This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Eskelinen, Juha

Comparison of variable selection techniques for data envelopment analysis in a retail bank

Published in:
European Journal of Operational Research

DOI:
[10.1016/j.ejor.2016.11.009](https://doi.org/10.1016/j.ejor.2016.11.009)

Published: 01/01/2017

Document Version
Peer reviewed version

Published under the following license:
CC BY-NC-ND

Please cite the original version:
Eskelinen, J. (2017). Comparison of variable selection techniques for data envelopment analysis in a retail bank. *European Journal of Operational Research*, 259(2), 778-788. <https://doi.org/10.1016/j.ejor.2016.11.009>

Comparison of variable selection techniques for data envelopment analysis in a retail bank

Juha Eskelinen

Aalto University, School of Business, Department of Information and Service Economy,

P.O. Box 21220, 00076 Aalto, Helsinki, FINLAND

E-mail: juha.p.eskelinen@aalto.fi, Tel. +358-400-813858

Abstract

If there are too few units compared to inputs and outputs, the efficiency evaluation based upon the data envelopment analysis suffers from a lack of discrimination. The literature has proposed various statistical techniques when the value judgments do not guide the selection of the inputs/outputs. Two techniques, the variable reduction procedure of Jenkins and Anderson (2003) and the approach based upon the efficiency contribution measure of Pastor, Ruiz, and Sirvent (2002), were compared in an empirical retail bank context. The objective was to select a representative set of outputs from the services the bank provides. As the techniques take different approaches to selecting influential variables, the output sets proposed by the techniques diverged. This created some significant differences in the efficiency evaluations of the bank branches. The bank management gave feedback on the techniques and the results from a practical perspective. The techniques led to different managerial interpretations of the performance complementing each other. Thus the techniques can be utilized to evaluate the units from multiple perspectives.

Keywords: Data envelopment analysis; OR in banking; Variable selection; Productive efficiency analysis

1. Introduction

Data envelopment analysis (DEA) is a method used to evaluate the efficiencies of comparable units with multiple inputs and/or outputs. It takes an optimistic perspective by choosing, for each decision-making unit (DMU), the most beneficial non-negative (i.e. those weights will maximize its efficiency score). This virtue has a consequence. Having a low number of units under assessment compared to inputs and outputs leads to finding a large portion of the units efficient or nearly efficient. The identification of many units falsely efficient and too optimistic efficiency scores is called a *lack of discrimination* (see e.g., Podinovski & Thanassoulis, 2007). Such an efficiency estimation may not serve the purpose of the evaluation.

Banks consume several resources (inputs) to produce several services or other outcomes (outputs). Even though there are large nationwide and international banks, the majority of the branch networks are small with a relatively low number of branches.¹ In addition, in the large banks, the branches are often clustered into smaller homogeneous subsets. Thus, the lack of discrimination can be an issue in bank branch efficiency evaluations (see, e.g., the discussion of Paradi & Zhu, 2013).

This essay focuses on the variable side (i.e., inputs and outputs) of the discrimination problem. The number of variables can be reduced by selecting the most important outputs/inputs to depict the activity. Numerous statistical techniques have been proposed for situations where judgmental knowledge is not available or not sufficient enough for selecting the variables. The following review illustrates the variety of techniques.²

Some selective techniques omit variables prior to DEA. Lewin, Morey, and Cook (1982), among others, used correlation and regression analysis to test redundant variables to be omitted. Jenkins and Anderson (2003) proposed a variable reduction procedure based on partial covariance (abbreviated as VR in this paper), where a combination of variables is selected based upon the proportion of the total variance retained. Gonzalez-Bravo (2007)

¹ In the United States, the commercial banks had, on average, 13.2 branches in 2011 (data source: U.S. Federal Deposit Insurance Corporation). In the United Kingdom, the monetary financial institutions have 28.7, in Germany 18.9, in France 31.8, and in Finland 3.8 branches (data source: European Central Bank 2011, UK data from 2010). The exceptionally low figure in Finland is explained by the high proportion of small local cooperative and savings banks.

² Besides selective techniques, that are the focus of this paper, there is another branch of statistical techniques that deals with the lack of discrimination by determining statistically the weights for summing up the data in a reduced set of variables. Various techniques ending up in aggregation have been proposed by Sengupta (1990), Sinuany-Stern and Friedman (1998), Ueda and Hoshiai (1997), Adler and Golany (2001, 2002), Bian (2012), and Amirteimoori, Despotis, and Kordrostami (2014).

proposed a procedure to help model specification by using an output/input ratio-based analysis prior to DEA.

Another sub-branch of selective techniques tests the impact of the retained variables on DEA results. Banker (1996) proposed the use of statistical tests for different inefficiency distribution assumptions, to indicate the significance of a variable in the production process (i.e., efficiency estimates). Simar and Wilson (2001) recommended bootstrapping to test whether the input or output variables are relevant. Pastor, Ruiz, and Sirvent (2002) introduced an efficiency contribution measure (ECM) to test for the significance of a variable in efficiency estimation. Fanchon (2003), Ruggiero (2005) and Sharma and Jin (2015) regressed the efficiency scores against the variables and tested the regression coefficient in dropping the insignificant variables. Wagner and Shimshak (2007) proposed a stepwise procedure that evaluated the impact of the variable set on the average efficiency scores.

There are a few published studies comparing some of the statistical variable selection techniques (see Sirvent, Ruiz, Borrás, & Pastor, 2005; Nataraja & Johnson, 2011). These comparisons are based upon Monte Carlo simulations, and the efficiency estimates are compared to the known true frontier.

We take another approach in this paper. The main interest is in the implications that the use of variable selection could have in the interpretation and acceptance of the DEA results for performance management. We use data that have been gathered during the research collaboration with a Finnish retail bank. The variable selection has a specific contextual focus: to specify a model for evaluating the operational efficiencies of the branches during 2007-2010 using DEA. The techniques and their outcomes are reviewed by the branch network management.

We focus on two techniques: VR and ECM. Our prior assumption is that both techniques could be helpful in our empirical case, where the output variables represent the sales quantities of banking services that are highly positively correlated. Both techniques arise from the critique against intuitive variable exclusion, based upon high correlations between variables, and present statistical approaches for variable selection (see Pastor et al. 2002, Jenkins & Anderson, 2003)³.

Both techniques aim to minimize information loss when omitting variables, but they use very different definitions of information loss. While the objective of VR is to minimize the

³ Even though we focus on the outputs, both techniques can be applied to inputs or both. Jenkins and Anderson (2003) present an example where they apply their approach to the entire set when there are high correlations between some of the inputs and outputs.

reduction in data variability, ECM focuses on minimizing distortion in the distribution of efficiency scores. This paper describes what kinds of consequences the different definitions can have in the case context.

