Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 175 (2020) 647–652

Procedia Computer Science

www.elsevier.com/locate/procedia

The 7th International Symposium on Emerging Inter-networks, Communication and Mobility (EICM)
August 9-12, 2020, Leuven, Belgium

Design and implementation of a system to determine tax evasion through de stochastic techniques

Noel Varela[a,*], Laura Patricia Carrasquilla Díaz[b], Omar Bonerge Pineda Lezama[c]

[a,b] *Universidad de la Costa, Barranquilla, Colombia*
[c] *Universidad Tecnológica Centroamericana (UNITEC), San Pedro Sula, Honduras*

Abstract

The dominant view on the study and practice of the tax service, mentions that taxpayers will always seek their own interest, for that reason taxes will be considered costs that could be avoided, unless the probability of being detected is high and the severity of the penalty is not an attractive option. This means that taxpayers would be motivated to develop tax evasion strategies that would increase their economic benefits. In this regard, the following two types of intentional tax evasion procedures are identified: one, where illegal practices are employed, such as failing to file a tax return without any legal justification to do so, and, another, where legal procedures are used to avoid filing a tax return. This paper establishes a methodology based on stochastic techniques, with the objective of creating a model that allows the identification of possible tax evaders of the value added tax (VAT) in Colombia.

Keywords: Tax evasion; Value added tax; Non-parametric stochastic model.

* Corresponding author. Tel.:+57-3235810446.
E-mail address: nvarela2@cuc.edu.co

10.1016/j.procs.2020.07.094

1. Introduction

1.1. Tax evasion in Colombia

This study focuses on the analysis of value-added tax (VAT), where each taxpayer is affected by this tax only in the proportion of value added to the final product, therefore, the final good will be taxed in the proportion of the final price [1]. This tax arose worldwide since the mid-20th century, except mainly for the United States, and is usually applied to all goods acquired, including those purchased outside the jurisdiction where this tax is declared.

The VAT that is paid by the consumers is collected by the producers, who have to declare it and pay it to the State. However, taxpayers could evade the declaration of these taxes by under-declaring or not declaring sales. This type of procedure is clearly illegal [2]. Another procedure for under-declaration of VAT is the creation of a tax credit, which is a legal procedure that serves to indicate that the VAT to be paid will actually be compensated with VAT generated on future purchases [3]. In Colombia, tax credit creation regulations do not regulate the time period for maintaining a tax credit or the limits that it may have, which leaves open the possibility that a tax credit may never be offset with VAT on future purchases [4].

VAT revenues are significant in Colombia and represent the highest percentage of total tax revenues [5]. This means that it is essential to develop mechanisms and strategies to prevent and mitigate VAT evasion. In [8], VAT evasion is estimated at 23.9% in 2017 and 24.3% in 2018. These authors made this estimate using the consumption method, which consists of calculating VAT from households' final consumption of goods and services taxed with VAT.

Finally, the present analysis of VAT evasion by taxpayers is restricted to the population related to the formal economy, that is, individuals related to the informal economy are excluded, see for example [9] and [10].

1.2 Identification of VAT evaders

The National Tax and Customs Directorate - DIAN makes continuous efforts to identify possible tax evaders, as well as analyzes the mechanisms used to carry out these frauds. Among the most recent fraud strategies detected by the DIAN are so-called "ghost companies," which are devices for producing sales receipts that reduce the value of taxes due, which led this control agency to reformulate Colombia's tax procedures [11]. In the specific case of detecting VAT evaders, the DIAN usually employs time-consuming procedures. Therefore, the use of automated rapid procedures is a necessity for this control agency.

This study proposes a methodology based on statistical techniques to identify VAT evasion in Colombia. The idea of this methodology is to associate atypicities in tax returns, or evasions, with unusual value behaviors that can be identified through regions of extreme values determined by Gertensgarbe and Werner's non-parametric statistical technique, which allows determining thresholds that differentiate usual values from those that are unusual. For this purpose, the behavior of an index associated to several individuals that keep some comparability is analyzed.

