Electricity Demand Modeling for Rural Residential Housing: A Case Study in Colombia

Wilson O. Achicanoy M.
Departamento de Electrónica
Universidad de Nariño
San Juan de Pasto, 520002, Colombia
Email: wilachic@udenar.edu.co

John Barco Jiménez
Departamento de Electrónica
Universidad de Nariño
San Juan de Pasto, 520002, Colombia
Email: johnbarco@gmail.com

***Abstract*—A model for the electricity demand in a rural area of Colombia is developed. The methodology used to build this model is based on the exploration of cross-categorical static and linear models based on robust regression, and using the most representative variables of the processes of electricity consumption in the region. The transversal observations used in the regressions were retrieved from the energy survey conducted in Nariño's rural area in 2013. The demand model will allow, in the future, proposing policies to increase coverage and/or use of alternative energies.**

***Index Terms*—Electricity demand, residential rural housing, econometrics, linear regression, alternative energy sources.**

I. Introduction

The Department of Nariño is located in southwestern Colombia, on the border with Ecuador. Its rural population is characterized because it makes use of renewable energy sources such as wood, or firewood, in daily processes such as cooking or water heating. Undoubtedly, and in the near future, the uncontrolled use of these renewable energy sources will cause, among other problems, the extinction of native forests in the region and the decline in the quality of life of people living in the homes. Respiratory diseases are common in family groups living in homes where cooking is done with firewood.

Therefore, the development of a model of electricity demand for rural residential housing in Nariño will help to understand the dynamics of the daily processes that the habitants of the rural area are developed; and especially, it will help to formulate policies or strategies for energy replacement or using of alternative energies, which, if they are framed in topologies of intelligent networks, can respond efficiently to the needs of the population and constraints imposed by electricity network distributors. The results of this modeling could even be extrapolated to other regions, especially in Latin America.

This work has been supported by ALTERNAR Project, BPIN 2013000100089, Acuerdo 005, OCAD-Fondo CTeI, SGR, Colombia.

W. Achicanoy is Assistant Professor of Departamento de Electrónica, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia (wilachic@udenar.edu.co).
J. Barco is Professional Researcher of Departamento de Electrónica, Universidad de Nariño, San Juan de Pasto, Nariño, Colombia (johnbarco@gmail.com).

The demand model proposed here is described in more detail in [1]. It uses only the information provided by the energy survey PERS-Nariño-2013, which was made during the months of June, July and August of the year 2013, on a representative sample of the Nariño's rural population. Due to the high level of dispersion and non-correlation in the potential explanatory variables and the power consumption observations reported in the survey, it is not an appropriate idea of getting a single model that explains all the consumption processes of rural households of Nariño; then, we propose a methodology based on categorical linear models to construct a more realistic model.

The most important processes in the energy consumption of Nariño's rural households can be summarized as follows: lighting, refrigeration, electric and electronic devices, water heating, cooking and climate control. This last process includes the use of heating systems; however, in the rural area, few homes have these types of systems. Also, the survey's observations show that the main energy source used for cooking is firewood, and the use of electric stoves is virtually nil.

Unlike to the study of residential demand in other countries, for instance the study of residential demand for electricity in Chile [2], or the short term micrometric analysis of residential electricity demand in Spain [3], the Nariño's electric rural residential model results from an exercise of categorization, which follows determined rules for classification of the survey's observations and uses dummy variables, which in turn provide high levels of goodness of fit of a set of proposed regressions.

The paper is organized as follows: Sec. II describes the process and variables associated to power consumption. Sec. III summarizes the best outcomes obtained from categorized and non-categorized regressions. Sec. IV explains the best selected models for demand, and in Sec. V conclusions of this work are presented.

II. Processes and variables associated to power consumption

The participation of each process in electricity consumption in Nariño's rural households is shown in Fig. 1. It is noted that the processes of greater participation correspond to (in descending order): (i) lighting, with 34%; (ii) refrigeration, with 33%; and (iii) electric and electronic devices, with 27%. These processes represent 94% of electricity consumption. The remaining 6% of electricity consumption is determined by (in descending order): (iv) water heating, with 3%; (v) cooking, with 2%; and (vi) climate control, with 1%.

