Using Data Envelopment Analysis to Ranking ICMS's Taxpayers

Sergio Augusto Para Bittencourt Neto[1], Simone Borges Simao Monteiro[2], Joao Carlos Felix Souza[2], Ricardo Matos Chaim[2]

[1]Federal District's Revenue Department, Brasilia, Brazil

[2]Department of Production Engineering, University of Brasilia, Brasilia, Brazil

Email address:
sergioaugustoneto@gmail.com (S. A. P. B. Neto), simoneborges@unb.br (S. B. S. Monteiro). jocafs@unb.br (J. C. F. Souza),
ricardoc@unb.br (R. M. Chaim)

Abstract: This study presents a practical methodology developed in the R software language, which makes use of Data Envelopment Analysis, in the Constant Returns of Scale model, to measure the tax collection efficiency of the ICMS taxpayers (Brazilian tax on commercial operations related to the movement of goods and interstate and inter-municipal transportation and communication services), using as input the component variables of the tax calculation function found in the amounts recorded in the Electronic Invoices (purchases and sales) and in billing obtained with sales made with Card (credit and debit mode). The data corresponding to a fiscal year are obtained in the databases of the Brazilian revenue agencies, tabulated and submitted to the DEA calculation (multipliers and the envelope models). Thus, in a process of monitoring taxpayers belonging to the same economic sector, the lower relative efficiency performances of the companies will raise suspicion and serve to identify those that deserve to be audited (fiscal audit). Two examples of application of the explained methodology are demonstrated (Department Stores sector and Retailing of Footwear sector), where it is possible to observe its positive results in the identification of the taxpayers with low efficiency in the tax collection and eligibility for the inspection action. Currently the methodology is in use in the Federal District Revenue (Brazil) as an instrument for selecting companies for auditing.

Keywords: DEA, Taxpayer's Efficiency, ICMS, Fiscal Audit

1. Introduction

Defined in Article 155, II, of the Federal Constitution of Brazil (1988) [1], the ICMS is a state tax that represents the Brazilian taxation on the value aggregated through commercial purchase and sale transactions, transport, and telecommunications services.

As the most representative tax for the states - since it is the largest source of public revenue - the fight against tax evasion denotes a desirable effort to balance the public finances of these entities. Thus, the optimized use of available resources for their surveillance is a desirable condition.

The current ICMS monitoring model in the Revenue Department is dedicated to analyzing the performance of taxpayers in their respective economic sectors. Therefore, it is important to hold a comparison parameter that can distinguish and expose companies with inconsistent behaviors when compares to others companies which are subject to the same pattern of taxation charge.

Thus, this research aims to propose a simple and practical appliance, making use of a non-parametrical optimization method called Data Envelopment Analysis, to measure and ranking the relative efficiency of the tax collection among ICMS's taxpayers who participate of the same economic sector, which will help maximize the tax monitoring task, aiding in the identification of possible tax evasion events.

That said, the goal of this paper is to present a DEA ranking tool that illustrate the present discrepancies in the tax collecting behavior of taxpayers.

The classical DEA model will be used, in particular the

Constant Scale Returns (CRS) model, in the multipliers (primordial) and envelopes (dual) and output-oriented versions, to distinguish firms according to their tax collection efficiencies.

2. The ICMS Taxpayers Monitoring Process

The process of monitoring ICMS taxpayers include encompasses activities such as: choosing the economic sector to monitor, study of the relative taxpayer's behavior, selecting suspect taxpayers, analysis of the suspicious behavior in search of fraud hypotheses, notifying and/or audit and assessment of any perceived irregularity.

This process can be summarized by the following diagram: Stages of the Taxpayers Monitoring Process
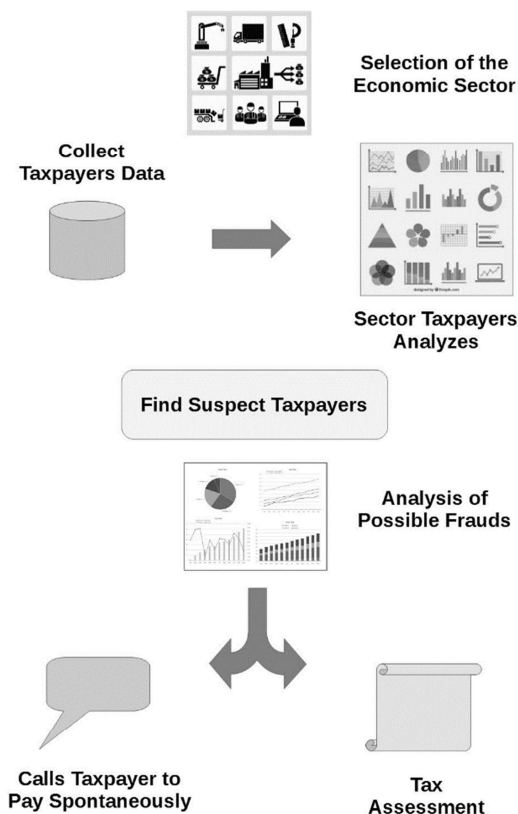

Figure 1. ICMS Taxpayers Monitoring Process.

As can be seen in the diagram, the leading activity of the monitoring process is completed with the selection of companies from the same economic sector of interest and the extraction of the pertinent fiscal data. The Collective fiscal information about these firms should be analyzed according to defined parameters and requirements - to select taxpayers who offer an outlier behavior worthy of in-depth investigation.

The audit of the tax data of the suspected taxpayers will ascertain the hypotheses of irregularities and tax frauds that may be practiced.

Some irregularities, of lesser offensive potential, can be remedied by notifying the taxpayer, encouraging him to spontaneously fulfill his tax obligation (without penalty).

In the case of noncompliance or in the case of fraud, a notice is filed that may lead to criminal prosecution.

Thus, the inaugural activity requires the classification of taxpayers for the selection of those who, due to their tax behavior, are shown to be interest for tax inspection.

One way of selecting to select taxpayers for a more detailed investigation is to compare their tax behavior to the performance of the tax collection among taxpayers of the same economic segment. It is correct to assume that these companies sharing the same parameters of tax requirements and tax applicable to the segment of economic activity condition that ensures homogeneity among agents.

