International Conference on Applied Energy 2019
Aug 12-15, 2019, Västerås, Sweden
Paper ID: 0323

CRITICAL ANALYSIS OF THE ECONOMETRIC MODELS USED IN THE REGULATORY FRAMEWORK OF NON-TECHNICAL LOSSES IN BRAZIL ELECTRIC SECTOR

Marcelo Maestrini[1], Murilo Miranda[1] *, Daniel Viana[1], Arthur Tavares[1], Ana Carolina Chaves[1], Luiz Ozório[1], Guilherme Dantas[1], Priscila Mendes[2], Jairo Eduardo[3]

1 GESEL - Grupo de Estudos do Setor Elétrico. Av Pasteur, 250, sala 226, CEP 22290-250, Urca, Rio de Janeiro, RJ, Brazil.

2 RGE, Rua Jorge de Figueiredo Correa, 1632, CEP 13087-397, Jardim Professora Tarcília Campinas, SP, Brazil.

3 CPFL Piratininga, Rua Jorge de Figueiredo Correa, 1632, CEP 13087-397, Jardim Professora Tarcília Campinas, SP, Brazil.

ABSTRACT

The non-technical losses in Electric Sector are all the wastes of energy that occurs in the electricity distribution grid because of commercials and manageable aspects. For example, this losses are higher associated with electric theft. To incentive the concessionaires to have a good performance in combat these losses, there is a regulatory treatment that imposes some targets for this issue. In Brazil, the regulatory framework adopt a benchmark model to determine the target. In this model, the regulator uses a regulator econometric approach with socioeconomic data to subsidize the comparison between companies based on the complexity of each concession area. This article execute a critical econometric analysis of this methodology. For this, the article discuss the effectiveness of the model based on econometric test of estimations and residuals. Moreover, there is a specific analysis for the specification of variables and for the autocorrelations in the model. In the end, the paper demonstrate the weaknesses of the model to represent the reality of fighting against losses in the sector. As a result, there are contests about the choice of this type of model and the biases of omitted variables.

Keywords: Econometric Analysis, Regulation of Electricity Non-technical Losses, Socioeconomic Model, Panel Data with Random Effects.

1. INTRODUCTION

The cost of the energy theft in Brazil represented almost US$2 billion of loss in 2015 and corresponded for 5% of the residential market [18]. The Brazilian Electrical Sector Regulator (National Electrical Energy Agency – ANEEL), has been working to improve this area of its regulation through constant improvements in methodologies of reviews and readjusts since 1997.

In this context, some studies have been worked with the variables that are used in the econometric estimates[1] [2] [3] [4] [5], and they have pointed the direct relation between energy losses and low level of socioeconomics indicators. The main contribution of this study is to suggest solutions to improve the regulation models currently in force in Brazil for the next Periodic Rate Case.

Initially, the economic problem of non-technical losses (NTL) can be showed in three ways. First, from consumers' perspective, it is a common pool problem with a death spiral. This occurs because of each consumer, that start to thief electricity, implies in an increase in bill value and in a stimulus for news thieves. Second, from the distribution utilities perspective, that usually does not have your property right respected in regulation methodology construction and, thus, does not receive the optimal encourage to seek efficient results. Third, from the regulator perspective, which consists in the focus of this paper and represents a classical information asymmetry problem.

The regulator considered that companies have a similar capacity to manage their resources, so he uses the incentive regulation to achieve efficient results. Nevertheless, the main problem in this approach is to put companies in an equitable way when establishes theirs regulatory targets. For this, the ANEEL uses an econometric approach with socioeconomic data to compare companies based on the complexity of each concession area.

This article has the objective to exam the econometric model adopted by the Brazilian regulator to determine the regulatory targets. This is important because an inaccurate methodology that uses benchmark analysis will produces mismatch target for the companies and inefficiency in sector.

2. ANEEL METHODOLOGY'S REVIEW:

In Brazil, the incentive regulation model were implemented through Price Cap model shortly after the creation of ANEEL in 1997[6] and for this model, the Utility has a cost target in a regulatory period. The Brazilian regulator defines the target for each year in the review period and, if the concessionaire will not perform equal or above the target, they will not receive payments for their additional costs over the limits. The methodologies for NTL – henceforth, Setup of the Reduction Target (SRT) – bases on three stages: i) target definition; ii) start point; and, iii) the fall trajectory speed.