The techniques should be considered as tools to help the management when deciding on inputs and outputs for efficiency evaluation. If the purpose is solely the correct classification of efficient and inefficient DMUs in a certain context, and the purpose is not variable selection, there may be other choices to reduce dimensionality.⁴

Even though the industry is less relevant in variable selection for DEA, it is worth noting that VR has been associated as an option to improve discrimination in bank and bank branch efficiency evaluation (Paradi, Yang, & Zhu, 2011; Paradi & Zhu, 2013). The reader may also be interested in the application of ECM in the evaluation of a Spanish bank branch network (Pastor, Lovell, & Tulkens, 2006).

The rest of this paper is organized as follows. Section 2 presents the two selected techniques used in this study. Section 3 introduces the empirical case, including the data used and the initial estimation without variable reduction. Section 4 describes how the techniques were applied to construct alternative specifications for efficiency estimation. Section 5 discusses the key findings. Section 6 summarizes the management feedback to the comparison. Section 7 makes concluding remarks.

2. Variable selection techniques

First, we briefly introduce a DEA estimator that provides the baseline for this study, and then we explain both techniques applied for variable selection in this comparison.

2.1. DEA estimator

The efficiency of a DMU can be evaluated using the generalized DEA estimator presented by Banker, Charnes, and Cooper (1984). We use the output-oriented multiplier form of DEA here. Assume there are n DMUs, each consuming m inputs and producing p outputs. The

⁴ ECM performed quite well in most tested scenarios when compared with a number of selection and aggregation-based approaches (Nataraja & Johnson, 2011). VR has been compared with a technique that combines principal component analysis (PCA) and DEA (Adler & Yazhemsky, 2010). The latter outperformed VR in the simulations. Even though the combination PCA and DEA reduces dimensionality it is an aggregation approach rather than a tool for guiding the selection of variables.

inputs are denoted by $\mathbf{x} \in \mathbb{R}_+^m$ and outputs by $\mathbf{y} \in \mathbb{R}_+^p$. Let $\mathbf{X} \in \mathbb{R}_+^{m \times n}$ and $\mathbf{Y} \in \mathbb{R}_+^{p \times n}$ be the matrices of observed inputs and outputs for the DMUs. DEA finds the most favorable input weights $\mathbf{v}$ and output weights $\boldsymbol{\mu}$ for the DMU under assessment, denoted by superscript r . Scalar u_o is associated with the returns to scale (RTS) assumption. If constant returns to scale (CRS) are assumed, u_o equals zero, and the estimator is called CCR (Charnes, Cooper, & Rhodes, 1978). If variable returns to scale (VRS) are assumed, u_o is free and the estimator is called BCC (Banker et al., 1984). The generalized DEA estimator is the following:

$$\begin{aligned} \min \quad & \mathbf{v}^T \mathbf{x}^r + u_o \\ \text{s.t.} \quad & \boldsymbol{\mu}^T \mathbf{y}^r = 1 \\ & -\boldsymbol{\mu}^T \mathbf{Y} + \mathbf{v}^T \mathbf{X} + \mathbf{1}^T u_o \geq \mathbf{0}^T \\ & \boldsymbol{\mu}, \mathbf{v} \geq \mathbf{0} \end{aligned} \tag{1}$$

The objective function provides an efficiency score for the DMU, the smaller the better because of the output orientation. If the score is one, and all variable weights $\mathbf{v}$ and $\boldsymbol{\mu}$ are strictly positive, the DMU is efficient.

The first constraint normalizes the sum of weighted outputs to one. This normalization has a convenient side-product, because the weighted value of the output $\mu_i y_i^r$ depicts directly the importance of the particular variable in the efficiency evaluation of the DMU. This weighted value is called the *virtual output*, and it is an important concept in this comparison. The second constraint limits the weights in such a way that the performance of the unit is scaled to the best performance among all units.

2.2. Variable reduction based on partial covariance

As variables in the efficiency evaluation are often highly correlated, an intuitive approach is to omit some of the highly correlated variables without a significant loss of information. Jenkins and Anderson (2003) criticized this approach because interrelations between variables are less obvious and cannot be determined directly from the correlation matrix. They wanted to introduce a systematic statistical procedure to help make decisions for variable selection. The authors have not given any specific name or abbreviation to their procedure. We call it *variable reduction based on partial covariance*, similarly to Alder and Yazhensky (2010).

VR is a multivariate statistical procedure used to find a reduced set of initial input or output variables to be included in DEA. The aim is to minimize the information loss of

original data when removing variables. The indicator of the information contained by a variable is the variance around the mean. The following summary of the procedure is based on Jenkins and Anderson (2003).

The conditional variance of a remaining i variable when the effect of variable i'' is removed is denoted by $\sigma_{i,i''}$. If the variables are perfectly correlated, then the variance remaining is $\sigma_{i,i''}=0$ and i'' can be removed without any loss of information. This conditioning can be extended to the sets of retained and omitted variables.

To illustrate the procedure, we apply it to the output variables. The data are normalized to zero mean and unit variance to treat all outputs equally. Then the sum of variances of the normalized outputs is equal to the number of outputs p .

The outputs are divided into the set of variables omitted $i=1, \dots, k$ and variables retained $i=k+1, \dots, p$, where the variables are in an arbitrary order. The variance covariance matrix of outputs denoted by $\mathbf{V}$ can be partitioned as variance-covariance matrices of omitted variables (denoted by $\mathbf{V}_{11}$) and retained variables ($\mathbf{V}_{22}$), covariance of omitted and retained variables ($\mathbf{V}_{12}$), and vice versa ($\mathbf{V}_{21}$). The partial variance-covariance matrix of omitted variables given the retained variables is computed $\mathbf{V}_{11.2} = \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{21}$. The trace of $\mathbf{V}_{11.2}$ represents the size of the remaining variance of omitted variables after conditioning the retained variables. The sum of trace measures the information loss. The proportion of information retained can be obtained by comparing the information loss to the original sum of variance p .

The test of the partial covariance is repeated for all combinations of retained and remaining variables. The information losses of the combinations are monitored. A reduced variable set with minimal loss of information is chosen, and this variable set is used with the appropriate DEA estimator.

Jenkins and Anderson (2003) do not give exact criteria for choosing the set of initial variables retained. The number of variables is externally defined to avoid the lack of discrimination. Jenkins and Anderson state that the ultimate decision to omit particular variables depends upon the importance of the variables to the management, and that the VR procedure plays a supportive role in this decision.

2.3. Variable selection based upon the efficiency contribution measure

Even though the ECM approach is also a selective technique, it differs from VR in the role of DEA. VR makes the variable selection using a statistical test, without any use of DEA. The ECM approach uses the appropriate DEA estimator to test the marginal contribution of each

variable to the efficiency estimation. Following Pastor et al. (2002), the test procedure is presented for an output-oriented radial efficiency estimation with efficiency scores $[1, \infty)$ where 1 is efficient.

The ECM approach is based on two variable specifications that are called total model and reduced model. The total model contains a set of variables. In the reduced model, one of the input or output variables, called the candidate variable (denoted by z), will be removed. As the first step, the data used in the total model is adjusted in such a way that all the DMUs become efficient. Then the candidate variable is removed and DEA is executed with the reduced model. DMU r 's efficiency score, denoted by ϕ^r , represents the change in the efficiency scores between the total model and the reduced model. ϕ^r is called the *efficiency contribution measure* of variable z to DMU r . If $\phi^r=1$, then we can conclude that the candidate variable z has no influence on the efficiency evaluation of the DMU. Analogously, $\phi^r=1.1$ indicates that the removal of the candidate variable has a 10% impact on the efficiency score of the DMU. These scores are calculated for all DMUs.