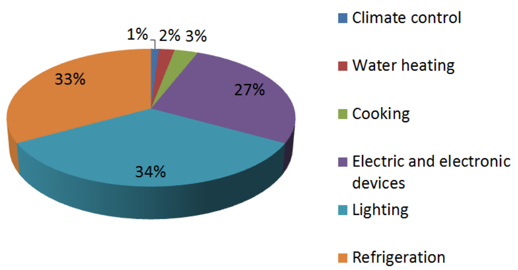

Fig. 1. Participation of each process in the consumption of electricity in Nariño's rural households (kWh/month). Source: SIPERSN: http://sipersn.udenar.edu.co:90/sipersn/.

The PERS-Nariño survey also reports the values of the consumption of other available energy sources in the area, such as liquefied petroleum gas (LPG), coal and firewood. In Fig. 2, the participations of each energy source in the energy consumption of rural housing of Nariño are shown. It is noted that the energy source of greater participation in energy consumption is firewood, with 87%; followed by LPG, with 8%, and electricity, with 5%. Use of coal in rural houses has virtually no involvement, it is 0%.

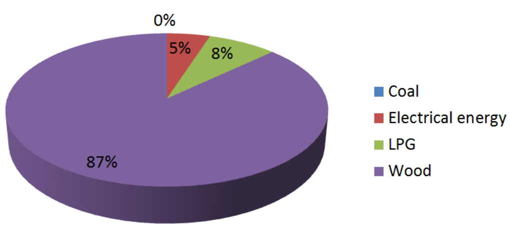

Fig. 2. Participation of each energy source in energy consumption of rural housing in Nariño (MCal/Month). Source: SIPERSN: http://sipersn.udenar.edu.co:90/sipersn/.

The variables that describe the processes of consumption of electric energy, other energy sources different from electricity, and other variables related to the household itself, as the number of people and rooms per house, have proven to be significant in models of residential demand [2]. Here, these variables are included as explanatory variables, quantitative and qualitative.

The dependent variable in the model proposed here corresponds to the electricity consumption reported by rural households. According to data reported by the survey, and that are detailed in Fig. 3, 97% of rural households (2,405) have the electrical service, and only 3% (74) of them do not have or do not use this service. From the 97%, the 86% (2,132) access the service through the public network, 7% (174) do this through a municipal diesel generator, 3% (74) take the service from a shared portable diesel generator, and 1% (25) has its own portable generator.

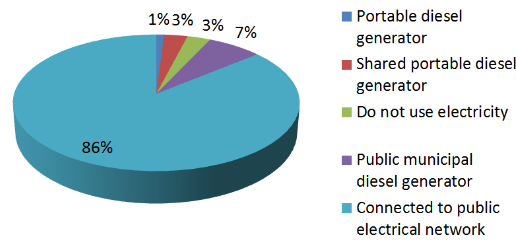

Fig. 3. Households with access to electricity service. Source: SIPERSN http://sipersn.udenar.edu.co:90/sipersn/.

From the 2,405 households reporting use of electricity, only $n = 1,397$ observations are included in the regressions, which correspond to 56% of the total households surveyed (2,479). This value corresponds to all those homes that facilitated the electric service bill for the registration of consumption. In Fig. 4 the histogram of the variable named *EE_Factura_CkWhMes*, which is the quantitative dependent variable that stores the observations of electricity consumption in rural houses, is shown.

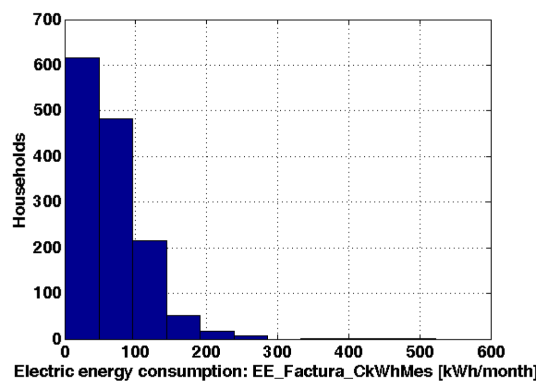

Fig. 4. Histogram of the dependent variable *EE_Factura_CkWhMes*. Source: SIPERSN http://sipersn.udenar.edu.co:90/sipersn/.

This histogram shows that 617 households recorded consumptions up to 48.5 kWh/month, followed by 482 households with consumptions in the range (48.5 to 95, 9] kWh/month, and then 216 homes with consumptions in the range (95.9 -. 143] kWh/month. In lesser amount, 51 households with consumptions in the range of (143-191] kWh/month, 18 households with consumptions in the range of (191 -. 238] kWh/month and 13 with values above the latter range. These last observations are considered as outliers in the regressions.

