IJEEP

INTERNATIONAL JOURNAL OF ENERGY ECONOMICS AND POLICY

EJ EconJournals

ISSN: 2146-4553

International Journal of Energy Economics and Policy

ISSN: 2146-4553

available at http: www.econjournals.com

International Journal of Energy Economics and Policy, 2018, 8(5), 202-211.

EconJournals

Determination of Models of Simple Regression and Multivariate Analysis for the Forecast of the Electricity Price in Colombia at 2030

Nelson Javier Hernández Bueno[1], María de los Ángeles Pinto Calderón[2], Yecid Alfonso Muñoz Maldonado[3], Adalberto Ospino Castro[4]*

[1]Department of Engineering, Universidad Autónoma de Bucaramanga, Colombia, [2]Department of Engineering, Universidad Autónoma de Bucaramanga, Colombia, [3]Department of Engineering, Universidad Autónoma de Bucaramanga, Colombia, [4]Department of Energy, Universidad de la Costa, Colombia. *Email: aospino8@cuc.edu.co

ABSTRACT

The electricity price in Colombia responds to demographic, economic, climatic changes, among others, that generate uncertainty and therefore risks in the electric production. Considering that the decision-making process has a great importance in the electricity market and that the participation of generators in energy auctions is usually based on intuition and previous experience, the need to study the possible alternatives and methods that minimize the risks before deciding some important matter can be appreciate. In this article, the estimation of the behavior of electrical energy prices in Colombia at the year 2030 for different scenarios and there are propose the following scientific models: (1) Simple regression; (2) econometric model. As result are obtained forecasts for each model, identifying that the econometric model has the lowest margin of error compared to the historical data that considers the behavior of different variables for the forecast.

Keywords: Econometric Modeling, Methods of Statistical Simulation, Forecasting, Electricity Price
JEL Classifications: C13, C15, Q47

1. INTRODUCTION

In Colombia, the electricity sector is known as one of the most dynamic sectors of the national economy, which shows a strong dependence of both petroleum and its derivatives, such as gas, coal and hydropower (El Heraldo, 2017). According to the *"Informe mensual de variables de generación y del mercado eléctrico colombiano"* (SIEL, 2016), the hydropower has the highest participation in the country's energy mix, with the 69.93% at December of 2016.

This sector is composed by entities and companies that fulfill various functions in the markets for generation, transmission, marketing and distribution of energy (Paez, et al. 2017). However, hydroelectric power plants are among the forms of generation that are more affected by natural phenomena that impact their normal development, due to the fact that the employee resource for their production is water and in times of drought the potential decreases as a function of flow rate or the reservoir (Goyeneche, 1995), while during the season of rainfall, reach their maximum level of production and let the excess water pass; these climatic variables mentioned above are some of the issues that have a direct impact on the production of electrical energy to the use of that resource. It is important to note that, in the case of those countries which are found in the equatorial zone, such as Colombia, there are two climatic phenomena that affect large proportions to this industry that are known as El Niño and La Niña phenomena (Dobias, 2015). These phenomena are opposite phases of the El Niño-Southern Oscillation (ENSO) cycle, where La Niña is referred to as the cold phase, therefore there are more rain than usual, and El Niño as the warm phase (NOAA, 2018).

Although climatic phenomena can not be predicted with total certainty, their economic consequences for the energy industries can be mitigated by employing a reliable tool for estimating the electricity price behavior in Colombia during the decision-making process. Both the forecasts are made based on the implementation of the statistical and econometric models and the analysis of the respective results for each of them, identifying which is the model with the smallest margin of error.

This research integrates the qualitative and quantitative approach, through deductive, inductive and sequential processes. From the qualitative approach, the various phenomena that affect the hydropower sector of Colombia and the prediction models to make forecasts of prices are explored from literature review in primary and secondary sources of information; from the quantitative approach, is carried out the data collection of the electricity prices, which are analyzed for structure two models: Statistical and econometric. On the other side, the research scope is of explanatory type due to, based on the work carried out, it is determined which model has best fit with the characteristics of the study variable from the forecast of electricity price in Colombia in 2030. The main result of the research is the structuring of a statistical tool, innovative and applicable to the Colombian electrical sector, replicable and adaptable to those economic sectors that are exposed to the risk of climate change market to industries for the variation of prices.

2. ELECTRICITY PRICE FORMATION

It is important to note that the Comisión de Regulación de Energía y Gas (CREG) is responsible of regulating the electricity market rates through resolutions. According to resolution No. 119 of 2007 of the (CREG, 2007), these rates are based on the unit cost (UC) of the energy supply's provision, expressed in Colombian pesos per kilowatt hour (COP/kWh), taking into account whether the consumption is industrial or residential. The unit cost is defined by equation (1):

$$UC=G+T+D+C+RL+R \quad (1)$$

UC: Unit Cost
G: Generation
T: Transmission
D: Distribution
C: Commercialization
RL: Recognized losses
R: Restrictions

In the case of residential users or regulated users, the rate depends by the socioeconomic status, for this reason, in some cases a subsidy is subtract, in others it is applied only the unit cost, or it is added a contribution or surcharge depending on the stratum, wich also is applied to industrial users regulated. Knowing the information mentioned, helps to identify that there is a variable cost that is impacted by different events, that is why it is appropriate to investigate what are those events and how causes fluctuations in the electricity price.

In order to have at least two models to compare among themselves and, subsequently, to analyze its percentage error respect to historical data to be used for the forecast of the electricity price in Colombia in 2030, it is decided to use stochastic or probabilistic and deterministic models. It should be noted that the independent variables that make up the forecast model are generally on an annual, therefore this is the periodicity selected for the study.