It should be noted, however important, that this homogeneity assumption is incorrect if ones compares firms with different taxes burden. Therefore, it is not prudent to compare companies of distinct tax regimes of taxation (like Normal vs Simplified vs Encourages scheme for taxation). For this reason, the monitoring activity will always due segregate the taxpayers by type of method of calculation of ICMS and by economic sector.

Following the idea of carrying out the evaluation among the participating taxpayers the Data Envelopment Analysis technique is the appropriate tool for this purpose.

3. Data Envelopment Analysis (DEA) and the Methodology Used

3.1. Data Envelopment Analysis (DEA)

The DEA is a multivariate technique used to analyze the productivity efficiency of the Decision Making Units (DMU), which establishes an indicator of the relative efficiency management of the input and output of these units (DMU) and providing quantitative data on possible directions to improve the performance of the units, when they are inefficient.

Also known as Frontier Analysis, DEA is based on nonparametric linear mathematical programming models, therefore, it does not make use statistical inferences, nor does it cling to measures of central tendency, coefficient tests or formalization of regression analysis. In this sense, DEA does not require the determination of functional relations between the inputs and the outputs and allows the use of discretionary, instrumental or decision variables, non-discretionary or exogenous and categorical variables (including dummy) in their applications.

The good reputation of this tool comes from its relative simplicity and the wide applicability in several problems found in the real world. Virtually any condition that has multiple units (DMUs) that operate in a similar way and that is concerned with the standardization of performance of these units can make use of this technique.

The DEA defines the relative competitive positioning of a

set of organizations or activities by comparing their technical, scale and allocative efficiencies or inefficiencies. That methodology evaluates multiple resource and multiple products for each DMU. The ability of that entity to generate outputs for certain inputs defines its efficiency. It is understood that less efficient DMUs can improve their efficiency to the limit of the best DMUs, whose efficiency is 1.00.

Among the attributes that make up the DEA model are: 1) the relative efficiency of each productive organization (DMU) summarized with a single number that synthesizes the interactions between multiple inputs and outputs; 2) the possibility of identifying input economies or production increases, that would allow inefficient DMUs to become efficient.

For tax purposes, has as its operational predicate: 1) to allow the classification of the relative contribution of taxpayers belonging to the same economic sector; 2) provides the selection of taxpayers of interest for a better investigation of their fiscal regularity, based on their low relative efficiency performance.

This exposed, the DEA corroborates with the proposal of this paper that seeks to establish a useful methodology for the selection of taxpayers who have contributory behavior incompatible with the economic sector to which they belong.

3.2. The Methodology Used

DEA models can be input or output oriented, and generally consider multiple inputs and outputs in multidimensional spaces.

The outputs and inputs grow in the same proportion as the tax contribution ratio, equivalent to a theoretical average rate corresponding to each economic sector. In other words, they follow the average incidence rate of ICMS on the products traded by the segment under study. It is for this reason that the proposed methodology makes use of the DEA-CRS model of constant returns to scale, since sectoral taxation follows a constant pattern of tax burden.

CCR/CRS, named after it's developers Chames, Cooper and Rhodes, it's the first and fundamental DEA model, built on the notion of efficiency as defined in the classical engineering ratio. The CCR ratio model calculates an overall efficiency for the unit in which both its pure technical efficiency and scale efficiency are aggregated into a single value.

In the tax collection model, what matters is the movement of output towards the efficiency frontier, given the economic conditions experienced and practiced by taxpayers.

The proposed model is product-oriented, assuming that inputs do not vary - they remain constant, since they are economic achievements of firms - as production varies to reach the efficient production frontier.

Accordingly, it is assumed that the respective values of the economic movement (purchases and sales) are maintained, and the variation in ICMS tax collection pertinent to each taxpayer according to their relative efficiency.

In the DEA-CRS technique the relative efficiency of the DMUs can be calculated under two models: 1) Algebraic model of multipliers, dedicated to establishing the efficiency frontier by algebraic optimization of the weights of each input and output component, and 2) Envelopment model (dual).

The CCR/CRS Multipliers Model Oriented output is given through the solution of the following optimization expression:

Knowing that y is the product, x is the input, P is productivity, E is efficiency and variables μ and v are the weights coefficients.

$$Minimize \ \frac{1}{P_0} = \frac{\sum_{i=1}^{r} v_i x_{i0}}{\sum_{j=1}^{s} \mu_j y_{j0}} \qquad (1)$$

Subject to:

$$\frac{\sum_{i=1}^{r} v_i x_{ik}}{\sum_{j=1}^{s} \mu_j y_{jk}} \geq 1 \ \forall \ k = 1, 2, 3, \dots, z$$

$$v_i, \mu_j \ \geq \ 0 \ \forall \ i \ and \ j$$

Transforming fractional to linear programming:

$$Minimize \ \frac{1}{E_{f0}} = \sum_{i=1}^{m} v_i x_{i0} \qquad (2)$$

Subject to:

$$\sum_{r=1}^{s} \mu_r y_{r0} = 1$$

$$\sum_{i=1}^{m} v_i x_{ik} - \sum_{r=1}^{s} \mu_r y_{rk} \ \geq 0 \ \forall \ k = 1, 2, 3, \dots, z$$

$$v_i, \mu_j \ \geq \ 0 \ \forall \ i \ and \ j$$

The model allows each DMU to choose the weights for each variable - input (v) or output (μ) - in the most benevolent manner, provided that those weights applied to the other DMUs do not generate a ratio less than 1.

The CCR/CRS Envelopment Model - Oriented output is given through the solution of the following expression: Knowing that Φ is the inverse of efficiency $\frac{1}{E}$ (such that $1 \leq \Phi \leq \infty$) and λ represents the weights.

$$Maximize \ \Phi \qquad (3)$$

Subject to:

$$x_{i0} - \sum_{j=1}^{n} x_{ij}\lambda_j \ \geq 0 \ \forall \ i = 1, 2, 3, \dots, m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - \Phi . y_{r0} \geq 0 \ \forall \ r = 1, 2, 3, \dots, s$$

$$\Phi \ and \ \lambda_k \geq 0 \ j = 1, 2, 3, \dots, n$$

Because they are dual, the multipliers model has the same value as the objective function of the envelopment model. In

these types of DEA models, DMU's products and virtual inputs are the products and inputs that result from the end of the process of minimization (multipliers) or maximization (envelopment) by linear mathematical programming.

