

---

# Propuesta de un Modelo con Redes Neuronales y Metodología Box & Jenkins para el Pronóstico del Precio de Bolsa de la Energía en Colombia

---



Presentado por

MSc. Ing. Ricardo Andrés Martín Mayorga

Ing. Luis Felipe Piñeros Sánchez

LOS LIBERTADORES  
FUNDACIÓN UNIVERSITARIA

**Fundación Universitaria Los Libertadores**

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2019



---

**Propuesta de un Modelo con Redes  
Neuronales y Metodología Box &  
Jenkins para el Pronóstico del Precio  
de Bolsa de la Energía en Colombia**

---

**LOS LIBERTADORES**

Presentado por

**MSc. Ing. Ricardo Andrés Martín Mayorga**

**Ing. Luis Felipe Piñeros Sánchez**

en cumplimiento parcial de los requerimientos para optar al título

de

**Especialista en Estadística Aplicada**

*Dirigida por*

**MSc. Juan Camilo Santana**

Profesor

**Fundación Universitaria Los Libertadores**

Facultad de Ingeniería y Ciencias Básicas

Especialización en Estadística Aplicada

Bogotá D.C, Colombia

2019



**Notas de aceptación**



---

---

---

---

---

# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

---

Firma del presidente del jurado

---

Firma del jurado

---

Firma del jurado





# LOS LIBERTADORES

## FUNDACIÓN UNIVERSITARIA

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.



# Dedicatoria

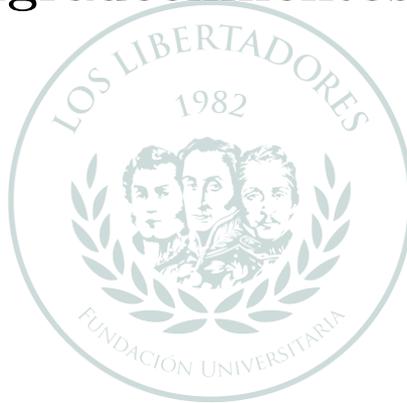


Este trabajo esta dedicado a nuestras familias quienes apoyaron y acompañaron nuestro compromiso con el estudio de la estadístico como parte integran de nuestra formación profesional.

LOS LIBERTADORES  
FUNDACIÓN UNIVERSITARIA



# Agradecimientos



# LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Agradecemos a nuestras familias por su paciencia y compañía. Así mismo, a los docentes de cada semestre que compartieron sus conocimientos y al director del trabajo de grado por su apoyo, paciencia y compromiso.



# Índice general

<b>1</b>	<b>Introducción</b>	<b>3</b>
<b>2</b>	<b>Planteamiento del Problema</b>	<b>5</b>
2.1	Formulación del problema	8
2.2	Pregunta de investigación	9
2.3	Objetivos	10
2.3.1	Objetivo General	10
2.3.2	Objetivos Específicos	10
2.4	Justificación	11
<b>3</b>	<b>Marco Teórico / conceptual</b>	<b>13</b>
3.1	El sector energético en Colombia	13
3.2	Los precios de energía en Colombia	14
3.3	Metodología Box & Jenkins	15
3.4	Redes Neuronales	19
<b>4</b>	<b>Marco Metodológico</b>	<b>21</b>
4.1	Diseño Metodológico	21
4.2	Etapa 1 Selección y Caracterización	22
4.3	Etapa 2 Definición del Modelo	22
4.4	Etapa 3 Propuesta y Desarrollo Metodológico	23
<b>5</b>	<b>Análisis y Resultados</b>	<b>25</b>
5.1	Selección de las variables	25
5.2	Definición de las variables de entrada	27
5.3	Definición del Modelo	28
5.3.1	Regresión Lineal Múltiple	31
5.3.2	Análisis de Componentes Principales (ACP)	32
5.4	Metodología Box & Jenkins	33
5.4.1	Arimax	36
5.4.2	VAR	41

5.5	Red Neuronal . . . . .	47
5.6	Discusión de resultados . . . . .	50
<b>6</b>	<b>Conclusiones y Recomendaciones</b>	<b>53</b>
6.1	Conclusiones . . . . .	53
6.2	Recomendaciones . . . . .	54

# Índice de figuras

2.1	Distribución de generación energía eléctrica en Giga Watt . . . . .	5
2.2	Marco Regulatorio . . . . .	6
2.3	Definición de tarifas por disponibilidad horaria . . . . .	7
2.4	Cadena de Valor Energía Eléctrica . . . . .	8
3.1	Serie de tiempo . . . . .	15
3.2	Red Neuronal . . . . .	19
5.1	Histogramas . . . . .	30
5.2	Boxplot . . . . .	30
5.3	ACP . . . . .	32
5.4	Selección de las Variables . . . . .	33
5.5	Serie de tiempo PBN . . . . .	34
5.6	Serie PBN Outliers . . . . .	34
5.7	Series Estructura con Tendencia . . . . .	35
5.8	Series Estructura Estacional . . . . .	35
5.9	Ajuste De los Datos . . . . .	38
5.10	ACF . . . . .	38
5.11	Residuales . . . . .	39
5.12	Pronóstico PBN ARIMAX . . . . .	40
5.13	Ajuste de los residuales del PBN . . . . .	43
5.14	Test CUSUM PBN . . . . .	43
5.15	Función de covarianza y correlación . . . . .	44
5.16	Pronóstico PBN VAR . . . . .	45
5.17	Impulso de respuesta . . . . .	46
5.18	Descomposición de la varianza . . . . .	46
5.19	Arquitectura Neuronal MLP . . . . .	48
5.20	Pronóstico PBN Red Neuronal . . . . .	49
5.21	Bandas de Confianza pronóstico PBN Red Neuronal . . . . .	50



# Índice de cuadros

5.1	Pruebas de Verificación . . . . .	31
5.2	Resultados ACP . . . . .	32
5.3	Prueba Dickey-Fuller . . . . .	36
5.4	Rezagos de las Variables . . . . .	37
5.5	Estimación del Modelo PBN . . . . .	37
5.6	Pruebas de Verificación . . . . .	39
5.7	Prueba de causalidad . . . . .	42
5.8	Pruebas de Verificación . . . . .	42
5.9	test de cointegración de Johansen . . . . .	44
5.10	Resultados de los errores de los modelos . . . . .	50



---

# Propuesta de un Modelo con Redes Neuronales y Metodología Box & Jenkins para el Pronóstico del Precio de Bolsa de la Energía en Colombia

## Resumen

El presente trabajo se ha desarrollado haciendo uso de información del sector de la energía eléctrica generada por hidroeléctricas, donde se tomo principalmente de *Xm información inteligente*. Dentro de los datos obtenidos se busca realiza el pronostico del precio de la bolsa de la energía.

Se propone realizar un modelo de pronostico por medio de la metodología Box & Jenkins y redes neuronales con perceptron multicapa (MLP), donde el resultado es comparar cual es el pronostico que mejor se ajusta se ajusta a la serie de datos.

**Palabras claves:** Precio de bolsa de la energía eléctrica, Pronósticos, Metodología Box & Jenkins, Redes Neuronales.



# Capítulo 1

## Introducción

En los últimos años se viene presentando situaciones asociadas al cambio climático, y el calentamiento global, que afectan distintos sectores de la economía Colombiana. Uno de los sectores que tienen un alto impacto es el sector energético que enfrenta importantes retos frente al fenómeno del niño, que hoy se sabe que es cada vez más fuerte, afectando de forma directa los niveles de los embalses de generación de energía eléctrica a partir de hidroeléctricas, que como se sabe es una de las grandes componentes de la matriz energética en Colombia. Situaciones como esta hace que los precios de la energía se vean afectados que en periodos de sequía asociados al fenómeno del niño genera incrementos de en los precios por kWh.

Estas situaciones hacen que la respuesta a la demanda se vea afectada dada la falta de capacidad de generación ofertada por los embalses, y es allí donde otras alternativas como la generación por parte de termoeléctricas y el hoy creciente auge de las energías renovables, supliendo las necesidades de la respuesta a la demanda.

Desde la creación de la ley 1715:14 se hace una apertura al mercado de la energía a fin de hacer mucho más competitivo este sector económico, sin embargo, esta apertura tendrá un crecimiento lento y progresivo que puede verse reflejado solo hasta unos años adelante. En este sentido el comportamiento de los negocios se verán afectados, por lo cual, es necesario lograr comprender de mejor manera el futuro de este sector.

Para la comprensión del sector se cuenta con una serie de datos que pueden llegar a definir nuevas estrategias tecnológicas, procesos de compra y venta de energía y posterior crecimiento que se espera conseguir con la nueva regulación Colombiana. Para este fin, el análisis del sector requiere la comprensión del comportamiento de ciertos factores que afectan al mercado, empresas y usuarios finales, donde uno de los factores posibles a tener en cuenta es el precio de bolsa nacional de la energía que depende de condiciones climáticas que inmediatamente se puede asociar a factores estacionales.

En el sector de la energía y con el crecimiento de la cuarta revolución industrial donde el manejo de los datos se hace un importante componente para la toma de decisiones,

y estas decisiones deben estar enfocadas en proyecciones hacia el futuro, con lo cual sea posible obtener estimaciones en las componentes del precio de la bolsa de la energía nacional.

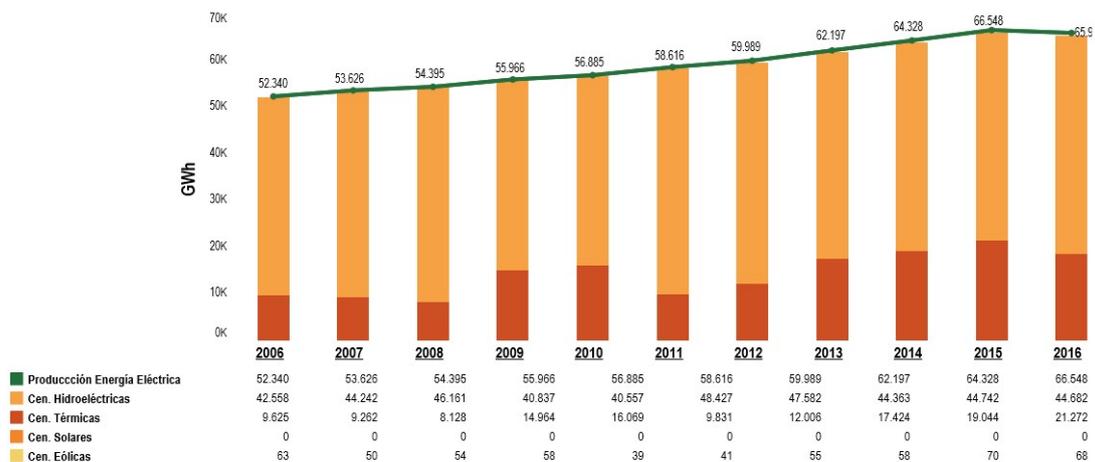
## Capítulo 2

# Planteamiento del Problema

En la actualidad a nivel mundial la energía cuenta con una serie de desafíos, los cuales se componen principalmente de: **(1)** atender correctamente la demanda que requieren los consumidores tanto a nivel residencial como industrial, **(2)** atender la demanda energética de forma medioambiental responsable reduciendo las emisiones de CO<sub>2</sub> y **(3)** suministrar dicha energía a precios razonables brindando bienestar social y económico.

Estos desafíos no son ajenos para Colombia, que actualmente cuenta con grandes fuentes generadoras de energía y cuya participación se muestra en **Figura 2.1**, estos valores representan las estadísticas desde el año 2010 al 2016, tomadas del portal de la Unidad de Planeación Minero Energética (UPME, *s.f. a*). Los datos muestran la distribución de consumo y producción de energía, permitiendo visualizar su comportamiento frente a combustibles como el carbón, petróleo, gas natural, entre otros.

Figura 2.1: Distribución de generación energía eléctrica en Giga Watt



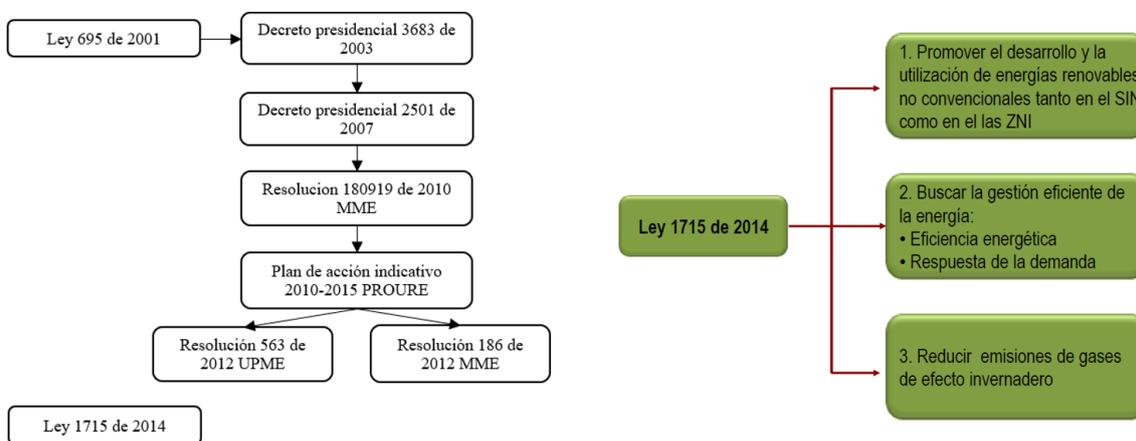
Fuente: (UPME, *s.f. a*)

Para el año 2016 se tiene un porcentaje de participación de centrales hidroeléctricas y

centrales térmicas de un 71 % y 28 % respectivamente y tan solo un 0,1 % para centrales eólicas. Por otra parte, se tiene un 1 % para auto y cogeneración.

Con los anteriores datos, se puede apreciar como la demanda energética viene creciendo en cada una de las fuentes generadoras de energía año tras año, lo cual ha generado preocupación en cómo se puede atender esta demanda según los desafíos que se mencionaron anteriormente. Es así, como en Colombia a través de una serie de regulaciones y en especial la ley 1715:14 ver **Figura 2.2** se ha incrementado la promoción de actividades en pro de atender la respuesta a la demanda tanto en kWh como en pesos colombianos COP frente al establecimiento de las tarifas de la energía y la reducción de las emisiones de efecto invernadero.

Figura 2.2: Marco Regulatorio



Fuente:(UPME, s.f. a)

Con respecto a los precios de la energía en el mercado colombiano, estos son regulados por el Estado, y en cuanto a la provisión del servicio participan la (generación, transmisión, distribución y comercialización). Para poder generar los precios adecuados actualmente a nivel internacional, principalmente en Alemania se han adoptado dos metodologías principales para el pronóstico de la demanda energética y del precio de oferta. El primero de ellos hace referencia al Bottom-Up, el cual es un modelo macroeconómico que permite establecer la demanda de energía en kWh de abajo hacia arriba, es decir, desde el consumo hacia la fuente, unos de los artículos cuyo principal exponente y que presenta un modelo de cadenas de Markov es (Shepero & Munkhammar, 2018), la segunda metodología es el Top-Down, el cual consiste en definir modelos estadísticos principalmente a fin de pronosticar la demanda energética de arriba hacia abajo, es decir, desde la generación hasta la comercialización y posterior consumo. En su trabajo (Richardson, Thomson & Infield, 2009), menciona diferentes técnicas estadísticas para lograr establecer tanto demanda energética kWh como precios de oferta.

En Colombia las tarifas se fijan con base en el costo unitario, el cual se ajusta, en cierto

grado, con los principales índices de precios de la economía, dichas tarifas son reguladas por la Comisión de Regulación de Energía y Gas (CREG) mediante resoluciones que en principio tienen vigencia por cinco años. Las tarifas definidas en pesos por kilovatio hora (\$/kWh), tiene como base de cálculo el costo unitario (Lozano-Espitia, Rincón-Castro, Rincón-Castro & Lozano-Espitia, 2010).

La forma en que actualmente Colombia realiza los despachos de energía con base en su demanda es a través de los agentes generadores quienes un día antes reportan al Centro Nacional de Despacho (CND) sus precios de oferta y su disponibilidad para las 24 horas del día siguiente ver **Figura 2.3**. Según (Gómez, 2017) quien cita a (Ossa López, 2012) “el CND ordena estos precios por orden de mérito de menor a mayor, y con base en estos, determina hora a hora qué plantas deben generar, hasta cubrir la demanda. El mayor precio de las ofertas correspondiente a la central flexible con el cual se cubre la demanda en una hora determinada, se denomina Máximo Precio Ofertado y es la base para la determinación del precio de bolsa”. Así mismo, se conoce que las plantas generadoras realizan vía subasta la oferta de la electricidad disponiendo de diferentes cantidades de energía a los precios a los que están dispuestos a generarlas, estas subastas, son el mecanismo utilizado para formar el precio de bolsa de la energía eléctrica. Lo anterior muestra que Colombia tiene ciertos mecanismos que permiten regular las tarifas de energía eléctrica.

