

# Chance-constrained Cost Efficiency in Data Envelopment Analysis model with random inputs and outputs

Rashed Khanjani Shiraz<sup>1</sup>, Adel Hatami-Marbini<sup>2</sup>, Ali Emrouznejad<sup>3</sup>, Hirofumi Fukuyama<sup>4</sup>

1 School of Mathematical Science, University of Tabriz, Tabriz, Iran, rashed.shiraz@gmail.com

2 Department of Strategic Management and Marketing, Leicester Business School, De Montfort University, Hugh Aston Building, The Gateway, Leicester LE1 9BH, UK, adel.hatamimarbini@dmu.ac.uk

3 Operations and Information Management Group, Aston Business School, Aston University, Birmingham, UK, a.emrouznejad@aston.ac.uk

4 Faculty of Commerce, Fukuoka University, 8-19-1 Nanakuma, Jonan-Ku, Fukuoka City, Fukuoka 814-0180, Japan, fukuyama@fukuoka-u.ac.jp

## Abstract

Data envelopment analysis (DEA) is a well-known non-parametric technique primarily used to estimate radial efficiency under a set of mild assumptions regarding the production possibility set and the production function. The technical efficiency measure can be complemented with a consistent radial metrics for cost, revenue and profit efficiency in DEA, but only for the setting with known input and output prices. In many real applications of performance measurement, such as the evaluation of utilities, banks and supply chain operations, the input and/or output data are often stochastic and linked to exogenous random variables. It is known from standard results in stochastic programming that rankings of stochastic functions are biased if expected values are used for key parameters. In this paper, we propose economic efficiency measures for stochastic data with known input and output prices. We transform the stochastic economic efficiency models into a deterministic equivalent non-linear form that can be simplified to a deterministic programming with quadratic constraints. An application for a cost minimizing planning problem of a state government in the US is presented to illustrate the applicability of the proposed framework.

**Keywords:** Data envelopment analysis; Weight restrictions; random input-output; Cost efficiency; Quadratic programming.

## 1. Introduction

Data envelopment analysis (DEA) is recognized as a powerful analytical tool that is widely used in measuring the relative efficiency of a group of decision making units (DMUs) with multiple inputs and multiple outputs. The first DEA model has been presented by Charnes et al. (1978) in the case of constant returns to scale (CRS) and later extended by Banker et al. (1984) in the case of variable returns to scale (VRS) for evaluating the technical efficiency of a set of comparable DMUs. A substantial number of DEA studies have been rapidly developed since 1978 and the evolution of the DEA scientific area can be found in Cook and Seiford (2009), Emrouznejad and De Witte (2010) and Liu et al. (2013).

In the conventional input-oriented DEA models<sup>1</sup>, the efficiency of the DMU to be evaluated is measured by making a comparison of its observed input vector with a projected point on an input-isoquant. That is, DEA emphasizes technical efficiency measurement by utilizing the radial measures, which are gauged relative to the input-isoquant by seeking the maximal equi-proportional reduction in all the inputs of the DMU that would be feasible for a given output vector. However, the radial input projection corresponding to the input-isoquant may not be located on the efficient (Pareto-Koopmans) input-frontier. Hence, after obtaining the radial projection, solving an additional optimization problem is needed. More precisely, in the radial-based DEA, it is common to use a two-phase procedure where the radial efficiency is estimated in the first phase and the input slack maximization problem is solved in the second phase. An alternative procedure is maximization of input slack by directly solving slack-based measures or input-oriented additive model. In either way, a projection point based on only input slacks (input surpluses) may not be a cost-minimizing vector, which is very important from the economic theoretic and managerial viewpoints. Cost minimization refers to the firm's decisions on the choice of input quantities given its output level and input prices.

Consider DMUs that minimize costs but do not maximize profits. Such DMUs include, not only non-profit organizations and cooperative firms, but also firms which for example struggle under economic depression and hence the output expansion is not possible. Under cost minimization, management chooses efficient input combinations but beyond that no particular criterion is implemented for choosing a specific output combination (Luenberger 1995). As is well-known, in competitive input and output markets profit maximization implies cost minimization but not vice versa.

According to microeconomic theory, marginal, average and total cost functions are major

---

<sup>1</sup>We can also consider output oriented models analogously.

tools for production analysis, implying that cost minimization is one of the basic norms in economic analysis. However, in real world situations, there exist inefficient firms due to excess use of their input-mix, in which case it is necessary to provide them with optimal input bundles.

For managers of economic entities too, taking cost performance into account is of great necessity because cost minimizing behaviour is at the core of managerial objectives. However, achieving such an objective may not be easy when they face uncertainty in data for which case the standard cost efficiency (CE) analysis cannot provide a practical solution. This necessitates implementing the stochastic nature in CE analysis.

A variety of production-based DEA models of stochastic programming have been developed for performance evaluation of DMUs in the various fields with different types of data such as deterministic, imprecise, interval and fuzzy data. Charnes and Cooper's (1959) stochastic programming<sup>2</sup> is one of the most commonly used methods. In order to enhance the practicality, Charnes and Cooper (1963) further presented a chance-constrained programming (CCP) model, in which a stochastic linear programming problem was transformed into a deterministic non-linear programming problem, and it was first utilized by Land et al. (1993) to deal with data stochasticity in production-based DEA.

Cooper et al. (2000a, b, 2004) further extended the CCP model into a congestion framework in DEA. While we recognise the usefulness of CCP, there does not exist any cost-based chance-constrained DEA model, which has the good theoretical foundation as well as good practicality. Therefore, we propose a cost-based DEA model that directly deals with stochastic input and output data for the purpose of increasing the realism of the CE analysis. To the best of our knowledge, the present study is the first one to develop a CE-DEA model using the stochastic input and output data along the line of chance-constrained programming introduced by Charnes and Cooper (1963) and Cooper et al. (1996). Our extension adapts Cooper et al.'s (2002a, b, 2004) production-based approach.

To show the applicability of our model we provide an illustrative example using the data provided by Ray et al. (2008). The dataset consists of US state-level data collected from the Economic Census. Although the original data values are nonstochastic and single-valued, we show how to use them for a planning purpose by introducing stochasticity in the data.

---

<sup>2</sup>Stochastic programming has been nowadays extended into numerous disciplines including operations research, mathematics and probability.

The remainder of this paper is organized as follows: In Section 2, we present a brief literature survey. In Section 3 we provide a review of the principle DEA models. Section 4 presents an extension of CE with stochastic input and output data. In Section 5, we develop a slack-based version of CE in the presence of stochastic data. Section 6 is dedicated to illustrate the application of the proposed approach using a case study for a cost minimizing planning problem of a state government in the US. Section 7 finally concludes our work and points out some directions of further work.

## **2. Selective literature**

Some observations can be located on the efficient frontier in the deterministic DEA, while some stochastic inputs and outputs are by definition allowed to be around the efficient frontier can be allowed with the aim of conceptualizing the stochastic nature of the data into the model to adapt the measurement and specification errors. Stochastic input and output variations in DEA have been studied within various input-output DEA contexts by many scholars (see e.g., Olesen and Petersen, 2015, Olesen and Petersen, 1995; Huang and Li, 1996; Cooper et al., 1996, 1998, 2002, 2004; Land et al., 1993; Morita and Seiford, 1999; Sueyoshi, 2000; Talluri et al., 2006; Olesen, 2006; Bruni et al., 2009; Wu and Lee, 2010; Tsionas and Papadakis, 2010; Udhayakumar et al., 2011).

Land et al. (1993) were the first to extend the chance-constrained programming (CCP) DEA proposed by Charnes and Cooper (1959), in order to compute efficiency in the presence of uncertainty in which inputs are assumed to be deterministic and outputs are jointly normally distributed. The CCP DEA is a non-parametric approach to evaluate the efficiency while stochastic frontier analysis (SFA) is an alternative parametric approach. The parametric approach puts emphasis on the production or cost function along with studying the characteristics of the functions under the presumption that all DMUs operate under rational behavior. In the literature, relatively few papers on efficiency analysis have analytically compared parametric and non-parametric approaches (e.g., Bjurek et al. 1990; Cooper and Tone 1997). Bjurek et al. (1990) argued that there is no significant difference between deterministic DEA and a loglinear parametric model. Cooper and Tone (1997) discussed DEA and stochastic cost functions, identifying some particular problems of bias in SFA approaches.

To measure the efficiency, several researchers have further extended the concept of the stochastic production function. Coelli (1996) presented a method to estimate the maximum likelihood estimator of SFA developed by Battese and Coelli (1992). In SFA, the inefficiency

effect can be used as how far the DMU operates below the frontier production function. Coelli (1996) indicated that the inefficiency effects can be referred to as the technical inefficiency.

Olesen and Petersen (1995) developed a chance constrained DEA model by imposing chance constraints on the multiplier model. Cooper et al. (1996) presented a joint chance constraints programming model in the multiplier DEA model. They used “satisficing concepts” presented by Simon (1957) to develop the potential applications of DEA models to situations where 100% efficiency can be replaced by aspired levels of performance. Huang and Li (1996) proposed a dominance structure to remove the anomalous (Pareto) efficient DMUs from the DEA envelopment side where input and output data are characterized by random variations. Cooper et al. (1998) introduced the “alpha-stochastic efficiency” and “alpha-stochastic efficiency dominance” of a DMU in stochastic DEA using the joint chance constraints where random disturbances are applied in the inputs and outputs on the assumption that the statistical distributions of data are known. Morita and Seiford (1999) studied robustness of the efficiency results when input and output data are subject to the stochastic measurement error. Sueyoshi (2000) put forward a stochastic DEA model to restructure strategy of a Japanese petroleum company, which is reformulated in the manner that ex ante information can be incorporated into the stochastic model. Cooper et al. (2002a) proposed a generalization of Cooper et al. (2002b)’s CCP model for identifying the technical efficiencies and inefficiencies. Cooper et al. (2004) extended congestion in DEA models based on CCP models. Talluri et al. (2006) applied chance constrained DEA model for vendor evaluation. Olesen (2006) presented a comparison of two different models (Land et al., 1993; Olesen and Petersen, 1995), both designed to extend DEA to the case of stochastic inputs and outputs. Lahdelma and Salminen (2006) proposed the SMAA-D method by combining DEA and SMAA-2 (stochastic multicriteria acceptability analysis) presented in (Lahdelma et al. 1998; Lahdelma and Salminen, 2001) which can be dealt with uncertain or imprecise data to provide stochastic efficiency measures. Bruni et al. (2009) proposed a stochastic DEA model based on the theory of joint probabilistic constraints to extend the concept of “Stochastic efficiency” to a measure called “alpha-stochastic efficiency”. Udhayakumar et al. (2011) exploited the genetic algorithm method to solve the chance constrained DEA model, which involves the concept of *satisficing*.

Tavana et al. (2012) developed three imprecise DEA models in the presence of probability-possibility, probability-necessity and probability-credibility constraints where fuzziness and randomness simultaneously are allowed to exist in an efficiency evaluation problem. Tavana et

al. (2013) introduced random fuzzy variables in DEA with the coexistence of randomness and vagueness. In this respect, they proposed three DEA models for measuring the radial efficiency of DMUs when the input and output data are random fuzzy variables with Poisson, uniform and normal distributions. Tavana et al. (2014) proposed a chance-constrained DEA model with birandom input and output data as well as formulating a super-efficiency model with birandom constraints that was solved by means of an equivalent non-linear deterministic model.

Although many contributions in CE have been reported in the DEA literature (see, e.g., Färe et al., 1985; Schaffnit et al., 1997; Jahanshahloo et al., 2008; Thompson et al., 1996; Camanho and Dyson, 2008), only a few researchers have considered uncertainty in the CE context (see e.g., Kuosmanen and Post 2001, 2002, 2003; Camanho and Dyson 2005; Mostafaei and Saljooghi, 2010). Kuosmanen and Post (2001) developed a DEA-based method to determine the upper and lower bounds for overall economic efficiency and allocative efficiency with incomplete price data in the form of a convex polyhedral cone. Kuosmanen and Post (2003) corrected a technical error concerning the lower bound of Kuosmanen and Post (2001). Kuosmanen and Post (2002) discussed necessary and sufficient first-order stochastic dominance (FSD) efficiency conditions for economic efficiency with regarding the preferences of the decision-maker and the statistical distribution of the prices. Camanho and Dyson (2005) developed weight restrictions methods for estimating cost efficiency bounds in complex scenarios of input price uncertainty under the assumption that input prices are in the form of ranges. However, Camanho and Dyson's (2005) model is not only computationally expensive but also may produce an infeasible solution (Mostafaei and Saljooghi, 2010). Amirteimoori et al (2006) proposed two DEA-based models to obtain the upper and lower bound of cost efficiency of each DMU with respect to the optimistic and pessimistic viewpoints, respectively. They then introduced a two-step procedure to improve the cost efficiency of DMUs.

Mostafaei and Saljooghi (2010) proposed a pair of two-level mathematical programming models to obtain the upper and lower bounds of the cost efficiency where the uncertain input prices are in the form of ranges. Bagherzadeh Valami (2009) provided an approach for generalizing the CE of DMUs when input prices are characterized by triangular fuzzy numbers. Fang and Li (2012) presented some counterexamples to show that the theorem of Mostafaei and Saljooghi (2010) on determination of the upper bound of CE is not correct in general and they then provided an alternative proof to clarify the reason. Fang and Li (2013)

studies the theoretical properties on the relationships and characteristics of the efficiency solutions between cone-ratio DEA models and CE models in situations of price uncertainty, where the upper and lower bounds of the input prices can be estimated for each DMU. They additionally developed a method and a lexicographic order algorithm based on the duality study to estimate the lower bounds of the CE measure.

