Bidding Prices Modeling in Colombian Electricity Market using Logistic Regression Model

Diana Jimenez, *Student Member, IEEE*, Luis Gallego, *Member, IEEE*, Martha Camargo, *Student Member*
Research Program on Acquisition and Analysis of Signals – PAAS, National University of Colombia

ABSTRACT-- **The main goal of this work is to show the application of an econometric model called Logistic Regression (LR) in a study of the bidding prices in the Colombian electricity market. The LR model is used to calculate the statistical significance of some variables of the market and the influence of changes in these variables on the estimated bidding price of Generation Companies GENCOs. The methodology is used to fit the bidding prices behavior of a hydraulic agent considering some exogenous variables such as: hydrological conditions, energy demand, spot prices and competitors behavior. The model obtained with LR shows a good agreement with the real data.**

***Index Terms*-- Electricity Market, Logistic Regression Model, Bidding Prices.**

I. INTRODUCTION

In Colombia, the deregulation process started in 1991 with the new constitution and later with the laws 142 and 143 in 1994. It was an important change to the regulatory of the electricity sector in Colombia. This was followed by the creation of the legal framework for the electricity Pool, enabling deregulation to take place and the market to operate from 1995 [1]. The deregulation was a process under the principle of increase efficiency in the provision of services, as well as ensuring the operational and financial viability of business.

One of the consequences of this structure has been the decentralization of decision making processes. In this sense, the participant companies in electricity markets are forced to make their decisions under several sources of uncertainty. Some of the most important uncertainty variables of the market are market prices, demand, hydrological conditions, etc. Among these variables, the market price is the most important variable because of its obvious impact on the electrical business [2][3].

The main objective of this work is to make a characterization of bidding prices to enable a better understanding of how different variables affect the formation of bidding price for each GENCO. Specifically, we will examine the idea that price evolves according to the value of system variables and therefore to the different states of the market.

In this manner, an analysis with logistic regression models could give us information to estimate price levels as well as to better understand the market operation. The logistic regression is a discrete choice model to solve the problem of explaining the behavior of a dependent variable as the probability of occurrence of an event, through a set of explanatory variables [4].

That is, the main objective of this approach is to model the influence of some system variables in the probability of occurrence of an event [5], in this case, the influence of the above mentioned variables (hydrological conditions, the demand of system, spot price and the competitors price) in an event that may be defined as the probability of a GENCO having a high, medium or low bidding price.

II. OPERATION OF ELECTRICITY MARKET

In the Colombian electricity market there are two mechanisms of energy transactions. One of this is the bilateral contracts, which is a financial agreement where is determined the delivery period, quantity and price of energy. The other one are the transactions in the electricity spot market, where the energy price is determined by a competition mechanism [6]-[7].

The spot market is a commercial figure, in which each participant must report a single bidding price (in $/ MWh) and a declaration of availability for each one of its generating units (in MW) for the 24 hours of the next day, before 8:00 a.m. The Market Operator -the Dispatch National Center (CND)- receives the offers and availability and provides the generation program taking into account the expected demand, so that for each hour using the resources of the lowest price.

Fig. 1 shows this process for a specific hour taking into account both, the available resources and the ordered bidding prices. The initial resources (following a bottom-up ordering) correspond to those needed to match the restrictions of the system. The resources are allocated depending on the price (from lowest to highest) to match the demand for the hour, then the CND announced to GENCOs the generation to the next day.

D. Jimenez is with the National University of Colombia - research group PAAS-UN (e-mail: dyjimeneza@unal.edu.co).
L. Gallego is professor of the National University of Colombia - and active researcher of group PAAS-UN (email: lgallegov@unal.edu.co).
M. Camargo is with the National University of Colombia - research group PAAS-UN (e-mail: mpcamargom@unal.edu.co).

