RAIRO-Oper. Res. 51 (2017) 903–920
DOI: 10.1051/ro/2016063

RAIRO Operations Research
www.rairo-ro.org

ESTIMATING THE IMPACT OF CONTEXTUAL VARIABLES ON THE PRODUCTIVITY: AN ENHANCED SLACK-BASED DEA MODEL

Linlin zhao[1], Zhangchen hu[2], Jun wang[3] and Yuliang zhuang[1]

Abstract. The contextual variable is an important issue that makes an indispensable impact on the productivities of decision making units (DMUs). Analyzing the contribution of such factors to productivity differences is an intriguing area of research in data envelopment analysis (DEA). We first investigate whether and how contextual variables impact performances of the DMUs based on slack-based measurement. We extend the implicit assumption of prior studies and suggest that contextual variables can be a catalyst to increase the productivity. Impact and error factors, which are derived from regression analysis and stochastic frontier analysis (SFA), are defined to better represent the composition of two contradictory impacts, catalyst and depressant, of contextual variables. A statistical analysis is provided to identify the significance of such impacts and recognize multi-collinearity among contextual variables. The two factors are also moderated flexibly by decision makers in accordance with various production scenarios. Accordingly, original inputs and outputs are appropriately adjusted. Further, modified slack-based DEA models are proposed to incorporate DEA and regression analysis within an integrated framework. Several properties and propositions are presented to better describe the characteristics of the models. An empirical example is shown to verify the feasibility of the proposed approach.

Mathematics Subject Classification. 90B030.

Received December 28, 2015. Accepted September 14, 2016.

1. Introduction

Contextual variables, such as the form of ownership, location characteristics, library and government regulations, are important factors which make indispensable impacts on the productivities of decision making units (DMUs). Analyzing the contribution of such factors to productivity differences is an intriguing content of research in Data Envelopment Analysis (DEA) [5]. Many efforts have been made by researchers to explore whether the impacts occur when contextual variables, characterized as exogenous or uncontrollable, exist and how their

Keywords. Data envelopment analysis, contextual variable, slack-based approach, productivity, stochastic frontier analysis.

[1] School of Management Science and Engineering, Nanjing Audit University, Yushan West Road 86, Nanjing, Jiangsu Province 211815, P.R. China. lin87@mail.ustc.edu.cn

[2] School of Management, University of Science and Technology of China, Jinzhai Road 96, Hefei, Anhui Province 230026, P.R. China.

[3] Bank of China, Jinzhai Road 131, Hefei, Anhui Province 230000, P.R. China.

Article published by EDP Sciences

effects are imposed on the productive and managerial practice. For example, Ray [23] identified various socio-economic factors which may influence the performance of school districts, and described the regressed DEA scores to show the implied effects. Camanho *et al.* [10] presented an enhanced DEA model to accommodate non-discretionary inputs and outputs and treated them differently as internal or external to the production process depending on the classification of these non-discretionary variables. Out of the DEA literature, the parametric estimator or specific parametric form is suggested to depict the relative performance of production (see Ref. [18]). Priori information or assumptions are needed to illustrate such parametric environment well. However, various parametric approaches may influence estimation results of the efficiencies [5]. Especially, when *a priori* knowledge is incomplete or deficient, the parametric method may not work well. Contrarily, DEA is an appropriate approach to deal with such situations. Besides, since contextual variables are exogenous and unable to be freely and discretionarily controlled by the DMUs, it is desired to incorporate additional considerations into existing DEA models so as to properly control these factors [3, 4, 11, 26].

As Fried *et al.* [13] mentioned, the approaches to specific characteristics of the environmental variables are to regard them as special inputs and outputs, but to restrict the performance evaluation to either inputs or outputs. Some pre-requirements may be necessary when those approaches are applied. Accordingly, two-stage or multiple-stage approaches, where efficiency scores are calculated by specific parametric or non-parametric models in the first stage and the variation of the scores is explained by using the regression analysis on account of the observed contextual variables in the second or multiple stage, are suggested to well illustrate the influence of contextual variables on productivity. For example, Pastor [22] introduced DEA into a two-stage framework where a specific-oriented DEA model is applied in either stage with different data sets and a comparison of the two efficiency scores depicts the impacts of environmental variables on productivity. Fried *et al.* [12] extended the two-stage approach by using Tobit regression analysis in the second stage to obtain a prediction of the impact of environmental variables. Since the environmental effects are statistically noise, it may not be appropriate to depict such effects with a deterministic DEA model. Accordingly, Fried *et al.* [13] expanded prior two-stage approaches to a three-stage analysis by considering random errors.

Within the framework of two or multiple-stage approaches, the arrangements of the DEA models, regression approaches and effects of contextual variables on the productivity discussed above raise a number of puzzling issues and need further consideration. First, does contextual variables affect technical efficiency or non-technical efficiency or simultaneously? If both, is the selected DEA model in previous researches such as Fried *et al.* [12,13] feasible and appropriate to better describe various influences of contextual variables, including technical and non-technical inefficiencies? If not, which model is a better choice? Second, do contextual variables affect the DMUs performances in a single and simplified way similar to that previous literature discussed? If not, what are the real effects of such variables? How do we characterize their positive or negative influences on the inputs or outputs or simultaneously? Third, which statistical or stochastic regression methods are appropriate to distinguish the positive and negative effects of the variables? In addition, if multiple contextual variables influence the productivity simultaneously, whether does multi-collinearity occur? To address these questions, we incorporate the specific aspects of contextual variables effects on the productivity into an integrated approach and use it to analyze the complex impacts of the variables to shed light on these empirical puzzles.

We first investigate whether and how contextual variables impact technical and non-technical performances of the DMUs simultaneously. Since the slacks from conventional CCR model can not reflect all the inefficiencies including non-technical inefficiency and technical inefficiency. Alternatively, we propose slack-based measure (SBM) to enable such identification of the effects on both technical and non-technical inefficiencies. Next, different from prior studies which assume contextual variables as part of the inputs and outputs (see Ref. [3]) or as an increase of them (see Ref. [13]), we make a helpful and feasible attempt to allow contextual variables to be a catalyst which can increase the productivity and the inputs could be reduced. Besides, we also make a necessary exploration on whether and how contextual variables have impacts on the outputs. We define impact factor and error factor on the basis of regression analysis and stochastic frontier analysis to better characterize contradictory impacts of contextual variables and random errors on various inputs and outputs as well. An appropriate statistical method is provided to identify the significance of such impacts. Since the regression analysis of multiple coefficients can result in the multi-collinearity and gives misleading information of the coefficients effects, we propose an approach to identify and manage the multi-collinearity in multiple regression analysis when multiple contextual variables co-exist and influence the productivity simultaneously. Moreover, since contextual variables and random error may have different importance on the impacts of the productivity, the two factors can be traded off by the decision maker with his preference and production practice. Accordingly, original inputs and outputs are moderately adjusted to depict individual inefficiency, contextual and random statistical error inefficiencies well. Further, we propose modified parametric slack-based DEA models to incorporate DEA and regression approaches into an integrated framework. Several properties and propositions are presented to better describe the characteristics of the models.

The paper is organized as follows. In Section 2, we briefly introduce the theoretical background of slack variables in DEA models and multiple-stage methods incorporating DEA and statistical models. A modified three-stage framework is proposed to evaluate the impact of contextual variables in Section 3. Section 4 provides a parametric slack-based DEA model incorporating three stages and Section 5 presents the algorithm of the method. An empirical example is illustrated in Section 6 to verify the feasibility of the proposed approach. Conclusions are made in the last section.

