European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Decision Support

Subsampling bootstrap in network DEA

Maria Michali[a], Ali Emrouznejad[b,*], Akram Dehnokhalaji[a], Ben Clegg[a]

[a] Operations and Information Management Department, Aston Business School, Aston University, Birmingham, UK
[b] Surrey Business School, University of Surrey, Guildford, UK

ARTICLE INFO

Article history:
Received 8 November 2021
Accepted 9 June 2022
Available online xxx

Keywords:
Data envelopment analysis
Network efficiency decomposition
Subsampling bootstrap
Monte carlo
Railways

ABSTRACT

Data Envelopment Analysis (DEA), provides an empirical estimation of the production frontier, based on an observed sample of decision making units (DMUs). Except for the single input-single output case, the asymptotic distribution of the DEA estimator can only be approximated through bootstrapping approaches. Therefore, bootstrapping techniques have been widely applied in the DEA literature to make statistical inference for the cases when the production process has a single-stage structure. However, in many cases, the transformation of inputs into outputs has an inner structure that needs to be considered. This paper examines the applicability of the subsampling bootstrap procedure in the approximation of the asymptotic distribution of the DEA estimator when the production process has a network structure, and in the presence of undesirable factors. Evidence on the performance of subsampling bootstrap is obtained through Monte Carlo experiments for the case of two-stage series structures, where overall and stage efficiency estimates are calculated using the additive decomposition approach. Results indicate great sensitivity both to the sample and subsample size, as well as to the data generating process. Subsampling methodology is then applied to construct confidence interval estimates for the overall and stage efficiency scores of railways in 22 European countries, where the railway transport process is decomposed into two stages and the railway noise pollution problem is considered as an undesirable output.

1. Introduction

Data envelopment analysis (DEA) is a non-parametric technique based on linear programming, which is used to assess the relative performance of a homogeneous set of decision making units (DMUs). Based on the seminal work of Farrell (1957), Charnes, Cooper, & Rhodes (1978) and Banker, Charnes, & Cooper (1984) introduced the two basic DEA models under the assumptions of the constant returns to scale (CRS) and the variable returns to scale (VRS) case, respectively.

In DEA, the efficiency of a DMU is measured as its distance from an empirically constructed efficient boundary, and therefore, its efficiency score depends on the available set of DMUs that shape the frontier. Assuming that the available set of DMUs is a sample generated from a population, the true efficient frontier is unknown, and statistical inference methods can be used to provide estimations for the true efficiency scores. Banker (1993) was among the first researchers to consider DEA as a consistent, but biased maximum likelihood estimator (MLE) of the efficient frontier. Since then, many studies have focused on deriving the convergence rate of DEA estimator and its asymptotic distribution. Several bootstrapping techniques have been developed and used to obtain the sampling distribution of DEA estimators, as in multi-dimensional settings, an analytical form is not possible to be derived.

Until now, studies on making statistical inference about DEA have been limited to production processes with one-stage structure. However, there is a high volume of studies in the DEA field, that have developed the Network DEA (NDEA) models to measure the efficiency of DMUs with more complex production processes that involve more than one stages to produce the final outputs.

The aim of this paper is to address this deficiency in the DEA literature by studying the performance of subsampling bootstrap in NDEA, through Monte Carlo simulations. Among the different bootstrapping approaches, subsampling bootstrap is the computationally easiest one, and therefore, was chosen as the most appropriate for NDEA where the dimensions of the model are usually higher compared to one-stage structures. In this paper, the general two-stage structure is being studied and the stage efficiency estimates are calculated using the additive decomposition approach, upon the assumption of VRS. Coverage probabilities of the

* Corresponding author at: Surrey Business School, University of Surrey, Guildford, UK.
E-mail address: a.emrouznejad@surrey.ac.uk (A. Emrouznejad).

https://doi.org/10.1016/j.ejor.2022.06.022
0377-2217/

confidence intervals for a fixed point coming from two data generating processes (DGPs) - defined on a five and a seven-dimensional input-output space, respectively - are calculated. We show that in NDEA coverage probabilities are more sensitive to the choice of subsample size than in one-stage structures. Finally, the subsampling methodology is applied on a data set referring to European railways to demonstrate the performance of subsampling bootstrap in real world cases.

The remainder of this paper is structured as follows. Section 2 is a review of the main literature in statistical inference about DEA and NDEA. In Section 3, the DEA estimator in the general two-stage production process is defined. In Section 4 subsampling methodology is adapted to NDEA. Section 5 includes details about the Monte Carlo simulations and discussion of the results. Section 6 is an illustration of subsampling to a data set, where the production model has a network structure. Finally, in Section 7, conclusions, limitations and future directions of this study are provided.

2. Literature review

Several studies focused on investigating the statistical properties of the CRS and VRS DEA estimators in different dimensions and established their rates of convergence. Kneip, Park, & Simar (1998) proved that the convergence rate of the VRS-DEA point estimator depends on the smoothness of the frontier. For the case when the frontier is twice differentiable, and under the consistency requirement that the input-output density is positive close to the frontier and strictly positive on the frontier, they found that the convergence rate is $n^{-\frac{2}{p+q+1}}$, where p and q are the number of inputs and outputs respectively. Under the global CRS assumption, Park, Jeong, & Simar (2010) proved that DEA estimator converges faster, at rate $n^{-\frac{2}{p+q}}$. In all cases, the rate at which the DEA estimator converges to the true frontier depends on the number of inputs and outputs; as the dimensions of the model increase, the number of data records should increase exponentially in order to achieve the same rate of convergence. Simar & Wilson (2008) provided a more detailed discussion on the curse of dimensionality of DEA estimators and a comparison with parametric estimators.

Gijbels, Mammen, Park, & Simar (1999) derived the analytical form of the asymptotic distribution of DEA estimator under the VRS, for the case of one input and one output. This is the only case where the asymptotic distribution can be used in practice to make inference. Jeong & Park (2006) extended their work to higher output dimensions. Under the global CRS assumption, Park et al. (2010) found that the DEA estimator follows an exponential distribution. However, in the multivariate cases it is difficult to estimate the distribution's parameters and thus, in practice, these results cannot be used for making inference.

In practice, except for the bivariate case, the only way to get the sampling distribution of the DEA estimators in higher dimensions is by using bootstrapping techniques. Bootstrap was first suggested by Efron (1979) as a method to obtain the sampling distribution of random variables through simulations. In bootstrap techniques, the observed sample X_N, which consists of N random and independent draws from a population, is assumed to mimic the population that it comes from. Therefore, a bootstrap sample X_N^* drawn from the original sample with replacement, can be treated as a sample generated from the population itself. This is known as naïve bootstrapping. By repeatedly imitating the data generating process (DGP) it is possible to get a sufficiently large number of bootstrap samples. The bootstrap sampling distribution obtained, mimics the original sampling distribution.

In the DEA framework, the first study that applied a bootstrap technique to approximate the distribution of the DEA estimator under the VRS assumption, was by Simar & Wilson (1998). DEA estimates obtained with naïve bootstrapping are inconsistent close to the boundary, i.e. as the sample size tends to infinity, the estimator does not converge to the true parameter. This happens because the naïve bootstrap estimate will equal with the sample estimate with non-zero probability, whereas the probability that the sample estimate equals with the true parameter is zero. See Simar & Wilson (1998; 2000) and Kneip, Simar, & Wilson (2008) for further discussion and proof.

To overcome the inconsistency problem, Simar & Wilson (1998) applied a homogeneous smooth bootstrap making the assumption that the distribution of inefficiencies is common for all DMUs. Although the homogeneity assumption is restrictive, this method can give good estimations even with a relatively small data set. Simar & Wilson (2000) extended their previous work to a heterogeneous smooth bootstrap, where the distribution of inefficiencies varies across the DMUs. In both studies, confidence intervals for the efficiency scores of a fixed point (x, y) are constructed using the bias-corrected estimator. Simar & Wilson (1999) applied the homogeneous bootstrap to estimate Malmquist indices and suggested a procedure for confidence interval construction without the explicit use of the bias-corrected estimates. Simar & Wilson (2002) developed hypothesis tests for examining the returns to scale and suggested bootstrap procedures for the estimation of the critical values for the test statistics.

Kneip et al. (2008) proved the consistency of two more bootstrapping techniques, based on smoothing and subsampling, respectively. The smoothing approach, requires smoothing both the input/output density function and the frontier estimate and is computationally very demanding. Kneip, Simar, & Wilson (2011) suggested a simplified version of the double-smoothed bootstrap. However, subsampling bootstrap - originally suggested by Swanepoel (1986) - is computationally easier. It consists of drawing $m = N^\kappa$ observations, usually with replacement, for $\kappa \in (0, 1)$. In this way, the frequency at which the sample maximum is drawn, is reduced, overcoming the inconsistency problem of naïve bootstrap. They also suggested a method to construct confidence intervals without using the bias-corrected estimates to avoid additional noise in the estimation procedure. Subsampling is easy to implement, however, its performance is sensitive to the choice of the subsample size m. Based on the minimum volatility criterion suggested by Politis, Romano, & Wolf (2001) for the subsample size selection, Simar & Wilson (2010) suggested an algorithm to choose an optimal subsample size in the DEA context. They also studied both subsampling with and without replacement and found that the first one yields better results.

Various bootstrapping techniques have been widely applied in one-stage DEA to obtain bias-corrected estimates and confidence intervals for the efficiency scores. For example, in the DEA transportation literature, Wanke (2012) used a Gaussian kernel to draw bootstrap samples and provide confidence interval estimates for the BCC efficiency scores of 68 Brazilian airports, and for testing the returns-to-scale. Marchetti & Wanke (2017) applied a second stage bootstrap truncated regression to assess the impact of contextual variables in the performance of the Brazilian rail concessionaires. In Nwaogbe, Wanke, Barros, & Azad (2017), the impact of contextual variables on 30 major Nigerian airports was assessed by combining first stage bootstrap efficiency estimations with a second stage censored quantile regression. Moradi-Motlagh & Emrouznejad (2022) provide a review of the main methodological developments and of the relevant software that was developed, as well as an extensive overview of the most impactful articles on the field.

In conventional DEA, the production of outputs is considered to occur in one stage and the inner structure of the production process is not taken into account. However, the operating process of a DMU may involve intermediate products which are outputs of one stage and inputs to the next stage. Färe & Whittaker (1995) and Färe & Grosskopf (1996) were among the first studies that considered the network structure of a DMU in its efficiency assessment.

There are two main efficiency decomposition approaches that are used in NDEA literature; the multiplicative and the additive approach. Kao & Hwang (2008) suggested decomposing the overall efficiency as the product of the two stage efficiencies. They linked the two stages assuming that the aggregated outputs of the first stage are introduced unchanged in the second stage. However, the multiplicative decomposition approach can only be applied under the CRS assumption, as under the VRS, the resulting models cannot be linearised. Another limitation of the conventional multiplicative approach is that it can be generalised to multi-stage series structures only in the cases when there are no stage specific inputs and outputs. For general network structures, alternative multiplicative approaches have been developed, such as converting the original model to a parametric linear one (see for example Zha & Liang, 2010). Chen, Cook, Li, & Zhu (2009) decomposed the overall efficiency of a DMU as the weighted average of the stage efficiencies. They defined the decomposition weights endogenously, so they can reflect the relative contribution of each stage to the overall process. The additive decomposition approach has the advantage that it can be used both under the CRS and the VRS assumptions. Cook, Zhu, Bi, & Yang (2010) extended the additive decomposition methodology to general multistage series structures. In NDEA, adjusting the inputs or outputs of a DMU by its efficiency scores will not necessarily project the DMU on the frontier. Chen, Cook, & Zhu (2010) suggested a model to get frontier projections in two stage structures. Chen, Cook, Kao, & Zhu (2013) demonstrated that in NDEA the duality between the envelopment and the multiplier model does not hold and suggested using the first one to get the frontier projections and the second one for calculating the overall and stage efficiency scores.

Despite the great number of applications of NDEA, there are very limited attempts for making statistical inference in NDEA. Trinh & Zelenuyk (2015), based on the work of Simar & Wilson (2002), developed hypothesis tests to examine whether the difference between the first moments and the difference between the density distributions of the efficiency scores in one-stage DEA and NDEA is significant. Bostian et al. (2018) suggested a statistical approach to make inference about NDEA based on a parametric Bayesian approach.

Dia, Takouda, & Golmohammadi (2020) applied a kernel smoothing-bootstrap in a three-stage NDEA assessing the efficiency of Canadian credit unions, where the CRS and VRS efficiency score of each stage are calculated independently. The overall efficiency is then calculated as the average or as the product of the stage efficiency scores. To the best of our knowledge, there is no study investigating the construction of confidence interval estimates for the overall and stage efficiency scores in network production structures taking into account the connection between the stages.

3. DEA estimator in two-stage production processes

Consider a general two-stage series structure depicted in Fig. 1. Suppose that a Decision Making Unit (DMU) in a first stage consumes P inputs $x_p = (x_1, \dots, x_P) \in \mathbb{R}_+^P$ to produce R final first stage outputs $l_r = (l_1, \dots, l_R) \in \mathbb{R}_+^R$ and Q intermediate outputs $z_q = (z_1, \dots, z_Q) \in \mathbb{R}_+^Q$. In the second stage, intermediate products obtained from the first stage, and external second stage inputs $g_t = (g_1, \dots, g_T) \in \mathbb{R}_+^T$ are consumed to produce S final outputs $y_s = (y_1, \dots, y_S) \in \mathbb{R}_+^S$.