Figura 2.3: Definición de tarifas por disponibilidad horaria

Tarifas Horarias	Precios de pico crítico	Precios de la energía en Tiempo Real
El precio del consumo depende de un rango de horas, el cual define tarifas para momentos de picos bajos, medios y altos. Las tarifas son fijas por cada periodo por lo que el cliente conoce cuáles serán los precios.	El precio de la energía puede ser de 3 a 10 veces más costoso durante los eventos de pico crítico. Estos eventos se dan a conocer con un día o algunas horas de anticipación y generalmente tienen una duración de entre 2 y 6 horas.	Las tarifas están basadas en los precios del mercado mayorista. Por lo general, los clientes reciben notificaciones con los precios con un día o una hora de anticipación.

*Fuente: (Inteligente, s.f.)*

En el análisis realizado por (Gómez, 2017) concluye con la afirmación que es incierto realizar un análisis de cómo se comportan los precios de oferta debido principalmente a estos pueden verse afectados por la hidrología y los fenómenos climáticos, variables internas del país, regulaciones asociadas al ajuste de los precios, la creación de impuestos, disponibilidad de los generadores, participación de los mismos y de los recursos, introduc-

ción de energías alternativas, precios de bolsa, la tasa de cambio, entre otras variables de influencia.

Es así como Colombia con la apertura del mercado de la energía, y el crecimiento de la demanda, tiene el reto de establecer mejores herramientas que permitan hacer un buen establecimiento de las tarifas convenientes en la cadena de suministro de la energía ver **Figura 2.4**, tanto por consumos picos durante el día, como en fines de semana, o proyecciones a largo plazo, entendiéndose largo plazo por períodos entre 6 y 12 meses dada la volatilidad de la energía en ciertas regiones. Estos consumos pueden ser analizados desde el precio de la bolsa el cual es uno de los factores que afecta el análisis del comportamiento de los precios de oferta de energía, según (Xm, s.f. d) : “En condiciones normales de operación, corresponde al mayor precio de oferta de las unidades con despacho centralizado que han sido programadas para generar en el despacho ideal y que no presentan inflexibilidad. Representa un precio único para el sistema interconectado en cada periodo horario”. Lo anterior define la importancia de realizar un análisis del precio de bolsa según lo sugerido por (Gómez, 2017) que es uno de las muchas variables que definen la dificultad del análisis del comportamiento de los precios de oferta.

Figura 2.4: Cadena de Valor Energía Eléctrica



Fuente:(Grupo, s.f.)

## 2.1 Formulación del problema

Según lo definido en el planteamiento del problema, en el cual se muestra la importancia de realizar un análisis del comportamiento de variables económicas y financieras

de sector de la energía, se selecciona como variable de interés el precio de bolsa nacional. En la que se identifican los siguientes cuestionamientos según (Henao, Resonsew & Souza, 2007):

1. ¿Cuáles son las propiedades de los precios medios mensuales?
2. ¿Es posible obtener pronósticos para los factores más importantes, al menos en el mediano plazo?
3. ¿Cuáles son los modelos más adecuados?

Con lo anterior, se define como situación problema el análisis del precio de bolsa, logrando identificar los factores que influyen en el pronóstico de estos precios y que permitan ser ajustados a un modelo estadístico adecuado para su análisis.

## 2.2 Pregunta de investigación

¿Cómo diseñar una propuesta de un modelo con Redes Neuronales y Metodología Box & Jenkins para el pronóstico del precio de bolsa de la energía eléctrica en Colombia?

## **2.3 Objetivos**

### **2.3.1 Objetivo General**

Realizar una propuesta de un modelo Redes Neuronales y Metodología Box & Jenkins para pronosticar el precio de bolsa de la energía eléctrica en Colombia.

### **2.3.2 Objetivos Específicos**

- Identificar los factores que influyen en el precio de bolsa de la energía eléctrica en Colombia.
- Seleccionar el modelo adecuado bajo la metodología Box & Jenkins para el análisis del pronóstico del precio de bolsa de la energía eléctrica en Colombia.
- Elaborar una red neuronal autorregresiva para realizar los pronósticos de precios.
- Comparar y discutir los resultados del modelo de Red Neuronal y el modelo bajo la metodología Box & Jenkins.

## 2.4 Justificación

El precio de bolsa de la energía eléctrica en Colombia, se caracteriza por ser uno de los Commodities con mayor fluctuación a nivel mundial. Por lo cual, llevar a cabo análisis o pronósticos se convierte en un verdadero reto como lo mencionan en la investigación adelantada por (Muñoz-Santiago, Urquijo-Vanstrahlengs, Castro-Otero & Lombana, 2017). La alta volatilidad que presenta el precio de bolsa, se debe a factores de diferentes orden: económico y climático.

Colombia se ha establecido en el mercado mayorista energético la siguiente premisa: *"hacer cada vez más competitivas las negociaciones, en aras de cubrir las demanda y lograr precios más bajos de la energía"* (Lozano-Espitia y col., 2010). Para lograr este objetivo se tiene que dar una armonía en la ejecución de los diferentes agentes del mercado (Generadores, Transmisores, Distribuidores y Comercializadores). Estas metas deben estar encaminadas con diferentes estrategias en el corto, mediano y largo plazo y el planteamiento de las mismas debe estar basado en la información del pasado.

A partir del análisis de estos datos históricos, se pueden generar métodos de predicción: pronóstico con series temporales y métodos de avanzada como lo pueden ser las redes neuronales artificiales. El resultante de estos modelamientos y sus respectivos análisis, convierten a los modelos de pronóstico en un vehículo contribuyente a construir la planeación estratégica del mercado mayorista energético.

Diversos estudios de precios de la energía eléctrica en Colombia, han generado una serie de discusiones y planteamientos desde el punto de vista gubernamental en donde se han establecido regulaciones que han permitido un mejor desarrollo del mercado energético, así mismo, se cuenta con estudios de investigación que buscan estudiar el comportamiento de estos precios generando distintas metodologías de pronóstico. Estos estudios llegan a conclusiones que son interesantes de evaluar y que justifican el desarrollo de estos trabajos.

En primer lugar, se reconoce la dificultad que enfrenta el país en materia de energía y este como afecta a los precios tanto de su generación como de la compra para su posterior distribución al usuario final, en el trabajo de (Rendón, Hinestroza & Moreno, 2011) se hace un análisis del mercado de la energía y su importancia de analizar e identificar los aspectos que determinan los precios para los mercados no regulados en Colombia, a partir de series de tiempo SARIMAX y que como variables estudiadas tiene (demanda, precio marginal del sistema y aportes hídricos). Del resultado de esta investigación se establece que el precio de bolsa muestra alta volatilidad debido a factores climáticos, y como principal variable para determinar el precio de bolsa se encuentra el precio marginal del sistema, teniendo comportamientos estacionales en los fines de semana. El estudio concluye con la recomendación de realizar más estudios que permitan mejorar los resultados propuestos, que incluyan variables basados en expectativas climatológicas y cómo afecta el desarrollo de mercados derivados de energía.

Es claro que se pueden incluir diferentes variables de análisis del precio de bolsa que permitan llegar a tener mejores resultados y conclusiones. En el trabajo de (Medina Hurtado, 2014) menciona una serie de beneficios que se tienen al realizar un adecuado análisis de estas variables y comportamientos; dentro los mencionados se encuentra: **(1)** permite realizar toma de decisiones a corto, mediano y largo plazo, **(2)** permite estimar la cobertura del mercado de bolsa o spot, **(3)** estructurar un portafolio energético para generadores, **(4)** cuantificación de los riesgos asociados a la volatilidad de los precios, y **(5)** la valoración de derivados en los precios. Lo anteriores beneficios son el resultado de un análisis conveniente según las variables o factores a estudiar, lo que conlleva a definir la importancia de una correcta selección de estos factores y que sean del alcance del estudio a realizar, brindando resultados alternativos a los beneficios mencionados y que agreguen valor a la toma de decisiones en el mercado energético.

Según el trabajo de (Barrientos, Rodas, Velilla, Lopera & Villada, 2012), indica que el IPC, la tasa de cambio real y el precio internacional de combustibles no afectan de forma significativa los precios de la energía debido a que la generación de la energía está basada principalmente hidroeléctricas. Por el contrario se menciona que las variables que tienen un efecto son el PIB, la demanda de electricidad y los niveles de los embalses, los cuales son factores sin ser tenidos en cuenta en otros estudios para el análisis de los precios. Sin embargo, el autor concluye que los niveles de los embalses tienen una relación inversa frente al precio, también se menciona que el precio se ve afectado por el crecimiento económico es decir el PIB. Estos son factores adicionales que permiten definir una mejor estructura para el análisis de los precios de bolsa y su importancia para el mercado nacional.

# Capítulo 3

## Marco Teórico / conceptual

### 3.1 El sector energético en Colombia

El sector energético en Colombia funciona a través de diferentes empresas y entidades que cumplen diversas funciones enmarcados en la generación, transmisión, comercialización y distribución de la energía eléctrica ver **Figura 2.4**. Este mercado energético es regulado por la Comisión de Regulación de Energía y Gas (CREG) *“es un organismo estatal colombiano, dedicado a regular las actividades de prestación de servicios públicos domiciliarios relacionados con energía eléctrica, gas natural, gas licuado de petróleo (GLP) y combustibles líquidos, con el fin de lograr que tales servicios se presten al mayor número posible de personas, al menor costo y equilibrando la compensación para las empresas prestadoras garantizando así calidad, cobertura y expansión”* (CGRE, s.f.). También se encuentra la Superintendencia de Servicios Públicos Domiciliarios (SSDP), que es la encargada de controlar y vigilar a las entidades prestadoras de servicios públicos domiciliarios, y por último Unidad de Planeación Minero Energética (UPME), es la unidad adscrita al Ministerio de Minas y Energías cuya función es la de elaborar los planes de expansión del sector eléctrico a nivel financiero, económico y ambiental.

Como se mencionó anteriormente existe una generación de la energía eléctrica la cual garantiza la producción energética en el país, siendo controlada por el Estado a fin de garantizar una estructura competitiva evitando los monopolios, esta generación es principalmente a través de hidráulica, es decir, centrales hidroeléctricas las cuales generan energía eléctrica por medio del almacenamiento de agua en embalses. Es por lo anterior que los cambios climáticos, por ejemplo, en temporadas secas hacen que los precios de la energía aumenten debido a la falta de agua para generar la oferta de energía la cual se traduce como la capacidad disponible por parte del generador que es capaz de entregar al Sistema Interconectado Nacional (SIN), esta es la red que conecta al generador con los usuarios finales. Es conveniente saber que existen usuarios que generan su propia energía, dentro de estos están los auto generadores, los cuales tienen la capacidad de generar su

propia energía por medio de sus procesos productivos o por contar con fuentes no convencionales de energía, y en segundo lugar están los cogeneradores quienes producen su energía en procesos de producción eléctrica y térmicas, en ambos casos el usuario no necesita estar conectado al SIN.

Luego que ya se cuenta con una oferta entra una segunda etapa la cual es la transmisión, esta actividad se encarga del transporte de la energía en grandes volúmenes de voltaje a través del SIN, este se compone principalmente de unas líneas de transmisión y subestaciones con equipos y transformadores que tienen capacidades de 220 y 500 Kv. La transmisión es regulada por el Estado brindando acceso a cualquier usuario. Como empresa destacable en Colombia esta ISA quien cuenta con el 75% de los activos de la red.

Posterior se cuenta con la distribución de la energía, es aquí donde los altos volúmenes de voltaje se transforman para poder ser entregados al usuario final a niveles menores a 220kV. El Estado regula este proceso a fin de brindar un servicio de calidad, y con precios razonables. Así mismo garantizando la cobertura de la distribución de la energía en el país.

Se finaliza con la comercialización, la cual opera realizando la compra de la energía al generador para venderla al usuario final. En este caso el Estado debe garantizar que se generen monopolios al igual que en la generación y se cree una actividad competitiva. En este caso la comercialización puede ser dada a usuarios regulados, en este caso es el que no puede contratar de forma directa la energía y son atendidos por una empresa comercializadora que brinde el servicio en ese mercado, y se tienen a los usuarios no regulados, que son lo que pueden contratar un comercializados de forma libre.

## **3.2 Los precios de energía en Colombia**

Las negociaciones de la energía se pueden dar para el caso de los usuarios regulados por medio de negociaciones bilaterales, que se refieren a generar contratos de largo plazo, y las de corto plazo por medio de mercados de precio bolsa de la energía. Esta es una figura comercial en donde se generan ofertas y demandas del mercado eléctrico y permitiendo la compra y venta de energía. Adicional a lo anterior los contratos bilaterales se pueden ver como un primer tipo de contrato definido como pague lo contratado, y que se refiere a que sin importar si no se consumió toda la energía el comercializador debe pagar al generador la suma total del contrato, y un segundo tipo de contrato llamado pague lo demandado o lo consumido, que como su nombre lo indica es un contrato que es pagado sólo por una demanda o un consumo. En ambos casos en que los requerimientos de energía sean mayores a los pactados, el comercializador deberá contratar el excedente de energía a través del precio de bolsa, las anteriores negociaciones son reguladas por la CREG. Para el caso de los usuarios no regulados, las negociaciones no están sujetas a un

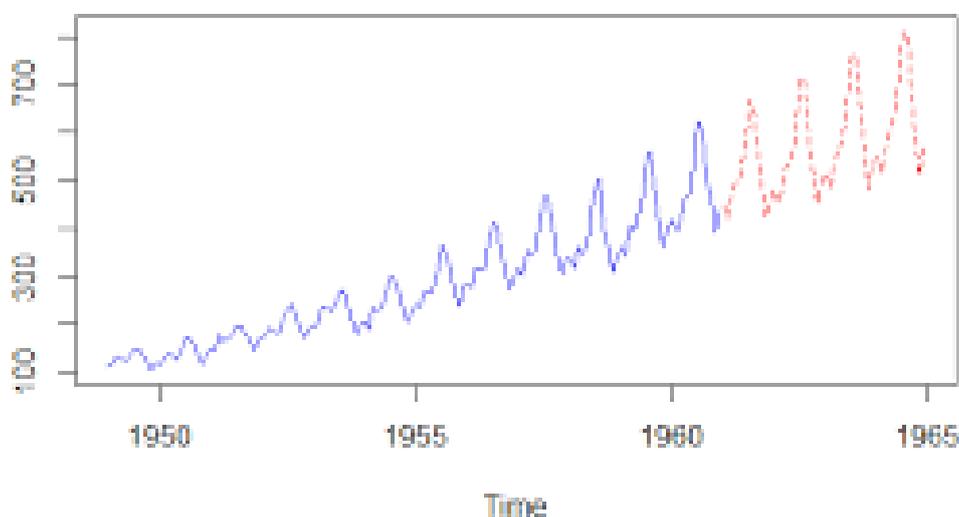
tipo de contrato en especial, y las partes son libres de definir los plazos de los contratos y los precios de venta, no requieren de una autorización por parte de la CREG.

Contextualizando más a fondo a cerca de la bolsa de energía llamada también como mercado spot, funciona con contratos a corto plazo, es decir, a menos de un día, que como se vio antes es con el fin de suplir necesidades no atendidas en contratos bilaterales de largo plazo. En este caso los generadores ofertan un día anterior unos valores únicos para el día siguiente. Este mecanismo funciona por medio de subastas de precio uniforme.

### 3.3 Metodología Box & Jenkins

La metodología **Box & Jenkins**, fue propuesta por George E. P. Box y Gwilym Jenkins durante los años 70s. Esta es una aplicación para el análisis de las series de tiempo enfocada al control y pronóstico, también conocida como metodología **ARIMA**, donde: **AR** se define como un proceso autorregresivo, **I** es el orden de integración para una serie de tiempo bajo estructura estacionaria, y **MA** de media móvil que se calcula a fin de suavizar la serie de datos (Muñoz-Santiago y col., 2017).

Figura 3.1: Serie de tiempo



*Fuente:(Geomatica, s.f.)*

Las observaciones que están distribuidas en el tiempo son definidas como series de tiempo, esta distribución ordenada puede estar dada como la observación  $n = 1$  en el tiempo  $t_1$  hasta la observación  $k$  en el tiempo  $t_n$ , donde  $k$  es el número de observaciones de una variable de interés a lo largo de un tiempo  $t$  que puede verse como horas, días, meses, años etc.

De acuerdo a la metodología **ARIMA**, la estructura de las observaciones a pronosticar es que la serie debe tener una media y varianza estable en el tiempo, con lo que se espera construir un modelo de pronóstico que con la información del pasado pueda predecir la información del futuro. Para lo anterior las observaciones deben estar ordenados de manera secuencial y mostrar una tendencia como se muestran en la **Figura 3.1**, la estacionalidad o ambas según sea el caso, lo anterior con el fin de poder analizar estos comportamientos (Montenegro, 2010), y lograr caracterizar las variables según su tendencia y estacionalidad.

La metodología **ARIMA** comprende cuatro etapas:

1. Identificación
2. Estimación
3. Verificación
4. Pronóstico

#### **Etapla identificación:**

Se busca encontrar los valores de  $p$ ,  $d$  y  $q$  del modelo **ARIMA** que está asociado a un tipo de proceso estocástico,  $p$  hace referencia los términos autorregresivos de la parte **AR** y  $q$  a los términos de la parte **MA** del modelo. Dentro de esta etapa es necesario lograr que la serie sea estacionaria, por lo cual se aplican pruebas de estacionariedad a fin de evaluar su comportamiento, la variable puede diferenciarse  $d$  veces hasta que su estructura sea estacionaria, lo anterior también se define como el orden de integración  $I(d)$  del modelo (Andes, s.f.).

