Predicting the efficiency of a surface coal mine for competitiveness

Mussa Daniel Budeba[a*], Johan Wilhelm Joubert[b], Ronny Webber-Youngman[a], and Shahriar Shafiee[c]

[a]Department of Mining Engineering, University of Pretoria
[b]Department of Industrial and Systems Engineering, University of Pretoria
[c]Senior Business Analyst, TAFE

17 March 2016

a∗ Coresponding author. mussaco@gmail.com

Abstract

Mining is a competitive business with many players. The survival of a mine in the business is determined by its efficiency and cost-effectiveness relative to the other producers. Both new and operating mines should select optimal technical variables, such as the production rate, that will make them competitive, taking into account mine unique project variables.

This paper describes a model for estimating the technical efficiency of surface mine for Coal Supply to Local and Export (CSLE). The application of the model and evaluation is shown using simulated data. It proposes a predictive model of the efficiency of a new project.

Keywords: DEA; surface coal mining competitiveness; technical efficiency

1 Introduction

Mine planning involves determining parameters such as production rate, and estimating capital and operating costs that will maximize Net Present Value (NPV). Mine plans should include optimal parameters that will make the mine efficient and cost-effective in a competitive business environment. An efficient mine, in this context, is one which uses optimum resources to generate maximum outputs at an effective cost, given both discretionary and non-discretionary operational conditions and unique deposit characteristics.

After fuel oil, coal is the second major source of primary energy in the world. It is expected to be the leading source of energy by 2030 and will be the only fossil fuel remaining after 2042 [34]. A major factor contributing to this is the continuously increasing demand for coal outside the Organisation for Economical Cooperation and Development (OECD), mainly driven by intensive industrial development in China and India. This increasing demand will drive the future supply of coal from both new operations and existing mines.

On the supply side, new and existing mine projects face challenges that can result in uncertainty in both their production performance and benefits. Various authors have highlighted some of the challenges among them, including a high stripping ratio, low cost to mine seam being located in remote areas with infrastructure problems, complex metallurgy, topographical challenges, inclement weather, and dip of the coal seam [33, 35]. The other factors include geology, variable coal quality, economics and legislation [6, 12, 13, 16, 32].

Despite the highlighted challenges, some mines are more efficient than others. These can be termed as best-practice mines and generate value on the investment while others are inefficient and ineffective. The best-practice mines become competitive while inefficient ones are less competitive, cannot deliver the expected benefits and face high risk of failing to survive in the business.

The question arises how to estimate the envelope of the best-practice mines and predict the technical efficiency, which will help a new mine to position itself competitively, given mine specific characteristics of the planned project and the mine supply market structure. The mine under evaluation are termed as Decision Making Units (DMUs). The DMU refers to any entity such as a firm or organisation for profit or non-profit making which is considered for measuring its efficiency relative to other units producing similar products or services. The envelope can be defined as that shape which is made by the piecewise lines connecting all the DMUs that are efficient with a score of 1. All DMUs lying on the envelope are termed as best - practice operations.

Planning a new mine should consider its competitiveness relative to operating mines. This can be achieved if the planned mine can predict its technical efficiency within a set of technical parameters relative to other existing producers. The plan can be optimized by iterative re-selection of the technical variables to determine those that will enhance its competitiveness.

Highlighting a number of cases, Lumley [24] states that forecasts in mine plans have not been met and returns on investment are lower than predicted. That is, 80-90% of projects will exceed their budgeted costs and will not deliver the benefits to stakeholders. More often the planned production rate has not been achieved due to technical deficiencies in the planning process, the planners' optimism and strategic misrepresentation. In addition, planning has often been based on ideal production rates, which are given by the Original Equipment Manufacturer (OEM). These production rates do not necessarily make the mine competitive. On the other hand, Bullock [5] showed some examples from mining projects studied between 1965 to 2002, indicating cost overruns, the lowest overrun being 22% and the highest 35%. In another study of 60 projects, 58% had overruns of between 15% and 100% of the capital cost.

More often, new mines plan to be the lowest cost producers in the industry. In this case cost estimation is done by averaging the cost of the mines with similar characteristics and adjusting the estimates to suit the operating conditions. In other cases, the mine evaluators tend to take a conservative approach, overestimating the costs to mitigate uncertainties [35].

This article describes the formulation of the model for estimating the efficiency of surface mine for Coal Supply to Local and Export - CSLE. The model is formulated using Data Envelopment Analysis (DEA) technique. DEA applies linear programming to determine the efficiency of a unit relative to other units, using similar inputs to generate similar outputs. The inputs and outputs quantities can vary from one unit to the other. Simulated data in this study are used to illustrate the application of CSLE model and to formulate a predictive model for the technical efficiency of a new mine. The mine could use the models to determine its efficiency and competitiveness relative to other operating mines, selecting the optimal parameters, such as production rate, for planning purposes. Mines would be able to identify and improve any stage of a supply chain which is causing inefficiency.

This article is organised as follows:

- Section 2 discusses the literature review considering mine competitiveness and the determination of the technical efficiency using the DEA method;
- Section 3 explains the formulation of CSLE model for the evaluation of the technical efficiency of a surface coal mine;
- Section 4 describes the simulation process of data and illustrates the evaluation of the CSLE model;
- Section 5 formulates a predictive model for the efficiency of a new surface coal mine; and
- Section 6 presents the conclusion.

2 Literature review

Mine planning starts once an economical block model has been made available. The major objective is to determine which tonnages, grades and associated costs will maximize NPV. In other cases, planning aims to maximize the rate of return (ROR) or achieve the lowest payback period for a given project. The production rate is always the first parameter to be specified and is used in selecting equipment and in determining the capital expenditures of the project. Long [22] describes the estimation of the production rate of a new mine, taking into account the expected tonnes to be mined. The production rate is represented by Equation 1 [22].

$$\text{Tonnes per day} = 0.014(\text{Expected tonnes})^{0.75} \tag{1}$$

Leinart and Schumacher [21], on the other hand, maintain that in mine planning, multiple scenarios should be used to estimate the production rate, determining the full cost estimate of each production rate and selecting the one that will generate maximum NPV. The authors suggest that before estimation, a conceptual mine plan should first be developed.

2.1 Competitiveness in the mines

The competitiveness of a mine is traditionally measured by its position on the cost curve. For example, Rudenno [30] suggests that investing in projects or resource companies at a given volatility of commodity prices calls for the consideration of that project's position on the cost curve. The cost curve refers to a plot of cash costs (US$/tonne) on the vertical axis against the cumulative production rate (Mt) on the horizontal axis of the mines involved in production. To illustrate the use of the cost curve, consider Figure 2 to represent a cost curve of the operating thermal coal mines. Assuming a free on board (FOB) price of US$100/tonne, mines A and B are regarded as competitive and mines C and D are considered to be high risk because their production costs are higher than the market price of the commodity.

