

ANÁLISIS Y MODELADO DEL PRECIO DE LA ENERGÍA ELÉCTRICA
PARA LA ADMINISTRACIÓN DE RECURSOS ENERGÉTICOS LOCALES

Presentado por
ROLANDO ANDRÉS RINCÓN SARAVIA
ALBERTH ANDRÉS PABA ÁLVAREZ

ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA
2016

ANÁLISIS Y MODELADO DEL PRECIO DE LA ENERGÍA ELÉCTRICA
PARA LA ADIMINISTRACIÓN DE RECURSOS ENERGÉTICOS LOCALES

Presentado por
ROLANDO ANDRÉS RINCÓN SARAVIA
ALBERTH ANDRÉS PABA ÁLVAREZ
*Trabajo de grado para optar por el título de
Ingeniero Electricista*

Director
IVÁN DAVID SERNA SUÁREZ
Magister en Ingeniería Eléctrica

Codirectores
GILBERTO CARRILLO CAICEDO
*Profesor Titular Laureado Emérito Universidad Industrial de Santander
Doctor Ingeniero Industrial*

GABRIEL ORDOÑEZ PLATA
*Profesor Titular Laureado Universidad Industrial de Santander
Doctor Ingeniero Industrial*

ESCUELA DE INGENIERÍAS ELÉCTRICA, ELECTRÓNICA Y DE TELECOMUNICACIONES
FACULTAD DE INGENIERÍAS FISICOMECAÑICAS
UNIVERSIDAD INDUSTRIAL DE SANTANDER
BUCARAMANGA
2016

Primero que todo gracias a Dios por permitirme culminar este proyecto, sin su voluntad nada habría sido posible. a Gloria mi abuela, quien con su amor me levantó y me dio las fuerzas para lograr mis metas, a Andrés y Alex mis hermanos, quienes con su apoyo me dieron el impulso para salir adelante, a Paola gracias por estar siempre ahí y a mis padres Albert y Aura, a quienes con su apoyo incondicional hicieron de mí una persona de bien y para bien, a ellos les debo lo que soy y lo que seré, porque nunca tendré mejores guía en la vida que ustedes dos. Los amo.

-Alberth

Agradezco a mis padres, a mi mamá por siempre apoyarme y nunca dudar de mí y a mi papá por sembrar en mí el carácter para nunca dudar de lo que podía hacer. A mis amigos que me acompañaron en todo el proceso de universidad, a mis profesores y a todos los que hicieron parte de mi vida universitaria.

-Rolando

AGRADECIMIENTOS

Agradecemos a nuestros profesores de la e3t, a nuestro director de prpoyecto Ivan Serna y a al profesor Jose Joaquín Alzate por su gran colaboración en este proyecto.

CONTENIDO

INTRODUCCIÓN	18
<hr/>	
1. ANÁLISIS DE SERIES DE TIEMPO	20
<hr/>	
1.1. PROBABILIDAD Y ESTADÍSTICA	21
1.2. TIPOS DE DATOS	21
1.3. FUENTE DE DATOS	23
1.4. COEFICIENTES DE CORRELACIÓN	23
1.4.1. Coeficientes de correlación parcial	24
1.4.2. Interpretación de los coeficientes de correlación	24
1.5. PROCESOS ESTOCÁSTICOS	25
1.5.1. Procesos estocásticos estacionarios	26
1.5.2. Procesos estocásticos no estacionarios	26
1.5.3. Pruebas de estacionalidad	27
1.5.4. Elección de la longitud del rezago	29
1.5.5. Prueba de raíz unitaria	29
2. PRONÓSTICO DE SERIES DE TIEMPO Y MODELADO.	31
<hr/>	
2.1. PRONÓSTICO	32
2.1.1. Enfoques de los pronósticos económicos	32
2.2. CREACIÓN DE MODELOS AR, MA, ARMA Y ARIMA	33
2.2.1. Modelo autorregresivo AR	33
2.2.2. Modelo de medias móviles MA	34
2.2.3. Proceso autorregresivo y de procesos móviles ARMA	34
2.2.4. Proceso autorregresivo integrado de promedios móviles ARIMA	35
2.2.5. Metodología de Box-Jenkins	35
3. PREDICCIÓN DEL PRECIO DE BOLSA	38
<hr/>	
3.1. MANEJO DE LOS DATOS	39

3.2. ESPECIFICACIÓN DEL MODELO	39
3.3. ESTIMACIÓN DE PARÁMETROS	47
3.4. DIAGNÓSTICO DEL MODELO	49
3.4.1. Valores de Likelihood, AIC y Varianza	49
3.4.2. Prueba cuantil-cuantil y prueba de Ljung-box	50
3.5. PRONÓSTICO	52
4. ANÁLISIS DE DATOS	57
<hr/>	
4.1. VENTANAS DE TIEMPO	58
4.2. ANÁLISIS DE COSTOS DE ENERGÍA EN UNA MICRO-RED	58
4.3. ANÁLISIS DE ERROR	62
5. CONCLUSIONES	68
<hr/>	
BIBLIOGRAFÍA	70
<hr/>	
ANEXOS	72
<hr/>	

LISTA DE FIGURAS

1.1. Correlograma simple con una semana de datos cada hora	29
2.1. Diagrama de flujos para el método Box - Jenkins	37
3.1. Análisis del comportamiento de un mes de datos mediante el uso de R	40
3.2. Gráfico de correlograma simple para un mes de datos	41
3.3. Gráfico de correlograma parcial para un mes de datos	42
3.4. Gráfico de correlograma simple en primera diferencia para un mes de datos. . .	43
3.5. Gráfico de correlograma parcial en primera diferencia para un mes de datos. . .	44
3.6. Gráfico de correlograma simple en segunda diferencia para un mes de datos. . .	45
3.7. Gráfico de correlograma parcial en segunda diferencia para un mes de datos. . .	46
3.8. Gráfico de acercamiento para correlograma parcial en primera diferencia para un mes de datos.	47
3.9. Resultado de prueba de q.q para un mes de datos	51
3.10. Resultado de prueba de Ljung-box para un mes de datos	52
3.11. Gráfico de franja de error de predicción para un mes de datos	54
3.12. Gráfico de valores de predicción de precio de 48 horas futuras para un mes de datos	56
4.1. Gráfico comparativo de precios en la semana 8	60
4.2. Gráfico comparativo de precios en la semana 1	61
4.3. Gráfico comparativo de precios en la semana 6	62
4.4. Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 4	64
4.5. Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 2	65
4.6. Comparación de aportes hidrológicos en 3 importantes zonas del país en las semana 6	66

4.7. Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 10	66
B.1. Análisis del comportamiento de cinco años de datos mediante el uso de R	82
B.2. Gráfico de correlograma simple para cinco años de datos	83
B.3. Gráfico de correlograma parcial para cinco años de datos	83
B.4. Gráfico de correlograma simple en primera diferencia para cinco años de datos.	84
B.5. Gráfico de correlograma parcial en primera diferencia para cinco años de datos.	84
B.6. Gráfico de correlograma simple en segunda diferencia para cinco años de datos.. . . .	85
B.7. Gráfico de correlograma parcial en segunda diferencia para cinco años de datos.	85
B.8. Resultado de prueba de q.q para cinco años de datos	87
B.9. Resultado de prueba de Ljung-box para cinco años de datos	88
B.10. Gráfico de franja de error de predicción para cinco años de datos con acercamiento al último mes.	89
B.11. Gráfico de valores de predicción de precio de 48 horas futuras para cinco años de datos.	90
B.12. Análisis del comportamiento de un año de datos mediante el uso de R	91
B.13. Gráfico de correlograma simple para un año de datos	92
B.14. Gráfico de correlograma parcial para un año de datos	92
B.15. Gráfico de correlograma simple en primera diferencia para un año de datos.	93
B.16. Gráfico de correlograma parcial en primera diferencia para un año de datos.	93
B.17. Gráfico de correlograma simple en segunda diferencia para un año de datos.	94
B.18. Gráfico de correlograma parcial en segunda diferencia para un año de datos.	94
B.19. Resultado de prueba de q.q para un año de datos.	97
B.20. Resultado de prueba de Ljung-box para un año de datos	97
B.21. Gráfico de franja de error de predicción para un año de datos con acercamiento al último mes.	99
B.22. Gráfico de valores de predicción de precio de 48 horas futuras para un año de datos.	100
B.23. Análisis del comportamiento de seis meses de datos mediante el uso de R	101
B.24. Gráfico de correlograma simple para seis meses de datos	101
B.25. Gráfico de correlograma parcial para seis meses de datos.	102
B.26. Gráfico de correlograma simple en primera diferencia para seis meses de datos.	102
B.27. Gráfico de correlograma parcial en primera diferencia para seis meses de datos.	103
B.28. Gráfico de correlograma simple en segunda diferencia para seis meses de datos.	103
B.29. Gráfico de correlograma parcial en segunda diferencia para seis meses de datos.	104

B.30. Resultado de prueba de q.q para seis meses de datos	105
B.31. Resultado de prueba de Ljung-box para seis meses de datos.	106
B.32. Gráfico de franja de error de predicción para seis meses de datos con acerca- miento al último mes.	107
B.33. Gráfico de valores de predicción de precio de 48 horas futuras para seis meses de datos.	108
B.34. Análisis del comportamiento de tres meses de datos mediante el uso de R . . .	109
B.35. Gráfico de correlograma simple para tres meses de datos	110
B.36. Gráfico de correlograma parcial para tres meses de datos	110
B.37. Gráfico de correlograma simple en primera diferencia para tres meses de datos.	111
B.38. Gráfico de correlograma parcial en primera diferencia para tres meses de datos.	111
B.39. Gráfico de correlograma simple en segunda diferencia para tres meses de datos.	112
B.40. Gráfico de correlograma parcial en segunda diferencia para tres meses de datos.	112
B.41. Resultado de prueba de q.q para tres meses de datos	114
B.42. Resultado de prueba de Ljung-box para tres meses de datos	115
B.43. Gráfico de franja de error de predicción para tres meses de datos con acerca- miento al último mes.	116
B.44. Gráfico de valores de predicción de precio de 48 horas futuras para tres meses de datos.	117
B.45. Análisis del comportamiento de una semana de datos mediante el uso de R . .	118
B.46. Gráfico de correlograma simple para una semana de datos	119
B.47. Gráfico de correlograma parcial para una semana de datos	119
B.48. Gráfico de correlograma simple en primera diferencia para una semana de datos.	120
B.49. Gráfico de correlograma parcial en primera diferencia para una semana de datos.	120
B.50. Gráfico de correlograma simple en segunda diferencia para una semana de datos.	121
B.51. Gráfico de correlograma parcial en segunda diferencia para una semana de datos.	121
B.52. Resultado de prueba de q.q para una semana de datos	124
B.53. Resultado de prueba de Ljung-box para una semana de datos	124
B.54. Gráfico de franja de error de predicción para una semana de datos.	126
B.55. Gráfico de valores de predicción de precio de 48 horas futuras para una semana de datos.	127
C.1. Gráfico comparativo entre valores pronosticados y valores reales en la semana 8	129
C.2. Gráfico comparativo entre valores pronosticados y valores reales en la semana 9	129
C.3. Gráfico comparativo entre valores pronosticados y valores reales en la semana 5	130
C.4. Gráfico comparativo entre valores pronosticados y valores reales en la semana 7	130

C.5. Gráfico comparativo entre valores pronosticados y valores reales en la semana 3	131
C.6. Gráfico comparativo entre valores pronosticados y valores reales en la semana 1	132
C.7. Gráfico comparativo entre valores pronosticados y valores reales en la semana 6	133
C.8. Gráfico comparativo entre valores pronosticados y valores reales en la semana 4	133
C.9. Gráfico comparativo entre valores pronosticados y valores reales en la semana 10	134
C.10. Gráfico comparativo entre valores pronosticados y valores reales en la semana 2	134

LISTA DE TABLAS

1.1. Serie de tiempo de la producción de huevos en Estados Unidos para los años 1990-1991. Donde X_1 : Precio por docenas en 1990 (centavos de dolar), X_2 : Precio por docenas en 1991 (centavos de dolar), Y_1 : Huevos producidos en 1990 (millones), Y_2 : Huevos producidos en 1991 (millones)	22
3.1. Parámetros arrojados por R para un mes de datos	48
3.2. Resultado de prueba de significancia para parámetros de un mes de datos . . .	49
3.3. Valores de σ^2 , log likelihood y AIC arrojados por R para un mes de datos . . .	50
3.4. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para un mes de datos.	53
3.5. Resultados finales de datos de predicción.	55
4.1. Tabla de comparación de errores entre costos reales y de predicción de 10 ventanas de tiempo de una semana de datos	59
B.1. Parámetros arrojados por R para cinco años de datos	86
B.2. Resultado de prueba de significancia para parámetros de cinco años de datos .	86
B.3. Valores de σ^2 , log likelihood y aic arrojados por R para cinco años de datos . .	87
B.4. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para cinco años de datos.	89
B.5. Resultados finales de datos de predicción.	90
B.6. Parámetros arrojados por R para un año de datos	95
B.7. Resultado de prueba de significancia para parámetros de un año de datos . . .	96
B.8. Valores de σ^2 , log likelihood y aic arrojados por R para un año de datos. . . .	97
B.9. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para un año de datos.	98
B.10. Resultados finales de datos de predicción.	99
B.11. Parámetros arrojados por R para seis meses de datos	104
B.12. Resultado de prueba de significancia para parámetros de seis meses de datos . .	105

B.13. Valores de σ^2 , log likelihood y aic arrojados por R para seis meses de datos . .	105
B.14. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para seis meses de datos.	107
B.15. Resultados finales de datos de predicción.	108
B.16. Parámetros arrojados por R para tres meses de datos	113
B.17. Resultado de prueba de significancia para parámetros de tres meses de datos .	113
B.18. Valores de σ^2 , log likelihood y aic arrojados por R para tres meses de datos .	114
B.19. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para tres meses de datos.	116
B.20. Resultados finales de datos de predicción.	117
B.21. Parámetros arrojados por R para una semana de datos	122
B.22. Resultado de prueba de significancia para parámetros de una semana de datos .	123
B.23. Valores de σ^2 , log likelihood y aic arrojados por R para una semana de datos .	124
B.24. Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para una semana de datos.	125
B.25. Resultados finales de datos de predicción.	126

LISTA DE ANEXOS

ANEXO A. MERCADO DE ENERGÍA ELÉCTRICA	73
ANEXO B. RESULTADOS DE MODELADO Y PREDICCIÓN	78
ANEXO C. GRÁFICOS COMPARATIVOS	128

RESUMEN

TÍTULO:

Análisis y modelado del precio de la energía eléctrica para la administración de recursos energéticos locales¹

AUTOR:

Rolando Andrés Rincón Saravia

Alberth Andrés Paba Álvarez²

PALABRAS CLAVE:

Series de tiempo, pronóstico, modelos ARIMA, sistemas de almacenamiento, mercados de energía.

