



Assessment of the technical efficiency of Brazilian logistic operators using data envelopment analysis and one inflated beta regression

Murilo Wohlgemuth¹ · Carlos Ernani Fries¹ ·
Ângelo Márcio Oliveira Sant'Anna² · Ricardo Giglio¹ ·
Diego Castro Fettermann¹

Published online: 24 November 2018

© Springer Science+Business Media, LLC, part of Springer Nature 2018

Abstract

In the past two decades logistics services providers have increased their sector participation due to growing outsourcing of these services. In current scenario, logistics operators typically offer service packages that include not only transport itself but also other services in supply chain and transport's service associated information. This work objective is to identify logistics services packages offered from logistics operators that lead to technical efficiency of operations observed in sector. For this analysis, the Data Envelopment Analysis (DEA) in two stages methodology was applied, where the first stage consists in the use of DEA models to obtain relative efficiency scores and the second stage consists in the use of one beta inflated regression to analyze the relationship between the technical efficiency scores obtained and the offered services. This study was made with secondary data base available on a logistics' sector specialized magazine, for the period 2007–2015. Results show that a relationship between the offer of logistics service packages and logistics service providers' technical efficiency exists. Different for each cluster, the statistically significant service packages vary as the magnitude of contribution on the efficiency measure. Most packages lead to negative contribution on the efficiency, while a few showed positive contributions.

Keywords Logistics · Data envelopment analysis · One inflated beta regression · Brazil

1 Introduction

Low investment and drop in consuming goods characterize the current scenario of the Brazilian economy. This potentiates the productive sector expenses with logistics because

✉ Diego Castro Fettermann
dcfettermann@gmail.com

¹ Department of Industrial and Systems Engineering, Federal University of Santa Catarina (UFSC), Rua Delfino Conti, s/n, Trindade, Florianópolis CP 88040-900, Brazil

² Polytechnic School, Federal University of Bahia (UFBA), Rua Prof. Aristides Novis, 02, Federação, Salvador CP 40210-630, Brazil

outsourcing the activities related to logistics lead to cost savings, such a goal that stands out in times of recession. Logistics outsourcing continues to be crucial for companies to reduce logistics costs, and it is an opportunity to create business value (Mello et al. 2008; Zacharia et al. 2011). A research conducted by the Brazilian Research Foundation with 142 companies of 22 industry segments (whose total sales amount to 15% of Brazilian GDP) supports this assertion. This study found that cost reduction is achieved by the outsourcing of logistics activities (Resende et al. 2016). There are similar results in other developing countries, such as Turkey (Aktas and Ulengin 2005; Aktas et al. 2011) and China (Yang et al. 2016) and India (Mitra 2006). Outsourcing of logistics services is adopted by companies that do not have logistics as a core business, transforming fixed costs into variable costs, i.e., causing logistics costs to be counted only when demand is verified. Based on Research Institute, in 2014 a growth of 11.7% in logistics costs was observed in the total net revenue of companies, while in 2010 this percentage was 10.6% (Ilos 2014). From this logistical nature spending of Brazilian companies, two-thirds are designated to pay logistics service providers, which highlights the importance of this market in the economy (CEL 2009). In general, logistics service providers offer service packages that include not only transport itself but also other services of the supply chain.

The logistic service providers supply many service options to their customers. Connectivity and communication requirements involved have pushed third-party logistic providers 3PLs into a more advanced level (Zacharia et al. 2011). The widespread dissemination of information and communication technology has opened new opportunities to create new roles in the supply chain (Evangelista and Sweeney 2006). However, creating and offering these new services might lead to considerable costs to the 3PLs. The identification of a proper service portfolio that drives the logistic operator is a crucial task to the agents involved (Lai 2004), with a notable impact on performance (Grönroos 2000). Such an issue is even more relevant for the logistic service providers, due to the growing complexity of dealing with specific outsourced services offered by other service providers (4PL and 5PL; Hosie et al. 2012). Despite that, literature is plenty of studies on the effect of outsourcing logistics services and on the selection of logistics providers (Aguzzoul 2014; Senthil et al. 2014; Guarnieri et al. 2015; Govindan et al. 2016). On the other hand, few studies were found on the performance of the logistic service providers, mainly focusing on the identification of their service portfolios.

The aim of this study was to identify the service packages offered by logistics operators leading to the greater efficiency observed in LSP's. For this analysis, the Data Envelopment Analysis (DEA) methodology was applied in two stages. First stage consists in using DEA models for efficiency scores while the second stage consists in using a regression model to assess the relationship between technical efficiency from DEA and the services offered by the respective companies. The study was based on secondary data available in a specialized journal of the logistics sector for the period of 2007–2015. Results show that the relationship between the offer of certain logistics services packages and technical efficiency of logistics operators could be verified and quantified. The definition of proper services portfolios allows the Logistic Service Providers to better drive investments towards performance improvement in all relevant business areas.