The above literature shows the lack of much attention to the CE analysis under uncertainty due to the degrees of complexity while real-world problems often include uncertain data, particularly the cases with stochastic input and output data. As can be seen in the aforesaid literature, no study has been dealt with stochastic data in CE. In this study, we strive to fill this gap by extending a CE-DEA model when inputs and outputs are stochastic. By exploiting the chance-constrained approach (Cooper et al., 1996) and converting the stochastic model to deterministic programming with quadratic constrains, we decrease the degree of complicated calculus and handle the embedded non-linearity.

### 3. Preliminaries

In this section, we first review the basic DEA models for measuring the technical efficiency, and we then present the non-parametric cost efficiency models.

#### 2.1 Technical efficiency

Suppose that there are  $n$  DMUs to be evaluated where each DMU produces  $s$  outputs using  $m$  inputs. Let  $X_j = (x_{1j}, x_{2j}, \dots, x_{mj}) \in \mathbb{R}_+^m$  and  $Y_j = (y_{1j}, y_{2j}, \dots, y_{sj}) \in \mathbb{R}_+^s$  be the observed input and output vectors of DMU $_j$  ( $j=1, \dots, n$ ). The production technology or production possibility set (PPS),  $T_c$ , is defined, as  $T_c = \{(x, y) \mid x \text{ can produce } y\}$ .

Accordingly, we consider the following assumptions to construct production technology without determining any functional form:

- a) Free disposability:  $(x, y) \in T_c, x' \geq x, 0 \leq y' \leq y \Rightarrow (x', y') \in T_c$ .
- b) Convexity:  $T_c$  is convex.
- c)  $r$  returns to scale:  $(x, y) \in T_c \Rightarrow (qx, qy) \in T_c, \forall q \in Y(r)$  where  $Y(crs) = \mathfrak{R}_0$ .

The production technology based on the observations and the assumptions (a-c) is expressed as follows:

$$T_c = \left\{ (x, y) : \sum_{j=1}^n \lambda_j y_j \geq y, \sum_{j=1}^n \lambda_j x_j \leq x, \lambda_j \geq 0, j = 1, \dots, n \right\} \quad (1)$$

We evaluate the technical efficiency of firm  $p$ , producing output  $y_{rp}$  using input  $x_{ip}$ , by examining whether and to what extent it can reduce its inputs without decreasing the outputs (so-called the input-oriented model) or augment its outputs without increasing the inputs (so-called the output-oriented model). Mathematically speaking, we measure the efficiency of DMU $_p$  under the assumption of constant returns to scale (CRS) using the following linear programming problems:

**Primal CCRmodel (input-oriented)**

$$\begin{aligned}
& \min \theta_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j x_{ij} \leq \theta_p x_{ip}, \quad i = 1, \dots, m, \\
& \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s, \\
& \lambda_j \geq 0, \quad j = 1, \dots, n.
\end{aligned} \tag{2}$$

**Dual CCRmodel (input-oriented)**

$$\begin{aligned}
& \max \theta_p = \sum_{r=1}^s u_r y_{rp} \\
& s.t. \\
& \sum_{i=1}^m v_i x_{ip} = 1, \\
& \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n, \\
& u_r, v_i \geq 0, \quad r = 1, \dots, s; i = 1, \dots, m.
\end{aligned} \tag{3}$$

where  $u_r$  ( $r = 1, \dots, s$ ) and  $v_i$  ( $i = 1, \dots, m$ ) in model (2) are the weights assigned to the  $r^{\text{th}}$  output and  $i^{\text{th}}$  input, respectively. The primal and dual of programs are called *envelopment* and *multiplier* DEA models, respectively. The formulations (2) and (3) can be converted to the DEA models under the assumption of variable returns to scale (VRS) by adding the convexity constraint  $\sum_{j=1}^n \lambda_j = 1$  and a free variable,  $u_0$ , respectively. Notice that the objective function of models (2) and (3) represents the best relative efficiency and the DMUs with  $\theta_p^* = 1$  ( $\phi_p^* = 1$ ), are called the technically input-efficient, and those units with  $\theta_p^* \neq 1$  ( $\phi_p^* \neq 1$ ) are called technically input-inefficient.

## 2.2 Cost efficiency



It is evident that technical efficiency alone does not necessarily imply cost minimization, whereas the opposite is true. To fully investigate the sources of inefficiency into its components of cost, allocative, scale and technical efficiency, we need to decompose the radial measures. In this section, we define the decomposition beginning from the most important; the radial cost efficiency measure.

Consider an empirically constructed PPS of (1) with  $n$  DMUs under the CRS assumption. Let  $w_{ip}$  be the given price for the  $i^{\text{th}}$  input of DMU <sub>$p$</sub> . Then, the minimum cost for DMU <sub>$p$</sub>  with  $w_{ip}$  input prices is obtained by solving the following linear program (Färe et al. 1985):

$$\begin{aligned}
& \min \sum_{i=1}^m w_{ip} x_i \\
& s.t. \\
& \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s, \\
& \sum_{j=1}^n \lambda_j x_{ij} \leq x_i, \quad i = 1, \dots, m, \\
& \lambda_j \geq 0, \quad x_i \geq 0, \quad j = 1, \dots, n; i = 1, \dots, m.
\end{aligned} \tag{4}$$

where  $\lambda_j$  and  $x_i$  are the decision variables. The second set of inequality constraints can be converted to equality constraints. The optimal solution of (4) yields  $x_i^*$  as an optimal level of input  $i$  for producing the current outputs at minimal cost. By the use of the objective function value of model (4), the cost efficiency of DMU <sub>$p$</sub>  is obtained as  $CE = \sum_{i=1}^m w_{ip} x_i^* / \sum_{i=1}^m w_{ip} x_{ip}$  where CE varies in  $[0, 1]$ .

CE evaluates the ability to produce current outputs at minimal cost by a firm and Farrell's decomposition of CE consists of multiplication of three components as

$$cost\ efficiency\ (CE) = technical\ efficiency\ (TE) \times allocative\ efficiency\ (AE)$$

where *cost efficiency* and *technical efficiency* are calculated by models (4) and (2). *Allocative efficiency* gauges to what extent the cost of the DMU can be scaled down when the selected inputs are most suitable for the input price ratio faced by the DMU in a given situation.

The market prices or managerial information enable us to determine bounds on ratio of pairs of weights. This is often referred to weight restrictions in DEA. Camanho and Dyson (2005) proposed a multiplier DEA-based model under weight restrictions to measure the CE of the DMUs in complex scenarios of price uncertainty as follows:

$$\begin{aligned}
& \max \quad \sum_{r=1}^s u_r y_{rp} \\
& s.t. \\
& \sum_{i=1}^m v_i x_{ip} = 1, \\
& \sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0, \quad j = 1, \dots, n, \\
& v_k = \frac{w_{kp}}{w_{gp}} v_g, \quad k < g, \quad k, g = 1, \dots, m, \\
& u_r \geq 0, \quad r = 1, \dots, s.
\end{aligned}$$

where  $v_k$  and  $v_g$  are the input weights of the cost efficiency, and  $w_{kp}$  and  $w_{gp}$  are the input prices of DMU<sub>p</sub>, for any two inputs  $k$  and  $g$  used by DMU. Jahanshahloo et al. (2008) relaxed some constraints of Camanho and Dyson (2005)'s model and proposed the following model with lower computational complexity.

$$\begin{aligned}
& \max \quad \sum_{r=1}^s u_r y_{rp} \\
& s.t. \\
& \sum_{r=1}^s u_r y_{rj} \leq \frac{\sum_{i=1}^m w_{ip} x_{ij}}{\sum_{i=1}^m w_{ip} x_{ip}}, \quad j = 1, \dots, n, \\
& u_r \geq 0, \quad r = 1, \dots, s.
\end{aligned} \tag{5}$$

On the basis of Schaffnit et al. (1997) it can be shown that model (5) is equivalent with the CE measure calculated by Camanho and Dyson (2005). The dual of the above problem can be written as follows:

$$\begin{aligned}
& \min \quad \sum_{j=1}^n \lambda_j \left( \frac{\sum_{i=1}^m w_{ij} x_{ij}}{\sum_{i=1}^m w_{ip} x_{ip}} \right) \\
& s.t. \\
& \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s, \\
& \lambda_j \geq 0, \quad j = 1, \dots, n.
\end{aligned} \tag{6}$$

By using a variable transformation  $\theta_p = \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ij} x_{ij} / \sum_{i=1}^m w_{ip} x_{ip} \right)$  model (5) is converted to the following model:

$$\begin{aligned}
& \min \theta_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ij} x_{ij} \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip}, \\
& \sum_{j=1}^n \lambda_j y_{rj} \geq y_{rp}, \quad r = 1, \dots, s, \\
& \lambda_j \geq 0, \quad j = 1, \dots, n.
\end{aligned} \tag{7}$$

Note that we take  $\sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ij} x_{ij} \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip}$  into account in model (7) instead of

$$\sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ij} x_{ij} = \theta_p \sum_{i=1}^m w_{ip} x_{ip}.$$

Based on the production technology (1), the cost-based production technology set can be defined as

$$T_C = \left\{ (\bar{x}, y) : \sum_{j=1}^n \lambda_j y_j \geq y, \sum_{j=1}^n \lambda_j \bar{x}_j \leq \bar{x}, \lambda_j \geq 0, j = 1, \dots, n \right\} \text{ where } \bar{x}_j = \sum_{i=1}^m w_{ip} x_{ij}.$$

As a result, model (7) is the alternative version of the CE model (4) and its optimal value is the CE of  $DMU_p$ .

Similarly, the revenue efficiency model in terms of the weight-restricted standard [output-oriented] DEA model can be formulated as follows:

$$\begin{aligned}
& \max \varphi_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \sum_{r=1}^s \rho_{rp} y_{rj} \geq \varphi_p \sum_{r=1}^s \rho_{rp} y_{rp}, \quad r = 1, \dots, s, \\
& \sum_{j=1}^n \lambda_j x_{ij} \leq x_{ip}, \quad i = 1, \dots, m, \\
& \lambda_j \geq 0, \quad j = 1, \dots, n.
\end{aligned} \tag{8}$$

where  $\rho_{rp}$  are the output prices. The procedure of obtaining model (8) represents in Appendix A.

Models (7) and (8) need the deterministic input and output data for each DMU although, in many real-world applications, the data often involve uncertainty.

#### 4. Cost efficiency with stochastic input-output data

As discussed before, in measuring efficiency of the firms, the data may involve stochastic variations and stochastic programming is one of the main models to deal with uncertainty in many decision-making problems (Charnes and Cooper, 1959). In this section, we extend the stochastic version of models (7) and (8) to evaluate the cost and revenue efficiency of units

with weight restrictions. Let us assume that  $\tilde{X}_j = (\tilde{x}_{1j}, \dots, \tilde{x}_{mj})^T \in \mathfrak{R}_+^m$  and  $\tilde{Y}_j = (\tilde{y}_{1j}, \dots, \tilde{y}_{sj})^T \in \mathfrak{R}_+^s$  are the random input and output vectors for DMU<sub>j</sub>,  $j = 1, \dots, n$ , and each of them has a normal distribution. Let us also assume,  $X_j = (x_{1j}, \dots, x_{mj})^T \in \mathfrak{R}_+^m$  and  $Y_j = (y_{1j}, \dots, y_{sj})^T \in \mathfrak{R}_+^s$  are the expected vectors of the inputs and outputs of  $\tilde{X}_j$  and  $\tilde{Y}_j$ , respectively. Let  $W_j = (w_{1j}, \dots, w_{mj})^T \in \mathfrak{R}_{++}^m$  be the input prices of DMU<sub>j</sub>.

When the stochastic input and output data are available the weight-restricted cost efficiency model (7) can be rewritten by the following stochastic programming:

$$\theta_p^* = \min \theta_p$$

s.t.

$$\begin{aligned} Pr \left[ \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) \leq \theta \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right) \right] &\geq 1 - \alpha, \\ Pr \left[ \sum_{j=1}^n \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{rp} \right] &\geq 1 - \alpha, \quad r = 1, \dots, s, \\ \lambda_j &\geq 0, \quad j = 1, \dots, n. \end{aligned} \tag{9}$$

where  $Pr$  denotes “probability” and “ $\sim$ ” presents the data as random variables with a normal distribution while  $\alpha \in (0, 1]$  is a pre-defined scalar for identifying an allowable chance of failing to satisfy the constraints. In Theorem 1 we will specify the deterministic form of model (9) that can be solved by the General Algebraic Modeling System (GAMS) software as well as studying the connection of (9) with the earlier discussed model (7).