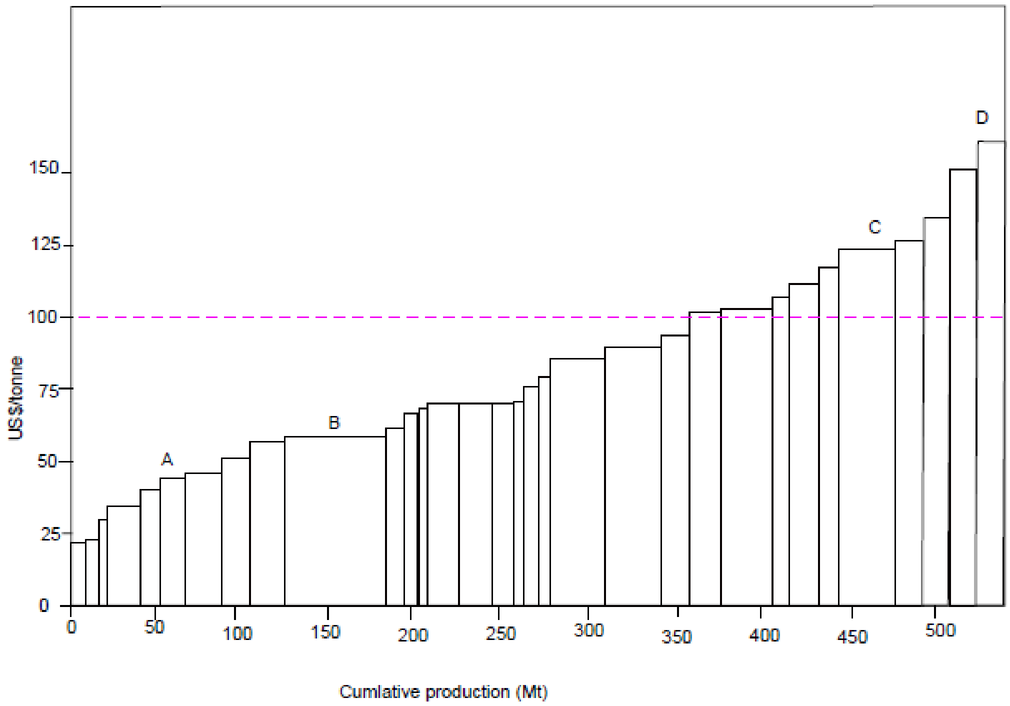

Figure 1: Example of mine cost curve (source: Authors' construction)

On the other hand, a challenge remains on how best a new mine positions itself competitively on the cost curve. This has not been fully investigated. The current practice involves the company comparing its production rates and associated costs with those of lower cost producers on the mine cost curve. According to [26], new mines are faced with the problem of combining technical design and economic parameters to generate value for the stakeholders. Planning to be a low-cost producer, however, has always been uncertain since each project has unique challenges and production characteristics which can critically affect efficiency and cost effectiveness. In addition, being a low-cost producing mine does not necessarily make a mine efficient. Besides cost curves, measuring the competitiveness of mines using relative technical efficiency can reflect how a mine is using the inputs to generate the target outputs at an effective cost.

The competitiveness measure of mines using the relative technical efficiency approach for a new mine is essential for three reasons. Firstly, as soon as the mine goes into operation, it will be exposed to competition. It will therefore need to compare itself to other mines in the market and determine whether it is on the envelope of best-practice, choosing appropriate operating parameters that will enable it to be competitive and hence survive in the business. Secondly, the mine should identify any stage of its supply chain which could cause overall inefficiency once it has commenced production, and take steps to minimize this. Thirdly, the mine should identify the best-practice mines on the envelope and use them to benchmark its own operation.

2.2 Determination of relative technical efficiency of a DMU

Measuring technical efficiency involves determining the levels of inputs that will be transformed by a unit such as a firm or organization to generate outputs. An operation is deemed efficient if it uses the minimum inputs to produce target output. A surface coal mine is no exception to this rule. It uses capital, labour and other technical parameters of a given project to generate a certain quantity of coal of a specific quality.

There are two major approaches to the measurement of efficiency, namely parametric and non-parametric. The parametric method involves determining efficiency on the basis of a linear function which is assumed to represent the plane of the efficient units. When a unit deviates from the plane, it is considered inefficient. The non-parametric method includes Data Envelopment Analysis (DEA), which uses a linear programming technique to define the envelope of the best-practice units. The differences between the parametric and non-parametric methods are illustrated by Figure 2 [18]. In Figure 2, the parametric

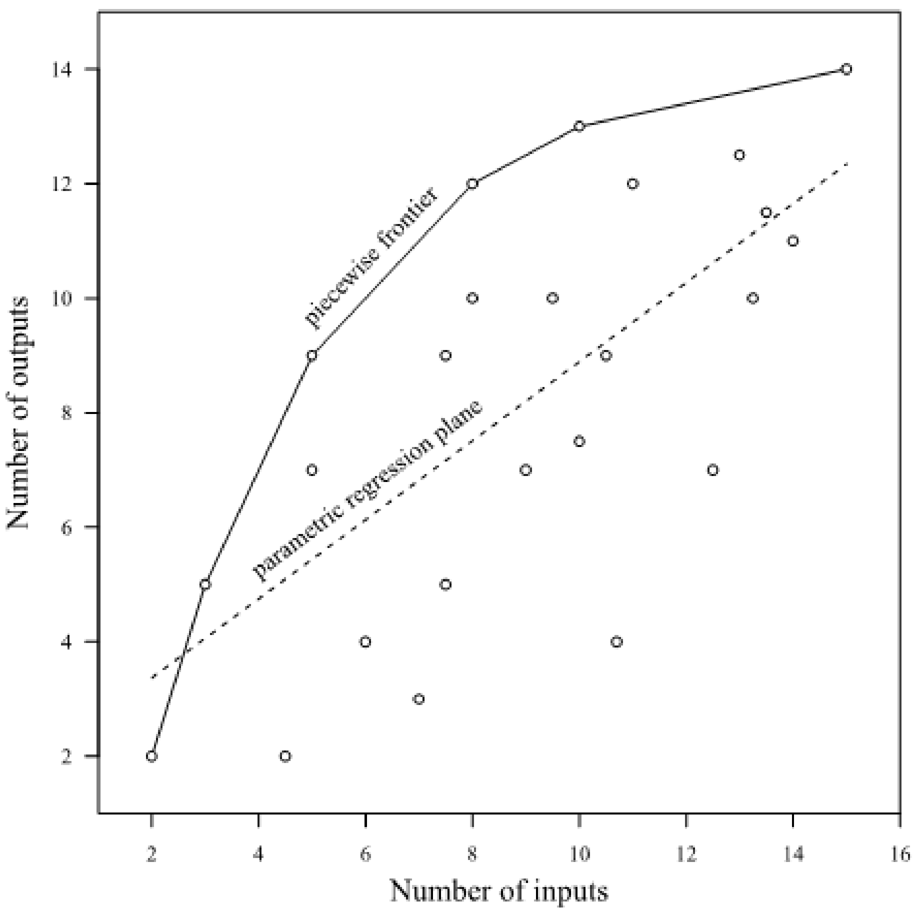

Figure 2: Comparison of DEA and regression (Source: Charnes et al. (1994))

model tries to fit a plane (dotted line) on the data points, while the DEA constructs the piecewise envelopment of best-practice using a linear programming technique. DEA is preferred to parametric method due to factors including;

- it does not require assuming the function;
- it can be used for measuring efficiency using multiple inputs and outputs; and
- it can be used even when there is insufficient data.

