. The sales might be very high when the rate of buying air conditioners in the city is highest. To claim that ice cream sales cause purchasing of air conditioners, or vice versa would imply a spurious relationship between the two.

8. **Stationary series:**.

   A stationary time series is one whose statistical properties such as mean and variance are constant over time.

   A non-stationary time series, on the other hand, may have its mean or the variance as a function of time (Wooldridge, 2015, Chapter 11).

9. **Wald test:**

   The Wald test works by testing the null hypothesis that a set of parameters is equal to some value. The null hypothesis is that the two coefficients of interest are simultaneously equal to zero. If the test fails to reject the null hypothesis, this suggests that removing the variables from the model will not substantially harm the fit of that model. More details of the Wald test, including its formula and mathematical derivation, are presented in (Fox, 1997, Chapter 9). The Wald test essentially tests how far the estimated parameters are from zero. Wald test can be used to test multiple parameters simultaneously.

Appendix B

Summary of Eikon data

In this chapter we present a summary of the data used in the case study in Section 2.4. Due to size, tables are presented on individual pages, starting with the next page.

Table B.1 Mean of the 55 dimensions from Eikon across Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017. All prices are in US dollars

	Brazil	India	China	USA	Canada	Mexico	SouthKorea	Japan
Total Assets	15505474539.06	62012169859.04	1214148198097.82	1146487076763.28	427797504191.02	38177690344.25	195205366056.09	722660173677.39
Full-Time Employees	12497.31	39569.07	230843.83	163295.07	49269.37	20052.42	10610.93	27949.66
Net Loans	6093079849.32	39092061228.24	618984979926.54	468402075538.06	206119599977.13	17594101744.68	134042911939.01	310033510429.56
Other Earning Assets	9600252776.41	19903479396.11	326839162227.61	512163524583.44	191991912819.32	14886641786.97	39798792447.62	309393708110.21
Long Term Investments	1023912732.04	545006250.21	1084638186.01	6103757354.64	1607687247.50	255840402.07	406014828.14	3343965999.38
Non-Interest Income, Bank	986392973.70	1274569466.89	9125309026.85	25204158687.13	7154864912.90	1013800753.59	8150989504.85	8349764721.25
Interest Income, Bank	2006406781.49	4721677224.51	46344214146.62	40338242283.90	12535104947.25	3237312935.84	8431821025.29	9228791349.14
Net Income Incl Extra Before Distributions	258914647.57	452436652.33	13238539974.23	9663672665.93	3375374043.51	677279522.40	1234279582.43	1637260822.88
Income Avail to Cmn Shareholders Excl Extra	257376358.03	451696954.49	13201118376.81	8924673478.24	3316723344.28	650929283.36	1175324536.94	1575279495.43
Income Avail to Cmn Shareholders Incl Extra	257414058.59	451590457.03	13201118376.81	8992063832.15	3246958490.89	677279522.40	1142224692.70	1575279495.43
Basic Weighted Average Shares	526973128.85	1688820496.65	192773471960.18	2857931732.34	888124169.48	4667668882.94	341125224.48	4116131537.14
Basic EPS Excluding Extraordinary Items	12.87	0.29	0.08	4.97	3.69	0.22	3.14	4.20
Basic EPS Including Extraordinary Items	12.87	0.29	0.08	5.28	3.67	0.22	3.14	4.20
Diluted Net Income	257414058.59	451554529.11	13164688809.09	9000441521.70	3282892458.78	677292403.82	1183965063.78	1595620692.50
Diluted Weighted Average Shares	527905780.75	1693123504.53	193644891168.76	2906466878.20	896872225.14	4669267261.91	353059813.78	4425932596.85
Diluted EPS Excluding Extraordinary Items	12.87	0.29	0.07	4.89	3.67	0.22	3.09	4.10
Diluted EPS Including Extraordinary Items	12.87	0.29	0.07	5.19	3.65	0.22	3.09	4.10
Net Income Before Extraordinary Items	258876947.02	452543149.79	13238539974.23	9596282312.02	3415890356.82	650929283.36	1200415197.76	1637260822.88
Other Expense	-754499270.71	-1005864279.35	-10419478026.69	-14837903714.50	-4016549020.96	-1079810264.07	-3281161805.14	-9890575079.87
Non-Interest Expense, Bank	-1177726410.75	-1732614147.60	-14590406123.98	-32609971306.92	-8794683067.25	-1416113123.09	-9408405938.29	-10643663737.59
Net Income Before Taxes	343963062.88	613367380.10	17417312615.30	13445810164.42	4306143801.21	815669043.16	1623424386.80	2849655671.86
Provision for Income Taxes	76572113.29	159476850.01	4001907313.93	3772928364.05	900790005.04	171821700.89	371067325.56	953471197.84
Net Income After Taxes	266664978.32	455266876.39	13415405301.37	9672881800.37	3434957723.38	643847342.27	1297025086.58	1896184474.02