**Theorem 1:** Consider the stochastic weight-restricted cost efficiency model (9). The corresponding deterministic equivalent of model (9) can be expressed as follows:

$$\begin{aligned}
& \min \theta_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha) v \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip} \\
& \sum_{j=1}^n \lambda_j y_{rj} + \Phi^{-1}(\alpha) u_r \geq y_{rp}, \quad r = 1, \dots, s \\
& v^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta_p^2 \sum_{i=1}^m w_{ip}^2 \text{Var}(\tilde{x}_{ip}) - \\
& \quad 2\theta_p \sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ip}), \quad i = 1, \dots, m, \\
& u_r^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) + \text{Var}(\tilde{y}_{rp}) - 2 \sum_{j=1}^n \lambda_j \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rp}), \quad r = 1, \dots, s, \\
& \lambda_j, v, u_r \geq 0, \quad r = 1, \dots, s; i = 1, \dots, m; j = 1, \dots, n.
\end{aligned} \tag{10}$$

where  $\Phi^{-1}$  is the inverse of cumulative distribution function (CDF) and  $\text{Var}(\cdot)$  and  $\text{Cov}(\cdot, \cdot)$  are the variance and covariance operators.

**Proof.** See Appendix B.

**Lemma:** If we assume that the outputs and inputs among different DMUs are independent, i.e.,  $\text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) = 0$  and  $\text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) = 0$ . This independence assumption in model (10) leads to the following model:

$$\begin{aligned}
& \min \theta_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha) v \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip}, \\
& \sum_{j=1}^n \lambda_j y_{rj} + \Phi^{-1}(\alpha) u_r \geq y_{rp}, \quad r = 1, \dots, s, \\
& v^2 = \sum_{j=1}^n \lambda_j^2 \sum_{i=1}^m w_{ip}^2 \text{Var}(\tilde{x}_{ij}) + (\theta_p^2 - 2\theta_p \lambda_p) \sum_{i=1}^m w_{ip}^2 \text{Var}(\tilde{x}_{ip}), \\
& u_r^2 = \sum_{j=1}^n \lambda_j^2 \text{Var}(\tilde{y}_{rj}) + \text{Var}(\tilde{y}_{rp}) (1 - 2\lambda_p), \quad r = 1, \dots, s, \\
& \lambda_j, v, u_r \geq 0, \quad r = 1, \dots, s; j = 1, \dots, n.
\end{aligned} \tag{11}$$

**Proof.** Let

$$\lambda_j = \begin{cases} 1, & j = p \\ 0, & j \neq p \end{cases}, \quad j = 1, \dots, n, \quad \theta_p = 1.$$

Obviously, it can be indicated that  $v=0$ . Therefore, this solution is feasible for model (10). ■

It should be emphasised that the factor  $\alpha$  embedded in the developed stochastic weight-restricted cost efficiency model plays an important role in determining the cost efficiency score for each DMU. We, therefore, focus on the role of  $\alpha$  in the following discussion to highlight the effect of the  $\alpha$  value on the cost efficiency score of a DMU.

**Proposition 2:** Let  $\alpha \leq 0.5$ . Then the objective function of model (10) varies in  $0 < \theta_p^* \leq 1$ .

**Proof.** Let  $\lambda_p = 1, \lambda_j = 0, \forall j \neq p$  and  $\theta = 1$  for all  $j \neq p$ . Then  $v_i = 0, u_r = 1$  and all constraints of model (10) will be satisfied by this solution. Due to the minimization of model (10), the upper bound of  $\theta_p^*$  is less than or equal to unity.

Assume that  $\theta_p^* > 0$ . Then,  $-\Phi^{-1}(\alpha) \geq 0$  and  $v \geq 0$  with regards to  $\alpha \leq 0.5$  and, as a result, the first constraint of model (10) is converted to the following inequality:

$$\theta_p \geq \frac{\sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right)}{\sum_{i=1}^m w_{ip} x_{ip}} \geq 0$$

Hence,  $\theta_p^* > 0$ .

Now let us assume  $\theta_p^* \leq 0$ . Then  $\lambda_j = 0$ . From the second constraint of model (6) we have  $y_{rp} \leq 0$  which contradicts with  $y_{rp} \geq 0$ . Therefore, the upper bound of  $\theta_p^*$  is bigger than zero.

This completes the proof. ■

**Remark:** When  $\alpha > 0.5$ , the objective function of model (10) may be negative (i.e.,  $\theta_p^* < 0$ ).

According to the above remark and proposition 2, the stochastic cost efficiency can be calculated using model (9) under  $\alpha \leq 0.5$  condition.

**Proposition 3:** The proposed stochastic cost efficiency model (10) always results in at least one efficient DMU i.e., there exists at least one  $k \in \{1, \dots, n\}$  such that  $\theta_k^* = 1$ .

**Proof.** See Appendix C.

**Proposition 4:** Let  $\alpha \leq 0.5$ . If  $\alpha$  decreases in model (10), then stochastic cost efficiency increases or unchanged.

**Proof:** See Appendix D.

**Proposition 5:** If  $\alpha \leq 0.5$ , the optimal solution of model (10) is unique.

**Proof:** See Appendix E. ■

**Definition (Stochastic cost efficiency):**  $DMU_p$  is stochastically cost efficient if and only if the objective function of model (10) is equal to 1, i.e.,  $\theta_p^* = 1$ .

Here, we extend the above definition to the *revenue* efficiency not only because revenue is a pivotal intent for both public and private firms, but also in many circumstances political pressure may push some organizations to sell products to domestic consumers at subsidized prices. However, the observed values of inputs and outputs in real-world problems are often uncertain. We determine the weight-restricted revenue efficiency model with stochastic data by re-formulation of model (8) as

$$\begin{aligned}
& \max \quad \varphi \\
& s.t. \\
& Pr \left[ \sum_{j=1}^n \lambda_j \sum_{r=1}^s \rho_{rp} \tilde{y}_{rj} \geq \varphi \sum_{r=1}^s \rho_{rp} \tilde{y}_{rp} \right] \geq 1 - \alpha, \quad r = 1, \dots, s, \\
& Pr \left[ \sum_{j=1}^n \lambda_j \tilde{x}_{ij} \leq \tilde{x}_{ip} \right] \geq 1 - \alpha, \quad i = 1, \dots, m \\
& \lambda_j \geq 0, \quad j = 1, \dots, n.
\end{aligned} \tag{12}$$

Similar to the prior formulation on the stochastic cost efficiency model, the deterministic equivalent to the stochastic revenue efficiency model (12) is formulated as:

$$\begin{aligned}
& \max \quad \varphi \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \left( \sum_{r=1}^s \rho_{rp} y_{rj} \right) + \Phi^{-1}(\alpha) u \geq \varphi \sum_{r=1}^s \rho_{rp} y_{rp}, \\
& \sum_{j=1}^n \lambda_j x_{ij} - \Phi^{-1}(\alpha) v \leq x_{ip}, \quad i = 1, \dots, m \\
& u^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \left( \sum_{r=1}^s \rho_{rp} Cov(\tilde{y}_{rj}, \tilde{y}_{rk}) \right) + \varphi^2 \sum_{r=1}^s \rho_{rp}^2 var(\tilde{y}_{rp}) - 2\varphi \sum_{j=1}^n \lambda_j \left\{ \sum_{r=1}^s \rho_{rp}^2 Cov(\tilde{y}_{rj}, \tilde{y}_{rp}) \right\} \\
& v_i^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k Cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + Var(\tilde{x}_{ip}) - 2 \sum_{j=1}^n \lambda_j Cov(\tilde{x}_{ij}, \tilde{x}_{ip}), \quad i = 1, \dots, m, \\
& \lambda_j \geq 0, u \geq 0, v_i \geq 0; i = 1, \dots, m; j = 1, \dots, n.
\end{aligned} \tag{13}$$

One of the referees asked us to investigate whether the proposed method can be extended into a non-radial efficiency measure along with decomposing cost efficiency into allocative and technical efficiency with stochastic input-output data. In response to the first question,

we develop a stochastic version of SBM measure by using the expected value operator in the ensuing section. For the second question, we develop a stochastic version of CCR (SCCR) technical efficiency measure (SCCR) and the corresponding decomposition of stochastic cost efficiency (SCE) as represented below.

As noted in Section 3, cost efficiency can be decomposed into *allocative efficiency* and *technical efficiency* in a non-stochastic situation. In the case of stochastic inputs and outputs, we first develop the following stochastic CCR technical efficiency model:

$$\begin{aligned}
 \theta_p^{SCCR} &= \min \theta_p \\
 \text{s.t.} \\
 Pr \left[ \sum_{j=1}^n \lambda_j \tilde{x}_{ij} \leq \theta_p \tilde{x}_{ip} \right] &\geq 1 - \alpha, i = 1, \dots, m, \\
 Pr \left[ \sum_{j=1}^n \lambda_j \tilde{y}_{rj} \geq \tilde{y}_{rp} \right] &\geq 1 - \alpha, r = 1, \dots, s, \\
 \lambda_j &\geq 0, \quad j = 1, \dots, n.
 \end{aligned} \tag{14}$$

where  $\theta_p^{SCCR}$  is called *stochastic technical efficiency* of  $DMU_p$ . Similarly, the deterministic equivalent of model (14) can be straightforwardly formulated for the purpose of calculating the optimal value of  $\theta_p^{SCCR}$ .

**Proposition 6:**  $\theta_p^* \leq \theta_p^{SCCR}$ .

**Proof:** Let us assume that  $(\lambda_j^{SCCR}, \theta_p)$  is a feasible solution. It is easy to show that  $(\lambda_j^{SCCR}, \theta_p)$  is also a feasible solution to the model (9), and consequently  $\theta_p^* \leq \theta_p^{SCCR}$ . ■

We define *stochastic cost efficiency* as the product of *stochastic technical efficiency* and *stochastic allocative efficiency*, viz.,  $\theta_p^* = \theta_p^{SCCR} \times \theta_p^{SAE}$ .

## 5. Extension to a slack-based method

Our cost efficiency method in Section 4 does not consider slacks in the output side. Hence, this section extends our method into a slacks-based measure (SBM) as one of the referees suggested. Let us adapt Tone (2000)'s SBM model with stochastic input and output data as presented below: as follows:



$$\begin{aligned}
& \min \quad \frac{1 - \frac{1}{m} \sum_{i=1}^m \frac{s_i^-}{\tilde{x}_{ip}}}{1 + \frac{1}{s} \sum_{r=1}^s \frac{s_r^+}{\tilde{y}_{rp}}} \\
& s.t. \\
& \sum_{j=1}^n \tilde{x}_{ij} \lambda_j = \tilde{x}_{ip} - s_i^-, \quad i = 1, \dots, m, \\
& \sum_{j=1}^J \tilde{y}_{rj} \lambda_j = \tilde{y}_{rp} + s_r^+, \quad r = 1, \dots, s, \\
& \sum_{j=1}^n \lambda_j = 1, \\
& \lambda_j \geq 0, \quad j = 1, \dots, n; \quad s_i^- \geq 0, \quad i = 1, \dots, m; \quad s_r^+ \geq 0, \quad r = 1, \dots, s.
\end{aligned} \tag{15}$$

It is well-known that the fractional program (15) can be solved via the Charnes and Cooper (1962) transformation as follows:

$$\begin{aligned}
& \min \quad \tau = t - \frac{1}{m} \sum_{i=1}^m \frac{\hat{s}_i^-}{\tilde{x}_{ip}}, \\
& s.t. \\
& \sum_{j=1}^n \tilde{x}_{ij} \bar{\lambda}_j = t \tilde{x}_{ip} - \hat{s}_i^-, \quad i = 1, \dots, m, \\
& 1 = t + \frac{1}{s} \sum_{r=1}^s \frac{\hat{s}_r^+}{\tilde{y}_{rp}}, \\
& t \tilde{y}_{r0} = \sum_{j=1}^n \bar{\lambda}_j \tilde{y}_{rj} - \hat{s}_r^+, \\
& \sum_{j=1}^n \bar{\lambda}_j = 1, \\
& \bar{\lambda}_j \geq 0, \quad j = 1, \dots, n; \quad \hat{s}_i^- \geq 0, \quad i = 1, \dots, m; \quad \hat{s}_r^+ \geq 0, \quad r = 1, \dots, s; \quad t > 0.
\end{aligned} \tag{16}$$

Since the stochastic SBM model (16) is more complicated than the stochastic cost and CCR models, we propose to use the expected values. The stochastic SBM model (16) consisting of the random variables can be converted into a deterministic model with the use of the expected value operator of random variables as follows:

$$\begin{aligned}
\min \quad & \tau = t - \frac{1}{m} \sum_{i=1}^m \frac{\widehat{s}_i^-}{E[\tilde{x}_{ip}]}, \\
s.t. \quad & \sum_{j=1}^n E[\tilde{x}_{ij}] \bar{\lambda}_j - E[\tilde{x}_{ip}] t + \widehat{s}_i^- = 0, \quad i = 1, \dots, m, \\
& t + \frac{1}{s} \sum_{r=1}^s \frac{\widehat{s}_r^+}{E[\tilde{y}_{rp}]} = 1, \\
& \sum_{j=1}^n \bar{\lambda}_j E[\tilde{y}_{rj}] - E[\tilde{y}_{rp}] t - \widehat{s}_r^+ = 0, \\
& \sum_{j=1}^n \bar{\lambda}_j = 1, \\
& \bar{\lambda}_j \geq 0, \quad j = 1, \dots, n; \widehat{s}_i^- \geq 0, \quad i = 1, \dots, m; \widehat{s}_r^+ \geq 0, \quad r = 1, \dots, s; t > 0.
\end{aligned} \tag{18}$$

It is worth noting that, due to Jensen's Operator Inequality, the objective function and the second constraint of the above model can be approximated as:

$$\begin{aligned}
& E\left[t - \frac{1}{m} \sum_{i=1}^m \frac{\widehat{s}_i^-}{\tilde{x}_{i0}}\right] \square t - \frac{1}{m} \sum_{i=1}^m \frac{\widehat{s}_i^-}{E[\tilde{x}_{ip}]}, \\
& E\left[t + \frac{1}{s} \sum_{r=1}^s \frac{\widehat{s}_r^+}{\tilde{y}_{rp}}\right] \square t + \frac{1}{s} \sum_{r=1}^s \frac{\widehat{s}_r^+}{E[\tilde{y}_{rp}]} = 1.
\end{aligned}$$

## 6. An empirical application: stochastic cost minimization planning for a state government in the US

In this section, we consider the cost minimization of 48 states in the US. The data for this example is constructed from the 2002 Economic Census – Manufacturing for the USA (United States Census Bureau, 2002) used in Ray et al. (2008) for cost minimization choice of production location. Analogously to Ray et al. (2008) we assume one output which is measured by the gross value of production and five inputs including 1) Production labour (L1), 2) non-production labour (L2), 3) capital (K), 4) energy (E) and 5) materials (M). We also assumed that all prices are fixed as listed in Table 1 but input and output variables follow normal distribution with known mean and standard deviation that are given in Table 2 and Table 3.

