Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 48 (2020) 3090–3106

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research – WCTR 2019, Mumbai, 26-30 May 2019

Framework for logistics performance index construction using DEA: an application for soybean haulage in Brazil

Isotilia Costa Melo[a*], Thiago Guilherme Péra[b], Paulo Nocera Alves Júnior[b], Daisy Aparecida do Nascimento Rebelatto[a], José Vicente Caixeta-Filho[b]†

[a]*São Carlos Engineering School (EESC), University of São Paulo (USP), Av. Trabalhador Sancarlense, 400, São Carlos – SP, 13566-590, Brazil*
[be]*Luiz de Queiroz College of Agriculture (ESALQ), University of São Paulo (USP), Avenida Pádua Dias, 11 – Piracicaba –SP, 13418-900, Brazil*

Abstract

Freight transportation is vital to a nation's long-term development and its performance needs to be carefully evaluated to ensure the effectiveness, efficiency, and equity of haulage infrastructure decisions. The previous attempts at benchmarking the transportation corridors and route efficiency through Data Envelopment Analysis (DEA) models violated homogeneity assumptions or did not provide an appropriate robustness analysis. The present paper integrates index creation and general DEA guidelines, and proposes a framework for the creation of a long-distance cargo haulage performance index, advancing towards the limitations of the previous efforts. The methodology is applied to the context of soybean transportation, one of the relevant Brazilian exporting products, during the harvest of 2015/2016, from the main mid-sized producing regions to the key exporting ports, by land transportation. The proposed approach and findings can provide insights into public and private long-term investment strategies and infrastructure policies in Brazil and other developing countries in a similar context.

Keywords: Data Envelopment Analysis (DEA), Slack-based model (SBM), performance index, freight tranportation, soybean, Brazil

* Corresponding author. *E-mail address: isotilia@usp.br*
* Corresponding author. Tel: +55 16 3373 9425; fax: +55 16 3373 9425.
E-mail address: isotilia@usp.br

10.1016/j.trpro.2020.08.179

1. Introduction

The transportation of cargo is one of the main pillars to a nation's long-term development and its performance needs to be carefully evaluated to ensure the efficiency and equity of haulage infrastructure decisions. The existence of a consolidated performance measurement methodology is a tool which provides the bases for an effective intervention of decision- and policy-makers in the regard of reducing national and regional inefficiencies.

In any system investigation, it is crucial to have an understanding of the whole process. For reducing subjectivity, a clear specification of the target to be investigated (based on previous literature and specialists' guidelines) is the principal point of discussion to guide the right choice and classification of the variables, besides the appropriate choice of the DEA model and its orientation, as consolidated by Cook, Tone and Zhu (2014). The previous applications of DEA for a context of freight efficiency investigation is scarce and thus far the literature on the topic does not appear to be consolidated, the known exceptions are Melo et al. (2018) and Oliveira and Cicolin (2016).

Melo et al. (2018) applied a type of DEA model, named slack-based model (SBM), for 102 soybean haulage routes (technically, each route is called decision-making unit, or simply DMU) in Brazil and in the USA, using 9 variables, based on data availability and literature review, and classified into *inputs*, *outputs*, *undesirable outputs*, and *uncontrollable variables*. The authors also integrated the SBM model with the tie-breaking method of the composite index for the final performance index generation. Though this paper may have extrapolated the assumption of homogeneity among DMUs, jointly comparing two different countries and water, rail and road routes. It would be recommended to use a non-homogeneous DEA model for such wide benchmark. In this regard, the current paper is focused on one unique country (Brazil) and one unique mode of transportation (road).

The same authors compared the result of their proposed model with technical international reports to demonstrate the most efficient DMUs pointed by SBM model are also pointed by the literature. Greco et al. (2018), focused on examining the variety of existing methodological approaches for composite indexes, emphasized the importance of a robustness test after the index creation. In this regard, the present paper integrates the sensitivity analysis for verifying the robustness of the results.

Aside from the application for the Brazilian context of road soybean haulage, this paper aims to provide guidelines for a logistics performance index (using DEA), in the format of a framework, that can be replicated in other contexts and can be used for directing infrastructural investment focus to those characteristics and units that are mostly impacting in inefficiency.

2. Background

2.1 The grain's profile logistics in Brazil

According to the Planning and Logistics Company (EPL, 2018), in Brazil, transportation infrastructure offers 1.563 million kilometers of roads (only 13.5% are paved), 30 thousand kilometers of railways (only one third in commercial operation), 41.6 thousand kilometers of navigable waterways (22 thousand kilometers of economically navigable routes). Specifically, for solid agricultural bulk, the estimated transport matrix is 60% road, 30% rail and 10% waterway.

In line with National Land Transportation Agency (ANTT, 2018), in 2010, grain production in Brazil reached 124.7 million tons, while railroad grain traffic was around 22.3 million tons - equivalent to 17.9% and the movement in the waterway was around 4.3 million tons - equivalent to 3.5% of the production. In 2017, the grain production increased to 211.8 million tons, grain movement in the railroad increased to 42.2 million tons (19.9% of production) and waterway to 16.9 million tons (8.0% of the production).

2.2 Index creation and the DEA literature

According to Greco et al. (2018), approaches for index construction may permit the *subjectivity* of letting the weight attributions to the decision-makers or may apply statistical methods - Data Envelopment Analysis (DEA), Principal Component Analysis (PCA), Factor Analysis (FA) - that automatically attribute weights, allegedly reducing arbitrariness. The goal of the current paper is the construction of an index without human weight attributions, and DEA is the only listed method exclusively focused on efficiency measurement.

The DEA is a non-parametric tool, created by Charnes, Cooper and Rhodes (1978), that compares decision-making units (DMUs), building a frontier of efficiency (*best-practice frontier*), based on a group of chosen variables (that represent the criteria of investigation interest and may be classified, at least, into *input*s and *outputs*). The first DEA model was named CCR, an acronym of its creators' family names, Charnes, Cooper and Rhodes (1978). The CCR model was constant to scale. In sequence, the BCC model was created by Banker et al. (1989), with variable scale. In both models, it is mandatory to choose an orientation: input minimization or output maximization. The additive DEA models permit simultaneous orientation. Among these kinds of models, the slack-based model (SBM), proposed by Tone (2001), returns directly to an efficiency rank, facilitating the interpretation. The SBM can have constant or variable to scale, depending on the interest of the analysis.