Al contar con la serie estacionaria se aplican se definen criterios de estimación de Función de Autocorrelación **ACF**, que en caso que los coeficientes de correlación sean altos indica que la serie es no estacionaria y la Función de Autocorrelación Parcial **PACF** como un complemento al **ACF** cuya finalidad es definir los parámetros  $p$  y  $q$ , donde puede ser necesario incluir rezagos estadísticamente significativos y que en cierta manera se hallan de forma empírica. Por último queda validar que el modelo es estacionario se realizan pruebas de Raíz Unitaria como puede ser bajo *Dickey-Fuller* que confirman o rechazan la hipótesis de estacionariedad.

#### **Etapla estimación:**

Luego de contar con criterios **ACF** estimados es necesario estimar los coeficientes autorregresivos **AR** y de media móvil **MA** que fueron incluidos en el modelo junto con los rezagos ya identificados. Durante esta estimación puede darse el caso que el modelo sólo contemple coeficientes **AR**, sin embargo, incluir la parte **MA** puede ayudar a mejorar la estimación, convirtiéndose así en un modelo **ARMA**, esto en caso que la serie ya tenga una estructura estacionaria, caso contrario el modelo sería **ARIMA**, con coeficientes **AR**, con orden de integración  $I(d)$  y coeficientes **MA**.

Un modelo **AR** (**p**) se define como:

$$x_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + w_t \quad (3.0)$$

donde  $x_t$  es estacionario,  $\phi$  son constantes diferentes de 0 y  $w_t$  es ruido blanco con media cero y varianza constante.

Un modelo **MA** (**q**) se define como:

$$x_t = w_t + \theta_1 w_{t-1} + \theta_2 w_{t-2} + \dots + \theta_p w_{t-p} \quad (3.0)$$

donde  $\theta$  son parámetros del modelo con  $\theta$  diferente de cero y  $w_t$  es un ruido blanco con varianza constante. A diferencia del **AR**, un **MA** es estacionario.

Un modelo **ARMA** (**p,q**) se define como:

$$x_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + w_t + \theta_1 w_{t-1} + \dots + \theta_q w_{t-q} \quad (3.0)$$

con  $\phi$  y  $\theta$  diferentes de cero y varianza mayor que cero.

La búsqueda de los coeficientes  $p$  y  $q$ , pueden ser de difícil identificación por lo cual se recomienda contar con varios modelos a fin de evaluar entre ellos cual es más apropiado para continuar con la siguiente etapa.

Una generalización de los modelos **ARIMA** son los modelos Vectores Autorregresivos **VAR**, el cual es un modelo de ecuaciones de forma reducida sin restringir, lo cual quiere decir que los valores contemporáneos de las variables del modelo no aparecen como variables explicativas.

Un modelo (**VAR**) se define como:

$$Y_t = C + \phi_1 Y_{t-1} + \phi_p Y_{t-p} + \alpha_t \quad (3.0)$$

### **Etapas de verificación:**

En esta etapa también llamada diagnóstico se evalúa si el modelo estimado se ajusta de forma apropiada a los datos, y como se sugirió en la etapa anterior contar con más de un modelo es posible que más de un modelo también se ajuste a los datos.

Se deben analizar los coeficientes para determinar que se cumpla con las condiciones de estacionariedad e invertibilidad, junto con que los rezagos tengan significancia estadística. En caso que los parámetros no sean significativos es necesario ir anulando parámetro por parámetro hasta lograr un mejor grado de ajuste de significancia. Si dos modelos son apropiados se seleccionará aquel que cuente con el menor error cuadrático medio, si los modelos tienen diferentes cantidad de parámetros se seleccionara el más sencillo basado en el principio de parsimonia (Muñoz-Santiago y col., 2017).

Es importante realizar una prueba de bondad de ajuste, para poder aplicar la selección del mejor modelo basado en el error  $R^2$  ajustado, **AIC** y **SC**, junto con un análisis a los

residuales.

Para el caso del  $R^2$  ajustado solo tiene validez para la comparación de modelos que presentan el mismo número de diferencias, debido a que la varianza total de cada modelo depende del número de diferencias realizadas.

El Akaike Information Criterion (**AIC**) y el Schwarz Criterion (**BIC**), son herramientas que permiten establecer el número apropiado de rezagos en los coeficientes  $p$  y  $q$ , esta prueba también es conocida como prueba de parsimonia (Andes, [s.f.](#)).

**AIC** se define como:

$$AIC = -2\log(\text{maximum likelihood}) + 2k \quad (3.0)$$

donde  $k$  es el numero de parámetros estimados del modelo junto con el máximo valor de la función likelihood.

La fórmula para **SC** es:

$$BIC = -2\log(\text{maximum likelihood}) + k\log(n) \quad (3.0)$$

El resultado de la aplicación de este diagnóstico es que el **AIC** y el **BIC** tengan valores lo menor posible entre los modelos comparados.

Para el análisis de los residuales, pueden ser evaluados bajo una gráfica de los errores que permita visualizar si los datos tienen varianza constante o correlación.

Una segunda manera de evaluar los residuales es por medio del estadístico *Ljung-Box-LB* que se evalúa por medio de un correlograma a fin de saber si los errores del modelo son ruido blanco, es decir, que su media es cero, la varianza es constante y es incorrelacionada (ver ecuación). En caso que el modelo no sea ruido blanco, da señal que la estructura del modelo no se logró capturar.

donde  $\alpha$  es el nivel de significancia. Si el estadístico *LB* es menor al nivel de significancia se puede afirmar que los errores del modelo siguen un proceso de ruido blanco, por tanto no se puede rechazar  $H_o$ .

Finalmente si los errores siguen una distribución normal se dice que es un modelo ruido blanco gaussiano. Para lo anterior puede usarse un histograma para validar la forma de su distribución, o utilizar el *qqPlot*.

Si el modelo cuenta con errores autocorrelacionados es necesario volver a la etapa de identificación para ajustar el modelo hasta que finalmente los errores cumplan con las condiciones mencionadas anteriormente.

#### **Etapas pronóstico:**

Aquí se define el pronóstico del modelo seleccionado según la aplicación de las etapas anteriores, en el periodo  $h$  de interés.

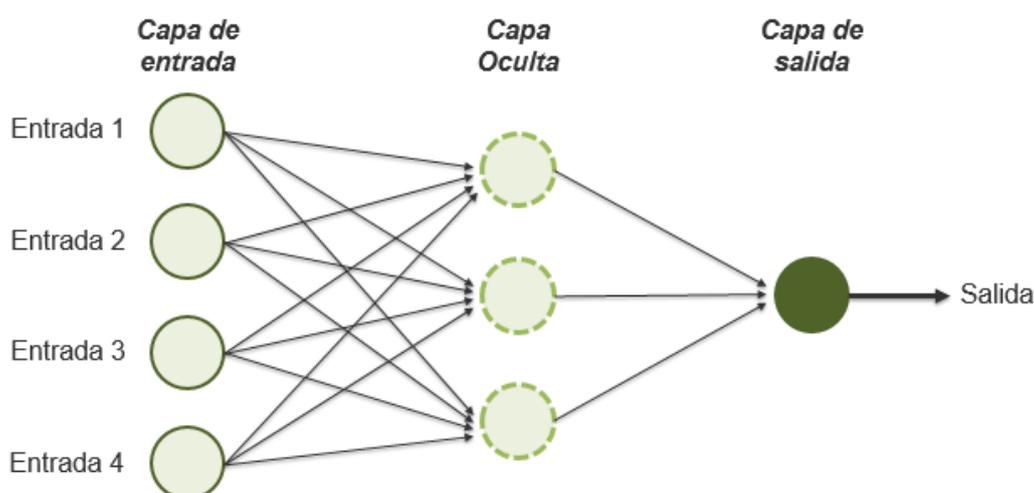
### 3.4 Redes Neuronales

Las redes neuronales artificiales (RNA) han sido utilizadas como métodos de pronóstico avanzados, en los cuales se ejecutan relaciones complejas entre la variable respuesta y sus predictores como lo mencionan en su trabajo (Hyndman & Athanasopoulos, 2018). La configuración y funcionamiento de las (RNA) está inspirado en redes neuronales biológicas.

*“Una neurona biológica es capaz de recibir información desde miles de otras neuronas, procesarla y luego generar una nueva información que enviará a otras neuronas con las que está conectada”* (Escobar, Valdes & Zapata, 2009).

Las redes neuronales artificiales se caracterizan por tener una arquitectura definida, la cual se puede considerar como una red de neuronas. Partiendo por las entradas, que hacen referencia a la información suministrada para el pronóstico, seguido se encuentran las capas ocultas o también llamados nodos ocultos, con el número de neuronas especificado, y finalmente se encuentran las salidas. En la **Figura 3.2** se ilustra la arquitectura de una red neuronal no lineal, con cuatro entradas, una capa oculta con tres neuronas y una salida.

Figura 3.2: Red Neuronal



Fuente:Hyndman y Athanasopoulos, 2018

El entrenamiento es un componente esencial en la redes neuronales, dado que es un proceso en donde se hace el aprendizaje de la misma, en el trabajo realizado por (Rojas, López & Isaza, 2004) se describe este proceso como:

*“Entrenar una red neuronal artificial (RNA) es un proceso que modifica el valor de los pesos y bias asociados a cada neurona [3, 6] con el fin de que la RNA pueda a partir de unos datos presentados en la entrada, generar una salida; en el caso del aprendizaje supervisado [8], se tiene un conjunto de datos pero no se conoce la función o relación*

*matemática que los representa; al propagar hacia delante cada uno de estos patrones se obtiene una respuesta en la salida de la RNA la cual se compara con la salida deseada, permitiendo obtener el error del desempeño de la red. Existen en la literatura especializada diversos algoritmos con capacidad de entrenar redes neuronales [3, 6], sin embargo, tienen como desventaja requerir grandes tiempos computacionales para redes de mediano y gran tamaño y en muchos casos se obtiene una pobre generalización” (Rojas y col., 2004).*

Para hacer pronósticos con series de tiempo en redes neuronales, los rezagos definidos pueden ser utilizados como variables de entrada, de la misma manera en que se utilizan en los modelos de predicción de auto regresión lineal. Estos modelos de pronóstico con (RNA) se conocen como modelos de autorregresión de red neuronal o modelo (NNAR). Éste tipo de modelos son equivalentes a un modelo ARIMA, sin embargo, no se aplican restricciones en los parámetros con el objetivo de asegurar su estacionalidad.

La aplicación de (RNA) en el pronóstico de series temporales, ha sido definida en cinco etapas en el trabajo realizado por (Escobar y col., 2009). **(1)** Búsqueda de las variables de entrada; en esta etapa se deben identificar los rezagos que tiene la serie, de forma, que puedan considerarse como entradas a la red. **(2)** Preparación del conjunto de datos; en esta etapa se realiza el escalamiento de los datos, es decir, se aplica una normalización a los mismos, generando intervalo  $[0, 1]$ . **(3)** Creación de la red; se definen cada elemento que debe comprender la arquitectura o diseño de la red. **(4)** Entrenamiento; en esta etapa se determina el algoritmo de entrenamiento a utilizar en la red. **(5)** Validación; esta etapa tiene como objetivo evaluar el nivel de ajuste en los pronósticos generados por la red, con respecto a los datos reales.

# Capítulo 4

## Marco Metodológico

### 4.1 Diseño Metodológico

Realizar la selección de la base de datos y su caracterización (**Etapa 1**): Se realizará la búsqueda de la base de datos. Basado en el planteamiento del problema y la justificación se realizará el estudio documental: libros, artículos científicos, revisiones, páginas web de entidades gubernamentales a nivel nacional, regional y local, videos, revisiones, e informes, que permitan caracterizar la base de datos en cuanto a la identificación de los factores que influyen en el precio de la bolsa de la energía eléctrica en Colombia.

En esta etapa, se permite dar un enfoque al cumplimiento del objetivo específico: **(1)** Identificar los factores que influyen en el precio de bolsa de la energía eléctrica en Colombia.

Definición del modelo (**Etapa 2**): Aquí es importante realizar un estudio de los modelos estadísticos que pueden ser convenientes para el procesamiento de la base de datos. Con lo anterior se definen la(s) técnica(s) apropiada(s) para la analizar el comportamiento de los precios de bolsa de la energía eléctrica. Se identifican y aplican los modelos estadísticos, y con base en esto se propone el más adecuado.

En esta etapa, permite dar un enfoque al cumplimiento del objetivo específico: **(2)** Seleccionar el modelo de series de tiempo adecuado para el análisis del pronóstico del precio de bolsa de la energía eléctrica en Colombia. Y **(3)** Elaborar una red neuronal autorregresiva para realizar los pronósticos de precios.

Propuesta y desarrollo metodológico (**Etapa 3**): En esta etapa se propone discutir los resultados del análisis del comportamiento del precio de bolsa, y cómo comparar los modelos para discutir los resultados del pronóstico de los precios de oferta.

En esta etapa propuesta y desarrollo metodológico da cumplimiento al objetivo específico: **(4)** Comparar y discutir los resultados del modelo de series de tiempo y de la red neuronal.

## **4.2 Etapa 1 Selección y Caracterización**

Durante esta etapa se realizara la búsqueda de los factores que influyen en la energía eléctrica en Colombia. Para lo cual, se seleccionaran las fuentes principales de información, que para este caso son aquellas que regulan y brindan las directrices de la prestación de este servicio.

Una de las fuentes es La Comisión de Regulación de Energía y Gas (GREG), la cual es una Entidad Estatal colombiana cuya principal función es la regulación de la prestación del servicio de la energía eléctrica, gas natural y gas licuado del petroleo, junto con el fortalecimiento del Ministerio de minas y Energías, y otras entidades gubernamentales que tengan una influencia en este aspecto, como lo es la Unidad de planeación Minero Energética (UPME).

Contando con el marco de las entidades gubernamentales, se realizara una investigación del negocio de la energía en Colombia el cual es mencionado de forma especifica en el Capitulo 3. Con lo anterior se emprenderá la búsqueda de la fuente generadora de los datos y con ello definir la base de datos.

Teniendo la base de datos se consultara bibliografía que permita establecer las variables que puedan influir en el precio de bolsa de la energía que es la variable de interés a estudiar, dicha bibliografía contara con la consulta de Tesis a nivel de pregrados y posgrados, artículos académicos, informes técnicos entre otros. Así mismo, permitirá caracterizar cada una de las variables y poder describir un posible comportamiento en la predicción de la variable de interés por medio de los modelos estadísticos a implementar.

## **4.3 Etapa 2 Definición del Modelo**

A nivel estadístico se definió realizar un modelo de regresión lineal múltiple y un modelo de Análisis Multivariado por medio de un análisis de componentes principales, con el fin de generar la selección de las variables que tienen relación y que puedan generar un mejor pronostico del precio de bolsa nacional y obtener el modelo más parsimonioso.

Continuando el desarrollo de esta etapa se definirán los modelos de series de tiempo por medio de la metodología Box & Jenkins que sean apropiados para el procesamiento de los datos, que para este caso sean aplicables a datos multivariantes debido al ingreso de diferentes variables que pretenden dar respuesta a una variable de interés (Precio de Bolsa Nacional).

En la metodología Box & Jenkins se cuenta con una serie de modelos de series de tiempo a nivel univariado y multivariado, dado que se cuenta con una serie de variables de entrada para el modelo de pronostico, se define tomar series de tiempo multivariadas. Contando con la definición anterior, y según la naturaleza de las variables, se identificaran aquellos modelos de series de tiempo multivariadas que sea apropiadas para realizar el pronostico.

Así mismo, se definirá el modelo de Red Neuronal que permita ser comparada con los modelos de la metodología Box & Jenkins y determinar cual es el mejor modelo para estimar la serie de tiempo y tenga el mejor pronóstico de la variable de respuesta.

#### **4.4 Etapa 3 Propuesta y Desarrollo Metodológico**

De acuerdo a la etapa dos, se desarrolla la propuesta de los modelos Box & Jenkins y la Red Neuronal identificando la estructura adecuada de cada una de las variables definidas como entrada para el pronóstico del Precio de Bolsa de la Energía. La estructura de los modelos Box & Jenkins se definen como los rezagos de acuerdo a si dicha estructura tiene tendencia y estacionalidad, para el caso de la Red Neuronal la estructura también contara con los rezagos de las variables y las Dummis.

Con los modelos identificados en la etapa dos y la estructura definida, se modela la serie de tiempo multivariada, buscando encontrar el modelo que tenga un mejor ajuste a los datos de predicción y que presente el menor error.

Con los resultados de los modelos obtenidos, se realizará la discusión y presentación de las conclusiones respectivas según los objetivos definidos, junto con las recomendaciones que se consideren que den a lugar.



