



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Forecasting Models Evaluation using A Slacks-based Context-dependent DEA Framework

Citation for published version:

Ouenniche, J, Bing, X & Tone, K 2014, 'Forecasting Models Evaluation using A Slacks-based Context-dependent DEA Framework', *Journal of Applied Business Research*, vol. 30, no. 5, pp. 1477-1484. <<http://www.cluteinstitute.com/ojs/index.php/JABR/article/view/8800>>

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

Published In:

Journal of Applied Business Research

Publisher Rights Statement:

© Ouenniche, J., Bing, X., & Tone, K. (2014). Forecasting Models Evaluation using A Slacks-based Context-dependent DEA Framework. *Journal of Applied Business Research*, 30(5).

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Forecasting Models Evaluation Using A Slacks-Based Context-Dependent DEA Framework

Jamal Ouenniche, University of Edinburgh, Edinburgh, United Kingdom & Business School, ESC Rennes, France

Bing Xu, Heriot-Watt University, UK

Kaoru Tone, National Graduate Institute for Policy Studies, Japan

ABSTRACT

Xu and Ouenniche (2012a) proposed an input-oriented radial super-efficiency Data Envelopment Analysis (DEA) based model to address a common methodological issue in the evaluation of competing forecasting models; namely, ranking models based on a single performance measure at a time, which typically leads to conflicting ranks. However, their approach suffers from a number of issues. In this paper, we overcome these issues by proposing a slacks-based context-dependent DEA framework and use it to rank forecasting models of oil prices' volatility.

Keywords: Forecasting Crude Oil Prices' Volatility; Performance Evaluation; Orientation-Free DEA

1. INTRODUCTION

The design of quantitative models for forecasting continuous variables in a wide range of application areas has attracted the attention of a large number of academics and professionals for some time; however, the performance evaluation of competing forecasting models has not received as much attention. Nowadays, although most published research involve using several performance criteria and measures to compare models, the performance evaluation exercise remains of a unidimensional nature; that is, models are ranked by a single measure and typically the obtained rankings are conflicting. Therefore, one cannot make an informed decision as to which model performs better under several criteria and their measures. To the best of our knowledge, the only papers that both raised concerns about this methodological issue and addressed it are the ones by Xu and Ouenniche (2011, 2012a, 2012b). A super-efficiency data envelopment analysis model has been proposed by Xu and Ouenniche (2011) to devise a multi-criteria ranking of competing forecasting models of oil prices' volatility; however, their approach suffers from the following issues. First, in many applications such as the ranking of forecasting models, the choice of an orientation is irrelevant. Second, under the variable returns-to-scale assumption, input-oriented scores can be different from output-oriented ones, which may lead to different rankings. Third, radial DEA models could only take account of technical efficiency and ignore potential slacks in inputs and outputs and thus may over-estimate efficiency scores. Fourth, radial super-efficiency DEA models may be infeasible for some efficient decision making units (DMUs) and would lead to unresolved ties. Finally, within a super-efficiency DEA framework, super-efficiency scores are used to rank order the efficient DMUs; however, the reference set changes from one efficient DMU evaluation to another, which in some contexts might be viewed as “unfair” benchmarking. In this paper, we overcome these issues by proposing a slacks-based context-dependent DEA (CDEA) framework (Morita, Hirokawa, & Zhu, 2005; Seiford & Zhu, 2003) for assessing the relative performance of competing volatility forecasting models.

The remainder of this paper is organized as follows. In Section 2, we describe the proposed slacks-based CDEA framework to evaluate the relative performance of competing volatility forecasting models. In Section 3, we report on our empirical findings. Section 4 concludes the paper.

2. A SLACKS-BASED CDEA MODEL FOR ASSESSING FORECASTING MODELS

In this paper, we propose a slacks-based CDEA framework to assess the relative performance of competing forecasting models for crude oil prices’ volatility. The proposed framework is a three-stage process which could be summarized as follows:

Stage 1 – Returns-to-Scale (RTS) Analysis: Perform RTS analysis to find out whether to solve a DEA model under constant returns-to-scale (CRS) conditions, variable returns-to-scale (VRS) conditions, increased returns-to-scale (IRS) conditions, or decreased returns-to-scale (DRS) conditions – see Banker, Cooper, Seiford, Thrall, and Zhu (2004) and Majid (2012) for details.