2. Theoretical background

2.1. Slacks in DEA models

The traditional CCR model, initiated by Charnes *et al.* [7], identified the technical efficiency through the score of θ:

$$
\begin{aligned}
&\min\ \theta \\
&\text{s.t.} \sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, i = 1, \ldots, m \\
&\quad \sum_{j=1}^{n} \lambda_j y_{rj} \geq y_{ro}, r = 1, \ldots, s \\
&\quad \lambda_j \geq 0
\end{aligned} \tag{2.1}
$$

A DMU is CCR-efficient if its ratio efficiency θ^* is equal to 1 and all slacks are equal to 0. Some researches tried to integrate θ and the slacks into a scalar measure. Charnes *et al.* [8] proposed an additive DEA model to deal with slack variables. Subsequent researches made further efforts to better illustrate the additive characteristics in DEA models (see Ref. [9]).

The efficiency score θ reflects the technical inefficiency and slack variables s_i^- and s_r^+ reflect non-technical inefficiency, such as allocative and managerial inefficiencies of the units. In Fried *et al.* [13] a CCR model is proposed to distinguish the slacks which are affected by contextual variables. The slacks they considered are a reflection of part of the inefficiency, whereas the inefficiency reflected by θ is ignored. In addition, as Aigner *et al.* [1] mentioned, errors of observation and measurement only constitute another source of statistical noise. In this way, Fried *et al.* [13] can not identify the impact of contextual variables on global productivity. Moreover, since traditional CCR model is input-oriented or output-oriented, it can only describe the influence of contextual variables on the inputs or outputs separately. If the inputs and outputs are influenced by contextual variables simultaneously, their models may not work well and further research is necessary to illustrate such situations.

Since the slacks in various DEA models may imply various inefficiencies, a proper model based on slack-based measure is suggested to determine various influences of contextual variables on technical and non-technical, *i.e.* allocative, and managerial performances. For this reason, the SBM DEA model [19, 27, 28], whose objective is to maximize the sum of the input and output slacks, is suggested in this study.

2.2. Multi-stage approaches incorporating DEA and statistical models

The basic idea of a multiple-stage approach identifying the influence of contextual variables is as follows. In the first phase, a DEA model is proposed to evaluate relative productivities of the observed units. Using regression analysis, such as ordinary least squares (OLS) proposed by Fried *et al.* [13], the efficiency scores are regressed in the second stage and the effects of contextual variables are investigated through the adjustments of the inputs and outputs. The significance of the influence is evaluated and the extent of such influence is obtained. The inputs and outputs are replaced with new post-adjusted data. Subsequently in the third stage, the first-stage model is applied again to acquire a group of new efficiency results eliminating the impact of contextual variables.

Prior researches focus their attention on two-stage or three-stage approaches to show the influence of contextual variables on performance. Ray [23] decomposed the effects with a second-stage regression. Ruggiero [24] extended traditional models to allow multiple non-discretionary inputs, and three-stage approaches based on regression analysis were applied to investigate the influences of the non-discretionary factors. Fried *et al.* [13] divided the slacks into three classifications, managerial inefficiency, the effect of contextual variables, and random errors, and suggested a three-stage approach incorporating stochastic frontier analysis (SFA) to illuminate the contextual influence on the performances of the units. Recently, Simar and Wilson [25] described a two-stage model incorporating data-generating process (DGP) and proposed single and double bootstrap procedures in the second-stage regression. Garca−Snchez [14] used a three-stage model to separate economic performance into three components which enables the author to detect specific strengths and weaknesses for each football club. Banker and Natarajan [5] and McDonald [20] discussed multiple-stage models applied in the analysis of the influence of contextual variables on performance so as to yield consistent estimators and superior parametric methods.

However, few researches take into account the complex impacts of contextual variables on the productivity, or consider an explicit illustration of the positive and negative influences simultaneously. Take Fried *et al.* [13] as an example, they proposed a situation where all the inputs are increased to yield the outputs given that the least favorable environment is taken as the base. Contrarily in production practice, contextual variables can affect the productivity in different ways. For some DMUs, the variables accelerate the production process and the inputs can be decreased to yield the desired outputs. Moreover, the outputs can also be affected by contextual variables in a positive or negative way. In this perspective, previous approaches should be extended to well illustrate the complex influences on both inputs and outputs.

3. Evaluating the impact of contextual variables in a three-stage framework

In an extension of Fried *et al.* [13], we first identify the slacks which include technical and non-technical by using slack-based measure (SBM). Incorporating stochastic frontier analysis (SFA) and maximum likelihood estimation (MLE), the slacks are classified into individual slacks, contextual slacks and random statistical error slacks. A statistical method is applied to detect multi-collinearity of independent variables and reflect the correlation between contextual variables and the input-output bundles. In addition, regression analysis is employed to determine the composite impacts of contextual variables. Subsequently, impact factor (IF) and error factor (EF) are defined to represent the effects of contextual variables and random errors on the productivity respectively. Further, a modified slack-based model is proposed to better incorporate DEA and regression analysis in an integrated framework.

3.1. Slack-based measure (SBM)

Considering n decision-making units (DMUs) with m inputs, s outputs and p contextual variables, the vectors $x_j = (x_{1j}, \ldots x_{mj})$, $y_j = (y_{1j}, \ldots y_{sj})$ and $z_j = (z_{1j}, \ldots z_{pj})$ denote the inputs, outputs and contextual variables of DMU_j, $j = 1, \ldots n$, respectively. The slack-based measure proposed by Tone [27] is as follows:

$$\begin{aligned}
&\min\ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}}{1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_r^+}{y_{ro}}} \\
&\text{s.t.} \sum_{j=1}^{n}\lambda_j y_{ro} - s_r^+ = y_{ro}, r = 1, \ldots, s \\
&\sum_{j=1}^{n}\lambda_j x_{io} + s_i^- = x_{io}, i = 1, \ldots, m \\
&\lambda_j, s_r^+, s_i^- \geq 0
\end{aligned} \tag{3.1}$$

Let $\rho = q$, $\lambda_j = \Lambda_j/t$, $s_i^- = S_i^-/t$ and $s_r^+ = S_r^+/t$, model (3.1) can be transformed into a linear form:

$$\begin{aligned}
&\min\ q = t - \frac{1}{m}\sum_{i=1}^{m}\frac{S_i^-}{x_{io}} \\
&\text{s.t. } t + \frac{1}{s}\sum_{r=1}^{s}\frac{S_r^+}{y_{ro}} = 1 \\
&\sum_{j=1}^{n}\Lambda_j x_{ij} + S_i^- = t x_{io}, i = 1, \ldots, m \\
&\sum_{j=1}^{n}\Lambda_j y_{rj} - S_r^+ = t y_{ro}, r = 1, \ldots, s \\
&\Lambda_j, S_r^+, S_i^- \geq 0
\end{aligned} \tag{3.2}$$

According to the theorem of Tone [27], a DMU is CCR-efficient if and only if it is SBM-efficient and the optimal SBM q^* is not greater than the optimal CCR θ^*, which reflects the fact that SBM accounts for all inefficiencies whereas θ^* accounts only for purely technical inefficiencies. In our study, the efficiency score and slacks are based on model (3.2).

The optimal solutions of model (3.2) provide $m + s$ slack variables for the evaluated unit which is subscripted as "o". Accordingly, these slack variables can explain the overall invalid degree of the units well. It is desirable to divide them into various effects, *i.e.* individual inefficiencies, contextual inefficiencies, and statistical noise, which are emphasized in the next stage.

3.2. Stochastic frontier analysis (SFA)

In the second stage, we use SFA to regress slack variables obtained in the first stage and divide them into various classifications to depict the influences of the observed contextual variables and error noise.

Initiated by Aigner *et al.* [1], the stochastic frontier production function is specified as:

$$t_j = f(z_j; \beta) + u_j + v_j \tag{3.3}$$

where u_j and v_j represent individual inefficiency and statistical noise inefficiency respectively, and $f(z_j; \beta)$ are deterministic feasible slack frontiers with parameter vector β to be estimated. In an economic logic, this specification implies that the production process is subject to two economically distinguishable random disturbances with different characteristics.