DEA was introduced by Charnes and Cooper in 1978 [8]. It has been used in different fields, for example in agriculture and economics. In mining, its application includes evaluating the technical efficiency of coal mines, the growth in productivity for both opencast and underground mines, and in assessing the efficiency of coal mine safety measures [20, 36, 38].

2.2.1 Mathematical representation of the basic DEA models

The DEA can be input-oriented, which refers to a DMU which uses a minimum of its inputs in order to achieve a specific level of outputs. The output-oriented, considers the DMU as using the same level of input and maximizing the outputs.

The mathematical representation of an input-oriented DEA is as follows. Assuming a set of DMUs given by $\boldsymbol{J} = \{1...n\}$, with each $j \in \boldsymbol{J}$ using m inputs to generate s outputs. Consider a set of inputs $\boldsymbol{I} = \{1...m\}$ and a set of output $\boldsymbol{R} = \{1...s\}$. Assume that the usage of input is x_{ij} and the amount of output is y_{rj}. The weights (multipliers) given to the inputs and outputs are v_i and u_r respectively. The efficiency score of a DMU under evaluation denoted by $DMU_j = o$ can be defined by Equation (2) [37]. The resulting relative technical efficiency of each DMU in fractional form is given by Equation 3-6.

$$\text{Efficiency} = \frac{\text{weighted sum of outputs}}{\text{weighted sum of inputs}} \tag{2}$$

$$\max h_o = \frac{\sum_{r \in \boldsymbol{R}} u_r y_{ro}}{\sum_{i \in \boldsymbol{I}} v_i x_{io}} \tag{3}$$

subject to

$$\tag{4}$$

$$\frac{\sum_{r \in \boldsymbol{R}} u_r y_{rj}}{\sum_{i \in \boldsymbol{I}} v_i x_{ij}} \leq 1 \qquad \forall j \in \boldsymbol{J} \tag{5}$$

$$v_i, u_r \geq 0 \qquad \forall i \in \boldsymbol{I}, \quad r \in \boldsymbol{R} \tag{6}$$

The above Fraction Programming (FP) equations can be transformed using Charnes and Cooper transformation into Linear Programming (LP) below.

$$\max g_o = \sum_{r \in \boldsymbol{R}} \mu_r y_{ro} \tag{7}$$

subject to

$$\sum_{r\in\boldsymbol{R}} \mu_r y_{rj} - \sum_{i\in\boldsymbol{I}} \nu_r x_{ij} \leq 0 \qquad \forall j \in \boldsymbol{J} \tag{8}$$

$$\sum_{i\in\boldsymbol{I}} \nu_i x_{io} = 1 \tag{9}$$

$$\nu_i, \mu_r \geq 0 \qquad \forall i \in \boldsymbol{I}, \quad r \in \boldsymbol{R} \tag{10}$$

$$\text{where} \quad t = \frac{1}{v_i x_{io}}, \quad \mu_r = t u_r \quad \text{and} \quad \nu_i = t v_i \tag{11}$$

The above Equation (7)-(10) can be written in dual form, and the efficiency score can then be solved. The dual form involves a minimum number of variables, but gives the same results as its primal linear programming counterpart. If the operation involves scale measures, the DEA can be considered in two formulations. The first of these, considers a constant return to scale (CRS) in which a DMU assumes that an increase in inputs results in a proportional increase in outputs, while the other is a variable return to scale (VRS) which assumes that the DMU increase in the inputs results in an unequal proportional increase in outputs.

3 Formulation of a Coal Supply for Local and Export (CSLE) model

The mathematical representation of the CSLE model are formulated from the generic structure of the mine system shown Figure 3. It is an extension of a model supplying coal for the export market as discussed by Budeba et al. [4]. In the formulation of the CSLE model in this research, consider a set of surface coal mine systems $\boldsymbol{J} = \{1, ..., n\}$. Each coal mine system $j \in \boldsymbol{J}$ is considered to be a DMU producing coal and supplying product to the local and export markets. The coal mine system consists of subsystems, including *mining* denoted by superscript m, *washing* represented by superscript b and *port* denoted by superscript p. These subsystems are interdependent functioning to produce coal for sales in the markets.

Let the following apply: A set of inputs to the mining-subsystem is $\boldsymbol{F} = \{1, ..., F\}$. A set of intermediate outputs from the mining-subsystem to the washing-subsystem is $\boldsymbol{K} = \{1, ..., K\}$. A set of inputs at the beginning of the washing-subsystem is $\boldsymbol{I} = \{1, ..., I\}$. A set of intermediate outputs from the washing-subsystem to the port-subsystem is $\boldsymbol{R} = \{1, ..., R\}$. A set of inputs at the beginning of the port-subsystem be $\boldsymbol{S} = \{1, ..., S\}$. A set of outputs from the port-subsystem be $\boldsymbol{T} = \{1, ..., T\}$ and a set of non-discretionary variables be $\boldsymbol{H} = \{1, ..., H\}$.

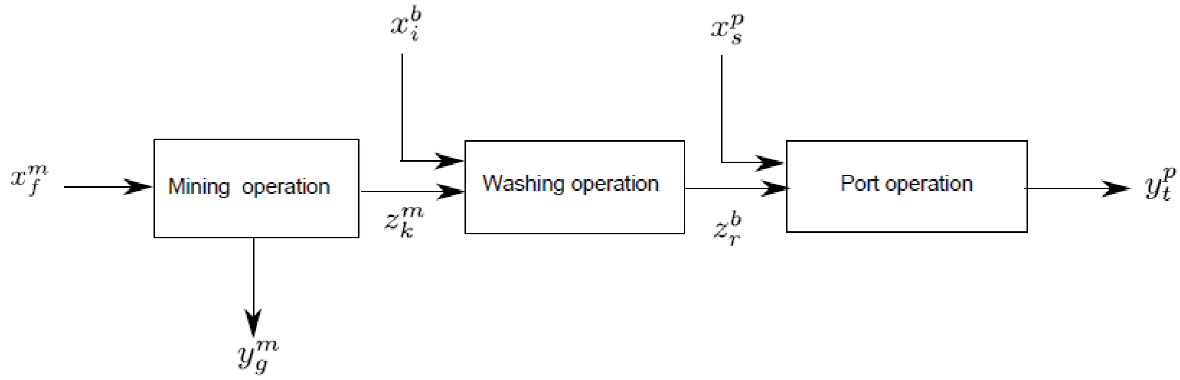

Figure 3: Generic mining to export coal supply chain structure

Consider the following mathematical notations that are used in the formulation of the models.

x_{fj}^m is the given usage of input $f \in \boldsymbol{F}$ by mining-subsystem m of a coal supply system $j \in \boldsymbol{J}$.

υ_f^m is the weight of input $f \in \boldsymbol{F}$.

y_{gj}^m is the given amount of output supplied to local market $g \in \boldsymbol{G}$ generated by mining-subsystem m of a coal mine system $j \in \boldsymbol{J}$.