The econometric models are mathematical expressions that establish the relationship between one or more endogenous and exogenous variables that explain the systematic or deterministic behavior of the model, in addition to the random noise, that describe the non-probabilistic part (Kmenta and Ramsey, 2014). According to (Wooldridge, 2006), the equation (2) describes a multiple regression model with economic variables.

$$Y_i=\beta_0+\beta_i X_{li}+ \; +\beta_k X_{ki}+u \quad (2)$$

Being *Yi* the dependent variable, or variable explained, *X* refers to each of the explanatory variables or independent variables. B_0 is the term independent of the model that reflects the value of the dependent variable when all other variables are zero. The other B_i are the coefficients of the model and refer to the relationship between the independent variables and the dependent variable.

To obtain an econometric model, using the methodology of the authors, it goes through the following steps: (1) Identification: It must identify what is the objective of this study, the factors and variables that are of the same type, passing through a selection of the most appropriate variables for the correct determination of the requirements; (2) Estimate: In this stage are obtained the numerical values of the parameters based on the information in the sample of each one of the model's variables; (3) Validation: Focuses on making parallels between actual values for periods that have already occurred with the forecasts for those same periods, therefore, the model is tested and; (4) Implementation: Then to make the relevant adjustments, the model is implemented in real cases.

Within the different statistical methods, firstly it is chosen between the simple average, moving average, linear estimation, estimation logarithmic and single, double and triple exponential moving average. These are chosen to demonstrate the difference between using simple methods, as it is the moving average, and other more complex that apply to data with trend, variability and/or seasonality. In order to know the statistical method that presents better adjustment to the historical electricity price behavior in Colombia, are evaluated the mentioned below, taking as a reference to (Ramírez, 2013; Bowerman et al., 2007; Grimaldo et al., 2017). It should be noted that, as mentioned by (Bowerman et al., 2007), the classifications of forecasting methods depending on their time horizon are: (1) short-term: <3 months, (2) Medium-term: 3 months-2 years and (3) Long term: 2 years or more.

To set the time horizon or period in which the forecast models are going to be projected, it is taken as a reference (Wolfram Research, 2007), which defines that the ideal is to use a forecast horizon equal to or less than the period of the historical data; the justification for this lies in the fact that one of the properties of the auto-correlation is parity, which means that the correlation between the time series X_t and X_{t-k} is equal to the correlation between the time series X_t and X_{t+k}, therefore the resulting graph is symmetric respect to the forecast axis, as shown in Figure 1, and the forecasts obtained are endorsed by the parity of the auto-correlation.

In view of the above, it is evident that for projecting the same amount of historical data should be extended until the year 2032, however, it is decided to raise the forecast until the year 2030 to be up to the beginning of the 1990s and in the same way 14 years remains a period close to the historic.

After this, it is determined that a probabilistic model is applicable as result of the technical analysis of the electricity price behavior over time and not only for purposes of comparison, also it is evident that there is a tendency, which allows to infer that can project implementing a statistical method of forecast. However, to collect historical data from the electricity price, there are certain periods of greater variability which can be caused by the impact of different events that the statistical methods mentioned above do not consider, which leads to opt for an econometric model to review the incidence of the different variables in the price of power in Colombia.

Similarly, it should be noted that, for the econometric model, the amount of data collected for each variable is not consistent among themselves due to the time series does not agree on the size of the arrangements, it means that they do not account with the same information in term of years. Therefore, it is appropriate to review the interval at which the known information of all the variables, obtaining that the period from the year 2001 to the year 2016, a total of 16 years, therefore, this same period is used as historical data in statistical and econometric models.

When it is run each of these methods with historical data, it can be determined that the method with a lower margin of error is the moving average, the logarithmic regression, followed by the simple and double moving average, then triple exponential and finally the remaining. This is from the point of view of technical analysis, however, it is important to make a graphical analysis, Figures 2 and 3, to see the behavior that each of the models follows.

Respect to the Figures 2 and 3, it can be seen that the simple average smooths the data, causing the behavior of the series be dimmed; With regard to the logarithmic regression, had been found in the technical analysis with low margin of error, however, when analyzing the graph it can be seen that their behavior is very linear and not assimilated to that of the series of historical data. On the other hand, with the moving average, it is noted that the triple exponential moving average is the one that has a similar behavior to that of the historical data.

In view of the above, choose the Triple Exponential Moving Average as a statistical method that best mirror the behavior of the series analyzed, taking into account that adapts to values with trend and seasonality, important features in this case knowing that El Niño and La Niña phenomena occur approximately every 4 years and have an impact on the variables that influence the electricity price as it is the precipitation of the country.

3. ECONOMETRIC MODEL

The econometric model integrates a fundamental analysis in its structure, which allows to identify, analyze and select the variables with the greatest impact to the Colombian electrical sector. From this, it is obtained that the natural phenomena such as El Niño and La Niña are the main random disturbance that allows to reason in probabilistic terms and not exact the behavior of the electricity sector in Colombia.

This model is unicuacional, the dependent variable is the electricity price, and non-linear since they are presented fluctuations. Similarly, the observed data for each of the variables that are intended to include in the model have different observations over time of the variable, therefore, it is a time series that can handle a periodicity time zone for each of the days, months, quarters or even years, always and when carrying out appropriate transformations for the type of variable to be handled. Once the characteristics of the model have identified, it is set out the different econometric parameters to be considered.

Based on the analysis of the annual time series, the socio-political, geopolitical, climatic and economic events that occurred during periods of fluctuation in the price of electricity are identified. Table 1 identifies the impact factors for the electric subsector and each of the variables with the highest incidence (Hernández and Sotelo, 2016).

