**A methodology for electricity option pricing. Case of study: Colombia**

# **Introduction**

It is known to us that the electricity has brought the ability of enhancing the production and availability of consumption options throughout the world (United Nations Development Programme, 2019), as well as it has allowed our society to connect via globalization (Arango, Arango, & Londoño, 2019). As a result of the importance of electric energy, its producers and consumers have been linked by markets that allow the interaction both between supply and demand in order to trade quantities of this commodity for a determined price, as it happens with many other commodities. Nonetheless, these electricity markets have suffered important changes in the last 30 years approximately as some of them have been deregulated, which allowed the negotiation of electricity through spot prices and, later, through future contracts and other derivatives.

Due to the importance that the electric energy has had throughout history, our society has become so dependent on it that the massive power outage that happened in the United States of America in 1965, even though it occurred due to the blind trust that people had in big interconnected electricity networks, and the electricity crisis that happened in the year 2000 in California hinted the importance of electricity availability in our society and of its market regulation. Due to all of the above, there has been multiple attempts in order to assure the supply of electric power about pricing electricity spot prices, future contracts and other derivatives.

# **Theoretical Framework**

## **Stochastic Processes**

A stochastic process is a variable that evolves over time in a way that is at least in part random (Dixit & Pindyck, 1994). These stochastic processes can vary in continuous time, as in the variable changes (or could change) its value at every possible moment time, or in discrete time, as in the variable could change its value only in some specific moments of time (ex. Every minute, every hour, every day). One of the simplest stochastic process, that will be shown here to illustrate the idea of what a stochastic process represents, is the discrete-state random walk, which can be expressed as following:

As we can see, in this case the variable x is a variable that starts with a known value and in every moment of time , it takes a jump from the precious value that is equal to , which is a stochastic variable that takes the value of 1 or -1, each with a probability of ½. It is worth noting that every jump that the variable suffers is independent of the previous jump, therefore is completely independent of and that is why can be written as the way we did.

### Ornstein-Uhlenbeck

A particular stochastic process that is going to be used in this work is the process known as Ornstein-Uhlenbeck. This stochastic process is a mean reversion process, which means that, in the long run, the stochastic variable is going to tend towards a mean. Mean reverting processes are often used to forecast spot prices of commodities such as copper or oil because, while in the short run this prices might fluctuate randomly up and down (due to fluctuation between demand and supply), in the longer run these prices should be drawn back towards the marginal cost of production for the commodity. If a variable follows a mean reverting process, its differential can be expressed as follows:

Here represents the speed of reversion and is the mean level of , which is the value to which tends to revert to. Therefore, the change of is going to depend of and . If is greater (smaller) than , is more likely to fall (rise) until it has reached its “normal” level. Also, note that if has a greater value, is less likely to drift away from its mean level as it will revert faster towards it.

If you wish to learn more about the Ornstein-Uhlenbeck process or about stochastic processes in general, please consult chapter 3 of (Dixit & Pindyck, 1994)

## **Options**

Financial options are financial assets that allows the owner to buy (Call Option) or sell (Put Option) an underlying asset for a certain price that was established previously (strike price). These options may have a variant in number of times that they can be exercised, the moment in which they can be exercised and/or the conditions that have to be met in order to make the Option exercisable; however, the most popular financial options that are traded in the market are American Options and European Options. American Options allow the owner to exercise their Option once at any time before its expiration date, while the European Options allow the owner to exercise it only during the day of its expiration. In order to access the option to buy or sell a certain amount of an asset at a strike price instead of at a spot price, the Option holder has to pay an Option premium to whoever is “selling” this Option.

If one owns a Call Option, one will exercise the option to buy the underlying asset at the strike price if it is lower than the spot price, otherwise, one will not exercise the option and buy the asset at the spot price. According to what was said above, the profit earned by owning a Call Option with a strike price of k is given by:

And the payoff profile for a Call Option with a strike price of k and an Option premium of y, the moment it can be exercised, is as follows:

Likewise, if one owns a Put Option, one will exercise the option to sell the underlying asset at the strike price if it is higher than the spot price, otherwise, one will not exercise the option and sell the asset at the spot price. Therefore, the profit earned by owning a Put Option with a strike price of k is:

And the payoff profile for a Put Option with a strike price of k and an Option premium of y, the moment it can be exercised, is as follows:

One of the easiest ways for pricing Options is through an estimation of the stochastic process that is followed by the prices of the underlying asset and, then, using Monte Carlo simulations to create multiple forecasts of paths that these prices might follow and one uses the profit function of said Option in order to calculate the average present value of this profit throughout the simulated price paths. Using this approach, there are two ways of pricing the options for a certain underlying asset: using real price growth and using risk neutral price growth; if neutral risk growth is used, one assumes that the mean of growth of the prices for the underlying asset is going to be equal to a risk free rate because there would be opportunity for arbitrage if one knew that the mean growth is going to be higher or lower than this risk free rate, on the other hand, if real price growth is used, one does not take into account the previous assumption and makes a new assumption that the prices, during the simulation time horizon, are going to have a mean growth that is equal to the mean growth that they have had historically.

# Electricity Production in Colombia

## **Electricity Production Technologies in Colombia**

Colombia is a Hispanic country that is located in the northern edge of South America, just above the line of the equator. Being in the tropical region, Colombia does not have traditional subtropical calendar seasons (e.g spring, summer, etc.) but it suffers tropical seasons of drought and wetness. As other tropical countries in South America, Colombia has a vast quantity of hydric resources which, among other things, has helped the country to produce big amounts of electric energy via hydroelectric plants. Of the 66,667 GWh of electric energy produced in Colombia in 2015, 80.3% of it was produced hydraulically (XM, 2017). Moreover, in 2017 the technologies for electricity production that were invested in for expansion were dominantly hydraulic (XM, 2017), as it is shown in *Figure 1*

Figure 1: Change in Colombian National Interconnected System by technology in 2017 according to expansion projects

Even though hydraulically generated electricity is relatively cheap due to it only needing an initial investment and a maintenance cost, the quantity of energy produced this way is deeply impacted by external events that rise or a decrease of available flowing water. This profound dependence of hydric resources in energy production in Colombia can be seen in *Figure 2* and *Figure 3* which shows the participation in the Colombian electricity matrix by technology when there is a wet period and a drought period, respectively (XM, 2017). What was said above makes Colombia highly vulnerable against the El Niño/Southern Oscillation, which is a major climatic phenomenon that impact the availability of water in Colombia.

Figure 2: Participation in the Colombian electric matrix by technology in a wet period

Figure 3: Participation in the Colombian electric matrix by technology in a drought period

## **El Niño Southern Oscillation (ENSO)**

The El Niño/Southern Oscillation is a natural phenomenon that involves a fluctuation in the temperature in the central and eastern equatorial Pacific, which is coupled with changes in the atmosphere. El Niño and La Niña are the oceanic components of this phenomenon and the Southern Oscillation is its atmospheric counterpart. This changes in ocean temperature and the atmosphere have a heavy influence over the climatic patterns around the world, creating heavy rains and floods in certain parts of the world and droughts in others.

The El Niño/Southern Oscillation contains three different phases: El Niño, neutral and La Niña. El Niño events often begin in the middle of the year with warming of the surface water in the central and eastern equatorial Pacific Ocean. Generally, El-Niño events peak around the November-January period and tends to decay over the first half of the following year. This phase of the phenomenon tends to occur every two to seven years and can last up to 18 months. As for Colombia, this phases bring abnormal deficits in rain level, which is shown by the over 50% of the Colombian territory being affected by a shortage of rainfall in 10 of the 12 El Niño events that happened form 1972 to 2010 (González & Hurtado)

La Niña, which is the opposite end of the ENSO cycle is an event in which there is a cooling of the surface temperatures of the central and eastern equatorial Pacific Ocean which comes with a reversal of the overall atmospheric and climatic conditions and variations compared to the El Niño. Therefore, La Niña events are known in Colombia for provoking a rise in the rainfall rate, which is shown by the 6 out of 8 La Niña events that happened between 1972 and 2010 which brought an excess of rainfall (González & Hurtado).Lastly, during the neutral phase of the ENSO cycle atmospheric patterns are controlled by other climate drivers in the world.

## Reliability Charge in Colombia

In order to hedge the risk that there is in electricity production in Colombia, this country’s government adopted in 1996 what they called a “Capacity Charge” method. This measure rewarded electricity generators as they increased the installed capacity of their plants. However, due to the high dependence of the National Interconnected System in hydroelectricity, an increase in installed capacity did not necessarily mean an increase in electricity generated (which also depends in climatic conditions and availability of water).