2.1 Initial target definition

The model uses a function of three vectors of variables to represent the NTL [10]:

$$NTL_i = C_i + X_i + IG_i \qquad (1.1)$$

Where:

- NTL_i is measured by electrical energy in low tension for the area i;
- C_i is a vector with variables that describe the intrinsic characteristics of the utilities that have influence in the level of no-technical losses for the area i;
- X_i is a vector with local characteristics of the concession area i;
- IG_i is a vector of variables relationship with utility manage capacity in area i;

As general rule, for estimate only the local difficulties that each regional utility deals with, an econometric model called "Complexity Model" (CM) is used. It compares different concessions areas through a benchmarking analysis based on the Yardstick Competition theory [7] to set the utilities' goals between rate cases. Used to estimate the relation of NTLs with socioeconomic problems for each utility, the CM uses selected socioeconomic variables in a random effects' panel data regression against the level of NTL measured in low-tension grid in each concession area (equation 1.2).

$$NTL_{i,t} = \beta_0 + \beta_1 * X_i + u \qquad (1.2)$$

Although the theoretical model for NTLs, presented by equation 1.1, contains the vectors of variables C_i and IG_i, the ANEEL model's just use X_i vector of variables to estimate CM. In that way, the regulator proposes three econometrical models (called C, G and K) that considered the following variables:

- vio: Violence - number of deaths due to aggression in area "i" in time "t" - DATASUS;
- pob2: % of people with per capita income less than 1/2 minimum salary in area "i" in time "t" - IBGE/IPEA;
- gini: Gini Index in area "i" in time "t" - IBGE;
- sub2: % of people in subnormal households in area "i" in time "t" - (Censo 2010) IBGE;
- lixo.u: % of households with garbage collection in area "i" in time "t" – IBGE;
- inad: Default of Credit Sector in area "i" in time "t" – BACEN;
- Mbr.Mb1Mbr: Low income market / (B1 total market) in area "i" in time "t" - SAMP;
- Mbr.Mbt: Low income market / (BT total market) in area "i" in time "t" - SAMP.

The estimation process executes a panel data with random effects and adjusts of autocorrelated errors. After this, the regulator calculates the probabilities for each concessionaire to be more complexity than other concessionaires are. Before probability matrix construction, the utilities are divided in two groups: i) group 1: companies with market major than 1.000 GWh/hour and more than 500 thousand clients, or companies with more than 15.000 Km of network; ii) group 2: other companies. In this way, group 1 represents large companies and group 2 represents small ones.

From the 3 previous models and through equation 1.5, there is one ranking for each model (C, G and K). The ranking classifies the firms by the complexity index. The highest scoring companies are those who find it more difficult to combat NTL in their concession area. These rankings are an important part of the methodology because they will determine the exception companies (incomparable ones).

$$\sum_{i=1}^{n} X_i^A * \beta_i \quad (1.5)$$

Where:

- X_i^A *= Variable value "i" for company "A";*
- β_i *= coefficient value of estimated regression for variable "i".*

The calculus of losses target for an area i for each company (as benchmark) uses the equation 1.6, where the smallest result is the target of the company i.

$$E_j = \mathrm{Pro}b_j * L_{bench} + [1 - \mathrm{Pro}b_j] * L_i \quad (1.6)$$

Where:

- $Prob_j$ *= probability of benchmarking to be in concession area more complex;*
- L_i *= no-technical losses divided by low tension Market of analyzed company;*
- $E(Company\ Target\ i)_j$ *= company "i" potential reduction in "j" model.*

Finally, the final target consists on the average of each target measured separately by models C, G and K.

$$E_{final,j} = \frac{\sum_{j=1}^{n} E(Target\ Company\ i)_j}{n} \quad (1.7)$$

Where:

- $E(Final\ Target\ Company\ i)$ *= potential reduction of the company i;*

2.2 Start point

Equations 1.8 and 1.9 present the general rule for the start point definition:

i. Group 1:

Maximum [7,50%; Minimum (3 cycle measured target, latest 4 years average)] (1.8)

ii. Group 2

Maximum [2,50%; Minimum (3 cycle measured target, latest 4 years average)] (1.9)

As exceptions, the model has three items showed below:

i. Companies that already have been performing with low losses
ii. Companies with low comparative probability
iii. Companies with target above than start point

2.3 The fall trajectory speed

The methodology consider three possibilities as below:

i. Group 1 companies with regulatory percentual (measured) above 7.50% and group 2 companies with regulatory percentual above 11.50%:

% Reduction Speed = % NTL regulatory/8 – 15/16 (1.10)

ii. Group 2 companies with regulatory percentual (measured) between 2.5% and 11.50%:

% Reduction Speed = 0,50% per year (1.11)

iii. Group 1 and 2 with calculated regulatory percent smaller than 7.5% and 2.5% respectively: will not be trajectory

3. METHODOLOGY

The propose of this is study is to examine the hold ANEEL methodology for NTL, and we will start the analysis by a criticism of that tool used by ANEEL in three ways: i) econometric analysis of CM, ii) criteria analysis about the variables used in the econometric model, and, iii) questions about the start point, goal point and fall trajectory definitions.