DESCRIPCIÓN:

En el presente, la generación de energía eléctrica se ve muy influenciada por las micro-redes, dando éstas las posibilidad de no depender al cien por ciento del suministro de energía dada por las fuentes centralizadas. Sin embargo, es necesario gestionar de manera eficiente el uso de la generación, analizando los costos que conlleva la utilización óptima de dichos recursos dependiendo del precio de bolsa de energía. Para llevar acabo la optimización y el análisis de costos, es de gran ventaja conocer el comportamiento futuro del precio de bolsa de energía. Para lograr tal objetivo, en este proyecto se aplica un modelo ARIMA al precio de bolsa de la energía eléctrica, con el fin de llegar a una predicción del precio de cuarenta y ocho horas futuras, para luego comparar el precio de predicción con datos reales y dar validez a los precios y modelos de predicción. En el presente libro, se toman cuatro capítulos para abarcar todo el tema: en el primero y segundo de ellos se toman todas las bases teóricas para el modelado en ARIMA; el siguiente capítulo muestra todo el proceso de modelado y predicción de datos; al final, en un cuarto capítulo se tiene todo lo relacionado con el análisis de datos y costos. Se concluye que el modelo ARIMA es una útil herramienta de predicción ya que, aunque no es posible determinar de manera precisa los futuros precios de la energía, los costos por consumo de energía eléctrica considerando almacenamiento se encuentran en una franja de error menor al cinco por ciento respecto al caso con el perfil de precios exacto

¹ Trabajo de grado.

² Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones. Director: Iván David Serna Suárez. Codirector: Gilberto Carrillo Caicedo. Codirector: Gabriel Ordoñez Plata.

ABSTRACT

TITLE:

Analysis and modeling of the price of electricity for the administration of local energy resources³

AUTHORS:

Rolando Andrés Rincón Saravia

Alberth Andrés Paba Álvarez⁴

KEY WORDS:

Simple correlogram, partial correlogram, econometrics, ARIMA model, time window limits Bartlett, R, variance, maximum likelihood, AIC, actual cost, cost prediction.

DESCRIPTION:

Nowdays, electricity generation is greatly influenced by the microgrids, due to its ability to avoid energy consumption from centralised energy sources. Therefore, there is a need to manage efficiently the generation scheduling through an analysis of the generation costs depending on the energy price. For the operation and cost analysis of such systems is of great advantage to know the future behavior of the price for electricity. In this project an ARIMA model is applied to the price of electricity, to get a price prediction for the next forty-eight hours and then compare the price prediction with actual data in order to validate the predictive models and prices. In this book, four chapters are taken to cover the whole issue: in the first and the second one all the theoretical basis for ARIMA modeling is taken; the next chapter shows the whole process of modeling and forecasting data; On chapter three the analysis of data and costs is presented. It is concluded that the ARIMA model is a useful predictive since, even though it can't predict precisely the energy price, the resulting energy consumption cost is within an error of five percent with respect the exact price profile case. In the same way, it is concluded that the model errors in different moments of time are due to abrupt climate changes and to the lack of information about the strategic behavior of the electricity market agents.

³Research work.

⁴Faculty of Physical-Mechanic Engineering. School of Electrical, Electronical and Telecommunications Engineering. Advisor: Iván David Serna Suárez. Co-advisor: Gilberto Carrillo Caicedo. Co-advisor: Gabriel Ordoñez Plata.

INTRODUCCIÓN

En el presente, la generación de energía eléctrica se ve muy influenciada por las micro-redes, dando éstas la posibilidad de independencia del suministro de energía centralizado, sin embargo, es conveniente decidir en que momento se conecta por completo a la generación alternativa y se desconecta del suministro de energía centralizado y viceversa, dependiendo del precio por hora de la energía y del costo que conlleva generar de manera eficiente mediante energía solar, eólica o cualquier otra fuente de generación que esté presente en la micro-red. Por tal motivo, es necesario gestionar de manera eficiente el uso de la generación, analizando los costos que conlleva la optimización de dichas micro-redes dependiendo del precio de bolsa de energía (comisión de regulación de energía y gas, 2007).

Para llevar a cabo la optimización y el análisis de costos, es de gran ventaja conocer el comportamiento futuro del precio de bolsa de energía, para lograr tal objetivo, el estudio matemático moderno presenta diferentes tipos de modelos dependiendo del comportamiento de los datos, los cuales brindan una herramienta poderosa y con alta eficiencia para tal propósito. En este proyecto se modela, mediante el uso de modelos ARIMA, el precio de bolsa de la energía eléctrica, con el fin de llegar a una predicción del precio de cuarenta y ocho horas futuras con la menor franja de error posible, para luego comparar el precio de predicción con datos reales y dar validez a los precios y modelos de predicción.

En el presente libro, se tienen tres capítulos para abarcar todo el tema. En el primer capítulo se toman todas las bases teóricas para el modelado en ARIMA, entre éstas, conceptos bases relacionados con probabilidad, estadística. El siguiente capítulo muestra todo el proceso de modelado y predicción de datos, comenzando con la especificación del modelo, pasando luego a la estimación de parámetros, el diagnóstico del modelo y la predicción como tal de

los valores de precio de bolsa de energía, mediante el análisis de la mejor ventana de tiempo para el análisis y modelado de los datos. Al final, en el tercer capítulo se tiene una comparación de costos de optimización entre datos reales contra datos de predicción para realizar un análisis de error y aprobar el uso de los modelos para la optimización, de igual manera se indaga en el por qué de los errores de predicción tomando como base el comportamiento del clima en dichas ventanas de tiempo. Para terminar, se tienen las conclusiones de los resultados, las cuales llevan a la aprobación de los modelos ARIMA para la optimización de micro-redes y muestran que los cambios abruptos de clima llevan a errores de predicción del precio de bolsa de energía.

CAPÍTULO 1

ANÁLISIS DE SERIES DE TIEMPO

Para el estudio de pronósticos de series de tiempo, se encuentran muchas herramientas que ayudan a tratar de predecir aquel dato que se quiere estimar, así que como base a un estudio de este tipo se necesita deducir si el pronóstico es válido o definitivamente está lejos de un resultado esperado. Por tal razón se comienza por estudiar la probabilidad y estadística para establecer la columna vertebral al estudio, continuado del análisis de los datos, el tipo de los datos y sus clasificaciones. Seguidamente para conceptualizar más profundamente, se enuncian los correlogramas y sus tipos, con sus respectivos análisis de resultados de coeficientes. Los procesos estocásticos son el siguiente tema a tratar, ya que su estudio conduce a analizar y clasificar todas las variables que influyen en un proceso, como lo son los procesos estocásticos estacionarios y los no estacionarios, con sus respectivas subdivisiones que facilitan entender el contorno general del tema. Seguidamente se estudia la estacionalidad y el análisis gráfico para llegar a la función de autocorrelación, donde se investiga acerca de la elección del rezago, la prueba de raíz unitaria y al tener claridad y dominio del tema, se analiza la parte concluyente que es el pronóstico y modelado de las series de tiempo.

1.1 PROBABILIDAD Y ESTADÍSTICA

El acopio de datos no es nada nuevo ya que esta práctica se realiza desde hace cientos de años, tomando gran fuerza e importancia durante la revolución industrial, lo que llevó al uso de los datos estadísticos en diferentes campos como la manufactura, desarrollo de productos alimenticios, farmacéuticos y software

La estadística permite determinar cosas tan simples como la cantidad de accidentes en un cruce de calles para estipular si se necesita o no un semáforo, o estudiar el gas emitido por un transformador en caso de fallas para determinar el origen de las mismas. Por el contrario de la probabilidad, que a partir de los datos tomados ya sean discretos o continuos, permite establecer a futuro, bajo ciertos criterios y utilizando las herramientas adecuadas cuantos accidentes ocurrirán si no se toman correctivos o cual será la vida útil de un transformador teniendo en cuenta las muestras obtenidas en cada experimento o suceso (Gujarati & Porter, 2010). Gran parte del éxito de un estudio econométrico, depende de los datos recopilados, conocer su origen, naturaleza y limitaciones para el análisis empírico.

1.2 TIPOS DE DATOS

Existen principalmente 3 tipos de datos para el análisis empírico. Series de tiempo, series transversales e información combinada. Esta última es la combinación de datos de series de tiempo y series transversales. Para este trabajo se profundiza en los datos de las series de tiempo ya que estos son la columna vertebral del estudio.

Datos de series de tiempo Una serie de tiempo son un conjunto de observaciones sobre una variable en diferentes momentos. La información para este tipo de series debe recopilarse en intervalos regulares dependiendo de su naturaleza u origen. Los datos pueden ser recopilados de manera diaria como el precio de las acciones, semanal como cifras de oferta monetaria, mensual como tasas de desempleo, trimestral como el producto interno bruto PIB, anual como los presupuestos del gobierno, decenal como el censo poblacional etc, pero para el caso del trabajo, tomar datos diarios no brindaría la certeza de obtener un pronóstico acertado ya que el mercado de energía eléctrica es tan dinámico que en cuestión de minutos fluctúa el precio en la bolsa y de ahí depende el éxito de las transacciones, por ende para el proceso se debe tomar datos como mínimo cada hora y hacer la serie de tiempo con esa regularidad para así asegurar la calidad de los resultados. La mayor parte del trabajo empírico con series de tiempo se hace bajo la suposición que es estacionaria. Para aclarar la idea se cita el siguiente cuadro acerca de la producción de huevos en Estados Unidos.

TABLA 1.1.: Serie de tiempo de la producción de huevos en Estados Unidos para los años 1990-1991. Donde X_1 : Precio por docenas en 1990 (centavos de dolar), X_2 : Precio por docenas en 1991 (centavos de dolar), Y_1 : Huevos producidos en 1990 (millones), Y_2 : Huevos producidos en 1991 (millones)

ESTADO	Y_1	Y_2	X_1	X_2	ESTADO	Y_1	Y_2	X_1	X_2
AL	2206	2186	92.7	91.4	MT	172	164	68	66
AK	0.7	0.7	151	149	NE	1202	1400	50.3	48.9
AZ	73	74	61	56	NV	2.2	1.8	53.9	52.7
AR	3620	3737	86.3	91.8	NH	43	49	109	104
CA	7472	7444	63.4	58.4	NJ	442	491	85	83
CO	788	873	77.8	73	NM	283	302	74	70
CT	1029	948	106	104	NY	975	987	68.1	64
DE	168	164	117	113	NC	3033	3045	82.8	78.7
FL	2586	2537	62	57.2	ND	51	45	55.2	48
GA	4302	4301	80.6	80.8	OH	4667	4637	59.1	54.7
HI	227.5	224.5	85	85.5	OK	869	830	101	100
ID	187	203	79.1	72.9	OR	652	686	77	74.6
IL	793	809	65	70	PA	4976	5130	61	52
IN	5445	5290	62.7	60.1	RI	53	50	102	99
IA	2151	2247	56.5	53	SC	1422	1420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48	45.8
KY	412	483	67.7	73.5	TN	277	279	71	80.7
LA	273	254	115	115	TX	3317	3356	76.7	72.6
ME	1069	1070	101	97	UT	456	486	64	59
MD	885	898	76.6	75.4	VT	31	30	106	102
MA	235	237	105	102	VA	943	988	86.3	81.2
MI	1406	1396	58	53.8	WA	1287	1313	74.1	71.5
MN	2499	2697	57.7	54	WV	136	174	104	109
MS	1434	1468	87.8	86.7	WI	910	873	60.1	54
MO	1580	1622	55.4	51.1	WY	1.7	1.7	83	83

Fuente: (?)

Datos transversales Los datos transversales consisten en datos de una o más variables recopilados en el mismo punto del tiempo, como el censo de población realizado por el DANE cada 10 años, las encuestas de gastos del consumidor o las encuestas de opinión. (Gujarati & Porter, 2010)

Datos Combinados Los datos combinados reúnen elementos de series de tiempo y transversales. Los datos de la figura 1 son datos combinados. Hay 50 observaciones transversales por año, y dos observaciones de series de tiempo sobre precios y producción de huevo por

estado: un total de 100 observaciones combinadas.

1.3 FUENTE DE DATOS

Los datos para el análisis empírico provienen de una dependencia gubernamental por ejemplo, la CREG, un organismo internacional como el Fondo Monetario Internacional (FMI) o el Banco Mundial, una organización privada o un particular. Hay varias agencias de este tipo que recopilan datos para uno u otro fin. Los datos recopilados por estas organizaciones pueden ser de naturaleza experimental o no experimental. En los datos experimentales, frecuentes en las ciencias naturales, el investigador suele recabar los datos con algunos factores constantes, con el fin de evaluar el efecto de otros en un fenómeno dado. Por ejemplo, al estimar el efecto de la obesidad en la presión arterial, el investigador recopilaría los datos y mantendría constantes los hábitos de las personas respecto de comer, fumar y beber para reducir la influencia de estas variables en la presión arterial. En las ciencias sociales, los datos por lo general son de naturaleza no experimental, es decir, no están sujetos al control del investigador. Por ejemplo, el investigador no ejerce ningún control directo sobre los datos del PIB, desempleo, precios de acciones, etc. Como se analiza, esta falta de control a menudo genera problemas especiales para el investigador al identificar la causa o causas precisas que afectan una situación particular.

1.4 COEFICIENTES DE CORRELACIÓN

El coeficiente de correlación r es la medida del grado de asociación lineal entre dos variables. Para el modelo de regresión con tres variables podemos calcular tres coeficientes de correlación: r_{12} (correlación entre Y y X_2), r_{13} (coeficiente de correlación entre Y y X_3) y r_{23} (coeficiente de correlación entre X_2 y X_3); observe que el subíndice 1 representa a Y por conveniencia notacional y los subíndices 2 y 3 representan a los factores X_2 y X_3 respectivamente, los cuales hacen referencia a las variables para establecer la relación lineal. Estos coeficientes de correlación se denominan coeficientes de correlación bruta o simple, o coeficientes de correlación de orden cero, y se calculan a partir de la siguiente definición. (Walpole *et al.*, 1999).

$$r = \frac{(\sum x_i y_i)}{\sqrt{[(\sum x_i^2) (\sum y_i^2)]}} \quad (1.1)$$

1.4.1 Coeficientes de correlación parcial No es posible afirmar que r_{12} mide el “verdadero” grado de asociación (lineal) entre Y y X_2 cuando existe una tercera variable X_3 que puede estar asociada a ellas. Suponga que el verdadero modelo de regresión es (1.2) pero se omite del modelo la variable X_3 , y sólo se hace la regresión Y sobre X_2 para obtener el coeficiente de la pendiente. En general, r_{12} tal vez no refleje el verdadero grado de asociación entre Y y X_2 en presencia de X_3 . De hecho, es probable que dé una falsa impresión de la naturaleza de la asociación entre Y y X_2 , como se demuestra en breve. Por consiguiente, lo que se necesita es un coeficiente de correlación que sea independiente de la influencia, si hay alguna, de X_3 sobre X_2 y Y . Dicho coeficiente de correlación se obtiene y se conoce como coeficiente de correlación parcial. En cuanto concepto, es similar al coeficiente de regresión parcial. Se define:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (1.2)$$

$r_{12,3}$ = coeficiente de correlación parcial entre Y y X_2 , manteniendo X_3 constante.

$r_{13,2}$ = coeficiente de correlación parcial entre Y y X_3 , manteniendo X_2 constante.