The relevance of the candidate variable z for the total model can be tested statistically using a binomial test. The null hypothesis is that the efficiency score distributions of the original model and the reduced model do not differ significantly and the candidate variable could be eliminated. The test has two parameters. The first one is a parameter for tolerated change between the models (denoted by $\bar{\varphi}$). Then T is the number of DMUs that exceed this tolerance. The second parameter is the probability level (denoted by p_0) for the maximum allowed proportion of DMUs exceeding the tolerance. Sirvent et al. (2005) suggest that $\bar{\varphi}=1.1$ and $p_0=0.15$ are a suitable conservative choice for practical purposes. These parameters mean that the proportion of DMUs affected by more than 10% should not exceed 15% at the given significance level. Otherwise, the variable cannot be eliminated.

The use of the ECM is a stepwise process. Pastor et al. (2002) propose both a backward procedure (variable elimination) and a forward procedure (variable incorporation). As the backward procedure is relevant for variable reduction, it will be our focus. First, all variables are tested. The candidate variable with the lowest value of T is eliminated. Then a new original model with the remaining variables is built. The process continues until no variables can be eliminated.

3. The empirical case

The case is an application to evaluate the operational efficiencies of the bank branches in a Finnish retail bank, Helsinki OP Bank Plc⁵. This is a retail bank operating in the Helsinki metropolitan area. It belongs to OP Group, which is a leading financial service provider in Finland.

The variable selection techniques are applied to a set of outputs (services) for operational efficiency estimation, when assuming that all outputs are initially equal. Two alternative output specifications are defined, one using the VR technique and one using the ECM technique. The impacts of the use of these techniques are assessed by the following questions:

- How discriminative are the alternative output specifications in relation to the distribution of efficiency scores and the identification of efficient units?
- What are the implications for the ranking of the units, based upon the efficiency scores of the alternative output specifications?
- What is the importance of each main group of services to the efficiency evaluation when alternative output specifications are used?

Management feedback on the techniques and the results will be also considered.

3.1. Outputs and inputs

The bank branches provide more than 50 banking products that are tracked in the bank's sales system. Many of them are only variants, such as different types of credit cards. The output set can be reduced to ten services by aggregating different product variants. These services, denoted by i , are presented in Table 1. The unit of measurement related to each service is the number of new agreements, such as new housing loans granted, mutual fund investments sold, or current accounts opened, during the period.

The services comprise three main groups: financing, investment, and daily banking. In addition to housing loans and consumer loans, financing services include payment protection insurance. The investment services include savings in an account or investment funds, either directly or as insurance savings accounts. Daily banking services include services that the customers need for payments and other daily affairs. Thus, new agreements for daily services

⁵ As of April 1 2016 the bank's name has been Helsinki Area Cooperative Bank.

reflect well the new customer base. Critical illness cover is actually insurance, but sold by banking professionals. It is a relevant output of the bank branch activity, but it is not significant enough to be a fourth group in this analysis. The bank management decided to locate it under investment services in the efficiency evaluation.

The initial inputs represent the working time of employees (full-time equivalent, FTE) in five different job roles in the sales activity. The bank had standard unit costs for different job roles. These unit costs were used to aggregate the cost of the workforce used as single input. The unit of measurement presented in this paper is working year (FTE), standardized with these unit costs.

Table 1. Initial outputs for the bank branch efficiency evaluation.

Notation (i)	Service	Service group
1	Housing loans	Financing
2	Consumer loans	Financing
3	Payment protection insurance	Financing
4	Mutual funds	Investment
5	Insurance savings	Investment
6	Savings accounts	Investment
7	Critical illness cover	Investment
8	Current accounts	Daily banking
9	Online banking agreements	Daily banking
10	Cards	Daily banking

3.2. Data

This study is based on annual data of the 25 bank branches from the years 2007-2010. As some of the branches were founded during the time frame, there are 95 observations. There are several temporal options for DEA when panel data is used (see Tulkens & Vanden Eeckaut, 1995). We use an intertemporal approach in the variable selection, meaning that all observations specify the model used in the efficiency estimation. If we applied a contemporaneous or sequential approach, each period would have a different specification, which would complicate the comparison of the techniques.⁶ In this study, DMUs are the observations of branch-year combinations.

⁶ Even though we select a single set of variables for all periods, changes over time can also be allowed. Chen and Johnson (2010) demonstrate how a variable selection technique, in their case ECM, can be used to study the change of the set of relevant variables, and bring new insights to the dynamics of the performance.

As the variable selection techniques are based upon statistical analysis, the distributions and correlations are relevant when explaining both the specifications and outcomes of the efficiency estimations. The cost of the work force is denoted by c . The outputs (i.e., services) are denoted by $y_i (i=1, \dots, 10)$. The descriptive statistics are presented in Table 2.

Table 2. Descriptive statistics of variables.

	Mean	Median	Std. Dev.	Skewness	Kurtosis	Min	Max
c Cost of sales force	12.0	10.3	6.3	0.73	-0.61	4.2	27.4
y_1 Housing loans	392.5	353	227.4	0.82	0.14	92	1074
y_2 Consumer loans	286.0	244	179.7	0.95	0.25	52	802
y_3 Payment protection	157.4	153	95.9	0.73	-0.01	33	432
y_4 Mutual funds	684.1	459	568.6	1.67	2.66	100	2781
y_5 Insurance savings	131.9	101	88.6	1.01	0.83	16	458
y_6 Savings accounts	918.4	836	670.7	1.26	1.26	153	3003
y_7 Critical illness cover	91.4	72	71.5	1.07	0.29	12	297
y_8 Current accounts	849.2	728	471.5	0.81	-0.01	277	2291
y_{19} Online banking	622.7	585	339.0	0.67	-0.35	172	1619
y_{10} Cards	896.7	756	528.8	0.69	-0.57	185	2206

$n=95$. For c , the unit of measurement is the working year (FTE). Outputs y_1 - y_{10} represent quantities of transactions.

The distributions of the variables are positively skewed. There are more small and medium-size branches than large branches in the network. The skewness differs by service. One reason for these differences is that in the panel data, the distribution is affected not only by the branch size, but also by the yearly fluctuation of the service.

The correlations between the variables are presented in Table 3. They are relatively high between the outputs and between the outputs and the input. There are, however, some lower correlations even between services within the same group. The lowest correlation is between y_4 mutual funds and y_6 savings accounts. The reason for this is in the customer's behavior during the turbulent times of 2007-2010. During the bull market before the financial crisis, customers sought profits from mutual funds, but as the share prices dropped, customers changed to less risky instruments, such as fixed-term savings accounts.

Table 3. Correlation matrix between variables.