# Capítulo 5

## Análisis y Resultados

### 5.1 Selección de las variables

Investigaciones previas relacionadas a pronósticos del sector energético hacen hincapié en la importancia de realizar modelos de predicción que se ajusten al mercado de la energía, una de las investigaciones a mencionar es la realizada por (Gómez, 2017) en su trabajo del análisis del precio de oferta de la energía eléctrica en Colombia, donde menciona que el precio de oferta puede verse afectado por fenómenos hidrológicos, climáticos, y otras variables asociadas directamente relacionadas con la demanda y generación, junto con factores económicos a nivel de Producto Interno Bruto (PIB) y el Índice de Precios al Consumidor (IPC) los cuales son variables importantes a modelar. Dentro de las recomendaciones en este trabajo se encuentra la relevancia de generar modelos de predicción no solo para el precio de oferta sino también al Precio de Bolsa a fin de conocer su impacto en el sector.

Con la anterior referencia de investigación se plantea realizar la predicción del Precio de Bolsa de la Energía Eléctrica en Colombia de ahora en adelante definido como PBN, y como variable económica de entrada se define el IPC para determinar su influencia en la variable de respuesta.

Dentro del mercado de la energía eléctrica se cuenta con una serie de actores dentro del sector que son quienes tienen la potestad de realizar su seguimiento y control, tal caso es la Unidad de Planeación Minero Energética que en su portal web (UPME, [s.f. b](#)) cuenta con el consolidado de indicadores de energía para:

1. Demanda.
2. Capacidad Efectiva.
3. Evolución de Embalses.
4. Evolución Precio de Bolsa.

## 5. Exportaciones e Importaciones de Energía.

De igual manera la UPME cuenta con información relativa a total del sector energético en cada uno de los componentes (Térmicas, Hidroeléctricas, Eólica, Solares), que para nuestro caso se evaluará la componente Hidroeléctrica del Sistema Interconectado de Energía (SIN).

Un segundo actor es la empresa Xm que es una filial de ISA, quien lleva los indicadores energéticos en tiempo real (Xm, [s.f. c](#)), llevando datos referente a:

1. Demanda SIN.
2. Aportes.
3. Vertimientos.
4. Importaciones.
5. Exportaciones.
6. Capacidad Efectiva.
7. Generación.
8. Precio de Bolsa Nacional (PBN).
9. Precio de Oferta.
10. Demanda no atendida.

Con base en la información proporcionada por las fuentes relacionadas anteriormente del mercado energético eléctrico, se seleccionan las variables mencionadas por la empresa Xm que guarda relación directa con los indicadores que maneja la UPME. De lo anterior se busca la relevancia que mantienen estas variables frente a cómo pueden verse asociadas al comportamiento del Precio de Bolsa Nacional, quien es la variable de respuesta a modelar.

Finalmente, la alta volatilidad que presenta el PBN asociado a diferentes factores como los son: **(1)** el mercado energético mayorista (MEM), **(2)** cambios hidrológicos, y **(3)** la demanda energética, hace necesario evaluar la relación e impacto de estas variables en los modelo de predicción frente a la variable de respuesta PBN. Dicha relación será evaluada en el apartado 5.3.

## 5.2 Definición de las variables de entrada

A continuación se realiza la definición de las variables de entrada del modelo.

### **Demanda SIN:**

Es el volumen de energía eléctrica necesaria para abastecer al Sistema Interconectado Nacional (SIN). *“Demanda de energía del SIN: se calcula con base en la generación neta de las plantas e incluye: hidráulicas, térmicas, plantas menores, cogeneradores, demanda no atendida, limitación del suministro e importaciones. Considera las plantas registradas ante el MEM. Demanda Energía SIN = Generación + Demanda No Atendida + Importaciones - Exportaciones”* (Xm, s.f. a).

Dado que en la demanda del (SIN), intervienen factores que determinan el precio de Bolsa, como lo son la generación, es relevante hacer el análisis de la relación entre estas dos variables.

### **Aportes:**

El Sistema Interconectado Nacional (SIN) se abastece de la cantidad de agua que llegan de los embalses a él. Éste volumen de agua es denominado “Aportes”. Son generados por el caudal de los diferentes ríos que se encuentran conectados al (SIN). Es de gran importancia evaluar su relación con el precio de bolsa, debido a que es una variable de orden hidrológico y como se ha mencionado anteriormente este factor ocasiona cambios muy bruscos en el (PBN).

### **Vertimientos:**

*“Los vertimientos se relacionan con la liberación al medio de sustancias líquidas que pueden afectar la calidad del suelo y/o el agua”* (COLOMBIA, s.f.). Están clasificados entre aguas industriales y domésticas. En paralelo con los Aportes, los vertimientos hacen parte del factor hidrológico, por consiguiente es relevante determinar su relación con el (PBN).

### **Importaciones:**

Durante los últimos 20 años, el Sistema Interconectado Nacional (SIN) ha presentado deficiencias para abastecer la demanda energética del país, lo que ha llevado a la necesidad de importar energía eléctrica de algunos países de la región. Por lo tanto, es considerable determinar la relación entre las compras de energía eléctrica y el comportamiento del precio de bolsa nacional.

### **Exportaciones:**

El aumento en la generación de eléctrica del país, ha permitido que en algunos periodos durante los últimos 20 años, Colombia pudiese vender energía eléctrica a otros países de la región, es decir, ha actuado como exportador de energía. Es importante analizar la relación que tienen estas ventas de energía con el (PBN).

### **Capacidad Efectiva Neta:**

Es la capacidad máxima de potencia con la que cuentan los agentes generadores. Es-

ta disponibilidad debe ser declarada en Megavatios (MW) para abastecer la demanda energética. Evaluar la relación entre la capacidad efectiva neta y el precio de bolsa nacional es de gran interés, debido a que, hace parte de la primera fase (Pre-operación) de determinación del (PBN)(CREG, s.f.).

#### **Generación:**

La generación es proceso de producción mediante el cual realiza el aprovechamiento diferentes fuentes para la elaboración de energía eléctrica. En Colombia los agentes generadores realizan este proceso productivo para la elaboración de energía Hidráulica, Biomasa, Solar, Eólica, entre otras. Como se explicó en el marco metodológico los agentes generadores son parte importante en la fijación del precio de bolsa, especialmente en la etapa (Pre-operación)(Xm, s.f. b). Por consiguiente, es conveniente evaluar la relación entre estos dos factores.

#### **Precio de Bolsa Nacional:**

*“La Bolsa de Energía es el sistema de negociación que existe en el mercado de energía eléctrica, donde los generadores hacen sus ofertas de precio y de cantidades de energía, y simplemente se cruza con la demanda, que la representan los comercializadores, y de esa forma se determinan, el precio y la cantidad que se va a entregar a la Bolsa”*(Portafolio, s.f.).

#### **Precio de Oferta:**

Como se menciona en el marco teórico el precio de oferta (PO) se genera en la fase de pre-operación del precio de bolsa. Éste es emitido por los agentes generadores, quienes realizan un reporte diario de la disponibilidad y valor por kilovatio hora (kWh) para día siguiente, este valor corresponde al (PO). Es indispensable tener en cuenta esta variable en el análisis y pronóstico del (PBN) debido a que, el máximo precio de oferta (MPO) de la última planta generadora para el abastecimiento de la demanda, es tomado como base para el cálculo del (PBN)(Gómez, 2017).

#### **Demanda No Atendida:**

Corresponde a la diferencia entre la Demanda del Sistema Interconectado Nacional (SIN) y la cantidad de energía producida por las plantas generadoras del (SIN). Esta cuantía se debe abastecer con la compra de energía a países de la región.

### **5.3 Definición del Modelo**

De acuerdo a la selección de variables que se generó en la sección 5.1 se relacionan las variables de entrada objeto de estudio, que para este caso son diez (10), junto con una variable de salida que es el PBN, para un total de once (11) variables definidas en el modelo. La base de datos contiene el histórico de información a partir de enero del año 2000 hasta agosto de año 2019, y la serie de tiempo se encuentra con periodicidad mensual, las variables numeradas del 1 al 9 y la 11 vienen representadas por kWh, y la

variable 10 en peso porcentual:

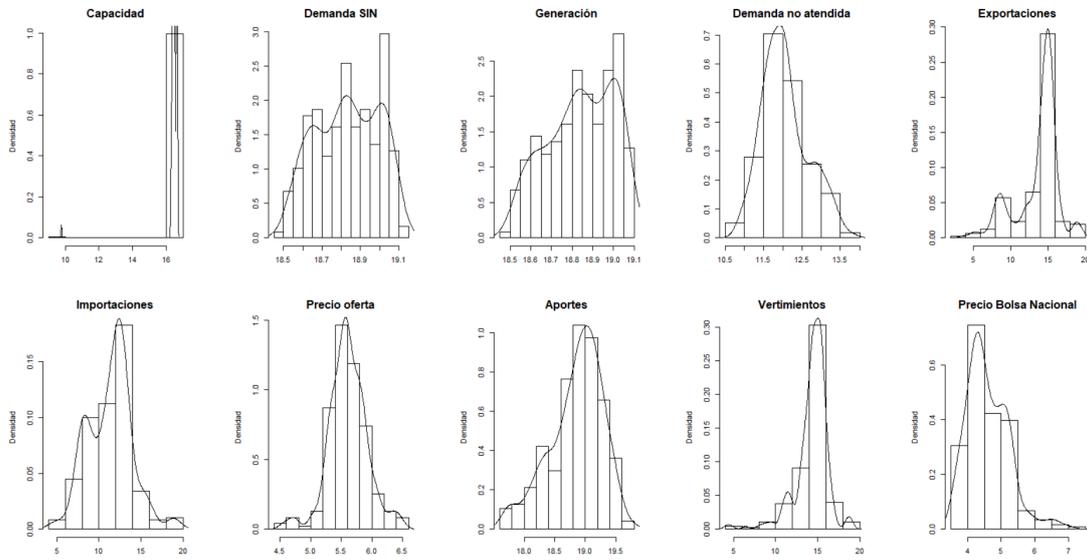
1. Capacidad efectiva
2. Demanda SIN
3. Generación
4. Demanda no atendida
5. Exportaciones
6. Importaciones
7. Precio de oferta
8. Aportes
9. Vertimientos
10. IPC – variación mensual
11. Precio de Bolsa Nacional (PBN)

Los factores entre el ítem uno (1) al cinco (5) están relacionados con flujo entre los agentes de generación, transmisión y consumo de la energía eléctrica. Con respecto a las variables (5, 6 y 7) corresponden al ejercicio comercial del (MEM). Finalmente las variables (8, 9) brindan la información sobre las fluctuaciones hidrológicas, la variable (10) guarda la información sobre variación mensual de la inflación (Banco, *s.f.*), y el PBN que es la variable de respuesta.

Para los análisis de la definición del modelo las variables ingresan con transformación por medio del *logaritmo* para facilitar su lectura.

Dentro de la base de datos seleccionada las variables *Exportaciones*, *Importaciones* y *Vertimientos* cuentan con valores perdidos para algunos meses, por lo cual, mediante el método aleatorio se realiza la imputación de estos valores. Para el caso de *Exportaciones* e *Importaciones* significa que en esos meses no se realizaron transacciones de energía, mientras que para los *Vertimientos* es debido a la no liberación de estas sustancias hídricas al medio.

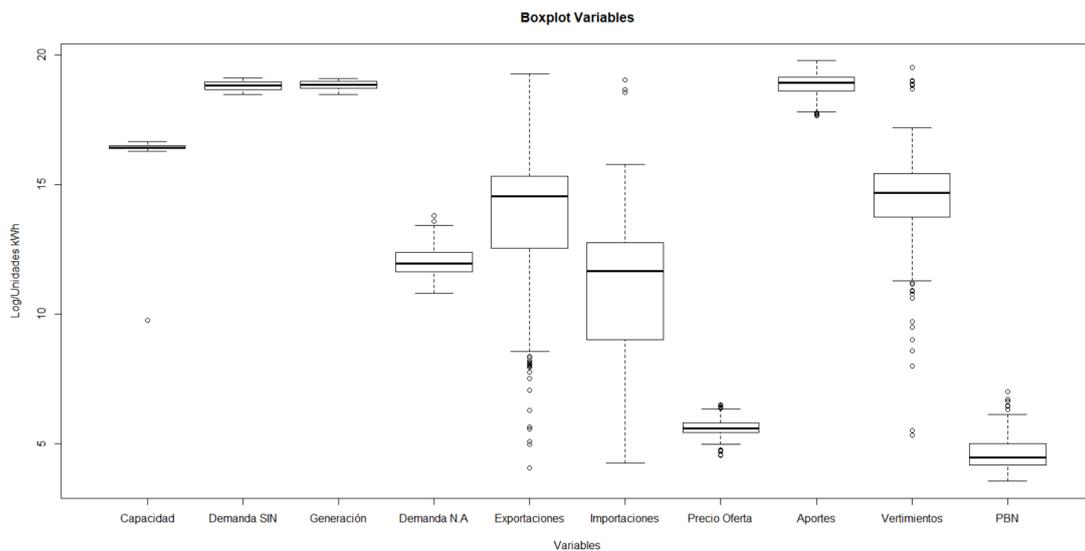
Figura 5.1: Histogramas



Fuente: Los Autores

Los gráficos de histogramas **Figura 5.1**, dan una idea de la distribución de las variables, junto con la densidad de los mismos, lo cual permite ver que varias variables tienen distribución multimodal. A primera vista el precio de bolsa nacional muestra que los datos son asimétricos hacia la derecha lo cual indica que estos pueden no tener una distribución normal.

Figura 5.2: Boxplot



Fuente: Los Autores

A fin de lograr visualizar si existen valores outliers **Figura 5.2**, se realiza un *Boxplot* de las variables donde se puede apreciar que la capacidad cuenta con un único outlier el cual está muy alejado de los demás valores.

Realizando el análisis de la base de datos frente a los outliers identificados se evidencia: la variable capacidad tiene un valor en extremo alejado del resto de los valores, el cual aparenta ser un error de digitación por lo cual se decide realizar una imputación a este datos. Para el resto de las variables los valores que están alejados son resultado del comportamiento normal del sector.

Respecto a las variables *Exportaciones*, *Importaciones* y *Vertimientos* es posible ver que tienen una alta variabilidad, y muestran asimetría hacia la izquierda como también lo evidencia la **Figura 5.1**.

### 5.3.1 Regresión Lineal Múltiple

Para hacer la evaluación y selección de las variables a incluir en los modelos, se determinó en primer lugar utilizar un modelo de regresión lineal múltiple, donde se ingresan las variables de manera contemporánea.

En el modelo de regresión lineal múltiple se utilizó el método “Stepwise” en dirección “both”, en donde se obtuvo el siguiente modelo con el mejor ajuste para la predicción del precio de Bolsa:

$$PBN = -1,655 + 5,957e^{-06}Demanda - 4,176e^{-06}Generación + 4,151e^{-01}Precio\ de\ oferta - 5,478e^{-07}Aportes + e_t$$

$$p - value = 2,2e^{-16}$$

$$R^2 = 41,08 \%$$

$$AIC = 2159,71$$

Cuadro 5.1: Pruebas de Verificación

TEST	P-Value
Durbin-Watson	< 1 %
Breusch-Pagan	< 1 %
Shapiro-Wilk	< 1 %

Fuente: Los Autores

De acuerdo al análisis de regresión lineal múltiple son seleccionadas las variables mencionadas en el modelo anterior PBN, sin embargo, utilizando este método la función “Stepwise” generó seis modelos donde la variable *Capacidad Efectiva Neta* estuvo pre-

sente. Por lo cual, se considera necesario evaluar si esta variable puede ser significativa para el modelo, por tanto se realiza un análisis de componentes principales (ACP). Las pruebas de verificación para la regresión lineal múltiple como se evidencia no superan los test de Autocorrelacion, Heterocedasticidad y Normalidad 5.1.

### 5.3.2 Análisis de Componentes Principales (ACP)

A continuación, se realiza el análisis de componentes principales de la serie de datos, donde se busca obtener un comparación de los resultados Vs la regresión lineal múltiple.