III. Best regressions

We have explored multiple linear regressions based on robust algorithms [4]-[6] and between the dependent variable and the explanatory variables. Non-linear relationships between these variables have not yet produced good results in the explanation of electricity consumption in Nariño's rural households. For each regression, we considered previous analysis presented in [2]-[3], such as a review of the dispersion and correlation between the variables, the analysis of distributions (histograms), and manual exclusion of outliers [1]. As a final result, five independent (non-nested) types of regressions were obtained.

A. Type 1

It is the best regression that tries to explain the consumption of electric energy in rural households through the greatest amount of explanatory variables available in the survey. In this regression, it was discarded households with higher consumption ($>$ 250 kWh/month) and that paid for the service more than 150,000 COP. This regression assumes that the electricity consumption is mainly explained by the linear contribution of: The frequency of use of incandescent light bulbs, called *Incandescente_Uso*; the frequency of use of compact fluorescent lamps (CFLs), *Ahorradora_Uso*; refrigerators size, *Refrigeracion_Tam*; the number of people living in each household, *Vivienda_Persona*; the number of rooms, *Vivienda_Cuarto*; monthly earnings in each of them, *Vivienda_Ingreso*; and the bill value of electricity service, *EE_Factura_Valor*.

B. Type 2

These are the best regressions explaining the electricity consumption of household when (i) they use only electricity and LPG, called Type 2.1; (ii) use only electricity and firewood, Type 2.2; and (iii) use electricity, LPG and firewood, but are categorized with a dummy variable that distinguishes two classes of households by energy source, Type 2.3. Due to there is no representative participation of electricity in the cooking process, it is not possible to make a regression that shows the incidence of demand for LPG and firewood in this process; however, it is possible to compare the energy consumption among households that use gas and households using firewood in an independent way.

In regression Type 2.1 it is assumed that the monthly electricity consumption is mainly explained by the linear contribution of: The refrigerator's size, *Refrigeracion_Tam*; the number of people living in each household, *Vivienda_Persona*; the number of rooms, *Vivienda_Cuarto*; and the bill value of the electricity service, *EE_Factura_Valor*. In the regression type 2.2, it is assumed that the monthly electricity consumption of these households is mainly explained by: The frequency of use of incandescent bulb lamps, *Incandescente_Uso*; the frequency of use of CFL, *Ahorradora_Uso*; refrigerator's size, *Refrigeracion_Tam*; the number of people living in each household, *Vivienda_Persona*; and the bill value of the electricity service, *EE_Factura_Valor*.

The regression Type 2.3 is categorical and facilitates the calculation of the difference in energy consumption between households that consume gas and those that consume firewood. To do this regression, rural households are categorized in two classes: Class A, households that consume electricity and firewood; and class B, households consuming electricity and gas. It is assumed that the monthly electricity consumption of the two class of households is mainly explained by: The frequency of use of CFL, *Ahorradora_Uso*, and the bill value of the electricity service, *EE_Factura_Valor*.

C. Type 3

It is the best regression which proposes a categorization of the model of demand for household's size and facilitates the calculation of the demand difference for electricity among households of the following classes: Class A, homes with more than 4 rooms; and class B, homes with 4 rooms or less. It is assumed that the monthly electricity consumption of the two classes of households is mainly explained by: The frequency of use of CFL, *Ahorradora_Uso*; refrigerator's size, *Refrigeracion_Tam*; the household's earnings, *Vivienda_Ingreso_COP*; and the bill value of the electricity service, *EE_Factura_Valor*.

D. Type 4

It is the best regression that explores the difference in electricity demand between households that are located in different earnings categories. The households are categorized in the following classes: Class A, households with earnings in the range [0, 350,000] COP; Class B, households with earnings in the range of (350,000, 700,000] COP; and class C, households with earnings of more than 700,000 COP. It is assumed that the monthly electricity consumption of the three classes of households is mainly explained by: the refrigerator's size, *Refrigeracion_Tam*, the number of rooms of each house, *Vivienda_Cuarto*, and the bill value of the electricity service, *EE_Factura_Valor*.

E. Type 5

Finally, the best regression that allows distinguishing the electricity consumption of households that belong to different social strata is obtained. The A class, households belonging to stratum 1; and class B, households belonging to stratum 2. No higher strata are reported in rural areas of Nariño. It is assumed that the monthly electricity consumption of the two classes of households is mainly explained by: the refrigerator's size, *Refrigeracion_Tam*; number of rooms of each house, *Vivienda_Cuarto*; and the bill value of the electricity service, *EE_Factura_Valor*.