Then, first, the taxpayers, which can be natural or legal persons, which are organized in economic conglomerates defined by the International Standard Industrial Classification, so that in each of these conglomerates these atypicals are analyzed in the monthly behavior of a taxpayer's sales index. The atypical values of this index would be related to unusual tax practices.

This methodology is evaluated in a population of taxpayers that represented over 90% of VAT collections during the fiscal periods between 2014 and 2017 [12]. In order to determine cases of tax evasion, 32 taxpayers were selected, which showed the highest values of the sales index analyzed. For each taxpayer in this economic conglomerate, the analysis is deepened by including VAT returns related to the invoicing of the taxpayer, its suppliers and its buyers. The comparison of these invoices allows the description of the mechanism of VAT evasion, if it exists.

2. Method

The purpose of the investigation carried out was to determine a procedure that would allow the identification of potential taxpayers who commit tax evasion related to value added tax (VAT). The method of analysis consisted in

identifying atypical behavior over time of an indicator on sales made by taxpayers. Subsequently, the companies carrying these atypicals were analyzed in terms of variables related to the value added tax (VAT), so as to identify cases of inconsistencies on information related to this tax. For this purpose, monthly VAT information from January 2014 to December 2017 was analyzed. Due to the excessive volume of information available, together with technicians from the DIAN, it was decided to choose 10,000 taxpayers who together represented more than 90% of VAT collection. Among these taxpayers were both companies and individuals.

As for the indicator on monthly sales made by taxpayers, for a given taxpayer, the following index was calculated [13]:

$$Ita = (\text{total sales in month t}) / (\text{total sales in the year a}) \quad (1)$$

The total sales considered concern both local sales and exports. Therefore, this indicator takes values between 0 and 1 each month. In order to identify atypicals of this indicator, all companies related to each other in terms of their economic activities were considered. These groupings of companies were made taking into account the first five alphanumeric characters of the ISIC code. In addition, only clusters with at least 100 taxpayers were considered so that atypicality is assessed in a considerable number of values of the indicator. These groupings are called economic clusters [14].

Assuming that the values of the indicator associated to the contributors of an economic conglomerate presented similar behaviors through time, since these contributors were related among themselves through an economic activity, the idea of atypicality that was considered was the identification of values of the indicator that show different behaviors to the values of this indicator that indicate the majority of contributors of the conglomerate in analysis. The technique for identifying these outliers is presented below. In addition, once the taxpayers with atypical values were identified, those taxpayers that presented these atypical values were selected systematically, that is, throughout the entire analysis period, from January 2014 to December 2017 [15].

Later, considering the previously selected taxpayers, the analysis of information related to VAT was deepened. In this sense, the VAT declared by the taxpayer was compared with the VAT paid by its suppliers and the VAT withheld by its buyers. According to the normal VAT declaration procedure, it is expected that, in purchases made, the value of the VAT declared by the analyzed taxpayer is equal to the value of the VAT declared by the supplier, and, in sales made, the value of the VAT declared by the analyzed taxpayer is equal to the value of the VAT declared by the buyer. However, this does not necessarily happen in practice, and therefore the analysis carried out concentrates on the VAT declarations of the analyzed taxpayer which are lower than those reported by the buyer. Additionally, there may be tax credit declarations, without necessarily paying this VAT, so the process of accumulation of VAT in this type of tax credit is also analyzed.

The outlier identification technique used was that of Gertensgarbe and Werner [16], who proposed a non-parametric procedure based on the well-known non-parametric Mann-Kendall statistic (Mann, 1945, pp. 245-259, and Kendall, 1975, pp. 1-200), but to identify outliers in time series of meteorological variables. It should be noted that this technique does not consider the temporality of the analyzed data. This technique has been applied in other fields, such as insurance (see for example [17]) where the data are related to claims. In this case, it has been adapted to create a method in which atypical companies can be detected to find out whether they might be evading tax collection.