After a debugging from the theoretical and conceptual point of view, the qualitative variables are discarded, as they are: Guerrilla Attacks, Oil Security, Wars and Natural Phenomena, due to that structure a model with very small intervals and, therefore, identify variables that are under the socio-political or geopolitical factors

Figure 1: Auto-correlation graphics

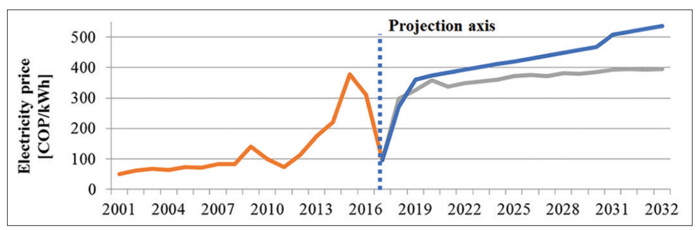


Figure 2: Behavior of the moving average and the logarithmic regression versus historical data

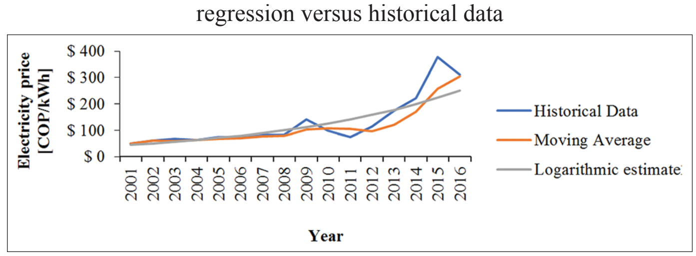
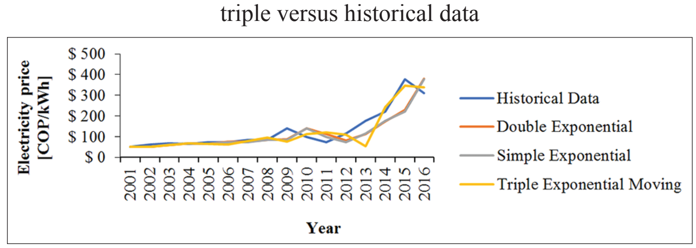


Figure 3: Exponential moving average behavior of single, double and triple versus historical data



would be complicated for a model so special, leading to handled quantitative variables. As a second instance, and as part of a theoretical-conceptual analysis, fundamental analysis is performed on each of the variables identified and then select which, from a conceptual point of view, has a greater impact on the electricity price.

Table 2 shows the periods of the independent variables. The range of study of the historical data of the variables is between the year 2001 and 2016.

Electricity price: Is the study variable or dependent variable. The time series of historical price of the annual electrical energy allows to observe the trend that has this variable over the years, it means, from 2001 until 2016, which is the defined range.

It is also evidence of a trend of upward growth with atypical behavior in the period between 2008 and 2011; it is called this way since it is a behavior that presents fluctuations of various recurrent and non-recurrent events, as is the weather, attacks on the infrastructure, the country's economic situation, movements in the exchange rate, population growth, among others.

The statistical analysis on the time series of the weighted average of the prices of the annual electricity and it can be concluded with 95% confidence for a sample of 16 years: The average annual temperature is 129.68 per kW/h, the annual standard deviation is 111.95 per kW/h, the maximum value is 377.55 per kW/h and the annual minimum value is 48.63 per kW/h.

Precipitation - Rainfall: Colombia has been characterized by being a country of great seasons of rains and droughts in different parts of the national territory or in some cases in all its geographical environment. In Figure 4 can be observed the weight of accumulated annual rainfall in the main cities of Colombia: Bogotá, Barranquilla, Cali and Medellín, in which is the annual rainfall accumulated in the city of Medellín is the region that weighs more in the accumulated precipitation of the years in Colombia and therefore its variation has a greater impact on the national energy market.

The Figure 5 shows the historical behavior of the cumulative annual rainfall of Medellín, which is selected as one of the major cities of Colombia, for being the one of the greatest impact on the cumulative annual rainfall of the country. In this series of time is observed that the data behave in a random way and is between 2007 and 2011 the period where is obtained a greater amount of rainfall in this sector of the country. The statistical analysis on the time series of annual precipitation accumulated of Medellín, it can be concluded with 95% confidence for a sample of 16 years: The annual average is 1770.04 mm annual, the standard deviation is 405.48 mm, the maximum annual value is 2554.20 mm and the annual minimum value is 1208.70 mm.

Representative Exchange Rate - TRM: The exchange rate is one of the variables with the greatest impact on the economy of the country and, therefore, each one of the economic activities that develop themselves. This has been a very useful variable to make innumerable events in the economic and financial field, however, to be a very sensitive, few dare to implement it. In the Figure 6 is show the historical behavior of the TRM, which presents a bearish trend. Additionally, it is present a high volatility, which can generate some noise at the time to describe the price of power in Colombia. It is possible to appreciate that it is a series of random time but not stationary, due to their statistical properties such as the mean and the variance do not remain constant over time.

The statistical analysis on the time series of the TRM annual it can be concluded with 95% confidence for a sample of 25 years: The annual average is 1927.73 COP, the annual standard deviation is 668.31 COP, the annual maximum value is 3050.97 COP and the annual minimum value is 737.98 COP.

Similarly, for the selected period which covers 16 years, it can be concluded with a 95% average annual rainfall is 2035.5 COP, the annual standard deviation is 308.98 COP, the annual maximum value is 3050.97 COP and the annual minimum value is 1768.23 COP.

To evaluate the other macroeconomic variables, the data is collected of the growths of these in Figure 7.