The DEA is based on the assumption of the homogeneity of the DMUs, i.e., the DMUs must be minimally homogeneous to be compared. The limits acceptable heterogeneity in a population still remain as a theme of debate. Although this research topic is beyond the scope of the present paper, one viable solution for the problem may be the adoption of a non-homogeneous DEA model, as proposed by Li et al. (2016).

One of the DEA pitfalls is the misjudgment of efficiency when inputs and outputs simultaneously deal with ratio and raw data. Cook, Tone and Zhu (2014) state the coexistence of the two types of data in the same DEA model is permissible under certain circumstances. In this way, the current paper does not assume this restriction as a condition for the index construction.

It is well known that the number of variables compared to the number of DMUs may also reduce the discrimination power of the DEA analysis. Banker et al. (1989) recommend that the number of DMUs may be, at least, three times the number of variables, independently on the type of DEA model. However, this rule is neither imperative nor based on a statistical evaluation, though accepted by convenience. In the effort for reducing the number of variables in DEA models, other techniques were incorporated to DEA, such as PCA proposed by Adler and Golany (2007). Cook, Tone and Zhu (2014) state that it is not mandatory to limit the number of variables, though the current paper assumed "1/3 of the number of DMUs = number of variables" as a desirable target for the index construction.

According to the same authors, DEA can be viewed as a tool for multi-criteria evaluation problem where DMUs are alternatives and each DMU performance is attributed by variables classified into, at least, inputs and outputs. In this regard, the concept of input and output may slightly differ from the first interpretation of a reader familiar to other statistical tools, such as linear regressions.

When DEA applied for benchmarking, the inputs are normally understood as the criteria desired to be minimized to improve efficiency. On the other hand, the outputs are the criteria desired to be maximized in the benefit of efficiency. In this way, there may be no implicit or explicit relationship of cause and consequence between inputs and outputs of a system.

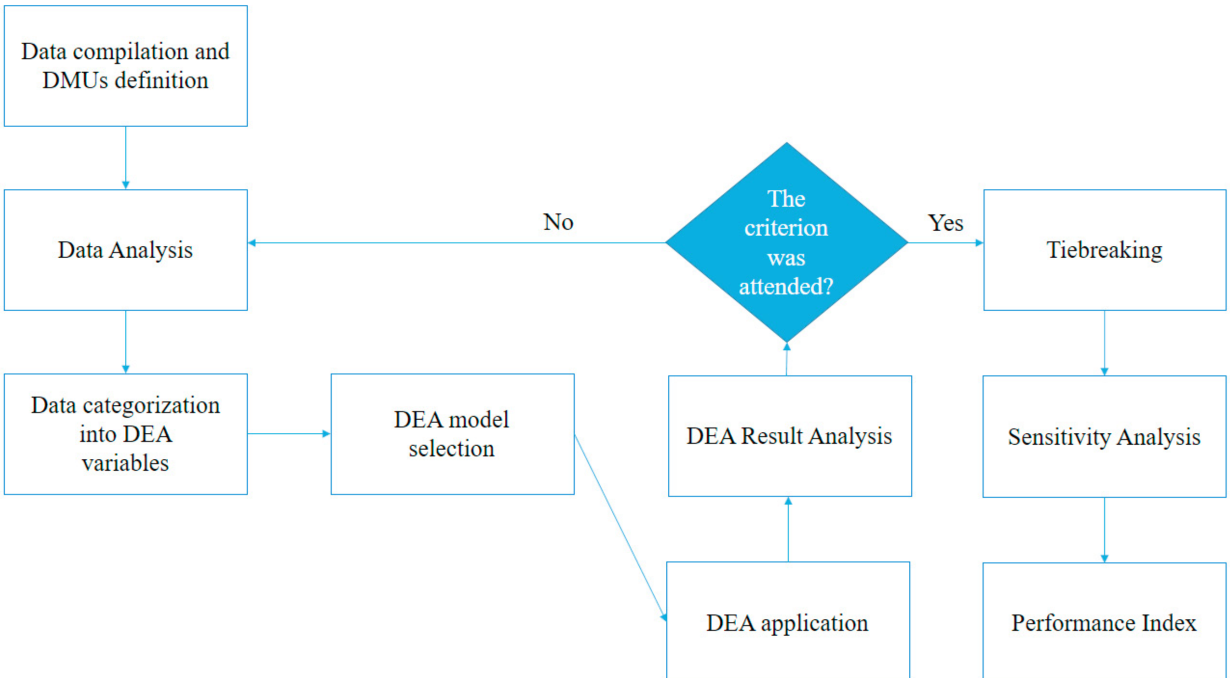
According to (Greco et al. 2018), the criticism on DEA application remains on the fact that, although the weights are not attributed by the decision-makers, the choice of variables remains on the subjectivity of those who designed the index. For reducing the subjectivity of categorization in DEA, one of the first proposed approaches was the use of correlation, as stated by Golany and Roll (1989). Inputs might be strongly correlated to outputs and weakly correlated with each other. Highly correlated inputs may imply in redundant variables. The same logic might be applied to outputs. It is important to remember that this argumentation is valid only if the correlation presents an accepted level of significance, otherwise, nothing can be stated. In parallel, correlation is also one of the tools pointed by Greco et al. (2018) for selecting variables in an index creation (independently on the chosen index creation method).

In addition, there are variables that can be classified as undesirable outputs, e.g., pollutions. The discussion about the possibility of treatments for those variables by DEA models goes beyond the scope of the present paper (an interested reader may consult Liu et al. (2010), Hua and Bian (2007), and Seiford and Zhu (2002)). For this context, it is relevant to mention that undesirable outputs may be inserted as inputs (for the benefit of minimization) or their inverse may be inserted as outputs (in this way that their maximization represents an actual value reduction).

There are also variables that affect the system and must be counted for efficiency, though they can be hardly changed by an intentional human effort (for example, the distance between two cities). These variables are called non-discretionary or non-controllable. The way of incorporation of this kind of a variable to a DEA model may differ depending to the analyzed case. Melo et al. (2018) incorporated the concepts of non-discretion of Saen (2005) to the SBM model proposed by Tone (2001). The present paper adopted the same approach, explicit in topic 3. Methodology. The current paper contributes to the advance of literature adding a sensitivity analysis to the approach, and showing an explicit framework for the logistics performance index construction.