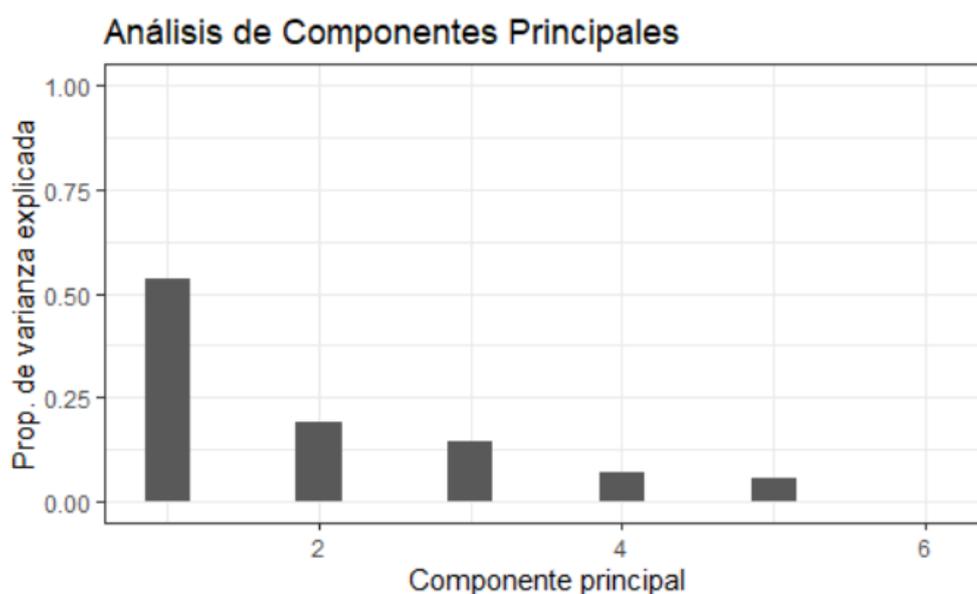
Cuadro 5.2: Resultados ACP

Variable	PC1	PC2
Capacidad Efectiva	0.429	0.140
Demanda Energía SIN	0.527	0.181
Generación	0.526	0.190
Precio Oferta	0.329	-0.115
Aportes	-0.061	0.844
Precio de Bolsa	0.387	-0.431

*Fuente: Los Autores*

En (ACP) indica que la variable (Capacidad Efectiva Neta) cuenta con un peso (0.4293) en la primera componente, por lo cual se evidencia la necesidad de ser incluida en los modelos de predicción.

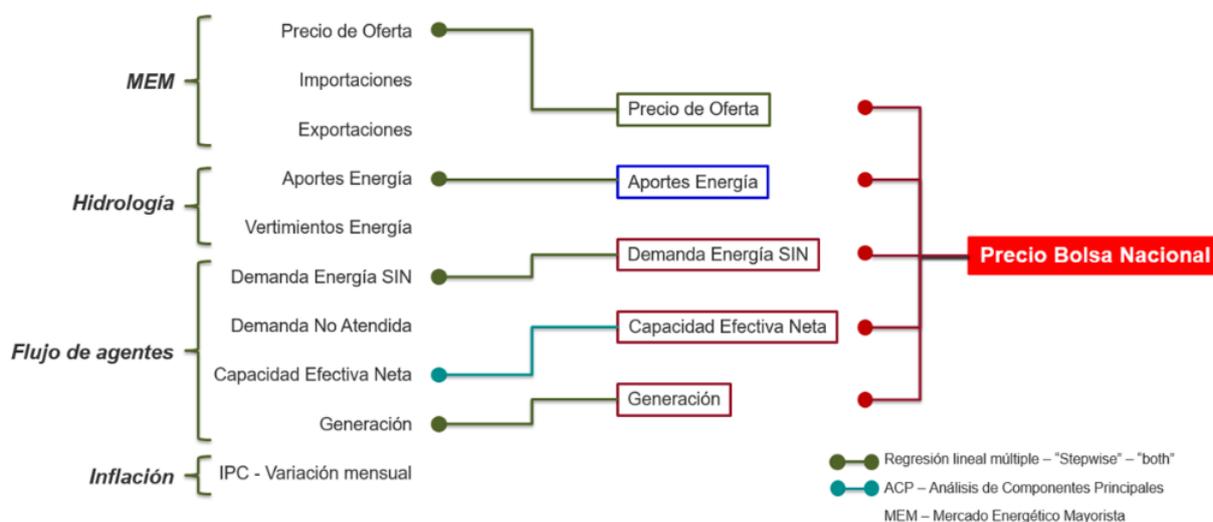
Figura 5.3: ACP



*Fuente: Los Autores*

Como se puede visualizar en la **Figura 5.3**, las componentes uno (1) y dos (2) contienen el 72.5 % de la varianza explicada de los datos.

Figura 5.4: Selección de las Variables



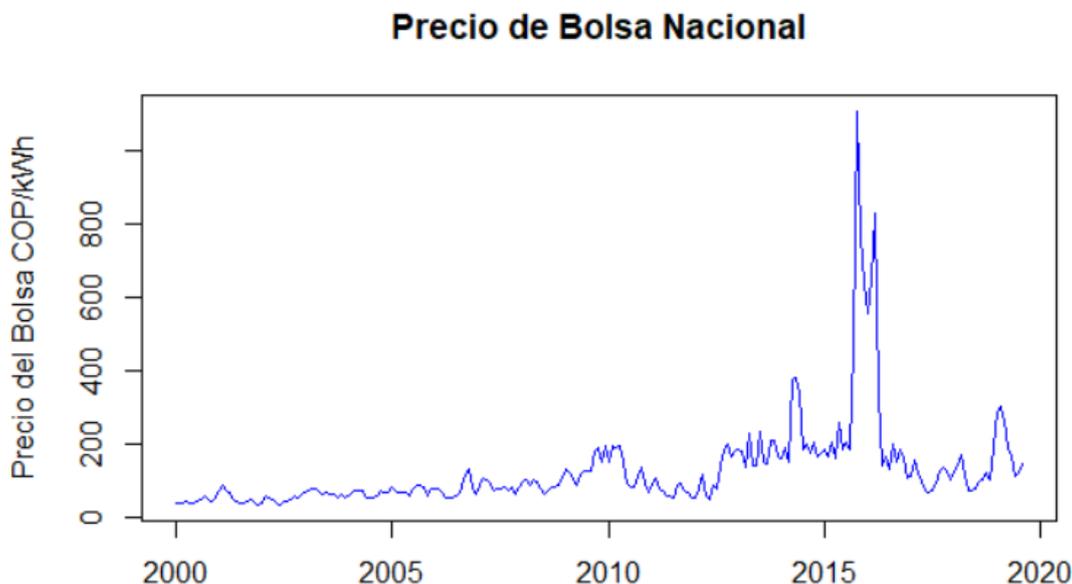
Fuente: Los Autores

En la **Figura 5.4** se resume el proceso de selección de las variables a incluir en el pronóstico del precio de bolsa de la energía eléctrica en Colombia.

## 5.4 Metodología Box & Jenkins

Una vez seleccionadas y definidas las variables de entrada a los modelos, es imperante hacer una caracterización de las mismas, identificando qué tipo estructura tienen; en tendencia, estacional o ambas.

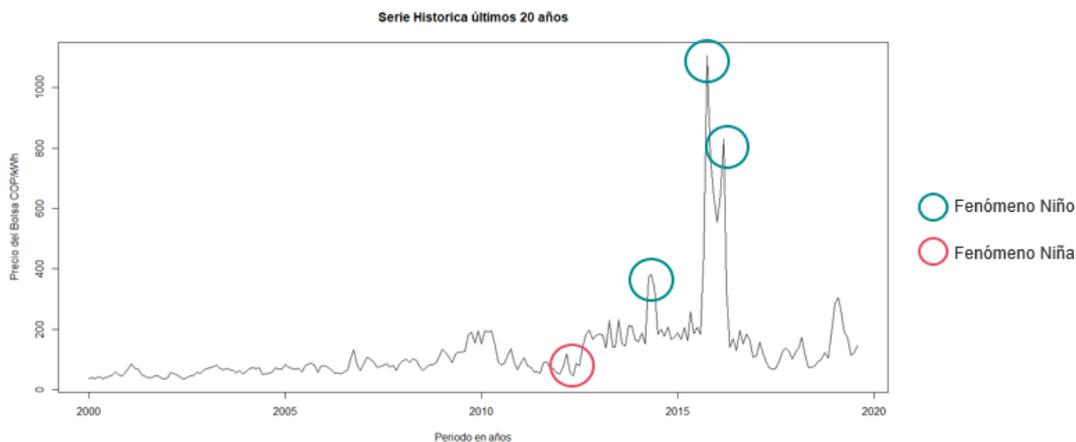
Figura 5.5: Serie de tiempo PBN



Fuente: Los Autores

En los últimos 20 años, el precio de bolsa nacional presenta una alta volatilidad, con fuertes incrementos en su valor. Estos aumentos han provocado cambios superiores al 50 % de un mes a otro en los años 2010 y 2016, los cuales han sido analizados en diferentes estudios sobre el comportamiento del precio de la energía en Colombia. Estas fluctuaciones estarían asociadas con cambios climáticos, provocados fundamentalmente por el fenómeno “El Niño”.

Figura 5.6: Serie PBN Outliers



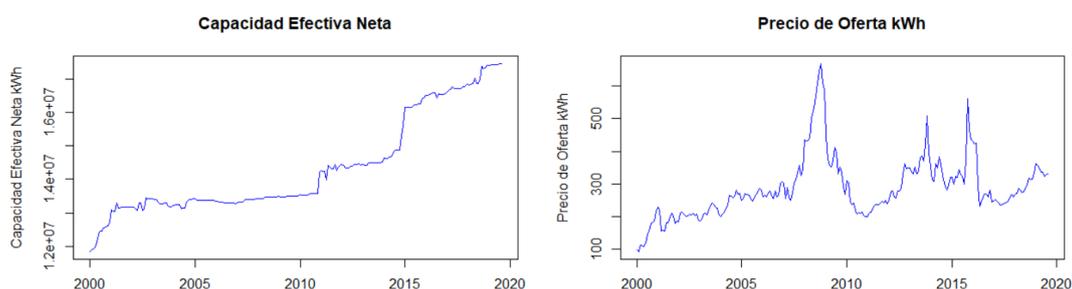
Fuente: Los Autores

“El fenómeno del niño de finales de 2009 y comienzos de 2010 fue especialmente fuerte, y se tradujo en una disminución de los embalses que obligó al sistema a hacer

una sustitución de generación hidráulica por térmica. Como las plantas térmicas no son capaces de abastecer la totalidad de la demanda de energía del país, enfrentan sobrecostos que terminan reflejados en menores transacciones de energía en bolsa y en mayores costos por kilovatio generado” (Lozano-Espitia y col., 2010).

Con respecto a la estructura que tiene la serie, el análisis descriptivo de la misma permite identificar una estructura en tendencia, en donde los precios han ido incrementando en las últimas dos décadas, y no se presencia una estructura estacional.

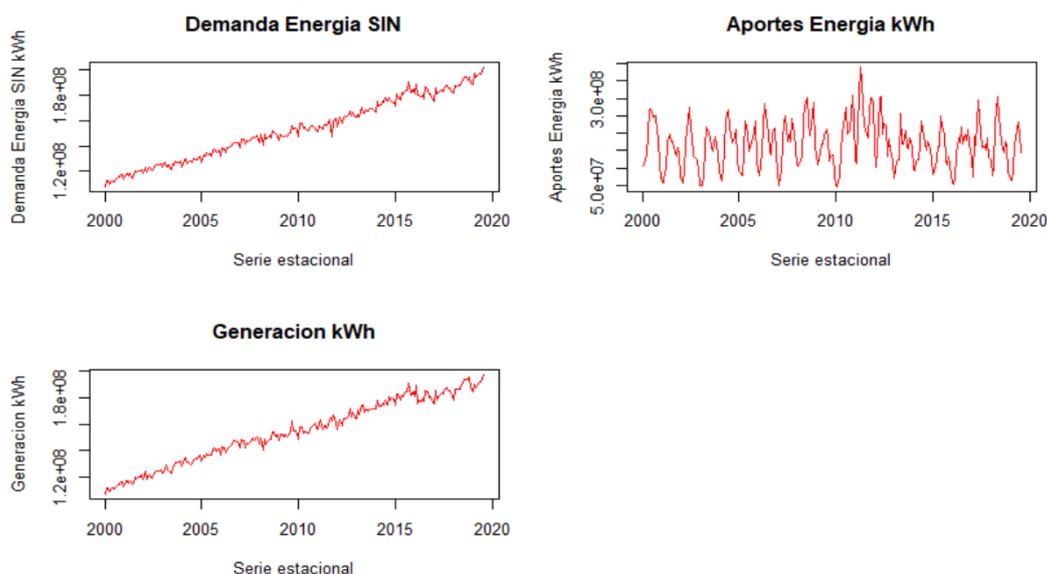
Figura 5.7: Series Estructura con Tendencia



Fuente: Los Autores

La capacidad efectiva neta y el precio de oferta de la energía eléctrica presentan estructuras con tendencia creciente en los últimos 20 años, no se evidencian estructuras estacionales 5.7.

Figura 5.8: Series Estructura Estacional



Fuente: Los Autores

La Demanda del Sistema Interconectado Nacional (SIN) y la Generación de energía eléctrica, comparten un comportamiento muy similar, en donde se evidencian estructuras

en tendencia creciente y estacionales. Con respecto a los Aportes, esta variable reporta una estructura estacional 5.8.

### 5.4.1 Arimax

Debido a que la base de datos cuenta con cinco (5) variables de entrada y una (1) variable de respuesta se define realizar la identificación de un modelo de serie de tiempo multivariado, y aplicando la metodología Box & Jenkins se define:

#### Etapa 1 Identificación:

Con base en la caracterización de las variables, el análisis gráfico de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) y la haciendo la prueba de raíz unitaria *Dickey-Fuller* a cada variable, los datos de algunas variables deben ser incluidas en los modelos con rezagos, es decir, aplicando diferencias en la serie.

Estas diferencias se aplican a las variables con los siguientes objetivos. **(1)** Eliminar la estructura en tendencia y convertirlas en estructura estacionaria, de forma, que cumplan la configuración requerida por los modelos ARIMAX y VAR. **(2)** Contemplar los efectos que tienen los cambios de algunas variables en meses posteriores al precio de bolsa, es decir, no se reflejan de forma contemporánea.

En aras de hacer más práctico el análisis de los modelos, y por ser el precio de bolsa una serie financiera, se hace la transformación logarítmica a los datos. Es importante aclarar que las conclusiones de los pronósticos corresponden a los valores originales serie.

Cuadro 5.3: Prueba Dickey-Fuller

Variable	$\tau$ 5 %	P-Value
Precio de Bolsa	-3.43	3.05 %
Capacidad Efectiva	-3.43	3.86 %
Demanda Energía SIN	-3.43	8.96 %
Generación	-3.43	10.36 %
Precio Oferta	-3.43	4.63 %
Aportes	-3.43	2.80 %

*Fuente: Los Autores*

Como se evidencia en el **Cuadro 5.3** todas las variables tienen un P-Value superior a TAO que es del 5% por lo tanto requieren de una diferencia.

Cuadro 5.4: Rezagos de las Variables

Variable	Rezago
Precio de Bolsa	14 meses
Capacidad Efectiva	1 mes
Demanda Energía SIN	12 meses
Generación	12 meses
Precio Oferta	Contemporánea
Aportes	1 mes

Fuente: Los Autores

Esta misma prueba se realizó para dos diferencias en donde el resultado para todas las variables indica que no hay evidencia estadística para la aplicación de una segunda diferencia. Así mismo, se muestran los rezagos de ingreso al modelo **Cuadro 5.4**, los retardos de cada variable ingresan de igual forma a la red neuronal.

### Etapa 2 Estimación:

En la elaboración del modelo de pronóstico del PBN, se tomo la serie de datos con un total de 236 observaciones y se dividió en dos partes, definidas como *In-Sample* y *Out-Sample*. La primera parte cuenta con 212 observaciones y la segunda con 24 observaciones. Lo anterior con el fin de modelar la parte *In-Sample* y contrastar el pronóstico a 24 casos, es decir, 2 años contra el *Out-Sample* para validar el nivel de ajuste del pronóstico a la serie original de salida.

Para realizar la estimación del modelo se hizo necesario evaluar diferentes estructuras tanto en la parte **AR**, como en la parte **MA**, con el fin de encontrar el modelo con el mejor ajuste a la serie de datos. Se recuerda que la serie del PBN no es estacionaria, por lo cual no se modela dicha estructura. Por lo tanto, se plantea un modelo que sea apropiado a la estructura de la serie temporal del PBN: **ARIMAX**.

Cuadro 5.5: Estimación del Modelo PBN

Modelo	Parámetros	BIC	R.B
ARIMAX (2, 1, 9)	AR(1,2) MA(1, 9)	-6.4923	SI
ARIMAX (2, 1, 9)	AR(1,2) MA(1,2, 3, 8, 9)	-10.526	SI
ARIMAX (2, 1, 14)	AR(2) MA(1, 2, 3, 8, 14, 9)	-12.922	NO
ARIMAX (0, 1, 14)	MA(1, 7, 14)	-22.062	SI
ARIMAX (0, 1, 14)	MA(1, 14)	-25.069	SI

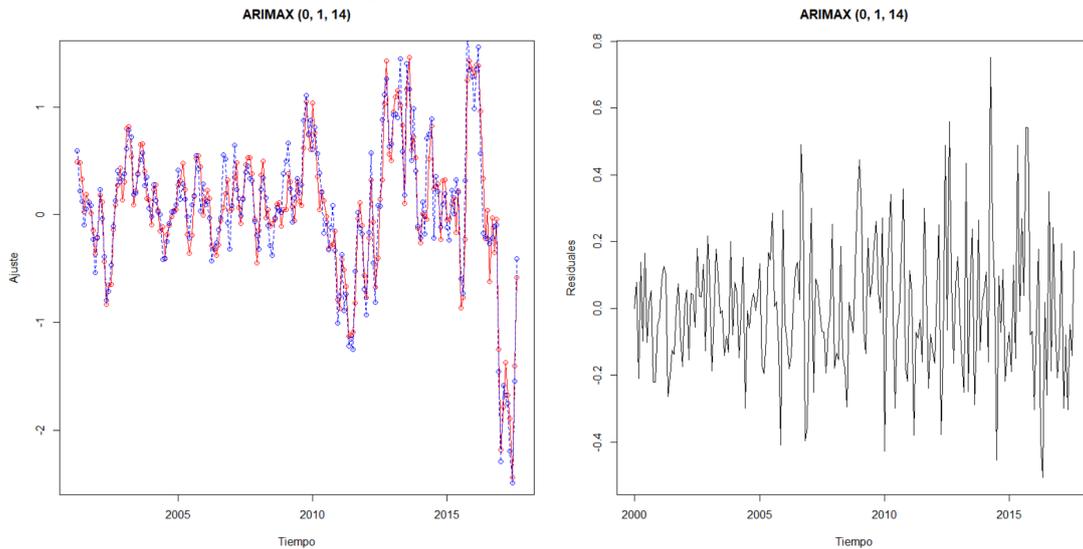
Fuente: Los Autores

Con base en el **cuadro 5.5** se determina que el mejor modelo con BIC de  $-25,069$  y el cual cuenta con estructura **ARIMAX (0, 1, 14)** en donde solo se modelan los parámetros **MA(1, 14)**.