	<i>c</i>	<i>y1</i>	<i>y2</i>	<i>y3</i>	<i>y4</i>	<i>y5</i>	<i>y6</i>	<i>y7</i>	<i>y8</i>	<i>y9</i>	<i>y10</i>
<i>c</i>	1										
<i>y1</i>	0.941	1									
<i>y2</i>	0.781	0.836	1								
<i>y3</i>	0.743	0.799	0.686	1							
<i>y4</i>	0.842	0.777	0.720	0.620	1						
<i>y5</i>	0.833	0.834	0.834	0.710	0.771	1					
<i>y6</i>	0.734	0.742	0.685	0.588	0.433	0.711	1				
<i>y7</i>	0.812	0.761	0.695	0.812	0.682	0.707	0.678	1			
<i>y8</i>	0.846	0.850	0.725	0.825	0.593	0.707	0.799	0.853	1		
<i>y9</i>	0.790	0.811	0.780	0.784	0.528	0.724	0.852	0.825	0.958	1	
<i>y10</i>	0.766	0.786	0.885	0.733	0.577	0.758	0.814	0.788	0.892	0.954	1

All correlations are significant at the 0.001 level.

3.3. Initial sales efficiency estimation

The baseline estimation utilized a sales efficiency model that consists of all ten outputs. We do not utilize the logical grouping of services. The branches' intent was to maximize the sales of banking products with their resources, rather than minimize the resources to a given demand. Thus, the output-oriented measurement of efficiency will be used.

We apply a radial output-oriented DEA estimator (1). Additive DEA estimators, based on the difference between weighted outputs and weighted inputs, are sometimes used in bank branch efficiency analysis. They allow negative input/output values, which we do not have⁷. The radial DEA is more appropriate for our purposes to rank the units. The efficiencies will be estimated using both CRS and VRS assumptions, and a test will be used to determine the appropriateness of the CRS assumption.

When applying the initial sales efficiency estimation with ten outputs, 17 units (18% of all units) are found to be efficient when CRS is assumed and 27 units (28%) when VRS is assumed. The average efficiency scores are 0.82 (CRS) and 0.87 (VRS). As this is a real

⁷ See, for example, Portela and Thanassoulis (2007), where the outputs include the change in the number of clients and differences in the monetary values of various financial services. These changes can obviously also be negative.

application, we cannot obviously say which of the units are actually correctly or incorrectly identified as efficient or inefficient.

As the purpose is to give help to managers in specifying the output set, we look at the initial estimation from their perspective. If they are satisfied with the outcome of the initial estimation, they would accept a high proportion of efficient units compared to a reduced variable set. When used in target-setting, less improvement would be required from the branch network. Some of the improvement potential may be lost, and there would be a risk of picking poor benchmark DMUs from which the others have little to learn. They would also accept that a unit can be determined to be efficient based upon high performance in one output, even if the larger number of other outputs would have significant space for improvement.

If the discrimination is increased, the benchmarks are fewer and they represent more likely good practices, but on the other hand some relevant benchmarks may be ignored. The targets based upon an efficient frontier tend to be more of a stretch. If the local management and employees consider the targets impossible to achieve, these targets may be potentially demotivating.

In our case, the management considered the variable selection to be justified.

4. Applying the variable selection techniques

This section presents how the reduced output specifications were constructed using the variable selection techniques.

4.1. Applying VR

In this study, VR was applied to select one output from each service category. Thus, the number of retained outputs was three. Appendix A presents the ranking of the 36 combinations meeting this criterion.

We computed the matrices of partial variance-covariance for each possible combination of retained/omitted variables and sorted the combinations according to the variance retained, as proposed by Jenkins and Anderson (2003).

The first ranked combination (y_3 Payment protection insurance, y_4 Mutual funds, and y_9 Online banking agreements) was selected for use in the comparison. We call it the *VR*

specification in this comparison. The differences between the top-ranked combinations were rather small. We will discuss the sensitivity of this selection later, in 5.3.

4.2. Applying the ECM

We used the backward procedure of the ECM to delete superfluous variables from the initial variable set. The DEA efficiency scores used as the efficiency contribution measure are potentially affected by the RTS specification. As we do not have a predefined RTS specification, the procedure was executed both for CCR and BCC estimators separately. Thus, it could be possible that the procedure ends up with a different reduced variable set.

As Sirvent et al. (2005) suggest, parameters $\bar{\varphi}=1.1$ and $p_0=0.15$ were used when eliminating outputs in the stepwise process. Binomial distribution $B(94, 0.85)$ was used with a 5% level of statistical significance, to determine if the output can be eliminated⁸. However, in some steps, two or more variables were tied for the highest effect with the initial tolerance level. In such cases, the tolerance was decreased ($\bar{\varphi}=1.05$, $\bar{\varphi}=1.03$ etc.) until the variable to be removed was identified.

The stepwise variable elimination process is presented in Appendix B for both CRS and VRS assumptions. Even though there were some differences between the intermediate steps, both assumptions ended up with the same three variables, after which no further variable eliminations were possible. When following the ECM procedure, the ideal combination for evaluating efficiency is variables y_1 Housing loans, y_4 Mutual funds, and y_8 Current accounts. The rest of the outputs can be considered superfluous. We call the combination of these three variables the *ECM specification* in this comparison.

This combination of y_1 , y_4 , and y_8 also meets the criterion to select one variable from each service group, similarly to our application of the VR procedure. It retains 83.53% of the total variance of the initial output. It is, however, only the 21th of the variable alternatives according to the VR procedure (see Appendix A).

At this stage, we can conclude that y_4 Mutual funds is a statistically highly influential service from the performance perspective, as it is included in the reduced variable set in both the VR and ECM variable sets.

⁸ Binomial distribution $B(94, 0.85)$ was used because there are only $n-1$ DMUs that can be affected. At least one of the DMUs remains efficient in the reduced model.

5. Findings

The efficiencies of the DMUs were estimated by employing both CRS and VRS assumptions. The descriptive statistics of the efficiency scores and the number of efficient units are presented in Table 4. The scores are presented in a form where scores below one represent inefficiency. High average, median, and minimum efficiency scores and low standard deviation are also indicators of possible overestimation of efficiencies.

Table 4. Descriptive statistics of efficiency scores and number of efficient DMUs under CRS and VRS assumptions.

	Number of efficient DMUs	Efficiency scores			
		Mean	Median	Std.Dev.	Min
Constant returns to scale (CCR)					
Initial estimation	17	0.82	0.82	0.13	0.57
VR specification	4	0.69	0.68	0.15	0.41
ECM specification	3	0.76	0.75	0.12	0.56
Variable returns to scale (BCC)					
Initial estimation	27	0.87	0.91	0.12	0.65
VR specification	16	0.77	0.76	0.16	0.42
ECM specification	13	0.82	0.82	0.12	0.57

n=95

The initial DEA estimation, which contains all ten variables, is, by default, the least discriminative specification. If VRS was assumed, 28% of the DMUs would be considered efficient. The discrimination can clearly be improved by adopting either of the selective techniques. The scores of the VR specification are, on average, lower than the scores of the ECM specification. On the other hand, the ECM specification finds less DMUs efficient than the VR specification.