For the first point, using the ANEEL current metrics and the public audience data's[8] [9] [10], the whole three models were estimate using Stata® software and some report problems with ANEEL documentation were found. As can be checked in public audience part at the technical notes, some agents already point out this problem with data for results documentation of K and G models.

Furthermore, the key point of the econometric analysis were made through the results of three tests: i) Hausman test [11], that checks if it is suitable to run random panel, ii) serial correlation in residues, that can denote if the model has some problem of omitted variable, and iii) residuals normality by histogram and Jarque-Bera test [12] [13].

Finally, to examine the two last points, this study brings some qualitative analysis about the set of variables used in section 5, and, the construction of the rules for start point, trajectory and goal point in the section 6.

4. ECONOMETRIC ANALYSIS OF COMPLEXITY MODEL:

The CM uses a panel data with random effects for the estimates, but it seems that without considering the best fit of the choice between random or fixed effect. [14] says that to decide between fixed or random effects, it can be run a Hausman test. Doing the test for the three models, all of then rejected the null hypothesis, thus, the correct way to estimates the model should be in fixed effect panel. By itself, this option made by ANEEL invalidates the whole methodology and its conclusions.

It is important to say, the correct way to work with random panels would be do the Hausman test for each new model specification, in other words, if at any time were necessary to adjust some model variable, would be necessary to run the test again to confirm the correct uses of random effect.

As a second test, the parameters for each model were estimated to compare the results. For first model – letter C – the results found match with the reported in [10], but for models G and K, the results did not match with reported in the same technical note. For the parameters, this problem is not a big problem, but as the methodology use the standard error to calculate the probability matrix, it can make some mismatch with the benchmark.

The serial correlation test of residuals proves that there are omitted variable in the model specification was done. However, even knowing about the problem, the regulator choose to do nothing in an adjustment way of the serial correlation, because, in his words, this problem does not influence the ordering of the concession areas. Therefore, isolated, this simplification of the treatment maybe could not do a relevant influence in simple ordering of area if it were considered just the coefficients, but, for probability matrix, that uses each standard error, it will be a relevant influence in the benchmark match.

Even though the residuals have approximately zero of average, it is clear that the residuals distributions are not normally distributed. The residuals normality can be seen either by [12] and [13], that shows the probability of the residuals be normal is approximately zero for all models.

5. VARIABLE ANALYSIS

The main problem can arise in the good variable availability and it is related with a desirable aggregation and frequency of updating. In this way, data institutions/reports, as IBGE, CENSO and PNAD, have been the best way to estimate the socioeconomic impact.

Since the second cycle, the selection of the socioeconomic variables set has varied according to the different tariff review periods. However, it is possible to verify that the socioeconomic variables present in ANEEL models normally represents 5 main dimensions [15]: (i) violence, (ii) income, (iii) precariousness, (iv) infrastructure and (v) impairment of income.

As a rule, the determination of the variables has followed some criteria's: information reliability, statistical significance of the coefficients; sign of the coefficients; and the increase in R^2. The first critic about the variables is that the last CENSO update was in 2010. The second critic is that there are variables with large correlation and probably that are capturing almost the same effect (e.g. poverty and infrastructure. Moreover, there is an unavailability of sufficiently disaggregated data (e.g. PNAD) or in the annual update frequency (e.g. CENSO). Thus, some companies may not be represented with the correct socioeconomic aspects involved in the NTL problem, mainly for the great complexity in this issue.

6. SETUP OF THE REDUCTION TARGET (SRT)

In SRT, the first important rule is the company's segregation between Group 1 and 2. However, there is no mention about the origin of the criteria adopted to share into these two groups. One important thing that must be said is the relevant impact of this rule in the benchmark concession area choice, since these groups restrict the universe of companies compared in probability matrix.

In this way, a cluster analysis were done to test if ANEEL segregation were efficient. So, the insure is the cluster formation must consider the maximum likelihood intergroup and minimum variance intragroup, and in this way were run the method K-means, defined in [16].

Through this analysis, the efficient segregation for 2 groups would have the following setup for Group 2: i) maximum of 702 thousand of market and maximum of 324 thousand of consumers; or, ii) maximum of 204 thousand of network. The other companies compose the Group 1.

Although the selection of number of groups may be subjective in some analyses, there are some techniques based on information criteria for the selection of the optimal number of clusters. These techniques penalize the increase in the number of clusters in while occur informational gain by a greater disaggregation. Based on

[17], the present paper defends that the optimal number of clusters would be 3 for distribution sector in Brazil.

7. CONCLUSIONS:

So, as mentioned above, the key points of this analysis is focus on the critique of the econometric model used in the estimation of the complexity model, since ANEEL model's is fragile when presenting an autoregressive AR1 process in the residues, which leads to believe that quadratic errors may be interfering in the probability matrix that identifies the Benchmarking. Another important econometric problem is the non-normality of model errors, which denotes the presence of heteroscedasticity.