$r_{23,1}$ = coeficiente de correlación parcial entre X_2 y X_3 , manteniendo Y constante.

$$r_{12,3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{[(1 - r_{13}^2)(1 - r_{23}^2)]}} \quad (1.3)$$

$$r_{13,2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{[(1 - r_{12}^2)(1 - r_{23}^2)]}} \quad (1.4)$$

$$r_{23,1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{[(1 - r_{12}^2)(1 - r_{13}^2)]}} \quad (1.5)$$

1.4.2 Interpretación de los coeficientes de correlación Para el caso en el que se presenten las dos variables, el coeficiente de correlación simple tiene un significado directo: medir el grado de asociación lineal, mas no causal, entre la variable dependiente Y y la variable explicativa X . Sin embargo, al presentarse el caso de dos variables, se debe prestar minuciosa atención a la interpretación del coeficiente de correlación simple. De (1.3), por ejemplo, se debe prestar atención a lo siguiente:

- Aunque $r_{12} = 0$, $r_{12,3}$ no será cero a menos que r_{13} o r_{23} , o ambos, sean cero.
- Si $r_{12} = 0$ y r_{13} y r_{23} son diferentes de cero y tienen el mismo signo, $r_{12,3}$ es negativo, mientras que si son de signos opuestos, es positivo. Un ejemplo aclara este punto. Sea

Y = rendimiento del cultivo, X_2 = la lluvia y X_3 = la temperatura. Suponga que $r_{12} = 0$, es decir, no hay asociación entre el rendimiento del cultivo y la lluvia. Tenga en cuenta, además, que r_{13} es positiva y r_{23} es negativa. Entonces, como lo indica (1.3), $r_{12,3}$ es positivo; es decir, con la temperatura constante, existe una asociación positiva entre el rendimiento del cultivo y la lluvia. Pero, este resultado, en apariencia paradójico, no es sorprendente. Como la temperatura X_3 afecta el rendimiento Y y también afecta la lluvia X_2 , con el fin de encontrar la relación neta entre rendimiento del cultivo y lluvia se debe eliminar la influencia o el peso de la “molesta” variable temperatura. Este ejemplo muestra la posibilidad que el coeficiente de correlación simple genere resultados equivocados.

- Los términos $r_{12,3}$ y r_{12} (y comparaciones similares) no necesitan tener el mismo signo.
- En el caso de dos variables r_2 se encuentra entre 0 y 1. La misma propiedad se cumple para los coeficientes de correlación parcial al cuadrado. Así, el lector debe verificar que es posible obtener la siguiente expresión a partir de (1.3):

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \quad (1.6)$$

Suponga que $r_{13} = r_{23} = 0$. no significa esto que r_{12} también es cero. La respuesta se desglosa de (1.6). El hecho de que Y y X_3 y X_2 y X_3 no estén correlacionadas no significa que Y y X_2 no lo estén. A propósito, observe que la expresión $r_{12,3}^2$ puede denominarse coeficiente de determinación parcial e interpretarse como la proporción de la variación en Y no explicada por la variable X_3 que se explica por la inclusión de X_2 en el modelo. Conceptualmente, es semejante a R^2 . Antes de continuar observe las siguientes relaciones entre R^2 , los coeficientes de correlación simple y los coeficientes de correlación parcial:

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \quad (1.7)$$

$$R^2 = r_{12}^2 + (1 - r_{12}^2) r_{13,2}^2 \quad (1.8)$$

$$R^2 = r_{13}^2 + (1 - r_{13}^2) r_{12,3}^2 \quad (1.9)$$

1.5 PROCESOS ESTOCÁSTICOS

Un proceso estocástico o aleatorio es una colección de variables aleatorias ordenadas en el tiempo. Si Y denota una variable aleatoria y es continua, se denota como $Y(t)$, como por ejemplo los electrocardiogramas, pero si es discreta se expresa como Y_t . (Gujarati & Porter,

2010). Como el producto interno bruto. La mayoría de los datos económicos se recopilan en puntos discretos de tiempo. La distinción entre el proceso estocástico y su realización es semejante a la diferencia entre población y muestra en datos de corte transversal. De la misma forma como se hacen inferencias sobre la población a partir de datos muestrales, se efectúan inferencias sobre el proceso estocástico subyacente en las series de tiempo mediante la realización. (Gujarati & Porter, 2010; Walpole *et al.*, 1999).

1.5.1 Procesos estocásticos estacionarios El proceso estocástico que ha recibido gran atención y ha sido objeto de investigación por parte de los observadores de series de tiempo es el proceso estocástico estacionario. En resumen, se dice que un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende sólo de la distancia o rezago entre estos dos periodos, y no del tiempo en el cual se calculó la covarianza. En la descripción sobre las series de tiempo, un proceso estocástico como éste, se conoce como proceso estocástico débilmente estacionario, estacionario covariante, estacionario de segundo orden o proceso estocástico en amplio sentido.

1.5.2 Procesos estocásticos no estacionarios Aunque nuestro interés se centra en las series de tiempo estacionarias, a menudo se topa uno con series de tiempo no estacionarias, cuyo ejemplo clásico es el modelo de caminata aleatoria (MCA). A menudo se dice que los precios de valores, como las acciones o las tasas de cambio, siguen una caminata aleatoria; es decir, son no estacionarios. Hay dos tipos de caminatas aleatorias: Caminata aleatoria sin deriva o sin desvío (es decir, sin término constante o de intercepto) y caminata aleatoria con deriva o con desvío (es decir, hay un término constante).

Caminata aleatoria sin deriva Suponga que u_t es un término de error de ruido blanco, con media 0 y varianza σ^2 . Entonces se dice que la serie Y_t es una caminata aleatoria si

$$Y_t = Y_{t-1} + u_t \quad (1.10)$$

En el modelo de caminata aleatoria, como se ve en (1.10), el valor de Y en el tiempo t es igual a su valor en el tiempo $(t - 1)$ más un choque aleatorio; por tanto, es un modelo AR(1). Es posible pensar que (1.10) es una regresión de Y en el tiempo t sobre su valor atrasado un periodo. Los defensores de la hipótesis del mercado de capital eficiente argumentan que los precios de las acciones son en esencia aleatorios y, por tanto, no hay lugar para la especulación rentable en el mercado de valores. Si se pudiese predecir el precio de las acciones del día siguiente con base en su precio del día anterior, todos seríamos millonarios.

Caminata aleatoria sin deriva Se modifica (1.10) de la siguiente forma:

$$Y_t = \delta + Y_{t-1} + U_t \quad (1.11)$$

Donde δ se conoce como el parámetro de deriva. El término deriva proviene del hecho de que, si se escribe la ecuación anterior como

$$Y_t - Y_{t-1} = \Delta Y = \delta + U_t \quad (1.12)$$

Se demuestra que Y_t se deriva o desvía hacia arriba o hacia abajo, según δ sea positiva o negativa. Observe que el modelo (1.11) también es un modelo AR(1).

Según el procedimiento analizado en la caminata aleatoria sin deriva, es posible demostrar que, para el modelo de caminata aleatoria con deriva (1.11) .

$$E(Y_t) = Y_0 + t\delta \quad (1.13)$$

$$\text{var}Y_t = t\sigma^2 \quad (1.14)$$

Como se observa, para el MCA con deriva, la media, al igual que la varianza, se incrementa con el tiempo, lo que viola de nuevo las condiciones de la estacionalidad (débil). En resumen, el MCA, con o sin deriva, es un proceso estocástico no estacionario.

1.5.3 Pruebas de estacionalidad Aunque existen varias pruebas de estacionalidad, se enfoca en las 2 más eficientes y utilizadas por los estudiosos de las series de tiempos, las cuales son el método gráfico y la función de autocorrelación (FAC) y correlograma.

Análisis gráfico Es importante recordar que antes de efectuar una prueba formal, siempre es aconsejable graficar la serie de tiempo en estudio, Estas gráficas proporcionan una pista inicial respecto de la posible naturaleza de las series de tiempo. Por ejemplo, si se considera una serie de tiempo de PIB (que por lo general tiende al alza) se observa que, a lo largo del periodo de estudio, el logaritmo del PIB se incrementa, es decir, muestra una tendencia ascendente, lo cual deja entrever que quizá esté variando la media del logaritmo del PIB. Esto tal vez indique que la serie logarítmica del PIB es no estacionaria, lo cual es más o menos verdadero para otras series de tiempo económicas. Esa intuición es el comienzo de una prueba más formal de estacionalidad.

Función de autocorrelación y correlograma Una prueba sencilla de estacionalidad se basa en la denominada función de autocorrelación (FAC). La FAC en el rezago k , denotada por ρ_k , se define como

$$\rho k = \frac{Y_k}{Y_0} = \frac{\text{covariana en rezago } k}{\text{varianza}} \quad (1.15)$$

Como la covarianza y la varianza se miden en las mismas unidades, ρk es un número sin unidad de medida, o adimensional. Se encuentra entre -1 y $+1$, igual que cualquier coeficiente de correlación. Si se gráfica ρk respecto de k , la gráfica obtenida se conoce como correlograma poblacional. Como, en la práctica, sólo se tiene una realización de un proceso estocástico (es decir, la muestra), sólo se calcula la función de autocorrelación muestral, ρk . Para ello, se debe calcular primero la covarianza muestral en el rezago k , Y_k , y la varianza muestral, Y_0 definidas como

$$Y_k = \frac{\sum(Y_t - Y)(Y_{t+k} - Y)}{n} \quad (1.16)$$

$$Y_0 = \frac{\sum(Y_t - Y)^2}{n} \quad (1.17)$$

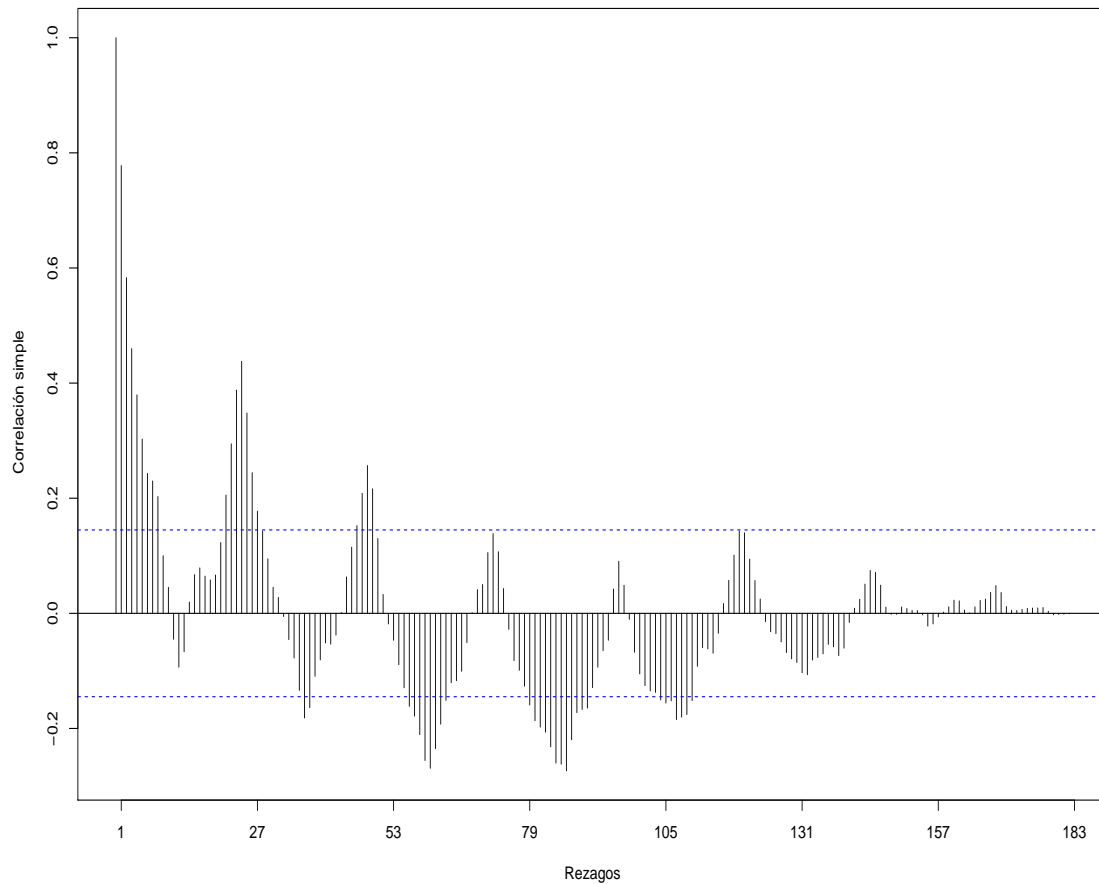
Donde n es el tamaño de la muestra y Y es la media muestral, por ende la función de autocorrelación en el rezago k es

$$\rho k = \frac{Y_k}{Y_0} \quad (1.18)$$

Que simplemente es la relación entre la covarianza muestral (en el rezago k) y la varianza muestral.

Para determinar con un correlograma si una serie de tiempo en particular es estacionaria, se presenta un correlograma muestral de un proceso puramente aleatorio de ruido blanco. Regresando al modelo de caminata aleatoria MCA sin deriva, ahí se genera una muestra de 500 términos de error, las u , a partir de la distribución normal estandarizada. El correlograma para estos 500 términos de error puramente aleatorios, se muestran en este correlograma hasta 30 rezagos. Más adelante se define cómo elegir la longitud del rezago. Por el momento, sólo observe la columna AC, que es la función de autocorrelación muestral, y el primer diagrama de la izquierda, llamado autocorrelación. La línea vertical continua de este diagrama representa el eje cero; las observaciones por arriba de esta línea son valores positivos, y los que están por debajo, negativos. Como resulta evidente a partir de este diagrama, para un proceso puramente de ruido blanco, las autocorrelaciones en distintos rezagos se ubican alrededor del cero. Ésta es una imagen de un correlograma de una serie de tiempo estacionaria. Por tanto, si el correlograma de una serie de tiempo real (económica) se parece al correlograma de una serie de tiempo de ruido blanco, se dice que dicha serie de tiempo es quizá estacionaria.

Figura 1.1.: Correlograma simple con una semana de datos cada hora



Fuente: ?

1.5.4 Elección de la longitud del rezago La elección de la longitud del rezago se trata de algo empírico, una regla práctica es calcular la función de autocorrelación FAC hasta un tercio o una cuarta parte de la longitud de la serie de tiempo. El mejor consejo es comenzar con rezagos bastante grandes y luego reducirlos por medios estadísticos.

1.5.5 Prueba de raíz unitaria Otra prueba sobre estacionalidad (o no estacionalidad) que se populariza cada vez más se conoce como prueba de raíz unitaria. El punto de partida es el proceso estocástico de raíz unitaria se inicia con

$$Y_t = \rho Y_{t-1} + u_t \quad -1 \leq \rho \leq 1 \quad (1.19)$$

Donde u_t es un término de error de ruido blanco. Se sabe que si $\rho = 1$, es decir, en el caso de la raíz unitaria, (1.19) se convierte en un modelo de caminata aleatoria sin deriva, del cual se conoce también que es un proceso estocástico no estacionario. Por consiguiente, si simplemente se hace la regresión de Y_t sobre su valor rezagado (de un periodo) Y_{t-1} y se averigua si la ρ estimada es estadísticamente igual a 1, de ser así, Y_t es no estacionaria. Ésta es la idea general de la prueba de raíz unitaria para la estacionalidad. Sin embargo, no se puede estimar la ecuación (1.19) por mínimos cuadrados ordinarios MCO y probar la hipótesis de que $\rho = 1$ por medio de la prueba t acostumbrada, porque esa prueba tiene un sesgo muy marcado en el caso de una raíz unitaria. Por tanto, se manipula (1.19) de la siguiente forma: se resta Y_{t-1} de ambos miembros de la ecuación (1.19) para obtener:

$$Y_t - Y_{t-1} = \rho Y_{t-1} - Y_{t-1} - u_t = (\rho - 1)Y_{t-1} + U_t \quad (1.20)$$

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (1.21)$$

Donde $\delta = (\rho - 1)$ y Δ es el operador de pequeñas diferencias.