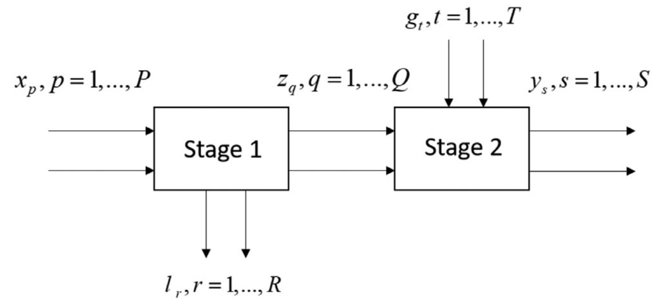

Fig. 1. General two-stage series network structure of a DMU.

In this network structure, the true production possibility set (PPS) of the overall production process is defined as

$$T = \left\{ (x, l, z, g, y) \in \mathbb{R}_+^{P+R+Q+T+S} \middle| x \text{ can produce } l, \text{and } z, \; z \text{ and } g \text{ can produce } y \right\}. \quad (1)$$

Consider the decomposition of the production process into its component stages and let T_1 and T_2 denote the PPSs of the first and second stage respectively. Then,

$$T_1 = \{(x, l, z) \in \mathbb{R}_+^{P+R+Q} | \exists (g, y) \in \mathbb{R}_+^{T+S} : (x, l, z, g, y) \in T\}, \quad (2)$$

$$T_2 = \{(z, g, y) \in \mathbb{R}_+^{Q+T+S} | \exists (x, l) \in \mathbb{R}_+^{P+R} : (x, l, z, g, y) \in T\}. \quad (3)$$

T, T_1 and T_2 can be described by their input or output correspondence sets, which inherit their properties. The input possibility sets for the overall process, the first and the second stage, respectively are

$$X(l, y) = \{(x, z, g) \in \mathbb{R}_+^{P+Q+T} | (x, l, z, g, y) \in T\}, \quad (4)$$

$$X_1(l, z) = \{x \in \mathbb{R}_+^P | (x, l, z) \in T_1\}, \quad (5)$$

$$X_2(y) = \{(z, g) \in \mathbb{R}_+^{Q+T} | (z, g, y) \in T_2\}, \quad (6)$$

and the output possibility sets of the overall process, the first and second stage, respectively, are

$$Y(x, z, g) = \{(l, y) \in \mathbb{R}_+^{R+Q+S} | (x, l, z, g, y) \in T\}, \quad (7)$$

$$Y_1(x) = \{(l, z) \in \mathbb{R}_+^{R+Q} | (x, l, z) \in T_1\}, \quad (8)$$

$$Y_2(z, g) = \{y \in \mathbb{R}_+^S | (z, g, y) \in T_2\}. \quad (9)$$

Concerning the properties of the input and output possibility sets, the assumptions discussed in Shephard (1970) and (Banker et al., 1984) are adopted in this study. Therefore, it is assumed that all input/output sets defined above are closed, the input (output) possibility sets are convex for all outputs (inputs), all inputs and outputs are strongly disposable, and that each of the input/output possibility sets defined above is the intersection of all the sets satisfying these three properties.

The efficient boundaries of the input possibility sets are defined as

$$\partial X(l, y) = \{(x, z, g) | (x, z, g) \in X(l, y), \theta^0 (x, z, g) \notin X(l, y), \forall 0 < \theta^0 < 1\}, \quad (10)$$

$$\partial X_1(l, z) = \{x | x \in X(l, z), \theta^1 x \notin X(l, z, y), \forall 0 < \theta^1 < 1\}, \quad (11)$$

$$\partial X_2(y) = \{(z,g) | (z,g) \in X(y), \theta^2(z,g) \notin X(y), \forall 0 < \theta^2 < 1\}. \tag{12}$$

Let DMU_{j_0} denote a DMU under evaluation. Then, the Farrell (1957) input efficiency measure of DMU_{j_0} for the overall process and the two stages, respectively can be defined as

$$\theta^0(x,l,z,g,y) = \inf\{\theta^0 | \theta^0(x,z,g) \in X(l,y)\}, \tag{13}$$

$$\theta^1(x,l,z) = \inf\{\theta^1 | \theta^1 x \in X_1(l,z)\}, \tag{14}$$

$$\theta^2(z,g,y) = \inf\{\theta^2 | \theta^2(z,g) \in X_2(y)\}. \tag{15}$$

For simplicity of notation, let $\theta^0(x,l,z,g,y) \equiv \theta^0$, $\theta^1(x,l,z) \equiv \theta^1$, and $\theta^2(z,g,y) \equiv \theta^2$. The Farrell input efficiency measure $\theta^{\phi}_{j_0}$, $\phi = 0,1,2$ of the DMU_{j_0} under evaluation is the euclidean distance of the point $(x_{j_0}, l_{j_0}, z_{j_0}, g_{j_0}, y_{j_0})$ from its projection to the overall or stage-efficient boundary, respectively, in direction parallel to the subspace defined by the overall or stage-specific input coordinates, respectively, while keeping the level of outputs fixed. The output efficiency measures can be defined in a similar way, using the overall and stage output feasibility sets, and can be found in the Appendix. Note also that the above formulations refer to the specific network structure, but they can be easily adapted to define the efficiency measure when the production process has a different structure.

In practice, the true production sets T, T_1, T_2 are unknown, and therefore, the efficiency scores $\theta^{\phi}_{j_0}$, $\phi = 0,1,2$ of a DMU_{j_0} need to be estimated based on the observed sample of DMUs. Let $S_N = \{(x_j, l_j, z_j, g_j, y_j) | j = 1,\dots,N\}$ be a random sample of N DMUs that is assumed to be generated by an unknown data generating process (DGP), $\mathcal{P}$. It is assumed that the DGP $\mathcal{P}$ is such that the DMUs included in S_N are i.i.d random variables belonging to the convex PPS T, and $(x_j, l_j, z_j) \in T_1$, $(z_j, g_j, y_j) \in T_2$, with $j = 1,\dots,N$. Let also $\widehat{\theta}^0_j$, $\widehat{\theta}^1_j$ and $\widehat{\theta}^2_j$ denote the estimators of the overall, the first stage and the second stage efficiency, respectively, for DMU_j, $j = 1,2,.,N$, with respect to the observed sample S_N.

Estimates can be obtained by solving the mathematical programmes for the observed sample of DMUs. As noted by Chen et al. (2013), in NDEA, the envelopment model can provide the overall efficiency estimates and the frontier projections, but no information on the stage efficiency estimates, while the multiplier model can be used to obtain the overall and stage efficiency estimates. Since the scope of this study is to make statistical inference about the overall and stage efficiency scores, the multiplier model is being used.

Under the VRS assumption, the independent efficiency score estimates of each stage for a DMU_{j_0}, in the input orientation, can be obtained by solving the following fractional programmes:

$$\begin{aligned} \max\ \widehat{\theta}^1_{j_0} &= \frac{\sum_{q=1}^{Q} \gamma^A_q z_{qj_0} + \sum_{r=1}^{R} \mu_r l_{rj_0} + u^A}{\sum_{p=1}^{P} \nu_p x_{pj_0}} \\ \text{s.t. } & \widehat{\theta}^1_j \le 1, \quad j = 1,\dots,N \\ & \nu_p, \mu_r, \gamma^A_q > 0, \\ & u^A \text{ free in sign} \end{aligned} \tag{16a}$$

$$\begin{aligned} \max\ \widehat{\theta}^2_{j_0} &= \frac{\sum_{s=1}^{S} \eta_s y_{sj_0} + u^B}{\sum_{q=1}^{Q} \gamma^B_q z_{qj_0} + \sum_{t=1}^{T} \pi_t g_{tj_0}} \\ \text{s.t. } & \widehat{\theta}^2_j \le 1, \quad j = 1,\dots,N \\ & \gamma^B_q, \pi_t, \eta_s > 0, \\ & u^B \text{ free in sign.} \end{aligned} \tag{16b}$$

In order to link the two stages, as in Kao & Hwang (2008), it is assumed that the optimal aggregated intermediate outputs of the first stage become inputs to the second stage, i.e. it is assumed that $\gamma^A_q = \gamma^B_q = \gamma_q$.

Adopting the additive decomposition approach (Chen et al., 2009), the overall efficiency $\theta^0_{j_0}$ of DMU_{j_0} is decomposed as the weighted sum of the stage efficiencies,

$$\widehat{\theta}^0_{j_0} = \widehat{w}_{1j_0}\widehat{\theta}^1_{j_0} + \widehat{w}_{2j_0}\widehat{\theta}^2_{j_0}, \quad \text{and} \quad \widehat{w}_{1j_0} + \widehat{w}_{2j_0} = 1. \tag{17}$$

By solving the system of these two equations, the decomposition weights can be obtained as functions of the model variables, as

$$\widehat{w}_{1j} = \frac{\sum_{p=1}^{P} \nu_p x_{pj}}{\sum_{p=1}^{P} \nu_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}, \tag{18}$$

$$\widehat{w}_{2j} = \frac{\sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}{\sum_{p=1}^{P} \nu_p x_{pj} + \sum_{q=1}^{Q} \gamma_q z_{qj} + \sum_{t=1}^{T} \pi_t g_{tj}}. \tag{19}$$

The decomposition weights are defined endogenously as the ratio of each stage's inputs to the total amount of inputs, to reflect the relative contribution of each stage to the overall process. Then, the overall efficiency $\widehat{\theta}^0_{j_0}$ of DMU_{j_0} is given by the fractional model

$$\begin{aligned} \widehat{\theta}^{0*}_{j_0} = \ \max\ & \frac{\sum_{q=1}^{Q} \gamma_q z_{qj_0} + \sum_{r=1}^{R} \mu_r l_{rj_0} + u^A + \sum_{s=1}^{S} \eta_s y_{sj_0} + u^B}{\sum_{p=1}^{P} \nu_p x_{pj_0} + \sum_{q=1}^{Q} \gamma_q z_{qj_0} + \sum_{t=1}^{T} \pi_t g_{tj_0}} \\ \text{s.t. } & \widehat{\theta}^1_j \le 1, \quad j = 1,\dots,N \\ & \widehat{\theta}^2_j \le 1, \quad j = 1,\dots,N \\ & \nu_p, \mu_r, \gamma_q, \pi_t, \eta_s > 0, \\ & u^A, u^B \text{ free in sign,} \end{aligned} \tag{20}$$

where $\widehat{\theta}^{0*}_{j_0}$ denotes the optimal objective value of model (20).

In the calculation of the optimal stage efficiency levels of a DMU_{j_0}, one of the two stages will be given pre-emptive priority. This stage's efficiency score will be maximised while the optimal overall efficiency is preserved. Let p denote the priority stage, and $\widehat{\theta}^{\phi p*}_{j_0}$, $\phi = 1,2$ denote the efficiency estimate of the priority stage for DMU_{j_0} and assume that the first stage is given priority over the second stage. Then,

$$\begin{aligned} \widehat{\theta}^{1p*}_{j_0} =& \max\ \widehat{\theta}^1_{j_0} \\ \text{s.t. } & \widehat{\theta}^1_j \le 1, \quad j = 1,\dots,N \\ & \widehat{\theta}^2_j \le 1, \quad j = 1,\dots,N \\ & \frac{\sum_{q=1}^{Q} \gamma_q z_{qj_0} + \sum_{r=1}^{R} \mu_r l_{rj_0} + u^A + \sum_{s=1}^{S} \eta_s y_{sj_0} + u^B}{\sum_{p=1}^{P} \nu_p x_{pj_0} + \sum_{q=1}^{Q} \gamma_q z_{qj_0} + \sum_{t=1}^{T} \pi_t g_{tj_0}} = \widehat{\theta}^{0*}_{j_0} \\ & \nu_p, \mu_r, \gamma_q, \pi_t, \eta_s > 0. \end{aligned} \tag{21}$$

Models (20) and (21) can be converted into linear ones using the Charnes-Cooper transformation (Charnes & Cooper, 1962) (see Appendix).

Let $(\widehat{\theta}^{0*}_{j_0}, \nu^*_p, \gamma^*_r, \mu^*_q, \pi^*_t, \eta^*_s)$ be the optimal solution to model (20). The optimal decomposition weights for DMU_{j_0} are calculated substituting the optimal multipliers into relations (18) and (19). Then, the efficiency estimate of the second stage is given from the optimal decomposition equation as

$$\widehat{\theta}^{2*}_{j_0} = \frac{\widehat{\theta}^{0*}_{j_0} - \widehat{w}^*_{1j_0}\widehat{\theta}^{1p*}_{j_0}}{\widehat{w}^*_{2j_0}}. \tag{22}$$

If the second stage was given pre-emptive priority, then second stage efficiency would be calculated first, in a similar way, maintaining the overall efficiency level estimate, and then the fist stage efficiency estimate would result from the optimal decomposition equation. Unless a DMU_j, $j = 1, \dots, N$ shows infeasibility problems, the additive decomposition of the overall efficiency is unique, and the choice of the priority stage does not affect the stage efficiency estimates.

In some cases, the optimisation process results in $\widehat{w}^*_{1j} = 0$ or $\widehat{w}^*_{2j} = 0$, which means that the contribution of one of the two stages to the overall process is ignored. In those cases, the decomposition weight restrictions $w_{1j}, w_{2j} \geq \kappa$, for some $\kappa \in (0, 0.5]$, can be added to model (20) or to its linear equivalent.

Definition 3.1. A DMU_{j_0} is considered to be overall efficient if and only if $\widehat{\theta}^{\phi*}_{j_0} = 1$, for $\phi = 0, 1, 2$. It is stage efficient if $\widehat{\theta}^{1*}_{j_0} = 1$ or $\widehat{\theta}^{2*}_{j_0} = 1$.

Because the efficiency measurement described above is radial, a DMU_{j_0} may be overall efficient, but be able to further improve its activity due to the existence of slacks (weak efficiency). However, in this study the existence of slacks is not examined.