2 Theoretical foundation

2.1 Productivity and efficiency

Productivity is a fundamental indicator of the economic evaluation of transformation processes. It is defined by the ratio between the volume of goods and/or services and the volume of inputs used in a transformation process. This relationship is trivial in the case of a process of transforming a single input into a single output. However, in the case of a general process, when multiple inputs are transformed into multiple outputs, the numerator and denominator of this relation must represent the characteristics of the production technology used. This general measure of productivity is referred to as Total Factor Productivity (TFP), which can be simplified as the weighted proportion of total output in relation to the weighted proportion of total inputs used in the production process (Coelli et al. 2005).

While productivity relates outputs and inputs, efficiency compares the productive performance with the best performance observed in the sector. To Norman and Stoker (1991), the assessment of how well an organizational unit transforms their inputs into outputs is made by the efficiency analysis of the transformation process used.

2.2 Data envelopment analysis

Data Envelopment Analysis (DEA) is a non-parametric mathematical programming approach to estimate efficiency frontiers of transformation processes of multiple inputs to multiple outputs. DEA is used to empirically assess the productive efficiency of a set of Decision Making Units (DMU's). The DMU's represent agents operating in an economic environment by producing goods and/or services. Such tool, based on linear programming, was originally proposed by Charnes et al. (1978) and able to deal with constant returns to scale only. It was then improved to accommodate variable returns to scale by Banker et al. (1984). Since then, DEA became a flexible way for the efficiency evaluation of processes, being often updated and extended to many different situations and scenarios.

DEA's basic model assess the efficiency of the DMUs by the weighted average of all output, divided by the also weighted average of all input, with no prior determination of factor scoring (Eq. 1). Each DMU is responsible for its production plan, which is the set of weights for input and output that it considers appropriated in order to maximize productivity (Cooper et al. 2007). For a given DMU o from a set of n production units using similar technologies (r inputs X and s outputs Y), the goal is to determine the set of weights v_i , ($i = 1, \dots, r$), u_j , ($j = 1, \dots, s$), which maximizes the weighted relation between input and output, $\sum_{j=1}^s u_j Y_{jo} / \sum_{i=1}^r v_i X_{io}$, subject to the restriction that for each DMU from the set to be analyzed, the weighted sum of their output is limited by the weighted sum of their input. The denominator of the objective function of this optimization problem above might be truncated to 1 in a way that it can be transformed into a linear programming problem:

$$\begin{aligned}
 \max_{u,v} \quad & e_o = \sum_{j=1}^s u_j Y_{jo} \\
 \text{subject to} \quad & \sum_{i=1}^r v_i X_{io} = 1 \\
 & \sum_{j=1}^s u_j Y_{jm} \leq \sum_{i=1}^r v_i X_{im} \quad m = 1, \dots, n \\
 & v_i \geq 0 \quad i = 1, \dots, r \\
 & u_j \geq 0 \quad j = 1, \dots, s
 \end{aligned} \tag{1}$$

				2° stage
Stages	Market classification	Selection of input and output variables for DEA analysis	DEA efficiency scores	Regression model
Data analysis	Cluster analysis (k-means cluster)	Pearson Correlation Principal Component Analysis	VRS model	One inflated Beta regression model

Fig. 1 Data analysis procedure

This linear programming problem is referred to in the literature as the constant returns to scale model (CRS). (Cooper et al. 2007). It can be solved for each unit m from the set of DMUs, where $e_m \leq 1$. The DMUs where $e_m = 1$ are operating with production plans in the efficiency frontier, while those where $e_m < 1$ are outside the efficiency frontier and so evaluated as inefficient. This model is called input-oriented because it calculates the distance between the production plan of the DMU o and the efficiency frontier by considering the lowest level of input that can be used to achieve the same level of output.

The addition of the convexity restrictions to the mathematical programming above makes the area of feasible solutions of the CRS model restricted to the convex combinations generated by the production plans of the observed DMUs. This model was proposed by Banker et al. (1984), and it has introduced the concept of variable returns of scale (VRS) to the DEA.

The use of DEA models is particularly advantageous in the assessment of units that production factors are not directly subject to market values or other cardinal measures of relative importance, because the analysis is done disregarding the functional form of the relationship between inputs and outputs (Fries 2013).

3 Proposed methodology

The data modeling for logistic operators consists in adopting structured stages during the execution activity, mainly in the most important stages of the data analysis: the DEA model and the Regression model. These ideas must focus on precision, in the sense of being a better solution to the problem and not just the simple identification operators' activities. The proposed scheme defines four stages for data modeling to logistic operators systems, as shown in Fig. 1.

The initial step of the first stage consists in gathering and processing of the data used in work. The Research Institute publishes annually an overview of logistics operators in Brazil with information about their characteristics and operating results. In the period under consideration from 2007 to 2015, it was counted 1207 records of Logistic Services Provider (LSP) companies. This group of companies consolidated the database for this study (Table 1).