ν_g^m is the weight given to output $g \in \boldsymbol{G}$.

x_{ij}^b is the given usage of input $i \in \boldsymbol{I}$ at the beginning of washing-subsystem b of a coal mine supply system $j \in \boldsymbol{J}$.

υ_i^b is the weight given to the input $i \in \boldsymbol{I}$.

z_{kj}^m is the amount of intermediate output $k \in \boldsymbol{K}$ generated by mining-subsystem m and is the usage of washing-subsystem b of a coal mine system $j \in \boldsymbol{J}$.

η_k^m is the weight given to intermediate output $k \in \boldsymbol{K}$.

x_{sj}^p is the given usage of input $s \in \boldsymbol{S}$ at the beginning of port-subsystem p of a coal mine supply system $j \in \boldsymbol{J}$.

υ_s^p is the weight given to the input $s \in \boldsymbol{S}$.

z_{rj}^b is the amount of intermediate output $r \in \boldsymbol{R}$ generated by washing-subsystem b and is the usage of port-subsystem p of a coal mine system $j \in \boldsymbol{J}$.

η_r^b is the weight given to intermediate output $r \in \boldsymbol{R}$.

y_{tj}^p is the amount of outputs $t \in \boldsymbol{T}$ generated by port-subsystem p and is the final output which is supplied to the export market.

ν_t^p is the weight given to output $t \in \boldsymbol{T}$.

x_{hj}^e is the given non-discretionary input $h \in \boldsymbol{H}$ with effects on a coal mine supply system $j \in \boldsymbol{J}$.

ω_h^e is the weight given to non-discretionary input $h \in \boldsymbol{H}$.

According to Cook et al. [7], the overall efficiency of a system involving subsystems is the weighted sum of the efficiencies of the subsystems. The weight of a subsystem is the ratio of the inputs of that system to the overall inputs of the system. Considering the assumption of VRS, the mathematical representation of the model for a mine supplying coal to both local and export markets is presented in fractional programming (FP) by Equations (12) -(20).

$$\max h_o = \frac{\sum_{g\in \boldsymbol{G}} \nu_g^m y_{go}^m + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{ko}^m + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{ro}^b + \sum_{t\in \boldsymbol{T}} \nu_t^p y_{to}^p + \pi^m + \pi^b + \pi^p}{\sum_{f\in \boldsymbol{F}} \upsilon_f^m x_{fo}^m + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{ko}^m + \sum_{i\in \boldsymbol{I}} \upsilon_i^b x_{io}^b + \sum_{s\in \boldsymbol{S}} \upsilon_s^p x_{so}^p + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{ro}^b} \tag{12}$$

subject to

$$\frac{\sum_{g\in \boldsymbol{G}} \nu_g^m y_{gj}^m + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{kj}^m + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{rj}^b + \sum_{t\in \boldsymbol{T}} \nu_t^p y_{tj}^p + \pi^m + \pi^b + \pi^p}{\sum_{f\in \boldsymbol{F}} \upsilon_f^m x_{fj}^m + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{kj}^m + \sum_{i\in \boldsymbol{I}} \upsilon_i^b x_{ij}^b + \sum_{s\in \boldsymbol{S}} \upsilon_s^p x_{sj}^p + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{rj}^b} \leq 1 \quad \forall j \in \boldsymbol{J} \tag{13}$$

$$\frac{\sum_{g\in \boldsymbol{G}} \nu_g^m x_{gj}^m + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{kj}^m + \pi^m}{\sum_{f\in \boldsymbol{F}} \upsilon_f^m x_{fj}^m} \leq 1 \quad \forall j \in \boldsymbol{J} \tag{14}$$

$$\frac{\sum_{r\in \boldsymbol{R}} \eta_r^b z_{rj}^b + \pi^b}{\sum_{i\in \boldsymbol{I}} \upsilon_i^b x_{ij}^b + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{kj}^m} \leq 1 \quad \forall j \in \boldsymbol{J} \tag{15}$$

$$\frac{\sum_{t\in \boldsymbol{T}} \nu_t^p y_{tj}^p + \pi^p}{\sum_{s\in \boldsymbol{S}} \upsilon_s^p x_{sj}^p + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{rj}^b} \leq 1 \quad \forall j \in \boldsymbol{J} \tag{16}$$

$$\sum_{f\in \boldsymbol{F}} \upsilon_f^m x_{fj}^m \geq \alpha \quad \forall j \in \boldsymbol{J} \tag{17}$$

$$\sum_{i\in \boldsymbol{I}} \upsilon_i^b x_{ij}^b + \sum_{k\in \boldsymbol{K}} \eta_k^m z_{kj}^m \geq \alpha \quad \forall j \in \boldsymbol{J} \tag{18}$$

$$\sum_{s\in \boldsymbol{S}} \upsilon_s^p x_{sj}^p + \sum_{r\in \boldsymbol{R}} \eta_r^b z_{rj}^b \geq \alpha \quad \forall j \in \boldsymbol{J} \tag{19}$$

$$\upsilon_f^m, \quad \nu_g^m, \quad \eta_k^m, \quad \upsilon_i^b, \quad \eta_r^b, \quad \upsilon_s^p, \quad \nu_t^p, \quad \geq \epsilon; \quad \pi^m, \quad \pi^b, \quad \pi^p \quad \text{are free in sign} \tag{20}$$

Taking into account the effect of non-discretionary variables e, Lotfi et al. [23] and Cooper et al. [8] suggested that non-discretionary variables have to enter the objective functions to account for their influence on the efficiency of the producing entity. Therefore, Equations (12) -(20) are transformed from FP to LP using the same form as the Charnes and Cooper transformation, and by including the influence of non-discretionary variables, with the resulting CSLE model being represented by Equations (21)-(31).

$$\max h_o = \sum_{g\in\boldsymbol{G}} \mu_g^m y_{gj}^m + \sum_{k\in\boldsymbol{K}} \gamma_k^m z_{ko}^m + \sum_{r\in\boldsymbol{R}} \gamma_r^b z_{ro}^b + \sum_{t\in\boldsymbol{T}} \mu_t^p y_{to}^p - \sum_{h\in\boldsymbol{H}} \omega_h^e x_{ho}^e + u^m + u^b + u^p \tag{21}$$

subject to

$$\sum_{g\in\boldsymbol{G}} \mu_g^m y_{gj}^m + \sum_{t\in\boldsymbol{T}} \mu_t^p y_{tj}^p - \sum_{f\in\boldsymbol{F}} \omega_f^m x_{fj}^m - \sum_{i\in\boldsymbol{I}} \omega_i^b x_{ij}^b - \sum_{s\in\boldsymbol{S}} \omega_s^p x_{sj}^p - \sum_{h\in\boldsymbol{H}} \omega_h^e x_{hj}^e + u^m + u^b + u^p \leq 0 \qquad \forall j \in \boldsymbol{J} \tag{22}$$

$$\sum_{f\in\boldsymbol{F}} \omega_f^m x_{fo}^m + \sum_{k\in\boldsymbol{K}} \gamma_k^m z_{ko}^m + \sum_{i\in\boldsymbol{I}} \omega_i^b x_{io}^b + \sum_{r\in\boldsymbol{R}} \gamma_r^b z_{ro}^b + \sum_{s\in\boldsymbol{S}} \gamma_s^p x_{so}^p = 1 \tag{23}$$