Due to this problem in the Capacity Charge measure, the Electricity and Gas Regulatory Commission (CREG), since December of 2006, opted towards creating to what is known today as the “Reliability Charge” to prevent possible electricity rationing periods during drought seasons. This new measure was also made to incentivize the investment on electricity production assets, but it was done in another way. Instead of rewarding an increase in installed capacity, the Reliability Charge measure assigns to different electricity generators, by an auction process, a firm energy obligation under scarcity conditions (OEF) which “commits the generating company to have available pre-specified quantities of energy in periods of scarcity, defined as periods in which the spot price exceeds the Scarcity Price.” (Harbord & Pagnozzi, 2008) This methodology is basically a Call Option with electricity as underlying asset which allows the government to buy a pre-determined amount of electric energy at a set price (the Scarcity Price, which is the Strike Price for this Option) instead of buying at Spot Price. This derivative is limited by the maximum amount of energy that the generator guarantees to generate under any climatic conditions and it must be supported by generation assets. As any other Option, the Colombian government pays an Option premium (Reliability Charge) to a generator for the option of buying energy at the Scarcity Price when the spot price surpasses it.

As the objective of this Reliability Charge is to be able to buy electricity at a Scarcity Price every time, during a time horizon of 10 years ahead, that its spot price is higher, this electricity derivative can be seen as a portfolio of an infinite number of European Options which maturation dates vary from now until the next ten years. Therefore, the payoff profile for each of this Options that make up the Reliability Charge portfolio is shown in *Figure 4.*

Figure 4: Payoff profile of each option with maturity date t. T: Time horizon for the Reliability Charge obligation. y:Option premium/ Reliability Charge. k: Strike price/Scarcity Price

Finally, the Option premium/Reliability Charge would have a value equal to the sum of all of these “infinite” European options that make up the portfolio. Therefore the price of this option would be the following:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |
|  |  |  |

It is worth noting that the actual Reliability Charge proposed the Scarcity Price to be equal to the spot price of electricity that is generated by the thermoelectric plant with the highest spot price.

# **Pricing Methodology Description**

In this paper, the prices are going to be estimated by 6 different ways in order to compare different ways of estimation and in order to choose the best model that can be used to simulate the prices of electricity in the Colombian energy stock exchange. The following steps are going to be developed in order to obtain the pricing models and to be able to do the final pricing of the “Reliability charge” for electricity in Colombia:

* Data collection and analysis: this steps consists in the explanation of what data was used, where it was obtained and an analysis of the main patterns present in this data. This is done with the purpose of identifying the main aspects that have to be included into the price forecasting models. Furthermore, in this step a division of the data into a training set and a validation set is also explained.
* Model Specification and Estimation: In this step six different price forecasting models and the methods of their estimation will be explained.
* Model Comparison: The models will be compared using the RMSE error measure.
* Monte Carlo Simulation: This step consists in the definition of the model and the parameters that are going to be used for simulation of future prices.
* Option pricing: Lastly, the electricity option pricing is going to be made by the simulation approach.

# **Data collection and analysis**

## ENSO Data

According to what was said about the El Niño and La Niña events, the ENSO is an extremely important phenomenon to be taken into account in order to predict electricity price in Colombia due to the impact that these ENSO cycle phases have on the availability of water in the country and the high reliance of Colombia in hydroelectric central in order to produce electric energy. Therefore, to take the ENSO phenomenon into account, the Oceanic Niño Index from the National Oceanic and Atmospheric Administration (NOAA) from the United States and probabilities for El Niño, La Niña and neutral ENSO events were obtained from the International Research Institute for Climate and Society (IRI) from the Earth Institute of the Columbia University and the Climate Prediction Center (CPC).

The Oceanic Niño Index is an index that measures the differences in sea surface temperature along the Niño 3.4 (A region along the Pacific Ocean that is shown in *Figure 5*). The probabilities for the ENSO are predictions that are made nine months ahead and show the estimate probability of there being La Niña, a neutral or El Niño event in a certain period of time (these probabilities always sum 1 for each estimation period). The probabilities are given by the CPC and IRI for overlapped three month periods (e.g., there is a January-February-March period and there also is a February-March-May period); however, for the sake of making things easier, the probability estimation for a three month period is going to be supposed as it is the estimation for the month in the middle, for example, if the estimation for January-February-March is 60% La Niña, 40% neutral and 0% El Niño this is going to be taken as following: there is going to be a 60% probability that there is a La Niña event in February and there is going to be a 40% chance that there is a neutral event in February, with there being no chance for an El Niño event happening.