To give greater reliability to the efficiency results and goals of each DMU, it is necessary to consider eventual slacks in the projection of the efficiency frontiers. Therefore, the best model should consider the slacks:

Being $S+$ output slacks and $S-$ input slacks.

$$Maximize\ \Phi + \varepsilon \left(\sum_{i=1}^{m} S_i^- + \sum_{r=1}^{s} S_r^+ \right) \qquad (4)$$

Subject to:

$$\sum_{j=1}^{n} x_{ij}\lambda_j + S_i^- = x_{i0} \ \forall\ i = 1,2,3,\dots,m$$

$$\sum_{j=1}^{n} y_{rj}\lambda_j - S_i^+ \geq \Phi y_{i0} \ \forall\ r = 1,2,3,\dots,s$$

$$\lambda_j \geq 0 \ j = 1,2,3,\dots,n$$

The model is based on the fact that, in the same economic segment, the result of the collection should remain at a reasonable minimum variation, given the behavior of purchases and sales (for other companies or for the final consumer).

Such inputs information represents the economic and fiscal movement of the company directly related to the level of expected taxes collection.

Obviously, several factors may explain a taxpayer's discrepancy and deserve to be thoroughly appreciated by pertinent tax auditing techniques.

The robustness of the CRS model determines that if the DMU is inefficient, it is actually relatively inefficient. This is a condition that is a criterion for the classification of suspected taxpayers.

This is not to say that there are no problems in companies considered efficient nor that the inefficiency noticed doesn't have a reasonable and fair explanation.

However, this does not detract from the use of the DEA as a preliminary indicator of possible irregularity worthy of investigation or monitoring, since that is the idea of the model presented.

4. References in the Scientific Literature

The use of DEA as an instrument for the measurement of relative efficiency has been the subject of academic research since the late 1970s, when it was developed by Charnes, Cooper and Rhodes (Charnes *et al.* [2] and Cooper *et al.* [3]). The DEA method is widely used in many areas such as: manufacturing production (*e.g.* Zhu *et al.* [4]), banking system (*e.g.* Wanke *et al.* [5] and Gulati and Kumar [6]), educational system (*e.g.* Jauhar *et al.* [7] and Gimemez *et al.* [8]), health system (*e.g.* Safdar *et al.* [9], Gholami *et al.* [10] and Bahari [11]), management assessments (*e.g.* Zare *et al.*

[12] and Jorgensen *et al.* [13]), commerce (*e.g.* Yang *et al.* [14]) and other industries and organizations, including services (*e.g.* Wang *et al.* [15] and Chowdhury and Zelenyuk [16]).

Introduced by Charnes, Cooper and Rhodes in 1978 [2], the DEA is a particular application of Operational Research, which offers an appropriate solution to the problem of the relative efficiency calculation, based on a linear programming model. That methodology can now be systematized and easily solved using computer programs. Its utility is attested by the development of many publications announcing practicals solutions constructed through the use of the tool (*e.g.* Emrouznejad [17], Cook and Seiford [18] and Charnes *et al.* [19]).

The work of Charnes *et al.* [2] and then Banker et al. [20] and Seiford and Thrall [21], it stands out as the foundation of the DEA method. In this method, each composite unit is a convex combination of its reference units constructing a hypothetical ideal reference of efficient.

As shown in Tone [22] and Zhu [4], in several applied studies, the DEA has been used to provide insights into various activities and the benchmark identification (Zhu [23]). The authors add the notice of since that DEA was first introduced in 1978, to present days, researchers from different fields of knowledge recognize DEA as an excellent, and simple, methodology for modeling operational processes of performance appreciation.

From the lessons of Chames, Cooper and Rhodes [2], the DEA can be explained as a non-parametric technique, constructed in linear programming, for the evaluation of organizational efficiencies and performance measurement of operational units (decision-makers DMUs), who operate in the same branch of activity, when the presence of multiple inputs and outputs makes comparison difficult.

A definition of the DEA, given by Zhu [23], is that it is a tool with support in mathematical programming being a method that offers the estimation of the best frontiers of production and benchmarking in relation to the efficiency of multiple entities.

In addition, the author (Zhu [24]) formally defines the DEA as a boundary methodology and not for central tendency limits. In contrast to trying to fit a plane through of the data center as in statistical regressions, the model defines a linear surface fractionated over the observations. Due to this peculiar perspective, the DEA is particularly suited to discover relationships that remain hidden from other methodologies.

Kassai [25], building an accounting application for the tool, offers the DEA in the perspective of an efficiency bend (or maximized productivity curve) considering the optimal relation between inputs and outputs. This curve can be determined as a frontier of efficiency. Thus, the units considered efficient will be in intersection with this paradigm curve, while inefficient ones will be located under it. That efficiency frontier will serve as a benchmark for an inefficient unit aims to become efficient.

According to Kassai [25], the solution coming from the

application of the DEA can be summarized in: a) an envelope surface formed by the best performing (efficient) DMUs, which represent the reference set for the other units; b) an performance index, which means the distance of each unit to the frontier and; c) projections of inefficient units at the border, composing targets for these units. In addition, the author says, DMUs can mean business groups, individual companies, administrative units, provided that they comply with the requirements that: a) they are comparable; b) act under the same conditions; c) and the factors of inputs/outputs are the same, differing only in intensity and magnitude.

In the lesson of Ferreira [26], the DEA is a mathematical programming approach, alternative for the classical parametric statistical methods based on average or hypothetical maximum efficiencies, which provides an estimate of the relative efficiency by a border limit (efficiency), which informs points limits of productivity where a hypothetical productive unit is technically efficient. The idea of the DEA technique is to construct a convex referential setting where the DMUs can be classified into efficient and inefficient profile, having as a reference the outline of this surface area.

According to Casu and Molyneux [27], DEA is a mathematical programming model for the definition of the frontier of production (maximized) and Observation of the individual relative efficiency measure compared to the constructed frontier. For Ibrahim [28], The DEA solution measures the relative efficiency of each DMU compared to the best results presented. The maximum achieved performances indicate the frontiers of empirical production that set the limits to the achievable results with a given set of resources. The efficiency factors of a DMU are measured from the relative positions in collation with the established boundaries. Each result represents the descriptive of the abilities and the objective restrictions of the unit, assuming that, bypassing the restrictions and increasing the abilities, the results can be improved.