The gross domestic product (GDP): Statistical analysis on the time series of quarterly and annual GDP it can be concluded with

Table 1: Impact factors in the electricity

Factors	Causes
Socio-political	Guerrilla attacks
	Oil Security
Geopolitical	Wars
Demographic	Population
Climate change	Temperature
	Precipitation/Rainfall
	Natural Phenomena: *"El Niño" and "La Niña"*
Economic	Gross domestic product (GDP)
	Unemployment
	Foreign direct investment (FDI)
	Consumer price index (CPI)
	Producer price index (PPI)
	Representative Exchange Rate (TRM)

Figure 4: Weighted by major cities in the cumulative total annual precipitation of colombia

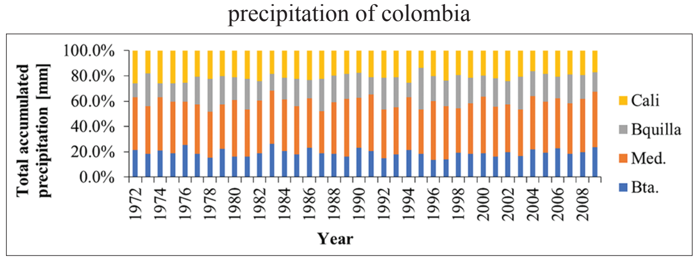
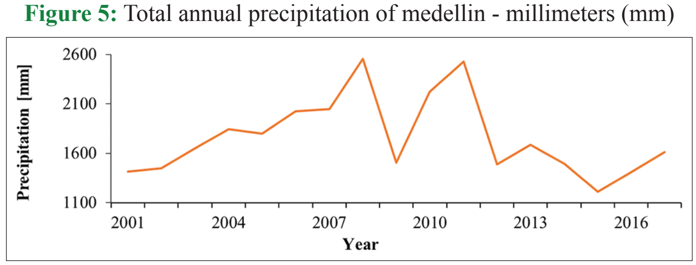


Figure 5: Total annual precipitation of medellin - millimeters (mm)



95% confidence for a sample of 22 years: The annual average is 3.39%, the annual standard deviation is 2.58%, the maximum annual value is 6.9% and the annual minimum value is −4.38%. Similarly, for the selected period which covers 16 years, it can be concluded with a 95% average annual 4.09%, the annual standard deviation is 1.69%, the maximum annual value is 6.9% and the annual minimum value is 1.70%.

The consumer price index (CPI): Statistical analysis on the time series of quarterly and annual CPI It can be concluded with 95% confidence for a sample of 25 years: The annual average is 9.57%, the annual standard deviation is 2.58%, the maximum annual value is 24.82%, and the minimum value is 2%. Similarly, for the selected period which covers 16 years, it can be concluded with a 95% average annual 4.9%, annual standard deviation is 1.9%, the maximum annual value is 7.39% and the minimum value is 2%.

The producer price index (PPI): Statistical analysis on the time series of quarterly and annual PPI It can be concluded with 95% confidence for a sample of 17 years: The annual average is 0.51%, the annual standard deviation is 0.705%, the maximum annual value is 2.16%, the minimum annual value is −0.80%. Similarly, for the selected period which covers 16 years, it can be concluded with a 95% average annual 0.51% annual, the standard deviation is 0.89%, the maximum annual value is 2.16% and the annual minimum value is −0.8%.

Population: The statistical analysis on the time series of quarterly and annual population it can be concluded with 95% confidence for a sample of 24 years: The annual average is 1.32%, the annual standard deviation is 0.22%, the maximum annual value is 1.93%, and the minimum annual value is 1.12%. Likewise, for the selected period which covers 16 years, it can be concluded with a 95% average annual 1.19%, the annual standard deviation is 0.04%, the maximum annual value is 1.28% and the annual minimum value is 1.12%.

Figure 6: Time series of the TRM

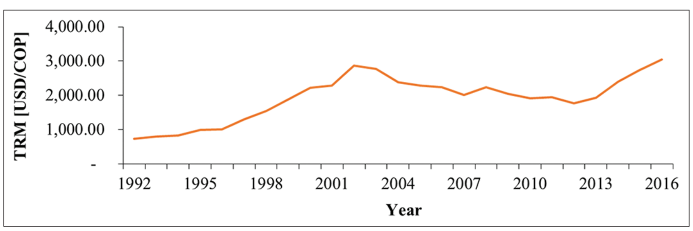
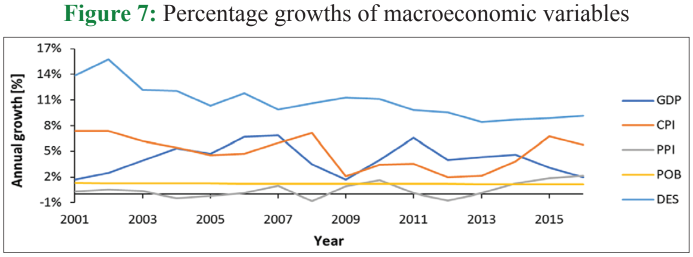


Figure 7: Percentage growths of macroeconomic variables



Unemployment - DES: Statistical analysis on the time series of quarterly and annual DES it can be concluded with 95% confidence for a sample of 16 years: The annual average is 10.85%, the annual standard deviation is 1.97%, the maximum annual value is 15.77%, and the minimum annual value is 8.44%. Similarly, for the selected period which covers 16 years, it can be concluded with a 95% average annual 10.85%, the annual standard deviation is 1.97%, the maximum annual value is 15.77% and the annual minimum value is 8.44%.

Correlation Coefficient: In the Table 3, the coefficient of correlation between the variables can be identified, where, according to the data, as well as the type of correlation between variables and the level of impact of this correlation, which are indicated in an abbreviated way.

To obtain the model that best describes the study variable and which at the same time comply with the tests of statistical rigor applied, 59 possible regression models are established from the combination of 2, 3, 4, 5, 6 and 7 independent variables.

The test of hypotheses about a population parameter only or also known t-test, which tests the hypothesis of the parameters of the functions of the regressions carried out, that according (Wooldridge, 2006) are described in equation 2.