IV. Selection of the best models

All the previous regressions were achieved maintaining the best possible indicators; e.g., coefficients with significance of 5%, explanation criteria, known as R^2 and adjusted R^2_a, greater than 70%, and root mean square error values, RMSE, less than 5%. Goodness of fit was also analyzed by Durbin-Watson test, Cook distance and presence of collinearity by means of the variance inflation factor (VIF). Given that the proposed models are independent, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) were used to choose the best models [1].

According to these tests, the best models, called M_1, M_2, and M_3, are: (i) The M_1 model, with higher quality of goodness of fit and acceptable level of information criterion, which corresponds to regression Type 4, which is the regression categorized by household's earnings. (ii) The M_2 model, of acceptable quality of goodness of fit and a higher level of information criterion, which corresponds to regressions Type 2.1 and Type 2.2, which are the regressions filtered for household groups consuming electricity and gas, and electricity and firewood, respectively. And (iii) the M_3 model, with lower quality of goodness of fit and minor information criteria, which corresponds to regression Type 1. The latest model is relevant because it includes the largest number of explanatory variables and can be set as a basis model for comparison with the other models.

A. Model M_1

This model describes the electricity demand for Class A, homes with earnings in the range [0, 350,000] COP; Class B, homes with earnings in the range (350,000, 700,000] COP; and C class, homes with earnings over 700,000 COP. The model's equations are:

$$\hat{y}_A = 4.88 + 1.38x_1 + 1.92x_2 + 0.003x_3 \tag{1a}$$
$$\hat{y}_B = 23.54 + 0.23x_1 + 1.92x_2 + 0.002x_3 \tag{1b}$$
$$\hat{y}_C = 18.65 + 1.38x_1 + 1.92x_2 + 0.003x_3 \quad , \tag{1c}$$

where $\hat{y}_A$ (kWh/month) is the estimated demand of households in class A, $\hat{y}_B$ (kWh/month) is the estimated demand for households in class B, and $\hat{y}_C$ (kWh/month) is the estimated demand for households in class C. x_1 (ft^3) is the refrigerator's size , x_2 (room/house) is the number of rooms that has each house, and x_3 (COP/month) is the bill value of the electricity service.

It is observed that: (i) The intercept of the model (kWh/month), mainly related to a minimum subsistence demand, is higher for class B. (ii) The contribution to the demand by refrigeration capacity (kWh/month/ft^3) is equal for Class A and C, and lower for Class B. (iii) The contribution to the demand for the number of rooms (kWh/month/room/house) is the same for all classes. (iv) The contribution to the demand for the amount paid for the service (kWh/COP) is the same for the B and C classes, and lower for class A. (v) Excluding the intercept, the demand for classes A and C is explained in the same way for each independent variable; however, the demand for class B differs from the previous ones in the contribution by the refrigerator's size and the value paid for the service (this contribution is less than that of the other classes). The behavior of these two variables is a determinant factor defining demand for electricity in rural households in Nariño.

Using average values for each variable contained in the model, it results that the energy demand for each rural household of Nariño is, respectively for the A, B and C classes, 78.5164 kWh/month, 70.5676 kWh/month, and 108.1462 kWh/month. The values of these demands are agreed with the categorization of households by earnings; that is, households with higher earnings have a higher electric power demand.

B. Model M_2

This model describes the electricity demand for Class A, hoouseholds that consume electricity and gas; and class B, households that consume electricity and wood. The model's equations are:

$$\hat{y}_A = 11.46 + 0.63x_3 + 1.32x_4 + 0.003x_5 + 1.15x_6 \tag{2a}$$
$$\hat{y}_B = 11.08 + 0.03x_1 + 0.07x_2 + 1.88x_3 + 1.58x_4 + 0.002x_5 \quad , \tag{2b}$$

where $\hat{y}_A$ (kWh/month) is the estimated demand for homes in class A, and $\hat{y}_B$ (kWh/month) is the estimated demand for homes in class B. x_1 (hours/month) is the frequency of use of incandescent bulb lamps, x_2 (hours/month) is the frequency of use of CFL, x_3 (ft^3) is the refrigerator's size, x_4 (person/house) is the number of people living in the house, x_5 (COP/month) is the bill value of the electricity service, and x_6 (room/house) is the number of rooms per house.