Note that the true PPS always includes the sample PPS, i.e $\widehat{T} \subset T$, $\widehat{T_1} \subset T_1$ and $\widehat{T_2} \subset T_2$. Therefore, it holds that $\widehat{\theta}^0 \geq \theta^0$, for all points $(x, l, z, g, y) \in \widehat{T}$, $\widehat{\theta}^1 \geq \theta^1$, for all $(x, l, z) \in \widehat{T_1}$, and $\widehat{\theta}^2 \geq \theta^2$, for all $(z, g, y) \in \widehat{T_2}$. Kneip et al. (1998) proved that the rate at which the estimator converges depends on the number of input and output variables and the degree of the true frontier smoothness. If it is assumed that the true frontier is twice differentiable then, the rate of convergence in the one-stage structure is $N^{-\frac{2}{(p-1)+q+2}}$, where N is the size of the sample and p, q are the number of inputs and outputs, respectively. This result can be easily extended to series network structures.

A common approach of demonstrating the convergence of an estimator is by proving that the estimator is bounded in probability. A set of random variables X_N is bounded in probability if there exist M_ϵ and N_ϵ such that for all $N > N_\epsilon$, and $\epsilon > 0$, $Prob(X_N > M_\epsilon) < \epsilon$. This is denoted by $X_N = O_p(1)$. Similarly, if $X_N = O_p(N^{-a})$ for $a > 0$ means that X_N/N^{-a} is $O_p(1)$. Then, it can be considered that X_N converges at a rate N^{-a}. The convergence rates of the overall and stage processes, respectively, are given in the following Proposition.

Proposition 3.1. *For a two-stage process with final first stage inputs and second stage outputs, under the assumptions of VRS and the true frontier being twice differentiable, it holds that*

$$\widehat{\theta^0} - \theta^0 = O_p(N^{-\frac{2}{P+Q+R+T+S+1}}) \tag{23}$$

$$\widehat{\theta^1} - \theta^1 = O_p(N^{-\frac{2}{P+Q+R+1}}) \tag{24}$$

$$\widehat{\theta^2} - \theta^2 = O_p(N^{-\frac{2}{Q+T+S+1}}) \tag{25}$$

Proof. The proof for the convergence rate for each stage results from the proof of the convergence rate in the one-stage structure provided by Kneip et al. (1998, pg. 7–9) considering the input and output possibility sets of each stage and the overall process, as those are defined in relations (4)–(6) and (7)–(9). □

In the next section, the subsampling bootstrap methodology used for the approximation of the distribution of $\widehat{\theta}^\phi/\theta^\phi = 0, 1, 2$ is discussed.

4. Bootstrapping with subsampling

As it was mentioned in the previous section, T, T_1, T_2, the DGP P and $\theta^0_j, \theta^1_j, \theta^2_j$ for a DMU_j, $j = 1, \dots, N$ are unknown and the observed sample S_N needs to be used to provide estimates for the true efficiency scores. However, when the dimensions increase, it is not possible to derive the sampling properties of the estimators analytically. Since S_N is the only sample available, bootstrapping techniques can be used to derive the sampling distribution of the estimators.

The main assumption in bootstrapping is that the original sample S_N generated from the unknown P, mimics the underlying population that it comes from. Therefore, a bootstrap sample $S^*_N\{(x^*_i, l^*_i, z^*_i, g^*_i, y^*_i)|i = 1, \dots, N\}$ generated from the original sample through a known DGP $\hat{P}$ can be used to estimate the unknown sampling distribution of $\widehat{\theta}^\phi_j$, $j = 1, \dots, N$. In other words, $\hat{\theta}$ is an estimator of θ obtained from the sample S_N through P, and $\widehat{\hat{\theta}}$ is an estimator of $\hat{\theta}$ obtained from the bootstrap sample S^*_N through $\hat{P}$. If $\hat{P}$ is a consistent estimator of P, i.e. $\hat{P}$ converges to P, for a given DMU_{j_0}, it holds that

$$\frac{\widehat{\hat{\theta}}^\phi_{j_0}}{\widehat{\theta}^\phi_{j_0}}\Big|\hat{P} \sim \frac{\widehat{\theta}^\phi_{j_0}}{\theta^\phi_{j_0}}\Big|P, \text{ for } \phi = 0, 1, 2. \tag{26}$$

The distribution of the right-hand side of relation (26) is unknown, but Monte Carlo simulations of the left-hand side can provide approximations of it. By generating a sufficiently large number B of bootstrap samples and applying the NDEA estimator to each one of those, a set of B estimates $\widehat{\hat{\theta}}^\phi_{j_0}$, $\phi = 0, 1, 2$ can be obtained. These can be used to derive the empirical distribution of the left-hand side in relation (26). As $B, N \to \infty$, the approximation of the right hand side becomes accurate.

In order for the relation (26) to hold, it is necessary that $\hat{P}$ is a consistent estimator of P. Naïve bootstrapping does not yield consistent boundary estimations (see Bickel & Freedman, 1981, pg. 1210). Smoothing and subsampling techniques have been proven to give consistent estimations of the production frontier. Subsampling allows for heterogeneity in the efficiency of DMUs and, among the different bootstrapping techniques, is computationally the easiest one, as it does not require any kernel smoothing to achieve consistency. In NDEA, where more than one stages are involved in the production process, the computational burden of kernel smoothing may be prohibitive and subsampling seems to be the most appropriate method for statistical inference.

The subsampling methodology suggested by Kneip et al. (2008) and Simar & Wilson (2010) can be adapted to the NDEA case, by considering the inner structure of DMUs. Kneip et al. proved that drawing pseudo-samples of size $m = N^k$, for $k \in (0, 1)$, where N is the original sample size, allows for consistent inference.

Consider the general two-stage network structure depicted in Fig. 1 and let $\mathcal{S}_N$ be the original sample of N DMUs. In each replication, a bootstrap subsample

$$\mathcal{S}^*_m = \{(x^*_j, l^*_j, z^*_j, g^*_j, y^*_j)|j = 1, \dots, m\}$$

consists of $m < N$ independent and identically distributed (iid) draws with replacement from the original sample.

For each bootstrap subsample, for a DMU_{j_0}, models (20), (21) and Eq. (22) are used to get the overall and stage bootstrap efficiency estimates $\widehat{\hat{\theta}}^{0*}_{j_0}, \widehat{\hat{\theta}}^{1*}_{j_0}$ and $\widehat{\hat{\theta}}^{2*}_{j_0}$, respectively. Kneip et al. (2008) proved for one-stage structures that the bootstrap distribution of $m^{2/(P+Q+1)}\left(\frac{\widehat{\hat{\theta}}_{j_0}}{\widehat{\theta}_{j_0}} - 1\right)$ approximates the distribution of $N^{2/(P+Q+1)}\left(\frac{\widehat{\theta}_{j_0}}{\theta_{j_0}} - 1\right)$, where P the number of inputs and Q the number of outputs. Let $A = 2/(P + R + Q + T + S + 1)$, $B = 2/(P + R + Q + 1)$ and $\Gamma = 2/(Q + T + S + 1)$. Adjusting the result of Kneip et al. (2008) to the specific general network structure, it holds that

$$m^{\tau}\left(\frac{\widehat{\widehat{\theta}}^{\phi}_{j_0}}{\widehat{\theta}^{\phi}_{j_0}}-1\right) \overset{approx}{\sim} N^{\tau}\left(\frac{\widehat{\theta}^{\phi}_{j_0}}{\theta^{\phi}_{j_0}}-1\right), (\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}. \tag{27}$$

Therefore, the bootstrap estimates of the overall and stage efficiencies, respectively, of a DMU_{j_0} can be used to approximate the right-hand side in relations (27) and construct the confidence intervals for the true overall and stage efficiencies.

For a DMU_{j_0}, for a chosen confidence level $a \in (0,1)$, the quantiles $c_{a/2,m}$ and $c_{1-a/2,m}$ of the unknown distributions can be estimated using the quantiles of the bootstrap distribution, obtained as

$$Prob\left(m^{\tau}\left(\frac{\widehat{\widehat{\theta}}^{\phi}_{j_0}}{\widehat{\theta}^{\phi}_{j_0}}-1\right) \leq \widehat{c}_{a/2,m}|S_N\right) = a/2$$

and

$$Prob\left(m^{\tau}\left(\frac{\widehat{\widehat{\theta}}^{\phi}_{j_0}}{\widehat{\theta}^{\phi}_{j_0}}-1\right) \leq \widehat{c}_{1-a/2,m}|S_N\right) = 1-a/2,$$

where $(\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}$.

In practice, this involves sorting the values $m^{\tau}\left(\frac{\widehat{\widehat{\theta}}^{\phi}_{j_0}}{\widehat{\theta}^{\phi}_{j_0}}-1\right)$ in ascending order and then assigning $\widehat{c}_{a/2,m}$ the first value that exceeds $(a/2)100\%$ of the observations and $\widehat{c}_{1-a/2,m}$ the value that is less than $(1-a/2)100\%$ of the observations. Then the true overall and stage efficiencies will lie in the confidence intervals

$$\left[\frac{\widehat{\theta}^{\phi}_{j_0}}{\left(1+N^{-\tau}\widehat{c}_{1-a/2,m}\right)}, \frac{\widehat{\theta}^{\phi}_{j_0}}{\left(1+N^{-\tau}\widehat{c}_{a/2,m}\right)}\right], \tag{28}$$

for $(\tau,\phi) \in \{(A,0),(B,1),(\Gamma,2)\}$, respectively.

5. Monte Carlo simulations

In order to examine the performance of subsampling bootstrap in network structures, Monte Carlo simulations were performed. Two sets of experiments were conducted for the general network structure (see Fig. 1) with five-dimensional and seven-dimensional DGPs, respectively, i.e. for $P=R=Q=T=S=1$ and for $P=R=T=1$, $Q=S=2$. Each experiment consists of 1000 Monte Carlo trials with 2000 bootstrap replications in each trial. Also, experiments were conducted for four sample sizes $N \in \{25,50,100,200\}$ and twelve subsample sizes $m=[N^k]$, for $k \in \{0.4,0.45,0.5,0.55,0.6,0.65,0.7,0.75,0.8,0.85,0.9,0.95\}$, where $[N^k]$ denotes the integer part of N^k.

In all the experiments, the confidence intervals for the true overall and stage efficiencies of a fixed point were estimated. Due to sampling noise, the true probability that the confidence interval contains the true efficiency score may differ from the nominal probability that the confidence level represents. Therefore, through the Monte Carlo trials, the true probability of the confidence interval containing the true parameter is approximated by calculating the coverage probabilities, i.e. the actual proportion of the estimated confidence intervals that include the true efficiency score.

In the first set of experiments (five-dimensional case), an efficient first stage input x_e is drawn from the uniform distribution $U[5,20]$. The efficient input is used to generate one final first stage output $l=(x_e)^{\beta}$ and one intermediate first stage output $z=(x_e)^{\gamma}$, where $\beta,\gamma>0$. Second stage-specific efficient input g_e is also drawn from the uniform distribution $U[5,20]$. The second stage efficient input and the intermediate product from the first stage, are used to generate one final output $y=\left((g_e)^{\zeta}z^{\xi}e^{-0.2|\epsilon|}\right)^{\nu}$, where $\epsilon \sim N(0,1)$ and $\zeta,\xi,\nu>0$. The term $e^{-0.2|\epsilon|}$ is added to the DGP to better reflect real-world scenarios by adding some stochasticity to the data.

For the experiments of the second set (seven-dimensional case), a similar DGP is used. An efficient first stage input x_e and a second stage-specific input g_e are drawn from the uniform distribution $U[5,20]$. One final first stage output is given by $l=(x_e)^{\beta}$ and the two intermediate first stage outputs are given by $z_1=(x_e)^{\gamma}$ and $z_2=(x_e)^{\delta}$, respectively, for some $\beta,\gamma,\delta>0$. The two final outputs are generated by $y_1=\left((g_e)^{\zeta}z_1^{\xi}z_2^{\lambda}e^{-0.2|\epsilon|}\right)^{\nu}$ and $y_2=\left((g_e)^{\zeta}z_1^{\lambda}z_2^{\xi}e^{-0.2|\epsilon|}\right)^{\nu}$, where $\epsilon \sim N(0,1)$ and $\zeta,\xi,\lambda,\nu>0$. Both in the five and the seven-dimensional case, the DGP ensures that at each stage, as well as in the overall process, an increase in stage-specific inputs results in a non-proportional increase in the stage outputs, to account for VRS.

Estimates of the overall and the stage efficiencies, respectively, were obtained through different experiments. In the set of experiments that were performed to estimate the overall efficiency, inefficiency was added to the first stage inputs $x=x_e e^{0.2|\epsilon|}$. The inefficient inputs were then used to calculate the sample overall efficiency of the fixed point under investigation. In the set of experiments concerning the true stage efficiencies, inefficiency was introduced to each stage's specific inputs, i.e., $x=x_e e^{0.2|\epsilon|}$ and $g=g_e e^{0.2|\epsilon|}$, and these inefficient first and second stage-specific inputs were used to calculate the first and second stage sample efficiencies of the fixed point. Since the input orientation is considered, the true overall inefficiency can be defined as the proportion of the input going to the production that exceeds the efficient input. Therefore, the true first and second stage efficiencies are defined as x_e/x and g_e/g, respectively. In the case of multiple inputs, since radial projections are used, all inputs are reduced by the same proportion, and the true efficiency score would be defined in a similar way.

Experiments were performed for $\beta=0.6,\gamma=0.7,\delta=0.8,\zeta=0.1,\xi=0.3,\lambda=0.5$ and $\nu=0.8$, and for the case when $x_e=g_e$. All the efficiency scores were obtained for the fixed point with $(x_{e0},g_{e0})=(13,13)$ and $\epsilon_0=1$, and with true overall and stage efficiency score $\theta^{\phi}_0=0.8187$, for $\phi=0,1,2$. This point lies about in the middle of the cloud of the generated points.