Other Revenue	116320949.93	189684991.31	806219183.38	2543168682.44	864165801.74	94558429.24	1241479731.06	3042995239.78
Non-Interest Income, Bank	986392973.70	1274569466.89	9125309026.85	25204158687.13	7154864912.90	1013800753.59	8150989504.85	8349764721.25
Net Interest Income	876204683.91	1666854016.42	27452819110.90	26797938920.33	6776532108.54	1645123574.46	4021060153.92	6127576012.33
Loan Loss Provision	351311879.08	625950724.85	4570610115.78	5946316136.12	783661363.98	429900117.85	760351496.29	1008121277.21
Net Interest Income After Loan Loss Provision	537359561.93	1042107630.81	22882409712.43	20851622784.20	5992870744.56	1215223456.61	3311594267.40	5143554688.20
Interest Income, Bank	2006406781.49	4721677224.51	46344214146.62	40338242283.90	12535104947.25	3237312935.84	8431821025.29	9228791349.14
Interest on Deposit	632374833.55	2729553759.77	16345863545.55	5655964104.56	4265413473.03	1417306981.20	2459211339.01	1461460973.13
Interest on Deposits	1328000745.27	161440993.76	4184789795.02	974442271.90	238104923.83	25044861.06	98623693.37	297996546.28
Other Interest Income	473086126.59	106758566.88	17748969805.72	2332465279.22	23948086.86	202846805.09	320450176.01	613391614.95
Interest & Fees on Loans	1095381958.47	3325526395.48	29464752079.62	26469810271.01	9349477712.01	3177077401.18	7038885879.96	6010917551.60
Interest And Dividends on Investment Secs	741878515.52	1275980037.47	9974439156.70	5575144946.34	2410031861.08	351307820.72	901142622.87	2014917094.29
Fed Funds Sold/Scty Prchd Under Resale Agrmnt	5995309145.10	0.00	807488479.38	2827140580.76	918542590.33	192393156.26	0.00	343604194.14
Total Equity	1705358805.56	4081674047.28	75854929285.59	107074903187.69	23093862063.67	4298601062.21	14408430044.81	29295913910.09
Total Liabilities And Shareholders Equity	15505474534.28	62012169859.32	1214148198099.03	1146487076762.43	427797505488.16	38177690342.24	196323125313.87	722660168320.66
Total Common Shares Outstanding	508496364.74	1729550732.85	195029329472.40	2900470880.90	892487955.05	4672767259.40	349705100.75	4207338389.92
Retained Earnings (Accumulated Deficit)	639371546.90	2759024989.07	40705074466.78	61275236644.86	13301470776.97	2102037019.02	6051004119.99	13243090178.13
Common Stock, Total	1079119209.28	99201227.96	25962868092.24	9389660414.18	7113148446.81	1603369349.73	1596573434.61	6622440509.10
Common Stock Other	1079119209.28	99201227.96	25962868092.24	9389660414.18	7113148446.81	1603369349.73	1596573434.61	6622440509.10
Other Liabilities, Total	3146619445.58	3896130240.64	26629624333.56	104665786060.00	52347429202.83	2718941705.44	15133802144.90	52615044871.10
Total Liabilities	13800115728.71	57930495812.03	1138293268813.44	1039411418364.37	404703643424.49	33879089280.04	180797074264.88	693364254410.57
Total Long Term Debt	1299648794.16	5527591369.01	19788693575.82	139159405956.99	9759217476.42	1677249680.41	22811954778.33	42487912219.28
Total Debt	4853147813.23	6497131841.28	33919976701.61	281666187684.73	48847859653.75	8391356216.88	44789126500.97	146446878798.06
Long Term Debt	1628706774.10	5645044482.08	21076413547.99	139279841826.61	9736040552.32	1928173409.49	30623978197.85	42721623006.38
Total Deposits	6803064593.29	52865418852.33	1055689319265.49	607573972952.36	285915957443.91	19328408591.49	114031076311.23	462514260191.51
Other Assets	1170506376.33	2223682786.47	31129870850.79	83552106140.02	13876816288.51	733254961.37	7672326187.33	23255791140.14
Other Assets, Total	1274625934.45	2647515779.07	37458248730.71	93130247678.94	14499955385.96	792168523.10	7890152152.37	23288977827.12
Total Assets, Reported	15505474539.06	62012169859.04	1214148198097.82	1146487076763.28	427797504191.02	38177690344.25	195205366056.09	722660173677.39
Property/Plant/Equipment Total - Net	117219061.35	450389756.45	11872538162.64	6298705758.94	1697676505.55	433157160.24	1954390573.23	5844408484.30
Net Loans	6093079849.32	39092061228.24	618984979926.54	468402075538.06	206119599977.13	17594101744.68	134042911939.01	310033510429.56
Other Earning Assets, Total	9600252776.41	19903479396.11	326839162227.61	512163524583.44	191991912819.32	14886641786.97	39798792447.62	309393708110.21
Total Investment Securities	5457409107.30	15145750760.21	236986822993.20	170572301601.99	61425409992.46	4945632230.23	27462403228.02	181552026316.12
Cash & Due From Banks	1550451856.84	6154019124.86	213446220473.92	27884796481.67	5635201318.31	4360737455.94	9654996766.43	67796267209.04