For the ith input of DMU_j, replace $f(z_j;\beta)$ and t_j with $z_j\beta^i$ and s_{ij} from model (3.2) in the first stage, and we can get the general form of SFA regressions:

$$s_{ij} = z_j\beta^i + u_{ij} + v_{ij}, i = 1,\ldots,m, j = 1,\ldots,n \tag{3.4}$$

Similarly, the general form of SFA regressions for the r th output of DMU_j is as follows:

$$s_{rj} = z_j\beta^r + u_{rj} + v_{rj}, r = 1,\ldots,s, j = 1,\ldots,n \tag{3.5}$$

In this study, regression analysis incorporating SFA has several significant advantages. First, the inefficiency of a unit can be attributed to contextual inefficiency, individual inefficiency and statistical noise inefficiency. A more comprehensive understanding of the inner causes which may account for inefficiency is obtained. Second, no matter whether the impact of contextual variable on the performance is positive or negative, it can be determined by the estimated regression parameters without any priori information or analysis. Such incorporation simplifies our analysis, and the significance of contextual impacts can also be easily obtained by applying conventional likelihood ratio tests.

Suppose v follows a normal distribution with a two-sided error term, *i.e.* $v_{ij} \sim N(0,\sigma_{vi}^2)$, $v_{rj} \sim N(0,\sigma_{vr}^2)$, and u follows the same distribution with a one-sided error term, *i.e.* $u_{ij} \sim N^+(0,\sigma_{ui}^2)$, $u_{rj} \sim N^+(0,\sigma_{ur}^2)$ [16], equations (3.4) and (3.5) are regressed $m+s$ times for each input and output using maximum likelihood techniques. FRONTIER 4.1 (see Ref. [6]) is applied and the results are obtained. In each regression, the parameters are estimated as $(\beta^i,\sigma_{vi}^2,\sigma_{ui}^2)$ and $(\beta^r,\sigma_{vr}^2,\sigma_{ur}^2)$.

From the above discussion, the error term in the stochastic frontier model follows the form $\varepsilon_{ij} = u_{ij} + v_{ij}$ and $\varepsilon_{rj} = u_{rj} + v_{rj}$, representing the errors of inputs and outputs respectively. u_{ij} and u_{rj} are non-negative error terms representing individual inefficiency, and v_{ij} and v_{rj} are normal error terms representing pure randomness. v_{ij} and u_{ij} are distributed independently of each other, which is the same as v_{rj} and u_{rj} [16]. The decomposition of error terms is to separate individual inefficiency and the normal error term, so as to clarify the real causes which affect the production process.

Given the error term ε, let us consider the conditional distribution of u. For notational simplicity, we ignore the subscript of ε, u and v, *i.e.* $\varepsilon = u + v$, and σ^2, u_* and σ_*^2 are defined as $\sigma^2 = \sigma_u^2 + \sigma_v^2$, $u_* = \sigma_u^2\varepsilon/\sigma^2$ and $\sigma_*^2 = \sigma_u^2\sigma_v^2/\sigma^2$. According to Jondrow *et al.* [16], the conditional distribution of u given ε follows a $N(u_*,\sigma_*^2)$ distribution which is truncated at zero. Thus we obtain:

$$E(u|v+u) = \sigma_*\left[\frac{f(\varepsilon\lambda/\sigma)}{1-F(\varepsilon\lambda/\sigma)} - \left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \tag{3.6}$$

where $\lambda = \sigma_u/\sigma_v$, $u_*/\sigma_* = \varepsilon\lambda/\sigma$ and $f(\cdot)$ and $F(\cdot)$ are standard normal density and distribution functions respectively.

Equation (3.6) can be converted to:

$$E(u|v+u) = \sigma_*\left[\frac{\exp(\varepsilon\lambda/\sigma)^2}{\sqrt{2\pi} - \int_{-\infty}^{\varepsilon\lambda/\sigma} \mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x} - \left(\frac{\varepsilon\lambda}{\sigma}\right)\right] \tag{3.7}$$

Thus, conditional estimators of v_{ij} and v_{rj} in equations (3.4) and (3.5) are obtained and equations (3.4) and (3.5) can be converted to the following:

$$E[v_{ij}|v_{ij}+u_{ij}] = s_{ij} - z_j\hat{\beta}^i - E[u_{ij}|v_{ij}+u_{ij}], i = 1,\ldots,m, j = 1,\ldots,n \tag{3.8}$$

$$E[v_{rj}|v_{rj}+u_{rj}] = s_{rj} - z_j\hat{\beta}^r - E[u_{rj}|v_{rj}+u_{rj}], r = 1,\ldots,s, j = 1,\ldots,n \tag{3.9}$$

Horsky and Nelson [15] proposed parameter significance tests to evaluate (3.8) and (3.9) and demonstrated that they are solvable when estimating the parameters. Subsequently, error terms are separated into two components, v and u.

3.3. Identifying multicollinearity in multiple regression analysis

In statistics, the greater the number of independent variables, the less discerning the analysis results of the multiple regression analysis is [17]. When two or more explanatory variables in a multiple regression model are highly correlated with one another, multi-collinearity often occurs among the variables. In practice, multi-collinearity can cause strange results when attempting to study how well individual independent variables contribute to a deep understanding of the impacts of the dependent variables. Once the collinear variables are identified, it can be helpful to study whether there is a causal link among the variables.

Mathematically, a set of independent variables $z^t = (z_{t1}, z_{t2}, \ldots, z_{tn}), t = 1, \ldots, p$ is perfectly multi-collinear if there exists one or more exact linear relationships among some of the variables.

$$\lambda_0 + \lambda_1 z_{1j} + \lambda_2 z_{2j} + \ldots + \lambda_p z_{pj} = 0$$

It holds for all observations $j, j = 1, \ldots, n$ where λ_t are constant and z_{tj} is the jth observation on the tth explanatory variable. However, in most applications, perfect multi-collinearity is usually unpractical. So we obtain a modified form of the equation with an error term b_j as follows:

$$\lambda_0 + \lambda_1 z_{1j} + \lambda_2 z_{2j} + \ldots + \lambda_p z_{pj} + b_j = 0$$

If the error term b_j is sufficient small for some set of the values, the variables $z^t, t = 1, \ldots, p$ are assumed nearly perfectly multi-collinear. Some authors have suggested a formal detection-tolerance or the variance inflation factor (VIF) for multi-collinearity (see Ref. [21]) as follows:

$$tolerance = 1 - R_t^2$$
$$VIF = \frac{1}{tolerance}$$

where R_t^2 is the coefficient of a regression of independent variable t on all the other independent variables. The bigger R_t^2, the bigger the standard error is, and the more z^t is correlated with the other independent variables. O'Brien [21] also gives the criterion for multi-collinearity problem that a tolerance less than 0.10 (or a VIF no less than 10) indicates that multi-collinearity problem may have a significant impact on the results.

One of the simplest ways to resolve multi-collinearity problems is to reduce the number of collinear variables. According to Jenkins and Anderson [17], if the independent variables $z^t, t = 1, \ldots, p$, are highly correlated, some of the variables can be omitted appropriately without loss of much information. The process is characterized simply as follows. First, the variables are normalized to have a mean of 0 and a variance of 1, which can make each variable selected equally important and have a total combined variance p. Second, denote $\sigma_{tt,t'}$ as the conditional variance of a variable representing the variance remaining in variable t when the effect of variable t' is removed. If variable t is perfectly correlated with t', then $\sigma_{tt,t'} = 0$. Third, if $t = 1, \ldots, k$ variables are perfectly correlated with $t = k + 1, \ldots, p$, then the variance of variable set $z^{k+1}, z^{k+2}, \ldots, z^p$ conditioning on $z^1, z^2, \ldots, z^k$ is equal to zero. Thus, variables $z^1, z^2, \ldots, z^k$ can be omitted with least loss of information. Since perfect correlation does not exist in any real data, a small error term $b_t \neq 0$ may always occur, and the residual variance in the conditioned variables can be easily detected as well. Thus it is easy to decide which variables can reasonably represent the information in all p variables by the residual variance.