In many subsamples, the fixed point shows infeasibility in one of the two stages. Infeasibility problems occur when one (or more) of the outputs of the fixed point (overall or stage-specific output) is (are) greater than the respective maximum output(s) of the points included in the subsample, and/or when one or more of the inputs of the fixed point is (are) less than the corresponding minimum input(s) of the points belonging to the subsample. Infeasibility problems in the subsamples occur more often in NDEA compared to one stage DEA because of the higher dimensions. Therefore, in this study, in all bootstrap replications the fixed point of interest was being added to the subsample.

Another way to treat infeasibilities is to set the efficiency estimate of the infeasible DMU equal to one. In the additive efficiency decomposition approach, in some cases the algorithm may give zero optimal multipliers and therefore, zero efficiency estimates. For the overall process this happens in cases where the DMU is overall efficient. Therefore, for DMUs with zero overall efficiency estimate, it can be set that their overall efficiency estimate is equal to one. However, it may also happen a DMU to be assigned a non-zero overall efficiency estimate, and get a zero efficiency estimate for the one of the two stages when a stage is considered as priority stage, but get a non-zero efficiency estimate - less than one - when the priority stage changes. Including the DMU under evaluation in the subsample significantly reduces the times when this happens and therefore, provides a partial solution to the problem, whereas

Table 1
Coverage probabilities for the confidence interval estimates when $P = R = Q = T = S = 1$ and the DMU under evaluation is included in the subsample.

		Overall			1st Stage			2nd Stage		
		1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.938	0.976	0.983	0.930	0.965	0.976	0.980	0.984	0.985
	0.45	0.984	0.988	0.988	0.965	0.982	0.985	0.986	0.987	0.987
	0.50	0.989	0.989	0.989	0.978	0.985	0.985	0.991	0.991	0.991
	0.55	0.989	0.989	0.989	0.979	0.983	0.985	0.990	0.991	0.991
	0.60	0.990	0.990	0.990	0.981	0.988	0.988	0.990	0.995	0.995
	0.65	0.984	0.992	0.992	0.981	0.990	0.993	0.959	0.985	0.996
	0.70	0.974	0.994	0.995	0.972	0.990	0.993	0.930	0.970	0.995
	0.75	0.943	0.975	0.996	0.938	0.973	0.992	0.837	0.917	0.988
	0.80	0.889	0.951	0.993	0.903	0.951	0.983	0.740	0.861	0.969
	0.85	0.833	0.910	0.978	0.879	0.923	0.977	0.647	0.783	0.944
	0.90	0.742	0.841	0.964	0.818	0.880	0.962	0.517	0.670	0.878
	0.95	0.646	0.787	0.924	0.766	0.857	0.931	0.410	0.576	0.808
50	0.40	0.986	0.999	0.999	0.920	0.986	1.000	0.999	0.999	0.999
	0.45	1.000	1.000	1.000	0.941	0.993	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	0.978	1.000	1.000	0.999	1.000	1.000
	0.55	1.000	1.000	1.000	0.987	1.000	1.000	0.995	0.999	1.000
	0.60	0.995	1.000	1.000	0.997	1.000	1.000	0.969	0.994	1.000
	0.65	0.987	0.998	1.000	0.994	0.999	1.000	0.906	0.973	0.999
	0.70	0.958	0.982	0.999	0.983	0.996	1.000	0.767	0.905	0.992
	0.75	0.910	0.966	0.994	0.969	0.987	1.000	0.618	0.787	0.968
	0.80	0.826	0.915	0.980	0.937	0.974	0.995	0.434	0.643	0.910
	0.85	0.707	0.826	0.954	0.900	0.942	0.984	0.296	0.466	0.771
	0.90	0.556	0.723	0.905	0.855	0.909	0.975	0.180	0.320	0.643
	0.95	0.412	0.568	0.821	0.769	0.859	0.949	0.104	0.205	0.481
100	0.40	1.000	1.000	1.000	0.868	0.967	1.000	1.000	1.000	1.000
	0.45	1.000	1.000	1.000	0.900	0.977	1.000	1.000	1.000	1.000
	0.50	1.000	1.000	1.000	0.958	0.993	1.000	0.989	0.998	1.000
	0.55	1.000	1.000	1.000	0.980	0.996	1.000	0.945	0.987	0.999
	0.60	0.996	1.000	1.000	0.988	0.997	1.000	0.791	0.941	0.999
	0.65	0.973	0.995	1.000	0.995	0.999	1.000	0.539	0.792	0.983
	0.70	0.903	0.959	0.997	0.988	0.996	1.000	0.261	0.488	0.884
	0.75	0.786	0.899	0.986	0.975	0.991	0.999	0.130	0.300	0.730
	0.80	0.563	0.789	0.947	0.944	0.978	0.996	0.047	0.140	0.489
	0.85	0.361	0.556	0.868	0.903	0.946	0.988	0.015	0.059	0.273
	0.90	0.210	0.363	0.717	0.839	0.908	0.973	0.007	0.025	0.131
	0.95	0.106	0.211	0.515	0.773	0.850	0.940	0.006	0.009	0.069
200	0.40	1.000	1.000	1.000	0.737	0.929	0.999	1.000	1.000	1.000
	0.45	1.000	1.000	1.000	0.819	0.958	0.999	0.995	1.000	1.000
	0.50	1.000	1.000	1.000	0.920	0.982	0.999	0.817	0.976	1.000
	0.55	0.998	1.000	1.000	0.958	0.994	1.000	0.430	0.791	0.995
	0.60	0.971	0.996	1.000	0.985	0.997	1.000	0.098	0.365	0.907
	0.65	0.868	0.964	0.999	0.995	0.997	1.000	0.014	0.103	0.621
	0.70	0.628	0.841	0.985	0.994	1.000	1.000	0.001	0.015	0.288
	0.75	0.333	0.569	0.903	0.987	0.996	1.000	0.000	0.000	0.077
	0.80	0.136	0.310	0.731	0.967	0.988	0.999	0.000	0.000	0.017
	0.85	0.037	0.124	0.442	0.925	0.964	0.997	0.000	0.000	0.002
	0.90	0.011	0.036	0.226	0.861	0.927	0.984	0.000	0.000	0.001
	0.95	0.007	0.013	0.087	0.773	0.846	0.949	0.000	0.000	0.001

setting the efficiency estimate of the infeasible DMUs equal to one does not treat this issue.

Confidence interval estimates for the true efficiency scores were obtained for three levels of significance, $\alpha \in \{0.1, 0.05, 0.01\}$ to get the 90%, 95% and 99% confidence intervals. Coverage probabilities represent the proportion of confidence intervals, i.e. the proportion of Monte Carlo trials, that the true efficiency score is included in the estimated confidence interval. The resulting coverage probabilities of the confidence intervals for the overall and stage efficiencies, for the five and seven-dimensional cases, respectively, when the DMU under evaluation is included in the subsample are given in Tables 1 and 2.

Results from Monte Carlo simulations when the DMU under evaluation is not included in the subsample, the efficiency score of infeasible DMUs is set equal to one and both stages are considered alternately as priority stages are provided in Tables A.2 and A.3 in the Appendix. The issue lies in the additive decomposition algorithm, which in some cases assigns zero efficiency scores. If this happens at a stage level, and not to the overall efficiency, then this mainly affects the coverage probabilities of that stage. In most of the cases, the confidence intervals obtained with this approach have higher range compared to those obtained when the DMU under evaluation is included in the subsample.

The results indicate that the choice of the subsample size is crucial for getting high coverage probabilities, irrespective of the original sample size. However, as the sample size increases, sensitivity on the subsample size seems to increase. Coverage probabilities of the confidence intervals for the true overall and first stage efficiency seem to rise as k increases and then, in most cases, after some point they show a downturn. These conclusions seem to be in accordance with those for one stage structures (see Kneip et al., 2008, pg. 1682–1683). According to the results, one of the two stages shows greater sensitivity to the subsample size, and for some subsample sizes confidence interval estimates can even have zero coverage probabilities. However, this sensitivity is not related to the choice of priority stage. Monte Carlo simulations were performed both by treating first stage as the priority stage, and then by treating the second stage as the priority stage, and coverage

Table 2
Coverage probabilities for the confidence interval estimates when $P = R = T = 1$, $Q = S = 2$ and the DMU under evaluation is included in the subsample.

		Overall			1st Stage			2nd Stage		
		1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.728	0.872	0.978	0.948	0.977	0.988	0.171	0.331	0.769
	0.45	0.882	0.957	0.996	0.970	0.985	0.989	0.312	0.517	0.901
	0.50	0.950	0.987	0.997	0.982	0.990	0.990	0.460	0.756	0.962
	0.55	0.945	0.985	0.997	0.982	0.989	0.990	0.450	0.751	0.964
	0.60	0.977	0.995	0.997	0.983	0.990	0.990	0.557	0.854	0.988
	0.65	0.993	0.998	0.998	0.986	0.992	0.992	0.872	0.957	1.000
	0.70	0.993	0.997	0.998	0.977	0.992	0.993	0.917	0.969	1.000
	0.75	0.972	0.993	0.998	0.958	0.982	0.992	0.960	1.000	1.000
	0.80	0.939	0.974	0.997	0.930	0.964	0.988	1.000	1.000	1.000
	0.85	0.908	0.951	0.994	0.895	0.942	0.980	1.000	1.000	1.000
	0.90	0.849	0.914	0.980	0.817	0.901	0.966	0.999	1.000	1.000
	0.95	0.774	0.871	0.952	0.748	0.845	0.943	0.996	0.998	1.000
50	0.40	0.598	0.808	0.980	0.908	0.984	1.000	0.056	0.193	0.585
	0.45	0.761	0.903	0.995	0.935	0.989	1.000	0.147	0.320	0.748
	0.50	0.910	0.978	0.999	0.977	1.000	1.000	0.303	0.491	0.893
	0.55	0.941	0.990	1.000	0.983	0.999	1.000	0.375	0.570	0.930
	0.60	0.980	0.996	1.000	0.995	1.000	1.000	0.497	0.785	0.965
	0.65	0.987	0.999	1.000	0.996	1.000	1.000	0.597	0.891	0.998
	0.70	0.990	0.996	1.000	0.991	0.997	1.000	0.847	0.946	1.000
	0.75	0.977	0.990	0.999	0.972	0.989	1.000	0.917	0.980	1.000
	0.80	0.947	0.976	0.996	0.935	0.973	0.997	0.954	1.000	1.000
	0.85	0.907	0.949	0.985	0.890	0.940	0.986	1.000	1.000	1.000
	0.90	0.851	0.910	0.969	0.830	0.898	0.964	1.000	1.000	1.000
	0.95	0.753	0.842	0.943	0.751	0.831	0.935	0.998	1.000	1.000
100	0.40	0.556	0.770	0.985	0.816	0.956	0.999	0.048	0.170	0.543
	0.45	0.652	0.845	0.992	0.858	0.972	1.000	0.075	0.242	0.618
	0.50	0.863	0.970	0.998	0.940	0.983	1.000	0.229	0.418	0.814
	0.55	0.929	0.986	1.000	0.969	0.995	1.000	0.317	0.508	0.887
	0.60	0.972	0.994	1.000	0.987	0.998	1.000	0.419	0.634	0.943
	0.65	0.991	0.998	1.000	0.992	1.000	1.000	0.566	0.791	0.976
	0.70	0.991	0.999	0.999	0.988	0.998	1.000	0.697	0.915	0.999
	0.75	0.983	0.993	0.999	0.976	0.990	0.999	0.886	0.962	1.000
	0.80	0.949	0.979	0.998	0.944	0.975	0.996	0.954	0.997	1.000
	0.85	0.894	0.943	0.991	0.899	0.943	0.984	0.996	1.000	1.000
	0.90	0.820	0.887	0.964	0.832	0.896	0.963	0.999	1.000	1.000
	0.95	0.726	0.808	0.923	0.746	0.832	0.928	0.995	0.999	1.000
200	0.40	0.385	0.642	0.969	0.627	0.868	0.997	0.020	0.111	0.497
	0.45	0.536	0.794	0.988	0.734	0.910	0.998	0.051	0.197	0.600
	0.50	0.753	0.918	0.999	0.860	0.970	1.000	0.160	0.370	0.733
	0.55	0.894	0.978	1.000	0.935	0.990	1.000	0.281	0.518	0.854
	0.60	0.964	0.991	1.000	0.974	0.994	1.000	0.449	0.642	0.926
	0.65	0.986	0.998	1.000	0.987	1.000	1.000	0.581	0.733	0.969
	0.70	0.993	0.998	1.000	0.993	0.998	1.000	0.690	0.880	0.992
	0.75	0.983	0.994	0.999	0.977	0.993	0.999	0.848	0.953	1.000
	0.80	0.962	0.984	0.997	0.944	0.976	0.994	0.934	0.987	1.000
	0.85	0.901	0.959	0.989	0.880	0.938	0.983	0.975	1.000	1.000
	0.90	0.819	0.887	0.970	0.828	0.876	0.956	0.999	1.000	1.000
	0.95	0.697	0.788	0.913	0.749	0.823	0.901	0.988	0.999	1.000

Table 3
Mode, mean and range of the optimal subsample size values among the different Monte Carlo trials, using the minimum volatility algorithm, in the five-dimensional case.

		Overall			1st Stage			2nd Stage		
		1-α			1-α			1-α		
N		0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	mode	5	5	5	5	5	5	5	5	5
	mean	8.149	7.874	7.297	9.423	9.144	9.126	9.835	8.458	6.952
	range	13	13	13	14	13	13	13	13	13
50	mode	33	33	8	33	33	33	33	33	8
	mean	25	21.5	15.3	23.67	22.67	20.56	25.67	24.33	19.74
	range	28	28	26	28	28	26	28	28	26
100	mode	63	63	63	63	63	63	63	63	63
	mean	55.4	50.04	41.54	48.79	45.32	44.94	51.04	50.28	41.48
	range	56	56	56	56	56	56	51	56	56
200	mode	117	117	117	117	117	117	117	117	117
	mean	105.6	104.4	96.71	91.26	86.74	84.37	97.59	96.99	91.59
	range	93	93	107	99	99	107	93	93	99

Table 4
Mode, mean and range of the optimal subsample size among the different Monte Carlo trials, using the minimum volatility algorithm, in the seven-dimensional case.