Where it is assumed that meets the assumptions of the classical linear model - MLC, in which it should be noted that all the βj are parameters that reflect an unknown feature of the study population that cannot be known with certainty, so it can be speculated these values to then make statistical inferences and to test the hypothesis, based on estimates and modeling. That is why the t-test is applied, since it allows to demonstrate that the random variables are independent, it means that they have no effect on the behavior of the study variable, to what is known as the null hypothesis.

The null hypothesis H_0 describes that β_j measures the partial impact of x_j on and, therefore, when controlling the other independent variables (x1, x2, ..., xj-1, xj+1, ...),x_j has no effect on the expected value of y. Thus:

Table 2: Periods in which each variable contains records of information

Variables	Definition	Period	Source
EPr	Electricity Price	1996-2016	(Derivex, 2016; XM, 2017)
PRECIP	Precipitation	1972-2016	(CGM, 2016, 2017)
TRM	Representative exchange rate	1992-2016	(El País, 2016)
GDP	Gross domestic product	1995-2016	(DANE, 2017a)
CPI	Consumer price index	1992-2016	(DANE, 2017b)
PPI	Producer price index	2000-2016	(DANE, 2017c)
POB	Population	1993-2016	(DANE, 2017e)
DES	Unemployment	2001-2016	(DANE, 2017d)

Table 3: Correlation Coefficient and level of impact between selected variables

Variable	PRECIP (%)	GDP (%)	CPI (%)	PPI (%)	POB (%)	DES (%)	TRM (%)	EPr (%)
PRECIP	100.0							
GDP	59.9	100.0						
CPI	−6.4	−19.81	100.0					
PPI	−37.8	−28.29	8.2	100.0				
POB	2.6	−6.47	47.0	−44.7	100.0			
DES	−7.3	−27.9	49.3	−20.8	86.2	100.0		
TRM	−47.9	−41.8	67.8	41.6	12.3	28.2	100.0	
EPr	−48.1	−31.1	−5.7	69.6	−78.5	−62.3	35.3	100.0
Relationship according to the color	**High|x|>65%**	**Middle 65%>|x|>35%**	**Under 35%>|x|**					
Direct								
Indirect								

Table 4: T-theoretical (t_{theo}) according to the number of variables in the regression

No. variables	*DF*	t_{theo}
2	13	2.53264
3	12	2.56003
4	11	2.59309
5	10	2.63377
6	9	2.68501
7	8	2.75152

- If H_0 is not rejected: The independent variable is not related to the dependent variable (it is not significant)
- If H_0 is rejected: The variable is accepted since it is related to the dependent variable.

Similarly, the t-test statistic is used to perform the rejection test of the null hypothesis, which requires an alternative hypothesis H_1 and the level of significance α chosen for the test. For this study, a level of confidence of 95% is established and, therefore, the level of significance α is 5% (probability of rejecting H_0 when it is true; however, it is considered that the independent variables have a *ceteris paribus* effect in y without specifying, positive or negative, so the alternative hypothesis is bilateral or two-tailed, meaning that α would be distributed 50% at each end of the distribution, dividing the confidence level by 2, which in this case is 2.5%, and placing it at the 97.5 percentile, so it is expected that the value of t of each variable is large enough to reject H_0 in favor of H_1.

According to the mentioned, for the purposes of this regression, determine a critical value or t-calculated (t_{calc}), so that the rejection of H_0 takes place for a significance level of 2.5% for each end of all the random samples when H_0 is true and the degrees of freedom ($DF = n - k - 1$) by establishing a critical value according to the number of variables, which can be evidenced in the Table 4.

Based on the foregoing analyzes the statistical t of each variable in each model generated, with the purpose of determining:

$t_{calc} > t_{theo}$: Is rejected H_0 and H_1 is accepted, so that the variable is significant.
$t_{calc} < t_{theo}$: Do not reject H_0 and H_1 is not accepted, so that the variable is not significant.

Based on the analysis made, for a confidence level of 95% it can be observed that:

- The cumulative annual rainfall of Medellin is one of the independent variables with the greatest impact on the electricity price in the Colombian market. This variable in 35 of the 59 regressions performed rejects H_0 so that the cumulative annual rainfall of Medellín is significant, indicating that it relates to the electricity price from Colombia. The precipitation of a country, or in this case of a city that has great participation in the country, precipitation, has a direct impact on the availability of water resources for the generation of electric power, due to the fact that Colombia generates most of electricity from hydroelectric plants, therefore fluctuating rainfall is impacted resources, the supply of electrical energy and therefore the price of the same, which to comply fully with the statistical tests applied to the time t is selected as one of the independent variables impact the electricity price.
- The GDP in any of the regressions rejects H_0 so it does not accept the alternative hypothesis, indicating that it is the independent variable for a confidence level of 95% is not significant, it being concluded that is not related to the electricity price from Colombia. Therefore, this independent variable, since it does not meet the basic statistical criteria tested.
- The CPI in only four regressions rejects H_0 so it is an independent variable that does not have a high significance respect to the electricity price of Colombia, which is excluded from the group of the impact variables.
- The PPI in only four regressions rejects H_0 so it is an independent variable that does not have a high significance respect to the electricity price from Colombia.
- The population is the independent variable with the highest number of times rejects H_0, since in 49 of the 59 regressions rejects the null hypothesis and accept the alternative, so it is clear that this variable is related to a large sense with the electricity price from Colombia. The above reflects the particularity of the law of demand indirectly due to the greater number of people, in this technological age, the greater the need for electric power, which would imply that the energy market must adapt to meet the demand, implying a direct impact to the electricity price in Colombia.