Table B.2 Standard Deviation of the 55 dimensions from Eikon across Brazil, Canada, China, India, Japan, Mexico, South Korea and the USA and for time period 2000-2017. All prices are in US dollars

	Brazil	India	China	USA	Canada	Mexico	SouthKorea	Japan
Total Assets	39402282173.75	45697940627.89	897955856811.74	309184644754.08	195826731259.31	13976732058.00	101648697034.17	355656579447.02
Full-Time Employees	16608.82	15728.59	107803.02	25715.49	11923.02	7372.15	8122.11	13005.50
Net Loans	9710112037.08	33092344482.36	468037838296.31	128771791565.37	101336773177.33	5869632706.74	68786919973.49	129724106093.70
Other Earning Assets, Total	26343958189.11	18043786142.34	230638567340.68	144707854265.46	84268313435.42	7991447261.65	23425654837.08	151172893978.31
Long Term Investments	561328155.28	860891368.27	711387543.82	4202849351.19	964234526.92	223532462.17	229434423.04	1979301047.10
Non-Interest Income, Bank	2342834193.42	1490084247.83	8297582284.10	6398710242.17	2661315841.54	435705165.25	5673852135.58	3947149427.03
Interest Income, Bank	4269107681.16	3865316154.22	34461120688.87	10702768879.96	4044809615.63	745751893.88	4580522308.79	3786322112.71
Net Income Incl Extra Before Distributions	596156773.39	344434059.95	10933792627.91	4220129460.40	1837161763.95	242330108.33	719028271.88	2564012078.17
Income Avail to Cmn Shareholders Excl Extra	596152728.63	343519756.34	10888330139.02	4425288194.13	1787183117.00	243146884.36	707016659.26	2559303439.34
Income Avail to Cmn Shareholders Incl Extra	596161848.23	343595855.79	10888330139.02	4254477496.21	1755997155.59	242330108.33	647313108.28	2559303439.34
Basic Weighted Average Shares	1051574544.77	864220789.56	81481011471.33	991214853.70	168446351.04	745903000.42	77862689.76	1927683649.14
Basic EPS Excluding Extraordinary Items	77.71	0.18	0.06	4.64	1.57	0.21	1.65	11.75
Basic EPS Including Extraordinary Items	77.71	0.18	0.06	4.91	1.57	0.21	1.65	11.75
Diluted Net Income	596161848.23	343559649.80	10847422892.07	4260363525.90	1807184292.43	242321230.03	709081088.26	2565517755.60
Diluted Weighted Average Shares	1054646724.34	869173443.54	81879501864.76	1010058701.45	168607454.29	745088777.83	70600420.01	1943401581.12
Diluted EPS Excluding Extraordinary Items	77.71	0.18	0.06	4.57	1.56	0.21	1.68	11.72
Diluted EPS Including Extraordinary Items	77.71	0.18	0.06	4.82	1.56	0.21	1.68	11.72
Net Income Before Extraordinary Items	596148072.25	344357871.13	10933792627.91	4428737643.27	1823828904.73	243146884.36	661550298.14	2564012078.17
Other Expense	1847879288.28	1111305456.48	6531731200.82	3755804437.83	1478188829.69	366023238.58	2411096620.99	4164265954.63
Non-Interest Expense, Bank	2766042243.67	1767209933.23	10534778433.68	8133648271.80	3109036026.67	447681979.98	6087387229.58	4425054160.69
Net Income Before Taxes	815084647.95	472216060.56	13987221822.32	6275480474.98	2186502732.90	327108931.90	899469403.14	3145175575.15
Provision for Income Taxes	216978996.77	135426575.02	3002872525.97	1983169883.46	440889172.52	95480890.69	234891154.61	601225870.73
Net Income After Taxes	601202744.63	347756658.50	11021162924.70	4451161388.26	1809099400.01	250037517.18	763584884.93	2664987998.02
Other Revenue	186339664.24	163780406.29	675736139.63	1652815264.29	357482114.38	77420287.46	1794289874.69	1436851468.82
Non-Interest Income, Bank	2342834193.42	1490084247.83	8297582284.10	6398710242.17	2661315841.54	435705165.25	5673852135.58	3947149427.03
Net Interest Income	1978926683.17	1411197236.49	19679680462.08	6533620528.58	2974050467.31	530084287.17	1894994897.37	2314701609.35
Loan Loss Provision	731219412.94	945612805.01	3801557952.06	5640442545.69	361089180.30	290232263.56	417257823.33	1170626638.26
Net Interest Income After Loan Loss Provision	1254191950.22	678979334.17	16551723932.52	6816572220.39	2711749926.52	306629063.45	1830857597.38	2432128430.71