3.1 Market classification

The first data analysis stage seeks to identify homogeneous groups of LSP regarding size and types of services provided by companies. This classification seeks to organize LSP's that have functional similarities into relatively homogeneous groups, which share both production factors and similar results. For this purpose cluster analysis was used, which aims to identify groups of elements with homogeneous characteristics and heterogeneous to one another (Hair et al. 2010; Tabachnick and Fidell 2013). This step is detailed to the con-

Table 1 Descriptive of the sample. *Source:* Castro (2015)

Variable	Average	SD	Coefficient of variation	Number of cases
Market activity (years)	23.0	20.3	0.9	131
Number of employees	1205.5	2635.4	2.2	127
Number of customers	173.5	541.1	3.1	116
Annual gross revenue (US\$)	67,846,962.22	189,816,848.6	2.8	87
Storage area—own (m ²)	233,235.7	1,350,594.6	58	111
Storage area—customer (m ²)	43,663.2	83,562.7	1.9	73
Storage area—customs (m ²)	54,581.7	199,001.5	3.6	53
Storage area—refrigerated (m ²)	23,466.5	116,564.0	5.0	60
Storage area—outdoor (m ²)	103,296.8	183,985.6	1.8	91
Storage area—total (m ²)	365,618.8	1,430,084.9	3.9	117
Number of warehouses—own	7.8	15.1	1.9	115
Number of warehouses—customers	4.6	7.1	1.5	77
Number of managed items	93,259,663.3	561,237,260.9	6.0	60
Weight of managed items	2,616,601.6 kg	8,486,663.4 kg	3.2	52

dition of homogeneity of DMU's sample, which is recommended for application of DEA models. Among the available techniques of classification, was opted the method k-means cluster analysis algorithm with the generalized expectation maximization (Hair et al. 2010). To perform the cluster analysis, were selected 14 dummy variables (presence/absence): (1) own transport fleet, (2) own fleet routing systems, (3) outsourced fleet routing systems, (4) tracking satellite technology for own fleet, (5) tracking satellite technology for outsourced fleet, (6) transport coordination services, (7) distribution services, (8) door to door transport, (9) transfer service (10) milk run transportation, (11) packaging services, (12) Assembly kits and SETS, (13) WMS, (14) use of customer warehouses, and (15) use of own warehouses.

To better assess the arrangement of the number of the clusters it was used the Silhouette coefficient (Si; Rousseeuw 1987), that indicates the best arrangement is achieved with three clusters. Thus, the sample of 1207 companies analyzed was divided into three clusters according to their composition of assets and offered services. The data was obtained in the survey carried out between the years of 2007 and 2015 (Castro 2015). Because of missing data on some variables, it was necessary to disregard 219 cases. This procedure is considered the first data disposal of the analysis. These clusters were named as: (1) Node, formed by 120 companies that mainly carry out logistics services on the premises, that is, in general warehouses and manufacturing facilities; (2) Network With Assets, formed by 626 companies that predominantly have their own fleet and provide logistics services throughout the network, i.e. the facilities and services associated with transport that enable connection between the logistics facilities, and (3) Network Without Assets, formed by 242 companies that predominantly do not have own fleet and offer logistics services throughout the network. Companies in this category usually outsource transportation activities to other logistics companies.

Table 2 Variables input and output selected

Cluster	Input/output	Variable	Unit
Node	Output	Annual gross revenue	US\$
	Input	Number of employees	
	Input	Storage area—total	m ²
Network with assets	Output	Annual gross revenue	US\$
	Input	Number of employees	
	Input	Storage area—total	m ²
Network without assets	Output	Annual gross revenue	US\$
	Input	Number of customers	
	Input	Storage area—total	m ²

3.2 Selection of variables for the DEA

In the following it is described the procedure of variable selection that was applied to the original set of potential input and output variables. The objective is to improve DEA's discriminatory power. For each of the three clusters detailed above, a Principal Component Analysis (PCA) was performed to help selecting the variables to represent the processing function of inputs into outputs. The need for variable selection happens when there are many variables as inputs and outputs, because many different units can be identified as efficient for different parameters, and therefore their scores will not have discriminatory desired power (Adler and Yazhemsky 2010). This technique enables reduction of variables with the least possible loss of information content.

It must be highlighted that the PCA was applied here only to help selecting the subset of the original variables that accounts for the largest part of the variance, and not to replace original DEA inputs by the PCA factor scores, what is known as PCA–DEA (Adler and Golany 2007). In the current case where a regression analysis is performed on top of DEA efficiency scores, it was more suitable to use raw rather than PCA-transformed variables. Dyson et al. (2001) point out that the interpretations of the efficiency scores and other benchmarks between peers can change under weights restrictions such as the PCA factor scores transformation.

From the factor loadings of each variable in each component it is verified the possibility of discarding some variables, after all, variables with similar loadings across the components convey basically the same information, and thus can be removed from the DEA problem for the sake of discriminatory power. The selected components have eigenvalues higher than 1.0 and represent more than 60% of the data variance, as suggested by the literature (Tabachnick and Fidell 2013). In this sense, from the factor loadings of the analysis of each variable in each component, variables have selected the ones to be used as input and output in DEA. At the end of this procedure, for each of the three clusters analyzed the following variables were selected (Table 2).

With the definition of the variables that are used to determine the efficiency of LSP's with the application of DEA models, the sample of the three clusters undergoes a new discard phase due to values not informed by the respondents of the selected variables. This second stage of data disposal refers to cases that do not present records referring to the input and output variables of their actuation cluster. Figure 2 shows the reduction related to the disposal suffered by the sample in the clustering step (1st disposal) and disposal suffered after the selection of input and output variables (2nd disposal).