- L1 is measured by the number of hours worked. The corresponding input price is wage paid per hour to production workers (w1). However, because different economy in different states, the input prices for production labour in different states are reported in the

column w1 in Table 1. For example, for Alabama (AL), the wage paid per hour is \$15.156.

- L2 is the number of non-production employees and its corresponding wage rate (w2) is total emolument per employee (e.g. for Alabama, the hourly salary per non-production employee is \$48).
- E is constructed by deflating the expenditure on purchased fuels and electricity by a state-specific energy price (e.g. for Alabama, the average energy price is 5.07).
- M is total expenditure on materials, parts, and containers is used as a measure of the materials input quantity, its input price assumed to be fixed (unity) for every state.
- K is the average of beginning and end of the year values of gross fixed assets and its corresponding input price is measured by the sum of depreciation, rent, and (imputed) interest expenses per dollar of gross value of capital (e.g. for Alabama, the average cost of capital (price of capital) is 0.121).

-----Insert Table 1 here-----

Table 3 presents the results of the stochastic cost efficiency model for the distinct  $\alpha$ -values. The value of  $\alpha$  represents the pre-determined minimum probability that each of the constraints of (11) fails to satisfy. If  $\alpha = 0.05$ , then the five percentage of unsatisfied constraints is allowed by the decision maker. Since  $\alpha$  is a predetermined acceptable risk, it can be used as a planning purpose.

The last column for  $\alpha = 0.5$  corresponds to the standard deterministic cost efficiency (Farrell cost efficiency), while other columns represent stochastic cost efficiency at given levels of  $\alpha$ . Comparison of the Farrell cost efficiency results with chance-constrained cost efficiency results provides interesting insight. Overall the chance constrained scores for each DMU are higher than (or equal to) the deterministic counterparts. It is evident from the reported results that DMUs have a higher efficiency score under  $\alpha = 0.001$  compared with other probability levels. This is also illustrated in Figure 1.

-----Insert Table 3 and Figure 1 here-----

Under all given probability levels, Delaware (DE) is the best performer. Louisiana (LA) is stochastically efficient for  $0.001 \leq \alpha \leq 0.3$  but not Farrell cost efficient. Interestingly, Table 3 and Figure 1 show that efficiency increases when assuming lower probability, i.e.

$$\text{Eff}_{\alpha=0.001} \geq \text{Eff}_{\alpha=0.01} \geq \text{Eff}_{\alpha=0.04} \geq \text{Eff}_{\alpha=0.05} \geq \text{Eff}_{\alpha=0.1} \geq \text{Eff}_{\alpha=0.3} \geq \text{Eff}_{\alpha=0.4} \geq \text{Eff}_{\alpha=0.5}.$$

Table 3 also shows that the number of stochastically cost efficient states increases as  $\alpha$  decreases. Only DE is stochastically cost efficient for  $\alpha = 0.5$ , equivalently Farrell cost efficient, while there are four and eighteen stochastically cost efficient states for  $\alpha = 0.05$  and  $\alpha = 0.001$ , respectively. Table 4 shows the benchmark units for each state. For stochastically cost inefficient states, DE is the only benchmark or contributes the obtained efficient target for every inefficient state. For  $\alpha = 0.1$  where DE and LA are stochastically cost efficient, the efficient targets of IN, ND, TX, WA and WI are formed by DE and LA.

For an illustrative purpose, suppose the state government of AL (Alabama), which tries to develop its cost-efficient state economy by maintaining the previous year's output value, and its hypothetical economic planning department (division) is in charge. In order to develop a cost-efficient state economy for the next year, the department attempts to provide an efficient target. However, the next year's overall economic situation is not certain and hence a practical production possibility set cannot be deterministic, i.e., the inputs and outputs of all states to be used to construct empirical technology sets representing various future production possibilities should be treated as stochastic. Now consider the situation where the economic planning department decides to use the stochastic cost efficiency criterion and then take the risk of  $\alpha = 0.01$  based on Tables 1-4 and other available economic information. Then the efficient target is obtained as the convex combination of DE and LA. The computation of (11) yields the following feasible optimal solution vector

$$(\theta_p^*, v^*, u^*, \lambda_7^*, v_{16}^*) = (0.721798, 58.13526, 1.467466, 0.56954, 0.131606),$$

where the numbers 7 and 16 in the subscripts represent DE and LA, respectively. Considering the first constraint in (11) leads to the following relationship:

$$\frac{\sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right)}{\sum_{i=1}^m w_{ip} x_{ip}} - \frac{\Phi^{-1}(\alpha) v}{\sum_{i=1}^m w_{ip} x_{ip}} \leq \theta_p,$$

Plugging the optimal solution and observed values in the left-hand side, we obtain

$$\begin{aligned}
& \frac{\lambda_7^* \left( \sum_{i=1}^{48} w_{i,1} x_{i,7} \right) + \lambda_{16}^* \left( \sum_{i=1}^{48} w_{i,1} x_{i,16} \right)}{\sum_{i=1}^{48} w_{i,1} x_{i,1}} - \frac{\Phi^{-1}(0.01) v^*}{\sum_{i=1}^m w_{i,1} x_{i,1}} \\
&= \frac{0.56954 \times 2119.675 + 0.131606 \times 2580.4736}{2330.413429} - \frac{-2.3263 \times 58.13526}{2330.413429} \\
&= 0.66376424 - (-0.05803384) = \theta_p^*
\end{aligned}$$

where the number 1 in the subscript indicates AL. The efficient input target vector for  $\alpha = 0.01$  is obtained as

$$\begin{pmatrix} L_1^* \\ L_2^* \\ E^* \\ M^* \\ K^* \end{pmatrix} = \begin{pmatrix} \left( \lambda_7^* L_{1,7} + \lambda_{16}^* L_{1,16} - \frac{\Phi^{-1}(0.01) v^*}{w_{L_1,1}} \times c_{L_1} \right) \\ \left( \lambda_7^* L_{2,7} + \lambda_{16}^* L_{2,16} - \frac{\Phi^{-1}(0.01) v^*}{w_{L_2,1}} \times c_{L_2} \right) \\ \left( \lambda_7^* E_7 + \lambda_{16}^* E_{16} - \frac{\Phi^{-1}(0.01) v^*}{w_{E,1}} \times c_E \right) \\ \left( \lambda_7^* M_7 + \lambda_{16}^* M_{16} - \frac{\Phi^{-1}(0.01) v^*}{w_{M,1}} \times c_M \right) \\ \left( \lambda_7^* K_7 + \lambda_{16}^* K_{16} - \frac{\Phi^{-1}(0.01) v^*}{w_{K,1}} \times c_K \right) \end{pmatrix} = \begin{pmatrix} 54.620 \\ 11.203 \\ 59.049 \\ 10.414 \\ 8.299 \end{pmatrix}$$

where  $c_{L_1}$ ,  $c_{L_2}$ ,  $c_E$ ,  $c_M$  and  $c_K$  are observed cost shares. Note that the risk factor  $-\Phi^{-1}(0.01) v^* / w_{L_1,1}$  weighted by  $c_{L_1}$  is added to  $\lambda_7^* L_{1,7} + \lambda_{16}^* L_{1,16}$  in order to obtain the efficient target for the  $L_1$  input. Similarly, the other input targets are obtained. The total minimum cost for  $\alpha = 0.01$  is 1682.085 and the observed cost is \$2330.413 million, in which case the stochastic cost efficiency is 0.7218. In contrast, for  $\alpha = 0.5$  the total cost is \$1201.550 million with the cost efficiency of 0.5156 which is equivalent to the Farrell cost efficiency score. See Table 5. Rather than taking the risk of allowing for the 50% violation of each constraint in (11) to obtain the efficiency score of 0.5156, it is desirable to take a smaller risk of the 1% violation according to the proposed stochastic cost efficiency analysis. The stochastic cost efficiency method can aid policy makers with the economic planning decisions since we live in the world of uncertainty.

Before concluding this section, we provide an illustration based on the 2002 Economic Census that we utilise in this section. According to Table 6, Alabama represented by AL, which is used as our illustrative example, is the least efficient in terms of the stochastic SBM model with a score of 0.5197, while its target DMUs DE and LA are the most efficient with the score of one. For the stochastic decomposition analysis, we focus on  $\alpha=0.5$ , 0.3 and 0.1 because some DMUs have a score more than one for  $\alpha$  less than 0.1 (see Table 6). For all chosen  $\alpha$ -values in Table 6, DE has always the best performance and LA is the best or the second best performance among all the states with respect to stochastic AE and stochastic CCR. This indicates that the early results of stochastic CE are consistent with those of stochastic AE and stochastic CCR. An additional important finding is that stochastic AE is less than stochastic CCR for all DMUs except for the stochastic cost efficient DMUs. This signposts that the allocative inefficiency arising from the wrong mix of inputs for given input prices, is more severe than the radial technical inefficiency.

-----Insert Table 6 here-----

We emphasize that the data used in this application and the analysis does not aim to represent an in-depth or meticulous study of the problem at hand, but rather to show the applicability of our method. The key feature of the models in this study is that they enable managers to view more appropriate economic efficiency measures.

## 7. Concluding remarks and future work

Cost efficiency evaluates the ability of DMUs to produce the current outputs at minimal cost, given exogenous input prices. Analogously, revenue efficiency provides an estimate of attainable maximum revenue for a given input intensity and a set of output prices. In the conventional deterministic view of economic efficiency, the ex post evaluation is made for the case of deterministic and known prices. However, as many production and process planning decision are made in anticipation of unknown and stochastic information, the evaluation introduces a bias with unknown properties. Naturally, a parametric stochastic approach such as SFA may be used to address the stochastic nature of the composite variables cost and revenue, but resorting to a parametric approach introduces additional strong assumptions of the entire production set and the distribution of the inefficiency. This paper focuses at the case of stable and known deterministic market prices, exposed to random influences only through

the production technology. Such applications are readily found in e.g. financial intermediation, banking, utilities, food processing and logistics.

Precisely speaking, this paper has developed economic efficiency, both cost and revenue efficiencies, when input and/or output data of the DMUs are considered to be stochastic whereas input prices are known and deterministic. The chance-constrained program proposed in this study requires the known mean and variance, along with assuming the normal distribution for the input/output data of each unit. We show that the deterministic equivalent of the stochastic model, can be converted to a quadratic problem. The key parameter in the model is the chance constraint parameter  $\alpha$ , also used in Charnes and Cooper (1959) and in the SDEA models (Land et al., 1993).

In the application on US state data, the findings show an interesting pattern when the stochastic model is applied. Whereas the deterministic model heavily penalizes the performance of certain states down to as much as 55% lower cost efficiency, the stochastic model shows broad ranges of states with comparable cost efficiency results in the ranges around 30-40% lower than best practice. Of course, the assumptions regarding the data generation process for the input prices in the application can be discussed, given regional patterns of population, unemployment and required skills. Nevertheless, we suspect that the distribution of the estimates from the stochastic model more closely mimics the true economic situation for a future decision than the deterministic frontier results. The decision maker must take into account not only the expected value for the input prices, but also their underlying variance, even in terms of the economic efficiency of the entities to be assessed.

Further work may concern the case of input and output prices that are stochastic. In addition, one may explore the determination of mean and variance and extending the model for non-standard normal distribution of data, for example, cases that data follows skewed or truncated normal distribution since in many real applications ‘sticky’ prices are primarily changing upwards. Since data in many real-world problems are relatively noisy, another future research direction would be to scrutinise the robustness of the results of the proposed model in this study in face of SFA.

## References

Amirteimoori, A., Kordrostami S., Rezaitabar A., (2006). An improvement to the cost efficiency interval: a DEA-based approach. *Applied Mathematics and Computation*, 181, 775–81.