Interest Income, Bank	4269107681.16	3865316154.22	34461120688.87	10702768879.96	4044809615.63	745751893.88	4580522308.79	3786322112.71
Interest on Deposit	892656370.32	2228057081.43	12097248269.35	4527703499.88	1749534056.43	688494324.34	1227122826.25	1021376628.87
Interest on Deposits	1763373717.01	68872955.66	3668056265.87	774740378.06	161746800.77	3132580.35	54110915.71	210006528.41
Other Interest Income	993914123.05	106512578.01	22753049116.75	2017730163.10	22251237.19	127330610.18	963010697.87	338267475.82
Interest & Fees on Loans	1801392992.95	2692202815.88	20330384728.43	5749191941.54	3296758075.67	787430851.00	3741262823.81	2359759057.87
Interest And Dividends on Investment Secs	1172334425.63	1005129995.10	8390279573.64	1837431005.11	815664836.56	222444888.60	348298581.62	815191673.74
Fed Funds Sold/Scty Prchd Under Resale Agrmnt	2121186715.32	0.00	918266178.12	2405071104.59	657936063.59	38860984.24	0.00	328828349.17
Total Equity	4137921676.37	3576512675.51	73034585529.21	40294852131.77	12562108861.18	1617899367.25	8521058299.88	17796661162.19
Total Liabilities And Shareholders Equity	39402282175.55	45697940627.61	897955856811.88	309184644756.28	195826731543.44	13976732055.40	103862456170.20	355656580111.54
Total Common Shares Outstanding	1033787036.86	884827005.39	83456510976.34	960350990.25	168792756.51	749435016.19	75327169.53	1969449958.90
Retained Earnings (Accumulated Deficit)	1478737085.01	2228335795.72	56162312568.32	22794954936.28	6598942779.89	1544889514.35	5082148393.07	10958755687.76
Common Stock, Total	2753268914.84	18499936.01	13390726632.67	9044491318.47	3758651272.66	521782890.28	361149214.32	3357501561.72
Common Stock Other	2753268914.84	18499936.01	13390726632.67	9044491318.47	3758651272.66	521782890.28	361149214.32	3357501561.72
Other Liabilities, Total	9378271079.62	3424361761.92	22389866759.70	27778863162.95	23296156268.40	3116780693.41	11701387259.86	27620914108.64
Total Liabilities	35270937675.27	42264269743.19	826650608269.17	271686846356.56	183417757306.38	12495710414.63	93206087416.00	338780524984.37
Total Long Term Debt	3231727226.13	5981765941.48	24252043031.59	46123300863.05	3983507435.08	1874759690.95	24883476056.47	28979776829.32
Total Debt	12392575959.84	7134432928.69	39437849573.75	80769077950.27	23734764758.25	5372912450.86	18834128039.74	75284051515.39
Long Term Debt	3269828431.82	5978604670.24	23980912893.77	45933037233.13	3992518936.93	1681441375.45	18557660783.79	28877674590.39
Total Deposits	12739084962.17	42787899449.25	752861470508.05	212419544524.53	134456496856.40	4492939300.38	62390328403.24	223478501598.99
Other Assets	2171878594.39	2777642972.92	24793910350.33	20582889920.37	7554086462.53	591360287.59	3788665831.02	13424043287.48
Other Assets, Total	2180729694.34	3234440150.77	27545741188.99	24345761304.30	7733945641.57	638237067.20	3873039989.75	13377369009.84
Total Assets, Reported	39402282173.75	45697940627.89	897955856811.74	309184644754.08	195826731259.31	13976732058.00	101648697034.17	355656579447.02
Property/Plant/Equipment Total - Net	267587828.56	347854041.06	7589573008.50	1613916392.72	493055233.10	97578975.36	750542688.19	2343509502.63
Net Loans	9710112037.08	33092344482.36	468037838296.31	128771791565.37	101336773177.33	5869632706.74	68786919973.49	129724106093.70
Other Earning Assets	26343958189.11	18043786142.34	230638567340.68	144707854265.46	84268313435.42	7991447261.65	23425654837.08	151172893978.31
Total Investment Securities	19602137939.07	12313161085.74	166193239669.33	64963091566.19	33963196558.83	3025141981.16	13731287039.48	90409741430.80
Cash & Due From Banks	4235760568.59	4432926882.02	178817946454.79	8140203739.57	2997920471.35	1140023768.17	5453646065.93	90550896249.97