3.4. Input and output adjustments

From the above analysis, we obtain a decomposition of the performance into three parts and the adjustments of the inputs are:

$$x_{ij}^A = x_{ij} + \left[\max_j(z_j\hat{\beta}^i) - z_j\hat{\beta}^i\right] + \left[\max_j(\hat{\upsilon}_{ij}) - \hat{\upsilon}_{ij}\right], i = 1, \ldots, m, j = 1, \ldots, n \tag{3.10}$$

where x_{ij} and x_{ij}^A are the observed and adjusted inputs, respectively. The formulas $\max_j(z_j\hat{\beta}^i) - z_j\hat{\beta}^i$ and $\max_j(\hat{\upsilon}_{ij}) - \hat{\upsilon}_{ij}$ represent the effects of contextual variables and random errors on the inputs respectively. Since $\max_j(z_j\hat{\beta}^i) - z_j\hat{\beta}^i$ and $\max_j(\hat{\upsilon}_{ij}) - \hat{\upsilon}_{ij}$ are no less than zero, Equation (3.10) makes it desirable to increase observed inputs. Such idea has been applied in several previous researches.

However, contextual variables may affect the inputs and outputs in various ways. Fried *et al.* [13] depicted a situation where the least favorable operating environment as the base to adjust the inputs. In this case, the data are adjusted by increasing the input levels for DMUs if contextual variables do not work. Contrarily, another situation may exist where contextual variables accelerate the production process and the inputs may be flexibly decreased to yield the desired outputs. Moreover, the outputs can also be affected by contextual variables in a positive or negative way. In this perspective, equation (3.10) is not appropriate to illustrate various influences of the inputs and outputs simultaneously.

In order to better depict the potential positive and negative influences of contextual variables on the inputs and outputs, and to avoid the possibility that some extremely disadvantaged (advantaged) contextual variables might have some inputs (outputs) adjusted downward as to become negative as well, we introduce two factors, impact factor and error factor, to represent the effects of contextual variables and random errors.

Definition 3.1. Impact factor (IF) is defined as the ratio estimator of input and output adjustment, which reflects the extent of how contextual variables impact them. It is influenced by the value of $z_j\beta^i$ or $z_j\beta^r$ derived from the result of regression analysis. Based on the values of $z_j\beta^i$, the impact factor of input i is formulized as follows:

If $\min_j(z_j\beta^i)$ is not less than zero,

$$IF_i = \frac{z_j\beta^i - \min_j(z_j\beta^i)}{\max_j(z_j\beta^i) - \min_j(z_j\beta^i)} \tag{3.11}$$

If $\min_j(z_j\beta^i)$ is less than zero and $\max_j(z_j\beta^i)$ is not less than zero,

$$IF_i = \frac{z_j\beta^i}{\max_j(z_j\beta^i) - \min_j(z_j\beta^i)} \tag{3.12}$$

If $\max_j(z_j\beta^i)$ is less than zero,

$$IF_i = \frac{z_j\beta^i - \min_j(z_j\beta^i)}{\max_j(z_j\beta^i) - \min_j(z_j\beta^i)} \tag{3.13}$$

The impact factor of output r is given by replacing the superscript i with r in equation (3.11) to (3.13).

Definition 3.2. Error factor (EF) is defined as the ratio estimator of input and output adjustment which reflects the extent of how random errors impact them. The value of v_{ij} or v_{rj} is obtained from the result of regression analysis. Based on various values of v_{ij}, the error factor of input i is formulized as follows:

If $\min_j(v_{ij})$ is not less than zero, then

$$EF_i = \frac{v_{ij} - \min_j(v_{ij})}{\max_j(v_{ij}) - \min_j(v_{ij})} \tag{3.14}$$

Otherwise,

$$EF_i = \frac{v_{ij}}{\max_j(v_{ij}) - \min_j(v_{ij})} \tag{3.15}$$

The error factor of output r is given by replacing the subscript i with r in equation (3.14) to (3.15).

Since the impact factor is defined as the proportional distance of $z_j\beta$ between the evaluated DMU and the reference DMUs, it can illustrate the influence of contextual variables on the units. In this perspective, it is feasible as an estimator to reasonably measure such influences. Similarly, the error factor is defined as the proportional distance of v_{ij} between evaluated DMU and the reference DMUs, it is applicable as an estimator to measure the overall errors.

In models (3.11), (3.12) and (3.13), if IF_i in the three models are less than zero, the inputs may be increased, indicating that contextual variables can benefit the unit and the inputs are consumed in a better way. In other words, contextual variables accelerate the productivity of the unit in a positive manner. Contrarily, if IF_i is greater than zero, it implies the inputs may be reduced, indicating that contextual variables be a detriment to the unit and consequently the inputs are consumed in a wasteful way. That is, contextual variables decelerate the productivity of the unit in a negative manner. A special case is that IF_i in the three models is equal to 0, which implies that contextual variables have no influence on the productivity. In addition, the statistical randomness can affect the outputs in a similar way.

Since IF and EF are obtained from regression analysis incorporating SFA, the influences of contextual variables on the inputs and outputs are made clear from the above analysis. The adjustment of the inputs can be reformulated as follows:

$$x_{ij}^A = \begin{cases} x_{ij}(1-\alpha\cdot IF_i-(1-\alpha)\cdot EF_i) & \text{if } \max_j(z_j\beta^i)\geq 0 \\ x_{ij}(1+\alpha\cdot IF_i-(1-\alpha)\cdot EF_i) & \text{otherwise} \end{cases} \quad \forall i,j \tag{3.16}$$

where α is a parameter whose purpose is to adjust the extent of various influences by contextual variables and random statistical errors. In some situations, statistical errors can impose a relatively significant impact on the inputs and outputs. Contrarily, the impact of statistical errors can be less significant than that of contextual variables. Decision makers could make a rational and flexible decision on the extents the two factors are contributing to the production process. In this perspective, α reflects a feasible trade-off between the two influences and is preliminarily given by decision makers. The first adjustment on the right side of equation (3.16) puts all production units into a common external environment and the second adjustment puts all production units into a common state of nature.

Apply equation (3.16) and x_{ij} is adjusted to x_{ij}^A. Correspondingly, the adjustment of outputs can be formulized similarly:

$$y_{rj}^A = \begin{cases} y_{rj}(1+\alpha\cdot IF_r+(1-\alpha)\cdot EF_r) & \text{if } \max_j(z_j\beta^r)\geq 0 \\ y_{rj}(1-\alpha\cdot IF_r+(1-\alpha)\cdot EF_r) & \text{otherwise} \end{cases} \quad \forall r,j \tag{3.17}$$

Using equation (3.17), y_{rj} is adjusted to y_{rj}^A. If y_{rj}^A is larger than y_{rj}, it implies that contextual variables have a positive impact on output r and the adjustment $z_j\hat{\beta}$ is not less than zero.

Equations (3.16) and (3.17) have the following properties.

Property 1. If $\max_j(z_j\beta^i)$ is not less than zero, then x_{ij}^A changes with $IF_i - EF_i$ and α.

(1) When $IF_i - EF_i$ is greater than 0, x_{ij}^A is decreasing in α.
(2) When $IF_i - EF_i$ is less than 0, x_{ij}^A is increasing in α.
(3) When $IF_i - EF_i$ is equal to 0, x_{ij}^A remain unchanged in α.