$$\sum_{g\in\boldsymbol{G}} \mu_g^m y_{gj}^m + \sum_{k\in\boldsymbol{K}} \gamma_k^m z_{kj}^m - \sum_{f\in\boldsymbol{F}} \omega_f^m x_{fj}^m + u^m \leq 0 \qquad \forall j \in \boldsymbol{J} \tag{24}$$

$$\sum_{r\in\boldsymbol{R}} \gamma_r^b z_{rj}^b - \sum_{i\in\boldsymbol{I}} \omega_i^b x_{ij}^b - \sum_{k\in\boldsymbol{K}} \gamma_k^m z_{kj}^m + u^b \leq 0 \qquad \forall j \in \boldsymbol{J} \tag{25}$$

$$\sum_{t\in\boldsymbol{T}} \mu_t^p y_{tj}^p - \sum_{s\in\boldsymbol{S}} \omega_s^p x_{sj}^p - \sum_{r\in\boldsymbol{R}} \gamma_r^b z_{rj}^b + u^p \leq 0 \qquad \forall j \in \boldsymbol{J} \tag{26}$$

$$\sum_{f\in\boldsymbol{F}} \omega_f^m x_{fj}^m \geq \beta \qquad \forall j \in \boldsymbol{J} \tag{27}$$

$$\sum_{i\in\boldsymbol{I}} \omega_i^b x_{ij}^b + \sum_{k\in\boldsymbol{K}} \gamma_k^m z_{kj}^m \geq \beta \qquad \forall j \in \boldsymbol{J} \tag{28}$$

$$\sum_{s\in\boldsymbol{S}} \omega_s^p x_{sj}^p + \sum_{r\in\boldsymbol{R}} \gamma_r^b z_{rj}^b \geq \beta \qquad \forall j \in \boldsymbol{J} \tag{29}$$

$$\omega_h^e \geq 0 \qquad \forall h \in \boldsymbol{H}, \quad \forall j \in \boldsymbol{J} \tag{30}$$

$$\omega_f^m, \quad \mu_g^m, \quad \gamma_k^m, \quad \omega_i^b, \quad \gamma_r^b, \quad \omega_s^p, \quad \mu_t^p \quad \geq \epsilon; \quad u^m, \quad u^b, \quad u^p \quad \text{are free in sign} \tag{31}$$

Where:
u is a difference of two positive numbers and accounts for VRS; ϵ is a smallest positive number with a value of $= 10^6$, it is used to ensure that there is no slacks in inputs when a DMU attains the efficiency score of one relative to other DMUs, that is a DMU of efficiency scores of one must be efficient [40].
β =0.05 is the minimum weighted input at each subsystem.

4 Data simulation and application to CSLE model

The major source of data for this study is Raw Materials Data (RMD) database subscribed in 2013 to 2014 [17]. The database consists of minerals and mining entities. Data for some mines (named DMUs) that were extracted are represented in Tables 1 and 2. Table 1 shows the data for selected major mine design and production variables while Table 2 indicates the deposit's unique characteristics for same mines. These were found to be the only mines having all values in the variables required for purpose of the study.

Table 1: Selected mine variables

DMUs	CAPEX (US$M)	SR	ROM (Mt/yr)	Cap (Mt/yr)	Age (yrs)
DMU1	631.33	2.00	11.07	15.00	8.00
DMU2	502.71	5.00	8.89	6.00	19.00
DMU3	15.88	10.30	14.80	11.00	17.00
DMU4	62.18	13.20	8.31	7.80	9.00
DMU5	2.00	7.80	0.60	4.00	2.00
DMU6	8.31	2.40	1.24	1.20	4.00
DMU7	424.99	5.20	4.00	2.40	3.00

Source: RMD, annual reports, media and company websites

Table 2: Location and deposit specific variables

DMUs	Ash (%)	Moisture(%)	Dist-port (Km)	Precipitation (mm)	CV (MJ/Kg)	Thickness (m)
DMU1	26.50	9.00	262.00	630.00	26.10	15.00
DMU2	10.10	11.00	275.00	656.00	27.80	5.50
DMU3	5.50	15.50	41.50	2809.00	25.80	8.70
DMU4	6.00	16.00	517.00	2905.00	28.90	5.00
DMU5	5.50	13.50	98.00	2121.00	27.60	3.50
DMU6	25.00	10.00	951.20	688.00	20.00	10.00
DMU7	13.30	2.90	570.00	683.00	27.80	3.00

Source: RMD, annual reports, media and company websites

Considering that all models formulated using DEA technique are sensitive to number of DMUs in relation to number of inputs and outputs. These models does not perform well when the sample size is small, they poorly discriminate the efficient from inefficient DMUs. Cooper et al. [8] explain the rule of thumb to determine the minimum number of DMUs needed for DEA studies. The number of DMUs should be the maximum of $(n * m, 3(m + n))$, where n is the number of inputs and m the number of outputs. The extracted mines for this study did not satisfy the suggested condition [8]. Simulation was proposed to generate the representative samples conditioned to the correlation among the variables of the mines from Table 1 and Table 2.

Before conducting the simulation, a bootstrap sampling technique was applied to infer the underlying distribution of each variable based on skewness and kurtosis statistical measures. Bootstrap sampling refers to the processes of drawing a representative sample from a population of data, then assuming that sample being a population from which other samples of the same size are drawn with replacement. The generated samples are known as bootstrap samples. Ankarali et al. [1] explain an example of the computation and construction of the distribution for skewness and kurtosis of bootstrap samples and more details about the bootstrap technique has been discussed in [10].

In this study, the calculated bootstrap mean values of skewness of most variables are greater or less than zero and the mean kurtosis values are less than 3. These are typical properties of non-normal data, suggesting that the variables for each mine were drawn from non-normal populations of their distributions.

Various methods are available for the simulation of multivariate non-normal data. The methods include the iterative method, also known as Sample and Iterate *Sample and Iterate* (SI) [31], the extended Fleishman's linear transformation [3, 39], the fifth order polynomial transformation [15], the generalized lambda methods [14], and the copula based method which was proposed as an alternative to the iterative method by Mair et al. [25].

In this research, the SI method was chosen due to the following reasons;

- it can be applied to any distribution in which other methods do fail;
- it does not require a specification of the moment but rather iterations to restore the correlation matrix;
- it can be used in sampling discrete distribution;
- it distinguishes populations with equal moments;
- no boundary conditions exist for defined moment; and
- it can accommodate distributions with undefined moments.

The simulation was conducted in two steps. Firstly, the determination of the correlation matrix indicated in Table 3 was determined from the samples of variables of mines collected. Secondly, the SI method was applied, whose implementation in R software is discussed in detail by Ruscio and Kaczetow [31]. The matrix of the simulated data is shown in Table 4. A comparison of the correlation matrix between the original samples and the simulated data has a root mean square residual of 0.099. This is the minimum average error that was obtained in reproducing the original matrix for all simulations conducted. It was achieved when the number of DMUs reached 60. These were used in the application and evaluation of the CSLE model and in the predictive modelling.