- Bagherzadeh Valami, H., (2009). Cost efficiency with triangular fuzzy number input prices: An application of DEA, *Chaos, Solitons and Fractals*, 42, 1631–1637.
- Banker, R.D., Charnes, A., Cooper, W.W., (1984). Some method for estimating technical and scale inefficiencies in data envelopment analysis, *Management Science*, 30 (9) 1078–1092.
- Battese, G.E., Coelli, T.J., (1992). Frontier production function, technical efficiency and panel data with application to paddy farmers in India. *Journal of Productivity Analysis*, 3, 153–169.
- Bjurek, H., Hjalmarsson, L., Førsund, F.R., (1990). Deterministic parametric and nonparametric estimating of efficiency in service production: A comparison, *Journal of Econometrics*, 46, 213–227.
- Bruni, M.E., Conforti, D., Beraldi, P., Tundis E., (2009). Probabilistically constrained models for efficiency and dominance in DEA. *International Journal of Production Economics*, 117(1), 219–28.
- Camanho, A.S., Dyson, R.G., (2005). Cost efficiency measurement with price uncertainty: A DEA application to bank branch assessments. *European Journal of Operational Research*, 161, 432–446.
- Camanho, A.S., Dyson, R.G., (2008). A generalization of the Farrell cost efficiency measure applicable to non-fully competitive settings. *Omega*, 36 (2008) 147–162.
- Charnes, A., Cooper, W.W., (1959). Chance constrained programming. *Management Science*, 6, 73–79.
- Charnes, A., Cooper, W.W., (1963). Deterministic equivalents for optimizing and satisfying under chance constraints. *Operations Research*, 11, 1 18–39.
- Charnes, A., Cooper, W.W., Rhodes, E., (1978). Measuring the efficiency of decision making units, *European Journal of Operational Research*, 2 (6), 429–444.
- Coelli, T.J., (1996). Simulators and hypothesis tests for a stochastic model: a Monte Carlo analysis. *Journal of Productivity Analysis*, 6, 247–268.
- Cook, W.D., Seiford, L.M., (2009). Data envelopment analysis (DEA) - Thirty years on, *European Journal of Operational Research*, 192, 1–17.
- Cooper, W. W., Tone, K., (1997). Measures of inefficiency in data envelopment analysis and stochastic frontier estimation. *European Journal of Operational Research*, 99, 72–88.
- Cooper, W.W, Huang, Z., Li, S.X., (1996). Satisfy DEA models under chance constraints. *The Annals of Operations Research*, 66, 279–295.
- Cooper, W.W., Deng, H., Huang, Z., Li, S., (2002b). A chance constrained programming approach to congestion in stochastic data envelopment analysis. *European Journal of Operational Research*, 53, 1–10.
- Cooper, W.W., Deng, H., Huang, Z., Li, S.X., (2002a). Chance constrained programming approaches to technical efficiencies and inefficiencies in stochastic data envelopment analysis. *Journal of the Operational Research Society*, 53, 1347–1356.
- Cooper, W.W., Deng, H., Huang, Z., Li, S.X., (2004). Chance constrained programming approaches to congestion in stochastic data envelopment analysis, *European Journal of Operational Research*, 155, 487–501.
- Cooper, W.W., Huang, Z.M., Lelas, V., Li, S.X., Olesen, O.B., (1998). Chance constrained programming formulations for stochastic characterizations of efficiency and dominance in DEA. *Journal of Productivity Analysis*, 9, 530–579.
- Dempster, M.A.H., (1980). *Stochastic Programming*. Academic Press, London.
- Emrouznejad, A. and De Witte, K. (2010). COOPER-framework: A unified process for non-parametric projects. *European Journal of Operational Research*, 207(3): 1573–1586



- Emrouznejad, A., B. R. Parker, Tavares G. (2008). Evaluation of research in efficiency and productivity: A survey and analysis of the first 30 years of scholarly literature in DEA. *Socio-Economic Planning Sciences* 42(3): 151-157.
- Fang, L., Li, H., (2012). A comment on “cost efficiency in data envelopment analysis with data uncertainty”. *European Journal of Operational Research*, 220, 588–90.
- Fang, L., Li, H., (2013). Duality and efficiency computations in the cost efficiency model with price uncertainty. *Computers & Operations Research*, 40, 594–602.
- Färe, R., Grosskopf, S., Lovell, C.A.K., (1985). *The measurement of efficiency of production*. Kluwer-Nijhoff Publishing: Boston.
- Farrell, M. J., (1957). The measurement of productive efficiency. *Journal of Royal Statistical Society, Series A*, 120, 253–281.
- Huang, Z., Li, S.X., (1996). Dominance stochastic models in data envelopment analysis. *European Journal of Operational Research*, 95 (2), 390–403.
- Jahanshahloo, G. R., Soleimani-damaneh, M., Mostafaei, A., (2008). A Simplified version of the DEA cost efficiency model. *European Journal of Operational Research*, 184, 814–815.
- Kall, P., Wallace, S. W., (1994). *Stochastic Programming*. John Wiley & Sons.
- Kuosmanen, T., Post, T., (2001). Measuring economic efficiency with incomplete price information: With an application to European commercial banks, *European Journal of Operational Research*, 134, 43–58.
- Kuosmanen, T., Post, T., (2002). Nonparametric efficiency analysis under price uncertainty: A first-order stochastic dominance approach, *Journal of Productivity Analysis*, 17 (3), 183–200.
- Kuosmanen, T., Post, T., (2003). Measuring economic efficiency with incomplete price information, *European Journal of Operational Research*, 144, 454–457.
- Lahdelma, R., Hokkanen, J., Salminen, P., (1998). SMAA—stochastic multiobjective acceptability analysis, *European Journal of Operational Research*, 106, 137–143.
- Lahdelma, R., Salminen, P., (2001). SMAA-2: Stochastic multicriteria acceptability analysis for group decision making, *Operations Research*, 49 (3), 444–454.
- Lahdelma, R., Salminen, P., (2006). Stochastic multicriteria acceptability analysis using the data envelopment model. *European Journal of Operational Research* 170, 241–252.
- Land, K. C., Lovell, C. A. K., Thore, S. (1993). Chance constrained data envelopment analysis. *Managerial and Decision Economics*, 14, 541–554.
- Liu, B. (1997). Dependent-chance programming: A class of stochastic programming. *Computers & Mathematics with Applications*, 34(12), 89–104.
- Liu, B. (1999). *Uncertain programming*. New York: John Wiley & Sons.
- Liu, J.S., Lu, L.Y.Y., Lu, W.M., Lin, B.J.Y. (2013) A survey of DEA applications, *OMEGA* 41(5): 893–902.
- Luenberger, D.G. (1995). *Microeconomic Theory*. McGraw-Hill, Boston.
- Morita, H., Seiford, L. M. (1999). Characteristics on stochastic DEA efficiency. *Journal of Operational Research Society of Japan*, 42(4), 389–404.
- Mostafaei, A., Saljooghi, F.H., (2010). Cost efficiency measures in data envelopment analysis with data uncertainty. *European Journal of Operational Research*, 202, 595–603.
- Olesen, O.B., (2006). Comparing and combining two approaches for chance constrained DEA. *Journal of Productivity Analysis*, 26 (2), 103–119.

- Olesen, O.B., Petersen, N.C., (1995). Chance constrained efficiency evaluation. *Management Science*, 41, 442–457.
- Olesen, O.B., Petersen, N.C., (2015). Stochastic Data Envelopment Analysis - A review. *European Journal of Operational Research*, In Press.
- Ray S. C., Chen L., Mukherjee L. (2008). Input price variation across locations and a generalized measure of cost efficiency, *International Journal of Production Economics*, 116 (2), 208–218
- Schafnits, C., Rosen, D., Paradi, J.C., (1997). Best practice analysis of bank branches: An application of DEA in a large Canadian bank. *European Journal of Operational Research*, 98, 269–289.
- Simon, H. A. (1957). *Models of man*. New York: John Wiley and Sons, Inc.
- Sueyoshi T., (2000). Stochastic DEA for restructure strategy: an application to a Japanese petroleum company. *Omega*, 28(4), 385–98.
- Talluri, S., Narasimhana, R., Nairb, A., (2006). Vendor performance with supply risk: a chance-constrained DEA approach. *International Journal of Production Economics*, 100, 212–222.
- Tavana, M., Khanjani Shiraz, R., Hatami-Marbini, A., Agrell, P., Paryab, K., (2012). Fuzzy stochastic data envelopment analysis with application to Base Realignment and Closure (BRAC). *Expert Systems with Applications*, 39 (15), 12247-12259.
- Tavana, M., Khanjani Shiraz, R., Hatami-Marbini, A., Agrell, P., Paryab, K. (2013). Chance-constrained DEA models with random fuzzy inputs and outputs. *Knowledge-Based Systems*, 52, 32–52.
- Tavana, M., Khanjani Shiraz, R., Hatami-Marbini, A. (2014). A new chance-constrained DEA model with birandom input and output data. *Journal of the operational research society*, 65, 1824–1839.
- Thompson, R.G., Dharmapala, P.S., Humphrey, D.B., Taylor, W.M., Thrall, R.M., (1996). Computing DEA/AR efficiency and profit ratio measures with an illustrative bank application. *Annals of Operations Research*, 68, 303–327.
- Tsionas, E.G., Papadakis, E.N., (2010). A Bayesian approach to statistical inference in stochastic DEA. *Omega*, 38, 309–314.
- Udhayakumar, A., Charles, V., Kumar, M., (2011). Stochastic simulation based genetic algorithm for chance constrained data envelopment analysis problems, *Omega*, 39, 387–397.
- Wu, D., Lee, C.G., (2010). Stochastic DEA with ordinal data applied to a multi-attribute pricing problem. *European Journal of Operational Research*, 207, 1679–1688.

## Appendix A.

---

$$\min \sum_{i=1}^m v_i x_{ip}$$

s.t.

$$\sum_{r=1}^s u_r y_{rp} = 1,$$

$$\sum_{i=1}^m v_i x_{ij} - \sum_{r=1}^s u_r y_{rj} \leq 0, \quad j = 1, \dots, n,$$

$$u_k = \frac{\rho_{kp}}{\rho_{gp}} u_g, \quad k < g, \quad k, g = 1, \dots, s, \quad (I)$$

$$u_r, v_i \geq \varepsilon, \quad r = 1, \dots, s; i = 1, \dots, m.$$

From the constraint (I) of the above model we have  $\frac{\rho_{1p}}{\rho_{2p}} = \frac{u_1}{u_2}, \frac{\rho_{2p}}{\rho_{3p}} = \frac{u_2}{u_3}, \dots, \frac{\rho_{s-1p}}{\rho_{sp}} = \frac{u_{s-1}}{u_s}$  and it

leads to  $\rho_{1p} = \rho_{2p} \frac{u_1}{u_2}, \rho_{2p} = \rho_{3p} \frac{u_2}{u_3}, \dots, \rho_{s-2p} = \rho_{s-1p} \frac{u_{s-2}}{u_{s-1}}, \rho_{s-1p} = \rho_{sp} \frac{u_{s-1}}{u_s}$ .

Thus,  $\prod_{k=1}^{s-1} \rho_{kp} = \rho_{2p} \times \dots \times \rho_{s-1p} \times \rho_{sp} \times \left( \frac{u_1}{u_2} \times \frac{u_2}{u_3} \times \dots \times \frac{u_{s-2}}{u_{s-1}} \times \frac{u_{s-1}}{u_s} \right) \Rightarrow u_s = u_1 \frac{\rho_{sp}}{\rho_{1p}}$ . Therefore,

$\sum_{r=1}^s u_r y_{rp} = \frac{u_1}{\rho_{1p}} \sum_{r=1}^s \rho_{1p} y_{rp}$ . Accordingly, the following equation is obtained using  $\sum_{r=1}^s u_r y_{rp} = 1$

from the first constraint:

$$1 = \frac{u_1}{\rho_{1p}} \sum_{r=1}^s \rho_{1p} y_{rp} \Rightarrow u_1 = \frac{\rho_{1p}}{\sum_{r=1}^s \rho_{1p} y_{rp}} \quad (\text{equivalently, } u_r = \frac{\rho_{rp}}{\sum_{r=1}^s \rho_{1p} y_{rp}})$$

Consequently, the following model is formulated using the above substitutions.

$$\min \sum_{i=1}^m v_i x_{ip}$$

s.t.

$$\sum_{i=1}^m v_i x_{ij} \leq \frac{\sum_{r=1}^s \rho_{rp} y_{rj}}{\sum_{r=1}^s \rho_{rp} y_{rp}}, \quad j = 1, \dots, n,$$

$$v_i \geq 0, \quad i = 1, \dots, m.$$

The dual of the above model is:

$$\max \sum_{j=1}^n \lambda_j \frac{\sum_{r=1}^s \rho_{rp} y_{rj}}{\sum_{r=1}^s \rho_{rp} y_{rp}}$$

*s.t.* :

$$\sum_{j=1}^n \lambda_j x_{ij} \leq x_{ip}, \quad i = 1, \dots, m$$

$$\lambda_j \geq 0, \quad j = 1, \dots, n.$$

By using a variable transformation  $\varphi = \sum_{j=1}^n \lambda_j \frac{\sum_{r=1}^s \rho_{rp} y_{rj}}{\sum_{r=1}^s \rho_{rp} y_{rp}}$  the following model can be

formulated:

$$\max \varphi$$

*s.t.* :

$$\sum_{j=1}^n \lambda_j \sum_{r=1}^s \rho_{rp} y_{rj} \geq \varphi \sum_{r=1}^s \rho_{rp} y_{rp}, \quad r = 1, \dots, s,$$

$$\sum_{j=1}^n \lambda_j x_{ij} \leq x_{ip}, \quad i = 1, \dots, m,$$

$$\lambda_j \geq 0, \quad j = 1, \dots, n.$$

## Appendix B.

### Proof of Theorem 1:

Let us assume

$$\tilde{h} = \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) - \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip}$$