In order to examine which RTS assumption should be used in the model, the non-parametric Kolmogorov-Smirnov (KS) test (Banker, 1993; Banker & Natarajan, 2011) is used. The test compares CCR and BCC efficiency score distributions. The null hypothesis is that the frontier exhibits CRS and the alternative hypothesis is VRS. The results of the KS test are presented in Table 5. The initial estimation retains the null hypothesis of CRS if the 5% statistical significance level is required. Both the VR and ECM specifications reject CRS for VRS. The statistical techniques do not only increase discrimination, but they may increase the difference between the scores of RTS assumptions, as less information is used.

Table 5. Kolmogorov-Smirnov test of the constant returns to scale specification

	Most extreme differences (D)	p-value
Initial estimation	0.179	0.095
VR	0.242	0.008
ECM	0.221	0.019

Critical D (5% sign.) is 0.197.

Based upon the RTS test, we assume VRS in the remaining part of this paper.

5.1. Ranking differences

Some significant differences occur in the rankings of individual DMUs between the output specifications selected by the techniques. Figure 1 plots the ranking of DMUs. The figure illustrates that the evaluations based upon the two variable sets identified by different techniques are related. The point in the lower left corner (ranking 1 with both the ECM and VR specifications) includes 11 observations that are identified as efficient using both techniques. There are, however, many significant differences.

The biggest difference in rankings was related to branch B23, which, according to the ECM specification, is efficient but the VR specification ranks it as 54th with an efficiency score 0.75. This is illustrated in Figure 1. The reason for the high ECM-based score was the high housing loan sales when the market peaked in 2008. However, the branch was not able to utilize that opportunity in the sales of payment protection insurance to the housing loan customers, which would have provided high scores when the VR specification was used, too. There are several similar examples, where the excellent performance of a branch becomes mediocre using the other technique – or a mediocre performance becomes poor.

Figure 1. Rankings of the DMUs by VR and ECM specifications.

The rank correlations of the DMUs are presented in Table 6. The rank correlation also indicates that even though the rankings based upon the two techniques are significantly correlated, they draw a different picture of the relative performance.

Table 6. Rank correlations between DMUs, VRS (Spearman's rank correlation coefficient).

	Initial estimation	VR	ECM
Initial estimation	1		
VR specification	0.796	1	
ECM specification	0.830	0.765	1

All correlations are significant at the 0.001 level.

The differences between the ECM and VR emerge from the different aims in the variable selection. VR focuses on the variance between the variables, and thus emphasizes the differences in outputs between the units. On the other hand, the ECM tends to retain variables that cause significant changes in the efficiency scores of the units. These variables that the ECM retains are not necessarily the ones with big differences between the DMUs. For example, y_1 Housing loans is a core service, in which most branches perform relatively well,

thus a large number of DMUs operate close to the frontier and y_I cannot be deleted without a significant change in the efficiency scores of a large proportion of DMUs.

As different techniques take different approaches to dealing with the dimensionality issue, some variations in the results are natural.

5.2. Importance of outputs in efficiency evaluations

The analysis of virtual outputs proposed by Boussofiane, Dyson, and Thanassoulis (1991) can be insightful. Virtual outputs depict the importance of a particular output for the unit under assessment, for the maximum efficiency rating. Virtual outputs can be utilized to identify units with good practices that could be disseminated. Table 7 presents the descriptive statistics of virtual outputs in the initial estimation, and the two alternative specifications.

Because a lot of attention is paid to the service groups in the case bank, a question arises about how important each main service group is to the efficiency evaluation. We use virtual outputs to provide management with information on this matter. We compare the expected virtual outputs and the mean of the actual virtual outputs on the service group level in the set of all DMUs. Even though the weights are free for an individual unit, in a service business like retail banking, it may not be desirable for the overall efficiency evaluation to be too strongly ruled by one group of services.

Table 7. Descriptive statistics of virtual outputs, all DMUs.

	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
Initial estimation										
Mean	0.221	0.0533	0.104	0.158	0.0455	0.0811	0.1122	0.128	0.0588	0.0400
Std.Dev.	0.284	0.128	0.225	0.199	0.091	0.187	0.193	0.245	0.156	0.113
Min	0	0	0	0	0	0	0	0	0	0
Max	1	0.762	0.937	0.922	0.444	0.921	0.839	0.833	0.834	0.674
VR specification										
Mean	-	-	0.197	0.278		-	-	-	0.525	-
Std.Dev.	-	-	0.258	0.258		-	-	-	0.287	-
Min	-	-	0	0		-	-	-	0	-
Max	-	-	1	1		-	-	-	1	-
ECM specification										
Mean	0.3699	-	-	0.268	-	-	-	0.3623	-	-
Std.Dev.	0.379	-	-	0.247	-	-	-	0.371	-	-
Min	0	-	-	0	-	-	-	0	-	-
Max	1	-	-	1	-	-	-	1	-	-

The weights of the efficient DMUs are not unique in DEA and may vary by solver used.

Table 8 presents the mean virtual outputs of each service group by alternative. There are clear differences between alternatives in how the service groups are weighted. The virtual outputs of the ECM specification are most equally balanced between the service groups. If balanced weighting of the service groups is wanted, the initial estimation underweights daily banking services. The VR specification underweights financing services and overweights daily services.

Table 8. Mean virtual outputs by service group, all DMUs.

	Financing services	Investment services	Daily banking services	t² test sta- tistics	p-value
Initial estimation	38%	40%	23%	7.511	0.028
VR specification	20%	28%	53%	46.021	0.000
ECM specification	37%	27%	36%	6.706	0.041

The test is Hotelling's t-squared statistics.⁹ The test is indicative only.

The importance of the variables among all DMUs depends on how far the observations generally are from the frontier. The distributional anomalies affect the mean virtual outputs. For example, positive skewness indicates that the majority of the variable's observations are far away from the tails defining the frontier. As the skewness differs by variable, it is generally more favorable to weight variables whose observations tend to be closer to the frontier. The VR procedure is especially sensitive to the distributional differences, as it is based upon the few retained variables selected prior to DEA¹⁰. The low sum of the virtual output of financing services indicates that a high performance in the representative variable, payment protection insurance, is rare. In addition, the ECM procedure is potentially affected by extreme performances in single service. However, the ECM procedure tends to delete outputs that have the vast majority of observations far away from the frontier, as it is favorable for the DMUs not to weight them.

⁹ The null hypothesis is that the vector of mean virtual outputs of the groups does not deviate from the expected mean vector. The expected mean is approximately 33% (1/3 of each service group) when VR and ECM specifications are used. When an initial specification is used, the expected mean for financing services and daily services is 30% (3/10), and 40% (4/10) for investment services. These tests are only very indicative, because the virtual outputs of service groups depend on each other.

¹⁰ VR assumes implicitly that the variable distributions are normal when using partial covariance to select variables. The issue of normality requirement for VR in empirical cases has been brought up previously by Adler and Yazhemsky (2010).