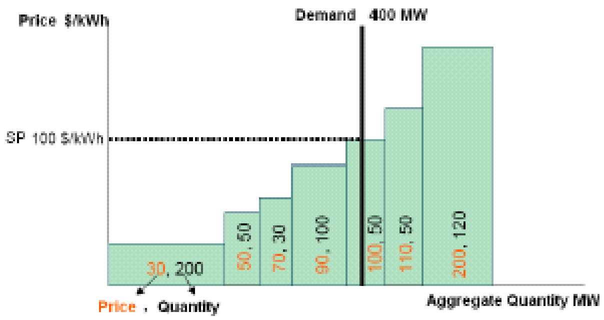

Fig. 1. Spot Price Process

This process determines the Spot price as the bidding price of the marginal resource to supply the demand [8].

III. Market's Variables

The variables used to model the bidding price of the GENCOs have a continuous domain, so it is necessary to make a discretization for each variable. The categorization was performed by means of a clustering method called kmeans that distributes n observations into k clusters in which each observation belongs to the cluster with the nearest mean [9]. The variables that were considered in this research are grouped into 3 groups: High (H), Medium (M) and Low (L).

The independent variables were taken into account in the model are: maximum demand daily, Reservoir to be offered, lagged Competitors Price (t-1), lagged Spot Price (t-1). On the other hand, the dependent variable is daily Bidding Price. The variables are described below:

Energy Demand (MD): Maximum daily demand in GWh.

Reservoir Offering (RO): Is the difference daily between reservoir level *R* and *MOS* "Minimum Operating Level", divided into his capacity total of reservoir *CR*.

$$RO(\%) = \frac{R - MOS}{CR} \quad (1)$$

Competitors Price (CP): Weighted average of daily bidding price *P* of the n competitor weighted by their installed capacity *C*. The CP is give in ($/kWh).

$$CP = \frac{\sum_{i=1}^{n} P_i \cdot C_i}{\sum_{i=1}^{n} C_i} \quad (2)$$

where i= 1,2,..n an

Spot Price (SP): Weighted average of hourly Spot Price *SPh* weighted by hourly demand of the system *D*. The units of SP are (S/kWh).

$$SP = \frac{\sum_{i=1}^{24} SPh_i \cdot D_i}{\sum_{i=1}^{24} D_i} \quad (3)$$

Bidding Price (BP): the daily Bidding Price is the weighted average of hourly bidding price *Pu* of each one of the k units of the GENCO, weighted by the capacity of each one of his resources *Cu*. The *BP* is give in $/kWh.

$$BP = \frac{\sum_{r=1}^{k} Pu_r \cdot Cu_r}{\sum_{r=1}^{k} Cu_r} \quad (4)$$

IV. Ordinal Logistic Regression Model

Logistic regression is a statistical tool for multivariate analysis. This method is suitable for situations where the dependent variable is categorical and is required to estimate the probability π that an individual belongs to a group. Besides, the regression analysis also identifies the most important independent variables that explain the difference between the groups [10][11].

Because the normal linear regression model assumes that the dependent variable *Y* increases linearly with the independent variables ***X***, is not a suitable model because it is necessary that the response of the model is between 0 and 1. Therefore is needed a probabilistic model, which is always some number between 0 and 1.

Given that the use of a distribution function ensures that the result of the estimate is between 0 and 1, an alternative is the logistic distribution function, which relates the probability $\pi(z)$ with the explanatory variables **X** as shown in Fig. 2.

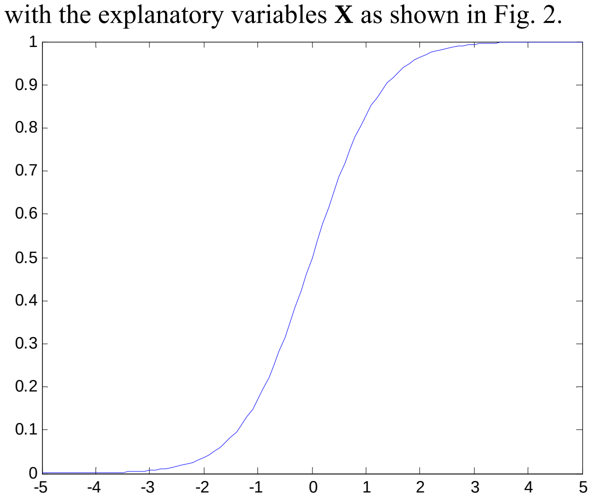

Fig. 2. Logistic function

$$\pi_i(Z_i) = \frac{e^{Z_i}}{1 + e^{Z_i}} \quad (5)$$

where $Z_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$

The equation (5) is nonlinear in the parameters, but using a mathematical transformation can be linearized as follows:

If π_i is the probability of occurrence of an event, then $1-\pi_i$ is the probability that the event does not occur. The ratio odds of occurrence of the event can write as:

$$\frac{\pi_i(Z_i)}{1-\pi_i(Z_i)} = \frac{1+e^{Z_i}}{1+e^{-Z_i}} = e^{Z_i} \quad (6)$$

If is calculated the natural logarithm of (6) gives the logarithm of the odds ratio called "logit" and has a linear model parameters

$$L_i = \ln\left(\frac{\pi_i(Z_i)}{1-\pi_i(Z_i)}\right) = Z_i \qquad (7)$$

where $Z_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki}$

When the independent variable to model has two or more categories and they are ordered is estimated the probability of each category, the specification model for Ordinal logistic model is:

$$\pi_i(Y_i = 0) = \Lambda(-\beta' X_i)$$
$$\pi_i(Y_i = 1) = \Lambda(\mu_1 - \beta' X_i) - \Lambda(-\beta' X_i)$$
$$\pi_i(Y_i = 2) = \Lambda(\mu_2 - \beta' X_i) - \Lambda(\mu_1 - \beta' X_i)$$
$$\pi_i(Y_i = J-1) = 1 - \Lambda(\mu_{j-2} - \beta' X_i) \qquad (8)$$

where $\mu_1, \mu_1, \ldots, \mu_{J-2}$ are the thresholds that divide the answer alternatives of the dependent variable.

$$\Lambda(-\beta' X_i) = \frac{e^{\beta' X}}{1+e^{\beta' X}}$$

J are the categories of the dependent variable

The proportional odds assumption is then that the β's are independent of j (note that the β's have no subscripts). In other words, it assumes that if we looked at (binary) logistic regressions of category 1 vs. 2, category 2 vs. 3, and so on, then the intercepts in the equations might vary, but the parameters would be identical for each category.

To Estimate the Logistic model was used the method of maximum likelihood. This method determines the most parsimonious and best fit describing the relationship between the dependent variable and explanatory a set of variables. The estimate of the parameters model by means of this method requires the use of iterative methods or algorithms that allow optimization and convergence of the values estimates, since the system of equations is obtained by maximizing the likelihood function is not linear.

Once the model is estimated is realized the validation of the individual statistical significance the explanatory variables incorporated in the model and measure the overall goodness fit. For this is used in the econometric literature the contrast ratio of likelihood, the Hosmer-Lemeshow test and measuring the percentage of successes [10].

In this work the results of the logistic model is the estimate of the probability that a GENCO offers high, medium or low, given the conditions of the generator itself and the market environment, and identify which of these variables are statistically significant at the formation of the bidding price of the GENCO in question.

V. Case Study

To show the application of the logistic regression model explained above, it was selected a hydraulic GENCO of the Colombian Market owing the major number of hydraulic plants of the system. The time period considered for this study is between 2001 and 2004.

As a first step it was necessary to eliminate undesired outliers in both, dependent and independent variables. In this case the three median absolute deviations (MAD) procedure was used.

Then, the independent variables were discretized and joined in a vector E that took the values of the categorical variables, so that E = [MD, RO, CP, SP] is a vector that represents a state of the GENCO. Thus, given that each variable is classified into 3 categories (high, medium and low) the number of possible states is given by $3^4 = 81$ for the market in the studied period. These states are representations of the perception of GENCO about the environment of the market [12].

Because of the discretization of the market variables was achieved taking into account the behavior of historical data, each state has a different probability of occurrence, therefore, any GENCO has a probability distribution of states that occurs in the time period.