Stage 2 – Classification of DMUs: Use the following algorithm to partition the set of DMUs into several levels of best-practice frontiers or evaluation contexts, say L :

Initialization Step

Initialize the performance level counter ℓ to 1 and the set of $DMUs$ to evaluate at level ℓ , say J^ℓ , to $\{DMU_k, k=1, \dots, n\}$. Use the relevant DEA model to evaluate J^ℓ and set the ℓ^{th} -level best-practice frontier E^ℓ accordingly; that is, $E^\ell = \{k \in J^\ell \mid Efficiency\ Score\ \rho_k^\ell = 1\}$. Exclude the current performance level best-practice frontier E^ℓ from the set of $DMUs$ to evaluate next; that is, set $J^{\ell+1} = J^\ell - E^\ell$, increment ℓ by 1 and proceed to the iterative step.

Iterative Step

While $J^\ell \neq \emptyset$ **Do**

{

Use the relevant DEA model to evaluate J^ℓ , set the ℓ^{th} -level best-practice frontier E^ℓ accordingly, set $J^{\ell+1} = J^\ell - E^\ell$, and increment ℓ by 1;

}

where the relevant DEA model to use is the slacks-based measure (SBM) model of Tone (2001):

$$\begin{aligned}
 \text{Min} \quad & \rho_k^\ell = \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{s_{i,k}^-}{x_{i,k}} \right) \bigg/ \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{s_{r,k}^+}{y_{r,k}} \right) \\
 \text{s.t.} \quad & \sum_{j \in J^\ell} \lambda_j x_{i,j} + s_{i,k}^- = x_{i,k}; \forall i \\
 & \sum_{j \in J^\ell} \lambda_j y_{r,j} - s_{r,k}^+ = y_{r,k}; \forall r \\
 & \lambda_j \geq 0, \forall j \in J^\ell; s_{i,k}^- \geq 0, \forall i; s_{r,k}^+ \geq 0, \forall r
 \end{aligned} \tag{1}$$

where the i^{th} input and r^{th} output of $DMU_j (j=1, \dots, n)$ are denoted by $x_{i,j} (i=1, \dots, m)$ and $y_{r,j} (r=1, \dots, s)$, respectively, λ_j is the weight assigned to DMU_j in constructing its ideal benchmark, $s_{i,k}^-$ and $s_{r,k}^+$ are slack variables associated with the first and the second sets of constraints, respectively, and ρ_k^ℓ denotes the SBM efficiency score of DMU_k achieved at performance level ℓ . If the optimal value of $\rho_k^\ell = 1$, then DMU_k is classified as efficient; otherwise DMU_k is classified as inefficient. Note that model 1 above is solved as it is if stage 1 reveals that the CRS conditions hold; otherwise, one would have to impose one of the following additional constraints depending on whether VRS, IRS, or DRS conditions prevail, respectively:

$$\sum_{j \in J^\ell} \lambda_j = 1; \sum_{j \in J^\ell} \lambda_j \geq 1; \sum_{j \in J^\ell} \lambda_j \leq 1 \tag{2}$$

Obviously, once DMUs have been partitioned into L efficient frontiers with different levels of performance, one could rank order them from best to worst starting with 1st-level efficient frontier DMUs as best and ending with the L^{th} -level efficient frontier DMUs as worst. Note that ties might exist between DMUs on the same efficient frontier and the next stage is designed to break those ties.

Stage 3 – Break Efficiency Ties: First, for each efficient frontier E^ℓ ($\ell = 2, \dots, L$), compute *relative progress scores* δ_k^1 s with respect to the best evaluation context,¹ E^1 , by solving the following model for each $DMU_k \in E^\ell$ and rank order DMUs on efficient frontier E^ℓ according to the values of these scores:

$$\begin{aligned} \text{Min } \delta_k^1 &= \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{t_{i,k}^-}{x_{i,k}} \right) \bigg/ \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{t_{r,k}^+}{y_{r,k}} \right) \\ \text{s.t.: } \sum_{j \in E^1} \lambda_j x_{i,j} &\geq x_{i,k} - t_{i,k}^-; \forall i \\ \sum_{j \in E^1} \lambda_j y_{r,j} &\leq y_{r,k} + t_{r,k}^+; \forall r \\ \lambda_j &\geq 0, \forall j \in E^1; t_{i,k}^- \geq 0, \forall i; t_{r,k}^+ \geq 0, \forall r \geq 0 \end{aligned} \tag{3}$$