The initial DEA model was constructed by Chames, Cooper and Rhodes [2] and is called by its initials CCR, is to this day the most widely used model. This model has support in the definition of total unit efficiency, established as a ratio, that works with Constant Scale Returns (CRS). In the CCR model, weight weighting is associated to the inputs and output variables associated with the DMUs. Each double weight establishes the importance of the DMU in the composition of the input-output variables of the composite unit. The composite unit is a combination of efficient units. Thus, a given DMU is inefficient if the dual CCR model can present a hypothetical composite unit that surpasses it.

As Ferreira [26] teaches, the CCR model can be oriented to the inputs or to the outputs. Coelli points out [29] that the orientation towards the inputs tries to solve the question: observed the output standard of the unit, what is the possible reduction in the input, so as to maintain the current level of outputs? Concerning the models oriented towards the outputs, the answer is to the question: given the level of inputs used, which is the highest level of outputs that can be achieved by maintaining the level of these inputs constant?

Over the years, the applicability of the DEA has expanded, making it necessary new mathematical models to supply this new range of applications in different sectors. With this evolution, the models started to present modifications to the original model stemming from the incorporation of new concepts to each model. The DEA currently has a variety of models that from the classical DEA models (described in this article), until the approaches that combine DEA models with Monte Carlo simulation methods, sophisticated statistical models and fuzzy logic as shown by Tone [22], Emrouznejad [30] and Ghasemi *et al.* [31].

By providing a solution for measuring efficiencies between companies that share economic similarity, the DEA is an ideal tool for the identification and selection of the best eligible contributors for the analysis of anomalies, since it will be possible to o find the taxpayers with the least efficiency in terms of their taxes contributions.

5. The Data

This research uses information extracted from the databases of the Department of Revenue of the Brazil's Federal District.

Considering the data available to the state tax offices and the variables that explain the economic function that produces the tax ICMS (economic movement with goods and services), it is fair to use the following data for the composition of the DEA model in presentation:

OUTPUT: As the purpose of the methodology is to define a comparative list of efficient of the amount collected from the tax for the companies participating in a common economic segment, the only product of interest for the measurement of efficiency to be relativized in the input-output equation is the annual total tax payment (ICMS) by the participating companies. Specifically, the amount collected in the year 2017 of ICMS, under the revenue code 1517 (regular ICMS).

Considering that the choice of inputs is based on the parameters of economic movement of the commercial companies (their purchases and their sales), the model proposes use of the following variables:

INPUTS: Sum of Account Value of Electronic Receipt (Nfe's), documents generated, certified and approved electronically, which explains the bargaining movement of goods in a commercial transaction of purchase and sale, as well as the value of the goods and the levied ICMS. It is considered:

1. Purchase's Nfe, those representing the formation of commercial stocks of companies - purchases for asset formation manufacture or resale.
2. Sale's Nfe, those who report the sales between corporate taxpayers (input or resale) or the sale to the final consumer (individual or corporation).

The Nfe's (purchase or sale) can represent:

1. Internal Operations - when operations are carried out by

local taxpayers within the same federated unit, or

2. Interstate Operations - when operations are carried out between companies from different federated units.

This study will only considerer valid Nfe's, that is, those that have not been canceled for any reason.

Furthermore, much of the internal operations of sale to final consumer (especially individual) in the Federal District, are performed with the use of Tax Coupons (that will remain until its complete replacement by Electronic Consumer Tax Receipts).

The values of this commercialization to the final consumer, with the use of tax coupons, are not normally contemplated in NFEs. Nevertheless, they hold important values representative of the economic activity of taxpayers subject to ICMS taxation.

In order to reduce this predicament, considering that coupon information is not easily available for Revenue Organs, the billing movement of credit or debit card as a means of payment is used as input, since this type of operation represents approximately 60% to 70% of sales to the final consumer (especially the individual).

Finally, considering the desire to maintain homogeneity between the DMUs, the application of the DEA should focus on a cross section dedicated to the full calendar year (in the examples worked with the year 2017) and with the participation of only taxpayers in the active situation and submitted to the same tax regime since the previous year (*in casu* the regular regime).

The remote extraction of data from ORACLE database and the application of DEA technique in this study was carried out in R software under the RSTUDIO platform (Version 1.0.136), using respectively RODBC and Benchmarking packages.

6. Results – Analysis Procedures with Real Examples

Once chosen an Economic sector to analyze, values will be extracted from the fiscal information bases (Databases) of the Federal District Revenue through the ODBC connection between the ORACLE database and the R Software (RStudio platform), performed through the RODBC package that allows SQL queries directly, corresponding to the period under observation (as already said, 2017).

In reasons of mandatory tax secrecy, imposed by article 198 of the Brazilian Tax Code [32], the identification of taxpayers under study will not be disclosed.

6.1. Example 1 - Department Stores or Magazines

This first example will provide an analysis of the contributory efficiency of the taxpayer companies of the economic segments represented by the CNAE - G471300100 - Department stores or magazines.

The following table (table 1) presents the values of the inputs and the product to be used in the proposed methodology.

Table 1. *Values of output and inputs variables.*

(R$ x 1000)

Company	Tax Collected ICMS (product)	Total of Purchases (input)	Total of Sales (input)	Total of Cards (input)
GM-1	722	11548	654	9265
GM-2	1837	20846	4694	18965
GM-3	2161	26649	2962	23309
GM-4	988	11014	872	11966
GM-5	812	8891	772	8884
GM-6	1217	15934	978	14388
GM-7	836	8991	626	8104
GM-8	665	9086	718	7212
GM-9	207	2881	137	2976
GM-10	211	3153	158	2948
GM-11	215	3340	184	3047
GM-12	209	2632	169	2896
GM-13	206	2943	131	2934
GM-14	564	7146	510	6709
GM-15	494	5934	298	5728
GM-16	198	2362	205	2198
GM-17	4279	24469	34860	43388
GM-18	1464	8917	7835	10440
GM-19	849	4377	2158	5711
GM-20	1423	9962	2221	11419
GM-21	105	938	334	765
GM-22	8740	47509	7440	51666
GM-23	3077	16084	2946	17426
GM-24	2878	13723	3038	17966
GM-25	2695	12683	2342	16313
GM-26	1622	8585	2468	9753
GM-27	2071	10348	1996	10186

Company	Tax Collected ICMS (product)	Total of Purchases (input)	Total of Sales (input)	Total of Cards (input)
GM-28	1705	8590	2008	11193
GM-29	2379	24941	26591	23735
GM-30	272	3382	1665	5713
GM-31	140	2018	644	3716
GM-32	1255	4016	692	6789
GM-33	972	3197	515	5679
GM-34	498	1630	243	2276
GM-35	664	2122	373	955
GM-36	4839	13917	11236	24848
GM-37	327	1365	277	2568
GM-38	225	831	211	1644

In the Federal District there are thirty-eight active taxpayers belonging to economic sector of Department Stores or Magazines that meet the criteria established for the study in 2017.