3. Results

The organization of clusters of contributors by the first 5 digits of the ISIC and considering clusters with at least 100 contributors in each of them, produced 23 clusters. In each of these clusters the index was constructed and is analyzed below using the Gertensgarbe and Werner technique.

Table 1 presents some statistics of four of the clusters formed, which will be followed up later as cases. The detail of the other clusters is omitted in order to keep the information used confidential. In the results of this table, it can be observed that there is great variability in the number of observations, a variable represented by N, that make up these clusters; in fact, a group that is not represented in this table has 19.3** observations. A relevant fact of the mean values presented, is that they vary very little among themselves, and in terms of maximum values this variability is

almost nonexistent. The presence of unusual values in these economic conglomerates is analyzed in the next subsection.

Table 1. Statistics of the selected sample of contributors, considering four economic clusters.

Case (group)	N (*)	Average	Maximum
1	5.9**	0,077854	1,00000
2	6.9**	0,088894	1,00000
3	4.9**	0,092541	1,00000
4	**8.3****	**0,086582**	**0,67985**

(*) The last two digits are omitted in full in order to keep the data confidentiality.

3.1. Estimation of thresholds to define unusual values

For estimating thresholds that identify regions with unusual values, the Gertensgarbe and Werner technique was applied to each of the 23 economic clusters formed, however, only the four cases introduced in Table 1 are analyzed below, in order to keep the information used confidential.

Figure 1 shows the increasing and decreasing curves of the ordered data of observed, and their crossings, for case 3 described in Table 1. The curves in this graph are standardized with respect to the number of observations represented on the x-axis, so that the information used is kept confidential.

The results of this figure show the presence of unusual values in the economic conglomerate analyzed for case 3. Similar results were obtained for the other conglomerates. In fact, the crossing points of the two graphed curves give the threshold at which the observations can be considered to be unusual values. This threshold is defined by the index value associated with the abscissa coordinate of that crossing point. The statistical significance of these thresholds is evaluated on the ordinate axis, through a value of the normal distribution corresponding to the crossing point. In this case, the null hypothesis is that the threshold detected is not statistically significant, so the association of the crossing point with ordinates greater than 3 or less than -3, leads to the rejection of this hypothesis, so it is concluded that the corresponding threshold is statistically significant. The presence of this type of value was notorious in the second case, but very marginal in the fourth case. In any case, the existence of these thresholds is important since it shows that, in terms of sales, some taxpayers behaved differently from the majority of taxpayers. It should be noted that, if the increasing and decreasing curves were not crossed, this would imply that no unusual values are detected in the analyzed economic group.

Table 2 shows statistical details of the thresholds found in each of the four analyzed economic clusters. The thresholds presented in this table are higher than the average values presented in Table 1, but much lower than the maximum values also presented in that table. On the other hand, according to the p-values of the thresholds indicated in Table 2, which allow the assessment of the null hypothesis that a threshold is not statistically significant, it is concluded that all the thresholds found are statistically significant, which means that the values above these thresholds can indeed be considered as unusual values. Additionally, it is observed that there are significantly small percentages of unusual values detected, as in the last two cases; while in the first two cases the unusual values are much more frequent.

3.2. Selection of taxpayers for VAT evasion analysis

In order to proceed with the analysis of VAT evasion, the taxpayers with the highest values in the index were selected and also those with at least 6 empirical standard deviations from the empirical average of this index. This produced 32 taxpayers.

For this group of 32 taxpayers, information on the invoicing of themselves, their suppliers and their purchasers was incorporated so that, based on the comparison of the monthly VAT returns of these actors, indications of VAT evasion are established. For this analysis, the amounts recorded in the credit notes related to the buyers are

eliminated from the VAT declared by the taxpayers, as well as the amounts recorded in the debit notes related to the suppliers are incorporated into this VAT.