where $t_{i,k}^-$ (respectively, $t_{r,k}^+$) denotes the amount by which input i (respectively, output r) of DMU_k should be decreased (respectively, increased) to reach the efficient frontier corresponding to evaluation context E^1 . Second, for DMUs belonging to the best efficient frontier E^1 , compute *relative attractiveness scores* γ_k^2 s with respect to the second best evaluation context,² E^2 , by solving the following model for each $DMU_k \in E^1$ and rank order DMUs on the best efficient frontier according to the values of these scores:

$$\begin{aligned} \text{Max } \gamma_k^2 &= \left(1 - \frac{1}{m} \sum_{i=1}^m \frac{t_{i,k}^+}{x_{i,k}} \right) \bigg/ \left(1 + \frac{1}{s} \sum_{r=1}^s \frac{t_{r,k}^-}{y_{r,k}} \right) \\ \text{s.t.: } \sum_{j \in E^2} \lambda_j x_{i,j} &\leq x_{i,k} + t_{i,k}^+; \forall i \\ \sum_{j \in E^2} \lambda_j y_{r,j} &\geq y_{r,k} - t_{r,k}^-; \forall r \\ \lambda_j &\geq 0, \forall j \in E^2; t_{i,k}^+ \geq 0, \forall i; t_{r,k}^- \geq 0, \forall r \end{aligned} \tag{4}$$

where $t_{i,k}^+$ (respectively, $t_{r,k}^-$) denotes the amount by which input i (respectively, output r) of $DMU_k \in E^1$ should be increased (respectively, decreased) to reach the frontier corresponding to evaluation context E^2 .

In the next section, we use the proposed procedure to rank order competing forecasting models of crude oil prices' volatility and report on our empirical findings.

¹ The rationale behind this choice is to set a common global target for all lower level efficient frontiers for the sake of fairness in benchmarking.

² The rationale behind this choice is to compare the most efficient DMUs with those that have the closest performance.

3. EMPIRICAL INVESTIGATION AND RESULTS

For comparison purposes with the results obtained by Xu and Ouenniche (2012a), we use the same data, inputs, outputs, and forecasting models.³ Note that RTS analysis revealed that VRS conditions hold and therefore models 1, 3, and 4 are augmented with the following constraint: $\sum_{j \in J'} \lambda_j = 1$. Table 1 (respectively, Tables 2 and 4) provide the unidimensional (respectively, multidimensional) rankings of 14 forecasting models of crude oil prices' volatility based on 9 measures of 3 criteria: *biasedness*, *goodness-of-fit*, and *correct sign*. Table 1 is a typical output presented by most existing forecasting studies – these unidimensional rankings are devised as follows: models are ranked from best to worst using the relevant measure of each of the criteria under consideration. Notice that different criteria led to different unidimensional rankings, which provides evidence of the problem resulting from the use of a unidimensional approach in a multicriteria setting. Table 2 summarizes multidimensional rankings, where the models are ranked from best to worst based on the corresponding super-efficiency scores obtained using both input-oriented and output-oriented radial super-efficiency DEA models. Notice that, under VRS conditions, the rankings of input- and output-oriented analyses are different and the rankings of output-oriented analysis show more ties. Table 3 provides the efficient frontiers obtained with SBM-CDEA. These results suggest that the best and the worst efficient frontiers are insensitive to adjusting biasedness measures for volatility. Note that any rankings based on these efficient frontiers would lead to a large number of ties. In order to break these ties, we use relative progress and attractiveness scores obtained by solving models 3 and 4, respectively, which results in the multidimensional rankings provided in Table 4 where models are ranked from best to worst based on these relative scores. Notice that Tables 2 and 4 reveal that the multicriteria rankings of models obtained by input- and output-oriented super-efficiency DEA analyses and SBM-CDEA analysis are different. These differences are due to the fact that input-oriented analysis minimizes inputs for fixed amounts of output and output-oriented analysis maximizes outputs for fixed amounts of input, whereas orientation-free analysis optimizes both inputs and outputs simultaneously. In addition, oriented super-efficiency analyses only take account of technical efficiency, whereas orientation-free CDEA analysis takes account of an additional performance component; namely, slacks. In fact, for our data set – see Table 4, the efficient model SMA20 maintained its best position in the rankings regardless of the type of DEA analysis, because it is always on the best efficient frontier and has zero slacks regardless of the performance measures used. As to the rankings of the remaining models, there are differences that are mainly due to the presence of slacks and the nature of benchmarks.