From these companies were extracted the annual values corresponding to all ICMS collection of the period. These values will be used as the OUTPUT variable for the DEA model.

Likewise, the values related to total purchases, total sales, and total of credit or debit card transactions are extracted to compose the DEA model INPUT variables.

Applying the DEA-CRS (product-oriented) the following results are obtained regarding the efficiency of taxpayers in the economic sector.

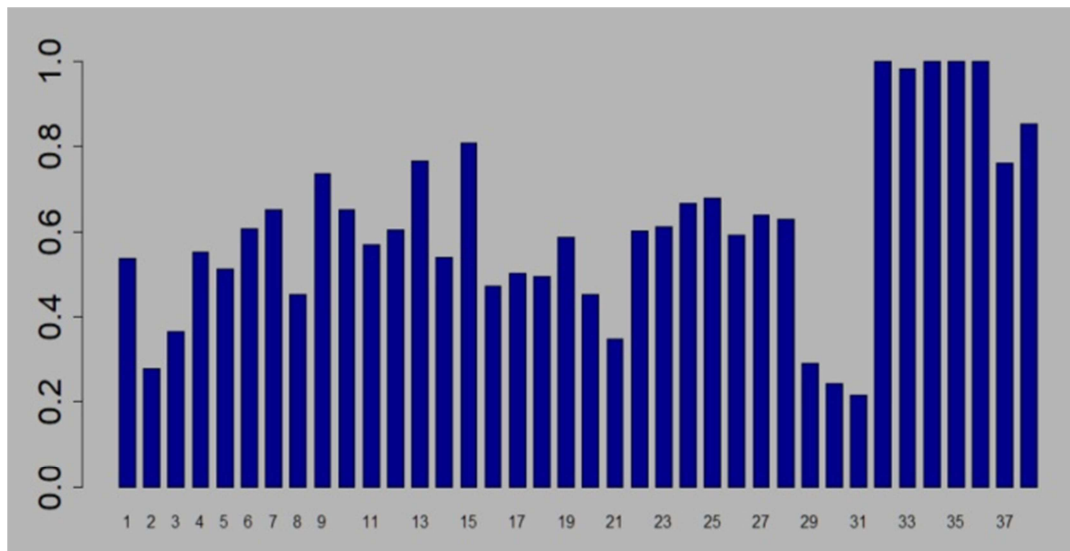DEA - Contributing efficiency in the economic sector of Department Stores or Magazines

Figure 2. Bar graph showing the result of the relative efficiencies of the 38 companies under study.

The graph (figure 2) demonstrates the individual result of the relative efficiency of each taxpayer in collecting the ICMS tax, as a weighted function of their commercial movement of purchases (inventory formation) and sales. Recalling that the paradigms of total efficiency are the companies that have efficiency index equal to 1.

The summary of the results of the efficiency indicators of the Department Stores or Magazines found by the DEA-CRS method can be seen as follows (table 2):

Table 2. Summary of Efficiencies.

Eff range	#	%	Min	0.2163
$0.2 \leqslant E < 0.3$	4	10.5	1st Qrt.	0.4971
$0.3 \leqslant E < 0.4$	2	5.3	Median	0.6024
$0.4 \leqslant E < 0.5$	4	10.5	Mean	0.6121
$0.5 \leqslant E < 0.6$	8	21.1	3rd Qrt.	0.7225
$0.6 \leqslant E < 0.7$	10	26.3	Max	1
$0.7 \leqslant E < 0.8$	3	7.9		
$0.8 \leqslant E < 0.9$	2	5.3		
$0.9 \leqslant E < 1$	1	2.6		
$E = 1$	4	10.5		

It is possible to say that, according to the results of the application of the DEA-CRS (output oriented) methodology, for the thirty-eight companies in the economic sector under analysis, only four companies had an efficiency index of 1, with the average efficiency for the sector equal to 0.6121, which is a bit far from the ideal expected.

The most efficient taxpayers in the sector in 2017 were: GM-32, GM-34, GM-35 and GM-36.

By analyzing the λs results (the largest participation as reference of the DMUs) it is possible to determine that the companie GM-34 serve as the Benchmarks of the model.

Also, no output slacks were reported in the model, *i.e.* the output slacks are all equal to zero.

The individualized result per taxpayer is shown in the following table (table 3) - considering $slacks_{output} = 0 \; \forall \; DMU$:

Table 3. *Projection of DMUs at the Border of Efficiencies.*

Taxpayer	Efficiency	Original	Expected	Dif	%
GM-1	0.54	722	1340	618	86
GM-2	0.28	1837	6580	4743	258
GM-3	0.37	2161	5895	3734	173
GM-4	0.55	988	1787	799	81
GM-5	0.51	812	1582	770	95
GM-6	0.61	1217	2004	787	65
GM-7	0.65	836	1283	447	53
GM-8	0.45	665	1471	806	121
GM-9	0.74	207	281	74	36
GM-10	0.65	211	324	113	54
GM-11	0.57	215	377	162	75
GM-12	0.6	209	346	137	66
GM-13	0.77	206	268	62	30
GM-14	0.54	564	1045	481	85
GM-15	0.81	494	611	117	24
GM-16	0.47	198	420	222	112
GM-17	0.5	4279	8500	4221	99
GM-18	0.49	1464	2958	1494	102
GM-19	0.59	849	1446	597	70
GM-20	0.45	1423	3143	1720	121
GM-21	0.35	105	302	197	188
GM-22	0.6	8740	14535	5795	66
GM-23	0.61	3077	5039	1962	64
GM-24	0.66	2878	4329	1451	50
GM-25	0.68	2695	3975	1280	47
GM-26	0.59	1622	2739	1117	69
GM-27	0.64	2071	3248	1177	57
GM-28	0.63	1705	2715	1010	59
GM-29	0.29	2379	8130	5751	242
GM-30	0.24	272	1117	845	311
GM-31	0.22	140	647	507	362
GM-32	1	1255	1255	0	0
GM-33	0.98	972	988	16	2
GM-34	1	498	498	0	0
GM-35	1	664	664	0	0
GM-36	1	4839	4839	0	0
GM-37	0.76	327	429	102	31
GM-38	0.85	225	264	39	17

The highlighted rows are the results of the most efficient DMUs (blue) and those with very low relative efficiency, qualified for further analysis of their fiscal situation (red).