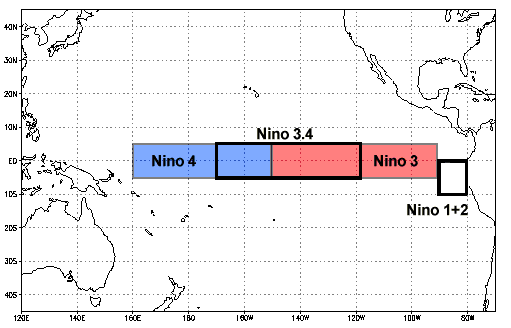


Figure 5 El Niño regions. Taken from the National Oceanic and Atmospheric Administration. Avalibale at: https://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst/

Even though the IRI and the CPC have been reporting this ENSO probabilities forecast since 2002, in this work only the forecasts made since 2012 were taken into account. This selection of the data was made because the data from 2002 to 2012 was published in a way that was very inconvenient for it to be merged into a database. It is worth noting that the IRI, since they adopted the new publishing format in 2012, publishes two different sets of probabilities for each three-month period they forecast; a set of probabilities that are obtained only by taking into account only a Predictions Plume of dynamic and stochastic models that forecast the Niño 3.4 Sea Surface Temperature anomalies and the official set of probabilities forecast that they publish, which is based on a consensus of CPC and IRI forecasters taking into account the predictive information of the ENSO Predictions Plume, observational information and human judgement. Both of these sets of probabilities will be used to create forecasting models for the electricity price in Colombia.

## Electricity Prices Data

For the estimation of the different models, the prices of electric energy, per kWh, were obtained from XM’s (the Colombian electricity market manager) database. This database contains the hourly prices of electricity 24 hours a day, seven days a week, in COP available from July 20, 1995 to April 2, 2019. However, according to what was said before, only the prices from 2012 to are going to be used in this work. This data was sub-setted into a training set consisting of the spot price from January of 2012 and December of 2016 and a validation set consisting of prices from January to December of 2017.

As it was determined that these prices were stationary (Cabrales & Uriza, 2016), the most dominant frequencies had to be determined. For this purpose, a method that was implemented in R by Robin J. Hyndaman (Professor of Statistics at Monash University) was used. This method determines the dominant frequency from a spectral analysis from a time series. First, a linear trend is removed and then the spectral density function is estimated from the best fitting autoregressive model (based on the AIC). If there is a large maximum in the spectral density at a frequency f, then the dominant period will be 1/f (rounded to the nearest integer). If there is no dominant frequency, this method will return 1 as the “dominant period”. (Package 'forecast', 2019) For obtaining the dominant frequencies, this method was used for different time series (that all go from January, 2012 to December, 2017) in order to evaluate if there are different seasons that affect the movement of the electricity prices in Colombia, the different time series that were used to determine these dominant frequencies were:

* The hourly prices’ time series (in order to determine a dominant period in hours).
* The daily average prices’ time series (in order to determine a dominant period in days).
* The monthly average prices’ time series (in order to determine a dominant period in months).
* The yearly average prices’ time series (in order to determine a dominant period in years).

The results of these estimations were the following:

As it can be seen **in**, the dominant period that resulted from the hourly prices analysis was that there is a strong intra-day behavior in which electricity prices rise early in the morning due to people waking up and using electricity to power up their home appliances, which provokes an increase in electric energy’s demand, then prices go down steadily and then they increase at night due to the need of using electricity to power domestic light bulbs, street lights and home appliances that are turned on by people which have returned to their homes. Second, we have a non-intuitive dominant period of 11 days. Third, there is another non-intuitive dominant. Lastly, we can see that there is no dominant yearly period because of its result being 1

# **Model Specification and estimation**

According to what was proposed **by** , the price behavior is going to be described by two components. One component that comprises all the predictable dynamics of the prices’ time series, such as cyclical behavior (that involves the dominant frequencies that were found in the **Data collection and analysis**), a deterministic upwards trend and the probabilities of the ENSO phases. The other component is purely stochastic and it was assumed that it follows a particular continuous diffusion process.