- The DES reject H_0 in 15 of 59 cases, from which it can be deduce that there is a relationship, not as strong, between this variable and the electricity price from Colombia, therefore, is considered to review its impact and behavior, since it partially fulfilled the test performed, however, this independent variable can be ruled out.
- The TRM in just eight regressions rejects H_0 so it is an independent variable that does not have a high significance respect to the electricity price from Colombia. This can be denoted since the daily volatility of the TRM is reflected in the annual average which is the figure recorded in the time series of this variable, so including this variable in the model would have a variable too sensitive and would have a model little stable, which is why it rejects any regression that considers the TRM.

Therefore, it is determined that the independent variables at this stage are: Annual rainfall accumulated in Medellín; population and unemployment. However, these regressions are performed with variables with power one without any alteration. For this reason, to obtain the model that best describes the electricity price in Colombia, it is incorporated non-linear relationships between dependent and independent variables.

The main statistical characteristics of the 23 regressions carried out between the following variables: Dependent Variable: EPr and Ln(EPr). Independent variables: PRECIP, POB, DES, Ln(PRECIP), Ln(POB) and Ln(DES), where "Ln" is the natural logarithm.

The t-test is performed in the first 16 combinations that include the linear or non-linear form of the three independent variables and the independent variable, in which unemployment does not meet the criteria stipulated and does not reject the H_0, so additional regressions 8 are made with combinations of linear and non-linear form of the explanatory variables: Rainfall and population, as well as the variable explained EPr, which is that the variables predictive of all regressions are in compliance with the t-test, rejecting the H_0 and therefore accepting the alternative hypothesis, considering that the final variables to structure a model that best describes the behavior of the price of the electrical energy in Colombia are the annual cumulative precipitation of Medellín and the population.

After doing the non-linear regressions between the dependent and independent variables, the statistics of the selected model are obtained, as shown in Table 5. The results mean that the variation of the independent variables explains the variation in the electricity price in Colombia. In the same way the R^2 set indicates that the number of variables does not affect the level of explanation of the independent variable. The standard error of the estimate (Se) is low, so it can be seen the level of the standard deviation between the dependent variables and the forecasts made in the regression.

The Table 6 presents the result of the variance analysis, which allows us to identify whether or not there is a significant relationship between the variables. The F statistic allows to test the H_0 that the population value of R is zero, so the regression model is equivalent to contrast the hypothesis that the slope of the regression line is zero. The critical level indicates that if R is equal to 0, R has a null probability of taking the value of 0.95 and therefore the variables are not related to each other, on the contrary, it can be shown that for the case of study R is greater to 0, indicating that the variables do relate to each other.

Table 5: Statistics of the regression

Statistics of the regression	
Multiple correlation coefficient	95.85%
The coefficient of determination R^2	91.87%
R^2adjusted	90.62%
Standard Error	18.84%
Comments	16

Table 6: Analysis of variance

Parameter	Degrees of freedom	Sum of squares	Average of the square	F	Critical value F
Regression	2	5.21	2.60	73.47	0.00
Waste	13	0.46	0.035		
Total	15	5.67			

The resulting econometric model is represented in equation (3):

$$E\,\mathrm{Pr} = -6516.22 - 203.07 * \ln(PRECIP) - 1843.66 * \ln(POB)$$

A model of a functional level-log, since the dependent variable is EPr and independent variables Ln(PRECIP) and Ln(POB). This model implies that an increase of 1% in annual rainfall of Medellín is associated with the decrease in the price of the annual electrical energy in Colombia of 0.20 pesos and in turn, an increase of one percent of the population of Colombia decreases 18,436 pesos the annual electricity price in Colombia.

To apply the triple exponential moving average it is establish that approximately every 4 years there is seasonality, since El Niño and La Niña phenomena, and assigning that value to the coefficient of periods of seasonality (PS). With this, the results of the coefficients used in the model are: Trend alpha (α)= 0.9983, bias statistic (β)= 0.9706 and seasonality gamma (γ) = 57.9897.

4.ESTIMATION OF ELECTRICITY PRICE IN COLOMBIA FOR 2030

For the forecast of the electricity price in Colombia in 2017-2030, it is necessary to determine the values of the variables that compose the econometric model, these being the total annual precipitation and Population, in order to make the necessary estimates; however, only knows the value of each variable until the year 2017, so it is necessary to project the variables for the period 2018-2030. To do this, statistical methods are used, verifying that the range consistent with the appreciation of the DANE and not exceeding the limits that this entity considers, obtaining as a result that The method with a lower margin of error respect to the historical data is the logarithmic regression, therefore, is selected and the method by which the data is derived from the Table 7.

The Figure 8 presents the time series of the electricity price in Colombia: Historical data, Projected with the econometric model and projected with the statistical model. During the period between the year 2001 and 2008 the electricity price in Colombia presents a slight upward trend without abrupt changes in their behavior. For the year 2009, the study variable adds around 60 COP/kWh compared to the previous year, however between 2009 and 2011 decreased the price, positioning the same around 74 COP/kWh. Since the year 2012, and until the year 2014, there is an upward trend, increasing annually at a rate of significant change, however is 2014 to 2015 when an atypical data that increases from 1 year to another 156 COP/kWh, which is the highest observed variation in the historical data, which puts the electricity price in Colombia for the year 2015 around the 377.55 COP/kWh and decreasing until the 311.28 COP/kWh in 2016.

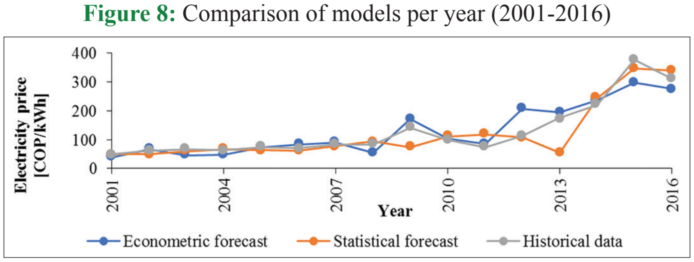
The Table 8 summarizes the resulting margin of error in the statistical and econometric model respect to the historical data that served as a basis for determining the models, it means, the time series that covers from the year 2001 up to 2016, finding as a result that the econometric model has less margin of error that the statistician, since it includes impact variables that describe more accurately the electricity price for forecast.