3. Methodology

Schema 1 shows the proposed framework for the construction of a freight corridor performance index, combining jointly the recommendation of Greco et al. (2018) for composite indicators in general and the guidelines for DEA application of Cook, Tone and Zhu (2014). The criterion for continuing on the loop of analysis or tiebreaking, depends on the system under analysis, the goals of the analysis, the literature review and/or the opinion of the specialists. The present paper adopted that the best desired DEA result analysis was represented by a model with the maximum incorporated characteristics (to be presented in subtopic 3.2 Data analysis) and a rank of efficiency with the minimal number of ties.



Schema 1 The proposed framework for a logistics performance index construction applying DEA.

3.1 Data compilation and DMUs definition

The data of the harvest of 2015/2016 were supplied by ESALQ-LOG (2018), the official agricultural logistics database for Brazil. There were routes with exclusive road transportation alternatives and routes by waterway and railway haulage. Due to the assumption of homogeneity of DEA, only DMUs with exclusively road transportation were analyzed, once this mode is the most frequent case in Brazil. For simplification, the present paper considers route, corridors, and DMUs as synonymous.

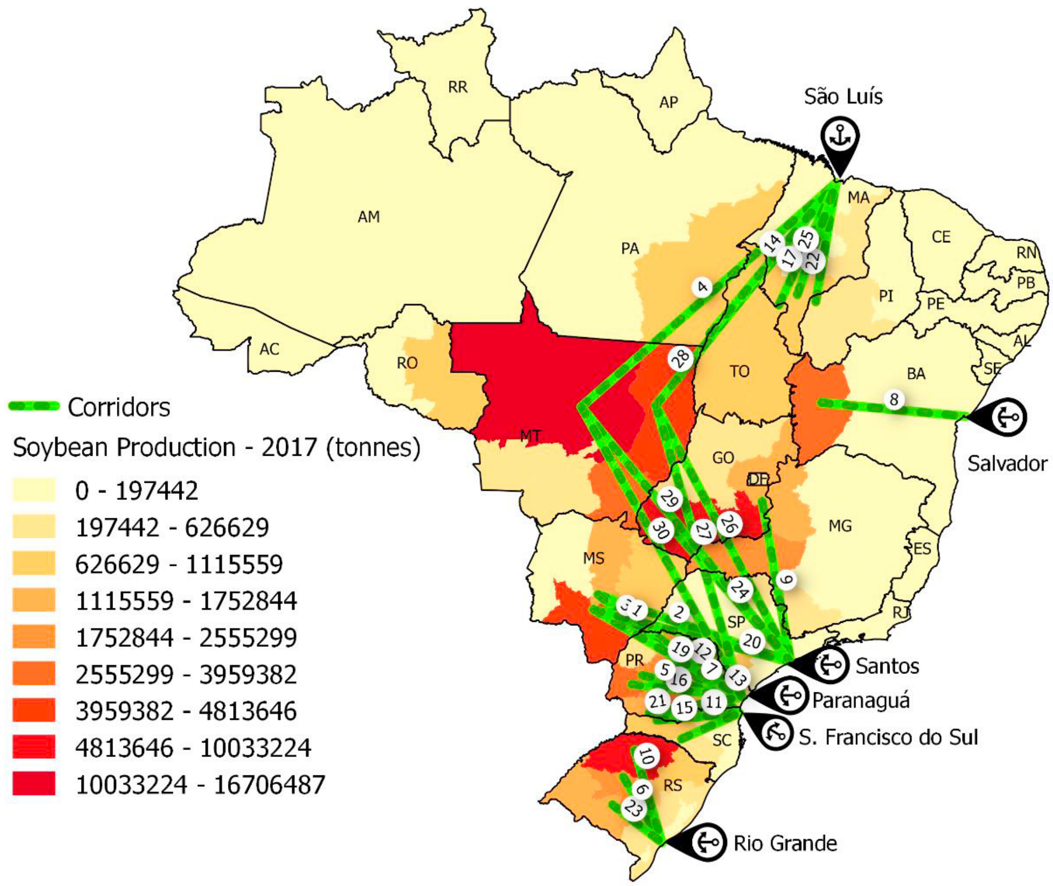


Fig. 1. Map of Brazil with selected DMUs (from the mid-sized regions to the exporting ports).

The Brazilian Institute of Geography and Statistics (IBGE, 2018) officially classifies Brazilian intra-estate areas, from lowest to highest level, into micro-, meso-, and macro-regions. In this paper, the term mid-sized region is adopted as a synonymous of the Portuguese technical term *meso-region*, used by IBGE (2018). Fig. 1 shows the state borders with black solid lines and analyzed mid-sized regions in different colors (inside the state area).

Considering the relevant mid-sized regions of the 10 main producing states - i.e., Bahia (BA), Goiás (GO), Maranhão (MA), Mato Grosso (MT), Mato Grosso do Sul (MS), Piauí (PI), Paraná (PR), Rio Grande do Sul (RS), Santa Catarina (SC), and Tocantins (TO) -, and six national exporting ports - i.e., Salvador (BR SSA), Santos (BR SSZ), São Luiz (BR SLZ), Paranaguá (BR PNG), São Francisco do Sul (BR SFS), Rio Grande (BR RIG) -, 30 routes (DMUs) were identified. Fig. 1 also presents an illustrative representation of the DMUs from the mid-sized regions of origin to the ports.

3.2 Data analysis

As there were 30 DMUs, the target for the total of variables of a DEA model was 10 (i.e., 1/3). First, the available variables were organized by the ESALQ-LOG specialists according to the main five characteristics (economic, operational, environmental, infrastructural, and productive). It was desirable to have at least one variable of each characteristic in the model. For providing conditions for results replicability, Table 1 shows variables and their descriptive statistics.

Table 1. Descriptive statistics of available variables, followed by their units and ESALQ-LOG specialists' classification.