In addition to the individual result of the calculated relative efficiency, table 3 shows the comparisons between the value of the tax collected and its projection at the efficiency frontier (expected value), giving the numerical difference and the percentage difference necessary for that the taxpayer reaches the maximum ideal of efficiency.

For the purposes of selecting taxpayers for auditing, in consideration of their relatively low relative performance, it is desirable that a threshold value be used to elect suspicious collection performance.

In the hypothesis under analysis, it is interesting to define that all the companies the relative efficiency results below 0.4 are considered with a suspicious performance being chosen for inspection.

Thus, six taxpayers with the worst relative efficiency performances, *i.e.* GM-2, GM-3, GM-21, GM-29, GM-30 and GM-31, should be evaluated in the audit process.

6.2. Example 2 - Retailing of Footwear

The second example will provide an analysis of the contributory efficiency of the taxpayer companies of the economic segments represented by the CNAE – G478220100 - Retailing of Footwear.

Table 4, below, considers the values of inputs and output required to apply the methodology to the chosen sector.

Table 4. *Values of output and inputs variables.*

(R$ x 1000)

Company	Tax Collected ICMS	Total of Purchases	Total of Sales	Total of Cards
	(product)	(input)	(input)	(input)
SP-1	830	2161	100	5777
SP-2	268	726	65	1906
SP-3	270	711	64	1752
SP-4	578	1830	64	3792
SP-5	357	911	41	2372
SP-6	267	719	80	1830
SP-7	609	1520	504	3410
SP-8	484	1149	451	2749
SP-9	93	1427	0	1567
SP-10	70	1083	0	1358
SP-11	127	1155	0	1463
SP-12	119	2967	0	2713
SP-13	570	2012	330	3772
SP-14	13	777	0	1195
SP-15	220	297	128	1084
SP-16	246	4152	131	4340
SP-17	116	1972	0	2161
SP-18	502	2082	267	3623
SP-19	191	655	38	1039
SP-20	4	734	0	1627
SP-21	295	932	45	1554
SP-22	420	2441	1053	3098
SP-23	437	2037	407	3519
SP-24	223	355	601	713
SP-25	1671	8074	1681	9256
SP-26	202	629	686	428
SP-27	512	3250	849	2643
SP-28	393	1713	290	2288
SP-29	899	4508	660	4480
SP-30	881	4155	614	5801
SP-31	714	2867	430	3305
SP-32	582	2565	454	2788
SP-33	676	2708	618	3102
SP-34	471	2516	513	3009
SP-35	702	3956	1262	6188
SP-36	720	3962	1137	6038
SP-37	8	1510	20	2017
SP-38	642	2761	9231	3168
SP-39	232	1441	143	1299
SP-40	218	990	613	813
SP-41	236	1199	451	1104
SP-42	301	1483	197	1670
SP-43	181	519	715	378
SP-44	412	2255	282	2602

In the economic sector of Retailing of Footwear, forty-four companies present themselves as candidates for the application of the DEA methodology, for the year 2017.

From these taxpayers are obtained the values: a) of the ICMS collected in the period and b) related to their operating economic variables of purchases and sales (including by card), to serve as output and input variables of the DEA model, respectively.

Applying the DEA (product-oriented) the following results are obtained regarding the efficiency of taxpayers in the economic sector.

DEA - Contributing efficiency in the economic sector of Retailing of Footwear.
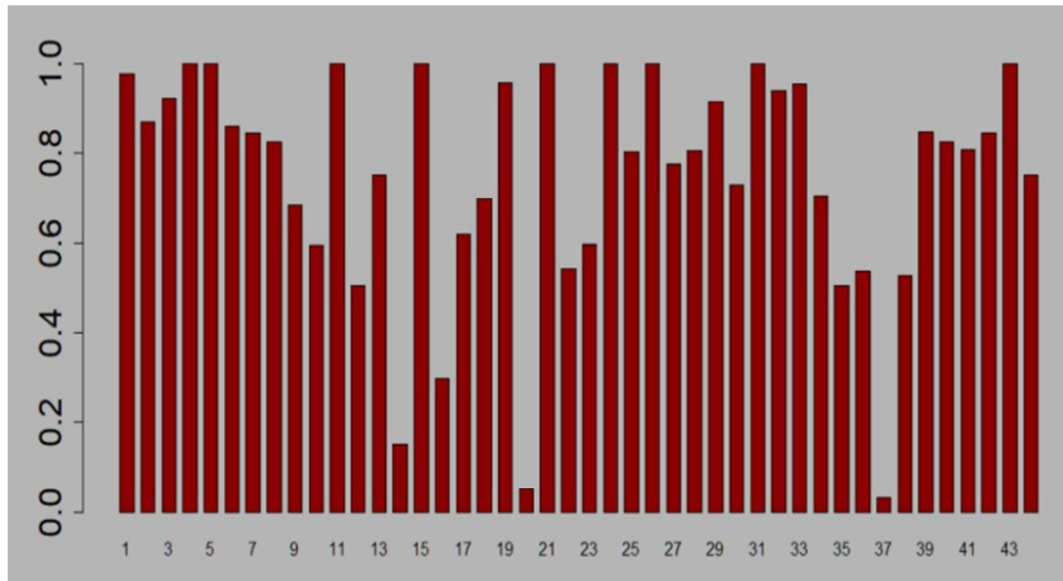
Figure 3. *Bar graph showing the result of the relative efficiencies of the 44 companies under study.*

Graphically (figure 3) it is demonstrated the disposition of the efficiency indices found by the DEA-CRS method application in the companies of the economic sector, obtained from the combination of the input and output factors presented.