³ Inputs (respectively, outputs) consist of performance metrics to be minimized (respectively, maximized) according to the principle of the less (respectively, the more) the better – for space limitations, description of performance metrics and forecasting models are provided in Xu and Ouenniche (2012).

Table 1: Unidimensional Rankings of Competing Forecasting Models

Criteria	Measures	Ranked From Best to Worst
Biasedness	Mean Error (ME)	3→5→11→9→14→10→12→4→13→6→2→8→7→1
	Mean Volatility-Scaled Errors (MVolScE)	3→5→9&10&11&12&14→4→6&13→2→8→7→1
Goodness-of-fit	Mean Absolute Error (MAE)	8→5→6→3→9&10→11→12→7→14→13→4→1→2
	Mean Absolute Volatility-Scaled Errors (MAVolScE)	5&8→3&6→9&10&11&12→7&14→13→4→1→2
	Mean Squared Error (MSE)	14→13→10→12→11→9→5→3→6→4→8→7→2→1
	Mean Squared Volatility-Scaled Errors (MSVolScE)	14→13→10→11&12→9→5→3→6→4→8→7→2→1
	Mean Mixed Error Under-estimation penalized (MMEU)	14→3→5→10→13→12→9→11→4→6→8→7→2→1
	Mean Mixed Error Over-estimation penalized (MMEO)	1→2→7→8→6→11→12→9→13→10→14→4→5→3
Correct Sign	Percentage of correct direction change predictions (PCDCP)	3→5→10&12→9→14→4→13→6→8→11→1&7→2

*¹RW [Random Walk]; ²HM [Historical Mean]; ³SMA20 [Simple Moving Average]; ⁴SMA60; ⁵SES [Single Exponential Smoothing]; ⁶ARMA(1, 1) [Auto Regressive Moving Average]; ⁷AR(1) [AutoRegressive]; ⁸AR(5); ⁹GARCH(1, 1) [Generalized Auto Regressive Conditional Heteroscedasticity]; ¹⁰GARCH-M(1, 1); ¹¹EGARCH (1,1) [Exponential GARCH]; ¹²TGARCH (1, 1) [Threshold GARCH]; ¹³PARCH (1, 1) [Power Auto Regressive Conditional Heteroscedasticity]; ¹⁴CGARCH(1,1) [Component GARCH]

Table 2: Input- and Output-oriented Super-Efficiency Rankings

Panel A: Combinations of Performance Measures used as Inputs Along With Output PCDCP in Input-Oriented Analysis		Panel B: Combinations of Performance Measures used as Inputs Along With Output PCDCP in Output-Oriented Analysis	
Inputs	Models Ranked from Best to Worst	Inputs	Models Ranked from Best to Worst
ME; MAE	3→ 5→8→6→9→10→11→12→14→7→13→4→1→2	ME; MAE	3&5&8→10&12→9→14→6→4→13→11→1→7→2
ME; MAVolScE	3→ 5→8→6→9&10&11&12→14→7→13→4→1→2	ME; MAVolScE	3&5→10&12→9→14→4→8→13→6→11→1→7→2
ME; MSE	3→ 14→5→11→10→12→9→13→6→4→8→7→2→1	ME; MSE	3→14→5→10→12→9→13→4→6→11→8→1&7→2
ME; MSVolScE	3→ 14→5→11→10→12→9→13→6→4→8→7→2→1	ME; MSVolScE	3→14→5→10→12→9→13→4→6→11→8→1&7→2
ME; MMEU	3→ 14→5→10→13→12→9→11→4→6→8→7→2→1	ME; MMEU	3&14→5→10&12→9→4→13→6→8→11→1→7→2
ME; MMEO	3→1→2→11→5→8→6→9→12→10→7→14→4→13	ME; MMEO	3→2→1→11→5→8→6→9→12→10→7→14→4→13
MVolScE; MAE	3→ 5→8→6→9&10→11→12→14→7→13→4→1→2	MVolScE; MAE	3&5&8→10&12→9→14→6→4→13→11→1&7→2
MVolScE; MAVolScE	3→ 5→8→6→9&10&11&12→14→7→13→4→1→2	MVolScE; MAVolScE	3&5→10&12→9→14→4→8→13→6→11→1→7→2
MVolScE; MSE	3→ 14→5→10→12→9→11→13→6→4→8→7→2→1	MVolScE; MSE	3→14→5→10→12→9→13→4→6→11→8→1&7→2
MVolScE; MSVolScE	3→ 14→5→10→12→9→11→13→6→4→8→7→2→1	MVolScE; MSVolScE	3→14→5→10→12→9→13→4→6→11→8→1&7→2
MVolScE; MMEU	3→ 14→5→10→13→12→9→11→4→6→8→7→2→1	MVolScE; MMEU	3&14→5→10&12→9→4→13→6→8→11→1→7→2
MVolScE; MMEO	3→1→2→11→5→8→6→12→9→10→14→7→4→13	MVolScE; MMEO	3→2→1→11→5→8→6→12→9→10→7→14→4→13