Characteristics	Variable	Obs.	Mean	Std. Dev.	Min	Max
Economic	Freight price (R$/t)	30	121.87	64.72	27.30	269.88
	Logistic loss (t)	30	73,561.69	82,052.68	9,236.80	307,516.00
	Fuel consumption (km/l)	30	1.91	0.13	1.77	2.15
	Length of the route (km)	30	982.40	581.55	191.59	2,299.25
Operational	Travel time (h)	30	16.17	10.29	3.02	45.80
	Average speed (km/h)	30	61.88	8.67	50.20	89.36
Environmental	Emissions (kg of CO_2/t of transported soybean)	30	1,383.54	777.13	287.01	3,291.49
Infrastructural	On-farm storage capacity (%)	30	0.12	0.12	0.00	0.52
	Off-farm storage capacity (%)	30	0.58	0.39	0.15	2.16
	Total storage capacity (%)	30	0.70	0.46	0.26	2.68
	On-farm storage capacity (t)	30	1,266,143.00	2,497,902.00	0.00	8,539,456.00
	Off-farm storage capacity (t)	30	4,156,065.00	5,102,062.00	381,571.00	17,400,000.00
	Total storage capacity (t)	30	5,422,208.00	7,440,085.00	474,300.00	25,900,000.00
	Port representativeness (%)	30	0.58	0.30	0.06	1.00
	Corridor exports (t)	30	1,010,714.00	1,276,207.00	155,428.80	6,312,151.00
	Port capacity (t)	30	8,289,251.00	3,290,840.00	2,693,166.00	13,000,000.00
	Port capacity related to mid-sized regional production (%)	30	3.91	2.75	0.28	9.42
Productive	Production (t)	30	4,279,674.00	4,949,615.00	490,028.00	17,700,000.00
	Percentage of the mid-sized regional production related to the state (%)	30	0.39	0.28	0.08	1.00
	Ratio between production of soybean and production of other grains (%)	30	0.58	0.14	0.35	0.90

3.3 Data categorization

The software STATA was applied to build the correlation matrix with all available variables. Due to this paper format restrictions, Table 2 shows a matrix exclusively with those variables that present accepted significant correlations or were judged relevant by the specialists (i.e., *corridor exportation* and *inverted emissions*).

In this context, for the first trial of the model (Model 1), freight price, logistic loss, and fuel consumption were considered inputs. The minimization of these variables was a goal due to economic reasons. The emissions are a sub-product of the system, once it will be generated even by efficient DMUs, though it is desired to be the least possible, hence it is considered an *undesirable output*. Among the possible treatments, it was chosen to insert emissions, initially, as an input (for minimization).

Theoretically, the minimal length of the route, the best for the freight (in other words, it may represent reduced freight price and travel time). However, the power of intervention of policy-makers in actually reducing the physical length of a route may be limited by the physical, economic, social, and cultural barriers. Due to this limitations, Melo et al. (2018) considered the length of the route as a *non-discretionary variable* and the present paper adopted the same procedure.

On-farm storage capacity and production were considered outputs because Brazil presents insufficient warehousing infrastructure. Once the farmers have few storage options, they are pressed to the sale and transport their production to the exporting ports during the harvest, causing congestions and traffic problems. Given this context, it is desirable to maximize on-farm storage capacity. The maximization of soybean production is an economic goal. It assumed the greatest the production, the greatest is the haulage, once it was not possible to measure the haulage directly.

Table 2. Correlation matrix of chosen variables due to significance and attributed relevance for the analysis.

	Freight price (R$/t)	**Logistic loss (t)**	**Fuel consumption (km/l)**	**On-farm storage capacity (t)**	**Emissions (Eq. CO2/transported t)**	**Length of the route (km)**	**Production (t)**	**Corridor exports (t)**	**Inverted Emissions**
Freight price (R$/t)	1.0000								
Logistic loss (t)	0.6020	1.0000							
	0.0004								
Fuel consumption (km/l)	0.5766	0.2773	1.0000						
	0.0009	0.1379							
On-farm storage capacity (t)	0.6332	0.9339	0.3339	1.0000					
	0.0002	0.0000	0.0713						
Emissions (Eq. CO2/transported t)	0.9552	0.6963	0.5761	0.7081	1.0000				

	0.0000	0.0000	0.0009	0.0000					
Length of the route (km)	0.9513	0.6914	0.6514	0.7126	0.9946	1.0000			
	0.0000	0.0000	0.0001	0.0000	0.0000				
Production (t)	0.6159	0.9978	0.3103	0.9482	0.7117	0.7104	1.0000		
	0.0003	0.0000	0.0951	0.0000	0.0000	0.0000			
Corridor exports (t)	-0.0362	0.4537	0.0376	0.2365	-0.0190	-0.0029	0.4289	1.0000	
	0.8493	0.0118	0.8438	0.2084	0.9205	0.9880	0.0180		
Inverted Emissions	-0.7653	-0.4261	-0.5403	-0.3624	-0.7873	-0.7825	-0.4342	-0.0858	1.0000
	0.0000	0.0189	0.0021	0.0491	0.0000	0.0000	0.0165	0.6520	

3.4 DEA model selection, application and loop investigations

The software MatLab was used to execute the SBM model and the SBM super-efficiency analysis (to be presented in the subtopic 3.6 Sensitivity analysis: SBM Super-efficiency). The choice of the model (SBM) was due to its characteristic of simultaneously maximize outputs and minimize inputs and its previous usage in similar contexts by Melo et al. (2018).

The SBM formulation with non-discretionary variables, as proposed by Saen (2005), is linearly programmed according to Equations 1 and 2, and subject to restrictions represented by Equations 3, 4, 5, 6, and 7, as stated by Tone (2001):

$$\text{Minimize } \tau = t - \left(\frac{1}{m}\right) \sum_{i=1}^{m} \frac{S_i^-}{x_{i0}} \tag{1}$$

$$\left(\frac{1}{s}\right) \sum_{r=1}^{s} \frac{S_r^+}{y_{r0}} + t = 1 \tag{2}$$

$$\sum_{i=1}^{m} \Lambda_k x_{ik} + S_i^- - t x_{i0} = 0 \quad k=1,2,\ldots, z \tag{3}$$

$$\sum_{i=1}^{m} \Lambda_k y_{rk} + S_r^+ - t y_{r0} = 0 \quad k=1,2,\ldots, z \tag{4}$$

$$S_i^- \leq \beta_i x_{i0} \quad i=1,2, \ldots, p \tag{5}$$

$$S_r^+ \leq \gamma_r y_{r0} \quad r=1,2, \ldots, q \tag{6}$$

$$\Lambda_k \geq 0,\ S_i^- \geq 0, S_r^+ \geq 0 \text{ and } t>0 \tag{7}$$

Where:
τ: is the efficiency.
S_i^-: is the slack of the ith input.
S_r^+: is the slack of the rth output.
Λ_k: is the contribution of the kth DMU to the analyzed DMU.
t: is the model linearization factor.

x_{i0}: is the ith input of the DMU under analysis.
x_{ik}:is the ith input of the kth DMU.
y_{rk}: is the rth output of the kth DMU.
m: is the number of inputs.
s: is the number of outputs.
z: is the number of DMUs.
p: is the number of non-discretionary inputs.
q: is the number of non-discretionary outputs.
β_i:is a constant of discretion for inputs (it from 0 to infinite, 0 represents a totally non-controllable input and infinite represents a totally controllable input, i.e. a standard SBM model).
γ_r: is a constant of discretion for outputs (it from 0 to infinite, 0 represents a totally non-controllable output and infinite represents a totally controllable output, i.e. a standard SBM model).