An interesting comparison from the managerial perspective is how the virtual outputs reflect the cost of the sales force. As the initial input data contained the work of the sales force by job role and the standard unit costs, we can make an indicative split of the costs of the sales force to the service groups. The cost of the team leaders is allocated to the service groups by the FTEs of the direct sales force. An indicative split of the cost of the sales force is: financing services 41%, investment services 30%, and daily banking services 28%. The VR specification violates this split clearly, while the ECM specification weights are quite well aligned with the resource allocation of the branch network.

The results show that the apparent equality of the variables may lead, in practice, to unexpected differences in the importance of the variables and their logical groups in the efficiency evaluation. This may be against the implicit knowledge of management. Thus, it is worth studying the virtual outputs (and inputs) generated through the variable selection and discussing the findings with management.

5.3. Sensitivity of the variable selection

Jenkins and Anderson (2003) have pointed out that the results of omitting or including variables in DEA are difficult to predict, even if the variables are highly correlated. The examples that the authors present show that one should be cautious and study the differences between the candidate variable sets. We checked the extent to which the top four ranked output combinations (see Appendix A) provide different results. These combinations are compared in Table 9.

Table 9. Comparison of four top output combinations according to the VR procedure

Output combination	Number of efficient DMUs				Spearman' s rank correlation with	
		Average	Median	Std. Dev	Min	1st selection
1st y_3, y_4, y_9	16	0.77	0.76	0.16	0.42	1
2nd y_2, y_4, y_9	12	0.75	0.75	0.15	0.42	0.889
3rd y_2, y_4, y_8	13	0.79	0.77	0.13	0.46	0.850
4th y_3, y_4, y_{10}	13	0.74	0.72	0.18	0.38	0.926

All rank correlations are significant at the 0.001 level.

Interestingly, the 1st output combination that we have used as the VR specification in the comparison does not provide the highest discrimination. It identifies more efficient DMUs than the other three selections. The 4th output combination has both lower average scores and

higher variance between the scores. However, the rank correlations indicate that the four options give fairly similar rankings to DMUs. The rank correlations of the 1st, 2nd, and 4th output combinations with the ECM specification are very similar (rank correlations 0.709-0.764), but the rankings of the 3rd combination are closer to the ECM rankings (rank correlation 0.877). Thus, we can conclude that the 2nd and 4th output combinations take a fairly similar perspective of efficiency to the 1st output combination, and the change of one variable does not have a significant impact in our case.

As described earlier in 4.2, the results of the ECM technique could potentially be impacted by the RTS assumption. However, both RTS assumptions converged to the same final variable sets.

The forward selection procedure of the ECM was also tested. The single variable y_4 Mutual funds was considered to be essential, and other variables were tested and added to the model in a stepwise process following the steps described by Pastor et al. (2002). The forward procedure also converged to the same the variable set y_1 , y_4 , and y_8 . Thus, in our case, the ECM technique ends up with a single variable set regardless of the RTS assumption or the form of procedure.

6. Management feedback

The management of the branch network reviewed the main results of the initial efficiency estimation and the alternative selection techniques. The managers were provided with a graphical presentation of the efficiency scores of branches. Because each branch had several yearly observations from 2007-2010, averages of efficiency scores were used. The branch-level results were more convenient for the managers. This, however, reduced the visibility of some extreme observed performances and yearly performance fluctuation. The management also reviewed the virtual outputs of the service groups and compared them to the costs of different job roles that were attached to these groups. In addition, the fundamentals of each technique were discussed with the management.

The management stated that even though it preferred the approach of evaluating efficiency using an aggregation of variables based upon the revenue value of services, an operational efficiency evaluation could be a supplementary analysis¹¹. If a wide data set is gathered, then it would attempt to use all that information. However, the management found it

¹¹ Eskelinen and Kuosmanen (2013) and Eskelinen, Halme, and Kallio (2014) present applications using the revenue value-based variable aggregation for DEA at the case bank.

hard to accept that a branch could claim itself to be efficient only on the basis of any individual service that the bank provides, as would be possible with the initial variable set. Thus, variable selection is justified.

The managers commented that the principle of using the output set with the largest overall differences between the observations was reasonable (i.e., variance retained). In addition, the set of variables selected by VR was justified, as the management had put emphasis on these services during the years under evaluation. However, the disadvantage in this case was that the results were impacted by the strongly overweighted daily services.

The principle of eliminating the least significant outputs according to their contribution to efficiency was considered attractive by the management. The output specification obtained from the ECM approach contained the core services with which the branches are constantly engaged. The virtual outputs of the services group were also balanced and aligned with the costs. Thus, the management favored the ECM over VR if only one reduced set of variables is to be selected.

7. Concluding remarks

It has been well known that the DEA results are sensitive to variable selection. Even in a case where the variables are highly correlated, the choice of the variables can affect the results of the efficiency evaluation significantly (Dyson et al., 2001). This study has demonstrated that the implications may concern not only the efficiency scores and rankings of individual DMUs, but also the importance that the variables and their logical groups have in the efficiency evaluation. Thus, it may affect the conclusions drawn from the results regarding the entire branch network.

On the other hand, this case study also suggests that some contradictions are natural. The techniques compared represented two strategies of variable selection. VR emphasized the combination of services with the largest performance differences between the units, while the ECM ended up with a set of core outputs in which a high number of units perform relatively well. The management found that the techniques resulted in different interpretations of the efficiency measurement.

The fundamental differences between the selection approaches can benefit decision-making in companies and public organizations. Cooper, Seiford, and Tone (2007, p. 166) state that: "If we cannot identify the characteristics of the production frontiers by preliminary surveys, it may be risky to rely on only one particular model." When the results from different

selection techniques are uniform, they confirm the basis for decisions. Contradictory results regarding an important issue indicate uncertainty, and further study may be needed before the decision.

Both techniques have benefits and drawbacks. VR has been criticized because variance retained has little value to managers (Pastor et al., 2006). In our case, the management appreciated the principle of VR, and the reduced variable set made sense to them. Thus, it might be insightful to consider which set of outputs contains the biggest differences, and to discuss whether these differences are relevant from the performance improvement point of view. The unpredictability of VR has been highlighted by Jenkins and Anderson (2003). The results are sensitive to extreme performances, skewness, and also noise, which Adler and Yazhemsky (2010) have recognized in their experiments. In our case, the sensitivity to anomalies caused an unbalanced weighting of services, contradicting the prior knowledge of management.

The ECM's criterion for variable selection, contribution to efficiency measurement, was straightforward to the management and thus it was easily acceptable in our case. As the technique avoids elimination with a large overall impact on the efficiency scores, it seems to emphasize existing strengths in the network, rather than areas where the units have the biggest opportunities to stretch their capabilities.

This study complements the previous simulation-based comparative studies of variable selection by employing an empirical case. The author hopes that this article can also be helpful to the practitioners of DEA-based performance management.