The objective of the analysis of these 32 taxpayers is to identify those who show signs of VAT evasion. The finding of a significant number of potential evaders among these 32 cases would allow, under the proposed methodology, to determine a high probability of finding this type of evaders in a group with a large number of taxpayers. It should be noted that these 32 cases do not represent normal expected behavior, since these selected cases belong to the region of unusual values determined with the statistical technique used.

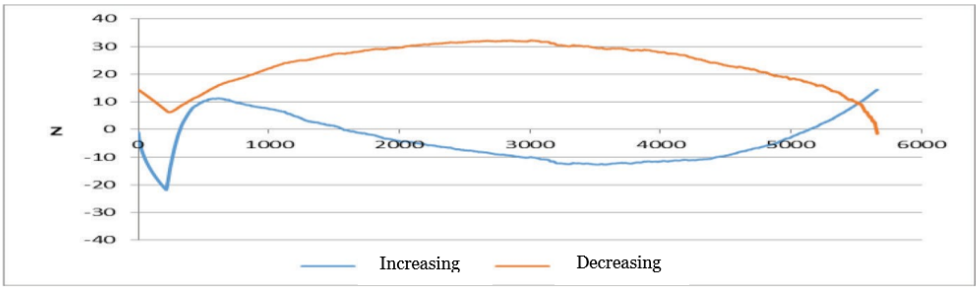

Fig. 1. Gertensgarbe and Werner curves, considering four economic conglomerates

All curves have been standardized on the number of observations

Table 2. Statistics on the thresholds found, considering four economic conglomerates.

Case (group)	Value of Ita	Threshold p-value	Percentage of unusual observations
1	0,13654	< 0,00001	8,7%
2	0,14794	< 0,00001	13,6%
3	0,19875	< 0,00001	3,7%
4	**0,13254**	**< 0,00001**	**2,2%**

3.3. Analysis of tax evasion due to failure to declare VAT.

One type of VAT evasion is the failure to declare VAT without reference to any legal arguments, which leads to the use of illegal procedures in the declaration of VAT. This type of evasion can be detected in accounting years by comparing the invoices made by the taxpayer and his clients or purchasers. More precisely, when a taxpayer sells goods or services to a buyer, then the VAT declared by both the taxpayer and the buyer must match. For the evasion analysis, it is interesting to identify the situations where the VAT declared by the taxpayer is lower than the VAT declared by the buyer.

Table 3. Statistics on evasion due to failure to declare VAT, amounts evaded over 30,000 USD

Case (group)	Maximum amount evaded (*)	Total amount evaded (*)
1	4*.***,**	10*.***,**
2	8*.***,**	25*.***,**
3	**10*.***,****	**18*.***,****

(*) The last four whole digits and the two decimal places are omitted in order to keep the data confidential and all periods are consolidated.

Table 3 presents a sample of the results obtained by analyzing the 32 companies indicated above with respect to evasion due to failure to declare VAT, amounts evaded in excess of 85,000 USD. For each of the three selected economic conglomerates, this table shows the maximum amount evaded and the total amount evaded. From this, the

maximum amount evaded can be important, but much more so the total amount evaded. It should be noted that the total number of taxpayers detected who practice this type of tax fraud constitutes almost half of the number of companies analyzed.

4. Conclusions

This study formulated a methodology based on stochastic techniques for the identification of taxpayers who present evidence of value added tax (VAT) evasion. This methodology considers conglomerates of taxpayers differentiated by economic activity, according to the coding of the International Standard Industrial Classification (ISIC). The application of this methodology to a population of taxpayers representative of the total VAT collection, showed the existence of taxpayers who presented unusual behavior in a sales indicator, which was denoted by Ita. From this latter population, a small sample was selected in order to examine the VAT returns on their invoices versus the VAT returns versus the invoices of their suppliers and buyers.

This analysis showed that a significant percentage of taxpayers do not declare the VAT reported by their clients, while an important group of taxpayers progressively accumulate tax credit, thus reducing their VAT returns.