As the present application requires a variable return of scale, it was necessary to add an additional restriction, according to the Equation 8.

$$\sum_{k=1}^{z} \Lambda_k = 1 \tag{8}$$

The optimum solution $(\rho^*, t^*, \Lambda_k^*, S_i^{-*}, S_r^{+*})$ is described by the conditions in Equation 9:

$$\rho^* = \tau^*,\ \lambda_k^* = \frac{\Lambda_k^*}{t^*},\ s_i^{-*} = \frac{S_i^{-*}}{t^*},\ s_r^{+*} = \frac{S_r^{+*}}{t^*} \tag{9}$$

In this model, a DMU will be considered efficient when $\rho^*=1$.

According to Tone (2001) , when DMUs present null values for one or multiple variables, the resulted final ranking may be stabled according to the following procedure: run SBM model with all DMUs and all variables (Model A); then run SBM with all DMUs again, but, this time, excluding variables with null values (Model B); finally, build the rank considering the efficiency of Model A for those DMUs which don't present a null value for any variable and the efficiency of Model B for those DMUs which present a null value for at least one variable. This procedure avoids issues caused by zero for x_{i0} and y_{r0} in Equations 1 and 2 (underestimation of efficiency).

An adopted criterion for choosing the configuration of the model was the minimum total of ties, keeping the physical coherence of the model and the approval of specialists. Table 3 presents the sequence of model configuration trials. The use of strikethrough text (i.e., ~~strikethrough~~) is to represent and to spotlight the exclusion of variable in comparison to the previous model.

Table 3. Loop of model trials, followed by their variable configurations and the resulting total of ties.

Investigation Loops	Variables	Total of ties
Model 1	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: Production, On-farm storage capacity (t) UNCONTROLABLE VARIABLE: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as input)	23
Model 2	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, On-farm storage capacity (t)	4

	UNCONTROLABLE VARIABLE: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as input)	
Model 3	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: Production, ~~On-farm storage capacity (t)~~ UNCONTROLABLE VARIABLE: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as input)	14
Model 4	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: Production, On-farm storage capacity (t) UNCONTROLABLE VARIABLE: ~~Length of the route~~ UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	15
Model 5	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: Production, ~~On-farm storage capacity (t)~~ UNCONTROLABLE VARIABLE: ~~Length of the route~~ UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	14
Model 6	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, Corridor exportation, ~~On-farm storage capacity (t)~~ UNCONTROLABLE VARIABLE: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	10
Model 7	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, Corridor exportation, On-farm storage capacity (t) UNCONTROLABLE VARIABLE: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	14
Model 8	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, Corridor exportation, On-farm storage capacity (t) UNCONTROLABLE VARIABLE: ~~Length of the route~~ UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	13
Model 9	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, Corridor exportation, ~~On-farm storage capacity (t)~~ UNCONTROLABLE VARIABLE: ~~Length of the route~~ UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	9
Model 10	INPUTS: Freight price, Logistic loss, Fuel consumption OUTPUTS: ~~Production~~, Corridor exportation, ~~On-farm storage capacity (t)~~ INPUT ~~UNCONTROLABLE VARIABLE~~: Length of the route UNDESIRABLE OUTPUT: Emissions (treated as ~~input~~ inverted output)	21

Model 1, based on the significant correlation results, presented 23 ties. Production had a high correlation with other variables. Despite knowing it was desired to have a variable that represents the cargo transportation, production was excluded as part of the investigation process. Model 2 (with production) presented only 4 ties. The second way of the investigation was the exclusion of other output, on-farm capacity instead of production (Model 3). It resulted into 14 ties.

After investigating the impact of outputs in ties, the next steps of investigation amended undesirable output and uncontrollable variable. Emissions is the only variable that represents an environmental characteristic. It is mandatory to keep it, whether it is desired to capture this aspect in the performance index. Considering this, it was chosen to amend the way this undesirable output is treated. Instead of being considered an input, its inverse was considered an output. In this way, the maximization of output (performed by the DEA model) results in better rank positions for those DMUs with the maximum inverted emission, i.e., minimal emissions. This approach was adopted from Model 4 on.

The exclusion of variable length of the route, in Model 4, resulted in 15 ties. The additional exclusion of the variable on-farm storage capacity (t), in Model 5, resulted in 15 ties. At this point in the analysis, it was clear that the most impacting variable (causing ties), that could be potentially exchanged or excluded, was production. The operational variables (travel time and average speed) were excluded from all models due to the lack of significant correlation. If production was also excluded, that would represent that two characteristics were excluded from the performance index. It is aimed to have the maximum characteristics measured and represented. In this way, it was understood that a non-significant correlation may represent that the analysis of correlation is invalid for classifying a variable, though it does not mean this variable may not be at the model at all. Exercising the DEA prerogative that the main view of the system has precedence over the model construction, it was accepted to exchange Production for Corridor exports in Model 6. This choice resulted in 10 ties.

In the opposite direction of the investigation, Model 7 covered the inclusion of on-farm storage capacity (based on Model 6 configuration). It resulted into 14 ties. Model 8 explored the possibility of keeping the two inputs (corridor exportation and on-farm storage capacity) and excluding the uncontrollable variable. It resulted into 13 ties. Model 9 investigated the possibility of elimination of variables on-farm capacity and length of the route. It resulted into 9 ties, though, in this case, the infrastructural characteristics would be excluded from the index. Model 10, finally, examined the remote possibility of treating the length of the route as an input (that could be assumedly reduced by public investments strategies). It resulted into 21 ties. Model 6 was adopted because its configuration presented the minimal number of ties with the maximum representativeness of characteristics.