		Overall			1st Stage			2nd Stage		
		$1-\alpha$			$1-\alpha$			$1-\alpha$		
N		0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	mode	5	5	5	5	5	5	5	5	5
	mean	6.926	7.010	7.617	7.913	7.827	8.67	10.1	10.63	10.8
	range	13	14	13	13	14	13	13	13	13
50	mode	33	33	8	33	33	8	8	8	12
	mean	24.41	19.28	12.06	21.90	20.54	16.94	15.15	16.5	17.67
	range	28	28	26	28	28	26	28	28	26
100	mode	63	63	63	63	63	63	12	12	12
	mean	55.08	51.66	34.63	52.16	47.91	42.93	22.92	21.05	23.47
	range	46	46	46	56	56	56	56	56	56
200	mode	117	117	117	117	117	117	10	10	10
	mean	101.7	100.8	95.59	96.43	95.84	93.6	39.97	34.07	27.09
	range	93	86	107	93	86	107	107	107	107

Table 5
Data Set.

	DMU	Invest. Costs (bn €)	O&M Costs(bn €)	TSIWagons	Total Wagons	Length of Lines (km)	Freight MT-km	Pass. M-km	$L_{den} \geq 55$ dB
1	Austria	2.61	1.66	6,511	23,345	5,491	21,361	12,497	1,081,900
2	Belgium	1.78	0.38	2,312	12,013	3,607	0	10,025	324,400
3	Bulgaria	0.30	0.25	568	16,915	4,029	3,434	1,455	42,300
4	Croatia	0.19	0.30	383	2,274	2,604	2,160	827	26,400
5	Czech Rep.	1.35	1.36	8,000	42,199	9,564	15,619	8,738	268,500
6	Denmark	0.39	0.13	225	366	2045	2,616	6,332	84,300
7	Estonia	0.06	0.14	0	20,849	1,161	2,340	316	6,100
8	Finland	0.41	0.18	200	9,942	5,926	9,456	3,868	119,400
9	France	5.09	3.67	8,558	77,678	28,364	32,569	90,612	3,780,000
10	Germany	7.74	3.92	59,626	165,653	38,623	126,686	95,465	6,390,500
11	Ireland	0.16	0.21	100	254	1,931	101	1,991	42,600
12	Latvia	0.11	0.17	0	11,888	1,860	15,873	584	40,600
13	Lithuania	0.22	0.31	0	14,828	1,911	13,790	280	11,600
14	Netherlands	2.73	1.02	9,000	21,226	3,058	6,641	17,483	312,500
15	Poland	3.50	0.69	2,750	83,500	19,132	50,650	19,067	419,700
16	Portugal	0.71	0.26	3,123	3,313	2,546	2,774	4,266	137,100
17	Slovenia	0.23	0.18	226	3,230	1,209	4,360	611	47,600
18	Spain	5.23	0.73	6,781	20,833	16,167	10,550	26,646	69,300
19	Sweden	1.07	0.45	931	11,000	10,882	21,406	12,800	549,400
20	UK	6.46	3.45	15,467	18,246	16,253	17,053	68,010	1,709,400
21	Norway	0.52	0.48	516	1,623	3,895	3,312	3,695	123,400
22	Switzerland	2.50	1.58	19,236	21,200	3,650	12,447	20,657	482,400

probabilities were not affected. Results for stage one being the priority stage and then for stage two being the priority stage, when the DMU under evaluation is not included in the subsample can be found in the Appendix. The number of bootstrap replications certainly affects the coverage probabilities in NDEA more than in conventional DEA due to the lower convergence rate. However, it is computationally difficult to further increase the replications to a number significantly higher than 2000.

In the five-dimensional case (see Table 1), coverage probabilities for the overall and the second stage efficiency are very high for lower values of k, but as k increases they get very poor and even tend to zero as the original sample size increases. In the cases when the estimations of the overall efficiency scores are poor, the coverage probabilities of the second stage seem to be also affected. For the first stage, $k = 0.60$ or $k = 0.65$ seems to result in higher coverages.

In the seven-dimensional case (see Table 2), for the overall and the first stage, subsample sizes resulting from $k = 0.65$ or $k = 0.70$, in most of the cases, yield higher coverage probabilities, however, for a wide range of k, coverage probabilities are still very high. Coverage probabilities for the second stage true efficiency are very poor for the first half values of k, especially for the larger sample sizes. Nonetheless, as k increases coverage probabilities for the second stage become very high; for this stage, in most of the cases, a value of k around 0.9 gives the best coverage probabilities. The difference between the results in the five and the seven-dimensional cases indicate that the coverage probabilities and the optimal subsample sizes strongly depend on the DGP.

Wrong choice of subsample size may result in totally misleading confidence interval estimates. Politis et al. (2001) suggested a minimum volatility criterion for the selection of the optimal subsample size. Let $I_{m,low}$ and $I_{m,up}$ be the lower and the upper bounds of a confidence interval estimate, resulting from subsampling bootstrap, with subsample size m. For a small $r \in \mathbb{Z}^+$, Politis et al. suggested calculating the volatility index $V_m = \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,low}, \ldots, I_{m+r,low}\} + \sum_{m=m_{small}}^{m_{big}} sd\{I_{m-r,up}, \ldots, I_{m+r,up}\}$ and then choosing the subsample size m that corresponds to the minimum V_m.

Results from applying the above algorithm to the confidence interval estimates of each Monte Carlo trial are given in Tables 3 and 4, for the five and seven-dimensional case, respectively.

According to the results, in the five-dimensional DGP (see Table 3), in almost all the cases, the mode for the optimal subsample size is the same for obtaining the overall and stage efficiency estimations and does not change for the different levels of significance. However, from the range of the optimal subsample values, it can be seen that the optimal subsample size can vary significantly among the different Monte Carlo trials, although all samples are generated through the same DGP, and it should be expected that the algorithm would yield similar values for the optimal subsample size.

Furthermore, as it was mentioned previously, for this five-dimensional DGP, values of $k = 0.60$ and $k = 0.65$ yielded the highest coverage probabilities in most of the cases. These values correspond to subsample sizes $[25^{0.60}] = 6$ and $[25^{0.65}] = 8$, respectively, when $N = 25$, subsample sizes 10 and 12 when $N = 50$, 15 and 19 when $N = 100$ and 24 and 31 respectively, when $N = 200$. These subsample sizes, however, are much lower than those indicated by the algorithm, especially when the sample size increases.

For the seven-dimensional DGP, there is a substantial difference between the optimal subsample size for obtaining the overall and first stage efficiency score, compared to the optimal subsample size for the second stage. That raises an important issue, as all the estimations, for the overall, first and second stage efficiencies, are based on the same subsample size, and it is not possible to use the stage-specific optimal subsample size without affecting the consistency of the results. This difference between the overall-first stage and the second stage optimal subsample sizes in this DGP, is also reflected in the coverage probabilities reported in Table 2, where the value of k with higher coverage probabilities for the second stage ($k = 0.90$), was significantly different compared to that for the overall and first stage estimations ($k = 0.65, 0.70$). Similarly to the five-dimensional case, the subsample sizes that resulted in higher coverage probabilities in Table 2, do not coincide with the optimal subsample sizes yielded by the minimum volatility algorithm.

These issues indicate the great sensitivity of the algorithm, and demonstrate the need for defining a more robust way for the subsample size selection in NDEA. This variability of the optimal subsample size among the different Monte Carlo trials and the different stages can be imputed to the zero-efficiency score issue of the NDEA algorithm that mainly affects the upper bounds of the confidence intervals. This issue is further discussed in the following Section were the subsampling bootstrap is applied in a real dataset, and some suggestions for dealing with it are provided.

6. Application

In this section, the subsampling methodology is applied in the efficiency analysis of railways in 22 European countries. The data set has been previously used by Michali, Emrouznejad, Dehnokhalaji, & Clegg (2021) and considers the noise-pollution problem arising from the operation of railways, with measurements referring to 2016–2017. A limitation of this data set is that it does not include information about all the European countries, as countries with missing data were excluded. As in DEA the efficient frontier is empirically constructed from the available set of DMUs, omission of DMUs may have an impact on its shape and result in some DMUs being assigned a higher efficiency score. Since the true frontier is unknown, the subsampling methodology can be used to provide estimations of the efficiency scores of European railways.

Michali et al. (2021) divided the railway transport process into assets and services related stages. Infrastructure investment and operating costs are used as inputs to the first stage, and to decrease the dimensions of the model, they are added in a single input that represents costs (x). The number of wagons in each country that are compliant with the noise emission standards set by the European Commission are final outputs from the first stage (l) and the total number of wagons (z_1) and the length of railway lines (z_2) in each country are the intermediate products that are introduced to the second stage. The second stage has two desirable outputs, the millions of passengers (y_1) and the million tonnes of freight (y_2) that travelled over one kilometre, and one undesirable output (y_b), the number of people exposed to high levels of railway noise ($L_{den} \geq 55dB$). The data set is provided in Table 5.

The undesirable output of the second stage is treated as input to that stage, as the aim is to proportionally decrease it, together

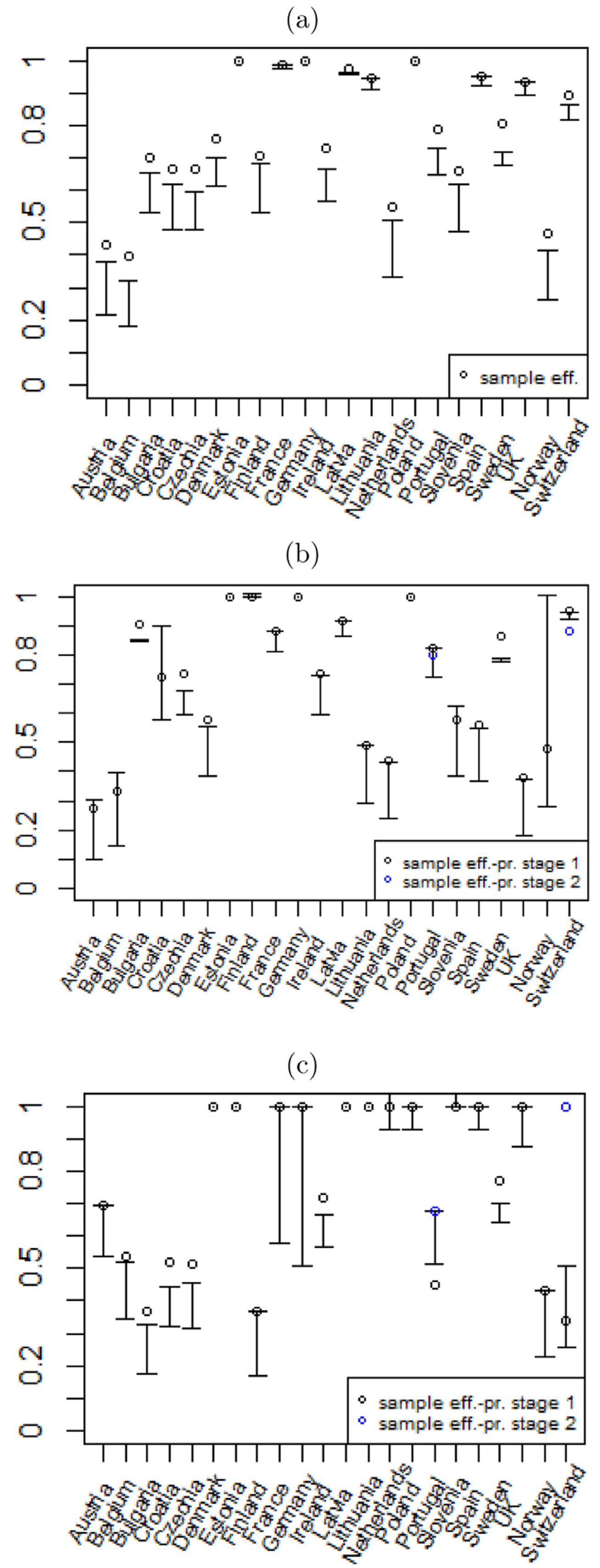

Fig. 2. Confidence interval estimates for (a) the overall, (b) the first stage, and (c) the second stage efficiency scores of European railways, for $k = 0.65$, $m = 7$, and $a = 0.1$, when the DMU under evaluation is included in the subsample.