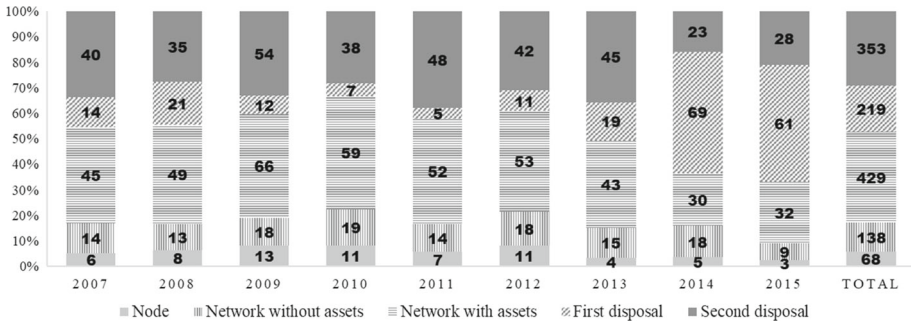


Fig. 2 Number of LSP's in each cluster and disposal after clustering and selection of input and output variables

3.3 Determination of DEA efficiency scores

The third step was the determination of the actual efficiency scores. Applies the mathematical programming model with the data for each year, for each cluster, and determines the scores of relative efficiencies of LSP's. In this step was applied the BCC model (model VRS) input oriented, or seek to minimize inputs used to generate the same amount of each of the outputs. For this application of DEA model, the GAMS software (General Algebraic Modeling System) was used. In the result of each year, scores of companies equal to one (1000) are on the efficient frontier for an efficiency given for each model, and thus qualified as efficient. Companies with scores below one (1000) qualify as inefficient, and their scores are the level of efficiency of the companies considered efficient in the reporting year. Figure 3 shows the histograms of the DMU's scores of the clusters; the nodes; the network with assets and; the network without assets, respectively. The histograms present a high concentration of values close to 1, what is a very frequent situation in Data Envelopment Analysis. This distribution of data makes the use of regression models based on the normality assumption not recommended.

3.4 Regression

The fourth and final stage (2nd stage) consisted in the application of regression models to investigate the relationship between the efficiency scores obtained in the previous step (dependent variable) and the services offered by logistics operators (independent variables) sample. The independent variables are a dummy (presence/absence) of the logistics service. In total were selected 20 different services offered by these logistics operators: (1) Storage (STO), (2) Inventory Control (INC), (3) Packaging (PAC), (4) Assembly Kits and Sets (AKS) (5) Third-Party Management (TPM), (6) Palletizing (PAL), (7) Cross-Docking (CRD), (8) Just-In-Time (JIT), (9) Import/Export and Customs Clearance (IEC), (10) Reverse Logistics (RLG), (11) Fiscal Support (FIS), (12) Projects Development (PRD), (13) Performance Monitoring (PEM), (14) Supply (SUP) (15) Coordination (COO), (16) Distribution (DIS), (17) Door-To-Door (DTD), (18) Transfer (TRA) (19) Milk Run (MKR) and (20) Intermodal Management (IMA).

The used regression models seek to identify the effect of the provision of logistics services packages (independent variables) in the technical efficiency measure (VRS model; dependent variable) on each identified cluster (network with assets, network without assets and node). The technical efficiency scores obtained by the DEA have values between 0 and 1. The