### Etapa 3 Verificación:

Para realizar la verificación del modelo se evalúa el ajuste de los datos, los gráficos ACF y PACF, junto con los residuales.

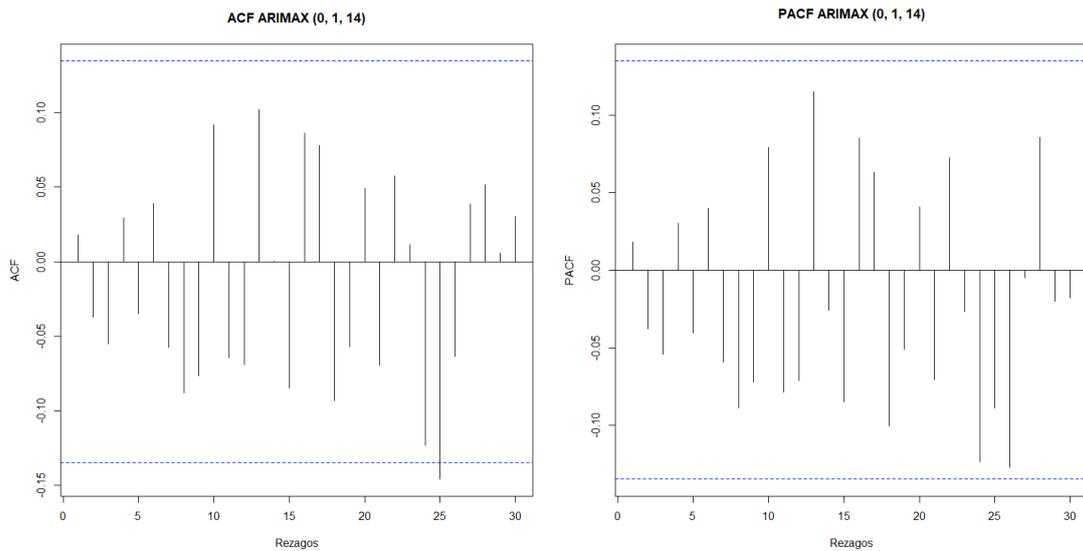
Figura 5.9: Ajuste De los Datos



Fuente: Los Autores

Los datos del modelo reporta un buen ajuste de acuerdo a la estructura identificada anteriormente, modelando correctamente los residuales 5.9.

Figura 5.10: ACF

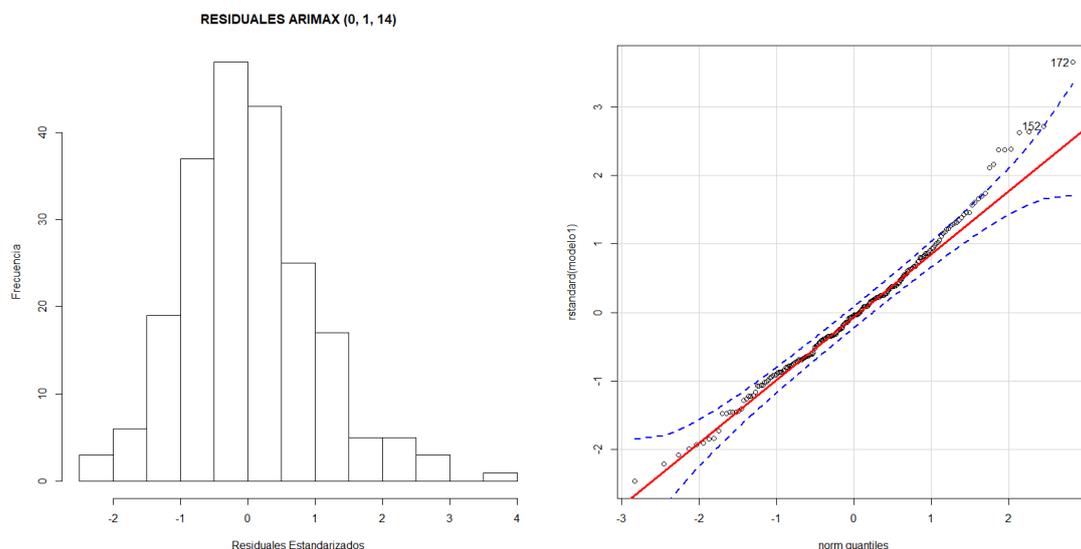


Fuente: Los Autores

Analizando la estructura del ACF y PACF, se puede visualizar que los rezagos a 24 meses están dentro de las bandas de confianza, para el caso del rezago 25 la barra logra

salir, sin embargo, es un modelo que se ajusta a largo plazo. En el PACF todos los rezagos ingresan dentro de las bandas de confianza 5.10.

Figura 5.11: Residuales



Fuente: Los Autores

Haciendo el análisis descriptivo de los residuales tanto en el histograma como en el gráfico de normalidad, el modelo no evidencia tener una distribución normal, debido a que la serie de datos presenta outliers, que se identificaron en 5.11.

Cuadro 5.6: Pruebas de Verificación

TEST	P-Value
Ljung-Box	66 %
Jarque Bera	< 1 %
Shapiro-Wilk	1.6 %

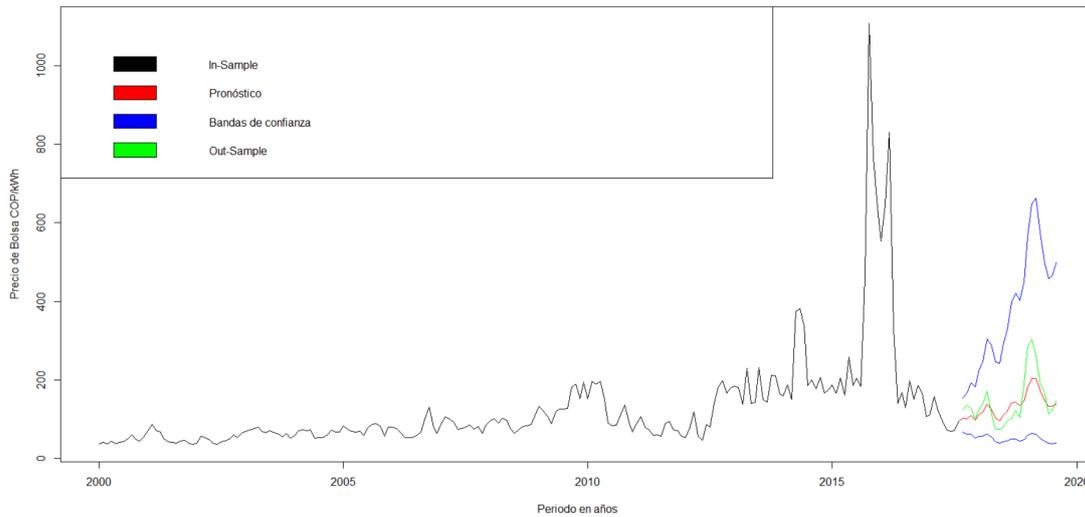
Fuente: Los Autores

La prueba de *Ljung-Box* muestra ausencia de autocorrelación serial. De acuerdo a las pruebas *Jarque Bera* y *Shapiro Wilk*, se tiene suficiente evidencia estadística para concluir que no existe normalidad en los datos. Lo cual es común debido a que la serie proviene del sector económico y financiero, en los cuales los datos con regularidad no se comportan con distribución normal.

#### Etapa 4 Pronóstico:

A continuación, se presenta el resultado del pronóstico realizado con el modelo seleccionado **ARIMAX (0, 1, 14)** y con la identificación de parámetros **MA(1, 14)** 5.5.

Figura 5.12: Pronóstico PBN ARIMAX



Fuente: Los Autores

Como se puede apreciar en la **Figura 5.12**, el pronóstico logra modelar los datos Out-Sample. En el primer periodo de pronóstico de (*septiembre de 2017 hasta abril de 2018*) lograr ajustarse con mucha cercanía a los datos observados. Sin embargo, al inicio del año 2019 en los meses de (*enero a marzo*) el PBN reporta un incremento en sus valores, en donde el pronóstico del modelo replica su comportamiento, pero no logra alcanzar los picos presentados. Con respecto al ajuste del pronóstico en las bandas de confianza con dos desviaciones estándar, el resultado es positivo dado que tanto el pronóstico como el Out-Sample están dentro de las mismas.

Finalmente se define la ecuación del modelo:

$$Z_t = \theta_1 e_{t-1} + \theta_{14} e_{t-14} + \beta_1 \log(\text{Capacidad}_{t1}) + \beta_2 \log(\text{Demanda}_{t12}) + \beta_3 \log(\text{Generación}_{t12}) + \beta_4 \log(\text{Precio Oferta}_t) + \beta_5 \log(\text{Aportes}_{t1}) + e_t$$

Donde:

$$\begin{aligned}
 Z_t &= \nabla \log(\text{Precio de Bolsa}_t) \\
 \widehat{\theta}_1 &= -0,3320(tt : -4,85) \\
 \widehat{\theta}_{14} &= -0,1355(tt : -2,21) \\
 \widehat{\beta}_{1_1} &= 0,32 \%(tt : 6,69) \\
 \widehat{\beta}_{2_{12}} &= -3,04 \%(tt : -2,23) \\
 \widehat{\beta}_{3_{12}} &= 3,05 \%(tt : -2,23) \\
 \widehat{\beta}_{4_t} &= 1,03 \%(tt : 7,15) \\
 \widehat{\beta}_{5_1} &= -0,28 \%(tt : -6,91) \\
 \widehat{\sigma}_e^2 &= 4,23 \% \\
 BIC &= -25,06912
 \end{aligned}$$

## 5.4.2 VAR

### Etapa 1 Identificación:

Como se mencionó en el análisis del **Cuadro 5.3**, en la que se realizó la prueba de raíz unitaria para cada una de las variables, en la que se evidencio que es necesario aplicar una diferencia a las variables exógenas y la variable endógena. Dado que los modelo VAR requieren que todas la variables tengas el mismo orden integración  $\sim \mathbf{I}(1)$ , dichas variables son ingresadas bajo este parámetro.

Para la identificación de los rezagos se parte del análisis realizado al inicio del inciso 5.4 en el cual se menciona que las variables Demanda, Aportes y Generación tienen estructura estacional. Por consiguiente el modelo VAR debe incluir rezagos con periodos a 12 meses.

### Etapa 2 Estimación:

Con el objetivo identificar una estructura apropiada para el modelo que mejor presente a la serie de datos se utiliza la función VAR, para capturar la estructura no estacionaria de las series se incorporan componentes de tendencia y ciclo, en la cual se obtiene como resultado un valor BIC quien arroja un orden de retraso igual a  $P = 3$ .

$$BIC = -3952,054 \tag{5.0}$$

Adicional, para adecuar la estructura del modelo, los parámetros de las variables que sean menores a 2 deben ser excluidos, que se define como el nivel de significancia para el modelo. Con lo cual se obtiene un nuevo BIC.

$$BIC = -4487,952 \tag{5.0}$$

La selección del BIC se basa en aquel que tenga un valor menor, donde el modelo excluye los parámetros inferiores al nivel de significancia estipulado.

Posterior a lo anterior se realizan las pruebas de causalidad al modelo.

Cuadro 5.7: Prueba de causalidad

Variable	P-Value	Causalidad
Precio de Bolsa	0.05873	SI
Capacidad Efectiva	0.8314	SI
Demanda Energía SIN	0.01176	NO
Generación	0.0006372	NO
Precio Oferta	0.05445	SI
Aportes	0.04288	NO

*Fuente: Los Autores*

De acuerdo a los resultados de las pruebas de causalidad a cada una de las variables, hay suficiente evidencia estadística para afirmar que el PBN, la capacidad efectiva, y el precio de oferta son variables que pueden explicar el comportamiento de las variables incluidas en el modelo. Caso contrario, con la demanda, generación y los aportes.

### Etapa 3 Verificación:

Para esta etapa, se realizan las pruebas de diagnostico al modelo, a fin de determinar su grado de ajuste a la estructura definida.

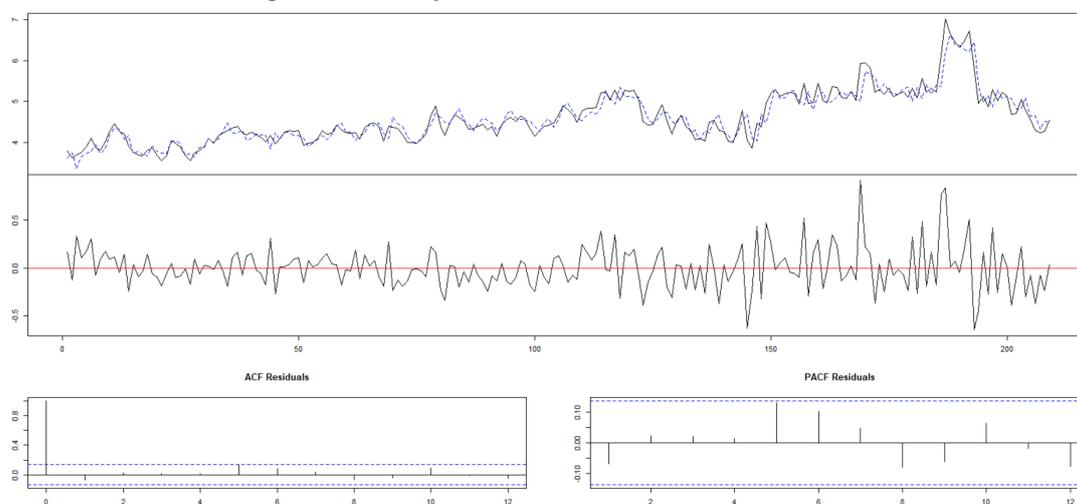
Cuadro 5.8: Pruebas de Verificación

TEST	P-Value
Portmanteau	6.4 %
ARCH	9.5 %
Kurtosis	< 1 %

*Fuente: Los Autores*

Los resultados del test de correlación serial (*Portmanteau*), permiten establecer que existe ausencia de correlación serial multivariada, la prueba de Heterocedasticidad (*ARCH*) muestra que los datos tienen estructura Homocedastica, finalmente la prueba de normalidad (*Kurtosis*) de igual manera que en el modelo **ARIMAX** corrobora ausencia de normalidad en los datos.

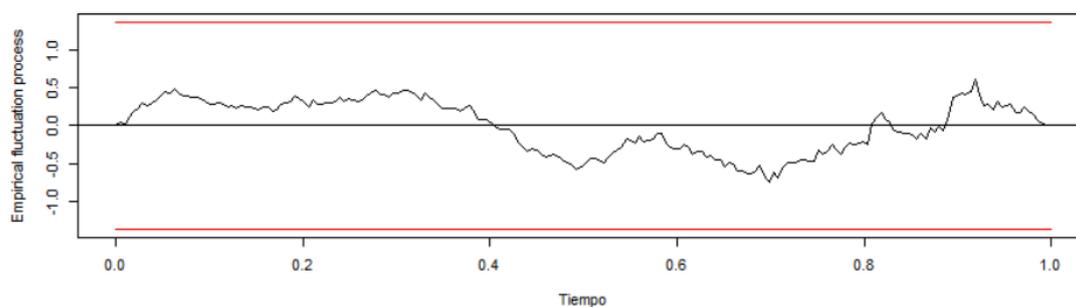
Figura 5.13: Ajuste de los residuales del PBN



*Fuente: Los Autores*

Como se puede apreciar en la **Figura 5.13**, los datos cuentan con un buen ajuste de los residuales y así mismo, los gráficos de ACF y PACF están dentro de las bandas de confianza.