3.5 Tie-breaking method: DEA composite index

The tie-breaking method proposed by Leta et al. (2005) was applied, according to Equation 10:

$$E_k^{composite} = \frac{\left[E_k^{standard} + \left(1 - E_k^{inverted}\right)\right]/2}{\max\ \{[E_k^{standard} + \left(1 - E_k^{inverted}\right)]/2]\}} \quad k=1,2,3,\ldots, z \tag{10}$$

Where:
$E_k^{standard}$: is the standard efficiency resulted from the application of the DEA model for the kth DMU;
$E_k^{inverted}$: is the inverted efficiency of the kth DMU, i.e., the resulted efficiency when inputs are inserted in SBM model as outputs and vice versa.

This tie-breaking method is named *composite index*. It represents an arithmetic average between standard and inverted efficiencies standardized by the maximum composite index of the analyzed population.

3.6 Sensitivity analysis: SBM Super-efficiency

According to Greco et al. (2018), the composite index construction must be followed by robustness investigation, especially when the indicator is intended to be used for directing political decisions; otherwise, it may lead to mistaken decisions. In this way, the current paper represents an advance of previous efforts of freight performance index construction using DEA, specifically Melo et al. (2018) and Oliveira and Cicolin (2016), which have not incorporated robustness tests as part of the index creation process.

Among the techniques associated with DEA, there is the super-efficiency analysis. It was initially thought as a tie-breaking method, then it was pointed as a technique for outlier's determination and, more recently, Zhu (2001) and Mozaffari and Gerami (2012) stated that super-efficiency is accepted as a technique of sensitivity analysis for DEA. The creator of the SBM model, Tone (2001), proposed, in a subsequent paper (Tone, 2002), the super-efficiency analysis in the linear form for SBM, according to Equation 11:

$$\delta^* = \min \tau = \frac{1}{m}\sum_{i=1}^{m}\frac{\tilde{x}_i}{x_{i0}} \tag{11}$$

Subject to Equations 12, 13, 14, 15, 16, and 17:

$$\frac{1}{s}\sum_{r=1}^{s}\frac{\tilde{y}_r}{y_{r0}} = 1 \tag{12}$$

$$\tilde{x} \geq \sum_{k=1,\ \neq 0}^{z} \Lambda_k x_k \tag{13}$$

$$\tilde{y} \leq \sum_{k=1,\ \neq 0}^{z} \Lambda_k y_k \tag{14}$$

$$\tilde{x} \geq t x_0 \tag{15}$$

$$\tilde{y} \leq t y_0 \tag{16}$$

$$\Lambda_k \geq 0,\ \tilde{x} \geq 0, \tilde{y} \geq 0 \text{ and } t > 0 \tag{17}$$

Where:
δ^*: is the optimized super-efficiency score.
δ: is the super-efficiency score.
$\tilde{x}$: is the linearized average of input expansion rate for the linear problem ($\tilde{x} = t\overline{x}$).
$\tilde{y}$: is the average output reduction rate for the linear problem ($\tilde{y} = t\overline{y}$).
$\overline{x}$: is the average input expansion rate.
$\overline{y}$: is the average output reduction rate.
The other elements are described in Equations 1, 2, 3, 4, 5, 6, and 7.
The optimum solution ($\delta^*, t^*, \lambda_k^*, \overline{x}^*, \overline{y}^*$) is described by the conditions in Equation 18.

$$\delta^* = \tau^*,\ \lambda_k^* = \frac{\Lambda_k^*}{t^*},\ \overline{x}^* = \frac{\tilde{x}^*}{t^*},\ \overline{y}^* = \frac{\tilde{y}^*}{t^*} \tag{18}$$

4. Results and Discussion

Table 4 shows the resulting efficiency of the Model 6, followed by the inverted efficiency and the composite index rank. The 14 most efficient DMUs (routes) present origin and the destination port at the same state. Thirteen of them are in the Southern region of Brazil, being the only exception DMU24 in the north-eastern state of Maranhão (MA). That may suggest that, besides the shorter length of the routes and the lack of cross-state taxes due to transportation inside the same federative unit, the states from Southern regions are more economically competitive than the other origins.

Before the application of the tie-breaking method, there are 10 ties (DMU6, DMU5, DMU22, DMU10, DMU9, DMU24, DMU11, DMU12, DMU15, and DMU20). After the tie-breaker, DMU6 is the unique considered 100% efficient, because it is efficient in standard DEA rank and, simultaneously, presents the best results for inverted efficiency. Among the 1% most efficient DMUs, there are DMU5, DMU22, DMU10, and DMU9, which also present a similar behavior to DMU6 for both efficiencies (standard and inverted). Among the 2% more efficient DMUs, there is exclusively DMU24. Among the 3% more efficient DMUs, there are DMU11 and DMU12, followed by DMU15 and DMU20, among the 4% more efficient.

Table 4. Results of DEA SBM for Model 6, considering standard and inverted efficiency, followed by composite index rank.

DMU	State of Origin	Mid-sized region of Origin	Destiny Port	Efficiency	1 - Inverted Efficiency	Average	Composite Index (CI)	CI Rank
DMU 6	PR	Centro Oriental Paranaense	Paranaguá (PR)	1.0000	0.9106	0.9553	1	1
DMU 5	RS	Centro Ocidental Rio-grandense	Rio Grande (RS)	1.0000	0.9104	0.9552	0.9999	2
DMU 22	RS	Sudoeste Rio-grandense	Rio Grande (RS)	1.0000	0.9096	0.9548	0.9995	3
DMU 10	SC	Norte Catarinense	São Francisco do Sul (SC)	1.0000	0.9092	0.9546	0.9993	4
DMU 9	RS	Noroeste Rio-grandense	Rio Grande (RS)	1.0000	0.9079	0.9540	0.9986	5
DMU 24	MA	Sul Maranhense	São Luís (MA)	1.0000	0.8791	0.9396	0.9835	6
DMU 11	PR	Norte Central Paranaense	Paranaguá (PR)	1.0000	0.8684	0.9342	0.9779	7
DMU 12	PR	Norte Pioneiro Paranaense	Paranaguá (PR)	1.0000	0.8665	0.9332	0.9769	8
DMU 15	PR	Oeste Paranaense	Paranaguá (PR)	1.0000	0.8516	0.9258	0.9691	9
DMU 20	PR	Sudoeste Paranaense	Paranaguá (PR)	1.0000	0.8372	0.9186	0.9616	10
DMU 4	PR	Centro Ocidental Paranaense	Paranaguá (PR)	0.6626	0.8637	0.7631	0.7988	11
DMU 17	SC	Serrana	São Francisco do Sul (SC)	0.6471	0.8784	0.7628	0.7984	12
DMU 14	SC	Oeste Catarinense	São Francisco do Sul (SC)	0.5051	0.8952	0.7002	0.7329	13
DMU 7	BA	Extremo Oeste Baiano	Salvador (BA)	0.4185	0.8330	0.6258	0.6550	14
DMU 21	PI	Sudoeste Piauiense	São Luís (MA)	0.2746	0.8195	0.5470	0.5726	15
DMU 23	GO	Sul Goiano	Santos (SP)	0.2573	0.7422	0.4997	0.5231	16