References

[1] Olsen, J., Kogler, C., Brandt, M. J., Dezső, L., & Kirchler, E. (2019). Are consumption taxes really disliked more than equivalent costs? Inconclusive results in the USA and no effect in the UK. Journal of Economic Psychology, 75, 102145.

[2] Mitra, A. (2020). The impact of Belgium federal reform on economic well-being: a parametric and non-parametric synthetic control approach.

[3] Nguyen, T. T., Prior, D., & Van Hemmen, S. (2020). Stochastic semi-nonparametric frontier approach for tax administration efficiency measure: Evidence from a cross-country study. Economic Analysis and Policy.

[4] Wang, J., Shen, G., & Tang, D. (2020). Does Tax Deduction Relax Financial Constraints? Evidence from China's Value-Added Tax Reform. Evidence from China's Value-Added Tax Reform (February 17, 2020).

[5] Chan, S. G., Ramly, Z., & Zulkhairi, M. M. (2019). Value-added tax and economic efficiency: Role of country governance. Panoeconomicus, (00), 20-20.

[6] Sok-Gee, C., Zulkufly, R., & Mohd, Z. M. (2018). Value-added Tax and Economic Efficiency: Role of Country Governance. Panoeconomicus, 1-37.

[7] Rahimikia, E., Mohammadi, S., Rahmani, T., & Ghazanfari, M. (2017). Detecting corporate tax evasion using a hybrid intelligent system: A case study of Iran. International Journal of Accounting Information Systems, 25, 1-17.

[8] Triana, M., & Olivo, J. (2018). Análisis de las causales más recurrentes de inadmisión, derivado del proceso de devolución de saldos a favor por IVA en Colombia.

[9] Ojala, H., Kinnunen, J., Niemi, L., Troberg, P., & Collis, J. (2019). What Turns the Taxman On? Tax Aggressiveness, Financial Statement Audits and Tax Return Adjustments in Small Private Companies. International Journal of Accounting, Forthcoming.

[10] Waseem, M. (2018). Taxes, informality and income shifting: Evidence from a recent Pakistani tax reform. Journal of public economics, 157, 41-77.

[11] Rodríguez, J. A., & Ávila Mahecha, J. (2017). The tax burden on labor and capital income in Colombia: The case of the income tax and the VAT. Cuadernos de Economía, 36(SPE72), 99-138.

[12] Benavides, E. S., Brochero, N. A., & Viloria, A. (2019). Competitiveness indicator in Colombia through of multivariable statistics. In Communications in Computer and Information Science (Vol. 1122 CCIS, pp. 496–505). Springer. https://doi.org/10.1007/978-981-15-1301-5_39.

[13] Tan, S. K., Salleh, M. F. M., & Kassim, A. A. M. (2020). The Mediating Effect of Ethical Perception on the Relationship between Tax Service and Tax Compliance Behavior Using Baron and Kenny and Bootstrapping Method. Journal of Business Management and Accounting, 9(2), 41-49.

[14] Waseem, M. (2019). Does Cutting the Tax Rate to Zero Induce Behavior Different from Other Tax Cuts? Evidence from Pakistan. Review of Economics and Statistics, 1-45.

[15] Vallés-Giménez, J., & Zárate-Marco, A. (2017). Tax Effort of Local Governments and its Determinants: The Spanish Case. Annals of Economics and Finance, 18(2), 323-348.

[16] Contreras Rodríguez, T. Y., Bernal García, Y. S., & Cortes Devia, L. Y. (2017). Análisis histórico de la reforma tributaria: breve asomo a su implementación en Colombia año 2017.

[17] Quitian, O. I. T., Lis-Gutiérrez, J. P., & Viloria, A. (2020). Supervised and unsupervised learning applied to Crowdfunding. In Advances in Intelligent Systems and Computing (Vol. 1108 AISC, pp. 90–97). Springer. https://doi.org/10.1007/978-3-030-37218-7_11