Since  $\tilde{x}_{ij}$  is distributed normally,  $\tilde{h}$  also has a random variable with normal distribution with  $h$  mean and  $Var(\tilde{h}) = \sigma_h^2$  variance:  $\tilde{h} \sim N(h, \sigma_h^2)$ . The expected value and variance of  $\tilde{h}$  are obtained as

$$h = E(\tilde{h}) = E \left( \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) - \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right) = \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \theta_p \sum_{i=1}^m w_{ip} x_{ip}$$

where  $x_{ij} = E(\tilde{x}_{ij})$  and  $x_{ip} = E(\tilde{x}_{ip})$ . To derive equations for  $Var(\tilde{h})$  note that

$$\begin{aligned} Var(\tilde{h}) &= Var \left[ \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) - \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right] = \underbrace{Var \left( \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) \right)}_{(i)} + \underbrace{Var \left( \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right)}_{(ii)} \\ &\quad - \underbrace{2Cov \left( \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right), \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right)}_{(iii)} \end{aligned}$$

where

$$(i) \quad Var \left( \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right) \right) = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k Cov \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij}, \sum_{i=1}^m w_{ip} \tilde{x}_{ik} \right) = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ik})$$

$$(ii) \quad Var \left( \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right) = \theta_p^2 \sum_{i=1}^m w_{ip}^2 Var(\tilde{x}_{ip})$$

$$(iii) \quad -2Cov \left( \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} \tilde{x}_{ij} \right), \theta_p \sum_{i=1}^m w_{ip} \tilde{x}_{ip} \right) = -2\theta_p \sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ip})$$

Therefore,  $Var(\tilde{h})$  is:

$$Var(\tilde{h}) = \sigma_h^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta_p^2 \sum_{i=1}^m w_{ip}^2 Var(\tilde{x}_{ip}) - 2\theta_p \sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ip})$$

We then subtract the expected value of  $\tilde{h}$  from both sides of the right-hand of the first inequality constraint in model (9) and divide through by the standard deviation of  $\tilde{h}$  to result in

$$Pr\left(\frac{\tilde{h} - E(\tilde{h})}{\sigma_{\tilde{h}}} \leq \frac{-E(\tilde{h})}{\sigma_{\tilde{h}}}\right) \geq 1 - \alpha$$

Note that  $z = (\tilde{h} - E(\tilde{h})) / \sigma_{\tilde{h}}$  has a standard normal distribution with zero mean and unit variance. Therefore, we can write

$$Pr\left(z \leq \frac{-\sum_{j=1}^n \lambda_j \left(\sum_{i=1}^m w_{ip} x_{ij}\right) + \theta_p \sum_{i=1}^m w_{ip} x_{ip}}{\sigma_{\tilde{h}}}\right) \geq 1 - \alpha \Rightarrow \Phi\left(\frac{-\sum_{j=1}^n \lambda_j \left(\sum_{i=1}^m w_{ip} x_{ij}\right) + \theta_p \sum_{i=1}^m w_{ip} x_{ip}}{\sigma_{\tilde{h}}}\right) \geq 1 - \alpha$$

where  $\Phi$  represents the normal cumulative distribution function. We utilize the fractal function of a standard normal distribution,  $\Phi^{-1}(\alpha)$ , to obtain

$$\sum_{j=1}^n \lambda_j \left(\sum_{i=1}^m w_{ip} x_{ij}\right) - \sigma_{\tilde{h}} \Phi^{-1}(\alpha) \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip}$$

Similarly, the second inequality stochastic constraint of model (9) can be ultimately transformed into

$$\tilde{h}_r^O - \Phi^{-1}(\alpha) \sigma_{\tilde{h}_r}^O \geq 0$$

where

$$\tilde{h}_r^O = \sum_{j=1}^n \tilde{y}_{rj} \lambda_j - \tilde{y}_{rp},$$

$$\sigma_{\tilde{h}_r}^O = \sqrt{Var(\tilde{h}_r^O)} = \sqrt{\sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k Cov(\tilde{y}_{rj}, \tilde{y}_{rk}) + Var(\tilde{y}_{rp}) - \sum_{j=1}^n \lambda_j Cov(\tilde{y}_{rj}, \tilde{y}_{rp})}$$

Thus, the deterministic equivalent model (9) can be represented as

$$\min \theta_p$$

s.t.

$$\sum_{j=1}^n \lambda_j \left(\sum_{i=1}^m w_{ip} x_{ij}\right) - \sigma_{\tilde{h}} \Phi^{-1}(\alpha) \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip},$$

(A)

$$\sum_{j=1}^n \lambda_j y_{rj} + \sigma_{\tilde{h}_r}^O \Phi^{-1}(\alpha) \geq y_{rp}, \quad r = 1, \dots, s,$$

$$\lambda_j \geq 0, \quad j = 1, \dots, n.$$

Due to  $\sigma_{\tilde{h}}$  and  $\sigma_{\tilde{h}_r}^O$  the above model is obviously a nonlinear programming model. By substitution of the quadratic equality constraint  $v = \sigma_{\tilde{h}}$  and  $u_r = \sigma_{\tilde{h}_r}^O$  in model (A), the following quadratic programming problem is formulated:

$$\begin{aligned}
& \min \theta_p \\
& s.t. \\
& \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha) v \leq \theta_p \sum_{i=1}^m w_{ip} x_{ip} \\
& \sum_{j=1}^n \lambda_j y_{rj} + \Phi^{-1}(\alpha) u_r \geq y_{rp}, \quad r = 1, \dots, s \\
& v^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) + \theta^2 \sum_{i=1}^m w_{ip}^2 \text{Var}(\tilde{x}_{ip}) - 2\theta \sum_{j=1}^n \lambda_j \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ip}), \\
& u_r^2 = \sum_{j=1}^n \sum_{k=1}^n \lambda_j \lambda_k \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) + \text{Var}(\tilde{y}_{rp}) - 2 \sum_{j=1}^n \lambda_j \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rp}), \\
& \lambda_j, v \geq 0, \quad j = 1, \dots, n, \\
& u_r \geq 0, \quad r = 1, \dots, s.
\end{aligned}$$

The proof is complete. ■

## Appendix C.

### Proof of proposition 3:

Suppose that  $\theta_k^* < 1$ ,  $\forall k \in \{1, \dots, n\}$ . **Let** us assume that  $\lambda_j^*$  and  $\theta_k^*$  are a feasible solution of model (10). We prove  $\lambda_k^* < \theta_k^*$  by contradiction.

Consider  $\lambda_j^*$  and  $\theta_k^*$  in the first constraint of model (10) as

$$\sum_{\substack{j=1 \\ j \neq k}}^n \lambda_j^* \left( \sum_{i=1}^m w_{ik} x_{ij} \right) - \Phi^{-1}(\alpha) v \leq (\theta_k^* - \lambda_k^*) \left( \sum_{i=1}^m w_{ik} x_{ik} \right)$$

Let us assume that  $\lambda_k^* \geq \theta_k^*$  that implies  $(\theta_k^* - \lambda_k^*) \left( \sum_{i=1}^m w_{ik} x_{ik} \right) \leq 0$ , hence, the right hand-side

of the above equation is non-positive this contradicts since  $-\Phi^{-1}(\alpha)$  is positive for  $\alpha \leq 0.5$ .

Therefore, we conclude that  $\lambda_k^* < \theta_k^* < 1$ .

To complete the proof, we discuss the feasibility of model (10). To see this, let's divide both sides of the first, second and third constraints of model (10) by  $1 - \lambda_k^*$  as

$$\text{Const.1.} \quad \sum_{\substack{j=1 \\ j \neq k}}^n \frac{\lambda_j^*}{1 - \lambda_k^*} \left( \sum_{i=1}^m w_{ik} x_{ij} \right) - \frac{v^*}{1 - \lambda_k^*} \Phi^{-1}(\alpha) \leq \frac{(\theta_k^* - \lambda_k^*)}{1 - \lambda_k^*} \left( \sum_{i=1}^m w_{ik} x_{ik} \right),$$

$$\text{Const.2.} \quad \sum_{\substack{j=1 \\ j \neq k}}^n \frac{\lambda_j^*}{1 - \lambda_k^*} y_{rj} + \frac{u_r^*}{1 - \lambda_k^*} \Phi^{-1}(\alpha) \geq y_{rk}, \quad r = 1, \dots, s.$$

$$\begin{aligned} \text{Const.3.} \quad & \left( \frac{v^*}{1 - \lambda_k^*} \right)^2 = \sum_{j=1}^n \sum_{k=1}^n \frac{\lambda_j^*}{1 - \lambda_k^*} \frac{\lambda_k^*}{1 - \lambda_k^*} \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ik}) + \left( \frac{\theta_p}{1 - \lambda_k^*} \right)^2 \sum_{i=1}^m w_{ip}^2 \text{Var}(\tilde{x}_{ip}) \\ & - 2 \frac{\theta_p^*}{1 - \lambda_k^*} \sum_{j=1}^n \frac{\lambda_j^*}{1 - \lambda_k^*} \sum_{i=1}^m w_{ip}^2 \text{Cov}(\tilde{x}_{ij}, \tilde{x}_{ip}). \end{aligned}$$

$$\text{Const.4.} \quad \left( \frac{u_r^*}{1 - \lambda_k^*} \right)^2 = \sum_{j=1}^n \sum_{k=1}^n \frac{\lambda_j^*}{1 - \lambda_k^*} \frac{\lambda_k^*}{1 - \lambda_k^*} \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rk}) + \text{Var}(\tilde{y}_{rp}) - 2 \sum_{j=1}^n \frac{\lambda_j^*}{1 - \lambda_k^*} \text{Cov}(\tilde{y}_{rj}, \tilde{y}_{rp})$$

Equivalently, we have

$$\text{Const.1.} \quad \sum_{j=1}^n \bar{\lambda}_j \left( \sum_{i=1}^m w_{ik} x_{ij} \right) \leq \bar{\theta}_k \left( \sum_{i=1}^m w_{ik} x_{ik} \right),$$

$$\text{Const.2.} \quad \sum_{j=1}^n \bar{\lambda}_j y_{rj} \geq y_{rk}, \quad r = 1, \dots, s.$$



$$Const. 3. \quad \bar{v}^2 = \sum_{j=1}^n \sum_{k=1}^n \bar{\lambda}_j \bar{\lambda}_k \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ik}) + \bar{\theta}_p^2 \sum_{i=1}^m w_{ip}^2 Var(\tilde{x}_{ip}) - 2\bar{\theta}_p \sum_{j=1}^n \bar{\lambda}_j \sum_{i=1}^m w_{ip}^2 Cov(\tilde{x}_{ij}, \tilde{x}_{ip}),$$

$$Const. 4. \quad (\bar{u}_r^*)^2 = \sum_{j=1}^n \sum_{k=1}^n \bar{\lambda}_j \bar{\lambda}_k Cov(\tilde{y}_{rj}, \tilde{y}_{rk}) + Var(\tilde{y}_{rp}) - 2 \sum_{j=1}^n \bar{\lambda}_j Cov(\tilde{y}_{rj}, \tilde{y}_{rp})$$

where  $\bar{\theta}_k = \frac{(\theta_k^* - \lambda_k^*)}{1 - \lambda_k^*}$ ,  $\bar{\lambda}_j = \frac{\lambda_j^*}{1 - \lambda_k^*}$ ,  $j \neq k$ ,  $\bar{v} = \frac{v^*}{1 - \lambda_k^*}$ ,  $\bar{u}_r = \frac{u_r^*}{1 - \lambda_k^*}$  and  $\bar{\lambda}_k = 0$  are the feasible

solution of model (10). This shows  $\bar{\theta} < \theta_k^*$  and which contradicts with the optimality of minimizing model (10),  $\theta_k^*$ . It deduces that there exists at least one efficient DMU (i.e.,

$\theta_k^* = 1$ ). ■

## Appendix D.

### Proof of proposition 4:

Let  $\bar{\lambda}_j$ ,  $\bar{v}$  and  $\bar{\theta}_p$  are a feasible solution of model (10) under probability level  $\alpha$ . Obviously  $\bar{\lambda}_j$ ,  $\bar{v}$  and  $\bar{\theta}_p$  are still the feasible solution of model (10) at probability level  $\alpha^*$  such that  $\alpha > \alpha^*$ . Owing to  $\alpha > \alpha^* \Rightarrow -\Phi^{-1}(\alpha^*) > -\Phi^{-1}(\alpha)$ , we have