Por tanto, en la práctica, en vez de estimar (1.19), se calcula (1.21) y se prueba la hipótesis (nula) de que $\delta = 0$, y la hipótesis alternativa es que $\delta < 0$. Si $\delta = 0$, entonces $\rho = 1$; es decir, se tiene una raíz unitaria, lo cual significa que la serie de tiempo en consideración es no estacionaria. Antes de proceder con la estimación de (1.21) se debe observar que si $\delta = 0$, entonces (1.21) se convertirá en

$$\Delta Y_t = (Y_t - Y_{t-1}) = u_t \quad (1.22)$$

Como u_t es un término de error de ruido blanco, entonces es estacionario, lo cual significa que las primeras diferencias de una serie de tiempo de caminata aleatoria son estacionarias, una observación que ya se ha enunciado.

CAPÍTULO 2

PRONÓSTICO DE SERIES DE TIEMPO Y MODELADO.

El pronóstico, junto con sus enfoques constituyen el primer análisis a realizar a los datos obtenidos en la serie de tiempo, es por eso que se revisa minuciosamente hasta llegar a la creación de modelos AR, MA, ARMA y ARIMA, este último también conocido como método BJ (Box-Jenkins), el cual es la base para el propósito final de este trabajo que es pronosticar el precio de la energía eléctrica en Colombia, sustentado por pruebas para dar certeza de la veracidad del pronóstico encontrado.

2.1 PRONÓSTICO

Los pronósticos son una parte importante en el análisis econométrico, y en algunos gremios definitivamente constituye el área más importante y concluyente. Variables tan significativas como el producto interno bruto, inflación, tasas de cambio y tasas de desempleo son analizadas por 2 métodos que han adquirido mucha popularidad por su certeza. El primero es el método autorregresivo de promedios móviles (ARIMA), también conocido como el método de Box-Jenkins y el segundo método es el de vectores autorregresivos VAR. (Gujarati & Porter, 2010).

2.1.1 Enfoques de los pronósticos económicos En términos generales, hay cinco enfoques de los pronósticos económicos basados en series de tiempo.

Métodos de suavizamiento exponencial Son métodos para ajustar una curva apropiada a datos históricos de una determinada serie de tiempo. Existen diversos tipos como el suavizamiento exponencial simple, el método lineal de Holt y el método de Holt-Winters así como sus variaciones.

Modelos de regresión uniecuacionales Como ejemplo de un modelo de regresión uniecuacional, considere la función de demanda de automóviles. La demanda de automóviles es función de sus precios, publicidad, ingresos del consumidor, tasa de interés y otras variables relevantes como tamaño de la familia o distancia al trabajo. A partir de las series de tiempo estimamos un modelo apropiado de demanda de automóviles (lineal, log-lineal o no lineal), el cual sirve para el pronóstico de la demanda en el futuro. Los errores de pronóstico aumentan rápidamente si se va demasiado lejos hacia el futuro.

Modelos de regresión de ecuaciones simultáneas En su momento de apogeo, durante las décadas de 1960 y 1970, los modelos elaborados para describir la economía de EEUU basados en ecuaciones simultáneas dominaron el campo de los pronósticos económicos. Pero más adelante, el encanto por ese tipo de pronósticos terminó debido a las crisis del precio del petróleo de 1973 y de 1979 y a la crítica de Lucas. El argumento de esta crítica es que los parámetros estimados de un modelo econométrico dependen de la política prevaleciente en el momento en que se estima el modelo y cambian conforme lo hace la política. En conclusión, los parámetros estimados no son invariantes ante cambios de política.

Modelos autorregresivos integrados de promedios móviles. Popularmente conocida como metodología de Box-Jenkins (BJ), pero técnicamente conocida como metodología ARIMA, el interés de estos métodos de pronósticos no está en la construcción de modelos uniecuacionales o de ecuaciones simultáneas, sino en el análisis de las propiedades probabilísticas, o estocásticas, de las series de tiempo económicas por sí mismas según la filosofía de que “los datos hablen por sí mismos”.

Modelos de vectores autorregresivos A primera vista, la metodología de vectores autorregresivos VAR se asemeja a los modelos de ecuaciones simultáneas, pues considera diversas variables endógenas de manera conjunta. Pero cada variable endógena se explica por sus valores rezagados, y por los valores rezagados de todas las demás variables endógenas en el modelo; usualmente no hay variables exógenas en el modelo.

2.2 CREACIÓN DE MODELOS AR, MA, ARMA Y ARIMA

Esta tesis se fundamenta en la aplicación de modelos autorregresivos AR y de medias móviles MA, que en algunos casos es necesario estacionalizar la series de tiempo con un método integrativo, de ahí la diferencia entre un proceso ARMA y un proceso ARIMA. Por tal razón se presenta este marco teórico.

2.2.1 Modelo autorregresivo AR Sea Y_t el logaritmo de la serie de tiempo en un periodo t . Si se modela Y_t como

$$Y_t - \delta = \alpha_1(Y_{t-1} - \delta) + u_t \quad (2.1)$$

Donde δ es la media de Y y u_t es un término de error aleatorio no correlacionado con media cero (0) y varianza constante σ^2 , es decir, ruido blanco. Se dice que Y_t sigue un proceso estocástico autorregresivo de primer orden, o AR (1), aquí el valor de Y en el tiempo t depende de su valor en el periodo anterior y de un término aleatorio, los valores de Y están expresados como desviaciones de su valor medio. En otras palabras, este modelo dice que el valor de pronóstico de Y en el periodo t es simplemente alguna proporción ($= \alpha_1$) de su valor en el periodo $(t-1)$ más una perturbación aleatoria en el tiempo t ; de nuevo, los valores de Y están expresados alrededor del valor de su media.

Pero si consideramos este modelo

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + u_t \quad (2.2)$$

Se dice que Y_t sigue un proceso autorregresivo de segundo orden o AR (2). Es decir, el valor de Y en el tiempo t depende de sus valores en los dos periodos anteriores, los valores de Y expresados alrededor del valor de su media δ .

En general se tiene

$$(Y_t - \delta) = \alpha_1(Y_{t-1} - \delta) + \alpha_2(Y_{t-2} - \delta) + \dots + \alpha_p(Y_{t-p} - \delta) + u_t \quad (2.3)$$

En cuyo caso Y_t es un proceso autorregresivo de orden p , o AR (p). Observe que en todos los modelos anteriores sólo se consideran los valores actuales y anteriores de Y , no hay otras regresoras. En este sentido, se dice que “los datos hablan por sí mismos”.

2.2.2 Modelo de medias móviles MA El proceso AR ya expuesto no es el único mecanismo que pudo generar a Y . Suponga que se hace un modelo de Y de la siguiente manera

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} \quad (2.4)$$

Donde μ es una constante y u , al igual que antes, es el término de error estocástico de ruido blanco. Aquí, Y en el periodo t es igual a una constante más un promedio móvil de los términos de error presente y pasado. Así, en este caso decimos que Y sigue un proceso de promedios móviles de primer orden, o MA (1).

Pero si Y sigue la expresión

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} \quad (2.5)$$

Es un proceso MA(2), en forma general se obtiene que

$$Y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (2.6)$$

Es un proceso MA(q). En resumen, un proceso de promedios móviles es tan sólo una combinación lineal de términos de error de ruido blanco.

2.2.3 Proceso autorregresivo y de procesos móviles ARMA Es muy probable que Y tenga características tanto de AR como de MA a la vez, por consiguiente sea un ARMA, así que si Y_t sigue un proceso ARMA (1,1) se describirá de la siguiente manera

$$Y_t = \theta + \alpha_1(Y_{t-1}) + \beta_0 u_t + \beta_1 u_{t-1} \quad (2.7)$$

Porque hay un término autorregresivo y uno de promedios móviles. Para este caso el símbolo θ representa un término constante.

En general, en un proceso ARMA (p, q) , hay p términos autorregresivos y q términos de promedios móviles.

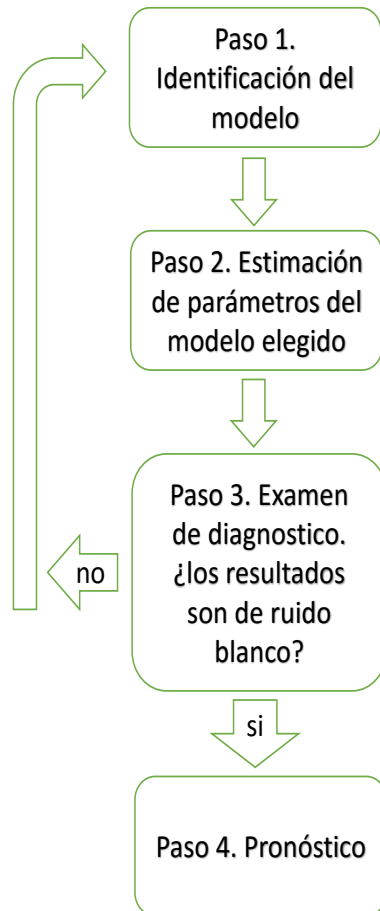
2.2.4 Proceso autorregresivo integrado de promedios móviles ARIMA Los modelos de series de tiempo analizados se basan en el supuesto de que las series de tiempo consideradas son (débilmente) estacionarias. En pocas palabras, la media y la varianza de una serie de tiempo débilmente estacionaria son constantes y su covarianza es invariante en el tiempo. Pero se conoce que muchas series de tiempo económicas son no estacionarias, es decir, son integradas, Sin embargo, si una serie de tiempo es integrada de orden 1 ($I(1)$), sus primeras diferencias son $I(0)$, es decir, estacionarias. En forma similar, si una serie de tiempo es $I(2)$, sus segundas diferencias son $I(0)$. En general, si una serie de tiempo es $I(d)$, después de diferenciarla d veces se obtiene una serie $I(0)$. Por consiguiente, si se debe diferenciar una serie de tiempo d veces para hacerla estacionaria y luego aplicarle el modelo ARMA (p, q) , se dice que la serie de tiempo original es ARIMA (p, d, q) , es decir, es una serie de tiempo autorregresiva integrada de promedios móviles, donde p denota el número de términos autorregresivos, d el número de veces que la serie debe diferenciarse para hacerse estacionaria y q el número de términos de promedios móviles. Así, una serie de tiempo ARIMA $(2, 1, 2)$ tiene que diferenciarse una vez ($d = 1$) antes de que se haga estacionaria, y la serie de tiempo estacionaria (en primeras diferencias) se modela como un proceso ARMA $(2, 2)$, es decir, tiene dos términos AR y dos términos MA. Desde luego, si $d = 0$ (es decir, si para empezar la serie es estacionaria), ARIMA $(p, d = 0, q) = \text{ARMA}(p, q)$. Observe que un proceso ARIMA $(p, 0, 0)$ significa un proceso estacionario AR(p) puro; un ARIMA $(0, 0, q)$ significa un proceso estacionario MA(q) puro. Con los valores de p, d y q se sabe de qué proceso se está haciendo el modelo.

2.2.5 Metodología de Box-Jenkins Al ver una serie de tiempo y determinar si es AR; MA, ARMA o ARIMA La metodología Box-Jenkins resulta útil a la hora de proceder. El método considera cuatro pasos:

1. Identificación. Es decir, encontrar los valores apropiados de p, d y q . Para este paso se basa en la información que muestran las tendencias del correlograma y el correlograma parcial.
2. Estimación. Tras identificar los valores apropiados de p y q , la siguiente etapa es estimar los parámetros de los términos autorregresivos y de promedios móviles incluidos en el modelo. Algunas veces, este cálculo se efectúa mediante mínimos cuadrados simples, pero otras hay que recurrir a métodos de estimación no lineal (en parámetros).

3. Examen de diagnóstico. Después de seleccionar un modelo ARIMA particular y de estimar sus parámetros, se trata de ver si el modelo seleccionado se ajusta a los datos en forma razonablemente buena, pues es posible que exista otro modelo ARIMA que también lo haga. Es por esto que el diseño de modelos ARIMA de Box-Jenkins es un arte más que una ciencia; se requiere gran habilidad para seleccionar el modelo ARIMA correcto, Por tanto, la metodología Box-Jenkins es un proceso iterativo.
4. Pronóstico. Una razón de la popularidad del proceso de construcción de modelos ARIMA es su éxito en el pronóstico. En muchos casos, los pronósticos obtenidos por este método son más confiables que los obtenidos de modelos econométricos tradicionales.

Figura 2.1.: Diagrama de flujos para el método Box - Jenkins



Fuente: (Gujarati & Porter, 2010)

CAPÍTULO 3

PREDICCIÓN DEL PRECIO DE BOLSA

En este capítulo se presenta de forma concisa y directa, los pasos dados en el capítulo inmediatamente anterior, en la sección 2.2.5 para llegar a la predicción del precio de bolsa de energía eléctrica, mediante el uso de los modelos autoregresivos integrados de media móvil, es decir, un modelo ARIMA, gran parte de los comandos utilizados se encuentran en el libro (Cryer & Chan, 2008). Por medio del uso de este modelo se quiere generar una predicción de 48 horas futuras a partir del último dato utilizado para realizar el modelado. El manejo dado al capítulo es el siguiente: En la sección inicial de este capítulo, se describe de manera corta la herramienta software utilizada para el análisis y modelado de los datos a trabajar, la cual se trata del software libre R y se enfatiza en el uso de la misma para el modelado en este capítulo. En la segunda parte del capítulo se presenta toda la información de los datos con los que se trabaja y las ventanas de tiempo en las que se dividen estos datos para determinar el mejor modelo a escoger. En la tercera parte del capítulo se especifica el modelo a utilizar, es decir, partiendo de que el modelo es un $ARIMA(p,d,q)$, se hallan los valores de p , d y q del mismo. La cuarta parte del capítulo lleva al lector a la estimación de los parámetros del modelo, los valores ϕ y θ que acompañan a los distintos ordenes p y q hallados en la sección tres. En la quinta parte se hace un diagnóstico del modelo para determinar si lo que se lleva hasta ahora en la tercera sección es válido o de lo contrario hay que mejorar el modelo. Por último, en la sexta sección del capítulo se realiza el pronóstico del precio de las siguientes 48 horas teniendo en cuenta el modelo hallado, corroborado y por ende aceptado para tal propósito.

3.1 MANEJO DE LOS DATOS

La cantidad de datos por hora que se tienen en este trabajo brindan la oportunidad de analizar distintas ventanas de tiempo, las cuales se dividen en: cinco años, un año, seis meses, tres meses, un mes y una semana. Estos intervalos brindan un poder comparativo para hallar el modelo que más se ajuste a la predicción deseada, es decir, que tenga la menor cantidad de errores y que los datos de predicción se encuentren, preferiblemente, dentro de una franja de certidumbre que es manejada por el error como tal del modelo. Como la cantidad de datos de las 24 horas del día durante cinco años es cuantiosa, 43.824 en total, los mismos se encuentran en un archivo adjunto a este trabajo.