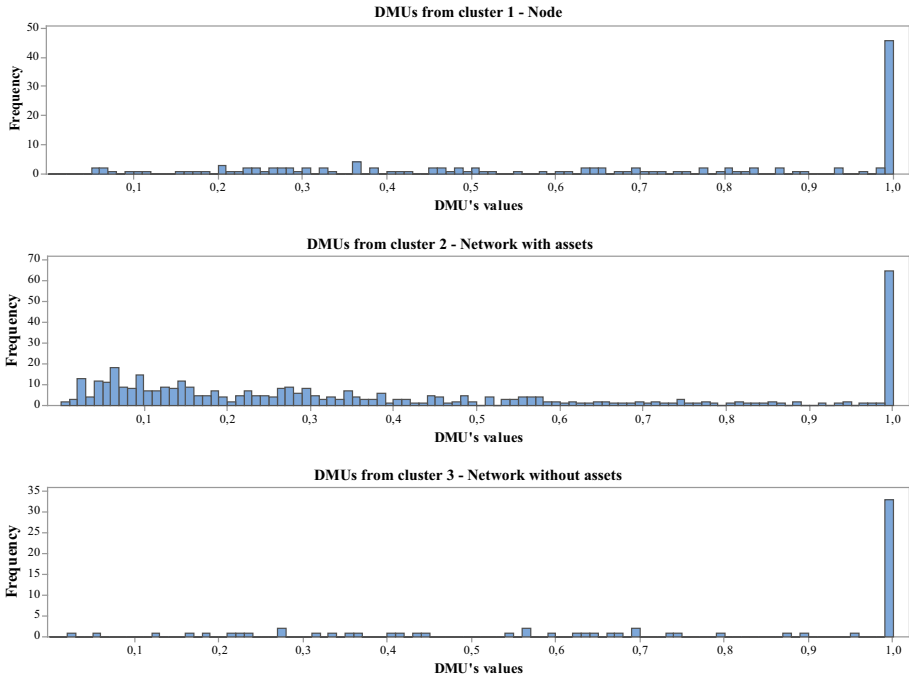


Fig. 3 DMU’s values of each cluster

Anderson–Darling test showed values that reject the hypothesis of adhesion to normal distribution of this data (dependent variable) for scores of clusters network without assets ($An = 57.50$; p value < 0.000), network with assets ($An = 107.93$; p value < 0.000) and node ($An = 34.04$; p value < 0.000). Thus, it is not recommended to use regression techniques based on normal distribution, such as OLS regression (Hair et al. 2010, Tabachnick and Fidell 2013). Noncompliance of efficiency scores obtained in the DEA to a normal distribution was also reported on the work of Hoff (2007) and Wu and Zhou (2015).

Alternatively, the Beta regression model was selected, since this model is useful for situations where the variable of interest is continuous, asymptotic and restricted to the interval (0, 1) and is related to other variables through a regression structure (Ferrari and Cribari-Neto 2004). This type of model uses the estimation of maximum likelihood and is based on the assumption that the response is beta distributed. The beta distribution is very flexible, and its density can have quite different shapes depending on the values of the parameters of the distribution. Often, proportions data include a non-negligible number of zeros and/or ones. When this is the case, the beta distribution does not provide a satisfactory description of the data, since it does not allow a positive probability for any particular point in the interval [0, 1] (Ospina and Ferrari 2010).

The Beta distribution with two parameters $\theta_1, \theta_2 > 0$ offers a robust approximation for a large variety of variables with several shapes and for modeling of data closed to interval [0, 1] (Johnson et al. 1995). The probability density function (pdf) of the Beta distribution is given by,

$$f(y; \theta_1, \theta_2) = \frac{\Gamma(\theta_1 + \theta_2)}{\Gamma(\theta_1)\Gamma(\theta_2)} y^{\alpha-1} (1 - y)^{\beta-1} \tag{2}$$

where $\Gamma(\theta)$ is the Gamma function evaluated as $\Gamma(\theta) = \int_0^{\infty} y^{\theta-1} e^{-y} dy, \theta > 0$. In the remaining of this section, we briefly describe the fundamentals of the Beta regression model proposed by Ferrari and Cribari-Neto (2004), with the reparameterization to mean and variability defined as $\mu = \frac{\theta_1}{\theta_1 + \theta_2}$ and $\phi = \theta_1 + \theta_2$, redefining with $\theta_1 = \mu\phi$ and $\theta_2 = (1 - \mu)\phi$, thus,

$$E(Y) = \mu \quad \text{and} \quad Var(Y) = \frac{Var(\mu)}{(1 + \phi)} \quad (3)$$

Finally, the Beta regression model can be represented in a traditional regression format by,

$$Y \sim Beta[\beta_0 + \beta_t X_t] = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k}}, t = 1, \dots, k \quad (4)$$

where Y is the output variable, β 's are unknown coefficients of X_t (x_1, \dots, x_k) input variables.

The data for technical efficiency scores (VRS model) obtained for the clusters analyzed in this study, as well as in other DEA applications have a higher density of values 1. This distribution is usually referred to one-inflated beta distribution and indicates the use of one inflated beta regression models.

4 Results and discussion

The beta regression models performed have the effect of providing 20 logistics services (independent variables) in the VRS efficiency score of DEA (dependent variable). The result obtained aims to identify which service packages provided by logistics operators that have a significant effect on the efficiency score. It conducted a regression model for each cluster of logistics operators analyzed, called node, network with assets and network without assets. For each cluster, was used the beta-inflated one regression model.

The results of the regression model for the logistics operators cluster (1) Node is presented in Table 3. Four service packages provided by logistics operators were identified that have significant relationship (p value < 0.10) with its efficiency score. Among the services that have a positive and significant effect on the efficiency score of logistics operators in cluster Node, can be mentioned the services: Palletizing (PAL) ($\beta = 1.442, p$ value < 0.0437), Distribution (DIS) ($\beta = 1.543, p$ value < 0.020) and Intermodal Management (IMA) ($\beta = 1.004, p$ value < 0.053). When assessing the coefficients of the variables we found that the provision of the distribution service (DIS) is important to determine the effect on the efficiency score in the cluster node. It was also found that the Milk Run service (MKR) has a significant negative effect ($\beta = -1076; p$ value < 0.073) in the efficiency score of logistics operators of this cluster. This result can be interpreted that the availability of this service in logistics operators type Node affects the transformation of inputs into outputs, resulting in reduced efficiency of logistics operators active in this segment.

The results of the regression model for the cluster logistics cluster operators (2) Network with assets is shown in Table 4. In the case of operators who are net with assets were identified seven service packages that present significant relationship (p value < 0.10) with its efficiency score.