DMU 8	GO	Leste Goiano	Santos (SP)	0.2146	0.6866	0.4506	0.4717	17
DMU 19	MS	Sudoeste de Mato Grosso do Sul	Santos (SP)	0.2246	0.6434	0.4340	0.4543	18
DMU 2	MS	Centro Norte de Mato Grosso do Sul	Santos (SP)	0.21000	0.6104	0.4102	0.4294	19
DMU 3	MS	Centro Norte de Mato Grosso do Sul	São Francisco do Sul (SC)	0.1784	0.5991	0.3888	0.4070	20
DMU 18	MS	Sudoeste de Mato Grosso do Sul	Paranaguá (PR)	0.2362	0.4170	0.3266	0.3419	21
DMU 16	TO	Oriental do Tocantins	São Luís (MA)	0.2453	0.2167	0.2310	0.2418	22
DMU 13	TO	Ocidental do Tocantins	São Luís (MA)	0.2542	0.0000	0.1271	0.1330	23
DMU 28	MT	Norte Mato-grossense	Santos (SP)	0.2052	0.0000	0.1026	0.1074	24
DMU 1	MS	Centro Norte de Mato Grosso do Sul	Paranaguá (PR)	0.1886	0.0000	0.0943	0.0987	25
DMU 25	MT	Nordeste Mato-grossense	Santos (SP)	0.1684	0.0000	0.0842	0.0881	26
DMU 29	MT	Norte Mato-grossense	Paranaguá (PR)	0.0939	0.0000	0.0470	0.0492	27
DMU 30	MT	Norte Mato-grossense	São Luís (MA)	0.0895	0.0000	0.0448	0.0469	28
DMU 27	MT	Nordeste Mato-grossense	São Luís (MA)	0.0675	0.0000	0.0337	0.0353	29
DMU 26	MT	Nordeste Mato-grossense	Paranaguá (PR)	0.0637	0.0000	0.0319	0.0334	30

For deeper investigation of the efficient DMUs, the super-efficiency was applied to the model. Among the 10 previously most efficient DMUs, nine present the results above 1 for super-efficiency; in decreasing order: DMU10, DMU9, DMU24, DMU22, DMU6, DMU5, DMU12, DMU15, and DMU20. DMU11 is the missing one. The super-efficiency may be interpreted as the "room" of movement of an efficient DMU that will not result in an exclusion of the efficiency frontier. In other words, whether DMU11 changes slightly its data, it will be considered inefficient. On the other hand, a change of up to 74.76% in its best variable (emissions) will not result in inefficiency for DMU10. It presents the minimal emissions of all population. DMU9 presents the greatest corridor exports. DMU6 presents minimal fuel consumption. DMU5, DMU22, DMU24, DMU12, DMU15, and DMU20 present a balanced combination of good performance for all analyzed variables.

Table 5. Results of super-efficiency.

DMU	CI	CI Rank	Super-efficiency	Super-rank
DMU1	0.0987	25	1	10
DMU2	0.4294	19	1	10
DMU3	0.4070	20	1	10
DMU4	0.7988	11	1	10
DMU5	0.9999	2	1.0014	6

DMU6	1.0000	1	1.0081	5
DMU7	0.6550	14	1	10
DMU8	0.4717	17	1	10
DMU9	0.9986	5	1.6456	2
DMU10	0.9993	4	1.7476	1
DMU11	0.9779	7	1	10
DMU12	0.9769	8	1.0005	7
DMU13	0.1330	23	1	10
DMU14	0.7329	13	1	10
DMU15	0.9691	9	1.0005	7
DMU16	0.2418	22	1	10
DMU17	0.7984	12	1	10
DMU18	0.3419	21	1	10
DMU19	0.4543	18	1	10
DMU20	0.9616	10	1.0003	9
DMU21	0.5726	15	1	10
DMU22	0.9995	3	1.0086	4
DMU23	0.5231	16	1	10
DMU24	0.9835	6	1.0460	3
DMU25	0.0881	26	1	10
DMU26	0.0334	30	1	10
DMU27	0.0353	29	1	10
DMU28	0.1074	24	1	10
DMU29	0.0492	27	1	10
DMU30	0.0469	28	1	10

5. Concluding remarks

The current paper successfully presented a framework for a freight performance index creation, using DEA models. The DEA is a data-driven method allegedly less subjective, once the decision-makers can not directly attribute weights to the index components.

The framework represents an advance in the literature, once it combines recommendations of Greco et al. (2018) for general index creations and general DEA guidelines of Cook, Tone and Zhu (2014), to the previous efforts of measuring efficiency of soybean corridors in Brazil of Melo et al. (2018) and Oliveira and Cicolin (2016).

After the data collection and the specialist opinions about which were the most relevant aspects to be analyzed, it was proposed to use the correlation as a criterion for the variable selection and categorization. The SBM with variable scale is argued to be the most adequate model to this application, once it does not require orientation choice. The next steps proposed by the framework involves DEA model trials with multiple variables configurations, focusing on the desired criterion of analysis, in the current application the focus was minimizing ties, as well as guaranteeing the measurement of the most relevant aspects.

Subsequently, the composite index of Leta et al. (2005) is applied as tie-breaking method for the final efficiency index construction. Finally, the robustness of the index is tested through super-efficiency analysis, as stated by Tone (2002). Although composite index was previously applied by Melo et al. (2018) for soybean freight efficiency measurement, this is the first registered effort to analyze robustness of the DEA efficiency rank for a logistics application.