Acknowledgments

The author would like to thank Jussi Huttunen and Marko Helin from Helsinki Area Cooperative Bank for the co-operation that has made this paper possible. He extends his thanks to Pekka Korhonen (Aalto University), C. A. Knox Lovell (University of Queensland), Maria C. A. S. Portela (Portuguese Catholic University), Andy L. Johnson (Texas A&M), and two anonymous referees for their suggestions for improving this paper.

References

- Adler, N., & Golany, B. (2001). Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe. *European Journal of Operational Research*, 132, 260-273.
- Adler, N., & Golany, B. (2002). Including principal component weights to improve discrimination in data envelopment analysis. *The Journal of the Operational Research Society*, 53, 985-991.
- Adler, N., & Yazhemsky, E. (2010). Improving discrimination in data envelopment analysis: PCA-DEA or variable reduction. *European Journal of Operational Research*, 202, 273-284.
- Amirteimoori, A., Despotis, D. K., & Kordrostami, S. (2014). Variable reduction in data envelopment analysis. *Optimization*, 63, 735-745.
- Banker, R. D. (1993). Maximum likelihood, consistency and data envelopment analysis: a statistical foundation. *Management Science*, 39, 1265-1273.
- Banker, R. D. (1996). Hypothesis test using data envelopment analysis. *Journal of Productivity Analysis*, 7, 139-159.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in Data Envelopment Analysis. *Management Science*, 30, 1078-1092.
- Banker, R. D., & Natarajan, R. (2011). Statistical tests based on DEA efficiency scores. In Cooper, W. W., Seiford, L., & Zhu, J. (Eds.), *Handbook on data envelopment analysis* (2nd ed., pp. 273-295). New York, Dordrecht, Heidelberg, London: Springer.
- Bian, Y. (2012). A Gram-Schmidt process based approach for improving DEA discrimination in the presence of large dimensionality of data set. *Expert systems with applications* 39, 685-705.
- Boussofiane, A., Dyson, R. G., & Thanassoulis, E. (1991). Applied data envelopment analysis. *European Journal of Operational Research*, 52, 1-15.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2, 429-444.
- Chen, W.-C., & Johnson, A.L. (2010). "The Dynamics of Performance Space of Major League Baseball Pitchers 1876-2006." *Annals of Operational Research*, 181, 287-302.

- Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis - A comprehensive text with models, applications, references and DEA-Solver Software* (2nd ed.). New York: Springer.
- Dyson, R. G., Allen R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, 245-259.
- Eskelinen, J., Halme, M., & Kallio, M. (2014). Bank branch sales evaluation using extended value efficiency analysis. *European Journal of Operational Research*, 232, 654-663
- Eskelinen, J., & Kuosmanen, T. (2013). Intertemporal efficiency analysis of sales teams of a bank: Stochastic semi-nonparametric approach. *Journal of Banking & Finance*, 37, 5163-5175
- Fanchon, P. (2003). Variable selection for dynamic measures efficiency in the computer industry. *International Advances in Economic Research*, 9, 175-188.
- Friedman, L., & Sinuany-Stern, Z. (1998). Combining ranking scales and selecting variables in the DEA context: The case of industrial branches. *Computers & Operations Research*, 25, 781-791.
- Gonzalez-Bravo, M. I. (2007). Prior-Ratio-Analysis procedure to improve data envelopment analysis for performance measurement. *Journal of the Operational Research Society*, 58, 1214-1222.
- Jenkins, L., & Anderson, M. (2003). A multivariate statistical approach to reducing the number of variables in data envelopment analysis. *European Journal of Operational Research*, 147, 51-61.
- Lewin, A. Y., Morey, R. C., & Cook, T. J. (1982). Evaluating the administrative efficiency of courts. *Omega*, 10, 401-411.
- Nataraja, N. R., & Johnson, A. L. (2011). Guidelines for using variable selection techniques in data envelopment analysis. *European Journal of Operational Research*, 215, 662-669.
- Paradi, J.C., Yang, Z., & Zhu, H. (2011). Assessing bank and bank branch performance - Modeling considerations and approaches. In Cooper, W. W., Seiford, L. M., & Zhu, J. (Eds.), *Handbook of data envelopment analysis* (2nd ed., pp. 315-362). New York, Dordrecht, Heidelberg, London: Springer.
- Paradi, J. C., & Zhu, H. (2013). A survey on bank branch efficiency and performance research with data envelopment analysis. *Omega*, 41, 61-79.

- Pastor, J. T., Lovell, C. A. K., & Tulkens, H. (2006). Evaluating the financial performance of bank branches, *Annals of Operations Research*, 145, 321-337.
- Pastor, J. T., Ruiz J. L., & Sirvent, I. (2002). A statistical test for nested radial DEA models. *Operations Research*, 50, 728-735.
- Podinovski, V. V., & Thanassoulis, E. (2007). Improving discrimination in data envelopment analysis: some practical suggestions. *Journal of Productivity Analysis*, 28, 117-126.
- Portela, M. C. A. S., & Thanassoulis, E. (2007). Comparative efficiency analysis of Portuguese bank branches. *European Journal of Operational Research*, 177, 1275-1288.
- Ruggiero, J. (2005). Impact assessment of input omission on DEA. *International Journal of Information Technology & Decision Making*, 4, 359-368.
- Sengupta, J.K. (1990). Test of Efficiency in data envelopment analysis. *Computers and Operations Research* 17(2), 121-132.
- Sharma, M. J. & Yu, S. J. (2015). Stepwise regression data envelopment analysis for variable reduction, *Applied Mathematics and Computation*, 253, 126–134.
- Simar, L., & Wilson, P. W. (2001). Testing restrictions in nonparametric efficiency models, *Communications in Statistics - Simulation and Computation*, 30, 159-184.
- Simar, L., & Wilson, P. W. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In Fried, A. O., Lovell, C. A. K., & Schmidt, S. S., (Eds.), *The measurement of productive efficiency* (pp.421-521). New York: Oxford University Press.
- Sinuany-Stern, Z., Friedman, L. (1998). DEA and the discriminant analysis of ratios for ranking units. *European Journal of Operational Research* 111, 470-478.
- Sirvent, I., Ruiz, J. L., Borrás, F., & Pastor, J. T. (2005). A Monte Carlo evaluation of several tests for the selection of variables in DEA models. *International Journal of Information Technology & Decision Making*, 4, 325-343.
- Tulkens, H., & Vanden Eeckaut, P. (1995). Non-parametric efficiency, progress and regress measures for panel data: Methodological aspects. *European Journal of Operational Research*, 80, 474-499.
- Ueda, T., & Hoshiai, Y. (1997). Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs. *Journal of the Operations Research Society of Japan*, 40, 466-478.

Wagner, J. M., & Shimshak, D. G. (2007). Stepwise selection of variables in data envelopment analysis: Procedures and managerial perspectives. *European Journal of Operational Research*, 180, 57-67.

Appendix A: Proportion of variance retained according to VR

Table A.1 presents the three output combinations between service groups. The combinations have been ranked by the proportion of variance retained, according to the variable reduction procedure of Jenkins and Anderson (2003).