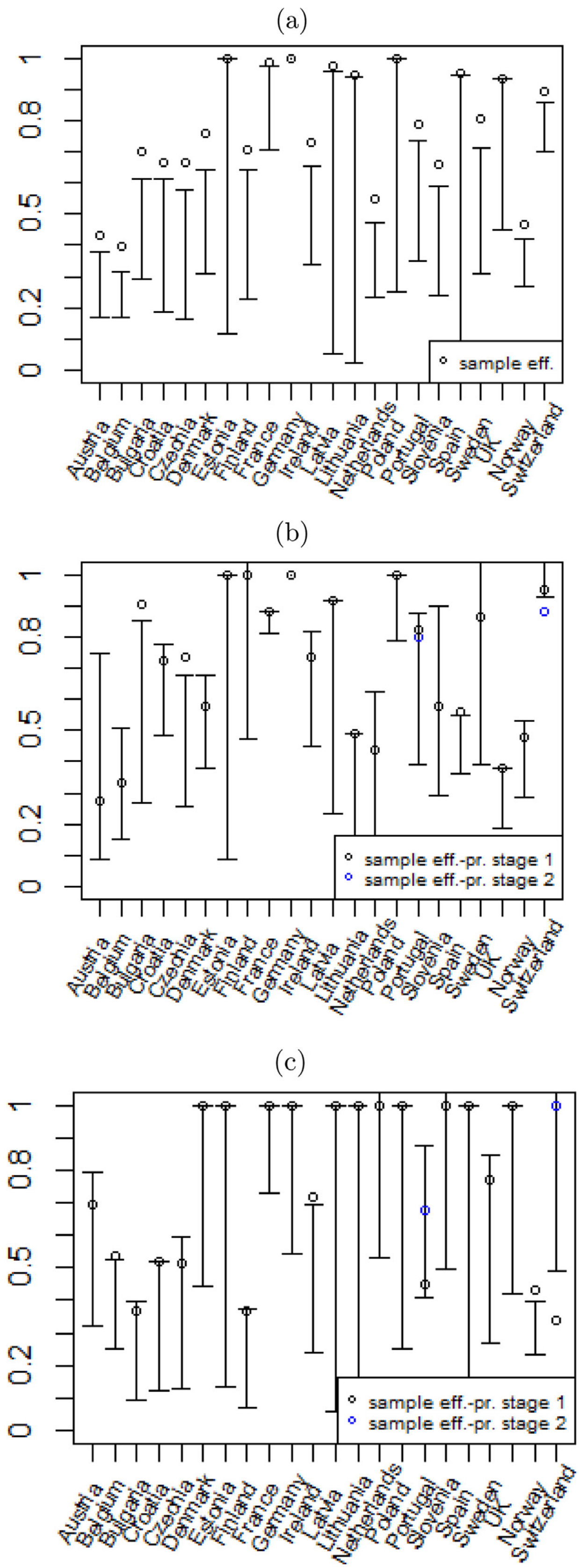

Fig. 3. Confidence interval estimates for (a) the overall, (b) the first stage, and (c) the second stage efficiency scores of European railways, for $k = 0.65$, $m = 7$, and $a = 0.1$, when the DMU under evaluation is not included in the subsample.

with the second stage inputs. In this case, the structure of the model is the same as the seven-dimensional case studied in the previous section.

Michali et al. (2021) performed a sensitivity analysis for the decomposition weights, for this specific data set, and suggested the use of the restrictions $\hat{w}_{1j}, \hat{w}_{2j} \geq 0.1$ in model (20). Therefore, in this analysis, the same decomposition weight restrictions are being used. Furthermore, although the additive approach suggested by Chen et al. (2009) usually provides a unique efficiency decomposition, in this data set, due to infeasibilities, Portugal and Switzerland do not have a unique efficiency decomposition, and their stage efficiency scores change depending on which one of the two stages is chosen as the priority stage.

Confidence interval estimates were obtained for different values of k, i.e. for various subsample sizes. Initially, the DMU under evaluation was included in the subsample. For very small subsample sizes, for many countries in this data set, the upper bound of the confidence interval for some stage efficiency score was either above one or negative. The negative bounds result very rarely, and are due to the boundary condition of the efficiency score on zero. An upper bound above one happens because in a significant number of bootstrap replications, the algorithm was returning zero values. This is an issue that can occur in efficiency decomposition approaches, but becomes more common when applying the subsampling bootstrap, because of the smaller size of the sample. If this happens in more than $(a/2)100\%$ of the bootstrap replications, the upper bound of the confidence interval is above one. This was more common in the cases when the subsample size was very small, but even for larger subsample sizes there were still some DMUs for which the upper bound exceeds one.

In some cases, obtaining the bootstrapped stage efficiency estimates while treating both stages as priority stages, and use the decomposition Eq. (22) - or the equivalent equation for the first stage- to obtain the efficiency estimates of the stage when the priority-stage-model yields infeasibility solves the zero efficiency estimate issue, without further affecting the results. However, this usually does not offer a solution to the problem, as both priority stages may yield zero bootstrap estimates. Therefore, removing the zero bootstrap estimates seems to be the only way to prevent them from distorting the upper bound of the confidence interval. Although this reduces the size of the bootstrap efficiency sample, if the number of bootstrap replications is large enough, it should not affect the quality of the results.

The subsample size that was used was $m = 7$, for $k = 0.65$. The choice of the subsample size was based on the coverage probabilities obtained from the Monte Carlo simulations and the results from the minimum volatility algorithm of Politis et al. (2001) for the overall process.

The model was implemented in R. In order to minimise the computational time, after drawing the 2000 bootstrap subsamples for a specific k, i.e. for a specific subsample size, the confidence intervals for each DMU were calculated separately and not in one *for* loop- but based on the same subsamples. In this way, it is possible to use parallel processing and reduce the computational cost. Although this dataset is small, in larger datasets this approach can make a significant difference in the computational time.

Figure 2 reports the sample overall and stage efficiency scores and their corresponding confidence interval estimates, for $k = 0.65$ and subsample size $m = \lceil 22^{0.65} \rceil = 7$, for the significance level $a = 0.1$, when the DMU under evaluation is included in the subsample. The specific values are provided in Table A.1 in the Appendix. The original sample efficiency scores are also depicted in Fig. 2 with small circles. In Fig. 3, the confidence interval estimates when the DMU under evaluation is not included in the subsample are provided.

For some efficient countries the confidence interval estimates converge to a single point, as in the majority of the bootstrap replications these countries yield an efficiency score equal to one, even for different subsample sizes.

According to the results, in this data set, the specific subsample size yields better overall efficiency estimates compared to stage efficiency scores, and estimations about the second stage are in most cases better compared to the first stage estimations. This is in accordance with Monte Carlo simulation results, where for one of the two stages coverage probabilities were more sensitive to the subsample size selection.

Confidence interval estimates reveal where the true efficiency score for each country lies. Although some efficiency scores appear to be very high, the lower bound estimation for their true efficiency is much lower, and sometimes a DMU with higher efficiency score from another might in reality be less efficient. For example, based only on the original sample, France and Germany are second stage efficient, whereas Sweden has a lower efficiency score. However, the lower bound estimation for France and Germany is lower than that of Sweeden. That means in reality, there is a chance that those two countries are less efficient in the second stage than Sweden. In this data set, confidence interval estimation is particularly useful in the second stage, where without the bootstrap estimations, 11 countries appear to be efficient. However, confidence interval estimates reveal that there might be differences in their true efficiency scores. For the second stage efficient countries Poland, Spain and Netherlands, for example, the lower bound remains above 0.9, whereas for the UK, the lower bound is about 0.8. For Germany and France, the lower bounds lie even lower. Therefore, confidence interval estimates should be considered from the countries to get a better insight into what is the main source of inefficiency for their railway network, to be able to form an effective improvement agenda. As it was indicated by the results, considering these estimates, provides higher discrimination among the different countries' railway-efficiency level.

As was discussed in the previous section, not including the DMU under evaluation in the subsample, increases the range of the confidence intervals. In this dataset, this approach resulted in infinite upper bounds for the majority of DMUs. The following adaptations resulted in avoiding the infinite upper bound, and significantly reduced the range of the confidence intervals: (i) setting the zero overall efficiency scores equal to one, (ii) treating both stages as priority stage and (iii) using the decomposition equation when the priority-stage-model yields zero efficiency score values Or alternatively, (i) setting the zero overall efficiency scores equal to one, and (ii) removing the stage bootstrap efficiency estimates that are equal to zero. Confidence interval estimations for $k = 0.65$ and $a = 0.1$ when the DMU under evaluation is not included in the subsample, are given in Fig. 3 in the Appendix.

7. Conclusion

The DEA approach, where the production frontier is constructed empirically, does not consider for sampling noise. Bootstrapping techniques are now well-established in the DEA literature, and are broadly used to make statistical inference about the efficiency scores in one stage production processes. However, in many cases, the production structure of DMUs involves sub-stages which need to be considered in the efficiency measurement. This paper examines the applicability and performance of bootstrapping in general two-stage structures, where the additive decomposition approach is used to calculate the overall and stage efficiency estimates, and the VRS is assumed.

Bootstrapping can be computationally demanding, especially in high-dimensional models. For this reason, among the different bootstrapping techniques, in this paper the subsampling methodology was considered, as it does not require any kernel smoothing and also allows for heterogeneity in the efficiency distributions among the different DMUs.

Monte Carlo simulations were performed, based on samples obtained through two DGPs, defined for a five and a seven dimensional two-stage series structure, respectively. According to the results, in network structures, the coverage probabilities are more sensitive to the DGP compared to single-stage structures. Similarly to one-stage processes, coverage probabilities are very sensitive both to the sample and subsample sizes, and get lower as the sample size increases. However, in contrast to conventional DEA, in NDEA for some subsample sizes, coverage probabilities tend to zero. That means that the subsample size should be selected very carefully, notably when the size of the original sample is large, as in any other case, the resulting confidence intervals could be misleading.

The selection of the subsample size is a limitation of applying subsampling bootstrap in NDEA. The algorithm suggested by Politis et al. (2001) may offer a rule of thumb, especially in one-stage structures, but in NDEA, it was shown that the optimal subsample size resulting from the minimum volatility algorithm is not always the one that yields the highest coverage probabilities. This issue was further investigated and it seems that a peculiarity of the additive efficiency decomposition approach that assigns zero optimal weights to some DMUs when considering some stage as priority stage affects the performance of the minimum volatility algorithm. Therefore, in this study it was suggested to either always include the DMU under evaluation in the subsample, or set the zero overall efficiency scores equal to one and use the priority-stage model to get the efficiency score of both stages, and then use the decomposition equation in the cases when the priority-stage-model yields zero optimal results.

Future studies could focus on defining a reliable method for the subsample size selection, as well as studying the performance of bootstrapping in other network structures, under different returns to scale assumptions, and/or considering substitution effects. Extension to the output orientation can also be considered, although the model orientation should not affect the results.

Appendix

The efficient boundaries of the output possibility sets are defined as

$$\partial Y(x,z,g) = \{(l,y) | (l,y) \in Y(x,z,g), \lambda^0(l,y) \notin Y(x,z,g), \forall \lambda^0 > 1\}, \tag{A.1}$$

$$\partial Y_1(x) = \{(l,z) | (l,z) \in Y_1(x), \lambda^1(l,z) \notin Y_1(x)\}, \tag{A.2}$$

$$\partial Y_2(z,g) = \{y | y \in Y_2(z,g), \lambda^2 y \notin Y_2(z,g)\}. \tag{A.3}$$

The Farrell (1957) output efficiency measure of DMU_{j_0} for the overall process and the two stages, respectively is defined as

$$\lambda^0(x,l,z,g,y) = \sup\{\lambda^0 | \lambda^0(l,y) \in Y(x,z,g)\}, \tag{A.4}$$

$$\lambda^1(x,l,z) = \sup\{\lambda^1 | \lambda^1(l,z) \in Y_1(x)\}, \tag{A.5}$$

$$\lambda^2(z,g,y) = \sup\{\lambda^2 | \lambda^2(y) \in X_2(z,g)\}. \tag{A.6}$$

Table A.1
Sample efficiency scores and bootstrap confidence interval estimates when stage 1 is the priority stage and for $a = 0.1$.

	DMU	θ^{0*}	lower b.	upper b.	θ^{1p*}	lower b.	upper b.	θ^{2*}	lower b.	upper b.
1	Austria	0.433637	0.215287	0.376686	0.274788	0.099529	0.301128	0.694873	0.537419	0.696555
2	Belgium	0.397365	0.182471	0.318035	0.333166	0.142657	0.399364	0.539735	0.344005	0.519700
3	Bulgaria	0.703788	0.530427	0.651790	0.908831	0.851818	0.851818	0.368453	0.171882	0.326891
4	Croatia	0.665272	0.478368	0.617430	0.727019	0.581371	0.899571	0.519453	0.321197	0.444597
5	Czech Rep.	0.667277	0.481022	0.594025	0.735596	0.593307	0.677312	0.511593	0.312524	0.452911
6	Denmark	0.762902	0.614582	0.700496	0.576000	0.386280	0.557439	1	1	1
7	Estonia	1	1	1	1	1	1	1	1	1
8	Finland	0.705130	0.532281	0.683309	1	1	1.010752	0.365703	0.169523	0.366258
9	France	0.988221	0.979159	0.988222	0.882214	0.810054	0.882215	1	0.577587	1
10	Germany	1	1	1	1	1	1	1	0.509655	1
11	Ireland	0.731261	0.568906	0.664852	0.737318	0.595714	0.731558	0.718122	0.569082	0.664763
12	Latvia	0.977333	0.960043	0.963409	0.918610	0.867337	0.918610	1	1	1
13	Lithuania	0.949317	0.911528	0.947552	0.493179	0.292572	0.493179	1	1	1
14	Netherlands	0.547971	0.334011	0.509601	0.438678	0.236643	0.433225	1	0.930982	1
15	Poland	1	1	1	1	1	1	1	0.930982	1
16	Portugal	0.787282	0.650746	0.730033	0.824983	0.722690	0.824984	0.447969	0.513432	0.675662
17	Slovenia	0.659973	0.471381	0.621496	0.576376	0.386729	0.623256	1	1	1.044866
18	Spain	0.955938	0.922906	0.954406	0.559388	0.366681	0.546500	1	0.930982	1
19	Sweden	0.805756	0.678702	0.721138	0.862833	0.780090	0.787411	0.769217	0.640919	0.701204
20	UK	0.938041	0.892275	0.937471	0.380410	0.182301	0.375570	1	0.874350	1
21	Norway	0.467649	0.259967	0.414903	0.481784	0.280487	1.008325	0.431738	0.229868	0.432147
22	Switzerland	0.893634	0.818026	0.863041	0.955176	0.926178	0.948139	0.339756	0.253971	0.509072

Table A.2
Coverage probabilities for the confidence interval estimates when $P = R = Q = T = S = 1$ and the DMU under evaluation is not included in the subsample.