Proof. If $\max_j(z_j\beta^i)$ is not less than 0, the adjustment of x_{ij}^A is as follows:

$$x_{ij}^A = x_{ij}(1-\alpha\cdot IF_i-(1-\alpha)\cdot EF_i) = x_{ij}\{[1-EF_i]-\alpha\cdot[IF_i-EF_i]\}$$

Since IF_i and EF_i are obtained from regression analysis, they remain unchanged in model (3.16). If $IF_i - EF_i$ is greater than 0, x_{ij}^A is decreasing in α. If $IF_i - EF_i$ is less than 0, x_{ij}^A is increasing in α. If $IF_i - EF_i$ is equal to 0, x_{ij}^A remain unchanged in α.

Property 2. If $\max_j(z_j\beta^r)$ is not less than zero, the adjusted output y_{rj}^A will change with $IF_r - EF_r$ and α.

(1) When $IF_r - EF_r$ is greater than 0, y_{rj}^A is increasing in α.
(2) When $IF_r - EF_r$ is less than 0, y_{rj}^A is decreasing in α.
(3) When $IF_r - EF_r$ is equal to 0, y_{rj}^A remain unchanged in α.

The proof is similar to that of Property 1.

Properties 1 and 2 imply that the adjustments of the inputs and outputs are influenced by the difference between IF and EF and the preference of decision makers, *i.e.* the parameter α.

Proposition 3.3. *If* $\min_j(z_j\beta^i) \geq 0$*, then* $IF_i > 0$*. When* $EF_i < 0$*,* x_{ij}^A *is decreasing in* α*.*

Proof. From Property 1, x_{ij}^A is influenced by $IF_i - EF_i$.

When $\min_j(z_j\beta^i) \geq 0$,

$$IF_i = \frac{z_j\beta^i - \min_j(z_j\beta^i)}{\max_j(z_j\beta^i) - \min_j(z_j\beta^i)} \geq 0$$

If $EF_i < 0$, then $IF_i - EF_i$ is greater than 0. According to Property 1, x_{ij}^A is decreasing in α.

Proposition 3.4. *If* $\min_j(z_j\beta^r) \geq 0$*, then* $IF_r > 0$*. When* $EF_r < 0$*,* y_{rj}^A *is increasing in* α*.*

Proof. From Property 2, y_{rj}^A is influenced by $IF_r - EF_r$.

When $\min_j(z_j\beta^r) \geq 0$,

$$IF_r = \frac{z_j\beta^r - \min_j(z_j\beta^r)}{\max_j(z_j\beta^r) - \min_j(z_j\beta^r)} \geq 0$$

If $EF_r < 0$, $IF_r - EF_r$ is greater than 0. According to Property 2, y_{rj}^A is increasing in α.

Following Propositions 3.3 and 3.4, decision makers can have a clear estimation of how to choose an appropriate α to reflect the effects of contextual variables and random errors on the adjustments of the inputs and outputs.

3.5. A modified slack-based DEA model

Based on regression analysis, original inputs and outputs are adjusted and slack-based approach is applied again to obtain the efficiencies of the units.

In order to better illustrate the adjustment of the regression results, we propose a modified slack-based DEA model as follows:

$$\begin{aligned}
&\min\ \rho = \frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^{M-}}{x_{io}^A}}{1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_r^{M+}}{y_{ro}^A}} \\
&\text{s.t.}\ \sum_{j=1}^{n}\lambda_j y_{rj}^A - s_r^{M+} = y_{ro}^A, r = 1,\ldots,s \\
&\quad \sum_{j=1}^{n}\lambda_j x_{ij}^A + s_i^{M-} = x_{io}^A, i = 1,\ldots,m \\
&\quad \lambda_j, s_r^{M+}, s_i^{M-} \geq 0, j = 1,\ldots,n
\end{aligned} \tag{3.18}$$

Similar to the transformation of model (3.1), model (3.18) can also be transformed into a linear form:

$$\begin{aligned}
&\min\ q = t - \frac{1}{m}\sum_{i=1}^{m}\frac{S_i^{M-}}{x_{io}^A}\\
&\text{s.t. } 1 = t + \frac{1}{s}\sum_{r=1}^{s}\frac{S_r^{M+}}{y_{ro}^A}\\
&\quad \sum_{j=1}^{n}\Lambda_j x_{ij}^A + S_i^{M-} = t x_{io}^A, i = 1,\ldots,m\\
&\quad \sum_{j=1}^{n}\Lambda_j y_{rj}^A - S_r^{M+} = t y_{ro}^A, r = 1,\ldots,s\\
&\quad \Lambda_j, S_r^{M+}, S_i^{M-} \geq 0, j = 1,\ldots,n
\end{aligned} \tag{3.19}$$

Where $\rho = q$, $\lambda_j = \Lambda_j/t$, $s_i^{M-} = S_i^{M-}/t$ and $s_r^{M+} = S_r^{M+}/t$. In addition, x_{io}^A and y_{ro}^A are adjusted input and output quantities, respectively.

4. A parametric slack-based DEA model incorporating three stages

In an effort to better describe slack-based approach and regression analysis, we attempt to unify the three stages in an aggregated model.

$$\begin{aligned}
&\min\ \rho = (1-k)\frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^-}{x_{io}}}{1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_r^+}{y_{ro}}} + k\frac{1 - \frac{1}{m}\sum_{i=1}^{m}\frac{s_i^{M-}}{x_{io}^A}}{1 + \frac{1}{s}\sum_{r=1}^{s}\frac{s_r^{M+}}{y_{ro}^A}}\\
&\text{s.t. } \sum_{j=1}^{n}\lambda_j[(1-k)y_{rj} + k y_{rj}^A] - [(1-k)s_r^+ + k s_r^{M+}] = (1-k)y_{ro} + k y_{ro}^A, r = 1,\ldots,s\\
&\quad \sum_{j=1}^{n}\lambda_j[(1-k)x_{ij} + k x_{ij}^A] + [(1-k)s_i^- + k s_i^{M-}] = (1-k)x_{io} + k x_{io}^A, i = 1,\ldots,m\\
&\quad k s_i^- = k\cdot(z_j\hat{\beta}^i + \hat{u}_{ij} + \hat{v}_{ij})\\
&\quad k s_r^+ = k\cdot(z_j\hat{\beta}^r + \hat{u}_{rj} + \hat{v}_{rj})\\
&\quad \lambda_j, s_r^{M+}, s_i^{M-}, s_i^-, s_r^+ \geq 0, j = 1,\ldots,n, k = \{0,1\}
\end{aligned} \tag{4.1}$$

where s_i^- and s_r^+ are slacks obtained from model (3.1), and k is a predefined parameter changing with various stages. model (4.1) represents parametric slack-based DEA model incorporating three stages.

In the first stage, k is set to be 0 and model (4.1) is equivalent to the slack-based model by Tone [27]. In the third stage, k is set to be 1 and model (4.1) is the same as model (3.18).

5. Algorithm

(1) Solve model (3.1) and calculate the slacks of the units;
(2) Quantify contextual variables if qualitative;
(3) Estimate whether contextual variables affect input and output variables with t-ratio analysis, and regress various effects of the contextual variables using FRONTIER software and MATLAB;
(4) Adjust the inputs and outputs through equations (3.16) and (3.17);
(5) Calculate model (3.18) and the efficiency is obtained by considering the impact of contextual variables.

6. An empirical example

In an effort to validate the feasibility of the proposed approach, we conduct simulations to estimate the impact of contextual variables on the scientific research performances of the universities in China (see Ref. [2]).

6.1. Data description

Data of 30 universities are collected where the inputs are Fixed Assets (10 000 Yuan) (x_1), Researchers (x_2), and Graduate Students (x_3), and the outputs are SCI Papers (y_1) and SCI Citation (y_2). The contextual variables are Size of University 10 000 square meters (z^1) and Research Funding (10 000 Yuan) (z^2). Their summary statistics are presented in Table 1.