Table 3: Correlation of original sample variables for CSLE

	ROM	CV	Thickness	Cap	CAPEX	SR	Precipitation	Age	Ash	Moisture	Dist.port	Export	Dquantity
ROM	1												
CV	0.2669	1											
Thickness	0.4244	-0.5155	1										
Cap	0.8336	0.259	0.6532	1									
CAPEX	0.2796	0.2859	0.3024	0.3755	1								
SR	0.2804	0.5409	-0.4961	0.0987	-0.5613	1							
Precipitation	0.2983	0.3463	-0.2785	0.2169	-0.693	0.9357	1						
Age	0.7908	0.2038	0.1389	0.4576	0.1939	0.2339	0.173	1					
Ash	-0.1403	-0.6565	0.7478	0.1007	0.4365	-0.8499	-0.7468	-0.3024	1				
Moisture	0.3345	0.1066	0.017	0.3139	-0.6135	0.6941	0.8166	0.4054	-0.5245	1			
Dist.port	-0.5209	-0.5835	0.037	-0.5802	-0.1004	-0.3373	-0.4205	-0.4404	0.5397	-0.4003	1		
Export	0.9362	0.2964	0.1692	0.6411	-0.0017	0.538	0.5265	0.8141	-0.4027	0.4875	-0.4893	1	
Dquantity	0.4901	-0.0849	0.8676	0.7938	0.6633	-0.4978	-0.3724	0.1641	0.6445	-0.143	-0.2019	0.1613	1

Table 4: Correlation matrix for simulated data for CSLE

	ROM	CV	Thickness	Cap	CAPEX	SR	Precipitation	Age	Ash	Moisture	Dist.port	Export	Dquantity
ROM	1												
CV	0.3122	1											
Thickness	0.4127	-0.4191	1										
Cap	0.9096	0.2424	0.5743	1									
CAPEX	0.1293	0.1607	0.2402	0.2761	1								
SR	0.26	0.455	-0.5007	0.0276	-0.5727	1							
Precipitation	0.2869	0.2113	-0.32	0.0477	-0.6453	0.8398	1						
Age	0.6634	0.2317	0.2886	0.5916	0.0255	0.2546	0.3237	1					
Ash	-0.1621	-0.6086	0.6104	-0.0265	0.3705	-0.8073	-0.6392	-0.2151	1				
Moisture	0.4498	-0.003	0.0911	0.2787	-0.7049	0.6474	0.745	0.4238	-0.3755	1			
Dist.port	-0.6143	-0.5633	0.0732	-0.4949	-0.0946	-0.4443	-0.3099	-0.3947	0.438	-0.3157	1		
Export	0.8473	0.3923	0.2072	0.7293	-0.1177	0.5271	0.5191	0.6509	-0.4025	0.6321	-0.6223	1	
Dquantity	0.4835	-0.0645	0.7168	0.6936	0.4818	-0.3999	-0.3366	0.3038	0.3571	-0.0562	-0.1854	0.2744	1

5 Results of the computation of efficiency using CSLE model

For computation of efficiency scores, the evaluation of the mines considered that the mines are subjected to similar economic conditions and legislation. The average export price of $90 per tonne of product for three years was used [19]. The export price was adjusted for actual heat content to account for the differing calorific value contents [9]. The assumed local coal supply price was approximated to be equivalent to $53 per tonne based on energy content. The minimum calorific value specification for export coal was considered to be 24.5 MJ/kg (5,850kcal/kg NCV) [11]. The allowable maximum carbon dioxide emission was assumed to be 25000 tonnes of Co_2 equivalent per annum. The CSLE model representing the mine system in Equations (21)-(31) was solved using the simulated DMUs from Section 4. The inputs and outputs of the subsystem used in solving the model are the following:

- Inputs in the mining-subsystem were CAPEX, SR, and mining employees, and the outputs were ROM, ash, moisture, local sales of coal.
- Inputs in the washing-subsystem were ROM, ash, moisture and the specication of plant capacity, and the estimated administrative and washing plant employees. The outputs included export tonnages.
- Inputs to the port-subsystem were export tonnages and allowable carbon emission, while the output was net revenue generated.
- Non-discretionary variables include precipitation, distance to the port, seam thickness and the age of the mine.

The CSLE model was solved using the computer code (program) developed by the authors in R software [27, 28].

The overall efficiency scores of the simulated surface coal mines' supply system are presented in Figure 4. The results show that the efficiency scores range from 0.691 to 1. Those with an efficiency score of 1 represent the envelope of the best-practice, while those with an efficiency score of less than 1 are deemed to be inefficient. The best-practice operations are those mines named DMU_7 , DMU_9, DMU_{15}, DMU_{17}, DMU_{33}, DMU_{44}, and DMU_{54}, with the remaining operations being deemed inefficiency. The best-practice mines are using minimal inputs to generate their target outputs.

The least efficient mine is DMU_{16} with technical efficiency of 69.1%. This DMU has to improve by 31.9% in order to be efficient. This can be achieved by reducing the inputs in its subsystems for an aggregate of 31.9 %. The DMU_{16} is inefficient because it is using excess inputs to produce the present outputs.