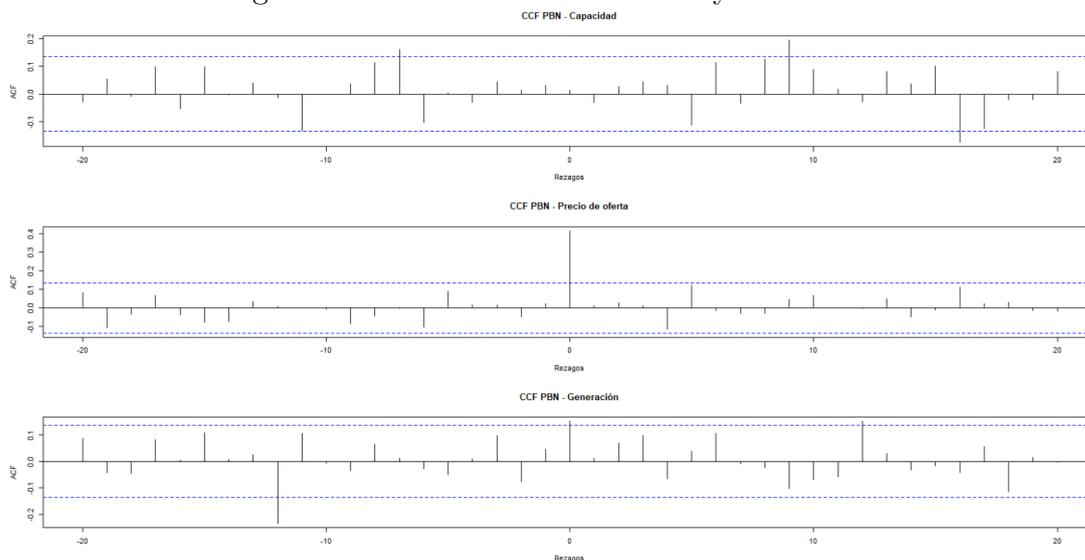
Figura 5.14: Test CUSUM PBN



*Fuente: Los Autores*

La **Figura 5.14** Test CUSUM, muestra la sensibilidad del modelo a la estimación paramétrica en diferentes ventanas de tiempo. Mientras este dentro de las bandas la estimación paramétrica es estable para la variable del PBN.

Figura 5.15: Función de covarianza y correlación



Fuente: Los Autores

Como se puede evidenciar en la **Figura 5.15** la capacidad presenta una correlación cruzada frente al PBN, en la cual se evidencia un choque en los rezagos 9, 11 en el pasado y 9 en el futuro, donde los valores del precio de bolsa presentan un cambio. Con respecto a la correlación cruzada frente al precio de oferta no se evidencia que los choques tenga influencia en el futuro o en el pasado, dado que el comportamiento de estas dos variables es contemporáneo. Para el caso de la generación la correlación cruzada, muestra que los choques afectan los rezagos cada doce meses, en donde el incremento en el periodo  $n$  genera fluctuaciones en los periodos siguientes.

**Ecuación de largo plazo:**

A fin de poder definir la ecuación, es necesario conocer la relación de cointegración del modelo.

Cuadro 5.9: test de cointegración de Johansen

<b>r</b>	<b>test</b>	<b>10pct</b>	<b>5pct</b>	<b>1pct</b>
$r \leq 5$	0.17	6.50	8.18	11.65
$r \leq 4$	4.62	15.66	17.95	23.52
$r \leq 3$	19.03	28.71	31.52	37.22
$r \leq 2$	41.12	45.23	48.28	55.43
$r \leq 1$	67.55	66.49	70.60	78.87
$r = 0$	112.66	85.18	90.39	104.20

Fuente: Los Autores

Como se puede apreciar en el **Cuadro 5.9**, quien es un resultante de la prueba *Johansen Procedure*, las variables del modelo reportan una relación de cointegración lo cual se

traduce en que a partir de una única ecuación se puede obtener información en el largo plazo para cada una de las variables. En seguida se relaciona la ecuación de largo de PBN.

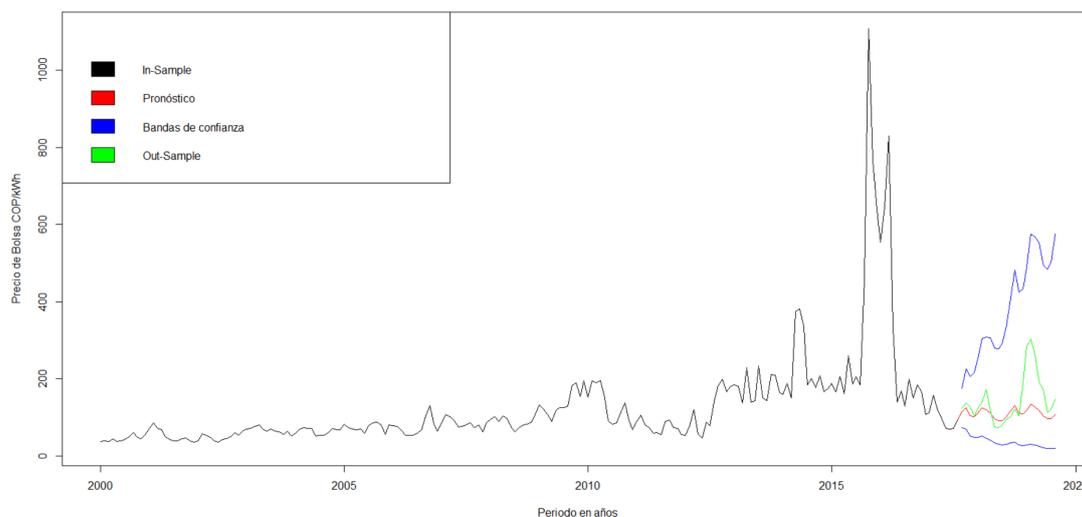
$$EC_t = 0 = z_{1,t} + 6,66z_{2,t} - 11,10z_{3,t} + 1,19z_{4,t} - 0,46z_{5,t} + 2,40z_{6,t} I(0)$$

$$z_{4,t} = -0,8403z_{1,t} - 5,59z_{2,t} + 9,32z_{3,t} + 9,32z_{4,t} + 0,3865z_{5,t}$$

#### Etapa 4 Pronóstico:

Finalizando, se presenta el pronóstico con el modelo VAR desarrollado.

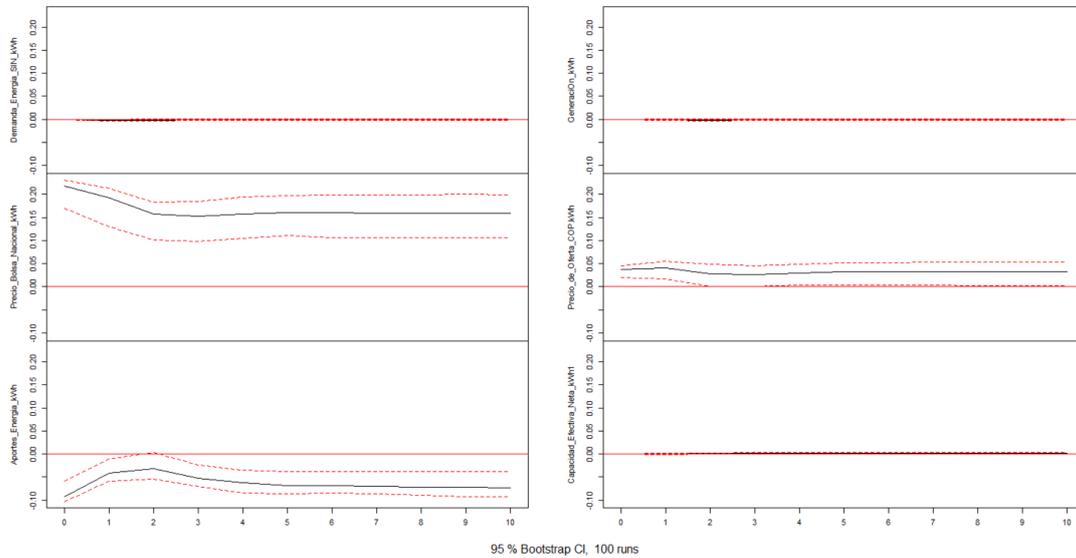
Figura 5.16: Pronóstico PBN VAR



Fuente: Los Autores

El pronóstico muestra para los primeros periodos un grado de ajuste bueno, posteriormente el pronostico suaviza la serie de datos y no logra ajustarse a los picos del *Out-Sample*. De igual manera, este pronóstico converge a la media de los datos y no logra replicar los comportamientos crecientes del PBN.

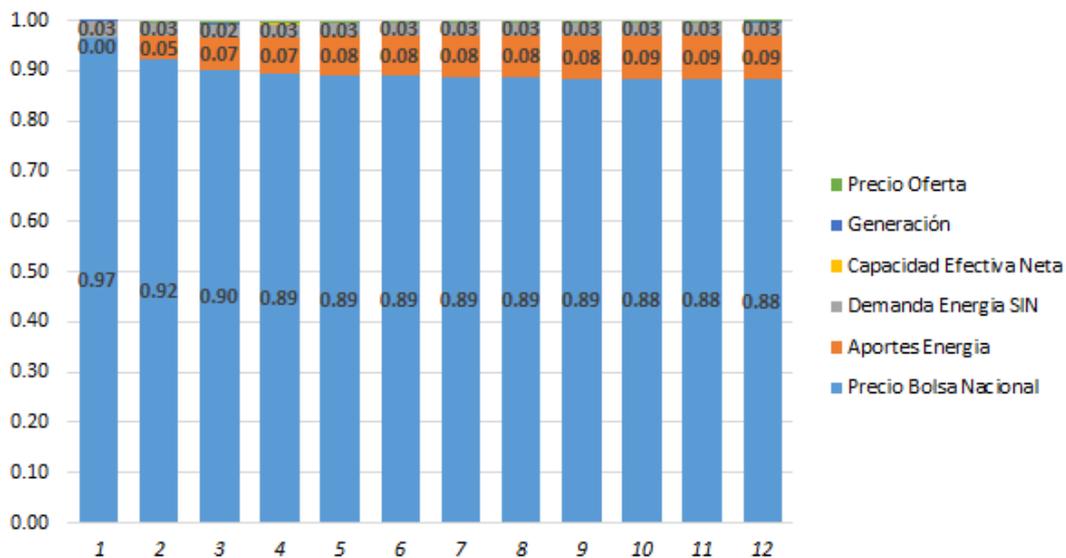
Figura 5.17: Impulso de respuesta



Fuente: Los Autores

Con respecto a la **Figura 5.17**, la presencia de un choque en el PBN las variables demanda, generación y capacidad no tienen cambios en el corto y mediano plazo. Las variables aportes y precio de oferta ante un choque en el PBN generará cambios. Para el caso de aportes el PBN en el periodo contemporáneo tiene una caída del 10% a partir de este periodo hasta el periodo 2 esta caída va disminuyendo paulatinamente, en adelante el comportamiento de los aportes es estable en un -7%, y este está dentro de las bandas de confianza. Con respecto al precio de oferta un choque del PBN genera incrementos cercanos al 4% los cuales se mantienen estables en este valor.

Figura 5.18: Descomposición de la varianza



Fuente: Los Autores

En la **Figura 5.18**, indica que el PBN en el periodo inmediato de generación no esta afectado por ninguna variable diferente a el mismo, solo el 3% de la varianza en este periodo esta depende de la demanda. sin embargo, a partir del periodo dos la participación de la variabilidad del PBN empieza a distribuirse de manera diferente, en donde los aportes comienzan a tomar cierta relevancia la cual va incrementándose paulatinamente hasta tomar el 9% de la variabilidad luego del periodo diez. Esto indica que la varianza en un periodo de doce meses depende de los aportes y de la demanda en un 9% y un 3% respectivamente.

La participación de las demás variables diferentes a la demanda y a los aportes no contribuyen en términos de variabilidad del modelo de pronostico del PBN.

## 5.5 Red Neuronal

Para este caso se trabajara con una red neuronal perceptron multicapa (MLP), el paquete a utilizar es Amore en R. *"Para el perceptron multicapas, el método de optimización no lineal más utilizado es el steepest descent sobre la función E. De esta forma, cada parámetro  $w$  de la red es modificado para cada patrón de entrada  $n$ "* (Santana, 2006). La selección de este tipo de red es debido a su capacidad como aproximador universal, facilidad en su uso y su capacidad de resolver problemas.

Para la construcción de la red se requiere que las neuronas estén agrupadas en diferentes capas de diferentes niveles, donde se tienen principalmente una capa de entrada o variables de ingreso, una segunda capa oculta y una tercera que es la salida de la red. Cada una de las capas esta conectada con la siguiente capa alimentándose hacia adelante, esto se desarrolla bajo una activación de propagación para ello se utiliza la función  $f$  *función de activación*, donde las mas utilizadas son la sigmoideal y la tangente hiperbólica.

Función sigmoideal:

$$f_{sigm}(x) = \frac{1}{1 + e^{-x}} \quad (5.0)$$

Función tangente hiperbólica:

$$f_{thip}(x) = \frac{1 - e^{-x}}{1 + e^{-x}} \quad (5.0)$$

Para definir el diseño de la arquitectura de la red, es necesario contar con el numero de neuronas y el numero de capas, las cuales son definidas por el modelo que se viene desarrollando, razón por la cual, se define una primera capa con 18 neuronas de entrada, tres capas ocultas con 12, 12 y 10 respectivamente neuronas y una capa de salida de una sola neurona.

El número de neuronas ocultas no es el parámetro principal, pues en este caso la

arquitectura es la que puede resolver el problema, que en nuestro caso son los rezagos quienes permiten la construcción del modelo. Adicional, es necesario que la red cuente con un entrenamiento, ya que el objetivo es que la salida de la red sea lo más aproximada posible a la serie original.

El algoritmo de retro-propagación es la regla para el aprendizaje de la red, que para el caso de un MLP esta dado por un algoritmo de aprendizaje supervisado, por lo cual durante el entrenamiento se pueden modificar los parámetros para que la salida estimada de la red sea lo mas parecida a la salida esperada. Por lo cual el algoritmo se formula como un problema de minimización dado por:

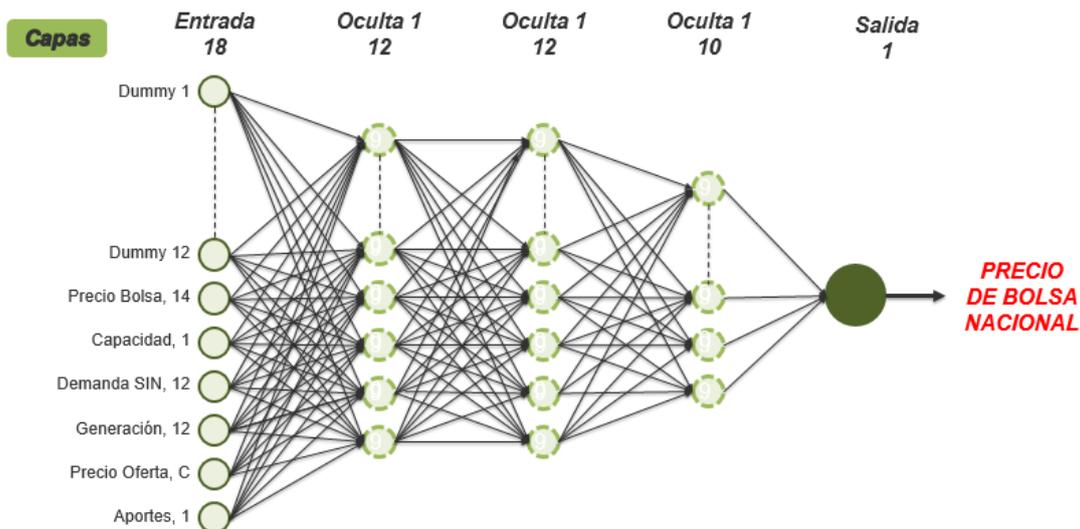
$$\text{Min}_w E \quad (5.0)$$

Siendo  $w$  los parámetros de la red y  $E$ , la función del error que compara las salidas de la red Vs la salida esperada. por lo cual, se puede concluir que busca encontrar un mínimo del error en la función definida.

$$E = \frac{1}{N} \sum_{n=1}^N e(n) \quad (5.0)$$

El entrenamiento de la red *In-Sample* cuenta con una serie de 212 valores iniciando con 18 neuronas de entrada; donde 6 son variables propias de la serie y 12 variables dummy que identifican cada mes del año, multiplicando estos valores se obtiene un total de 3816 observaciones, la red tendrá una única salida con un total de 24 observaciones.