*¹RW; ²HM; ³SMA20; ⁴SMA60; ⁵SES; ⁶ARMA(1, 1); ⁷AR(1); ⁸AR(5); ⁹GARCH(1,1); ¹⁰GARCH-M(1,1); ¹¹EGARCH(1,1); ¹²TGARCH(1,1); ¹³PARCH(1, 1); ¹⁴CGARCH(1,1)

Table 3: Efficient Frontiers With Different Performance Levels

Efficient Frontiers	Panel A: Combinations of Performance Measures used as Inputs Along With Output PCDCP					Panel B: Combinations of Performance Measures used as Inputs Along With Output PCDCP				
	ME & MAE	ME & MAVolScE	ME & MSE; ME & MSVolScE	ME & MMEU	ME & MMEO	MVolScE & MAE	MVolScE & MAVolScE	MVolScE & MSE; MVolScE & MSVolScE	MVolScE & MMEU	MVolScE & MMEO
E^1	{3,5,8}	{3,5}	{3,5,14}	{3,14}	{1,2,3,5,6,8,11}	{3,5,8}	{3,5}	{3,5,14}	{3,14}	{1,2,3,5,6,8,11}
E^2	{6,9,10,11}	{6,8,9,10,11}	{9,10,11,13}	{5}	{7,9,10,12}	{6,11}	{6,8,10,12}	{10,13}	{5}	{7,12}
E^3	{12,14}	{12,14}	{12}	{9,10,11}	{13,14}	{9,12}	{9}	{12}	{10}	{9,10}
E^4	{4,7,13}	{4,7,13}	{4,6}	{12,13}	{4}	{4,7,13}	{11,14}	{9,11}	{12,13}	{13,14}
E^5	{1,2}	{1,2}	{2,8}	{4}		{1,2}	{4,7,13}	{4,6}	{9}	{4}
E^6			{7}	{6}			{1,2}	{2,8}	{4,11}	
E^7			{1}	{2,8}				{7}	{6}	
E^8				{7}				{1}	{2,8}	
E^9				{1}					{7}	
E^{10}									{1}	

*¹RW; ²HM; ³SMA20; ⁴SMA60; ⁵SES; ⁶ARMA(1, 1); ⁷AR(1); ⁸AR(5); ⁹GARCH(1,1); ¹⁰GARCH-M(1,1); ¹¹EGARCH(1,1); ¹²TGARCH(1,1); ¹³PARCH(1, 1); ¹⁴CGARCH(1,1)

Table 4: SBM-CDEA Rankings

Panel A: Combinations of Performance Measures used as Inputs Along With Output PCDCP		Panel B: Combinations of Performance Measures used as Inputs Along With Output PCDCP	
Inputs	Models Ranked from Best to Worst	Inputs	Models Ranked from Best to Worst
ME; MAE	3→5→8→6→9&10→11→12→14→13→4→7→2→1	MVolScE; MAE	3→5→8→6→10→9→12→14→11→13→4→7→2→1
ME; MAVolScE	3→5→11→9→10→6→8→14→12→4→13→7→2→1	MVolScE; MAVolScE	3→5→10&12→6→8→9→14→11→4→13→7→2→1
ME; MSE	3→5→14→9→11→10→13→12→4→6→8→2→7→1	MVolScE; MSE	3→5→14→10→13→12→9→11→4→6→8→2→7→1
ME; MSVolScE	3→5→14→9→11→10→13→12→4→6→8→2→7→1	MVolScE; MSVolScE	3→5→14→10→13→12→9→11→4→6→8→2→7→1
ME; MMEU	3→14→5→9→10→11→12→13→4→6→8→2→7→1	MVolScE; MMEU	3→14→5→10→12→13→9→11→4→6→8→2→7→1
ME; MMEO	3→5→2→11→1→6→8→9,10&12→14→13→4	MVolScE; MMEO	3→5→2→1→11→6→8→12→7→9&10→14→13→4