Fig. 3 shows the probability of each state of the GENCO under study, and it can be observed that considering the 81 possible states, only 65 states were presented during the considered time period. In addition, TABLE I shows the most likely states to occur.

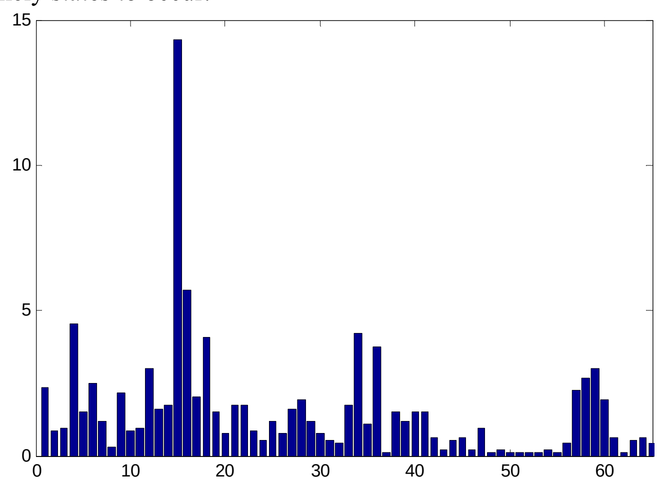

Fig. 3. Probability of the states of GENCO

TABLE I
STATES MOST POSIBLES TO GENCO

E = [DM, RO, CP, SP]	Probability (%)
E_{16} = [M, L, H, H]	14,3
E_{17} = [H, L, H, H]	6,35
E_{6} = [M, M, M, M]	3,55
E_{33} = [L, M, M, M]	3,34

On the other hand, Fig. 4 shows the behavior of the bidding price categorized by means of kmeans algorithm and table II shows the percentage of data found in each category.

TABLE II
CLASIFICATION TABLE

	N	% marginal
High	262	28,23
Medium	378	40,73
Low	288	31,03
Total Data	**928**	**100**

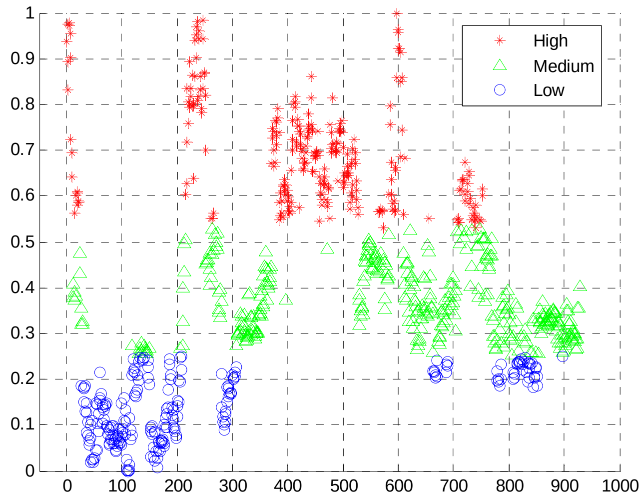

Fig. 4. Bidding price of GENCO categorized using k-means algorithm

In this case was necessary to make a transformation of the categories H, M and L to dummy variables. A dummy variable is a numerical variable used in regression analysis to represent subgroups of the variable of study. In this research a dummy variable was designed to distinguish different groups of each variable. TABLE III shows the applied transformation.

TABLE III
DUMMY VARIABLE

	V_1	V_2
Low	1	0
Medium	0	1
High	0	0

Applying the logistic regression model to categorized data, statistically reliable results were obtained after carrying out several tests from the data sample: A summary of these results is shown in TABLE IV.