Appendix C

Results of unit root tests

Table C.1 Panel Unit Root Test on TobinQ ratio from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	3.26524	0.9995
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	3.41664	0.9996
PP - Fisher Chi-square	2.80616	0.9999

Automatic lag length selection based on AIC: 0 to 2
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.2 Panel Unit Root Test on Efficiency calculated from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	1.27029	0.8980
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	18.9884	0.1654
PP - Fisher Chi-square	28.3023	.0130

Automatic lag length selection based on AIC: 0 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.3 Panel Unit Root Test on Effectiveness calculated from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	1.29154	.9017
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	10.6231	0.7154
PP - Fisher Chi-square	21.0399	0.1006

Automatic lag length selection based on AIC: 1 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.4 Panel Unit Root Test on log($TobinQ$) from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-4.40771	0.0000
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	34.6376	0.0045
PP - Fisher Chi-square	35.9576	0.0029

Automatic lag length selection based on AIC: 0 to 2
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.5 Panel Unit Root Test on log($Efficiency$) calculated from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-3.98417	0.0000
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	49.9638	0.0000
PP - Fisher Chi-square	95.9901	0.0000

Automatic lag length selection based on AIC: 0 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.6 Panel Unit Root Test on log($Effectiveness$) calculated from two-stage DEA model of Kumar and Gulati (2010)

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	.39032	0.6519
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	38.6487	0.0004
PP - Fisher Chi-square	233.819	0.0000

Automatic lag length selection based on AIC: 1 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.7 Panel Unit Root Test on TobinQ ratio from the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	3.26524	0.9995
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	3.41664	0.9996
PP - Fisher Chi-square	2.80616	0.9999

Automatic lag length selection based on AIC: 0 to 2
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.8 Panel Unit Root Test on Efficiency calculated from the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-2.88257	0.0020
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	27.9765	0.0318
PP - Fisher Chi-square	72.0144	0.0000

Automatic lag length selection based on AIC: 0 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.9 Panel Unit Root Test on Effectiveness calculated from the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-2.17594	0.0148
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	16.5236	0.4171
PP - Fisher Chi-square	37.6946	0.0017

Automatic lag length selection based on AIC: 1 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.10 Panel Unit Root Test on log($TobinQ$) for the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-4.40771	0.0000
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	34.6376	0.0045
PP - Fisher Chi-square	35.9576	0.0029

Automatic lag length selection based on AIC: 0 to 2
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.11 Panel Unit Root Test on log($Efficiency$) calculated from the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-2.62626	0.0043
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	26.2957	0.0500
PP - Fisher Chi-square	86.1522	0.0000

Automatic lag length selection based on AIC: 0 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Table C.12 Panel Unit Root Test on log($Effectiveness$) calculated from the best two-stage DEA model of Figure 4.3

Method	Statistics	Prob
Null: Unit root (assumes common unit root process)		
Levin, Lin & Chu t *	-4.94845	0.0000
Null: Unit root (assumes individual unit root process)		
ADF - Fisher Chi-square	58.1560	0.0000
PP - Fisher Chi-square	95.3418	0.0000

Automatic lag length selection based on AIC: 1 to 3
Newey-West automatic bandwidth selection and Bartlett kernel

Appendix D

Results of lag length selection

Table D.1 Lag length selection for Brazil

VAR Lag Order Selection Criteria
Endogenous variables: $\log(EFFECTIVENESS)$ $\log(EFFICIENCY)$ $\log(TOBINQ)$
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-86.24351	NA*	14.04544	11.15544	11.30030*	11.16286
1	-75.01009	16.85014	10.96936*	10.87626*	11.45570	10.90593*
2	-70.04715	5.583300	21.26519	11.38089	12.39492	11.43282

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	7.642302	9	0.5706	0.845680	(9,9.9)	0.5947
2	9.074484	9	0.4304	1.065222	(9,9.9)	0.4583