		Overall			1st Stage						2nd Stage					
					Stage 1 Priority			Stage 2 Priority			Stage 1 Priority			Stage 2 Priority		
		1-α			1-α			1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.967	0.994	1.000	0.991	0.994	0.998	0.985	0.991	0.998	0.993	0.997	1.000	0.989	0.997	1.000
	0.45	0.996	1.000	1.000	0.991	0.993	0.998	0.982	0.987	0.997	0.976	0.994	0.999	0.974	0.991	0.999
	0.50	0.998	1.000	1.000	0.990	0.994	0.999	0.983	0.989	0.998	0.959	0.985	0.998	0.958	0.983	0.998
	0.55	0.999	1.000	1.000	0.989	0.994	0.998	0.984	0.990	0.997	0.959	0.986	0.998	0.957	0.983	0.998
	0.60	0.998	1.000	1.000	0.983	0.993	0.997	0.974	0.988	0.996	0.917	0.979	0.998	0.914	0.975	0.998
	0.65	0.989	0.997	1.000	0.948	0.987	0.996	0.943	0.983	0.995	0.821	0.920	0.996	0.820	0.919	0.995
	0.70	0.971	0.993	1.000	0.928	0.978	0.994	0.924	0.970	0.993	0.761	0.883	0.995	0.764	0.885	0.994
	0.75	0.935	0.975	0.999	0.893	0.948	0.994	0.891	0.945	0.993	0.661	0.810	0.967	0.670	0.814	0.965
	0.80	0.891	0.945	0.993	0.873	0.915	0.990	0.870	0.913	0.988	0.572	0.727	0.938	0.572	0.733	0.940
	0.85	0.830	0.909	0.983	0.833	0.896	0.979	0.831	0.894	0.977	0.492	0.654	0.904	0.490	0.659	0.904
	0.90	0.733	0.851	0.954	0.786	0.865	0.958	0.786	0.865	0.957	0.399	0.560	0.820	0.390	0.558	0.816
	0.95	0.636	0.775	0.922	0.738	0.825	0.933	0.736	0.824	0.932	0.336	0.488	0.744	0.331	0.479	0.748
50	0.40	0.994	1.000	1.000	0.970	0.999	1.000	0.971	0.999	1.000	0.990	0.999	1.000	0.988	0.999	1.000
	0.45	1.000	1.000	1.000	0.975	0.997	1.000	0.973	0.997	1.000	0.974	0.998	1.000	0.972	0.998	1.000
	0.50	1.000	1.000	1.000	0.981	0.998	1.000	0.979	0.997	1.000	0.875	0.971	0.999	0.878	0.971	1.000
	0.55	0.999	1.000	1.000	0.978	0.995	1.000	0.977	0.995	1.000	0.816	0.943	0.997	0.825	0.947	0.997
	0.60	0.995	1.000	1.000	0.970	0.985	1.000	0.969	0.985	1.000	0.679	0.859	0.985	0.685	0.865	0.985
	0.65	0.985	0.996	1.000	0.960	0.980	0.998	0.962	0.980	0.998	0.559	0.771	0.959	0.578	0.780	0.966
	0.70	0.962	0.986	1.000	0.936	0.969	0.996	0.936	0.969	0.996	0.418	0.627	0.922	0.431	0.635	0.920
	0.75	0.917	0.968	0.994	0.912	0.955	0.988	0.911	0.955	0.988	0.321	0.522	0.853	0.327	0.526	0.861
	0.80	0.844	0.924	0.987	0.879	0.929	0.979	0.878	0.931	0.979	0.235	0.403	0.751	0.237	0.409	0.757
	0.85	0.717	0.855	0.964	0.842	0.903	0.967	0.842	0.903	0.969	0.168	0.302	0.654	0.172	0.302	0.658
	0.90	0.587	0.744	0.920	0.801	0.873	0.946	0.800	0.871	0.946	0.115	0.230	0.532	0.121	0.227	0.531
	0.95	0.427	0.599	0.839	0.749	0.823	0.919	0.747	0.822	0.919	0.075	0.174	0.430	0.066	0.168	0.425
100	0.40	1.000	1.000	1.000	0.975	0.996	1.000	0.976	0.997	1.000	0.845	0.984	1.000	0.851	0.984	1.000
	0.45	1.000	1.000	1.000	0.982	0.998	1.000	0.983	0.998	1.000	0.732	0.942	1.000	0.747	0.943	1.000
	0.50	1.000	1.000	1.000	0.986	0.996	1.000	0.985	0.996	1.000	0.363	0.704	0.986	0.387	0.735	0.989
	0.55	0.999	1.000	1.000	0.979	0.994	0.998	0.980	0.994	0.998	0.212	0.541	0.943	0.230	0.561	0.952
	0.60	0.997	0.999	1.000	0.977	0.989	0.997	0.977	0.989	0.997	0.110	0.326	0.849	0.123	0.346	0.861
	0.65	0.972	0.996	1.000	0.965	0.981	0.997	0.966	0.981	0.997	0.056	0.175	0.668	0.059	0.186	0.690
	0.70	0.904	0.961	0.997	0.947	0.971	0.992	0.947	0.971	0.992	0.018	0.090	0.463	0.019	0.096	0.480
	0.75	0.781	0.910	0.989	0.933	0.959	0.985	0.933	0.959	0.985	0.009	0.050	0.315	0.009	0.053	0.323
	0.80	0.567	0.772	0.957	0.900	0.946	0.981	0.898	0.946	0.981	0.008	0.031	0.189	0.007	0.029	0.199
	0.85	0.355	0.555	0.872	0.854	0.912	0.964	0.853	0.909	0.964	0.005	0.014	0.118	0.004	0.014	0.118
	0.90	0.205	0.367	0.700	0.796	0.871	0.953	0.796	0.870	0.953	0.002	0.012	0.077	0.001	0.009	0.076
	0.95	0.104	0.211	0.511	0.734	0.828	0.929	0.734	0.826	0.928	0.001	0.009	0.055	0.002	0.006	0.053
200	0.40	1.000	1.000	1.000	0.936	0.993	1.000	0.936	0.993	1.000	0.012	0.269	0.982	0.020	0.294	0.982
	0.45	1.000	1.000	1.000	0.957	0.996	1.000	0.957	0.996	1.000	0.004	0.056	0.833	0.004	0.071	0.846
	0.50	1.000	1.000	1.000	0.982	0.997	1.000	0.982	0.997	1.000	0.000	0.003	0.325	0.001	0.005	0.357
	0.55	0.996	0.999	1.000	0.986	0.997	1.000	0.986	0.997	1.000	0.000	0.003	0.104	0.000	0.002	0.128
	0.60	0.970	0.993	1.000	0.978	0.997	1.000	0.978	0.996	1.000	0.000	0.001	0.025	0.000	0.001	0.029
	0.65	0.858	0.965	0.998	0.974	0.990	0.999	0.973	0.990	0.999	0.000	0.000	0.005	0.000	0.000	0.009
	0.70	0.613	0.842	0.988	0.960	0.981	0.998	0.960	0.981	0.998	0.000	0.000	0.004	0.000	0.000	0.003
	0.75	0.303	0.567	0.914	0.929	0.971	0.995	0.929	0.970	0.995	0.000	0.000	0.000	0.000	0.000	0.000
	0.80	0.106	0.279	0.721	0.898	0.945	0.988	0.897	0.943	0.988	0.000	0.000	0.000	0.000	0.000	0.000
	0.85	0.041	0.103	0.442	0.848	0.906	0.977	0.847	0.904	0.976	0.000	0.000	0.000	0.000	0.000	0.000
	0.90	0.022	0.048	0.200	0.780	0.865	0.957	0.780	0.864	0.957	0.000	0.000	0.000	0.000	0.000	0.000
	0.95	0.010	0.024	0.078	0.714	0.799	0.915	0.714	0.799	0.914	0.000	0.000	0.000	0.000	0.000	0.000

Table A.3
Coverage probabilities for the confidence interval estimates when $P = R = T = 1, Q = S = 2$ and the DMU under evaluation is not included in the subsample.

		Overall			1st Stage						2nd Stage					
					Stage 1 Priority			Stage 2 Priority			Stage 1 Priority			Stage 2 Priority		
		1-α			1-α			1-α			1-α			1-α		
N	k	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99	0.90	0.95	0.99
25	0.40	0.810	0.931	0.995	0.929	0.951	0.955	0.929	0.951	0.955	0.911	0.964	0.988	0.938	0.980	0.987
	0.45	0.930	0.982	1.000	0.957	0.965	0.968	0.957	0.965	0.968	0.963	0.982	0.991	0.979	0.988	0.990
	0.50	0.972	0.996	1.000	0.967	0.972	0.972	0.967	0.972	0.972	0.976	0.993	0.993	0.989	0.992	0.992
	0.55	0.971	0.995	1.000	0.967	0.970	0.972	0.967	0.970	0.972	0.976	0.990	0.993	0.987	0.990	0.992
	0.60	0.982	0.998	0.999	0.970	0.976	0.976	0.971	0.976	0.976	0.981	0.995	0.996	0.991	0.994	0.995
	0.65	0.989	0.996	0.999	0.968	0.984	0.986	0.968	0.984	0.986	0.990	0.994	0.997	0.988	0.993	0.996
	0.70	0.980	0.994	0.999	0.957	0.980	0.986	0.957	0.981	0.987	0.971	0.993	0.998	0.972	0.993	0.997
	0.75	0.956	0.982	0.996	0.934	0.962	0.987	0.935	0.962	0.987	0.949	0.979	0.998	0.947	0.978	0.997
	0.80	0.921	0.958	0.994	0.902	0.949	0.985	0.902	0.950	0.985	0.923	0.954	0.993	0.922	0.958	0.995
	0.85	0.877	0.935	0.987	0.871	0.930	0.973	0.872	0.930	0.973	0.891	0.938	0.985	0.890	0.936	0.981
	0.90	0.810	0.887	0.965	0.817	0.885	0.960	0.815	0.885	0.960	0.838	0.901	0.965	0.832	0.904	0.965
	0.95	0.746	0.835	0.944	0.740	0.850	0.941	0.739	0.849	0.942	0.788	0.869	0.954	0.776	0.861	0.954
50	0.40	0.696	0.859	0.989	0.960	0.995	1.000	0.972	0.995	1.000	0.933	0.987	0.998	0.955	0.997	1.000
	0.45	0.818	0.932	0.999	0.972	0.995	1.000	0.976	0.996	1.000	0.974	0.993	1.000	0.991	1.000	1.000
	0.50	0.936	0.987	1.000	0.983	0.997	1.000	0.984	0.998	1.000	0.988	0.998	1.000	0.997	1.000	1.000
	0.55	0.962	0.996	1.000	0.986	0.998	1.000	0.985	0.997	1.000	0.991	0.997	1.000	0.997	0.999	1.000
	0.60	0.984	0.998	1.000	0.984	0.994	1.000	0.984	0.994	1.000	0.989	0.999	1.000	0.993	0.999	1.000
	0.65	0.990	0.997	1.000	0.971	0.991	1.000	0.971	0.990	1.000	0.977	0.995	0.999	0.982	0.995	0.999
	0.70	0.986	0.993	0.997	0.958	0.981	0.996	0.959	0.980	0.995	0.961	0.979	0.999	0.957	0.982	0.999
	0.75	0.972	0.988	0.997	0.928	0.966	0.991	0.927	0.967	0.991	0.943	0.968	0.992	0.937	0.967	0.993
	0.80	0.939	0.974	0.996	0.885	0.943	0.985	0.885	0.943	0.986	0.893	0.949	0.984	0.896	0.944	0.986
	0.85	0.893	0.938	0.985	0.835	0.903	0.973	0.835	0.901	0.973	0.837	0.905	0.972	0.829	0.904	0.973
	0.90	0.840	0.900	0.970	0.785	0.853	0.956	0.785	0.853	0.956	0.767	0.854	0.955	0.763	0.851	0.953
	0.95	0.742	0.839	0.940	0.710	0.805	0.924	0.711	0.805	0.925	0.691	0.782	0.912	0.684	0.782	0.912
100	0.40	0.682	0.850	0.991	0.947	0.986	1.000	0.946	0.987	1.000	0.970	0.996	1.000	0.987	0.999	1.000
	0.45	0.753	0.913	0.995	0.953	0.990	1.000	0.953	0.990	1.000	0.983	0.997	1.000	0.995	1.000	1.000
	0.50	0.902	0.976	0.999	0.980	0.995	1.000	0.980	0.995	1.000	0.993	1.000	1.000	0.997	1.000	1.000
	0.55	0.955	0.989	1.000	0.988	0.995	1.000	0.988	0.995	1.000	0.994	1.000	1.000	0.998	0.999	1.000
	0.60	0.977	0.994	1.000	0.988	0.996	1.000	0.988	0.996	1.000	0.990	0.998	1.000	0.993	0.998	1.000
	0.65	0.987	0.999	1.000	0.976	0.994	1.000	0.975	0.994	1.000	0.984	0.993	1.000	0.985	0.993	1.000
	0.70	0.988	0.994	1.000	0.951	0.982	0.997	0.952	0.981	0.997	0.957	0.984	0.997	0.953	0.984	0.998
	0.75	0.975	0.989	0.999	0.914	0.961	0.994	0.915	0.961	0.994	0.916	0.962	0.996	0.923	0.960	0.996
	0.80	0.938	0.974	0.994	0.880	0.924	0.986	0.880	0.923	0.986	0.862	0.929	0.985	0.862	0.932	0.986
	0.85	0.873	0.933	0.983	0.830	0.882	0.969	0.831	0.883	0.970	0.785	0.871	0.963	0.781	0.866	0.964
	0.90	0.800	0.871	0.961	0.764	0.843	0.929	0.765	0.843	0.929	0.696	0.793	0.929	0.688	0.794	0.934
	0.95	0.697	0.803	0.918	0.689	0.781	0.894	0.684	0.781	0.893	0.610	0.724	0.879	0.588	0.705	0.876
200	0.40	0.557	0.789	0.986	0.869	0.984	1.000	0.871	0.984	1.000	0.986	0.999	1.000	0.997	1.000	1.000
	0.45	0.686	0.874	0.994	0.918	0.994	1.000	0.918	0.994	1.000	0.996	1.000	1.000	1.000	1.000	1.000
	0.50	0.860	0.955	1.000	0.973	0.998	1.000	0.973	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	0.55	0.926	0.986	1.000	0.987	0.998	1.000	0.987	0.998	1.000	0.997	1.000	1.000	0.999	1.000	1.000
	0.60	0.971	0.997	1.000	0.989	0.998	1.000	0.991	0.998	1.000	0.992	0.998	1.000	0.993	0.999	1.000
	0.65	0.989	0.997	1.000	0.983	0.997	1.000	0.983	0.997	1.000	0.965	0.990	1.000	0.970	0.991	1.000
	0.70	0.990	0.996	0.999	0.979	0.990	0.998	0.977	0.990	0.998	0.928	0.970	0.997	0.933	0.974	0.998
	0.75	0.972	0.989	0.996	0.949	0.978	0.997	0.949	0.978	0.997	0.872	0.932	0.992	0.885	0.936	0.993
	0.80	0.941	0.969	0.994	0.900	0.957	0.988	0.901	0.955	0.988	0.816	0.886	0.972	0.823	0.889	0.971
	0.85	0.882	0.938	0.982	0.869	0.905	0.979	0.869	0.906	0.979	0.740	0.825	0.930	0.743	0.829	0.933
	0.90	0.787	0.871	0.955	0.806	0.878	0.954	0.805	0.879	0.953	0.642	0.742	0.892	0.641	0.748	0.894
	0.95	0.664	0.777	0.902	0.743	0.803	0.917	0.744	0.803	0.918	0.553	0.658	0.822	0.538	0.656	0.822