TABLE IV
COEFICIENTS OF REGRESSION LOGISTIC MODEL – INITIAL MODEL

Var.	beta	se	Wald	dfe	Sig	95% Confidence Interval	
C1	1,23	0,39	10,25	1	0,00	-1,99	-0,48
C2	1,98	0,38	27,12	1	0,00	1,24	2,73
MD1	-0,09	0,20	0,21	1	0,64	-0,48	0,30
MD2	-0,15	0,18	0,65	1	0,42	-0,50	0,21
RO1	-3,12	0,36	77,16	1	0,00	-2,42	-3,82
RO2	-1,27	0,32	15,45	1	0,00	-0,64	-1,91
CP1	0,71	0,21	11,59	1	0,00	1,11	0,30
CP2	-0,07	0,17	8,55	1	0,06	-0,27	-0,42
SP1	2,70	0,27	100,78	1	0,00	3,23	2,17
SP2	2,18	0,19	130,67	1	0,00	2,55	1,81

From the regression results it was concluded that the demand variable was not statistically significant. In contrast, remaining explanatory variables were significant. Therefore, a new logistic regression excluding the variable demand were achieved which results are shown in TABLE V.

TABLE V
COEFICIENTS OF REGRESSION LOGISTIC MODEL –FINAL MODEL

Var.	beta	se	Wald	dfe	Sig.	99% Confidence Interval	
C1	1,13	0,36	9,67	1	0,00	1,85	0,42
C2	2,08	0,36	33,41	1	0,00	1,38	2,79
RO1	-3,12	0,35	77,54	1	0,00	-2,43	-3,82
RO2	-1,28	0,32	15,75	1	0,00	-0,65	-1,92
CP1	0,70	0,20	11,58	1	0,00	1,10	0,30
CP2	-0,08	0,17	2,08	1	0,05	-0,26	-0,41
SP1	2,70	0,27	101,79	1	0,00	3,23	2,18
SP2	2,18	0,19	133,69	1	0,00	2,55	1,81

The remain variables were statistically significant with such a confidence level to reject the null hypothesis that the parameter is equal to 0, over 99% in all cases. The sign of the parameter estimate, which measures the relationship between the variable and the probability of having a high, medium and low price, has been coherent for all the incorporated variables.

Finally, the equation that quantifies the probability of a low, medium or high bidding price was modeled with the following expressions:

$$\pi(BP=L)=\frac{1}{1+e^{-[1.13-3.12RO_1-1.28RO_2+0,7CP_1-0.08CP_2+2,7SP_1+2.18SP_2]}} \quad (9)$$

$$\pi(BP=M)=\pi(BP=L)-\frac{1}{1+e^{-[2.08-3.12RO_1-1.28RO_2+0,7CP_1-0.08CP_2+2,7SP_1+2.18SP_2]}}$$

$$\pi(BP=H)=1-\frac{1}{1+e^{-[2.08-3.12RO_1-1.28RO_2+0,7CP_1-0.08CP_2+2,7SP_1+2.18SP_2]}}$$

Furthermore, Wald statistics represents the weight of the independent variables over the dependent ones. In this case the most important variable in predicting the bidding price was the spot price with a weight greater than 100, followed by the reservoir to be offered and bidding price of competitors, while the demand was not a significant contribution to explain bidding prices.

Observing the signs of the estimated parameters it may be concluded that the greater the spot prices, the greater the probability of an increase in bidding prices, which indicates that GENCO follows the spot price of the market.

Besides, when the competitors price increase from a low level, the probability of increasing GENCO bidding prices also increases but in contrast, when competitors prices increase from a medium level, the probability of increasing GENCO bidding prices smoothly decreases (factor of 0.08). Regarding reservoir to be offered, the GENCO increases the probability of having higher prices when its reservoir level decreases.

In general, the goodness of fit is considered adequate. The value obtained in the calculation of statistical Mc-Fadden R2 was 0,57 %, which is acceptable in this type of models. On the other hand, the results obtained with the Pearson and Deviance test allows to accept the null hypothesis of a good fit of the model. In terms of the percentage of success, the percentage of correct predicted prices was about 66.5% over the sample data. The results of the goodness fit tests are shown in TABLE VI.

TABLE VI
GOODNESS FIT TEST RESULTS

	Chi-square	dfe	Sig.
Pearson	99,763	40	,000
Deviance	100,050	40	,000

TABLE VII
TABLE OF PREDICTED

	High	Medium	Low	% Correct
High	162	97	3	61,83
Medium	61	240	77	63,49
Low	7	65	216	75,00
% global	24,78	43,32	31,90	**66,51**

From the above table it is concluded that the bidding price for the GENCO under study may be accurately predicted for the low, medium and high level in 75%, 62% and 63% respectively.