With real data from 2001 until 2016 to make estimates of the statistical model and structuring the econometric model, it is appropriate to apply the models to project the electricity price in the Colombian market from 2017 until 2030, which can be seen in Table 9. However, this result shows the effect of atypical variations in precipitation, as is the case of the year 2015 and 2016 where prices were above 300 COP/kWh since the period of drought that was presented in the colombian territory due to the presence of El Niño phenomenon, therefore it is evidence of a country's vulnerability in the face of extreme changes in the resource's booking of the main source of electrical power generation.

The electricity price registered by (Derivex, 2016) in 2017 is compared with the results of the models in that year to know the margin of error obtained in the data immediately presented, indicating that this is the first and only comparable real data with the forecasts made to date.

The Table 10 shows the margin of error of the projected values of the electricity price in Colombia by the econometric and statistical models for the year 2017. The actual price given for that year is around 97 COP/kWh, the econometric estimation for that same year is around 272.58 COP/kWh leaving a margin of error of 180.87%, while the price forecast for the year 2017 thrown by the statistical model selected is of 65.12% generating an error margin of 32.9%. It is important to note that the year 2015-2016 was a period of drought, while in 2017 there was a strong rainy season, so the price reacted and declined to levels below the 100 COP/kWh.

Figure 8: Comparison of models per year (2001-2016)



In the econometric model is obtained a lower price respect to the forecast of the previous year, however, did not conform exactly to the year 2017 due to the logarithmic model integrates the impact of both the population, that reflects the impact of the demand respect to the electrical energy, and precipitation, which measures the supply potential for the above. For this reason, is that by

Table 7: Forecast of population growth and total annual precipitation medellin 2018-2023

Year	Total annual precipitation forecast medellin (mm)	The population growth forecast (%)
2018	1616.57	1.10
2019	1605.14	1.08
2020	1593.79	1.07
2021	1582.52	1.08
2022	1571.33	1.07
2023	1560.22	1.07
2024	1549.18	1.07
2025	1538.23	1.06
2026	1527.35	1.06
2027	1516.55	1.06
2028	1505.83	1.06
2029	1495.18	1.06
2030	1484.61	1.06

Table 8: Percentage Error of the statistical and econometric model respect to the 2001-2016 time series

Parameter	Econometric	Statistical
Error respect to the historical data (%)	19.42	20.26

Table 9: Forecast of the electricity price in Colombia 2017-2030

Year	Econometric forecast model (COP/kWh)	Statistical forecast model (COP/kWh)
2017	272.58	65.12
2018	296.77	270.87
2019	326.54	361.19
2020	358.02	374.43
2021	338.07	383.67
2022	348.07	392.91
2023	354.67	402.16
2024	359.56	411.40
2025	371.38	420.64
2026	376.29	429.88
2027	372.53	439.13
2028	380.91	448.37
2029	378.88	457.61
2030	385.53	466.85

Table 10: The relative error between actual and projected for the year 2017

Parameter	Real price	Econometric	Statistical
Price (COP/kWh)	97.05	272.58	65.12
Relative error (%)	-	180.87	32.90

increasing the precipitation increases the reason of that variable, while the growth of the population decreases due to the model understands that increases the capacity of producing energy, and thus there is an increase in the product or service to offer, causing a state of supply and, therefore, that the price decrease at the rate of the beta model; on the other hand is that despite the decrease of the population and, as a consequence, demand, their variation is minimal so there is no impact on the projected price. In the case of the statistical model, the forecast of the price is significantly reduced compared to the previous year, closer to the real price of 2017 that the econometric analysis.

5.CONCLUSIONS

Within the national panorama, is was evidenced that the colombian energy sector has increased its participation in the Gross Domestic Product (GDP), reflecting the great opportunities that the country has in this sector. Additionally, it was demonstrated the direct dependence of the energy mix with the country's water resources, since the electricity generation is given mostly by the hydraulic generation. In the same way, it was demonstrated that regardless of the behavior of the sectors, the population is increasing slowly, it means that despite the increase in the number of people, the population growth decreases its speed which could lead to result in Colombia a future population aging.

By the econometric model, it was possible to note that the projected values of the electricity price have less margin of error that the statistical method previously selected, reflecting that it could be obtained a model with a greater accuracy by considering more variables for the forecast; however, it is necessary to consider the values of each of the independent variables, which makes the study more extensive since many times there is no relevant data and, as the model requires, it is necessary to look for the forecasts of these variables made by entities such as the DANE and the Bank of the Republic (5-year forecasts of the current period) or Implement statistical methods to perform the forecasts of these variables and feed the model.

The fact that the margin of error for both models was approximately 20%, generates a high inherent risk due to the level of uncertainty and therefore is subjective, the reader and the purpose of the study, accept it as an appropriate model. However, it could be considered that the margin of error of the models would have been lower if the electricity price between 2014 and 2015 had not increased suddenly, presenting the greatest variation observed in the historical data and causing the models were not so successful with these outliers.

It was determined that, among the analyzed variables, the precipitation is the most influencial on the electricity price thanks to the good correlation presented. This makes sense due to the electricity price is one of the most vulnerable to exogenous events such as the climatic variations, including La Niña phenomenon, in which the electricity price is lower, given that hydropower plants can reach its maximum storage capacity and therefore their participation increase over the thermoelectric plants; on the other hand is El Niño phenomenon where the opposite happens, demonstrating that there is an inverse relationship between the price and the level of the reservoirs in a country dependent on hydropower as Colombia, in addition to showing the need of diversify the country's energy mix with alternative sources more stable in the face of such variations.