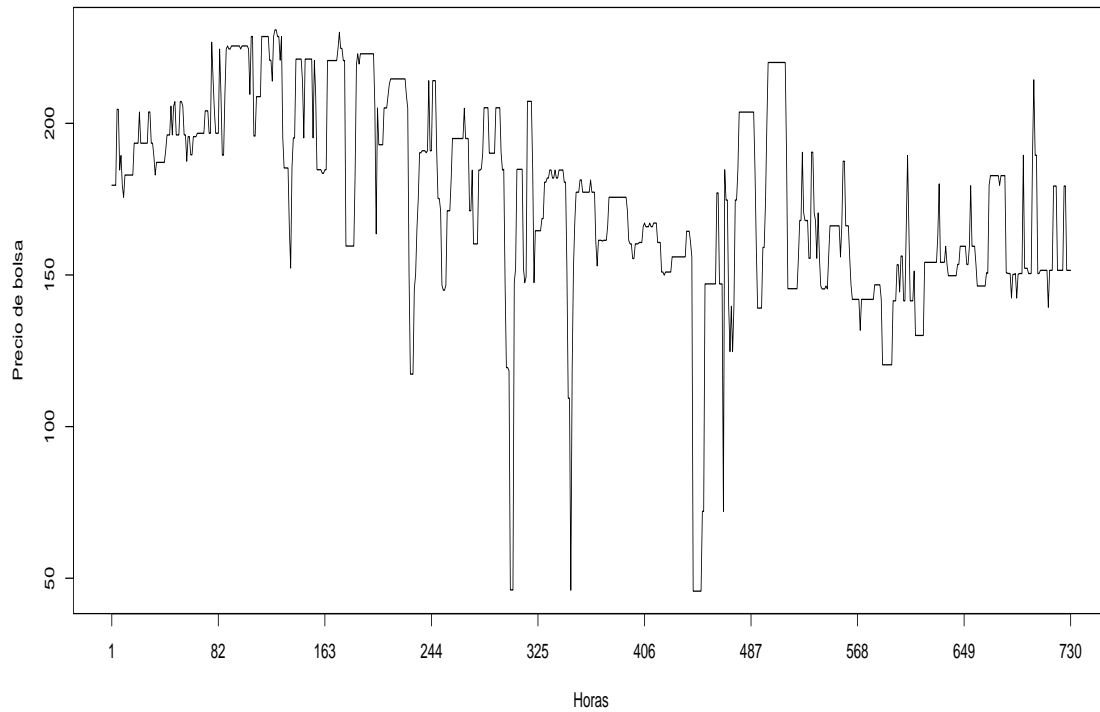
Para mayor entendimiento del trabajo que se realiza en este proyecto, durante todo el capítulo se toma como ejemplo la ventana de tiempo de un mes de datos, a la cual se le hace todo el análisis respectivo en R con el script que se muestra en el anexo a este capítulo.

3.2 ESPECIFICACIÓN DEL MODELO

En este paso, lo que se desea hallar son los órdenes p , d y q del proceso que determinan el comportamiento de el modelo $ARIMA(p,d,q)$ que se va a utilizar. Lo primero es determinar si el comportamiento de los datos es estacionario o no estacionario, para eso se analiza la correlación que hay entre ellos, lo cual se hace con ayuda de la primera parte del script en los siguientes pasos:

1. Graficar los datos para analizar su comportamiento estocástico.

Figura 3.1.: Análisis del comportamiento de un mes de datos mediante el uso de R.



Mes de datos comprendido entre el primero y el treinta y uno de Diciembre del año 2014

2. Una vez se analice el comportamiento y se determine que no es un comportamiento estacionario, se examinan los correlogramas de correlación simple y parcial para confirmar la no estacionalidad de los datos. Estos correlogramas dan el orden del modelo, siendo la correlación simple (Figura 3.2) la que dé el orden de la parte de los MA del modelo y la correlación parcial (Figura 3.3) la parte de los AR. Se observa en la Figura 3.3 la posibilidad de trabajar con la correlación parcial directa de los datos para hallar la cantidad de AR que expresan el modelo, pero ya que el gráfico de correlación simple se encuentra tan diverso, es necesario suavizar este último con una diferencia de los datos y con esto llegar a un modelo ARIMA.

Figura 3.2.: Gráfico de correlograma simple para un mes de datos

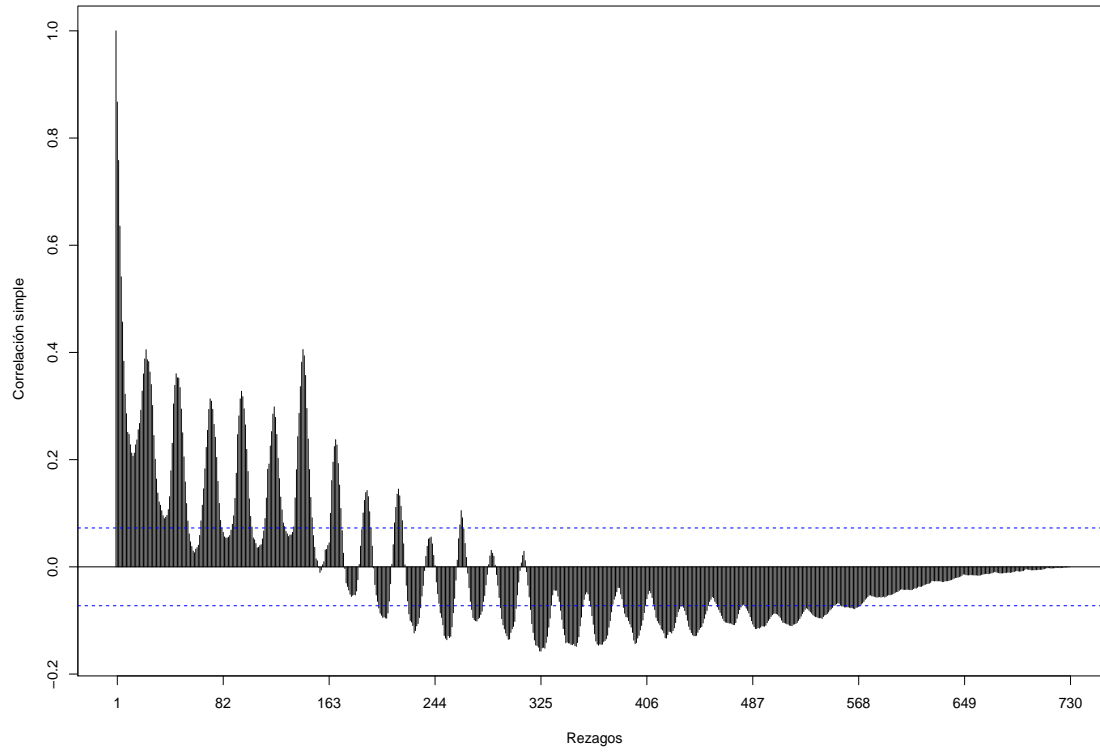
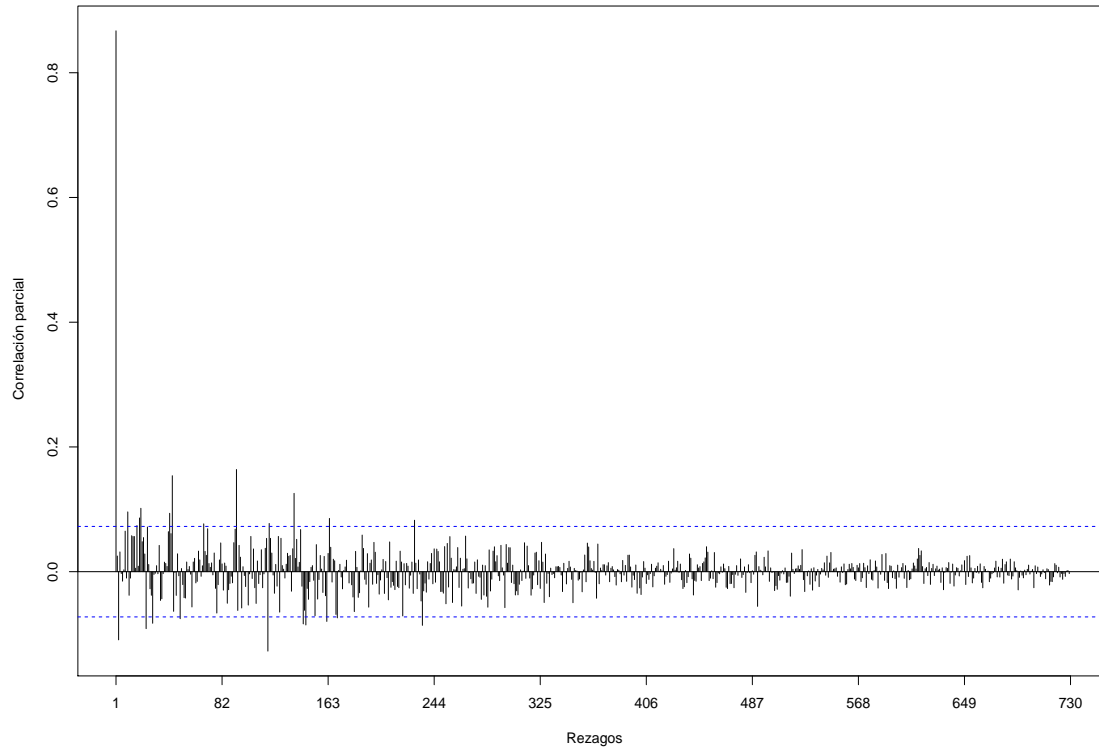


Figura 3.3.: Gráfico de correlograma parcial para un mes de datos



3. Luego del análisis es necesario suavizar los datos, el siguiente paso es pasar de un modelo ARMA a un modelo ARIMA, es decir, aplicar una o dos diferencias a los datos hasta que sean estacionarios y se puedan sacar de ellos el orden del modelo sobre el cual se está trabajando. En este punto, se aplica una diferencia a los datos y se grafica su comportamiento referente a la correlación entre ellos, es decir, hallar nuevamente dos correlogramas, simple y parcial, pero esta vez de la diferencia de los datos y con ello sel análisis del comportamiento de los mismos. Se observa en la Figura 3.4 los datos arrojados por la correlación simple, siendo ya unos datos estacionarios con los cuales se puede trabajar; también, aunque no mejor que la correlación parcial sin diferencia, se ve como los datos de la Figura 3.5 llevan a analizar la estacionalidad de los datos en correlación parcial pero ya siendo estacionario todo el modelo de datos a trabajar. Aunque ya se puede trabajar con la primera diferencia de los datos, es decir, con un modelo $ARIMA(p,1,q)$, es necesario corroborar, por medio de una segunda diferencia, que aceptamos el primer orden de diferencia del modelo.

Figura 3.4.: Gráfico de correlograma simple en primera diferencia para un mes de datos.

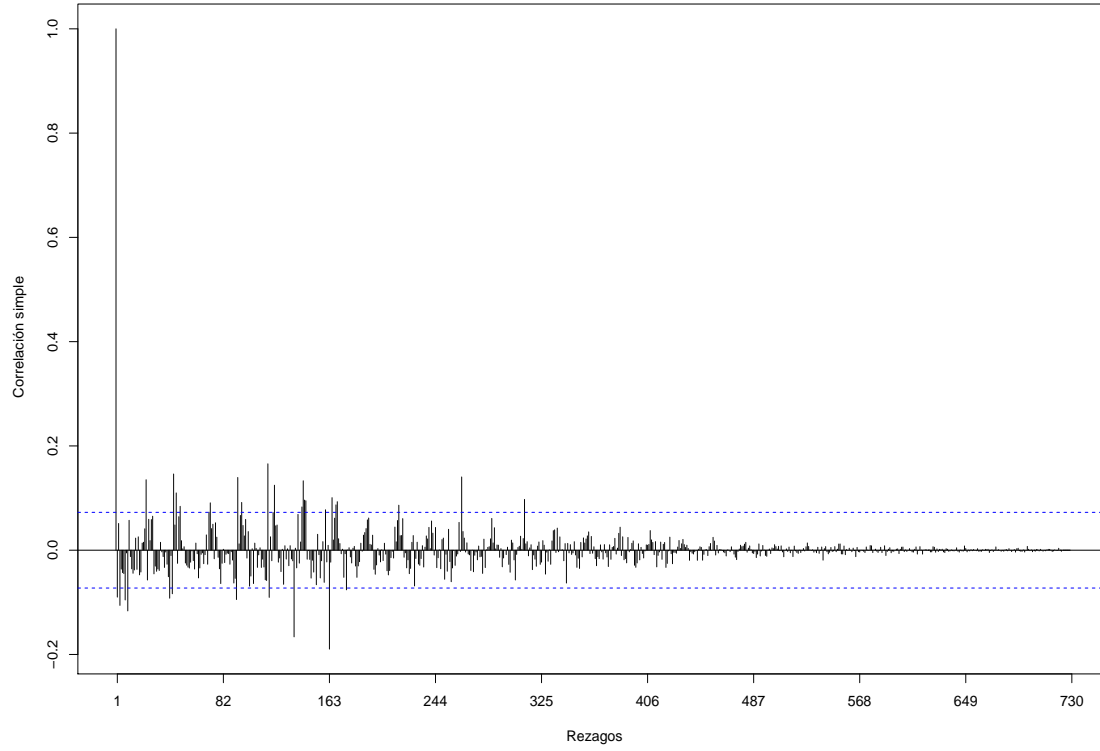
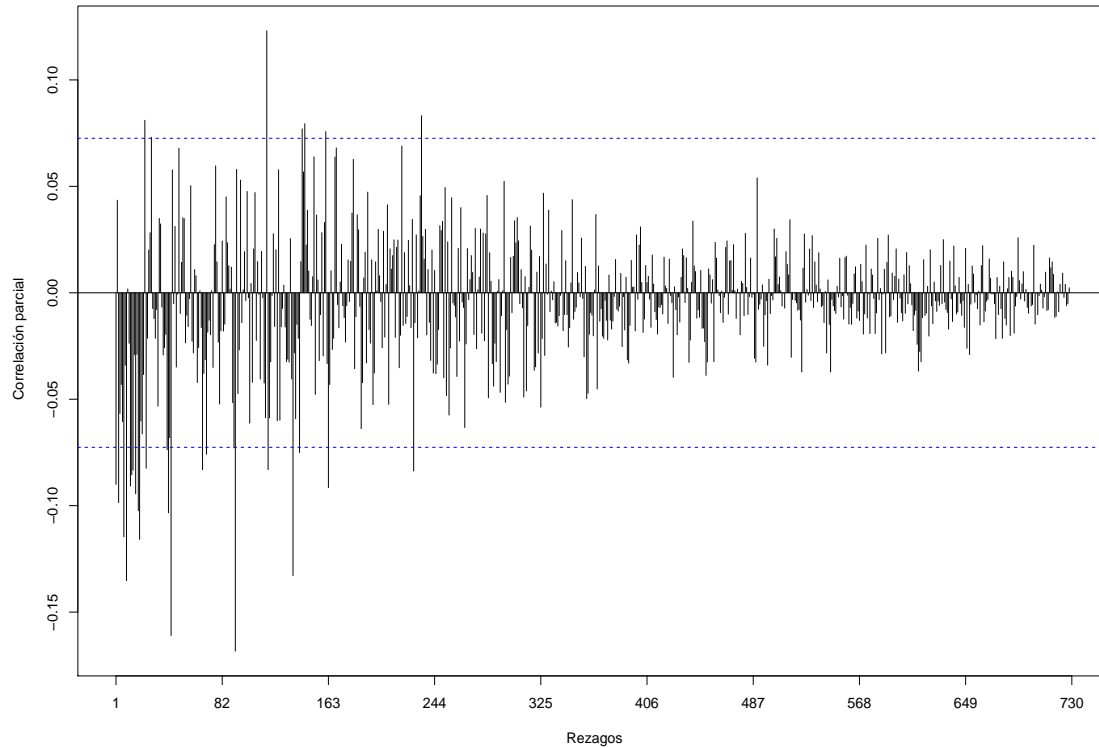


Figura 3.5.: Gráfico de correlograma parcial en primera diferencia para un mes de datos.