Finally, figures 5 to 7 shows a good fit between the responses of regression logistic model and the real frequency of data for each category of the GENCO bidding prices.

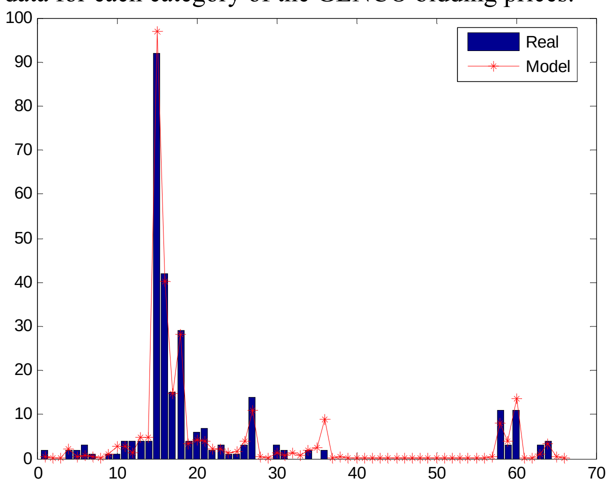

Fig. 5. Responses of the High bidding price

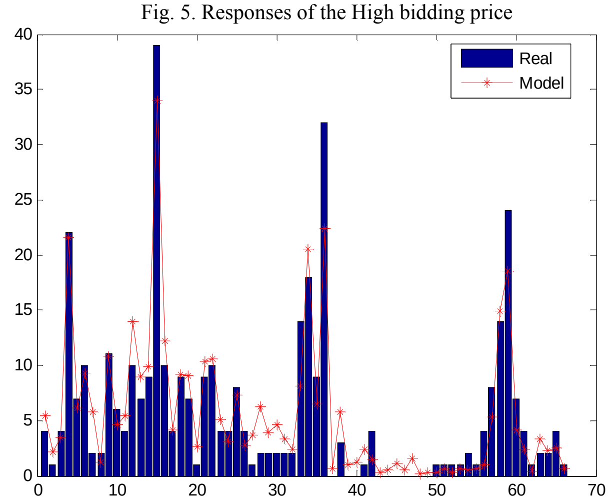

Fig. 6. Responses of the Medium bidding price

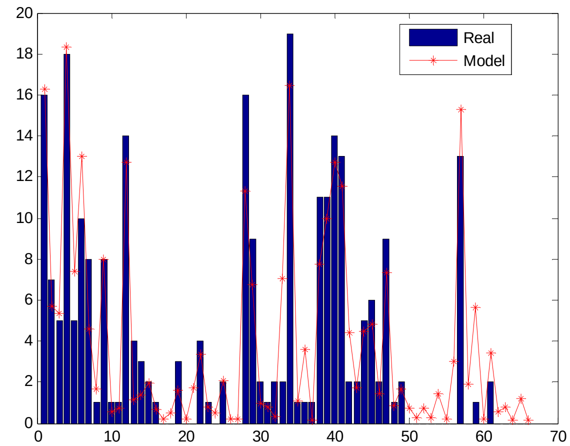

Fig. 7. Responses of the Low bidding price

VI. Conclusions

The results of the estimates, in terms of both goodness of fit and percentage of correct predicted prices are satisfactory. With the logistic regression was possible to model adequately the behavior of the GENCO with the explanatory variables considered.

The regression model presented in this paper may be used as a tool to reduce uncertainty in the context of electricity markets. Furthermore, the adopted methodology allows to include more explanatory variables in order to improve the accuracy of the model.

This paper presented a new modeling tool in the area of competitive electricity markets to estimate bidding prices under several scenarios or states of the market, based on a small set of explanatory variables.

This model can be applied to the prediction of not only bidding prices, but also different market events, such as probabilities to be dispatched given its bidding price.