Furthermore, the random panel used by ANEEL is inadequate because uses NTL from the whole distribution utilities in the sector, when this kind of econometric analysis just should be used with data from smaller samples. Thus, it makes sense the caught of aleatory effects that would be originated from aleatory sample.

Additionally, other problem addressed by the distributors regarding the limits for classification in small/large companies, which the cluster analysis showed his inefficiency, and has demonstrated the optimal number of groups would be three instead two as defined for ANEEL (using the same criteria of regulator). Is important to highlight the use of a larger number of groups could have positive effects on the treatment of outliers, which could imply into more robust results of the complexity model.

Finally, there is a multicollinearity among the variables with the closest dimensions and the databases periodicity updates. All these issues together demonstrate the weaknesses of the model to represent the reality of fighting against losses in the sector. As a result, there are contests about the choice of this type of model and the biases of omitted variables.

ACKNOWLEDGEMENT

The authors would like to thank CPFL group for the technical and financial support, through the Research & Development project with R&D ANEEL resources.

REFERENCE

[1] SMITH, Thomas B. Electricity theft: a comparative analysis. Energy policy, v. 32, n. 18, p. 2067-2076, 2004.

TENDÊNCIAS. Consultoria Integrada. Determinantes Estruturais de Perdas Não-Técnicas no Setor Elétrico Brasileiro 2005.

[2] ALMEIDA, A. C. O Componente Social das Perdas e Inadimplência na área de concessão da CERJ. XVI Seminário Nacional de Distribuição de Energia Elétrica, 2004.

[3] ARAUJO, A. C. M (2007). Perdas e Inadimplência na Atividade de Distribuição de Energia Elétrica. Tese (Doutorado em Ciências e Planejamento Energético) - Coordenação dos Programas de Pós-Graduação de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro.

[4] AEQUUS. Estudo sobre as condições socioeconômicas dos municípios que compõem a área de concessão da Espírito Santo Centrais Elétricas S.A. (Escelsa), 2007.

[5] FRANCISCO, Eduardo de Rezende et al. Development of an indicator of propensity to energy commercial losses using geospatial statistical techniques and socio-economic data: the case of AES Eletropaulo. RAM. Revista de Administração Mackenzie, v. 11, n. 4, p. 178-197, 2010.

[6] RESENDE, M., 2002. Relative efficiency measurement and prospects for yardstick competition in Brazilian electricity distribution. Energy Policy, v. 30, n. 8, p. 637-647.

[7] Shleifer, A., 1985. A theory of yardstick competition. The RAND Journal of Economics, p. 319-327. https://www.jstor.org/stable/2555560

[8] ANEEL - Agência Nacional de Energia Elétrica. Nota Técnica nº 342/2008. Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica, segundo ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica – Resultado Final. Superintendência de regulação econômica (SER). Brasília, 2008.

[9] ANEEL - Agência Nacional de Energia Elétrica. Nota Técnica nº 298/2011. Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica, terceiro ciclo de revisão tarifária periódica das concessionárias de distribuição de energia elétrica – Final. Superintendência de regulação econômica (SER). Brasília, 2011.

[10] ANEEL - Agência Nacional de Energia Elétrica. Nota Técnica nº 106/2015. Metodologia de tratamento regulatório para perdas não técnicas de energia elétrica - Audiência Pública 023/2014 Final. Superintendência de regulação econômica (SER). Brasília, 2015.

[11] HAUSMAN, J. A. (1978). Specification tests in econometrics. Econometrica: Journal of the econometric society, 1251-1271.

[12] JARQUE, C., BERA, A., 1980. Efficient tests for normality homoscedasticity and serial independence of regression residuals. Econometric Letters 6, 255–259.
[13] JARQUE, C., BERA, A., 1987. A test for normality of observations and regression residuals. International Statistical Review 55, 163–172.
[14] GREENE, W. H. (2003). Econometric analysis. Pearson Education India.
[15] ANEEL - Agência Nacional de Energia Elétrica. Nota Técnica nº 188/2014. Atualização do índice de Complexidade Socioeconômico adotado para definição do nível regulatório de perdas não técnicas na distribuição - Audiência Pública. Superintendência de regulação econômica (SER). Brasília, 2014.
[16] HARTIGAN, J. A. (1975). Clustering Algorithms. New York, NY: Wiley, 1975. - 364 p.
[17] T. CALINSKI AND J. HARABASZ, "A dendrite method for cluster analysis," Communications in Statistics, vol. 3, no. 1, pp. 1–27, 1974. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/03610927408827101
[18] BRASIL, I. A. Perdas comerciais e inadimplência no setor elétrico. Instituto Acende Brasil, Tech. Rep, v. 18, 2017.