4. Ya en este paso se va a, como se mencionó en el paso anterior, corroborar por medio de una segunda diferencia de los datos, que se va a trabajar con el modelo $ARIMA(p,1,q)$. Observando las gráficas se puede tomar una decisión, teniendo en cuenta el gran desmejoramiento del correlograma en la parte de la correlación simple (Figura 3.6) con respecto al de la primera diferencia y tomando el hecho de que, aunque la correlación parcial mejora un poco, es necesario que ambas correlaciones presenten un modelo estacionario con el que se pueda trabajar, podemos definir como modelo el $ARIMA(p,1,q)$ para el ejemplo de ventana de tiempo de un mes. Se debe aclarar que este análisis se hizo para todas las ventanas de tiempo y en todos los casos se obtuvo un modelo de primera diferencia.

Figura 3.6.: Gráfico de correlograma simple en segunda diferencia para un mes de datos.

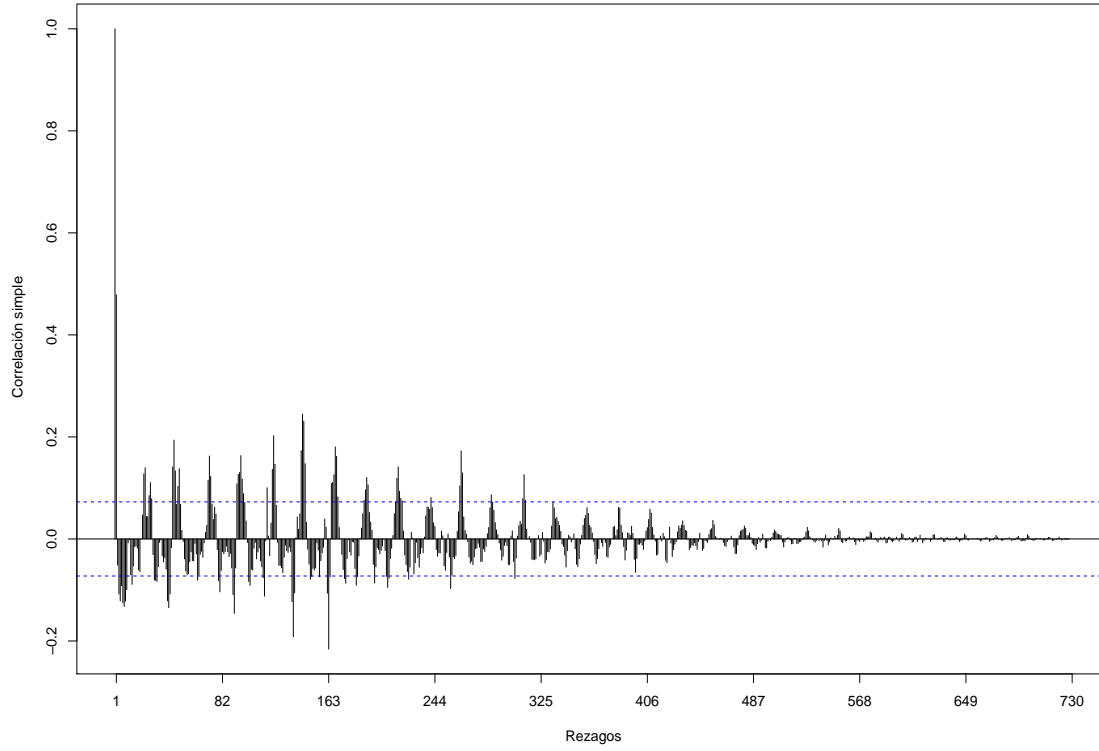
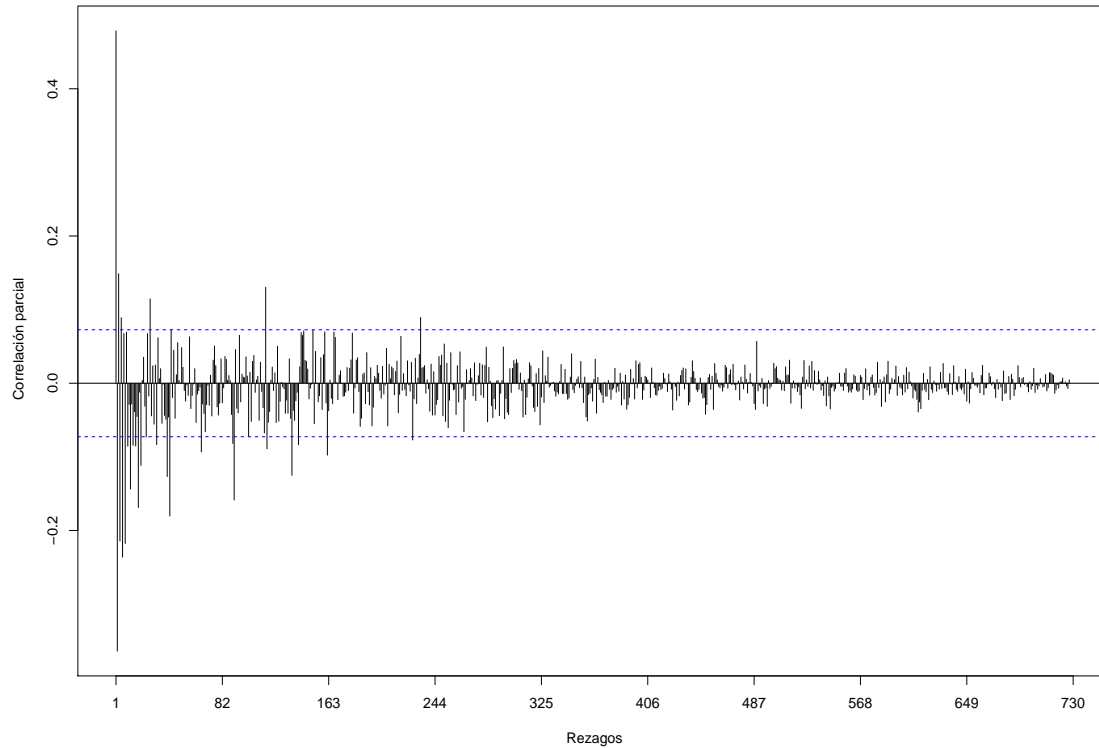
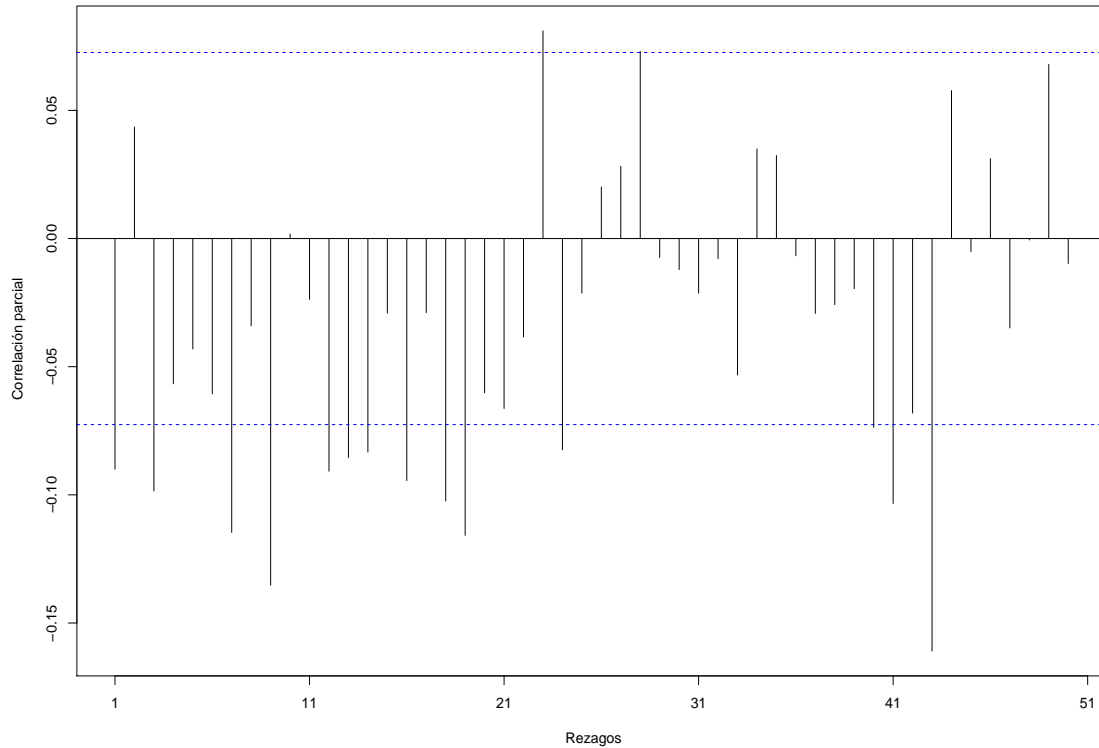


Figura 3.7.: Gráfico de correlograma parcial en segunda diferencia para un mes de datos.



5. Luego de escoger la diferencia adecuada para el modelo, se observa más de cerca los correlogramas y se escoge el orden de líneas que salen de los intervalos de Barlet, es decir, se toman todos los datos hasta antes de entrar en los intervalos, cuando ya entran, los que salen de nuevo en el sentido contrario a los primeros no se tienen en cuenta. Teniendo esto claro, se puede observar entonces que, es la primera línea la cual sale en el correlograma de correlación simple, de ahí, las siguientes líneas se encuentran dentro de los intervalos y la siguiente línea que sale está en el sentido contrario a la primera, por ende, no se tiene en cuenta y se puede afirmar que; hay un MA(1). De manera paralela, se analiza el correlograma de correlación parcial y, al haber datos tan diversos y cercanos, se determina la necesidad de un acercamiento (Figura 3.8); en este podemos notar que tenemos los AR: 7, 9, 12, 13, 16, 18, 19 y 20; con lo cual se define para el modelo general que tenemos un AR(20).

Figura 3.8.: Gráfico de acercamiento para correlograma parcial en primera diferencia para un mes de datos.



6. Al definir la diferencia que se aplica a los datos y el número máximo de q y p , se define como modelo para nuestro análisis un $ARIMA(20,1,1)$. Es así que, al tener claro como se llega a este punto y al definir el modelo a trabajar, el paso siguiente es la estimación de los parámetros que acompañan cada p y q del modelo.

3.3 ESTIMACIÓN DE PARÁMETROS

En esta parte del proceso, ya se tiene claro el modelo a utilizar, para el caso de la ventana de tiempo de un mes, el modelo seleccionado es $ARIMA(20,1,1)$, se debe a continuación, estimar los parámetros del proceso, es decir, los ϕ y θ que acompañan en la ecuación a cada p y q , siendo los ϕ los que acompañen a la parte de los p , es decir, a la parte de los AR y, los θ los que acompañen a las q , siendo estos los MA.

Para lograr tal fin, se emplea la herramienta software R. Con el objetivo de suavizar un poco más la muestra de datos y que la varianza sea lo más cercano a cero posible, se aplica

un logaritmo a la variable de datos. En el siguiente cuadro se observan todos los parámetros que acompañan a todos los valores p y q.

TABLA 3.1.: Parámetros arrojados por R para un mes de datos

AR	1	2	3	4	5	6	7	8	9
ϕ	0.3933	0.0359	-0.1546	-0.0449	-0.0542	-0.0871	-0.1097	-0.0132	-0.1520
s.e.	0.2401	0.0538	0.0406	0.0580	0.0509	0.0500	0.0539	0.0598	0.0472
AR	10	11	12	13	14	15	16	17	18
ϕ	0.0358	-0.0612	-0.1073	-0.0680	-0.0501	-0.0162	-0.0872	0.0077	-0.0938
s.e.	0.0630	0.0446	0.0453	0.0546	0.0524	0.0490	0.0438	0.0498	0.0409
AR	19	20							
ϕ	-0.0607	-0.0070		MA	1				
s.e.	0.0492	0.0545		θ	-0.5527				
				s.e.	0.2374				

Es importante recalcar que solo se toman los datos de los AR y MA que sobresalen de los límites generados por los intervalos de Bartet, para hacer más énfasis en esto, se realiza un prueba de significancia para cada parámetro hallado, la cual consiste en dividir el valor del parámetro sobre su error y si el resultado es mayor igual que dos entonces el parámetro es significativo.

TABLA 3.2.: Resultado de prueba de significancia para parámetros de un mes de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	0.3933	0.2401	1.6380	-0.5527	0.2374	2.3281
2	0.0359	0.0538	0.6673			
3	-0.1546	0.0406	3.8078			
4	-0.0449	0.0580	0.7741			
5	-0.0542	0.0509	1.0648			
6	-0.0871	0.0500	1.7420			
7	-0.1097	0.0539	2.0352			
8	-0.0132	0.0598	0.2207			
9	-0.1520	0.0472	3.2203			
10	0.0358	0.0630	0.5682			
11	-0.0612	0.0446	1.3821			
12	-0.1073	0.0453	2.3686			
13	-0.0680	0.0546	1.2454			
14	-0.0501	0.0524	0.9561			
15	-0.0162	0.0490	0.3306			
16	-0.0872	0.0438	1.9908			
17	0.0077	0.0498	0.1546			
18	-0.0938	0.0409	2.2933			
19	-0.0607	0.0492	1.2337			
20	-0.0070	0.0545	0.1284			

Es importante resaltar que con los parámetros más significativos se determina la fórmula final, la cual, para el caso del ejemplo de la ventana de un mes es:

$$y(t) = 0,1546y(t-3)+0,1097y(t-7)+0,152y(t-9)+0,1073y(t-12)+0,0938y(t-18)+0,5527e(t-1)$$

3.4 DIAGNÓSTICO DEL MODELO

Existen varias maneras de examinar si el modelo hallado es el correcto, este análisis se hace de una manera sencilla en dos grupos: el primero está dado por los valores likelihood (máxima verosimilitud), AIC y varianza que arroja el modelo al momento de hacer la estimación de parámetros. El segundo grupo corresponde a dos pruebas realizadas al modelo para analizar el comportamiento del mismo; la primera de ella es la prueba de cuantil-cuantil y la segunda la prueba de Ljung-box.

3.4.1 Valores de Likelihood, AIC y Varianza Son los valores arrojados por R al momento de utilizar el comando `arima()` y que se interpretan de la siguiente manera:

Likelihood: El valor de la máxima verosimilitud brinda uno de las tres bases para aprobar el modelo. Es sencillo determinar que el valor es acertado, teniendo en cuenta que este, mientras mayor sea, brinda una mejor confiabilidad del modelo, de esta forma se deja claro que, este número no puede dar cero y mucho menos arrojar valores negativos.

AIC: Llamado también criterio de información de Akaike, tiene como propósito ayudar a escoger el modelo con menor valor de AIC arrojado, teniendo como valor aceptado un número mucho menor a cero, es decir, un modelo cuyo valor de AIC sea mayor a aproximadamente 6, es un modelo no aceptado.

Varianza: Es importante que la varianza del modelo sea cercana a cero, para la aprobación del mismo, utilizaremos como criterio que la varianza (σ^2) que arroja R sea menor a 0.05.