It is possible to see a better uniformity in the behavior of taxpayer efficiencies in this economic sector compared to the first example.

The results obtained by the application of the DEA-CRS method for the entire Footwear Retail sector are summarized below (table 5), to provide a better picture of the distribution of their performance.

Table 5. *Summary of Efficiencies.*

Eff range	#	%	Min	0.0316
0 ≤E< 0.1	2	4.5	1st Qrt.	0.6131
0.1 ≤E< 0.2	1	2.3	Median	0.8169
0.2 ≤E< 0.3	0	0	Mean	0.7516
0.3 ≤E< 0.4	1	2.3	3rd Qrt.	0.9576
0.4 ≤E< 0.5	0	0	Max	1
0.5 ≤E< 0.6	6	13.6		
0.6 ≤E< 0.7	3	6.8		
0.7 ≤E< 0.8	6	13.6		
0.8 ≤E< 0.9	10	22.7		
0.9 ≤E< 1	6	13.6		
E = 1	9	20.5		

The performance obtained, based on the DEA-CRS (product-oriented) method for the forty-four companies participating in the economic segment (table 5), make evident that nine companies that have a maximum relative efficiency of 1, namely: SP -4, SP-5, SP-11, SP-15, SP-21, SP24, SP-26, SP-31 and SP-43.

By analyzing the λs (the largest participation as reference of the DMUs) it is possible to determine that the companies

SP-15, SP-21, SP-26 and SP-31 serve as the Benchmark of the model.

The average of the efficiencies presented by the companies was 0.7516. The output slacks are all equal to zero. There is no slack in the outputs.

The individualized result per taxpayer is shown in the following table (table 6) - considering $slacks_{output} = 0 \ \forall \ DMU$:

Table 6. *Projection of DMUs at the Border of Efficiencies.*

Taxpayer	Efficiency	Original	Expected	Dif	%
SP-1	0.98	830	849	19	2
SP-2	0.87	268	308	40	15
SP-3	0.92	270	292	22	8
SP-4	1	578	578	0	0
SP-5	1	357	357	0	0
SP-6	0.86	267	311	44	16

Taxpayer	Efficiency	Original	Expected	Dif	%
SP-7	0.85	609	720	111	18
SP-8	0.83	484	586	102	21
SP-9	0.68	93	136	43	46
SP-10	0.59	70	118	48	69
SP-11	1	127	127	0	0
SP-12	0.51	119	236	117	98
SP-13	0.75	570	758	188	33
SP-14	0.15	13	85	72	554
SP-15	1	220	220	0	0
SP-16	0.3	246	825	579	235
SP-17	0.62	116	188	72	62
SP-18	0.7	502	721	219	44
SP-19	0.96	191	199	8	4
SP-20	0.05	4	81	77	1925
SP-21	1	295	295	0	0
SP-22	0.54	420	773	353	84
SP-23	0.6	437	731	294	67
SP-24	1	223	223	0	0
SP-25	0.8	1671	2080	409	24
SP-26	1	202	202	0	0
SP-27	0.78	512	659	147	29
SP-28	0.81	393	488	95	24
SP-29	0.92	899	981	82	9
SP-30	0.73	881	1208	327	37
SP-31	1	714	714	0	0
SP-32	0.94	582	618	36	6
SP-33	0.96	676	706	30	4
SP-34	0.7	471	668	197	42
SP-35	0.51	702	1386	684	97
SP-36	0.54	720	1339	619	86
SP-37	0.03	8	253	245	3062
SP-38	0.53	642	1215	573	89
SP-39	0.85	232	274	42	18
SP-40	0.83	218	264	46	21
SP-41	0.81	236	292	56	24
SP-42	0.85	301	356	55	18
SP-43	1	181	181	0	0
SP-44	0.75	412	547	135	33

Table 6 presents the individual result of the relative efficiencies obtained by each DMU under analysis, also offering the efficiency frontier projection and the desirable tax collection targets for each taxpayer to achieve.

The highlighted blue lines show the companies that obtained the highest relative efficiency index (equal to 1).

Also, in red are the taxpayers who, by the established rule of result of suspicious efficiency, obtained index less than 0.4. Remember that this rule is the analyst's choice. In the example four companies are eligible to undergo fiscal audit procedures: SP-14, SP-16, SP-20 and SP-37.

With these two examples, it was possible to demonstrate the practicality and simplicity of the methodology proposed as an instrument for the monitoring of taxpayers and the selection of those who, due to their suspicious tax collection behavior, seem to be eligible for tax investigation.

7. Conclusion

This work agrees with the objective of constructing a useful analytical solution for the fiscal monitoring activity, which facilitates the selection of suspicious taxpayer companies, corroborating the fight against tax evasion that is more representative of state revenues.

The proposed tool consisted in applying the DEA to build a relative classification of collection efficiency by the company of an economic sector and thus to be able to identify the anomalous behaviors that require auditing.

As seen in the explored examples, the DEA model offered identifies companies that present the worst tax collection efficiency (ICMS), choosing them for audit research. It also makes possible the knowledge of the economic segment by the average behavior of all its participants.

For all of the above, the presented model corroborates to increase the *modus operandi* of the fiscal programming activity, since it provided a methodology of quantitative and objective selection of taxpayers. This means that the solution presented here contributes to the increase of successful results in the fight against ICMS tax evasion, since it has the impact of rationalizing the focus of the audits.

Certainly, the model should evolve, especially in order to contemplate electronic consumer tax documents - when it is fully mandatory in 2019, as well as progress is expected to be made to incorporate the peculiarities of each economic segment (when necessary).

Likewise, there is an opportunity for greater use of mathematical programming methods in tax audit procedures, since optimization solutions are desirable in a context of resource constraints, especially in human terms.

As a result of its practical application in the Federal District (Brazil) Revenue Office, it was possible to identify more than a hundred taxpayers offering extravagant tax practices and the respective companies were duly selected for auditing in 2018.

Finally, the model presented in this study was implemented as a working procedure for the monitoring of ICMS taxpayers in the Federal District - Brazil tax jurisdiction.