As a practical result from the current application, it was demonstrated that the Southern states of Brazil are more efficient in soybean transportation and should be considered as benchmarks for other regions. In the Northeastern region, only Maranhão state (MA) is already a benchmarking. When considering the sensitivity analysis, MA (DMU24) remains among the efficient case (third position at the super-efficiency rank). A DMU from the Southern region (DMU11, from Paraná state) does not present evolution in the super-efficiency rank, showing that it is at the border of efficiency, and Paraná (PR) may be a state with diverse conditions of transportation among the mid-sized regions. Public decision-makers from PR are recommended to focus on equalization of transport efficiency among mid-sized regions, while federal public decision-makers are recommended to focus on investments for improving transportation efficiency from Center-Western states (MS, MT, and GO), Northern states (TO), and two Northeastern states (BA and PI).

For further studies, it is recommended to incorporate more data collection, for example, measuring more productive characteristics. Whether total costs, and temporal levels of demand, reduction and inventories are available, a dynamic version of the allocative control DEA model, proposed by Alves Junior et al. (2018), could be used for a temporal index creation. This framework is recommended for the measurement of efficiency of other bulk freight transportations in the context of developing economies. It may also be applied for multi-modal corridors, once homogeneity assumptions are not violated. Based on the index result, decision-makers are expected be able to focus on investments in the inefficient DMUs and in the variables that are responsible for their inefficiency.

Acknowledgements

The authors would like to thank the **São Paulo Research Foundation (FAPESP), grant #2018/20436-1**, for supporting this research study through Post-Doctoral Scholarship. The authors also thank **Coordination for the Improvement of Higher Education Personnel (CAPES 001)** and **São Paulo Research Foundation (FAPESP), grant #2015/14791-5** for their financial support.

References

Adler, N., Golany, B., 2007. PCA-DEA In: “Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis”, Boston, MA: Springer US, 139–53

Alves Junior, P. N., Melo, I. C., Celestini, G., Rebelatto, D. A. N., 2018. Allocative Control Data Envelopment Analysis (DEA) Model to Evaluate Inventory Control Systems When There Is a Relationship among Variables. GEPROS.Gestão da Produção 11: 143–60.

ANTT – National Land Transportation Agency, 2018. http://www.antt.gov.br/ (July 5, 2018)

Banker, R. D., Charnes, A., Cooper, W.W., Swarts, J., 1989. An Introduction to Data Envelopment Analysis with Some of Its Models and Their Uses. Research in Governmental and Non-Profit Accounting 5: 125–63

Charnes, A., Cooper, W. W., Rhodes, E., 1978. Measuring the Efficiency of Decision Making Units. European Journal of Operational Research 2(6): 429–44.

Cook, W. D., Tone, K., Zhu, J., 2014. Data Envelopment Analysis: Prior to Choosing a Model. Omega 44, 1–4. http://dx.doi.org/10.1016/j.omega.2013.09.004

EPL - Planning and Logistics Company, 2018. https://www.epl.gov.br/ (July 5, 2018).

ESALQ-LOG, 2018. Research and Extension Group in Agroindustrial Logistics. http://esalqlog.esalq.usp.br/ (July 5, 2018).

Greco, S., Ishizaka, A., Tasiou, M., Torrisi, G., 2018. On the Methodological Framework of Composite Indices: A Review of the Issues of Weighting, Aggregation, and Robustness. Social Indicators Research: 1–34. http://link.springer.com/10.1007/s11205-017-1832-9 (June 12, 2018).

Golany, B., Roll, Y., 1989. An Application Procedure for DEA. Omega 17(3): 237–50.

Hua, Z., Bian, Y., 2007. DEA with Undesirable Factors In: "Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis", Boston, MA: Springer US, 103–21

IBGE - Brazilian Institute of Geography and Statistics, 2018. Regional Division of Brazil. https://ww2.ibge.gov.br/home/geociencias/geografia/default_div_int.shtm

Leta, F. R., Soares de Mello, J. C. C. B., Gomes, E. G., Angulo Meza, L., 2005. Turning machine static evaluation using methods for ranking inprovement of DEA models. Investigação Operacional 25(2): 229–42.

Li, W., Liang L., Cook, W. D., Zhu, J., 2016. DEA Models for Non-Homogeneous DMUs with Different Input Configurations. European Journal of Operational Research 254: 946–56.

Liu, W. B., Meng, W., Li, X. X., Zhang, Z. Q., 2010. DEA Models with Undesirable Inputs and Outputs. Annals of Operations Research 173(1): 177–94.

Melo, I. C., Alves Júnior, P. N., Périco, A. E., Guzman, M. G. S., Rebelatto, D. A. N., 2018. Benchmarking Freight Transportation Corridors and Routes with Data Envelopment Analysis (DEA). Benchmarking: An International Journal 25(2): 713–42. http://www.emeraldinsight.com/doi/10.1108/BIJ-11-2016-0175

Mozaffari, M. R., Gerami, J., 2012. Super-Efficiency and Sensitivity Analysis on Input-Oriented DEA-R. Journal of Mathematical Extention 6 (0). http://www.ijmex.com/index.php/ijmex/article/view/100

Oliveira, A. R., Cicolin, L. O. M., 2016. Evaluating the Logistics Performance of Brazil's Corn Exports: A Proposal of Indicators. African Journal of Agricultural Research 11(8): 693–700. http://www.academicjournals.org/AJAR (July 5, 2018).

Saen, R. F., 2005. Developing a Nondiscretionary Model of Slacks-Based Measure in Data Envelopment Analysis. Applied Mathematics and Computation 169(2): 1440–47.

Seiford, L. M., Zhu, J., 2002. Modeling Undesirable Factors in Efficiency Evaluation. European Journal of Operational Research 142(1): 16–20

Tone, K., 2001. A Slacks-Based Measure of Efficiency in Data Envelopment Analysis. European Journal of Operational Research 130(3): 498–509.

———., 2002. A Slacks-Based Measure of Super-Efficiency in Data Envelopment Analysis. European Journal of Operational Research 143(1): 32–41.

Zhu, J., 2001. Super-Efficiency and DEA Sensitivity Analysis. European Journal of Operational Research 129(2): 443–55.