This model shows that: (i) The intercept (kWh/month), mainly related to a minimum subsistence demand, is virtually the same for both classes. (ii) The contribution to the demand for incandescent lamps and CFL (kW) is significant only for class B. (iii) The contribution to the demand for refrigeration equipment capacity (kWh/month/ft^3) is greater for class B. (iv) The contribution to the demand for the number of people living in the house (kWh/month/person/house) is slightly greater for class B. (v) The contribution to the demand for the amount paid for the service (kWh/COP) is slightly greater for class A. (vi) the contribution to the demand for the number of house's rooms (kWh/month/room/house) is only significant for class A. (vii) Both models are virtually independent and the main difference in the model of the power is given by the variables used in the process of lighting.

With the average values of each variable included in the model, it is concluded that the demand for households in class A, 83.2658 kWh/month, is greater than the demand for households in class B, 48.0349 kWh/month; that is, the demand for electricity in rural households using gas is greater than the demand of households using firewood.

C. Model M_3

This model describes the electricity demand for all rural households in Nariño and without any categorization. It is described by the following equation:

$$\hat{y} = 4.41 + 0.02x_1 + 0.05x_2 + 0.98x_3 + 1.01x_4 + 1.69x_5 + 0.000004x_6 + 0.003x_7 \quad , \tag{3}$$

where $\hat{y}$ (kWh/month) is the estimated demand, x_1 (hours/month) is the frequency of use of incandescent bulb lamps, x_2 (hours/month) is the frequency of use of CFL, x_3 (ft^3) is the refrigerator's size, x_4 (person/house) is the number of people living in the household, x_5 (room/house) is the number of rooms per house, x_6 (COP/month) is the earnings of each house, and x_7 (COP/month) is the price paid for the electricity service.

Taking the average values for each variable, this model states that the value of the demand for each rural household in Nariño is 65.6133 kWh/month; this value is lower than the reported by the M_1 model and is approximately the average of model M_2.

V. Conclusion

We have obtained a model for the electricity energy demand for rural households in Nariño. This model, which is derived from categorical and non-categorical regressions can be used for future projects of power generation, network interconnection plans, consumption control strategies, and/or energy substitution efforts, especially those related to the use of alternative energy sources.

Since the main source of energy in Nariño's rural area is the wood, the effects of deforestation of tropical and native forests can be avoided if electricity demand models are used together with scenarios, or substitution strategies, in typical power consumption processes. For example, the use of wood in cooking process can be optimized by using smart stoves, or the use of alternative energies such as solar, or wind, or pellets.

In future work, we will propose possible scenarios to control some of the variables that explain the electric demand in Nariño's rural households. In addition, such scenarios will consider alternative power supply sources, for example the supply given by wing generators.

Acknowledgment

The authors would like to thank Universidad de Nariño, Departamento de Electrónica, and PERS-Nariño. This work has been supported by the project “Analisis de oportunidades energticas con fuentes alternativas en el Departamento de Nariño” (ALTERNAR), BPIN 2013000100089, Acuerdo 005, OCAD-Fondo CTeI, SGR, Colombia.

References

[1] W. Achicanoy, *Modelo de demanda de energía eléctrica residencial rural para El Departamento de Nariño obtenido a través de la información de la encuesta PERS-Nariño 2013. PERS-Nariño*, San Juan de Pasto, Colombia. [Online]. http://sipersn.udenar.edu.co:90/sipersn/index.php?per=s&od=s

[2] C. Agostini and C. Plottier, and E. Saavedra. *La demanda residencial por energía eléctrica en Chile.* Unpublished. ILADES-Georgetown University Working Papers inv240, Ilades-Georgetown University, Universidad Alberto Hurtado/School of Economics and Bussines, 2009.

[3] L. Fernández, *La demanda residencial de electricidad en España: un análisis microeconométrico.* Ph.D. dissertation. Departament de Política Económica i Estructura Económica Mundial, Universitat de Barcelona, 2006.

[4] W. H. DuMouchel, and F. L. O'Brien. *Integrating a Robust Option into a Multiple Regression Computing Environment.* Computer Science and Statistics: Proceedings of the 21st Symposium on the Interface. Alexandria, VA: American Statistical Association, 1989.

[5] P. W. Holland, and R. E. Welsch. *Robust Regression Using Iteratively Reweighted Least-Squares.* Communications in Statistics: Theory and Methods, pp. 813827, 1977.

[6] P. J. Huber, *Robust Statistics*. Hoboken, NJ: John Wiley & Sons Inc., 1981.