Table A.1. Proportion of variance retained, three output combinations

Ranking of combination	Combinations of three output variables			Proportion of variance retained %
1	y ₃	y ₄	y ₉	87.10%
2	y ₂	y ₄	y ₉	86.98%
3	y ₂	y ₄	y ₈	86.92%
4	y ₃	y ₄	y ₁₀	86.35%
5	y ₃	y ₅	y ₉	85.83%
6	y ₁	y ₄	y ₉	85.78%
7	y ₁	y ₄	y ₁₀	85.14%
8	y ₂	y ₇	y ₉	84.92%
9	y ₁	y ₅	y ₉	84.88%
10	y ₁	y ₇	y ₉	84.83%
11	y ₃	y ₅	y ₁₀	84.70%
12	y ₂	y ₅	y ₈	84.61%
13	y ₂	y ₆	y ₈	84.53%
14	y ₁	y ₇	y ₁₀	84.46%
15	y ₂	y ₅	y ₉	84.31%
16	y ₃	y ₅	y ₈	84.24%
17	y ₂	y ₇	y ₈	83.90%
18	y ₁	y ₆	y ₉	83.87%
19	y ₁	y ₆	y ₁₀	83.77%
20	y ₃	y ₄	y ₈	83.70%
21	y ₁	y ₄	y ₈	83.53%
22	y ₁	y ₅	y ₁₀	83.52%
23	y ₁	y ₅	y ₈	83.43%
24	y ₂	y ₇	y ₁₀	83.43%
25	y ₂	y ₄	y ₁₀	83.38%
26	y ₁	y ₆	y ₈	83.16%
27	y ₂	y ₆	y ₉	82.54%
28	y ₁	y ₇	y ₈	82.39%
29	y ₃	y ₆	y ₁₀	81.85%
30	y ₂	y ₅	y ₁₀	81.49%
31	y ₃	y ₇	y ₁₀	80.96%
32	y ₃	y ₇	y ₉	80.20%
33	y ₃	y ₆	y ₉	80.04%
34	y ₂	y ₆	y ₁₀	79.68%
35	y ₃	y ₆	y ₈	79.40%
36	y ₃	y ₇	y ₈	77.29%

Appendix B: ECM-based backward elimination procedure by step

Table B.1 presents the steps of the ECM-based backward elimination of variables for both the constant and variable returns to scale assumptions. The critical value for binomial test $B(94,0.85)$ at the 5% statistical significance level is 19. The null hypothesis states that the distributions of the efficiency scores in the original model (i.e., the model from the previous step) and the reduced model are statistically indistinguishable. Thus, if $T > 19$ at tolerance level $\bar{\varphi} = 1.1$, the null hypothesis is rejected, and the output cannot be eliminated. If there are several outputs that could be removed, the output with the lowest T is eliminated. If necessary, the tolerance level is reduced in order to identify the variable to be eliminated.

Table B.1. ECM-based backward elimination procedure by step

<i>T</i> , Constant returns to scale			<i>T</i> , Variable returns to scale					
	$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05		$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05			
Step 1: All 10 outputs								
delete y_1	18	28	delete y_1	10	18			
delete y_2	0	0	delete y_2	0	1			
delete y_3	1	6	delete y_3	0	6			
delete y_4	11	26	delete y_4	14	27			
delete y_5	0	2	delete y_5	0	3			
delete y_6	1	2	delete y_6	0	2			
delete y_7	6	18	delete y_7	4	13			
delete y_8	7	12	delete y_8	3	9			
delete y_9	0	1	delete y_9	0	1			
delete y_{10}	0	2	delete y_{10}	0	0			
Reduce to 9 outputs (eliminate y_2)			Reduce to 9 outputs (eliminate y_{10})					
	$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05		$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05			
Step 2: 9 outputs								
delete y_1	20	30	delete y_1	10	18			
delete y_3	1	7	delete y_2	0	3			
delete y_4	13	28	delete y_3	0	6			
delete y_5	1	2	delete y_4	16	27			
delete y_6	1	4	delete y_5	0	3			
delete y_7	6	18	delete y_6	0	4			
delete y_8	7	12	delete y_7	5	14			
delete y_9	0	1	delete y_8	3	9			
delete y_{10}	0	4	delete y_9	0	1			
Reduce to 8 outputs (eliminate y_9)			Reduce to 8 outputs (eliminate y_9)					
	$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05		$\bar{\varphi}$ =1.1	$\bar{\varphi}$ =1.05	$\bar{\varphi}$ =1.03	$\bar{\varphi}$ =1.01	
Step 3: 8 outputs								
delete y_1	20	31	delete y_1	10	18	25	35	
delete y_3	1	5	delete y_2	0	3	3	8	
delete y_4	13	28	delete y_3	0	6	8	13	
delete y_5	1	2	delete y_4	16	27	34	42	
delete y_6	1	4	delete y_5	0	3	3	6	
delete y_7	6	17	delete y_6	3	7	9	14	
delete y_8	9	14	delete y_7	6	14	17	23	
delete y_9	2	1	delete y_8	8	13	15	19	
Reduce to 7 outputs (eliminate y_5)			Reduce to 7 outputs (eliminate y_5)					

<i>T</i> , Constant returns to scale	<i>T</i> , Variable returns to scale	
$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.05$
Step 4: 7 outputs		
delete y_1 19	delete y_1 11	20
delete y_3 3	delete y_2 1	4
delete y_4 24	delete y_3 1	6
delete y_6 4	delete y_4 17	30
delete y_7 6	delete y_6 6	9
delete y_8 10	delete y_7 7	16
delete y_9 1	delete y_8 9	13
Reduce to 6 outputs (eliminate y_9)	Reduce to 6 outputs (eliminate y_2)	
$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.1$	
Step 5: 6 outputs		
delete y_1 22	delete y_1 15	
delete y_3 2	delete y_3 2	
delete y_4 32	delete y_4 24	
delete y_6 6	delete y_6 6	
delete y_7 6	delete y_7 6	
delete y_8 15	delete y_8 10	
Reduce to 5 outputs (eliminate y_3)	Reduce to 5 outputs (eliminate y_3)	
$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.1$	
Step 6: 5 outputs		
delete y_1 21	delete y_1 22	
delete y_4 36	delete y_4 24	
delete y_6 6	delete y_6 5	
delete y_7 5	delete y_7 7	
delete y_8 18	delete y_8 12	
Reduce to 4 outputs (eliminate y_7)	Reduce to 4 outputs (eliminate y_6)	
$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.1$	
Step 7: 4 outputs		
delete y_1 27	delete y_1 21	
delete y_4 31	delete y_4 23	
delete y_6 9	delete y_7 10	
delete y_8 26	delete y_8 16	
Reduce to 3 outputs (eliminate y_6)	Reduce to 3 outputs (eliminate y_7)	
$\bar{\varphi} = 1.1$	$\bar{\varphi} = 1.1$	
Step 8: 3 outputs		
delete y_1 25	delete y_1 25	
delete y_4 45	delete y_4 38	
delete y_8 36	delete y_8 26	
End up with 3 outputs	End up with 3 outputs	