Table D.2 Lag length selection for India

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-79.22326	NA*	5.840202*	10.27791*	10.42277*	10.28533*
1	-71.38759	11.75351	6.974732	10.42345	11.00289	10.45312
2	-65.15270	7.014246	11.53361	10.76909	11.78311	10.82101

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	5.768273	9	0.7629	0.591697	(9,9.9)	0.7784
2	7.032899	9	0.6337	0.759157	(9,9.9)	0.6554

Table D.3 Lag length selection for China

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-66.31602	NA	2.073213	9.242136	9.383746	9.240628
1	-51.14030	22.25772	0.946208	8.418707	8.985147	8.412674
2	-34.20198	18.06754	0.399092	7.360265	8.351535	7.349705
3	0.200682	22.93511*	0.024427*	3.973242*	5.389343*	3.958158*

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	21.88068	9	0.0093	4.533797	(9,9.9)	0.0139
2	15.44141	9	0.0795	2.383014	(9,9.9)	0.0972
3	8.918947	9	0.4448	1.040226	(9,9.9)	0.4724

Table D.4 Lag length selection for USA

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-75.10589	NA	3.490675	9.763237	9.908097	9.770655
1	-60.43861	22.00093*	1.774771*	9.054826*	9.634267*	9.084498*
2	-56.43771	4.501008	3.880219	9.679714	10.69374	9.731640

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	21.88011	9	0.0093	4.533555	(9,9.9)	0.0139
2	6.732022	9	0.6650	0.717869	(9,9.9)	0.6854

Table D.5 Lag length selection for Canada

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-63.03741	NA	1.339038	8.804988	8.946598	8.803480
1	-42.94137	29.47420*	0.317117	7.325516	7.891956*	7.319482
2	-31.81612	11.86693	0.290347*	7.042150*	8.033420	7.031591*
3	-23.00502	5.874069	0.539031	7.067336	8.483436	7.052251

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	5.656932	9	0.7737	0.577697	(9,9.9)	0.7886
2	9.934932	9	0.3558	1.208870	(9,9.9)	0.3844
3	10.97829	9	0.2772	1.395909	(9,9.9)	0.3054

Table D.6 Lag length selection for Mexico

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-68.58709	NA	2.806415	9.544945	9.686555	9.543436
1	-50.56684	26.42969	0.876556	8.342246	8.908686	8.336212
2	-31.74117	20.08072*	0.287460	5.373415*	8.023426	7.021596
3	-10.30061	14.29370	0.099073*	7.032155	6.789515*	5.358330*

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	14.02848	9	0.1213	2.034261	(9,9.9)	0.1430
2	2.780536	9	0.9724	0.253594	(9,9.9)	0.9745
3	15.97452	9	0.0674	2.524469	(9,9.9)	0.0837

Table D.7 Lag length selection for South Korea

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-59.91043	NA*	0.882515	8.388057	8.529667*	8.386549
1	-49.77938	14.85887	0.789189*	8.237251	8.803691	8.231217
2	-40.32553	10.08411	0.902947	8.176737	9.168008	8.166178
3	-29.13299	7.461695	1.220281	7.884398*	9.300499	7.869314*

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	4.755562	9	0.8551	0.468547	(9,9.9)	0.8652
2	11.05579	9	0.2719	1.410391	(9,9.9)	0.3000
3	10.88894	9	0.2834	1.379314	(9,9.9)	0.3116

Table D.8 Lag length selection for Japan

VAR Lag Order Selection Criteria
Endogenous variables: log($EFFECTIVENESS$) log($EFFICIENCY$) log($TOBINQ$)
Exogenous variables: C

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-58.45532	NA*	0.726880*	8.194043	8.335653*	8.192535
1	-51.30736	10.48368	0.967521	8.440982	9.007422	8.434948
2	-41.47622	10.48656	1.052674	8.330162	9.321432	7.809636*
3	-28.68540	8.527210	1.149587	7.824720*	9.240820	8.319603

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

VAR Residual Serial Correlation LM Tests

Lag	LRE* stat	df	Prob.	Rao F-stat	df	Prob.
1	6.193624	9	0.7204	0.646265	(9,9.9)	0.7381
2	8.412644	9	0.4932	0.960845	(9,9.9)	0.5197
3	21.06323	9	0.0124	4.200134	(9,9.9)	0.0180

Appendix E

PCA results

Table E.1 Dot product results of the top 7 dimensions from Eikon across each of the 7 eigenvectors of the covariance of the dataset