6.2. Simulation process

The initial efficiencies of 30 universities are calculated by applying model (3.2), and the input and output slacks are listed in Table 2. The initial DEA model does not provide a good measure of managerial performance. It may penalize the departments who operate in an unfavorable external environment and reward the departments who operate in a favorable external environment.

FRONTIER software is applied to proceed SFA, and the variables in equations (3.4) and (3.5) are obtained. Taking input and output slacks as dependent variables and contextual variables as independent variables, t-ratio analysis is followed to test the significance of contextual effects on the inputs and outputs. The results represented in Table 3 imply that contextual variables affect the inputs and outputs simultaneously, and all of them should be adjusted.

In order to detect multi-collinearity among independent variables, we conducted a multi-collinearity test by computing the tolerance and the variance inflation factor (VIF) for multi-collinearity as shown in Table 4.

Take the variable z^1 as an example. It has a tolerance of 0.863, which means that if running a multiple regression on z^1 as the dependent variables and z^2 as the independent variables, the R-square value is 0.137. As shown in Table 4, multi-collinearity does not appear to be a significant problem in our dataset. Since there are two contextual variables in our case, the results of the tolerance and VIF are the same.

The statistical results of regression analysis are summarized in Table 5. The statistical properties “MAX” and “MIN” reports the upper and lower bounds of IF and EF. β is the estimated parameter vector and σ^2 is the covariance of combination error. Since the values of γ are very small, it suggests that the contextual variables

Table 1. Descriptive statistics.

	Variables	Mean	Std.Dev	Max	Min
Inputs	x_1	255037.25	100971.49	540411.95	110783.23
	x_2	2915.90	1053.61	5135.00	1428.00
	x_3	11967.40	4017.43	21252.00	5310.00
Outputs	y_1	1023.97	715.39	3034.00	71.00
	y_1	3053.40	2744.75	8647.00	187.00

TABLE 2. Input and output slacks.

DMUs	s_1^-	s_2^-	s_3^-	s_1^+	s_2^+
1	0.0	0.0	0.0	0.0	0.0
2	141826.5	1069.0	0.0	242.9	7755.8
3	34990.7	0.0	3142.4	875.7	5313.0
4	0.0	820.3	2629.7	236.2	271.5
5	0.0	0.0	0.0	0.0	0.0
6	35643.0	743.3	0.0	611.6	9136.7
7	0.0	324.3	1507.9	612.7	5774.8
8	0.0	1292.0	34.3	1670.6	9138.0
9	0.0	1457.9	4337.6	2417.0	12797.2
10	0.0	2224.0	9105.6	1184.2	7315.4
11	0.0	722.3	2613.2	494.9	3881.0
12	37979.5	41.4	0.0	1765.9	10184.4
13	0.0	2355.2	9855.8	287.0	3120.9
14	0.0	1441.8	4406.6	1783.0	10634.4
15	22682.0	0.0	1305.0	2075.0	14310.0
16	0.0	1013.8	4134.7	246.8	4262.2
17	0.0	1023.5	3290.0	1245.0	8669.3
18	0.0	621.9	3759.2	571.4	5152.8
19	0.0	1287.3	5408.8	1257.1	8687.7
20	0.0	646.4	1169.9	759.0	6028.2
21	0.0	39.0	1519.4	986.4	6534.3
22	0.0	890.5	2716.8	1030.3	6875.2
23	0.0	254.9	1162.3	1559.4	7847.8
24	0.0	127.6	1234.1	775.0	5368.9
25	0.0	786.5	2530.1	1720.8	10024.3
26	0.0	1589.0	722.5	1256.3	7601.6
27	0.0	754.3	1763.8	1311.4	7659.9
28	19327.5	166.3	0.0	1495.7	8446.8
29	3864.5	252.8	0.0	1064.4	5943.5
30	0.0	637.7	961.9	552.8	3800.4

TABLE 3. T-ratio analysis of the impacts of contextual variables.

t-ratio	x_1	x_2	x_3	y_1	y_2
β^1	5.75E + 04	−1.25E + 01	−8.93	−1.29E + 01	−1.91
β^2	2.36E + 04	1.02E + 01	4.52	8.54	9.09E + 01

TABLE 4. The results of collinearity statistics.

Coefficients(a)			
Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	z^1	0.863	1.159
	z^2	0.863	1.159
a. Dependent Variable: s			

TABLE 5. Statistical properties of the result of regression analysis.

Item	Statistical property	Dependent Variable				
		FA(x_1)	R(x_2)	GS(x_3)	SP(y_1)	SC(y_2)
Independent Variable z	MAX	72677.16	1060.93	1961.38	991.66	6596.47
	MIN	6819.76	−296.83	−1507.23	−200.14	298.34
	β^1	57462.39	−1119.56	−3164.22	−958.65	−2372.94
	β^2	23646.53	1276.92	2572.07	1180.12	7103.89
	σ^2	522740560	311960.01	5720151.80	313692.42	12142585
	γ	0.04	0.03	0.02	0.04	0.50
Random noises v	MAX	79231.54	2021.94	9533.65	1971.06	5561.84
	MIN	−67916.76	−883.50	−1565.32	10.08	−466.05

TABLE 6. Comparison results of statistical properties of different weights.

Weights	0.5	0.6	0.7	0.8	0.9
Mean efficiency	0.3725	0.3753	0.3814	0.4077	0.4229
StDev.	0.2545	0.2574	0.2646	0.3052	0.3274
Max.	1.0000	1.0000	1.0000	1.0000	1.0000
Min.	0.0308	0.0293	0.0278	0.0262	0.0247

and statistical noise explain virtually all of the variations in the input and output slacks, and consequently the selected regression analysis method is remarkable. Table 5 also illustrates that input and output slacks are attributable to both contextual variables and statistical random error.

6.3. Discussion

As mentioned above, various values of α depict the trade-off of decision makers between contextual inefficiency and random statistical error inefficiency. The proposed model allows decision makers to set prior weights on the basis of their preferences or specific production processes. In an effort to better depict how the weights may influence adjustments of the inputs and outputs and the efficiencies of the DMUs, five weights are randomly chosen and the efficiency scores with different weights are listed in Tables 6 and 7. Table 6 reflects the variance of efficiency scores for different α and Table 7 depicts the detailed efficiencies and ranks of the DMUs. Obviously, the efficiency score changes as the weight of the impact factor alters, which illustrates that the impacts of contextual variables and random errors on the efficiency are related to the importance of contextual variables and random errors.

Figure 1 shows an obvious trend between the weights and the variation of mean efficiency of the DMUs.

Figure 2 shows the changes of randomly chosen DMUs with the variance of the weights. Obviously, the weights have various impacts on the DMUs. Consequently, the rankings of the efficiency scores are shown in Figure 3. Intuitively, the efficiency rankings of some DMUs change greatly, which implies that contextual variables impose a significant catalyst or depressant on inputs consumption and outputs production.

With the increase of parameter α, input adjustment on account of random errors is getting smaller, and efficiency scores of the DMUs change distinctly. Contrary to prior researches where the impacts of contextual variables and random errors are equally important implicitly, our approach can trade off the influences of contextual variables and random errors and thus shows more flexibility.