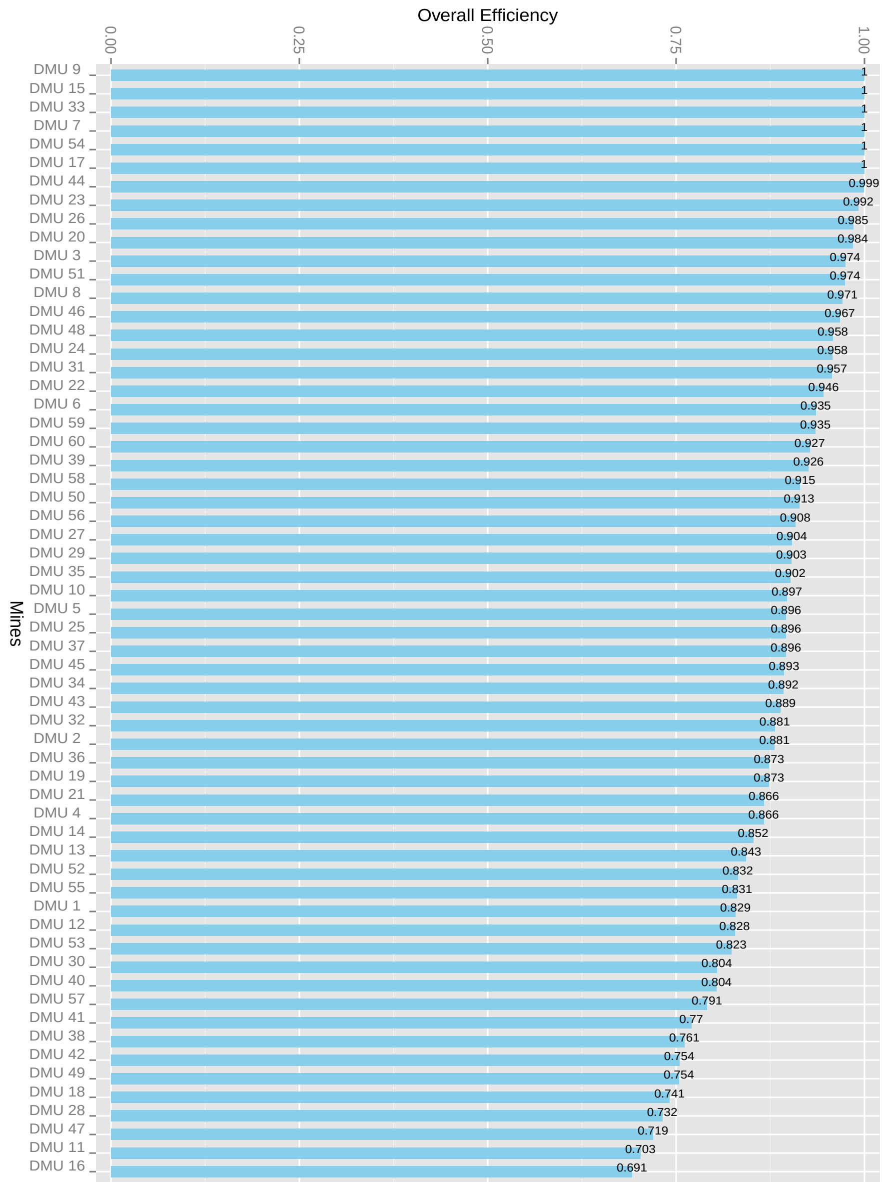

Figure 4: Overall efficiency scores for each DMU for CSLE model

5.1 Prediction of efficiency of a new surface coal mine

Methods for predicting DEA efficiency scores are addressed by Radovanović et al. [29]. The authors highlight the available methodologies, including statistical algorithms, such as linear regression and least median square regression, and machine learning algorithms, such as neural networks and support vector machines. They claim that predicting DEA efficiency scores saves re-computation of scores when a new DMU is added in the production. One could also argue that, in the absence of enough mines to be compared, the developed predictive model could help new mines to predict their technical efficiency within a given set of discretionary and non-discretionary variables.

In building the predictive function, the simulated DMUs were randomly divided into two groups, training sets (48 DMUs) and test sets (12 DMUs). Bootstrap of the training data sets is done so as to create independent efficiency scores that were obtained by relative computation among the DMUs. The training data were used to specify and estimate the model parameters using the stepwise akaike information criterion (AIC) based on bootstrap, for selecting the final model in linear regression modelling [2]. In the first place all variables are used and upon application of stepwise bootstrap the final model is specified. The test set was used to evaluate the model. The general form of the function considered is presented in Equation 32.

$$\hat{\theta}_{overall(j)} = \alpha_o + \sum_{i=1}^{n} \alpha_i x_{ij} \tag{32}$$

Where $\hat{\theta}_{overall(j)}$ is the overall efficiency of a mine supply system j and x_{ij} are the input variables consisting of both discretionary and non discretionary variables and α_i represents the coefficient of input i in the model.

The variables that were found to be useful in building the model at the 5% significant level are summarized in Table 5. These variables are used in specifying the final useful model represented by Equation 33.

Table 5: Regression Results

	Dependent variable:
	Efficiency
EmployeesM	0.0022^{***}
	(0.0006)
Thickness	0.0139^{***}
	(0.0030)
Precipitation	0.00004^{***}
	(0.00001)
Export	0.0117^{**}
	(0.0054)
EmployeesP	-0.0207^{***}
	(0.0050)
Constant	0.9899^{***}
	(0.0461)
Observations	48
R^2	0.626
Adjusted R^2	0.581
Residual Std. Error	0.057 (df = 42)
F Statistic	14.030^{***} (df = 5; 42)
Note:	*p<0.1; **p<0.05; ***p<0.01

$$\hat{\theta} = 0.9899 + 0.0022 \times (\text{EmployeesM}) + 0.0139 \times (\text{Thickness}) + 0.00004 \times (\text{Precipitation}) + 0.0117 \times (\text{Export}) - 0.0207 \times (\text{EmployeesP}) \quad (33)$$

Where:

EmployeesM: total number of employees in mining subsystem.

Thickness: average coal seam thickness (m).

Precipitation: annual precipitation (mm).

Export: annual export tonnages (Mt/yr).

EmployeesP: total number of employees in washing subsystem.

The evaluation of the performance of the predictive model was achieved by re-estimating the efficiency scores of the test sets using Equation 33. The computed efficiencies for selected test sets, as shown in Figure 4, are compared to the predicted efficiency scores using Equation 33, as indicated Table 6. The predictive model generates good estimates that are closer to the computed ones and the model root mean square error of 0.057. The model can be used by the new mine to assess the likely technical efficiency for the given predetermined parameters in the feasibility study.

Table 6: Comparison between the DEA efficiency scores of test sets and the predicted scores

Mines	Employees	Thickness (m)	Precipitation(mm)	Export(Mt/yr)	EmployeesP	Calculated Efficiency	Predicted Efficiency
DMU 2	992	8.7	2809	4.80	124	0.8806	0.8542
DMU 3	440	5.0	2809	6.00	59	0.9743	0.9698
DMU 4	832	5.5	2809	6.00	105	0.8662	0.8718
DMU 10	266	5.5	683	0.58	40	0.8971	0.8469
DMU 13	946	3.0	2809	6.99	118	0.8428	0.8260
DMU 31	61	3.5	656	0.58	13	0.9569	0.9339
DMU 34	320	10.0	683	6.00	49	0.8924	0.9034
DMU 39	170	5.0	656	0.58	30	0.9258	0.8385
DMU 42	812	5.5	683	6.00	104	0.7541	0.7657
DMU 47	1250	5.5	2905	6.99	154	0.7194	0.7762
DMU 50	345	15.0	656	6.00	53	0.9132	0.9430
DMU 51	440	5.0	2809	4.80	59	0.9742	0.9558

6 Conclusion

This study formulated a CSLE model based on a DEA technique for evaluating the relative technical efficiency of a surface mine supplying thermal coal to both local and export markets. The model can be used to evaluate a mine's competitiveness and to suggest the optimal inputs. The competitiveness of a new mine depends on its efficiency and cost effectiveness. Optimal competitiveness can be achieved by selecting and combining discretionary input variables to generate maximum output without the excessive use of inputs, taking into account the influence of non-discretionary variables in the area in which the mine project is located.

It could also be concluded that, before investing in a project, a new mine can predict its technical efficiency based on the choice of predetermined variables in a feasibility study. However, the predictive model has some limitations, for example, it can generate large errors if the mine variables are out of the range of the data used in specifying it. It is recommended that simulation studies should be done for operating mines with variables outside the range of the simulated DMUs for this study and then use the CSLE model to compute the efficiency of each DMU. In addition, one can collect more data from operating mines and compute their efficiency scores using the CSLE model.

Management can use the CSLE models to identify those mines that can be used as benchmarks for an inefficient mine. It can also help in choosing a better project for investment purposes from a given list of projects, taking into account their comparable technical efficiencies.

Acknowledgements

This study forms part of PhD research work at the university of Pretoria. We would like to acknowledge all people who provided support and the the Raw Materials Group which was the source of data subscribed to in 2013-2014. We are also grateful to all scholars for their works referenced, they made a significant contribution in this research study.