Al tomar los criterios anteriormente mencionados, se procede a analizar los valores arrojados por R para la ventana de tiempo de 1 mes:

TABLA 3.3.: Valores de σ^2 , log likelihood y AIC arrojados por R para un mes de datos

σ^2	Log likelihood	AIC
0.01282	552.8	-1061.59

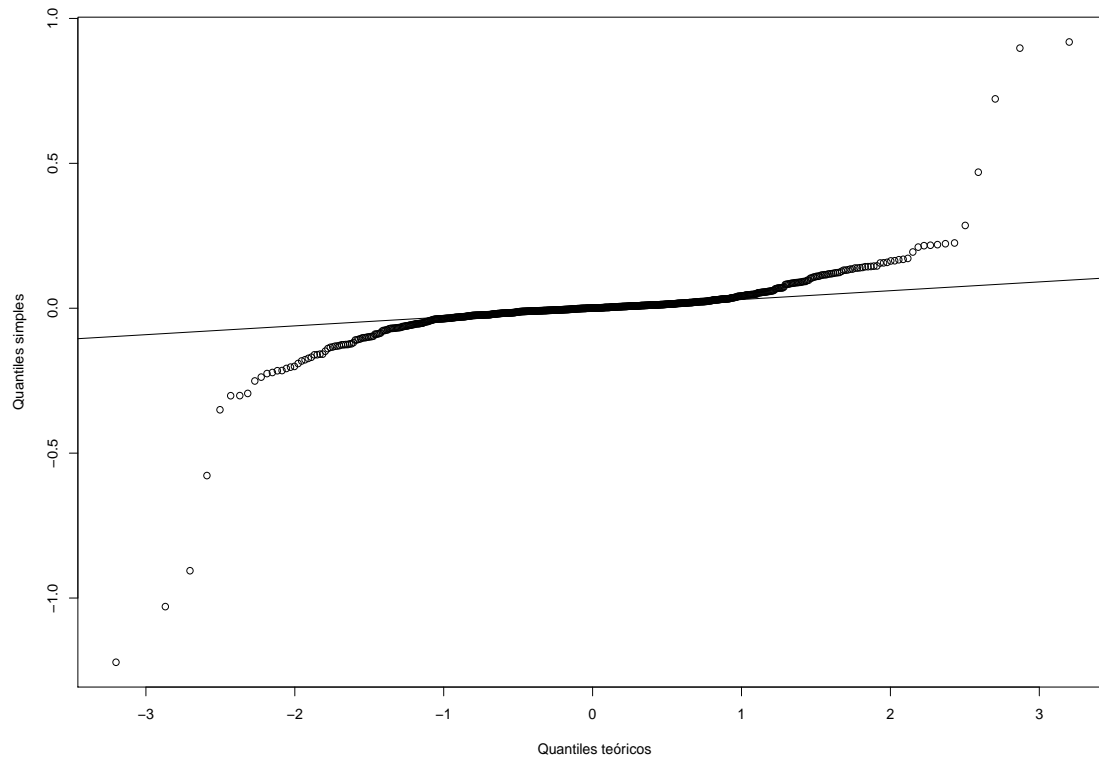
Al observar detalladamente cada dato, se identifica claramente lo siguiente:

- El valor de la varianza (σ^2), es claramente menor a 0.05, por lo que se acepta el modelo desde este punto de aprobación.
- El valor de likelihood es por mucho mayor a cero, por lo tanto, el modelo se acepta por este medio de prueba.
- El valor de AIC arrojado por R es mucho menor a cero, de manera inmediata ese acepta el modelo gracias al análisis dado por esta prueba.

3.4.2 Prueba quantil-quantil y prueba de Ljung-box Estas pruebas se utilizan para determinar el estado de los residuos del modelo, es decir, el objetivo es definir si los residuos replican el comportamiento de un ruido blanco, siendo esto que, su media es próxima a cero y su desviación estándar constante(Casimiro, 2009) .

Prueba quantil-quantil. Al aplicar la prueba al modelo obtenido en la sección 1.3 tenemos el siguiente resultado:

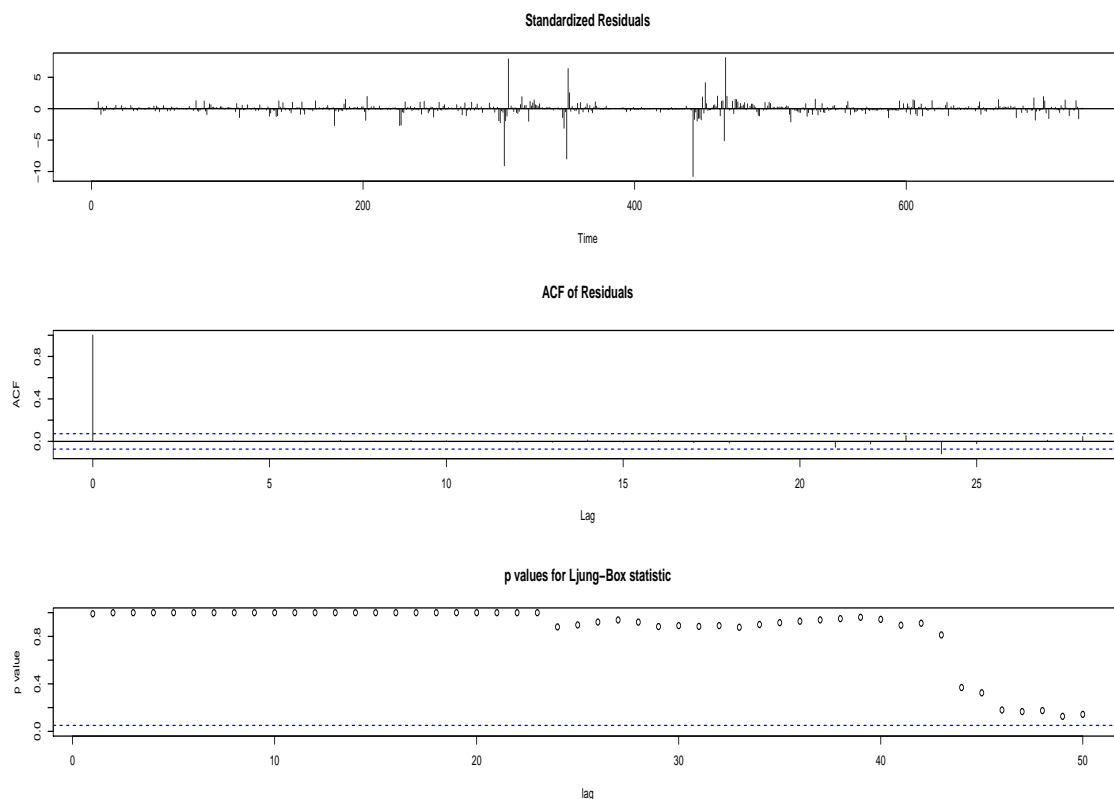
Figura 3.9.: Resultado de prueba de q.q para un mes de datos



La gráfica muestra que el modelo funciona de manera eficiente para datos cercanos a la media.

Prueba de Ljung-box. Al aplicar el comando respectivo a la prueba en R, se tiene el resultado de la prueba de Ljung-box.

Figura 3.10.: Resultado de prueba de Ljung-box para un mes de datos



En esta prueba se observa la correlación que hay entre los residuos, siendo ésta cercana predominantemente para el primer dato residual, es decir, su comportamiento depende a gran medida del comportamiento de los valores pasados, lo cual muestra que el modelo se puede ajustar un poco más para que la correlación residual baje.

3.5 PRONÓSTICO

Ya establecido el modelo y sus parámetros, el paso a seguir es el pronóstico del precio de bolsa de la energía eléctrica las siguientes 48 horas, para lo cual se usa el comando que brinda R. Vale aclarar que, ya que se toma el logaritmo de los datos para suavizar el modelo y con esto hallar los parámetros del mismo, la predicción que arroja el R está dada en valores menores a los deseados para la predicción, por lo cual, al finalizar la predicción, se le aplica una función exponencial a los datos arrojados para así cancelar el efecto del logaritmo y deje ver los valores que realmente predice. Se pasa entonces a aplicar el comando que brinda la

siguiente predicción:

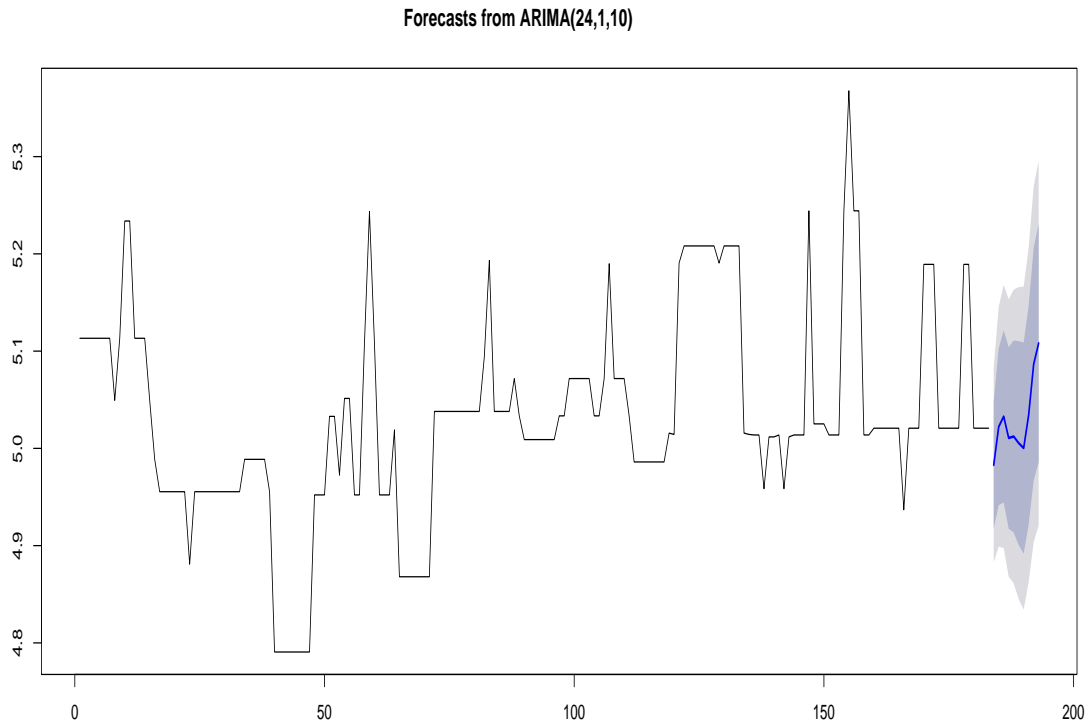
TABLA 3.4.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para un mes de datos.

[1 - 8]	5.0188	5.0212	5.0268	5.0314	5.0284	5.0462	5.0256	5.0327
[9 - 16]	5.0511	5.0702	5.0703	5.0711	5.0655	5.0515	5.0534	5.0639
[17 - 24]	5.0663	5.0643	5.0612	5.0597	5.0567	5.0501	5.0479	5.0433
[25 - 32]	5.0419	5.0432	5.0441	5.0428	5.0420	5.0433	5.0458	5.0486
[33 - 40]	5.0517	5.0530	5.0535	5.0541	5.0550	5.0557	5.0555	5.0557
[41 - 48]	5.0558	5.0556	5.0551	5.0541	5.0527	5.0515	5.0506	5.0500
[1 - 8]	0.1132	0.1479	0.1743	0.1889	0.1990	0.2060	0.2106	0.2128
[9 - 16]	0.2149	0.2154	0.2167	0.2178	0.2186	0.2191	0.2194	0.2201
[17 - 24]	0.2206	0.2215	0.2222	0.2227	0.2233	0.2246	0.2267	0.2295
[25 - 32]	0.2328	0.2368	0.2407	0.2447	0.2484	0.2515	0.2541	0.2564
[33 - 40]	0.2584	0.2602	0.2618	0.2631	0.2644	0.2655	0.2667	0.2678
[41 - 48]	0.2689	0.2700	0.2711	0.2723	0.2736	0.2750	0.2766	0.2784

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho.

Es claro notar que cada predicción trae consigo un error, el cual da la posibilidad de que tan lejos se encuentra el dato predicho de el valore real. Para notar de manera más sencilla lo anteriormente mencionado, se utiliza un comando en R, el cual muestra la franja de error en la que es válido el valor arrojado por la predicción y muestra también, una línea que representa a los datos hallados y su comportamiento.

Figura 3.11.: Gráfico de franja de error de predicción para un mes de datos



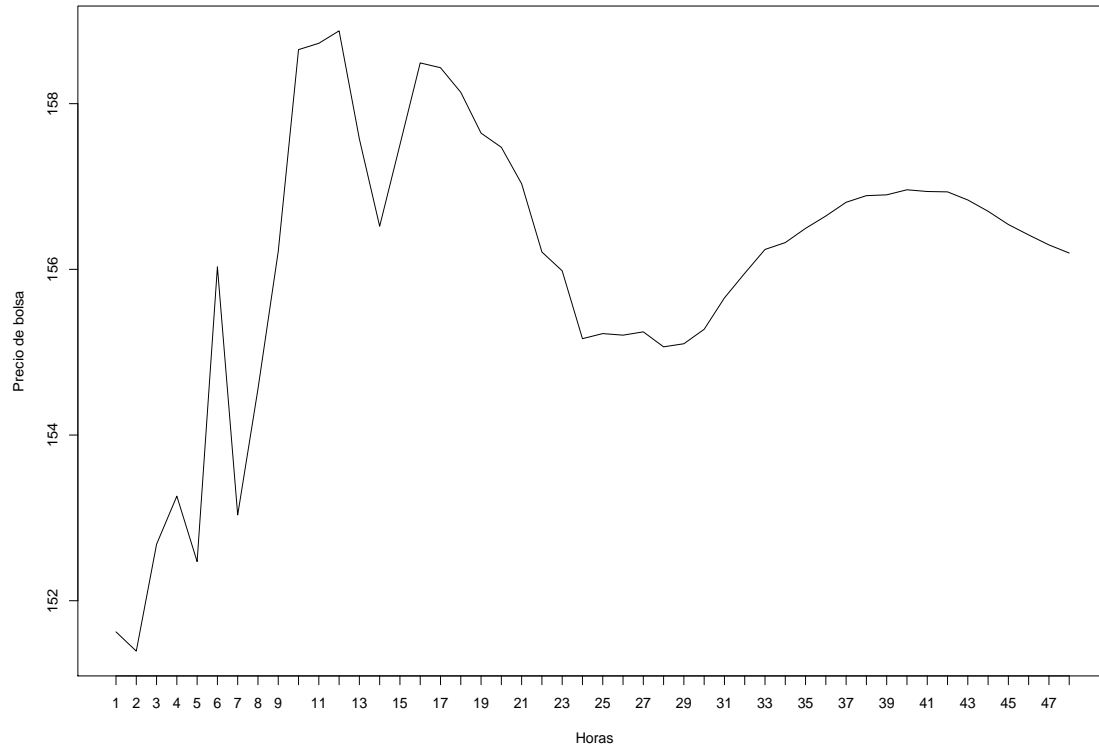
Ya hallados los datos de la predicción se procede a aplicarles exponente para observar su valor real, esto se hace de manera sencilla con los comandos que brinda R para tal fin, paso siguiente se exportan los datos a un archivo de texto plano con lo cual se observa el resultado final:

TABLA 3.5.: Resultados finales de datos de predicción.

1	151.624034566637	17	158.434142145192	33	156.240267706779
2	151.391924387476	18	158.136882040849	34	156.322940673467
3	152.681797378315	19	157.644579141709	35	156.494834023362
4	153.262785442705	20	157.473155111186	36	156.644044687775
5	152.4724615895	21	157.033632162102	37	156.809704979051
6	156.030265804651	22	156.209491405606	38	156.889698325159
7	153.035206076885	23	155.982059900713	39	156.899425788004
8	154.55638790326	24	155.163282821293	40	156.960628497814
9	156.217458292821	25	155.224739648939	41	156.9397541224
10	158.65245623368	26	155.205648180111	42	156.935359870805
11	158.72862769236	27	155.245541157722	43	156.837305915938
12	158.879173778983	28	155.064785412938	44	156.70091679766
13	157.574127761889	29	155.10200542766	45	156.541163350635
14	156.52018824011	30	155.276438094555	46	156.415511252534
15	157.50402287453	31	155.655308239956	47	156.295117666349
16	158.491822670152	32	155.952426124629	48	156.196838949233

Y utilizando el comando `plot()` obtenemos un gráfico de datos de predicción, con el cual vemos que no hay gran variabilidad entre ellos, por lo que descartamos esta ventana de tiempo para realizar las pruebas de optimización de la micro red.