$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$	$\mathbf{e_6}$	$\mathbf{e_7}$
0.43	0.35	0.28	0.36	0.32	0.34	0.30
0.35	0.31	0.25	0.32	0.28	0.32	0.29
0.34	0.31	0.22	0.30	0.28	0.30	0.28
0.33	0.31	0.22	0.28	0.27	0.29	0.27
0.28	0.30	0.21	0.26	0.26	0.29	0.27
0.27	0.28	0.20	0.21	0.24	0.28	0.25
0.25	0.27	0.20	0.19	0.24	0.27	0.25

Table E.2 Top 7 dimensions from Eikon whose dot product is the largest across each of the eigenvectors of the covariance of the dataset

$\mathbf{e_1}$	$\mathbf{e_2}$	$\mathbf{e_3}$	$\mathbf{e_4}$	$\mathbf{e_5}$	$\mathbf{e_6}$	$\mathbf{e_7}$
25	52	1	16	46	29	6
26	54	2	17	47	30	1
27	53	3	18	52	31	7
28	41	9	19	50	32	2
37	55	10	20	48	33	3
38	42	46	21	53	34	4
39	43	4	22	49	35	5

Appendix F

Summary of six permutations of two-stage DEA model

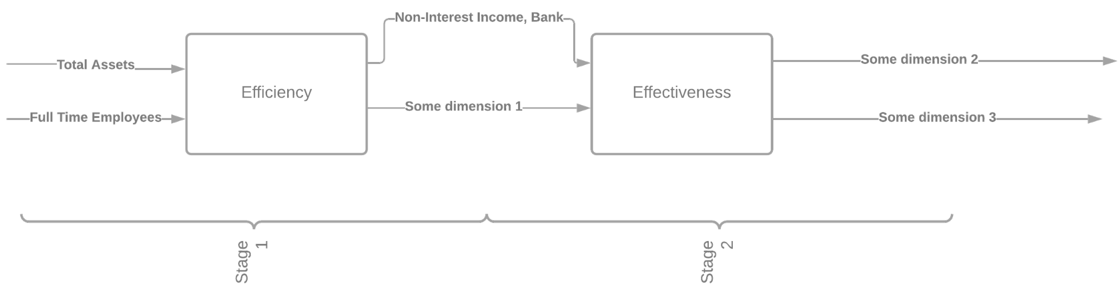

Figure F.1: In permutation 1, 'Total Assets' and 'Full time Employees' always occur as input variables of efficiency. 'Non Interest Income, Bank' always occurs as output variable of efficiency. 'Some dimension 1', 'Some dimension 2' and 'Some dimension 3' are three other variables chosen from the remaining 52 dimensions.

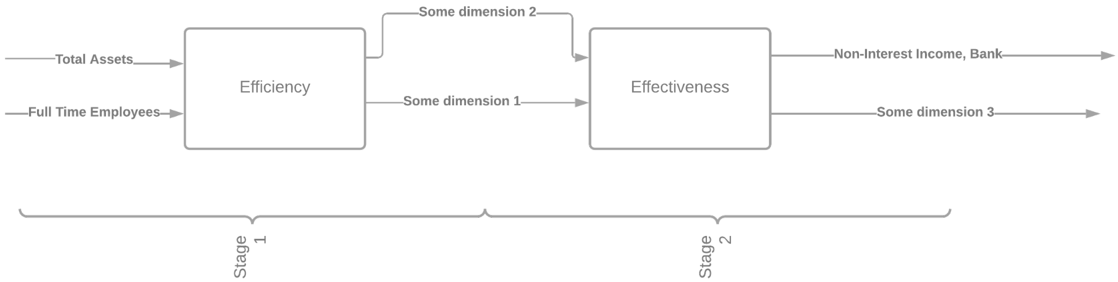

Figure F.2: In permutation 2, 'Total Assets' and 'Full time Employees' always occur as input variables of efficiency. 'Non Interest Income, Bank' always occurs as output variable of effectiveness. 'Some dimension 1', 'Some dimension 2' and 'Some dimension 3' are three other variables chosen from the remaining 52 dimensions.

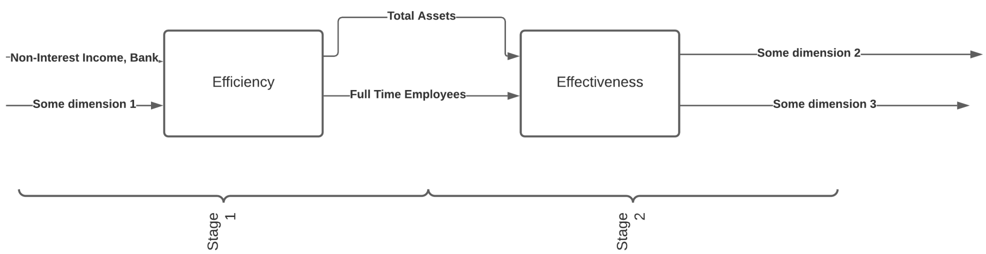

Figure F.3: In permutation 3, 'Total Assets' and 'Full time Employees' always occur as output variables of efficiency. 'Non Interest Income, Bank' always occurs as input variable of efficiency. 'Some dimension 1', 'Some dimension 2' and 'Some dimension 3' are three other variables chosen from the remaining 52 dimensions.