Among the services that present a significant effect on the efficiency score of logistics operators in cluster net with assets, can be mentioned the services: Storage (STO) ($\beta = 1.604, p$ value < 0.027), Palletizing (PAL) ($\beta = 1.053, p$ value < 0.006),

Table 3 One inflated Beta regression for cluster of Node

Variables	Estimate	SE	t-value	p value
(Intercept)	-1.568	1.039	-1.509	0.137
STO	-0.306	1.033	-0.296	0.768
TPM	0.303	0.804	0.377	0.707
PAL	1.442	0.729	1.977	0.043**
CRD	0.549	0.471	1.165	0.249
JIT	-0.078	0.431	-0.183	0.855
IEC	-0.480	0.392	-1.222	0.227
RLG	0.053	0.403	0.133	0.894
FIS	0.3896	0.430	0.905	0.370
PRD	-0.921	1.064	-0.866	0.390
SUP	0.187	0.428	0.437	0.663
COO	-0.667	0.606	-1.101	0.276
DIS	1.543	0.643	2.399	0.020**
DTD	-0.444	0.483	-0.919	0.362
TRA	0.005	0.402	0.014	0.989
AKS	-1.076	0.588	-1.828	0.073*
IMA	1.004	0.507	1.978	0.053*
Observations	68			
Global deviance	64.91787			
AIC	102.917			

*Significant at 10%, **significant at 5% and ***significant at 1%

Cross-docking (CRD) ($\beta = -0.719$; p value < 0.016), import/export and customs clearance (IEC) ($\beta = -0.288$; p value < 0.014), Supply (SUP) ($\beta = -0.052$; p value < 0.022), Distribution, Door-to-door (DTD) ($\beta = -0.0306$; p value < 0.043) and Transfer service (TRA) ($\beta = -0.720$; p value < 0.039). Among the services provided by logistics operators of cluster Network with Assets, it is found that the Storage Services (STO) and Palletizing (PAL) are those with positive effect on the efficiency score, or significantly contribute to the better efficiency in transformation of inputs into outputs in the operation. The others present negative contribution in the efficiency score and impair the level of efficiency of logistics operators operating in the Network with assets segment.

The results of the regression model for the logistics operators of cluster (3) Network without Assets is shown in Table 5. Five service packages provided by logistics operators were identified that have significant relationship (p value < 0.10) with their efficiency score. Only the assembly kits and sets service (AKS) ($\beta = 0.148$, p value < 0.030) showed a significant and positive contribution to the performance efficiency. Other services, such as Packaging (PAC) ($\beta = -0.223$; p value < 0.025), Third-party management (TPM) ($\beta = -0.110$; p value < 0.040), Performance monitoring (PEM) ($\beta = -0.106$; p value < 0.028) and Intermodal management (IMA) ($\beta = -0.734$; p value < 0.007) showed a significant negative contribution to the performance of logistics operators that operate in Network without Assets.

Table 4 One inflated Beta regression for cluster of Network with assets

Variables	Estimate	SE	t-value	p value
(Intercept)	−0.024	0.766	−0.032	0.974
STO	1.604	0.725	2.212	0.027**
INC	−0.794	0.314	−2.527	0.011
PAC	−0.070	0.185	−0.381	0.703
AKS	−0.207	0.229	−0.905	0.365
TPM	0.016	0.173	0.094	0.925
PAL	1.053	0.383	2.746	0.006***
CRD	−0.719	0.299	−2.406	0.016**
JIT	−0.020	0.137	−0.150	0.881
IEC	−0.288	0.117	−2.464	0.014**
RLG	−0.023	0.176	−0.131	0.895
FIS	0.032	0.145	0.225	0.821
PRD	0.136	0.252	0.541	0.589
PEM	−0.020	0.236	−0.086	0.931
SUP	−0.552	0.241	−2.287	0.022**
COO	0.005	0.219	0.027	0.978
DIS	−0.282	0.227	−1.239	0.216
DTD	−0.306	0.182	−1.680	0.043**
TRA	−0.720	0.347	−2.070	0.039**
AKS	0.130	0.134	0.976	0.329
IMA	−0.057	0.134	−0.430	0.667
Observations	429			
Global deviance	136.075			
AIC	184.075			

*Significant at 10%, **significant at 5% and ***significant at 1%

4.1 Contributions to theory

In this paper, we use the one-inflated Beta model in the DEA 2nd stage as alternative to traditional regression models. The problem of non-adherence of DEA efficiency scores to Normal distribution harm parameter estimation in the using of ordinary least squares method. The possibility of data transformation, although reducing the previous problem, affects the interpretation of the coefficients in the equation (Simar and Wilson 2007). Among the models reported in the literature, despite the reported problems, the use of the OLS method regression is still often (Liu et al. 2013; Dokas et al. 2014; Hou et al. 2017). The use of Tobit regression with maximum likelihood method is also used in literature (Hoff 2007; McDonald 2009). However, Tobit partially solves the problem, since it is limited to the nonnegative interval (McDonald and Moffitt 1980), not being restricted to the interval of the DEA scores [0,1]. Even so, the use of OLS or Tobit regression is not recommended for the second stage of DEA analysis (Simar and Wilson 2007; Liu et al. 2013; Banker et al. 2015). This study seeks to solve the problem of the second stage regression through this model based on the Beta distribution. The general family of Beta distributions sounds more suitable, because it is also restricted to the [0, 1] interval, and, also, it can be parametrized to assume virtually