Table 8 represents the difference of the impact between input or output adjustment individually and their joint adjustment, through which significant differences can be seen among the various adjustment methods. As a result of considering the contextual variables and statistical errors, both the average efficiency and the standard deviation of the efficiency score increased. The increase of average efficiency suggests that without controlling for the contextual variables and statistical errors, the penalty to the DMUs under unfavorable circumstances is

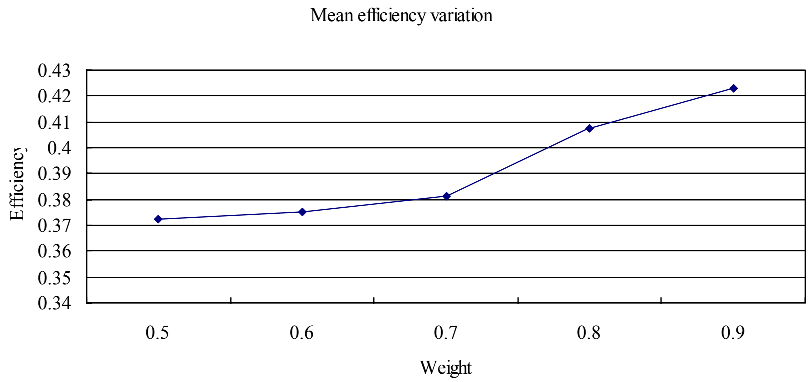

FIGURE 1. Variation of mean efficiency with change of weights.

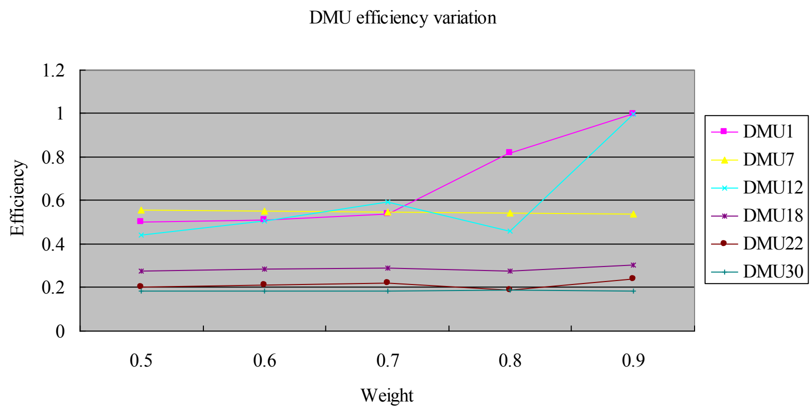

FIGURE 2. Variation of individual efficiency with change of weights.

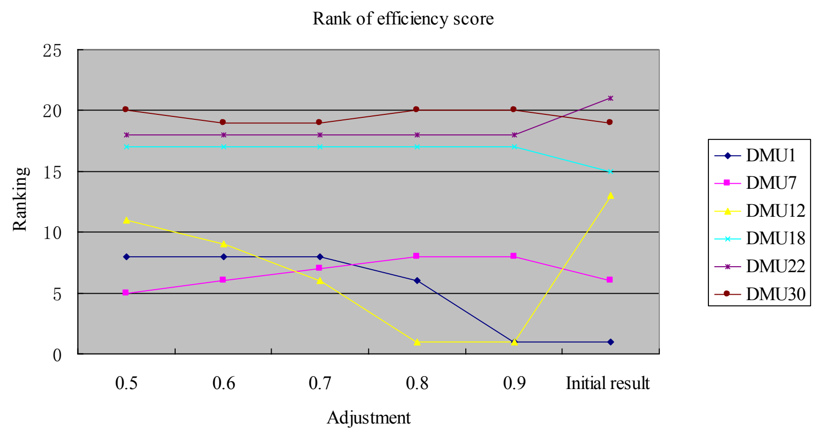

FIGURE 3. Variation of rankings with change of weights.

TABLE 7. Efficiency scores with various weights.

DMUs	$\alpha = 0.5$		$\alpha = 0.6$		$\alpha = 0.7$		$\alpha = 0.8$		$\alpha = 0.9$	
	efficiency	rank	efficiency	rank	efficiency	rank	efficiency	rank	efficiency	rank
1	0.4994	8	0.5095	8	0.5390	8	0.6153	6	1.0000	1
2	1.0000	1	1.0000	1	1.0000	1	1.0000	1	1.0000	1
3	0.6798	4	0.6441	4	0.6096	5	0.5759	7	0.5475	7
4	0.8023	3	0.8352	3	0.8702	3	0.9080	5	1.0000	1
5	1.0000	1	1.0000	1	1.0000	1	1.0000	1	1.0000	1
6	0.5364	6	0.5348	7	0.5330	9	0.5310	9	0.5288	9
7	0.5578	5	0.5526	6	0.5473	7	0.5417	8	0.5360	8
8	0.4355	12	0.4415	12	0.4474	11	0.4531	10	0.4588	10
9	0.5286	7	0.5884	5	0.6866	4	1.0000	1	1.0000	1
10	0.4909	9	0.4673	10	0.4486	10	0.4339	11	0.4224	13
11	0.4331	13	0.4251	13	0.4169	13	0.4084	14	0.3997	15
12	0.4394	11	0.5060	9	0.5913	6	1.0000	1	1.0000	1
13	0.4899	10	0.4605	11	0.4389	12	0.4230	12	0.4114	14
14	0.3833	14	0.3957	14	0.4089	14	0.4228	13	0.4375	12
15	0.1557	26	0.1536	26	0.1524	25	0.1524	24	0.1542	24
16	0.3567	15	0.3498	15	0.3432	16	0.3369	16	0.3307	16
17	0.1633	24	0.1545	25	0.1460	26	0.1379	26	0.1300	26
18	0.2779	17	0.2836	17	0.2894	17	0.2954	17	0.3015	17
19	0.1835	21	0.1817	20	0.1801	20	0.1786	21	0.1772	21
20	0.2801	16	0.3110	16	0.3466	15	0.3883	15	0.4392	11
21	0.1884	19	0.1803	21	0.1720	22	0.1636	22	0.1551	23
22	0.2019	18	0.2105	18	0.2194	18	0.2287	18	0.2384	18
23	0.0308	30	0.0293	30	0.0278	30	0.0262	30	0.0247	30
24	0.1788	22	0.1695	23	0.1603	24	0.1510	25	0.1418	25
25	0.1269	28	0.1212	28	0.1156	28	0.1102	28	0.1048	27
26	0.1582	25	0.1594	24	0.1605	23	0.1616	23	0.1627	22
27	0.1641	23	0.1709	22	0.1781	21	0.1857	19	0.1938	19
28	0.1292	27	0.1230	27	0.1167	27	0.1104	27	0.1040	28
29	0.1166	29	0.1129	29	0.1090	29	0.1050	29	0.1008	29
30	0.1861	20	0.1862	19	0.1861	19	0.1857	20	0.1851	20

greater than the benefit under favorable circumstances. The increase of the standard deviation of the efficiency scores may reflect the fact that without considering the contextual variables and statistical errors, the efficiency scores of the DMUs in favorable circumstances are biased downward, and the efficiency scores of the DMUs in unfavorable circumstances are biased upward.

7. Conclusions

In many real-life cases, contextual variables have a significant influence on the performances of decision making units, which brings difficulties in performance evaluation. In this paper, we argue that various DEA models may reflect various inefficiencies of the DMUs. An appropriate DEA model has a direct and significant effect on the DMUs by analyzing the impact of contextual variables. Slack-based measure is suggested to enable such illustration.

Moreover, contextual variables may have a direct influence on both the inputs and the outputs simultaneously. T-ratio, a statistical analysis, is induced to identify whether the influences exist and how they act. Multi-collinearity among contextual variables is also identified in the regression analysis. Based on the results of regression analysis and SFA, two factors are defined to represent the composition of various impacts, where the impact factor depicts the effect of contextual variables and the error factor illustrates the effect of random

TABLE 8. Results of different adjustments ($\alpha = 0.5$).