References

[1] Ankarali, H., ANKARALI, S. et al. [2009], 'A bootstrap confidence interval for skewness and kurtosis and properties of t-test in small samples from normal distribution', *Balkan Medical Journal* **2009**(4).

[2] Austin, P. C. and Tu, J. V. [2004], 'Bootstrap methods for developing predictive models', *The American Statistician* **58**(2), 131–137.

[3] Bendayan, R., Arnau, J., Blanca, M. J. and Bono, R. [2014], 'Comparación de los procedimientos de fleishman y ramberg et al. para generar datos no normales en estudios de simulación', *Anales de Psicología* **30**(1), 364–371.

[4] Budeba, M., Joubert, J. and WebberYoungman, R. [2014], 'Modelling and determining the technical efficiency of a surface coal mine supply chain', *The Southern African Institute of Mining and Metallurgy* .

[5] Bullock, R. L. [2011], 'Accuracy of feasibility study evaluations would improve accountability', *Mining Engineering* **63**(4), 78–85.

[6] Chan, M. [2008], Coal supply and cost under technological and environmental uncertainty, PhD thesis, Carnegie Mellon University. Available at `http://search.proquest.com/docview/304667434`.

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

[8] Cooper, W., Seiford, L. and Tone, K. [2007], *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software*, 2nd edn, John Wiley & Sons, p. 2.

[9] Docker, D. [2011], Coal sale and purchase agreements the buyers perspective, *in* '27th International Coal Supply Contracts & Transport Logistics'.

[10] Efron, B. and Tibshirani, R. [1994], *An Introduction to the Bootstrap*, Taylor & Francis.

[11] globalCOAL [2015], 'Scota specifications'. Available at `https://www.globalcoal.com/Brochureware/standardTradingContract/specifications/`.

[12] Gordon, R. L. [1976], 'The economics of coal supply-the state of the art', *Energy* **1**(3), 283 – 289.

[13] Haftendorn, C., Holz, F. and Hirschhausen, C. [2012], 'The end of cheap coal? a techno-economic analysis until 2030 using the coalmod-world model', *Fuel* **102**(0), 305 – 325.

[14] Headrick, T. C. and Mugdadi, A. [2006], 'On simulating multivariate non-normal distributions from the generalized lambda distribution', *Computational Statistics & Data Analysis* **50**(11), 3343–3353.

[15] Headrick, T. C. and Sawilowsky, S. S. [1999], 'Simulating correlated multivariate nonnormal distributions: Extending the fleishman power method', *Psychometrika* **64**(1), 25–35.

[16] Höök, M., Zittel, W., Schindler, J. and Aleklett, K. [2010], 'Global coal production outlooks based on a logistic model', *Fuel* **89**(11), 3546 – 3558.

[17] IntierraRMG [2014], 'Raw materials data coal'. Available at `www.intierrarmg.com`.

[18] Joubert, J. [2010], Data envelopment analysis: an overview. Working paper, 14, Optimisation Group, Industrial and Systems Engineering,University of Pretoria.

[19] Kolesnikov, I. [2015], 'Coal Prices: Long Term Forecast to 2020 Data and Charts'. Available at `http://knoema.com/xfakeuc/coal-prices-long-term-forecast-to-2020-data-and-charts`.

[20] Kulshreshtha, M. and Parikh, J. K. [2002], 'Study of efficiency and productivity growth in opencast and underground coal mining in india: a dea analysis', *Energy Economics* **24**(5), 439 – 453.

[21] Leinart, J. B. and Schumacher, O. L. [2010], The role of cost estimating in mine planning and equipment selection, *in* 'Mine Planning and Equipment Selection - MPES 2010', The Australasian Institute of Mining and Metallurgy (The AusIMM), pp. 69–80.

[22] Long, K. R. [2009], 'A test and re-estimation of taylors empirical capacity–reserve relationship', *Natural resources research* **18**(1), 57–63.

[23] Lotfi, F. H., Jahanshahloo, G. and Esmaeili, M. [2007], 'Non-discretionary factors and imprecise data in dea', *International Journal of Math Analysis* **1**(5), 237–246.

[24] Lumley, G. [2011], 'Mine planners lie with numbers'. Available at `http://www.scribd.com/doc/81005939/White-Paper-Mine-Planners-Lie-With-Numbers`.

[25] Mair, P., Satorra, A. and Bentler, P. M. [2012], 'Generating nonnormal multivariate data using copulas: Applications to sem', *Multivariate Behavioral Research* **47**(4), 547–565.

[26] Mohnot, J. K., Singh, U. K. and Dube, A. K. [2001], 'Formulation of a model for determining the optimum investment, operating cost and mine life to achieve planned profitability', *Mining Technology* **110**(2), 129–132.

[27] Pessanha, J. F. M., Marinho, A., da Costa Laurencel, L. and do Amaral, M. R. d. S. [2013], Implementing dea models in the R program, *in* '11th International Conference on Data Envelopment Analysis', Samsun, Turkey. Conference Presentation.

[28] R Core Team [2015], *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. Available at `http://www.R-project.org/`.

[29] Radovanović, S., Radojičić, M. and Savić, G. [2014], 'Two-phased dea-mla approach for predicting efficiency of nba players', *Yugoslav Journal of Operations Research ISSN: 0354-0243 EISSN: 2334-6043* **24**(3).

[30] Rudenno, V. [2009], *The Mining Valuation Handbook: Mining and Energy Valuation for Investors and Management*, 3rd edn, John Wiley & Sons, pp. 135–140.

[31] Ruscio, J. and Kaczetow, W. [2008], 'Simulating multivariate nonnormal data using an iterative algorithm', *Multivariate Behavioral Research* **43**(3), 355–381.

[32] Schneid, R. J. and Torries, T. F. [1991], 'Competitive costs of foreign and u.s. coal in north atlantic markets', *Mining Science and Technology* **13**(1), 89 – 104.

[33] Shafiee, S., Nehring, M. and Topal, E. [2009], 'Estimating average total cost of open pit coal mines in australia', *Australian Mining* **134**.

[34] Shafiee, S. and Topal, E. [2008], 'An econometrics view of worldwide fossil fuel consumption and the role of {US}', *Energy Policy* **36**(2), 775 – 786.

[35] Shafiee, S. and Topal, E. [2012], 'New approach for estimating total mining costs in surface coal mines', *Mining Technology* **121**(3), 109–116.

[36] Shu-Ming, W. [2011], 'Evaluation of safety input-output efficiency of coal mine based on dea model', *Procedia Engineering* **26**(0), 2270 – 2277.

[37] Talluri, S. [2000], 'Data envelopment analysis: models and extensions', *Decision Line* **31**(3), 8–11.

[38] Tong, L. and jia Ding, R. [2008], 'Efficiency assessment of coal mine safety input by data envelopment analysis', *Journal of China University of Mining and Technology* **18**(1), 88 – 92.

[39] Vale, C. D. and Maurelli, V. A. [1983], 'Simulating multivariate nonnormal distributions', *Psychometrika* **48**(3), 465–471.

[40] Zhu, J. [2008], *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*, Springer.