Table 5 One inflated Beta regression for cluster of Network without assets

Variables	Estimate	SE	t-value	p value
(Intercept)	0.301	1.808	0.167	0.868
STO	0.996	0.725	2.212	0.827
INC	0.815	1.963	0.415	0.678
PAC	-0.223	1.937	-1.154	0.025**
AKS	0.148	0.750	1.978	0.030**
TPM	-0.110	0.672	-1.638	0.040**
PAL	0.378	1.491	0.254	0.800
CRD	0.988	1.140	0.867	0.387
JIT	-0.201	0.367	-0.550	0.583
IEC	0.253	0.262	0.966	0.335
RLG	0.366	0.367	0.996	0.321
FIS	0.476	0.511	0.932	0.353
PRD	-0.036	0.858	-0.042	0.966
PEM	-0.106	0.597	-1.774	0.028**
SUP	0.561	0.435	1.291	0.199
COO	-0.458	0.591	-0.775	0.439
DIS	-0.111	0.711	-1.575	0.117
DTD	0.652	0.463	1.407	0.162
TRA	0.467	0.592	0.790	0.430
AKS	0.089	0.349	0.257	0.797
IMA	-0.734	0.270	-2.716	0.007***
Observations	138			
Global deviance	142.3534			
AIC	186.3534			

*Significant at 10%, **significant at 5% and ***significant at 1%

all distribution formats within the $[0, 1]$ range (Johnson et al. 1995). Secondly, we might consider that, also by construction, DEA scores are more concentrated at the proximity of 1 than of 0. Probabilistically, this effect gets stronger (i.e., more concentration right below or at 1) as the number of variables in a DEA problem grow. This specific characteristic of some Beta distributions is better captured by the One Inflated Beta distribution, which precisely models the truncation effect right before 1 (Ospina and Ferrari 2012).

4.2 Contributions to practice

This study permits the identification of service packages determinants to improve the operation efficiency of logistics operators. It was found that several available service packages significantly impair the level of companies' efficiency. This result presents important information relating to the decision to make available a suitable portfolio to the activity of logistics operators in Brazil. The results can also be used in other countries, especially those of continental dimensions where logistics costs tend to be an important part of the final consumer prices.

LSPs play a crucial role in making the operations of many industries more effective and efficient. The services provided by LSPs should be offered as efficiently as possible, allowing their managers to evaluate and monitor their operations continually. The search for efficiency must incorporate the selection of service packages that direct the operation to the efficiency frontier to ensure profitability of LSPs and therefore provide lower costs for shippers and end customers. Since logistics have become one of the determining factors in the competitiveness of the economy, the results of this study can contribute positively in this sense.

5 Conclusions

This study aimed to identify the service packages offered by logistics operators leading to the greater efficiency observed in LSP's sector. The classification of LSP's market identified three well defined clusters: (1) "Node", comprising the LSP's which mainly provide warehousing services; (2) "Network with assets", comprising the LSP's with own fleet operating in both storage services as transport; and (3) "Network without assets", which consists of LSP's with the same performance of the previous group, but without its own fleet.

With the determination of efficiencies by applying DEA models and the use of regression models was possible to identify service packages significant to the efficiency of the three logistics service clusters. Different for each group, the significant service packages vary in magnitude of effect in the efficiency measure. This effect tends to contaminate negatively efficiency measures of logistics service providers analyzed, while others affect positively. This observation suggests the existence of negative effect in certain packages on the service portfolio.

We also provided empirical evidence that DEA model as a classifier of relevant variables using the Principal Components Analysis and as modeling of important DEA scores using Beta Regression model, which are robust mathematical techniques.

References

- Adler, N., & Golany, B. (2007). PCA–DEA. In J. Zhu & W. D. Cook (Eds.), *Modeling data irregularities and structural complexities in data envelopment analysis*. Boston, MA: Springer.
- Adler, N., & Yazhemy, E. (2010). Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction. *European Journal of Operational Research*, 202(1), 273–284.
- Aguezoul, A. (2014). Third-party logistics selection problem: A literature review on criteria and methods. *Omega*, 49, 69–78.
- Aktas, E., Agaran, B., Ulengin, F., & Onsel, S. (2011). The use of outsourcing logistics activities: The case of turkey. *Transportation Research Part C: Emerging Technologies*, 19(5), 833–852.
- Aktas, E., & Ulengin, F. (2005). Outsourcing logistics activities in Turkey. *Journal of Enterprise Information Management*, 18(3), 316–329.
- Banker, R. D., Charnes, A., & Cooper, W. W. (1984). Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078–1092.
- Banker, R., Natarajan, R., & Zhang, D. (2015). Estimation of the impact of contextual variables in stochastic frontier production function models using data envelopment analysis: Two-stage versus bootstrap approaches. Working paper, Temple University.
- Castro, O. D., Jr. (2015). Performance of logistics operators 2007–2015. *Revista Tecnológica*, 252, 38–49.
- CEL. (2009). *Center for studies in logistics. Logistics overview: Outsourcing logistics in Brazil*. Rio de Janeiro: CEL/COPPEAD.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429–444.
- Coelli, T. J., Rao, D. S. P., O'Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis* (2nd ed.). New York: Springer.

- Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis, a comprehensive text with models, applications, references and DEA-solver software* (2nd ed.). New York: Springer.
- Dokas, I., Giokas, D., & Tsamis, A. (2014). Liquidity efficiency in the Greek listed firms: A financial ratio based on data envelopment analysis. *International Journal of Corporate Finance and Accounting (IJCFA)*, 1(1), 40–59.
- Dyson, R., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001). Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132, 245–259.
- Evangelista, P., & Sweeney, E. (2006). Technology usage in the supply chain: The case of small 3PLs. *The International Journal of Logistics Management*, 17(1), 55–74.
- Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31(7), 799–815.
- Fries, C. E. (2013). *Evaluation of the impact of the use of information and communication technologies on the efficiency of logistic service providers*. 193 f. Ph.D. thesis. Industrial and Systems Engineering, Federal University of Santa Catarina, Florianópolis, Brazil.
- Govindan, K., Khodaverdi, R., & Vafadarnikjoo, A. (2016). A grey DEMATEL approach to develop third-party logistics provider selection criteria. *Industrial Management & Data Systems*, 116(4), 690–722.
- Grönroos, C. (2000). *Service management and marketing: A customer relationship management approach* (p. 394). London, UK: Wiley.
- Guarnieri, P., Sobreiro, V. A., Nagano, M. S., & Serrano, A. L. M. (2015). The challenge of selecting and evaluating third-party reverse logistics providers in a multicriteria perspective: A Brazilian case. *Journal of Cleaner Production*, 96, 209–219.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (2010). *Multivariate data analysis* (7th ed.). New Jersey: Prentice Hall.
- Hoff, A. (2007). Second stage DEA: Comparison of approaches for modeling the DEA score. *European Journal of Operational Research*, 181(1), 425–435.
- Hosie, P., Sundarakani, B., Tan, A. W. K., & Koźlak, A. (2012). Determinants of fifth party logistics (5PL): Service providers for supply chain management. *International Journal of Logistics Systems and Management*, 13(3), 287–316.
- Hou, C. E., Lu, W. M., & Hung, S. W. (2017). Does CSR matter? Influence of corporate social responsibility on corporate performance in the creative industry. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-017-2626-9>.
- Ilos. (2014). *Institute of logistics and supply chain* (p. 2014). Rio de Janeiro: ILOS.
- Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd ed.). New York: Wiley.
- Lai, K. H. (2004). Service capability and performance of logistics service providers. *Transportation Research Part E: Logistics and Transportation Review*, 40(5), 385–399.
- Liu, J. S., Lu, L. Y., Lu, W. M., & Lin, B. J. (2013). Data envelopment analysis 1978–2010: A citation-based literature survey. *Omega*, 41(1), 3–15.
- McDonald, J. (2009). Using least squares and Tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, 197(2), 792–798.
- McDonald, J. F., & Moffitt, R. A. (1980). The uses of Tobit analysis. *The Review of Economics and Statistics*, 62(2), 318–321.
- Mello, J. E., Stank, T. P., & Esper, T. L. (2008). A model of logistics outsourcing strategy. *Transportation Journal*, 47(4), 5–25.
- Mitra, S. (2006). A survey of third-party logistics (3PL) service providers in India. *IIMB Management Review*, 18(2), 159–174.
- Norman, M., & Stoker, B. (1991). *Data envelopment analysis: The assessment of performance*. New York: Wiley.
- Ospina, R., & Ferrari, S. L. P. (2010). Inflated beta distributions. *Statistical Papers*, 51, 111–126.
- Ospina, R., & Ferrari, S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics & Data Analysis*, 56(6), 1609–1623.
- Resende, P., Sousa, P. R., & Oliveira, P. (2016). *Pesquisa de custos logísticos no Brasil*. Brasil: Fundação Dom Cabral.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Senthil, S., Srirangacharyulu, B., & Ramesh, A. (2014). A robust hybrid multi-criteria decision making methodology for contractor evaluation and selection in third-party reverse logistics. *Expert Systems with Applications*, 41(1), 50–58.
- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, 136(1), 31–64.

- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). New York: Pearson Education.
- Wu, J., & Zhou, Z. (2015). A mixed-objective integer DEA model. *Annals of Operations Research*, 228(1), 81–95.
- Yang, Q., Zhao, X., Yeung, H. Y. J., & Liu, Y. (2016). Improving logistics outsourcing performance through transactional and relational mechanisms under transaction uncertainties: Evidence from China. *International Journal of Production Economics*, 175, 12–23.
- Zacharia, Z. G., Sanders, N. R., & Nix, N. W. (2011). The emerging role of the third-party logistics provider (LSP) as an Orchestrator. *Journal of Business Logistics*, 32(1), 40–54.