## **Price Modeling (Mean Reversion)**

The stochastic process followed by the electricity spot prices are expressed as the sum of the deterministic component, which is a function that depends on the time and a forecast for the set of probabilities P for each of the phases of the ENSO cycle for that specific time, and the stochastic component . Therefore, the spot prices are going to be modelled as:

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Where:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Where and stands for an increment to a standard Brownian motion. Therefore, follows a mean reversion process, also known as Ornstein-Uhlenbeck, which was explained before, with a long run mean of zero and a reversion speed. Therefore, the strike prices tend to a mean function in the long run.

In the same way, the natural logarithm of the spot prices are going to be modelled as:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Where is another deterministic function that depends on the time and a set of ENSO probabilities forecast and follows a Ornstein-Uhlenbeck process of the form:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Similarly, is going to tend to a mean function in the long run. In both cases, and are going to converge faster to the mean as takes a higher value.

### Deterministic Component

It is necessary that the deterministic component of the price models and capture the most relevant predictable behavior of the electricity prices. According to the analysis that was made in section 4.2, the inclusion of cyclical behaviors and an upwards trend to these deterministic components seems to be appropriate. Furthermore, according to what has been said about the effects of the ENSO cycle in the supply of electricity in Colombia, taking this climatic phenomenon into account in the deterministic component of the prices’ modelling seems adequate as well. Therefore, different deterministic functions are proposed.

First we have a function that depend only on time:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

Where:

Based on the analysis made in section *4.2*, a lineal trend is added with the term , this would reflect the growth of electricity prices over the years (Figure). Also, having identified a difference in prices between a weekday and a holiday/weekend (Figure), a dummy variable models if a certain day corresponds to a holiday/weekend or not. Finally, some Fourier series were added in order to capture the seasonal patterns that prices follow:

* The term represents an intra-year cycle that the prices follow (Figure).
* represents the seasonal behavior that has a X month period
* The term captures the intra-day seasonal behavior that electricity prices follow (Figure).
* The term captures a seasonal behavior that has an 11 day period.
* The term

It is worth noting that these four terms are Fourier series that have their respective frequency and that have an order of 1.

Then, there is a function that depends on time and on the set of ENSO probabilities. However, as there are nine different sets of ENSO probabilities forecast for each month (except for), nine price forecasting models are proposed. As the time-dependent part of the model equation is the same as the previous model, a reference to the function is going to be used for expression simplicity. Furthermore, as these nine models are basically the same, just that the coefficients are estimated using different sets of probabilities, these models are expressed in equation (7). However, it is vital to keep in mind that nine different models are expressed there.

|  |  |  |
| --- | --- | --- |
|  |  | (7) |
|  |  |  |

It is worth noting that for these probabilities every date was reduced to its month and year (e.g., the date “5:00:00 of January 23rd, 2012” would be taken just as “January,2012”). Therefore, and would take the same value for every date that would belong to the same month and year. Likewise, a forecast published in November, 2012 (no matter the calendar day) is taken in this work as a forecast done one month before for the probabilities of El Niño and La Niña for all the dates that belong to December, 2012. These probabilities are added to the model in order to capture the effect that the ENSO cycle has on the electricity prices (Figure). However, the models through are kept as independent from to be able to compare the effect of taking the ENSO probabilities into account in the models.

Now, the deterministic functions used to model the natural logarithm of the electricity prices is very similar to the functions through ; however, as it was shown in , the most dominant frequencies present in the time series are different than the ones present in the time series. Hence, the deterministic functions that are going to be used to model the log prices are proposed as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

## **Price models and parameter estimation**

In order to estimate the parameters of the model using the discrete data available, the continuous-time equations were discretized as follows:

|  |  |  |
| --- | --- | --- |
|  |  | (9) |
|  |  |  |
|  |  | (10) |

Where the stochastic parts are independent and identically distributed (i.i.d.) for every hour . Now, by implementing this discretization and the deterministic component that were explained before, the following models are proposed:

***Model 1.*** *Price- Time-related variables only*

***Models 2. through 10.*** *Price- Time-related variables and ENSO probabilities forecast*

Where, if m=2, the second model is being represented, if m=3, the third model is being represented and so on.

***Models 11. Through 19.***  *Natural Logarithm of price- Time-related variables and ENSO probabilities forecast*

Where, if m=2, the second model is being represented, if m=3, the third model is being represented and so on.

**Monte Carlo Simulation**