It is very important to stress that for the approach of the model it was not considered the entry of alternative sources of energy, which would reduce the dependence on hydroelectrical energy and, as mentioned above, the price of power would be less impacted by precipitation.

6. ACKNOWLEDGMENTS

Thanks to ECOPETROL S.A for the finantial support to this study, to the *Grupo de Investigación en Recursos, Energía, Sostenibilidad – GIRES*, as well as the researcher Iván Ordóñez for providing valuable information for the development of a statistical model.

REFERENCES

Bowerman, B., O'Connell, R., Koehler, A. (2007), Pronósticos, Series de Tiempo y Regresión. 4th ed. México: Internacional Thomson editores.

CGM. (2016), Indicadores Ambientales Medellín. Contraloría General de Medellín. Available from: http://www.cgm.gov.co/cgm/Paginaweb/IP/InformesAmbientales/indicadoresambientalesmedellin2016.pdf.

CGM. (2017). Indicadores Ambientales Medellín. Available from: http://www.cgm.gov.co/cgm/Paginaweb/IP/Informes Ambientales/indicadoresambientalesmedellin2017.pdf.

CREG. (2007). Resolución No. 119 de 2007. Available from: http://www.apolo.creg.gov.co/Publicac.nsf/1aed427ff782911965256751001e9e55/c63f06a9114e1a150525785a007a6fa2.

DANE. (2017a), Cuentas Económicas Nacionales Trimestrales, Producto Interno Bruto-PIB. from: http://www.dane.gov.co/index.php/estadisticas-por-tema/cuentas-nacionales/cuentas-nacionales-trimestrales. [Last retrieved on 2018 Jan 27].

DANE. (2017b), Índice de Precios al Consumidor -IPC- Base 2008. Available from: http://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc.

DANE. (2017c), Índice de Precios del Productor (IPP). Available from: http://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-del-productor-ipp. [Last retrieved on 2018 Jan 27].

DANE. (2017d). Mercado Laboral. Available from: https://www.dane.gov.co/files/investigaciones/boletines/ech/ech/pres_web_empleo_rueda_prensa_dic_17.pdf.

DANE. (2017e), Reloj de Población. Available from: http://www.dane.gov.co/reloj/.

Derivex. (2016), Informes Mensuales del Precio Energía Eléctrica. Available from: http://www.derivex.com.co/SaladePrensa/Informes Mensuales/Forms/AllItems.aspx. [Last retrieved on 2017 Jul 18].

Dobias, J. (2015), El Fenómeno de El Niño y su Impacto en la Generación Hidroeléctrica de Chixoy. CNEE. Available from: http://www.cnee.gob.gt/pdf/informacion/ElNinoImpactoChixoy082015.pdf.

El Heraldo. (2017), Energía y Gas, un Sector Que Crece al Servicio de la Economía Nacional. Available from: https://www.elheraldo.co/mas-negocios/energia-y-gas-un-sector-que-crece-al-servicio-de-la-economia-nacional-324334.

El País. (2016), Dólar Alto Impacta las Importaciones en Colombia. Available from: https://www.elpais.com.co/economia/dolar-alto-impacta-las-importaciones-en-colombia.html. [Last retrieved on 2018 Aug 06].

Goyeneche, M. (1995), Hidroeléctricas, Problemas o Solución. Available from: http://www.eltiempo.com/archivo/documento/MAM-416567.

Grimaldo, J.W., Mendoza, M.A., Reyes, W.P. (2017), Forecast electricity demand model using predicted values of sectorial gross domestic product: Case of Colombia. Revista Espacios, 38(22), 3.

Hernández, B.N.J., Sotelo, S.O.A. (2016), Estructuración de un Derivado Climático Como Instrumento de Cobertura Para el Subsector Energético de Mayor Vulnerabilidad en Colombia. Bucaramanga: Universidad Autónoma de Bucaramanga.

Kmenta, J., Ramsey, J.B. (2014), Evaluation of Econometric Models. 1st ed. New York: Academic Press.

Paez, A., Muñoz, Y., Ospino, A., Hernandez, N., Conde, E., Pacheco, L., Gonzalez, W., Sotelo, O. (2017), Future scenarios and trends of energy demand in Colombia using long-range energy alternative planning. International Journal of Energy Economics and Policy, 7(5), 178-190.

NOAA. (2018), What are El Niño and La Niña? Available from: https://www.oceanservice.noaa.gov/facts/ninonina.html.

Ramírez, A. (2013), Métodos Utilizados Para el Pronóstico de Demanda de Energía Eléctrica en Sistemas de Distribución. Universidad Tecnológica de Pereira. Available from: http://www.repositorio.utp.edu.co/dspace/bitstream/handle/11059/3862/62131A719.pdf?sequence=1&isAllowed=y.

SIEL. (2016), Informe Mensual de Variables de Generación y del Mercado Eléctrico Colombiano. Available from: http://www.siel.gov.co/portals/0/generacion/2016/Segui_variables_dic_2016.pdf.

Wolfram Reserch. (2007), Mathematica Time Series. A Fully Integrated Environment for Time-Dependent data analysis. 1st ed. New York: Springer.

Wooldridge, J.M. (2006), Introducción a la Econometría: Un Enfoque Moderno. 2nd ed. España: Editorial Paraninfo.

XM. (2017), Precio en Bolsa Nacional. Available from: http://www.informacioninteligente10.xm.com.co/transacciones/Paginas/HistoricoTransacciones.aspx?RootFolder=%2Ftransacciones%2FHistoricoTransacciones%2FPrecios&FolderCTID=0x0120000E28F31AB3520A4C8D8CB6881FFA0AFE&View=%7B75B6225D-9047-44EC-8EC5-DEA7F7AA139C%7D.