The linear equivalents of models (20) and (21), respectively, are

$$\begin{aligned}
\widehat{\theta}_{j_0}^{0*} = \max\ & \sum_{q=1}^{Q} \mu'_q z_{qj_0} + \sum_{r=1}^{R} \gamma'_r l_{rj_0} + u^1 + \sum_{s=1}^{S} \eta'_s y_{sj_0} + u^2 \\
\text{s.t.}\ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} + \sum_{q=1}^{Q} \mu'_q z_{qj_0} + \sum_{t=1}^{T} \pi'_t g_{tj_0} = 1, \\
& \sum_{q=1}^{Q} \mu'_q z_{qj} + \sum_{r=1}^{R} \gamma'_r l_{rj} - \sum_{p=1}^{P} \upsilon'_p x_{pj} + u^1 \le 0, \quad j = 1, \dots, N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} - \sum_{q=1}^{Q} \mu'_q z_{qj} - \sum_{t=1}^{T} \pi'_t g_{tj} + u^2 \le 0, \quad j = 1, \dots, N \\
& \upsilon'_p, \gamma'_r, \mu'_q, \pi'_t, \eta'_s > 0, \\
& u^1, u^2 \text{ free in sign.}
\end{aligned}$$

and

$$\begin{aligned}
\widehat{\theta}_{j_0}^{1p*} = \max\ & \sum_{q=1}^{Q} \mu'_q z_{qj_0} + \sum_{r=1}^{R} \gamma'_r l_{rj_0} + u^1 \\
\text{s.t.}\ & \sum_{p=1}^{P} \upsilon'_p x_{pj_0} = 1 \\
& \sum_{q=1}^{Q} \mu'_q z_{qj} + \sum_{r=1}^{R} \gamma'_r l_{rj} + u^1 - \sum_{p=1}^{P} \upsilon'_p x_{pj} \le 0, \quad j = 1, \dots, N \\
& \sum_{s=1}^{S} \eta'_s y_{sj} + u^2 - \sum_{q=1}^{Q} \mu'_q z_{qj} - \sum_{t=1}^{T} \pi'_t g_{tj} \le 0, \quad j = 1, \dots, N \\
& (1 - \widehat{\theta}_{j_0}^{0*}) \sum_{q=1}^{Q} \mu'_q z_{qj_0} - \widehat{\theta}_{j_0}^{0*} \sum_{t=1}^{T} \pi'_t g_{tj_0} + \sum_{r=1}^{R} \gamma'_r l_{rj_0} \\
& + \sum_{s=1}^{S} \eta'_s y_{sj_0} + u^1 + u^2 = \widehat{\theta}_{j_0}^{0*}
\end{aligned}$$

$$\nu'_p, \gamma'_r, \mu'_q, \pi'_t, \eta'_s > 0,$$
$$u^1, u^2 \text{ free in sign.}$$

List of notations:	
j	Index of DMUs.
j_0	DMU under evaluation.
N	Number of DMUs in the sample.
m	Number of DMUs in the subsample
ϕ	Stage number indicator
$x_{pj} = (x_{1j}, \ldots, x_{Pj})$	Vector of first stage inputs for DMU$_j$.
$l_{rj} = (l_{1j}, \ldots, l_{Rj})$	Vector of first stage outputs for DMU$_j$.
$z_{qj} = (z_{1j}, \ldots, z_{Qj})$	Vector of intermediate products for DMU$_j$.
$g_{tj} = (g_{1j}, \ldots, g_{Tj})$	Vector of second stage inputs for DMU$_j$.
$y_{sj} = (y_{1j}, \ldots, y_{Sj})$	Vector of second stage outputs for DMU$_j$.
x_e	Vector of first stage efficient inputs
g_e	Vector of second stage efficient inputs
$\nu_p = (\nu_1, \ldots, \nu_P)$	Vector of multipliers for the first stage inputs in the fractional model.
$\mu_r = (\mu_1, \ldots, \mu_R)$	Vector of multipliers for the first stage outputs in the fractional model.
$\gamma_q = (\gamma_1, \ldots, \gamma_Q)$	Vector of multipliers for the intermediate products in the fractional model.
$\pi_t = (\pi_1, \ldots, \pi_T)$	Vector of multipliers for the second stage inputs in the fractional model.
$\eta_s = (\eta_1, \ldots, \eta_S)$	Vector of multipliers for the second stage outputs in the fractional model.
$\nu'_p = (\nu'_1, \ldots, \nu'_P)$	Vector of multipliers for the first stage inputs in the linear model.
$\mu'_r = (\mu'_1, \ldots, \mu'_R)$	Vector of multipliers for the first stage outputs in the linear model.
$\gamma'_q = (\gamma'_1, \ldots, \gamma'_Q)$	Vector of multipliers for the intermediate products in the linear model.
$\pi'_t = (\pi'_1, \ldots, \pi'_T)$	Vector of multipliers for the second stage inputs in the linear model.
$\eta'_s = (\eta'_1, \ldots, \eta'_S$	Vector of multipliers for the second stage outputs in the linear model.
w_{1j}	Decomposition weight of the first stage.
w_{2j}	Decomposition weight of the second stage.
θ^0_j	True overall efficiency score of DMU$_j$.
θ^1_j	True first stage efficiency score of DMU$_j$.
θ^2_j	True second stage efficiency score of DMU$_j$.
$\widehat{\theta^0}_j$	Estimation of the overall efficiency score of DMU$_j$.
$\widehat{\theta^1}_j$	Estimation of the first stage efficiency score of DMU$_j$.
$\widehat{\theta^2}_j$	Estimation of the second stage efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^0}}_j$	Bootstrap estimation of the overall efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^1}}_j$	Bootstrap estimation of the first stage efficiency score of DMU$_j$.
$\widehat{\widehat{\theta^2}}_j$	Bootstrap estimation of the second stage efficiency score of DMU$_j$.

References

Banker, R. (1993). Maximum likelihood, consistency and data envelopment analysis: A statistical foundation. *Management Science, 39*(10), 1265–1273.
Banker, R., Charnes, A., & Cooper, W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science, 30*(9), 1078–1092.
Bickel, P., & Freedman, D. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics, 9*(6), 1196–1217.
Bostian, M., Daraio, C., Färe, R., Grosskopf, S., Izzo, M., Leuzzi, L., … Weber, W. (2018). Inference for nonparametric productivity networks: a pseudo–likelihood approach. *Technical Report 6*. Sapienza, Università di Romà.
Charnes, A., & Cooper, W. (1962). Programming with linear fractional functionals. *Naval Research Logistics Quarterly, 9*(3–4), 67–88.
Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
Chen, Y., Cook, W., Kao, C., & Zhu, J. (2013). Network DEA pitfalls: Divisional efficiency and frontier projection under general network structures. *European Journal of Operational Research, 226*(3), 507–515.
Chen, Y., Cook, W., Li, N., & Zhu, J. (2009). Additive efficiency decomposition in two-stage DEA. *European Journal of Operational Research, 196*(3), 1170–1176.
Chen, Y., Cook, W., & Zhu, J. (2010). Deriving the DEA frontier for two-stage processes. *European Journal of Operational Research, 202*(1), 138–142.
Cook, W., Zhu, J., Bi, G., & Yang, F. (2010). Network DEA: Additive efficiency decomposition. *European Journal of Operational Research, 207*(2), 1122–1129.
Dia, M., Takouda, P., & Golmohammadi, A. (2020). Assessing the performance of Canadian credit unions using a three-stage network bootstrap DEA. *Annals of Operations Research*.
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics, 7*(1), 1–26.
Farrell, M. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society, Series A (General), 120*(3), 253–281.
Färe, R., & Grosskopf, S. (1996). Productivity and intermediate products: A frontier approach. *Economics Letters, 50*(1), 65–70.
Färe, R., & Whittaker, G. (1995). An intermediate input model of dairy production using complex survey data. *Journal of Agricultural Economics, 46*(2), 201–213.
Gijbels, I., Mammen, E., Park, B., & Simar, L. (1999). On estimation of monotone and concave frontier functions. *Journal of the American Statistical Association, 94*(445), 220–228.
Jeong, S., & Park, B. (2006). Large sample approximation of the distribution for convex-hull estimators of boundaries. *Scandinavian Journal of Statistics, 33*(1), 139–151.
Kao, C., & Hwang, S. (2008). Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *European Journal of Operational Research, 185*(1), 418–429.
Kneip, A., Park, B., & Simar, L. (1998). A note on the convergence of nonparametric DEA efficiency measures. *Econometric Theory, 14*(6), 783–793.
Kneip, A., Simar, L., & Wilson, P. (2008). Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. *Econometric Theory, 24*(6), 1663–1697.
Kneip, A., Simar, L., & Wilson, P. (2011). A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators. *Computational Economics, 38*(4), 483–515.
Marchetti, D., & Wanke, P. (2017). Brazil's rail freight transport: Efficiency analysis using two-stage DEA and cluster-driven public policies. *Socio-Economic Planning Sciences, 59*, 26–42.
Michali, M., Emrouznejad, A., Dehnokhalaji, A., & Clegg, B. (2021). Noise-pollution efficiency analysis of European railways: A network DEA model. *Transportation Research Part D: Transport and Environment, 98*, 102980.
Moradi-Motlagh, A., & Emrouznejad, A. (2022). The origins and development of statistical approaches in non-parametric frontier models: A survey of the first two decades of scholarly literature (1998–2020). *Annals of Operations Research*. https://doi.org/10.1007/s10479-022-04659-7.
Nwaogbe, O., Wanke, P., Barros, I., & Azad, A. (2017). Efficiency driver in Nigerian airports: A bootstrap DEA-censored quantile regression approach. *Socio-Economic Planning Sciences, 59*, 26–42.
Park, B., Jeong, S., & Simar, L. (2010). Asymptotic distribution of conical-hull estimators of directional edges. *Annals of statistics, 38*(3), 1320–1340.
Politis, D., Romano, J., & Wolf, M. (2001). On the asymptotic theory of subsampling. *Statist Sinica, 11*(4), 1105–1124.
Shephard, R. (1970). *The theory of cost and production functions*. Princeton: Princeton University Press.
Simar, L., & Wilson, P. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science, 44*(1), 49–61.
Simar, L., & Wilson, P. (1999). Estimating and bootstrapping Malmquist indices. *European Journal of Operational Research, 115*(3), 459–471.
Simar, L., & Wilson, P. (2000). A general methodology for bootstrapping in non-parametric frontier models. *Journal of Applied Statistics, 27*(6), 779–802.
Simar, L., & Wilson, P. (2002). Non-parametric tests of returns to scale. *European Journal of Operational Research, 139*(1), 115–132.
Simar, L., & Wilson, P. (2008). Statistical inference in nonparametric frontier models: Recent developments and perspectives. In I. H. Fried, C. A. K. Lovell, & S. S. Schmidt (Eds.), *The measurement of productive efficiency and productivity change*. Oxford Scholarship Online.
Simar, L., & Wilson, P. (2010). Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis, 36*(1), 33–53.
Swanepoel, J. (1986). A note on proving that the (modified) bootstrap works. *Communications in Statistics: Theory Methods, 15*, 3193–3203.
Trinh, K., & Zelenyuk, V. (2015). Bootstrap-based testing for network DEA: Some theory and applications. *Working Paper Series WP05/2015*. Australia: School of Economics, University of Queensland.
Wanke, P. (2012). Capacity shortfall and efficiency determinants in Brazilian airports: Evidence from bootstrapped DEA estimates. *Socio-Economic Planning Sciences, 46*, 216–229.
Zha, Y., & Liang, L. (2010). Two-stage cooperation model with input freely distributed among the stages. *European Journal of Operational Research, 205*(2), 332–338.