*¹RW; ²HM; ³SMA20; ⁴SMA60; ⁵SES; ⁶ARMA(1, 1); ⁷AR(1); ⁸AR(5); ⁹GARCH(1,1); ¹⁰GARCH-M(1,1); ¹¹EGARCH(1,1); ¹²TGARCH(1,1); ¹³PARCH(1, 1); ¹⁴CGARCH(1,1)

4. CONCLUSION

Xu and Ouenniche (2012a) proposed an input-oriented radial super-efficiency DEA-based framework to evaluate the performance of competing forecasting models of crude oil prices' volatility, which delivers a single ranking based on multiple performance criteria; such a framework suffers from several issues that were overcome in this paper. The main results may be summarized as follows. First, models that are on the efficient frontier and have zero slacks regardless of the performance measures used (e.g., SMA20) maintain their ranks regardless of the choice of DEA analysis and its orientation. Second, the multicriteria rankings of the best and the worst models seem to be relatively robust to changes in most performance measures; in sum, SMA20 is the best across the board and, for the remaining models, differences in rankings were mainly due to the presence of slacks and the nature of benchmarks. Finally, when under-estimated forecasts are penalized, most GARCH types of models tend to perform well – suggesting that they often produce forecasts that are over-estimated. On the other hand, when over-estimated forecasts are penalized, averaging models such as RW, HM, SES tend to perform very well – suggesting that these models often produce forecasts that are under-estimated.

AUTHOR INFORMATION

Dr. Jamal Ouenniche is a Reader in Management Science at the University of Edinburgh, Business School, Edinburgh, UK. His research focuses on the design and implementation of mathematical programming-based and artificial intelligence-based methods. His papers have been published in journals such as *Operation Research*, *European Journal of Operational Research*, and *International Journal of Production Research* amongst others. E-mail: Jamal.Ouenniche@ed.ac.uk (Corresponding author)

Dr. Bing Xu is a Lecturer at the School of Management & Languages; Accountancy, Economics and Finance; Heriot-Watt University; Edinburgh; EH14 4AS, UK. Her papers have been published in *Expert Systems with Applications*, *Energy Economics*, and *Applied Financial Economics* amongst others. E-mail: b.xu@hw.ac.uk

Professor Kaoru Tone is Professor Emeritus at the National Graduate Institute for Policy Studies, Tokyo, Japan. His papers have been published in *European Journal of Operational Research*, *Journal of the Operational Research Society*, *Omega*, *International Transactions in Operational Research*, and *Annals of Operations Research* amongst others.

REFERENCES

1. Banker, R. D., Cooper, W. W., Seiford, L. M., Thrall, R. M., & Zhu, J. (2004). Returns to scale in different DEA models. *European Journal of Operational Research*, 154, 345-362.
2. Majid, S.-D. (2012). On a basic definition of returns to scale. *Operations Research Letters*, 40, 144-147.
3. Morita, H., Hirokawa, K., & Zhu, J. (2005). A slack-based measure of efficiency in context-dependent data envelopment analysis. *Omega*, 33, 357-362.
4. Seiford, L. M., & Zhu, J. (2003). Context-dependent data envelopment analysis – measuring attractiveness and progress. *Omega*, 31, 397-408.
5. Tone, K. (2001). A slacks-based measure of efficiency in data envelopment analysis. *European Journal of Operational Research*, 130, 498-509.
6. Xu, B., & Ouenniche, J. (2011). A multidimensional framework for performance evaluation of forecasting models: context-dependent DEA. *Applied Financial Economics*, 21, 1873-1890.
7. Xu, B., & Ouenniche, J. (2012a). A data envelopment analysis-based framework for the relative performance evaluation of competing crude oil prices' volatility forecasting models. *Energy Economics*, 34, 576-583.
8. Xu, B., & Ouenniche, J. (2012b). Performance evaluation of competing forecasting models: A multidimensional framework based on MCDA. *Expert Systems with Applications*, 39, 8312-8324.

NOTES