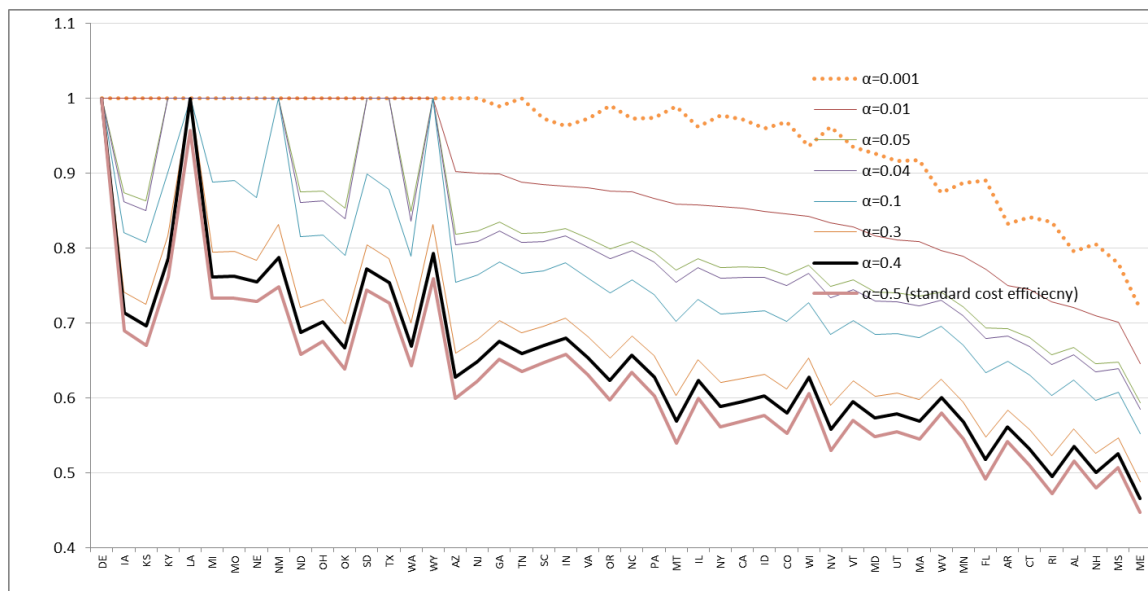
$$\begin{aligned} \sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha) v &\leq \theta_p \sum_{i=1}^m w_{ip} x_{ip} \\ \sum_{j=1}^n \bar{\lambda}_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha) \bar{v} &\leq \sum_{j=1}^n \bar{\lambda}_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \Phi^{-1}(\alpha^*) \bar{v} \leq \bar{\theta}_p \sum_{i=1}^m w_{ip} x_{ip} \\ \sum_{j=1}^n \bar{\lambda}_j y_{rj} + \Phi^{-1}(\alpha) \bar{u}_r &\geq \sum_{j=1}^n \bar{\lambda}_j y_{rj} + \Phi^{-1}(\alpha^*) \bar{u}_r \geq y_{rp} \end{aligned}$$

This shows that  $\bar{\lambda}_j$ ,  $\bar{v}$  and  $\bar{\theta}_p$  are the feasible solution at  $\alpha^*$ . Then we conclude that the efficiency at  $\alpha^*$  is greater or equal to efficiency at  $\alpha$ . ■

## Appendix E.

### Proof of proposition 5:

First, model (10) is a convex programming problem since  $\sum_{j=1}^n \lambda_j y_{rj} - y_{rp} - \sigma_{\tilde{h}_r}^O \Phi^{-1}(\alpha)$  and  $\sum_{j=1}^n \lambda_j \left( \sum_{i=1}^m w_{ip} x_{ij} \right) - \sigma_{\tilde{h}} \Phi^{-1}(\alpha) - \theta \sum_{i=1}^m w_{ip} x_{ip}$  are the convex functions where  $\sigma_{\tilde{h}}$  and  $\sigma_{\tilde{h}_r}^O$  are the convex functions and  $\Phi^{-1}(\alpha) \leq 0$  for  $\alpha \leq 0.5$ . We therefore conclude that model (10) has the global optimal solutions. The proof is complete. ■



**Figure 1:** Stochastic efficiency of US firms increase when probability reduce

**Table 1:** Manufacturing random average output, average inputs, and input prices for individual states in the US

State	Output* \$ million	Inputs*					Inputs Prices				
		L1	L2	E	M	K	w1	w2	PE	PM	PK
AL	13.087	85.72	12.46	82.7	6.648	7.289	15.156	48.509	5.07	1	0.121
AZ	8.442	41.76	12.81	11.25	2.842	3.592	15.633	61.057	8.47	1	0.147
AR	14.71	104.99	12.72	55.5	7.001	6.832	13.555	43.837	5.94	1	0.14
CA	7.805	41.34	12.1	13.26	3.155	2.951	15.66	59.732	8.68	1	0.158
CO	6.449	35.22	10.07	10.44	2.604	3.175	17.248	57.09	6.79	1	0.153
CT	8.29	46.61	16.53	10.43	2.564	3.624	18.038	60.108	10.4	1	0.135
DE	23.087	72.71	15.7	47.5	14.049	10.057	18.293	56.644	7.3	1	0.127
FL	5.164	31.73	8.68	12.33	2.064	2.228	14.423	54.015	6.51	1	0.146
GA	14.249	78.82	12.14	44.72	6.743	5.872	14.78	51.099	5.21	1	0.126
ID	8.371	50.31	7.96	34.8	3.797	5.63	15.748	52.347	6.47	1	0.151
IL	11.135	60.92	13.68	25.07	4.964	4.906	16.55	55.667	6.96	1	0.137
IN	17.443	92.7	15.1	54.25	8.141	7.798	18.058	53.257	5.34	1	0.125
IA	17.06	84.74	15.73	44.14	7.575	6.708	15.847	50.422	6.7	1	0.129
KS	15.697	80.21	15.82	33.97	8.298	5.204	16.118	53.369	6.6	1	0.132
KY	20.659	94.13	14.39	67.2	11.543	8.124	17	52.309	6.27	1	0.127
LA	25.469	62.57	11.69	206.65	15.402	16.322	18.954	54.601	4.95	1	0.112
ME	7.343	51.11	9.69	51.94	2.929	5.436	17.911	49.792	5.08	1	0.119
MD	9.058	46.05	14.02	27.24	3.537	4.08	17.537	57.893	7.34	1	0.135
MA	8.735	47.09	15.64	11.27	3.103	3.254	16.851	61.662	10.48	1	0.161
MI	14.59	71.56	12.28	25.45	7.593	5.896	20.764	56.791	7.09	1	0.136
MN	9.886	56.36	14.55	22.1	4.432	4.026	16.541	54.706	6.61	1	0.143
MS	13.651	99.43	13.17	63.57	6.988	7.289	13.545	46.478	6.1	1	0.11
MO	12.877	63.95	11.15	21.65	6.635	4.729	17.561	54.261	7.31	1	0.133
MT	4.045	21.53	3.92	32.78	2.366	2.271	15.846	45.036	5.07	1	0.126
NE	15.519	83.08	11.92	28.71	8.806	5.159	13.976	46.421	6.77	1	0.14
NV	4.819	31.7	8	8.14	1.883	2.139	15.543	54.822	13.61	1	0.164
NH	6.908	48.02	13.36	9.22	2.647	3.22	15.936	58.449	11.95	1	0.154
NJ	8.998	43.82	12.38	16.1	3.578	3.606	17.241	58.436	8.32	1	0.142
NM	6.45	29.03	5.85	13.81	2.386	3.236	16.432	52.156	7.18	1	0.166
NY	6.996	39.63	10.11	13.66	2.456	2.778	16.24	55.539	7.6	1	0.137
NC	14.577	86.19	14.04	31.82	5.743	5.727	14.068	50.299	6.84	1	0.128
ND	9.343	45.3	8.77	38.23	5.279	3.994	14.439	44.624	3.81	1	0.135
OH	13.888	72.73	13.08	30.44	6.747	5.718	18.237	52.873	7.68	1	0.133
OK	9.939	53.33	10.12	25.59	5.181	4.044	16.01	47.058	6.79	1	0.122
OR	8.228	45.48	9.59	22.61	3.003	3.716	16.797	54	7.35	1	0.147
PA	10.881	60.45	12.48	28.33	4.595	4.692	16.236	53.646	7.08	1	0.127
RI	5.063	39.98	9.13	6.34	1.898	2.134	14.457	54.63	11.72	1	0.143
SC	18.155	98.97	15.65	61.72	8.297	11.377	15.565	53.528	6.11	1	0.121
SD	11.539	61.06	9.11	16.93	5.409	2.956	13.16	41.818	6.44	1	0.136
TN	15.759	88.79	14.52	44.46	7.699	7.621	15.595	51.748	5.97	1	0.126
TX	14.551	54.91	12.47	80.36	7.87	7.5	16.321	55.635	5.17	1	0.118
UT	8.188	46.54	11.76	16.97	3.751	3.416	15.152	51.259	6.24	1	0.164
VT	8.051	48.34	13.09	9.59	2.966	5.076	16.496	53.245	12.58	1	0.17
VA	14.197	78.21	13.91	40.59	5.387	6.417	15.784	52.787	5.68	1	0.121
WA	10.315	44.32	12.93	35.02	5.097	4.537	18.079	52.591	5.54	1	0.133
WV	12.804	68.63	11.4	86.68	6.329	9.032	16.696	52.806	5.21	1	0.12
WI	12.556	71.6	13.97	28.74	5.562	5.211	16.979	52.578	6.72	1	0.136
WY	7.3	24.75	4.17	58.81	4.026	4.494	18.269	47.707	5.52	1	0.12

\* Source: Ray *et al.* (2008)

**Table 2:** Standard deviation of outputs and inputs for individual states in the US

State	Output* \$ million	Inputs*				
		L1	L2	E	M	K
AL	1.22	2.63	0.92	2.28	0.67	0.69
AZ	1.02	1.74	0.89	0.9	0.43	0.49
AR	1.25	2.73	0.98	2.1	0.71	0.75
CA	0.91	1.72	0.92	0.92	0.49	0.43
CO	0.79	1.57	0.9	0.82	0.44	0.49
CT	0.88	1.73	1.08	0.89	0.45	0.53
DE	1.39	2.25	1.13	1.79	1.08	0.9
FL	0.69	1.53	0.8	0.96	0.38	0.38
GA	1.15	2.32	0.92	1.69	0.71	0.64
ID	0.93	1.78	0.81	1.7	0.51	0.63
IL	1.06	2.23	1	1.39	0.62	0.61
IN	1.26	2.64	1.03	2	0.72	0.77
IA	1.31	2.33	0.99	1.73	0.74	0.66
KS	1.35	2.5	1.02	1.57	0.77	0.63
KY	1.46	2.79	1.04	2.09	0.9	0.71
LA	1.47	2.01	0.95	3.59	1.04	1.14
ME	0.8	1.95	0.83	2.04	0.46	0.61
MD	1.04	1.85	1.08	1.42	0.51	0.54
MA	1.01	1.83	1.01	0.89	0.46	0.45
MI	1.33	2.36	0.93	1.29	0.77	0.62
MN	1.04	1.95	0.96	1.2	0.54	0.51
MS	1.13	2.85	0.98	2.16	0.69	0.7
MO	1.2	2.1	0.84	1.2	0.65	0.58
MT	0.6	1.3	0.51	1.6	0.39	0.38
NE	1.14	2.35	0.92	1.38	0.8	0.64
NV	0.66	1.56	0.8	0.82	0.37	0.4
NH	0.77	1.85	1	0.76	0.43	0.47
NJ	0.87	1.67	0.94	1.08	0.51	0.55
NM	0.84	1.45	0.67	1.05	0.43	0.49
NY	0.89	1.79	0.81	1.06	0.43	0.45
NC	1.14	2.58	0.95	1.43	0.68	0.63
ND	1.03	1.89	0.79	1.78	0.64	0.5
OH	1.28	2.36	1	1.47	0.66	0.65
OK	1.11	1.98	0.81	1.44	0.64	0.54
OR	0.9	1.7	0.8	1.24	0.44	0.52
PA	1.11	2.06	0.9	1.36	0.57	0.56
RI	0.68	1.61	0.79	0.65	0.37	0.4
SC	1.34	2.81	1.1	1.99	0.81	0.94
SD	0.99	1.97	0.81	1.11	0.62	0.48
TN	1.39	2.51	1.01	1.88	0.76	0.74
TX	1.29	2.04	0.93	2.33	0.73	0.71
UT	0.85	1.94	0.9	1.14	0.53	0.49
VT	0.84	1.85	0.92	0.84	0.46	0.58
VA	1.2	2.51	1.04	1.69	0.66	0.68
WA	1.03	1.92	0.93	1.71	0.65	0.59
WV	1.08	2.2	0.85	2.67	0.65	0.81
WI	1.03	2.15	1.03	1.46	0.68	0.6
WY	0.82	1.27	0.55	2.04	0.54	0.59