DMUs	Input adjustments	Ranking	Output adjustments	Ranking	Input and output adjustments	Ranking	Initial results	Ranking
1	0.8194	4	1.0000	1	0.4994	8	1.0000	1
2	1.0000	1	1.0000	1	1.0000	1	0.5544	5
3	0.6600	5	0.6787	5	0.6798	4	0.5921	4
4	0.8245	3	1.0000	1	0.8023	3	0.7346	3
5	1.0000	1	1.0000	1	1.0000	1	1.0000	1
6	0.5444	7	0.5528	7	0.5364	6	0.4373	7
7	0.5731	6	0.6139	6	0.5578	5	0.5184	6
8	0.3978	13	0.4392	11	0.4355	12	0.3139	11
9	0.4833	8	0.4276	12	0.5286	7	0.2430	14
10	0.3980	12	0.3995	13	0.4909	9	0.2801	12
11	0.4467	10	0.4777	9	0.4331	13	0.4230	8
12	0.4581	9	0.4393	10	0.4394	11	0.2569	13
13	0.4434	11	0.4817	8	0.4899	10	0.3846	9
14	0.3326	15	0.3399	15	0.3833	14	0.2216	16
15	0.1536	24	0.1498	26	0.1557	26	0.1080	28
16	0.3722	14	0.3891	14	0.3567	15	0.3327	10
17	0.1419	25	0.1539	25	0.1633	24	0.1326	22
18	0.2772	17	0.2961	16	0.2779	17	0.2298	15
19	0.1571	23	0.1667	23	0.1835	21	0.1221	23
20	0.2814	16	0.2886	17	0.2801	16	0.1956	17
21	0.1733	20	0.1884	20	0.1884	19	0.1771	20
22	0.1882	19	0.2018	19	0.2019	18	0.1448	21
23	0.0274	30	0.0306	30	0.0308	30	0.0277	30
24	0.1707	21	0.1818	21	0.1788	22	0.1776	18
25	0.1115	28	0.1222	28	0.1269	28	0.0995	29
26	0.1412	26	0.1548	24	0.1582	25	0.1158	25
27	0.1577	22	0.1687	22	0.1641	23	0.1177	24
28	0.1211	27	0.1295	27	0.1292	27	0.1141	26
29	0.1097	29	0.1203	29	0.1166	29	0.1081	27
30	0.1904	18	0.2062	18	0.1861	20	0.1776	19
Mean.	0.3719	–	0.3933	–	0.3725	–	0.3114	–
StDev.	0.2696	–	0.2909	–	0.2545	–	0.2527	–
Max.	1.0000	–	1.0000	–	1.0000	–	1.0000	–
Min.	0.0274	–	0.0306	–	0.0308	–	0.0277	–

errors. Original inputs and outputs can be moderately adjusted to depict contextual, individual, and statistical error inefficiencies respectively.

Further, we propose modified slack-based DEA models to uncover inner relationships of the three stages by incorporating DEA and regression method. An integrated slack-based DEA model is also presented to combine the three stages by presetting a designated parameter. An empirical application to universities in China is reported to justify the feasibility of the models.

Finally, this research addresses a larger concern about the extended role of efficiency evaluation. Our study points to the important and expanding role of the adjustments of inputs and outputs in the contextual variable domain. Moreover, feasible methods in the integrated contexts of DEA, and statistical and stochastic optimization theories must be appropriately selected and modified to accomplish this role. The encouraging feedback from this case study, together with the applicability of our efficiency results, leads us to believe that these results are helpful across the full range of decision contexts.

Acknowledgements. The authors thank the editor and two reviewers for their helpful comments. This work was supported by the Project Founded by Priority Academic Program Development of Jiangsu Higher Education Institutions. This work was supported by National Natural Science Foundation of China (Grant Nos. 71701102)

References

[1] D.J. Aigner, C.A.K. Lovell and P. Schmidt, Formulation and estimation of stochastic frontier production function models. *J. Econometrics* **6** (1977) 21–37.
[2] A. Amirteimoori and A. Emrouznejad, Flexible measures in production process: a DEA-based approach. *RAIRO: RO* **45** (2011) 63–74.
[3] R.D. Banker and R.C. Morey, Efficiency analysis for exogenously fixed inputs and outputs. *RAIRO: RO* **34** (1986a) 513–521.
[4] R.D. Banker and R.C. Morey, The use of categorical variables in data envelopment analysis. *Manage. Sci.* **32** (1986b) 1613–1627.
[5] R.D. Banker and R. Natarajan, Evaluating contextual variables affecting productivity using data envelopment analysis. *Oper. Res.* **56** (2008) 48–58.
[6] J. Beasley, Determining teaching and research efficiencies. *J. Oper. Res. Soc.* **46** (1995) 441–452.
[7] A. Charnes, W.W. Cooper and E. Rhodes, Measuring the efficiency of Decision-Making Units. *Eur. J. Oper. Res.* **2** (1978) 429–444.
[8] A. Charnes, W.W. Cooper, B. Golany, L. Seiford and J. Stutz, Foundation of data envelopment analysis and Pareto-koopmans empirical production functions. *J. Econometrics* **30** (1985) 91–107.
[9] W.W. Cooper, L.M. Seiford and J. Zhu, A unified additive model approach for evaluating inefficiency and congestion with associated measures in DEA. *Socio-Econ. Plan. Sci.* **34** (2000) 1–25.
[10] A.S. Camanho, M.C. Portela and C.B. Vaz, Efficiency analysis accounting for internal and external non-discretionary factors. *Comput. Oper. Res.* **36** (2009) 1591–1601.
[11] S.M. Estelle, A.L. Johnson and J. Ruggiero, Three-stage DEA models for incorporating exogenous inputs. *Comput. Oper. Res.* **37** (2010) 1087–1090.
[12] H.O. Fried, S.S. Schmidt and S. Yaisawarng, Incorporating the operating environment into a nonparametric measure of technical efficiency. *J. Pro. Anal.* **12** (1999) 249–267.
[13] H.O. Fried, C.A.K. Lovell S.S. Schmidt and S. Yaisawarng, Accounting for environmental effects and statistical noise in data envelopment analysis. *J. Pro. Anal.* **17** (2002) 157–174.
[14] I.M. Garca-Snchez, Efficiency and effectiveness of Spanish football teams: a three-stage-DEA approach. *Central Eur. J. Oper. Res.* **15** (2007) 21–45.
[15] D. Horsky and P. Nelson, Testing the statistical significance of linear programming estimators. *Manage. Sci.* **52** (2006) 128–135.
[16] J. Jondrow, C.A.K. Lovell, I.S. Materov and P. Schmidt, On the estimation of technical inefficiency in the stochastic frontier production function model. *J. Econometrics* **19** (1982) 233–238.
[17] L. Jenkins and M. Anderson, A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *Eur. J. Oper. Res.* **147** (2003) 51–61.
[18] K.P. Kalirajan, On measuring the contribution of human capital to agricultural production. *Indian Econ Rev.* (1989) 247–261.
[19] F.F. Liu and C.L. Chen, The worst-practice DEA model with slack-based measurement. *Comput. Ind. Eng.* **57** (2009) 496–505.
[20] J. McDonald, Using least squares and tobit in second stage DEA analyses. *Eur. J. Oper. Res.* **197** (2009) 792–798.
[21] R.M. Obrien, A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **41** (2007) 673–690.
[22] J.T. Pastor, How to account for environmental effects in DEA: an application to bank branches. Working Paper University of Alicante, Spain (1995).
[23] S.C. Ray, Resource-use efficiency in public schools: a study of Connecticut data. *Manage. Sci.* **37** (1991) 1620–8.
[24] J. Ruggiero, Non-discretionary inputs in data envelopment analysis. *Eur. J. Oper. Res.* **111** (1998) 461–469.
[25] L. Simar and P.W. Wilson, Estimation and inference in two-stage, semi-parametric models of production processes. *J. Econometrics* **136** (2007) 31–64.
[26] M.J. Syrjanen, Non-discretionary and discretionary factors and scale in data envelopment analysis. *Eur. J. Oper. Res.* **158** (2004) 20–33.
[27] K. Tone, A slack-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **130** (2001) 498–509.
[28] K. Tone, A slacks-based measure of super-efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **143** (2002) 32–41.