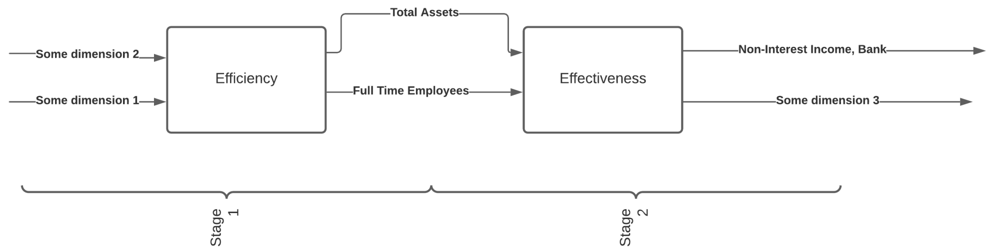

Figure F.4: In permutation 4, ‘Total Assets’ and ‘Full time Employees’ always occur as output variables of efficiency. ‘Non Interest Income, Bank’ always occurs as output variable of effectiveness. ‘Some dimension 1’, ‘Some dimension 2’ and ‘Some dimension 3’ are three other variables chosen from the remaining 52 dimensions.

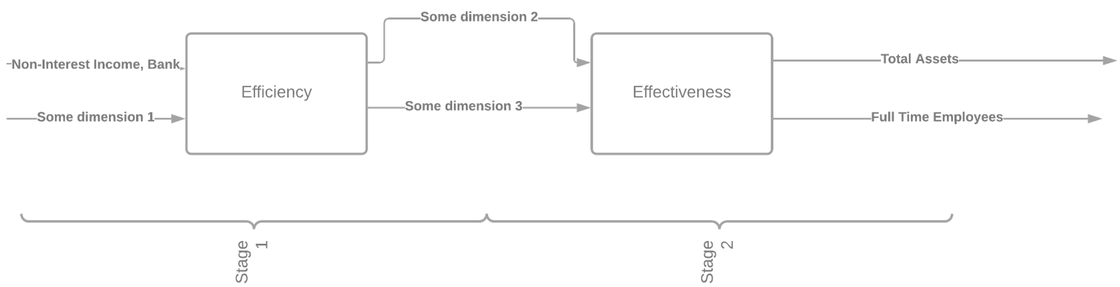

Figure F.5: In permutation 5, ‘Total Assets’ and ‘Full time Employees’ always occur as output variables of effectiveness. ‘Non Interest Income, Bank’ always occurs as output variable of efficiency. ‘Some dimension 1’, ‘Some dimension 2’ and ‘Some dimension 3’ are three other variables chosen from the remaining 52 dimensions.

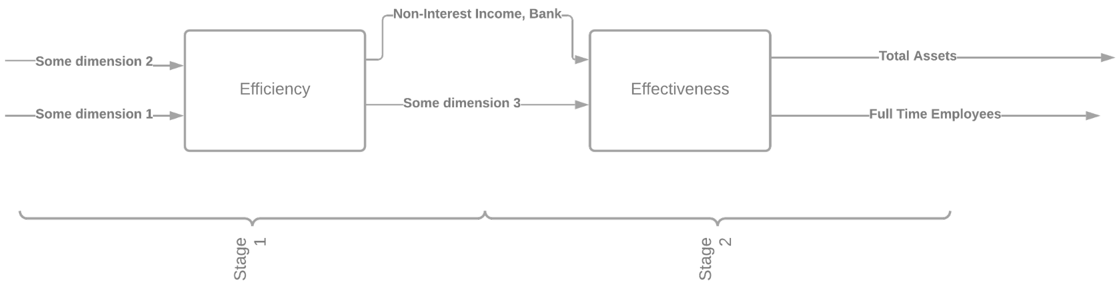

Figure F.6: In permutation 6, 'Total Assets' and 'Full time Employees' always occur as output variables of effectiveness. 'Non Interest Income, Bank' always occurs as output variable of efficiency. 'Some dimension 1', 'Some dimension 2' and 'Some dimension 3' are three other variables chosen from the remaining 52 dimensions.