**Table 3:** The stochastic cost efficiency of US firms

State	$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.04$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$
AL	0.7957	0.7218	0.6676	0.6575	0.6237	0.5578	0.5357	0.5156
AZ	1.0000	0.9039	0.8189	0.8041	0.7547	0.6596	0.6279	0.5992
AR	0.8329	0.7504	0.6929	0.6828	0.6492	0.5838	0.5618	0.5418
CA	0.9726	0.8549	0.7748	0.7609	0.7145	0.6252	0.5955	0.5687
CO	0.9692	0.8474	0.7647	0.7503	0.7025	0.6110	0.5807	0.5533
CT	0.8421	0.7460	0.6803	0.6689	0.6307	0.5570	0.5324	0.5102
DE	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
FL	0.8900	0.7728	0.6938	0.6800	0.6345	0.5473	0.5184	0.4924
GA	0.9899	0.9005	0.8350	0.8229	0.7819	0.7026	0.6760	0.6519
ID	0.9596	0.8496	0.7739	0.7606	0.7165	0.6313	0.6028	0.5770
IL	0.9626	0.8585	0.7865	0.7739	0.7320	0.6510	0.6239	0.5994
IN	0.9637	0.8840	0.8266	0.8162	0.7803	0.7057	0.6806	0.6579
IA	1.0000	0.9451	0.8746	0.8622	0.8210	0.7409	0.7140	0.6897
KS	1.0000	0.9375	0.8637	0.8508	0.8078	0.7242	0.6962	0.6708
KY	1.0000	1.0000	0.9589	0.9456	0.9015	0.8159	0.7871	0.7611
LA	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9873	0.9573
ME	0.7192	0.6461	0.5938	0.5844	0.5520	0.4879	0.4665	0.4470
MD	0.9267	0.8169	0.7421	0.7290	0.6855	0.6016	0.5736	0.5483
MA	0.9173	0.8096	0.7360	0.7231	0.6803	0.5975	0.5699	0.5448
MI	1.0000	1.0000	0.9512	0.9366	0.8880	0.7938	0.7622	0.7336
MN	0.8873	0.7902	0.7222	0.7103	0.6707	0.5939	0.5681	0.5448
MS	0.7791	0.7020	0.6482	0.6388	0.6074	0.5463	0.5257	0.5072
MO	1.0000	1.0000	0.9555	0.9406	0.8911	0.7951	0.7630	0.7338
MT	0.9891	0.8605	0.7706	0.7549	0.7029	0.6030	0.5698	0.5398
NE	1.0000	0.9992	0.9246	0.9115	0.8679	0.7834	0.7551	0.7295
NV	0.9627	0.8353	0.7493	0.7343	0.6848	0.5900	0.5586	0.5303
NH	0.8054	0.7107	0.6462	0.6349	0.5975	0.5254	0.5013	0.4796
NJ	1.0000	0.9007	0.8230	0.8094	0.7643	0.6774	0.6484	0.6223
NM	1.0000	1.0000	1.0000	1.0000	0.9609	0.8309	0.7878	0.7490
NY	0.9776	0.8564	0.7740	0.7596	0.7118	0.6197	0.5889	0.5612
NC	0.9730	0.8757	0.8089	0.7972	0.7582	0.6826	0.6573	0.6343
ND	1.0000	0.9531	0.8754	0.8616	0.8152	0.7205	0.6878	0.6582
OH	1.0000	0.9537	0.8767	0.8632	0.8182	0.7310	0.7018	0.6753
OK	1.0000	0.9353	0.8533	0.8390	0.7911	0.6983	0.6672	0.6390
OR	0.9909	0.8772	0.7995	0.7859	0.7407	0.6533	0.6240	0.5976
PA	0.9747	0.8678	0.7945	0.7816	0.7388	0.6557	0.6279	0.6027
RI	0.8353	0.7294	0.6574	0.6448	0.6031	0.5227	0.4960	0.4718
SC	0.9731	0.8860	0.8207	0.8091	0.7705	0.6956	0.6705	0.6478
SD	1.0000	1.0000	0.9627	0.9480	0.8990	0.8045	0.7728	0.7442
TN	1.0000	0.8894	0.8197	0.8074	0.7664	0.6868	0.6601	0.6359
TX	1.0000	1.0000	0.9388	0.9254	0.8792	0.7859	0.7546	0.7263
UT	0.9169	0.8124	0.7405	0.7279	0.6862	0.6059	0.5790	0.5548
VT	0.9347	0.8294	0.7575	0.7449	0.7030	0.6221	0.5950	0.5706
VA	0.9730	0.8811	0.8132	0.8009	0.7601	0.6810	0.6545	0.6306
WA	1.0000	0.9248	0.8495	0.8358	0.7895	0.6999	0.6699	0.6429
WV	0.8745	0.7971	0.7412	0.7311	0.6958	0.6248	0.6010	0.5795
WI	0.9364	0.8436	0.7777	0.7662	0.7277	0.6533	0.6284	0.6058
WY	1.0000	1.0000	1.0000	1.0000	0.9446	0.8315	0.7937	0.7595

**Table 4: Benchmarks**

State	$\alpha=0.001$	$\alpha=0.01$	$\alpha=0.04$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.3$	$\alpha=0.4$	$\alpha=0.5$
AL	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE	DE
AZ	AZ	DE	DE	DE	DE	DE	DE	DE
AR	DE, LA	DE	DE	DE	DE	DE	DE	DE
CA	DE	DE	DE	DE	DE	DE	DE	DE
CO	DE, LA	DE	DE	DE	DE	DE	DE	DE
CT	DE	DE	DE	DE	DE	DE	DE	DE
DE	DE	DE	DE	DE	DE	DE	DE	DE
FL	DE	DE	DE	DE	DE	DE	DE	DE
GA	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE	DE
ID	DE, LA	DE	DE	DE	DE	DE	DE	DE
IL	DE, LA	DE	DE	DE	DE	DE	DE	DE
IN	DE, LA	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE
IA	IA	DE	DE	DE	DE	DE	DE	DE
KS	KS	DE	DE	DE	DE	DE	DE	DE
KY	KY	KY	DE	DE	DE	DE	DE	DE
LA	LA	LA	LA	LA	LA	LA	DE	DE
ME	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE	DE
MD	DE, LA	DE	DE	DE	DE	DE	DE	DE
MA	DE	DE	DE	DE	DE	DE	DE	DE
MI	MI	MI	DE	DE	DE	DE	DE	DE
MN	DE, LA	DE	DE	DE	DE	DE	DE	DE
MS	DE, LA	DE	DE	DE	DE	DE	DE	DE
MO	MO	MO	DE	DE	DE	DE	DE	DE
MT	DE, LA	DE, LA	DE	DE	DE	DE	DE	DE
NE	NE	DE	DE	DE	DE	DE	DE	DE
NV	DE	DE	DE	DE	DE	DE	DE	DE
NH	DE	DE	DE	DE	DE	DE	DE	DE
NJ	NJ	DE	DE	DE	DE	DE	DE	DE
NM	NM	NM	NM	NM	DE	DE	DE	DE
NY	DE	DE	DE	DE	DE	DE	DE	DE
NC	DE, LA	DE	DE	DE	DE	DE	DE	DE
ND	ND	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE
OH	OH	DE	DE	DE	DE	DE	DE	DE
OK	OK	DE	DE	DE	DE	DE	DE	DE
OR	DE	DE	DE	DE	DE	DE	DE	DE
PA	DE, LA	DE	DE	DE	DE	DE	DE	DE
RI	DE	DE	DE	DE	DE	DE	DE	DE
SC	DE, LA	DE, LA	DE	DE	DE	DE	DE	DE
SD	SD	SD	DE	DE	DE	DE	DE	DE
TN	TN	DE, LA	DE	DE	DE	DE	DE	DE
TX	TX	TX	DE, LA	DE, LA	DE, LA	DE	DE	DE
UT	DE, LA	DE	DE	DE	DE	DE	DE	DE
VT	DE	DE	DE	DE	DE	DE	DE	DE
VA	DE, LA	DE, LA	DE	DE	DE	DE	DE	DE
WA	WA	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE
WV	DE, LA	DE, LA	DE, LA	DE, LA	DE, LA	DE	DE	DE
WI	DE, LA	DE	DE	DE	DE	DE	DE	DE
WY	WY	WY	WY	WY	DE	DE	DE	DE
#of efficient states	18	9	4	4	2	2	1	1



**Table 5: Input Targets for AL**

	$L_1$	$L_2$	$E$	$M$	$K$	total cost	cost efficiency
$\alpha=0.01$	54.620	11.203	59.049	10.414	8.299	1682.085	0.7218
$\alpha=0.5$	41.216	8.900	26.926	7.964	5.701	1201.550	0.5156
observed	85.72	12.46	82.7	6.648	7.289	2330.4134	

**Table 6: Decomposition Analysis for selected  $\alpha$  -values and stochastic SBM**

	$\alpha=0.5$			$\alpha=0.3$			$\alpha=0.1$			
State	SCE	SAE	SCCR	SCE	SAE	SCCR	SCE	SAE	SCCR	SSBM
AL	0.5156	0.6084	0.8474	0.5578	0.5877	0.9492	0.6237	0.6237	1	0.5197
AZ	0.5992	0.5992	1	0.6596	0.6596	1	0.7547	0.7547	1	1
AR	0.5418	0.5859	0.9248	0.5838	0.5838	1	0.6492	0.6492	1	0.5786
CA	0.5687	0.5940	0.9574	0.6252	0.6252	1	0.7145	0.7145	1	0.8548
CO	0.5533	0.6226	0.8887	0.611	0.6110	1	0.7025	0.7025	1	0.8991
CT	0.5102	0.5102	1	0.557	0.5570	1	0.6307	0.6307	1	1
DE	1	1.0000	1	1	1.0000	1	1	1.0000	1	1
FL	0.4924	0.5507	0.8941	0.5473	0.5554	0.9854	0.6345	0.6345	1	1
GA	0.6519	0.6934	0.9402	0.7026	0.7026	1	0.7819	0.7819	1	0.6382
ID	0.577	0.6498	0.8879	0.6313	0.6313	1	0.7165	0.7165	1	0.637
IL	0.5994	0.6730	0.8907	0.651	0.6568	0.9912	0.732	0.7320	1	0.6776
IN	0.6579	0.7039	0.9347	0.7057	0.7057	1	0.7803	0.7803	1	0.634
IA	0.6897	0.7141	0.9658	0.7409	0.7409	1	0.821	0.8210	1	0.6747
KS	0.6708	0.7192	0.9327	0.7242	0.7242	1	0.8078	0.8078	1	0.6757
KY	0.7611	0.7726	0.9851	0.8159	0.8159	1	0.9015	0.9015	1	0.7937
LA	0.9573	0.9573	1	1	1.0000	1	1	1.0000	1	1
ME	0.447	0.5022	0.89	0.4879	0.4936	0.9884	0.552	0.5520	1	0.6024
MD	0.5483	0.5839	0.939	0.6016	0.6016	1	0.6855	0.6855	1	0.7274
MA	0.5448	0.5448	1	0.5975	0.5975	1	0.6803	0.6803	1	0.9324
MI	0.7336	0.7722	0.95	0.7938	0.7938	1	0.888	0.8880	1	0.7065
MN	0.5448	0.6171	0.8828	0.5939	0.6039	0.9834	0.6707	0.6707	1	0.687
MS	0.5072	0.6028	0.8414	0.5463	0.5845	0.9347	0.6074	0.6074	1	0.5236
MO	0.7338	0.7697	0.9534	0.7951	0.7951	1	0.8911	0.8911	1	1
MT	0.5398	0.6787	0.7954	0.603	0.6883	0.8761	0.7029	0.7029	1	1
NE	0.7295	0.7504	0.9721	0.7834	0.7834	1	0.8679	0.8679	1	0.7169
NV	0.5303	0.5895	0.8995	0.59	0.5923	0.9962	0.6848	0.6848	1	1
NH	0.4796	0.5098	0.9407	0.5254	0.5254	1	0.5975	0.5975	1	0.8496
NJ	0.6223	0.6308	0.9865	0.6774	0.6774	1	0.7643	0.7643	1	0.8312
NM	0.749	0.7490	1	0.8309	0.8309	1	0.9609	0.9609	1	1
NY	0.5612	0.5612	1	0.6197	0.6197	1	0.7118	0.7118	1	1
NC	0.6343	0.6343	1	0.6826	0.6826	1	0.7582	0.7582	1	0.7104
ND	0.6582	0.7446	0.884	0.7205	0.7307	0.9861	0.8152	0.8152	1	0.6692
OH	0.6753	0.7438	0.9079	0.731	0.7310	1	0.8182	0.8182	1	0.6609
OK	0.639	0.7280	0.8778	0.6983	0.7099	0.9837	0.7911	0.7911	1	0.6686

OR	0.5976	0.6021	0.9925	0.6533	0.6533	1	0.7407	0.7407	1	0.8298
PA	0.6027	0.6531	0.9229	0.6557	0.6557	1	0.7388	0.7388	1	0.6949
RI	0.4718	0.4718	1	0.5227	0.5227	1	0.6031	0.6031	1	1
SC	0.6478	0.6924	0.9356	0.6956	0.6956	1	0.7705	0.7705	1	0.6063
SD	0.7442	0.7442	1	0.8045	0.8045	1	0.899	0.8990	1	1
TN	0.6359	0.7186	0.8849	0.6868	0.6898	0.9956	0.7664	0.7664	1	0.6096
TX	0.7263	0.7648	0.9497	0.7859	0.7859	1	0.8792	0.8792	1	0.6458
UT	0.5548	0.6348	0.874	0.6059	0.6253	0.969	0.6862	0.6862	1	0.7332
VT	0.5706	0.5706	1	0.6221	0.6221	1	0.703	0.7030	1	1
VA	0.6306	0.6335	0.9954	0.681	0.6810	1	0.7601	0.7601	1	0.6886
WA	0.6429	0.6717	0.9571	0.6999	0.6999	1	0.7895	0.7895	1	0.6871
WV	0.5795	0.6588	0.8796	0.6248	0.6374	0.9802	0.6958	0.6958	1	0.5398
WI	0.6058	0.6673	0.9078	0.6533	0.6541	0.9988	0.7277	0.7277	1	0.6687
WY	0.7595	0.7645	0.9934	0.8315	0.8315	1	0.9446	0.9446	1	1