References

[1]   Brasil, Constituicao da Republica Federativa do Brasil. 1988.

[2]   Charnes, A.; Cooper, W. W.; Rhodes, E., Measuring the efficiency of decision making units, European Journal of Operational Research, no. 2(6), pp. 429444, 1978.

[3]   W. W. Cooper, L. M. Seiford, and J. Zhu, Handbook on Data Envelopment Analysis. Springer Science & Business Media, 2011.

[4]   J. Zhu, Data Envelopment Analysis: A Handbook of Empirical Studies and Applications. Springer, 2016.

[5]   P. Wanke, C. P. Barros, and A. Emrouznejad, Assessing productive efficiency of banks using integrated Fuzzy-DEA and bootstrapping: A case of Mozambican banks, European Journal of Operational Research, vol. 249, no. 1, pp. 378389, Feb. 2016.

[6]   R. Gulati and S. Kumar, Analysing banks intermediation and operating efficiencies using the two-stage network DEA model: The case of India, International Journal of Productivity and Performance Management, vol. 66, no. 4, pp. 500516, Apr. 2017.

[7]   S. K. Jauhar, M. Pant, and A. K. Nagar, Sustainable educational supply chain performance measurement through DEA and differential evolution: A case on Indian HEI, Journal of Computational Science, vol. 19, pp. 138152, Mar. 2017.

[8]   V. Gimnez, C. Thieme, D. Prior, and E. Tortosa-Ausina, An international comparison of educational systems: a temporal analysis in presence of bad outputs, J Prod Anal, vol. 47, no. 1, pp. 83101, Feb. 2017.

[9]   K. A. Safdar, A. Emrouznejad, and P. K. Dey, Assessing the Queuing Process Using Data Envelopment Analysis: an Application in Health Centres, Journal of Medical Systems, vol. 40, no. 1, Jan. 2016.

[10] R. Gholami, D. An Hign, and A. Emrouznejad, Hospital performance: Efficiency or quality? Can we have both with IT?, Expert Systems with Applications, vol. 42, no. 12, pp. 53905400, Jul. 2015.

[11] Bahari A., A. Emrouznejad, Influential DMUs and outlier detection in Data Envelopment Analysis with an Application to Health Care, Annals of Operations Research, vol. 223 (1), pp. 95108, 2014.

[12] M. A. Zare, M. T. T. Fard, and P. Hanafizadeh, The Assessment of Outsourcing IT Services using DEA Technique: A Study of Application Outsourcing in Research Centers, IJORIS, vol. 7, no. 1, pp. 4557, Jan. 2016.

[13] P. W. Jorgensen, D. C. Trotter, and T. R. Hill, Ecosystem services assessments in local municipal decision making in South Africa: justification for the use of a business-based approach, Journal of Environmental Planning and Management, vol. 59, no. 2, pp. 263279, Feb. 2016.

[14] Z. Yang, Y. Shi, and H. Yan, Scale, congestion, efficiency and effectiveness in e-commerce firms, Electronic Commerce Research and Applications, vol. 20, pp. 171182, Nov. 2016.

[15] S. Wang, E. Cheng, J. Zhu, C. Fu, and W. Wang, Using DEA Models to Measure the Performance of Public Culture Services in China, in 2016 International Conference on Computational Science and Computational Intelligence (CSCI), 2016, pp. 447452.

[16] H. Chowdhury and V. Zelenyuk, Performance of hospital services in Ontario: DEA with truncated regression approach, Omega, vol. 63, pp. 111122, Sep. 2016.

[17] A. Emrouznejad, Advances in data envelopment analysis, Ann Oper Res, vol. 214, no. 1, pp. 14, Mar. 2014.

[18] W. D. Cook and L. M. Seiford, Data envelopment analysis (DEA) Thirty years on, European Journal of Operational Research, vol. 192, no. 1, pp. 117, Jan. 2009.

[19] A. Charnes, W. W. Cooper, A. Y. Lewin, and L. M. Seiford, Data Envelopment Analysis: Theory, Methodology, and Applications. Springer Science & Business Media, 2013.

[20] R. D. Banker, A. Charnes, and W. W. Cooper, Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis, Management Science, vol. 30, no. 9, pp. 10781092, Sep. 1984.

[21] L. M. Seiford and R. M. Thrall, Recent developments in DEA, Journal of Econometrics, vol. 46, no. 12, pp. 738, Oct. 1990.

[22] K. Tone, Advances in DEA Theory and Applications: With Extensions to Forecasting Models. John Wiley & Sons, 2017.

[23] J. Zhu, Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets. Springer, 2014.

[24] J. Zhu, Data envelopment analysis: let the data speak for themselves. Leipzig: Amazon Distribution, 2014.

[25] KASSAI, Slvia, Utilizacao da Analise por Envoltoria de Dados (DEA) na Analise de Demonstracoes Contabeis, Departamento de Contabilidade e Atuaria. Faculdade de Economia, Administração e Contabilidade. Universidade de Sao Paulo, Sao Paulo: USP, 2002.

[26] C. M. D. C. F. FERREIRA, Introducao a analise envoltoria de dados: teoria, modelos e aplicacoes. UFV, 2009.

[27] B. Casu, C. Girardone, and P. Molyneux, Productivity change in European banking: A comparison of parametric and non-parametric approaches, Journal of Banking & Finance, vol. 28, no. 10, pp. 25212540, Oct. 2004.

[28] O. H Ibrahim, Handbook of Research on Strategic Performance Management and Measurement Using Data Envelopment Analysis. IGI Global, 2013.

[29] T. Coelli, A multi-stage methodology for the solution of orientated DEA models, Operations Research Letters, vol. 23, no. 35, pp. 143149, Oct. 1998.

[30] A. Emrouznejad and M. Tavana, Eds., Performance Measurement with Fuzzy Data Envelopment Analysis, vol. 309. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[31] M.-R. Ghasemi, J. Ignatius, S. Lozano, A. Emrouznejad, and A. Hatami-Marbini, A fuzzy expected value approach under generalized data envelopment analysis, Knowledge-Based Systems, vol. 89, pp. 148159, Nov. 2015.

[32] Brasil, Codigo Tributario Nacional. Diario Oficial da Uniao. 1966.