Figura 5.19: Arquitectura Neuronal MLP



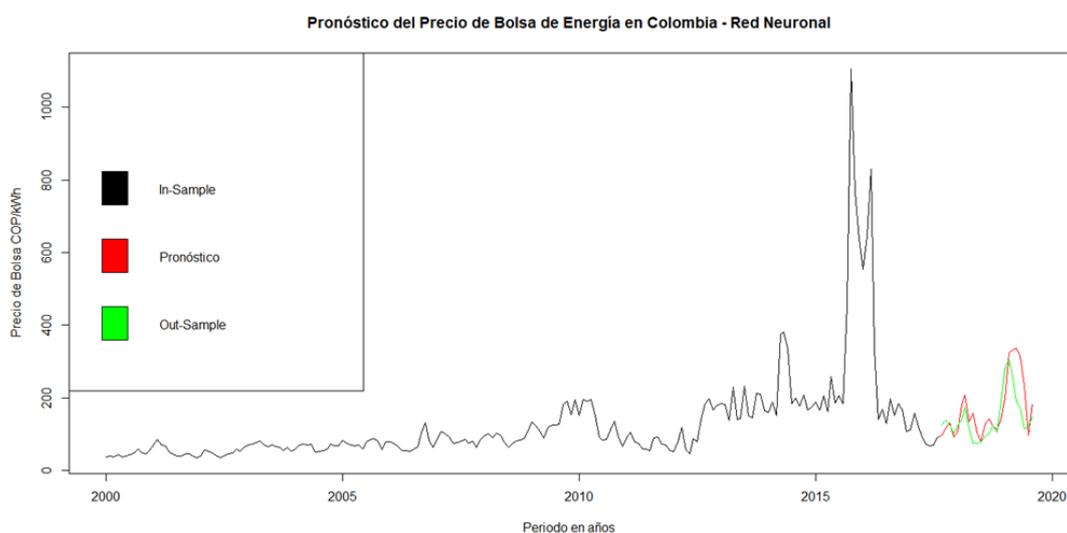
Fuente: Los Autores

La **Figura 5.19** muestra la arquitectura con el mejor desempeño de la red que se propone para realizar el pronóstico del PBN. En la capa de entrada se muestran las 12

dummy, y las variables de entrada con sus respectivos rezagos.

Para el pronóstico de la red, el conjunto de entrenamiento que equivale a 3816 observaciones fue entrenada bajo una tasa de aprendizaje de 0.01, con las entradas de capas definidas en 5.19, el conjunto de validación que se describió con 24 observaciones se definió como la función para evaluar el error tras la salida de la red.

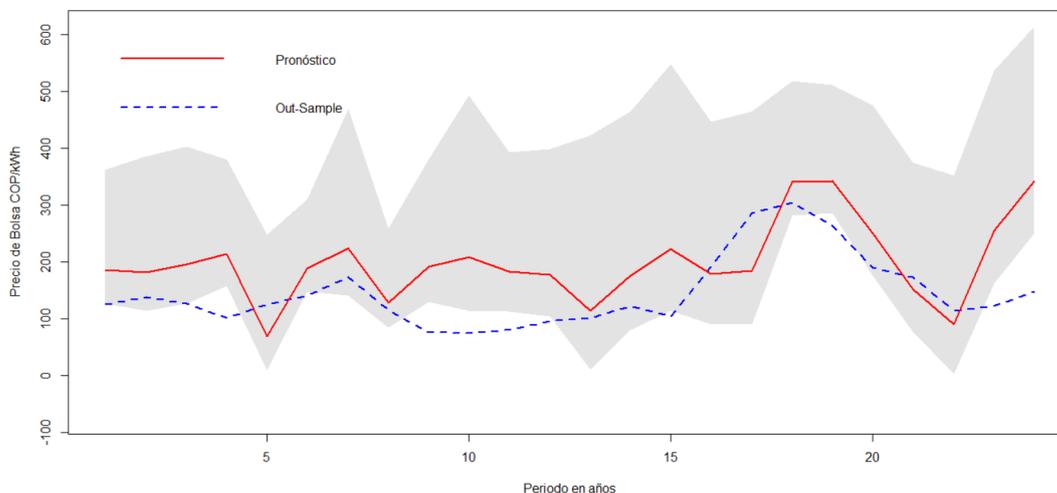
Figura 5.20: Pronóstico PBN Red Neuronal



Fuente: Los Autores

Como se puede ver en la **Figura 5.20**, el conjunto de entrenamiento se define como la serie de datos *In-Sample* (Color Negro), el conjunto de validación como *Out-Sample* (Color Verde), y la salida de la red es el pronóstico del PBN (Color Rojo). Según los resultados de salida, esta logra perseguir de forma adecuada el *Out-Sample*, llegando a modelar los picos de la serie.

Figura 5.21: Bandas de Confianza pronóstico PBN Red Neuronal



*Fuente: Los Autores*

En la **Figura 5.21** se generan las bandas de confianza para el pronóstico de la red neural con el siguiente método; (1) Se crean diez réplicas de la red con la misma arquitectura y parámetros definidos para la predicción del PBN.(2) Con el resultado de estas replicas se generan las bandas de confianza a tres desviaciones estándar, tanto para el límite superior como inferior. (3) Es incluida la serie de datos original definida como muestra Out-Sample, para evaluar el ajuste de la red.

## 5.6 Discusión de resultados

De acuerdo a los modelos definidos se logró evidenciar que los tres modelos logran ajustarse a la serie de los datos, sin embargo, a fin de tomar una decisión se muestra los errores de cada modelo:

Cuadro 5.10: Resultados de los errores de los modelos

	<b>ARIMAX</b>	<b>VAR</b>	<b>NNAR</b>
RMSE	39.92	65.22	60.87

*Fuente: Los Autores*

Como se pudo apreciar en cada una de las secciones dentro del desarrollo de la propuesta del modelo de pronóstico, se generaron aproximaciones a la serie de datos Out-Sample. Los resultados son los siguientes:

- Para la selección de las variables fue posible identificar un modelo de regresión lineal múltiple donde se definieron las variables que explican al modelo con respecto

al precios de bolsa nacional (PBN). De forma adicional se planteó un análisis de componentes principales (ACP) en el cual se pudo incluir una nueva variable de interés para la predicción del modelo.

- Respecto al modelo ARIMAX se identificó dentro de seis estructuras una que permite modelar de mejor manera la serie de datos. Los resultados del pronóstico son satisfactorios logran superar las pruebas de verificación. El ajuste a los datos que muestra el pronóstico es que logra perseguir la serie de datos, y hasta cierto punto lograr alcanzar los picos del out-Sample.
- El modelo VAR se definió con una estructura de 12 rezagos debido a que ciertas variables cuentan con estructura estacional. Así mismo, se logró evidenciar que supera las pruebas de verificación. Respecto a la salida del pronóstico, este no logra modelar los picos de la serie de datos del Out-Sample. De forma positiva el VAR muestra los choques que genera el PBN respecto a las demás variables y su relación de causalidad. Lo que lo hace un buen modelo a tener en cuenta por la información que entrega.
- Por último, la arquitectura de la red neuronal género una buena relación frente al conjunto de datos de entrenamiento Vs el de verificación. Frente al pronóstico se evidencia que logra perseguir la serie de datos e incluso alcanza a modelar los picos.

Gráficamente se puede evidenciar que el modelo ARIMAX tiene el mejor desempeño, seguido de la red neuronal y por último el modelo VAR. Comparando con los errores de los modelos el modelo seleccionado es el ARIMAX con error cuadrático medio de 39.92 pesos kWh.



## Capítulo 6

# Conclusiones y Recomendaciones

### 6.1 Conclusiones

- Se logró realizar la identificación de los factores que influyen en el precio de bolsa nacional de la energía eléctrica en Colombia, lo cuales están asociados con los diferentes agentes que intervienen en su determinación, como lo son variables hidrológicas, fenómenos climáticos y componentes económicos. Para esto se utilizó la regresión lineal múltiple y análisis de componentes principales, obteniendo las variables de entrada para el pronóstico de los modelos con metodología Box & Jenkins y la Red Neuronal.
- La selección de los modelos según la metodología Box & Jenkins se realizó basado en la series de datos multivariados, junto con la estructura que permitió definir un modelo ARIMAX y un modelo VAR, para el análisis del pronóstico del precio de bolsa de la energía.

Como resultado de los análisis se encontró que el precio de bolsa nacional no tiene una estructura estacional, lo cual es congruente debido a la naturaleza del sector económico. De igual manera las variables que pueden generar choques en la variable de respuesta son los aportes y la demanda, lo anterior según los resultados del impulso de respuesta y la descomposición de la varianza.

- Los resultados de la red neuronal autorregresiva con perceptrón multicapa, muestra que logra capturar de mejor forma más adecuada la volatilidad de los precios, pese a no tener el mejor ajuste entre los modelos propuestos. La estructura de la red neuronal es clave para poder encontrar la salida esperada, para lo cual es aconsejable incluir Dummies en la capa de entrada, junto con los rezagos de cada variable, esto genera un mejor ajuste en la red y un mayor grado de confianza en los resultados del pronóstico.

- Realizando la comparación de los metodologías propuestas se evidencio que el mejor pronóstico fue generado por el modelo ARIMAX, seguido de la red neuronal y por último el VAR. Sin embargo, es interesante analizar cómo cada uno de los modelos entregó información relevante para el entendimiento de los valores predichos, por lo cual, se puede concluir que no solo es válido conocer el mejor modelo de pronóstico, también es importante aplicar en conjunto estos modelos, obteniendo así resultados que permitan tomar decisiones a un mayor contexto.

## **6.2 Recomendaciones**

- Se recomienda a futuros estudiantes e investigadores interesados en temas relacionados a esta investigación, la generación de un modelo con la serie de tiempo en un periodo diario y según el alcance de los trabajos es posible proponer el pronóstico del precio de bolsa nacional en horas.
- El análisis de las variables a incluir en el modelo se puede convertir en un proceso bastante laborioso, por lo cual se recomienda hacer una clasificación previa por cada tipo de factor frente al precio de bolsa nacional o cualquier otra variable a pronosticar, de forma que la decantación no se haga con todos las variables y datos seleccionados para el estudio.
- La red neuronal demostró tener un comportamiento bastante interesante, esencialmente en su homologación a los picos del precio de bolsa, por lo tanto se recomienda continuar investigaciones en este tipo de metodologías de avanzada, evaluando otras arquitecturas y modelos tanto paramétricos como no paramétricos.
- Como resultado de esta investigación se obtuvo el mejor pronóstico con un método tradicional de series temporales, sin embargo, el comparativo con la metodología de avanzada dio resultados interesantes. En consiguiente, se recomienda incursionar en métodos que permitan la integración de las dos metodologías, es decir, metodologías híbridas.
- Como se mencionó en el desarrollo de la investigación, es imperante contar con diferentes variables afectan la determinación del PBN. Es de bastante interés ampliar la búsqueda de diversas variables (climáticas, hidrológicas y económicas, entre otras) diferentes a las tenidas en cuenta para los modelos propuestos y evaluar su relación e impacto con el PBN, haciendo uso de metodologías híbridas como se mencionó en la recomendación anterior.

# Bibliografía

- Andes, U. (s.f.). Metodología Box-Jenkins. Recuperado desde [https://economia.uniandes.edu.co/files/profesores/ramon\\_rosales\\_alvarez/docs/econometria2/Salidas%5C%20y%5C%20Ejercicios/EJC202220Metodologa20Box20-20Jenkins.pdf](https://economia.uniandes.edu.co/files/profesores/ramon_rosales_alvarez/docs/econometria2/Salidas%5C%20y%5C%20Ejercicios/EJC202220Metodologa20Box20-20Jenkins.pdf)
- Banco, d. l. R. (s.f.). IPC Variación mensual. Recuperado desde <https://www.banrep.gov.co/es/estadisticas/indice-precios-consumidor-ipc>
- Barrientos, J., Rodas, E., Velilla, E., Lopera, M. & Villada, F. (2012). Modelo para el pronóstico del precio de la energía en Colombia. *Lecturas de Economía*, 91-127. Recuperado desde [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0120-25962012000200003&nrm=iso](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-25962012000200003&nrm=iso)
- CGRE. (s.f.). ¿Qué es la CREG? Recuperado desde <https://www.creg.gov.co/>
- COLOMBIA, U. N. D. (s.f.). Vertimientos. Recuperado desde [https://bogota.unal.edu.co/fileadmin/noticias/Aspectos\\_ambientales\\_Sede\\_Bogota.pdf](https://bogota.unal.edu.co/fileadmin/noticias/Aspectos_ambientales_Sede_Bogota.pdf)
- CREG. (s.f.). Capacidad Efectiva Neta. Recuperado desde <http://apolo.creg.gov.co/Publicac.nsf/1c09d18d2d5ffb5b05256eee00709c02/7476f11b89e3dfde0525785a007a70c6?OpenDocument>
- Escobar, L., Valdes, J. & Zapata, S. (2009). *Redes Neuronales Artificiales en predicción de Series de Tiempo. Una aplicación a la Industria*. Working Paper Series Universidad de Palermo.
- Geomatica. (s.f.). Serie de Tiempo. Recuperado desde <http://armillary-geomatica.blogspot.com/2015/03/series-de-tiempo-en-r.html>
- Gómez, L. P. O. (2017). *Análisis del comportamiento de los precios de oferta de energía eléctrica en Colombia del año 2000 al 2016* (Tesis de maestría, Universidad Nacional de Colombia - Sede Medellín). Magister en Ingeniería - Sistemas Energéticos. Recuperado desde <http://bdigital.unal.edu.co/64462/>
- Grupo, E. B. (s.f.). Cadena de Valor de la Energía. Recuperado desde <https://www.grupoenergiabogota.com/eeb/index.php/empresa/preguntas-frecuentes>
- Henao, J. D. V., Resonsew, I. D. & Souza, R. C. (2007). ¿Por qué es tan difícil obtener buenos pronósticos de los precios de la electricidad en mercados competitivos? *Cuadernos de Administración*, 20(34), 259-282.
- Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.

- Inteligente, X. I. (s.f.). Metadata Pública. Recuperado desde [http://informacioninteligente10.xm.com.co/Pages/Metadatapublica.aspx?Paged=TRUE&p\\_ID=147&PageFirstRow=91&&View=%5C%7BCC29D6E9-E6F6-499F-A4F2-0E4DD762D41B](http://informacioninteligente10.xm.com.co/Pages/Metadatapublica.aspx?Paged=TRUE&p_ID=147&PageFirstRow=91&&View=%5C%7BCC29D6E9-E6F6-499F-A4F2-0E4DD762D41B)
- Lozano-Espitia, L. I., Rincón-Castro, H., Rincón-Castro, H. & Lozano-Espitia, I. (2010). Formación de las tarifas eléctricas e inflación en Colombia. *Borradores de Economía; No. 634*.
- Medina Hurtado, S. (2014). Análisis de la Correlación de Largo Plazo del Precio Spot en el Mercado Eléctrico Colombiano. *Cuadernos de Administración, 27*, 67-96.
- Montenegro, A. (2010). Análisis de series de tiempo. *Bogotá: Pontificia Universidad Javeriana*.
- Muñoz-Santiago, A., Urquijo-Vanstrahlengs, J., Castro-Otero, A. & Lombana, J. (2017). Pronóstico del precio de la energía en Colombia utilizando modelos ARIMA con IGARCH. *Revista de Economía del Rosario, 20(1)*, 127-161.
- Portafolio, R. (s.f.). Precio de Bolsa Nacional. Recuperado desde <https://www.portafolio.co/economia/finanzas/funciona-venta-energia-colombia-459928>
- Rendón, J. G., Hinestroza, A. G. & Moreno, L. S. (2011). Determinantes del precio de la energía eléctrica en el mercado no regulado en Colombia. *Revista Ciencias Estratégicas, 19(26)*, 225-246.
- Richardson, I., Thomson, M. & Infield, D. (2009). A high-resolution domestic building occupancy model for energy demand simulations. Recuperado desde [https://repository.lboro.ac.uk/articles/A\\_high\\_resolution\\_domestic\\_building\\_occupancy\\_model\\_for\\_energy\\_demand\\_simulations/9563585](https://repository.lboro.ac.uk/articles/A_high_resolution_domestic_building_occupancy_model_for_energy_demand_simulations/9563585)
- Rojas, G. A. G., López, J. C. H. & Isaza, H. S. (2004). Entrenamiento de una red neuronal artificial usando el algoritmo simulated annealing. *Scientia et Technica, 1(24)*.
- Santana, J. C. (2006). Predicción de series temporales con redes neuronales: una aplicación a la inflación colombiana. *Revista Colombiana de Estadística, 29(1)*, 77-92.
- Shepero, M. & Munkhammar, J. (2018). Spatial Markov chain model for electric vehicle charging in cities using geographical information system (GIS) data. *Applied Energy, 231*, 1089-1099. doi:10.1016/j.apenergy.2018.09.175
- UPME. (s.f. a). Balance de Energía Eléctrica SIN. Recuperado desde <https://www1.upme.gov.co/Paginas/Indicadores-de-Energia.aspx>
- UPME. (s.f. b). Indicadores de Energía Eléctrica. Recuperado desde <http://www1.upme.gov.co/InformacionCifras/Paginas/PETROLEO.aspx>
- Xm. (s.f. a). Demanda de energía nacional. Recuperado desde <http://informesanuales.xm.com.co/2017/SitePages/operacion/4-1-Demanda-de-energia-nacional.aspx>
- Xm. (s.f. b). Generación de la energía. Recuperado desde <https://www.xm.com.co/Paginas/Generacion/tipos.aspx>
- Xm. (s.f. c). Indicadores de Energía. Recuperado desde <http://informacioninteligente10.xm.com.co/pages/default.aspx>

Xm. (s.f. d). Precio de Bolsa. Recuperado desde <https://www.xm.com.co/Paginas/Mercado-de-energia/precio-de-bolsa-y-escasez.aspx>