VII. Acknowledgment

The authors would like to thank to the research group PAAS UN of National University of Colombia for the supporting of this work, the Graduate School of Electrical Engineering as well as XM SA, the Colombian Independent Operator, for providing the historical market data.

VIII. References

[1] Erik R. Larsena, Isaac Dyner. "Lessons from deregulation in Colombia: successes, failures and the way ahead". *Energy Policy* pag. 1767–1780 Vol. 32 2004.

[2] Unidad de Planeación Minero Energética –UPME- "Una visión del mercado eléctrico colombiano" República de Colombia Bogotá, July 2004.

[3] García, A, Arbeláez, L. E. Market power analysis for the Colombian electricity market. Energy Economics. pp. 217-229 Vol 24. No. 3. 2002.

[4] Alfred DeMaris, "A tutorial in Logistic Regression". Working *Paper Series Economic Development No 81*. University of Santiago de Compostela.

[5] Departamento de Estadística- Facultad de Ciencias Valladolid España. " Tratamiento Estadístico de Datos –Modelos de Regresión Logística-". 2006/1007.

[6] Garcia A., Arbelaez L.,"Market power analysis for the Colombian electricity market" Energy Policy pag 217-229 Vol 24 2002.

[7] Gallego L. "Moldeamiento del comportamiento de la oferta de energía eléctrica en el mercado colombiano". PhD Thesis. National University of Colombia. 2008

[8] A. Delgadillo, L. Gallego., D. Jimenez, M. Camargo, "Agent learning methodology for generators in an electricity market" *Power And Energy Society General Meeting - Conversion And Delivery Of Electrical Energy* In The 21st Century, 2008 Ieee ISSN: 1932-5517 ed: v.1 fasc.1 p.1 - ,2008.

[9] L. Gallego., O. Duarte. "Modeling of bidding prices using soft Computing Techniques". IEEE T&D Latin America 2008.

[10] David W. Hosmer, Stanley Lemeshow, *Applied Logistic Regression*, vol 2. Wiley Series in Probability and Statistics, p. 31.

[11] Menard S., *Applied Logistic Regression Analysis*, vol 2. Sage University Paper. P 63.

[12] L. Gallego., O. Duarte., A. Delgadillo. "Strategic bidding in Colombian Electricity market using a multi-agent learning approach" 17th Conference of the Electric Power Supply IndustryCEPSI 2008. Macau-China. October 2008.

IX. Biographies

Diana Jiménez Eng. Diana Jiménez was born in Bogota, Colombia, in 1981. She finished her undergraduate studies in the Department of Electrical Engineering of the National University of Colombia. She actually is candidate to obtain a Master in Electrical Engineering in the same institution. Eng. Jiménez has been research assistant in the Research Program of Acquisition and Analysis of Electromagnetic Signals of the National University of Colombia - PAAS -UN Group since 2003. Her research interests are electricity markets, econometrics model and power quality analysis.

Luis Gallego PhD. MSc Eng. Luis Eduardo Gallego Vega was born in Bogota, Colombia, in 1976. He finished his undergraduate, Master and PhD studies in the Department of Electrical Engineering of the National University of Colombia. Eng. Gallego has been researcher in the Research Program of Acquisition and Analysis of Electromagnetic Signals of the National University of Colombia - PAAS -UN since 2000, working in research projects mainly related to power quality, power markets and lightning protection. Eng. Gallego has been enrolled in teaching activities related to power quality and power markets as professor of National University of Colombia. His research interests are power quality analysis, power markets and computational intelligence applied to power systems modeling.

Martha Camargo Eng. Martha Camargo was born in Bogota, Colombia, in 1981. She finished her undergraduate studies in the Department of Electrical Engineering of the National University of Colombia. She actually is candidate to obtain a Master in Electrical Engineering in the same institution. Eng. Camargo has been research assistant in the Research Program of Acquisition and Analysis of Electromagnetic Signals of the National University of Colombia - PAAS -UN Group since 2003. Her research interests are electricity markets and soft computing techniques.