Figura 3.12.: Gráfico de valores de predicción de precio de 48 horas futuras para un mes de datos



A manera de conclusión, ya que se tiene un análisis similar a este para las otras 5 ventanas de tiempo, análisis mostrado junto con gráficos y datos en el anexo de este capítulo, se toma como ventana adecuada, cuyos datos sirven de una mejor manera, para comprobar si un modelo de predicción ARIMA contribuye a la optimización de una micro red, a la ventana de una semana de datos, ya que en los datos de predicción se nota un comportamiento más adecuado a un perfil de costos de precio de energía eléctrica en un día normal. También se toma el hecho importante de que con una semana de datos es suficiente para la predicción de 48 horas futuras.

CAPÍTULO 4

ANÁLISIS DE DATOS

En el presente capítulo se presenta un análisis del comportamiento de costos de energía eléctrica, tomados de diez distintas ventanas de tiempo, en las cuales se tiene una comparación de resultados de datos reales contra datos de predicción. El capítulo se divide en cuatro secciones, en la primera de ellas se analiza el comportamiento de las ventanas de tiempo tomadas, dejando claro por qué se escogen las mismas y, el modelo ARIMA con el que se va a trabajar en ellas. En la segunda parte del capítulo se analiza el impacto de la variabilidad del precio de la energía eléctrica en cada ventana de tiempo sobre los costos de energía de una micro-red, así también, se toma en cuenta el error absoluto entre los datos reales y de predicción para aceptar el modelo tomado para la optimización. En la tercera parte del capítulo se realiza un análisis del error, más que nada, del por qué en algunas ventanas de tiempo no se logra realizar la predicción correcta para la optimización, de la misma manera, cómo perturba el comportamiento climático al precio de la energía eléctrica, afectando esto al modelo ARIMA y su predicción.

4.1 VENTANAS DE TIEMPO

El tratamiento dado en el capítulo inmediatamente anterior a la base de datos en general, ha dado como resultado que, la ventana de tiempo que mejor se comporta para el análisis de costos, a base de una predicción de 48 horas, es la de una semana de datos. Teniendo en cuenta lo anterior, en este capítulo se toman diez ventanas de tiempo de una semana a través de los cinco años de datos, es decir, se halló para diez distintas ventanas de tiempo, un modelo ARIMA con el cual realizar una predicción de 48 horas y, comparar los datos de costos de predicción contra los datos de costos de datos reales para cada semana y, ya con esto hallar el error absoluto para determinar la veracidad de los modelos ARIMA para la optimización de una micro-red. Las semanas tomadas para hallar los modelos ARIMA son las siguientes:

- Semana 1: Comprende los días entre el 18 y 24 de Julio del 2014
- Semana 2: Comprende los días entre el 19 y 25 de Junio del 2014
- Semana 3: Comprende los días entre el 10 y 16 de Agosto del 2013
- Semana 4: Comprende los días entre el 31 de Octubre y el 6 de Noviembre del 2013
- Semana 5: Comprende los días entre el 25 de Julio y el 1 de Agosto del 2010
- Semana 6: Comprende los días entre el 25 de Noviembre y el 1 de Diciembre del 2012
- Semana 7: Comprende los días entre el 25 y 31 de Octubre del 2013
- Semana 8: Comprende los días entre el 10 y 16 de Octubre del 2014
- Semana 9: Comprende los días entre el 1 y 7 de Diciembre del 2013
- Semana 10: Comprende los días entre el 26 de Octubre y el 1 de Noviembre del 2013

Los datos de predicción de cada ventana, junto con los datos reales de 48 horas siguientes de cada semana y un gráfico comparativo entre ambos, datos de predicción y reales, se encuentran en un anexo a este capítulo.

4.2 ANÁLISIS DE COSTOS DE ENERGÍA EN UNA MICRO-RED

En esta sección se plantea, a grandes rasgos, la utilidad de conocer de antemano, el valor de las horas futuras del precio de bolsa de energía, a partir de una hora de la semana, para la optimización de una micro red, enfocándose, más que todo en la comparación del costo de

operación de datos reales y datos de predicción. Para el análisis, se evalúa el comportamiento del costo de operación de la micro-red, primero, con los datos que brinda XM para el precio de energía 48 horas después de la semana tomada como ventana, posterior a esto, se evalúan los datos arrojados por R como predicción de 48 horas futuras para la ventana de una semana escogida. Estos costos están basados en el proceso de optimización lineal de una micro-red, los cuales se calculan mediante la herramienta software python, con ayuda de las librerías Cvxopt, Sympy y Cvxpy. Los cálculos dan como resultado de cuantificar la cantidad de potencia que debe inyectar el almacenamiento, es decir, se toma la cantidad de potencia total inyectada por cada uno de los precios de las 48 horas futuras a la ventana de tiempo, ya sean las reales o los valores de predicción, lo que da como resultado el costo total con el que se realiza el análisis. La descripción del problema de optimización a resolver se puede encontrar en (Serna-Suarez *et al.*, 2015).

En relación a lo anteriormente mencionado, se presentan los resultados de comparación de costos basados en precio real contra los costos basados en precio de predicción y, el error que hay entre los dos valores.

TABLA 4.1.: Tabla de comparación de errores entre costos reales y de predicción de 10 ventanas de tiempo de una semana de datos

Semana	Costos datos reales	Costos datos predicción	% Error
1	1288133.159	1220677.07	5.2367
2	1231400.946	1954022.546	58.6829
3	1141714.774	1192779.677	4.4726
4	1524839.79	1888245.461	23.8324
5	535518.5115	549040.6525	2.5251
6	853687.4728	1008286.433	18.1096
7	2648999.609	2724932.097	2.8665
8	1423033.356	1424015.52	0.0690
9	1283724.962	1289426.731	0.4442
10	1971895.731	3047817.528	54.5628

Este error se halla mediante la expresión:

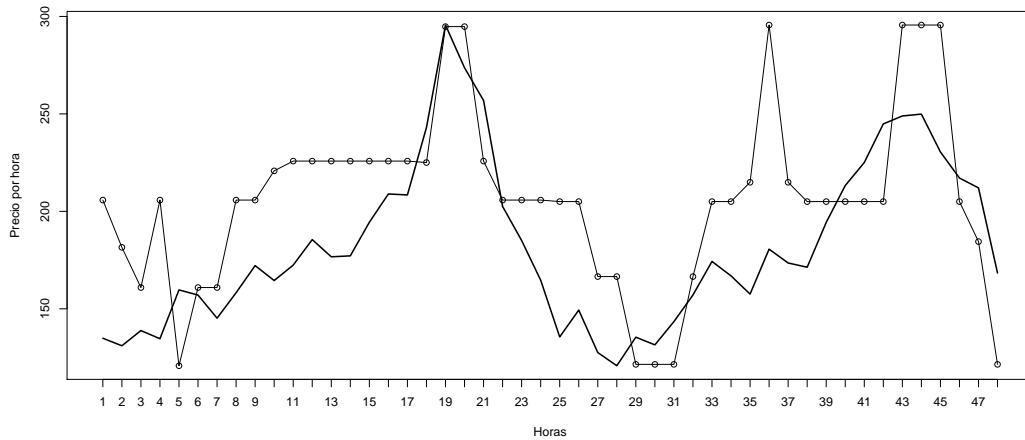
$$\text{Error} = \frac{\text{Costo real} - \text{Costo predicción}}{\text{Costo real}} \times 100 \%$$

El valor del error indica que tan eficiente es el modelo ARIMA para pronosticar los precios de bolsa de la energía y, si es aceptado para la optimización de una micro red. El método de aprobación de los datos de predicción se basan en lo siguiente:

- Si el error es menor al 5 % el modelo es completamente aceptado, ya que la diferencia de

comportamiento del precio de predicción no tiene una variación alta respecto al precio real. como ejemplo de esto, se muestra un gráfico comparativo de precios en la semana 8:

Figura 4.1.: Gráfico comparativo de precios en la semana 8



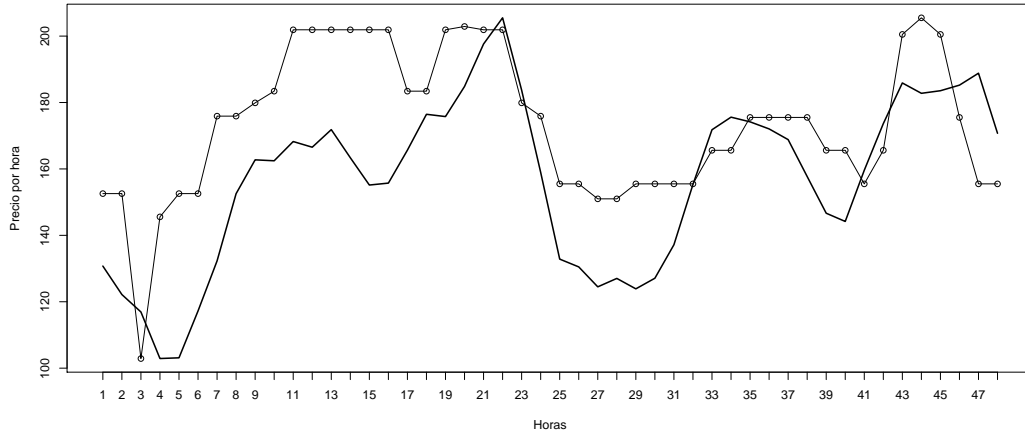
línea continua: datos de predicción

línea con puntos: datos reales

Fuente datos reales: (XM, 2014b)

- Si el error se encuentra entre el 5 % y el 10 % el modelo es igualmente aceptado, ya que cabe la posibilidad de que un cambio en el estado climático del país en esos días haya afectado la predicción. A modo de ejemplo, se muestra un gráfico comparativo de precios en la semana 1:

Figura 4.2.: Gráfico comparativo de precios en la semana 1



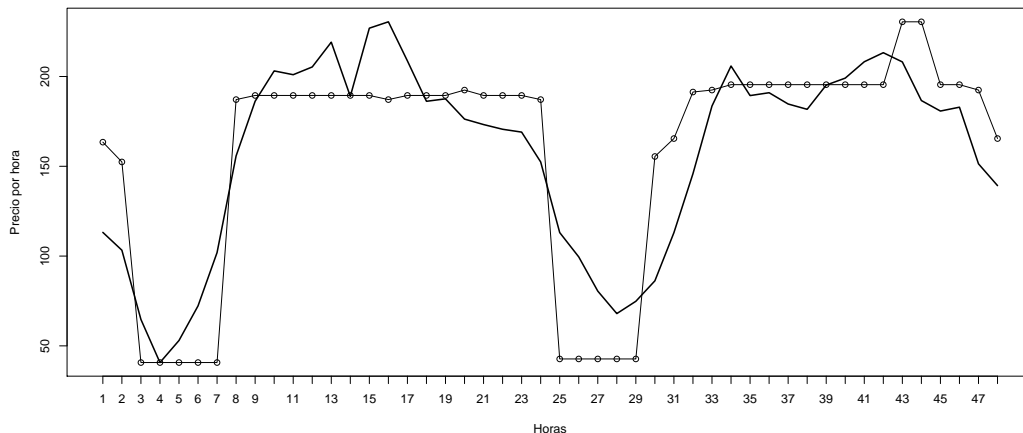
línea continua: datos de predicción

línea con puntos: datos reales

Fuente datos reales: (XM, 2014b)

- Si el error es mayor al 10% el modelo no es aceptado y se debe analizar el por qué del error, lo cual se hace en la siguiente sección. Ejemplo de este comportamiento es la semana 6:

Figura 4.3.: Gráfico comparativo de precios en la semana 6



línea continua: datos de predicción

línea con puntos: datos reales

Fuente datos reales: (XM, 2012)

4.3 ANÁLISIS DE ERROR

En esta sección se analiza por qué algunas ventanas de tiempo producen pronósticos con errores absolutos superiores al 10 % los cuales no son datos significativos, así mismo, se demuestra que dichos errores no se deben a fallas de los modelos si no a los cambios abruptos que modifican de manera negativa la condición de normalidad. En el estudio de error en los costos de predicción y costos reales de la semana 4 comprendida entre los días 31 de octubre y 6 de noviembre del año 2013, se obtiene un error absoluto de 23.83 % el cual es un valor inaceptable a la hora de pronosticar un precio, y aunque se usa la misma metodología para pronosticar el costo de todas las semanas en estudio, el error fuera de rango se presenta por las siguientes razones.

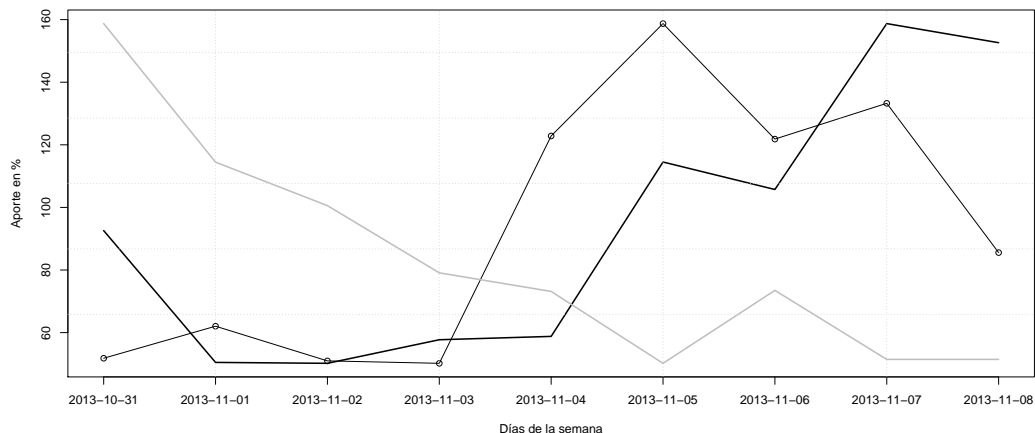
El pronóstico por medio del método ARIMA, consiste básicamente de valores pasados o rezagos, de la diferenciación para convertir la serie de tiempo no estacionaria a estacionaria y de unas medias móviles que amortizan el error de la predicción. Los valores de AR y MA son tomados de los pacf y acf respectivamente, con la diferenciación que exija el proceso de estacionalización de la serie de tiempo, es decir, depende de los valores pasados al pronóstico (rezagos) y de la tendencia que tenga el comportamiento del valor del precio en la semana en estudio, por ende, cualquier alteración ajena al pronóstico como los cambios climáticos o las

crisis financieras que afectan los precios de bolsa, están alterando el valor real de la energía eléctrica y también el valor pronosticado con el promedio de la tendencia.

Una de las causas principales por las que los modelos se ven afectados por el cambio climático, es por la alteración de la normalidad de los residuos, ya que al analizar la prueba de Ljung-box, vista en el capítulo 3, para cada modelo, se observa que en la gráfica de correlación simple de los residuos, ésta tiene un valor predominante, lo cual significa que el valor futuro pronosticado por el modelo depende en gran manera del comportamiento de los rezagos anteriores, por ende, al seguir un comportamiento que depende de los datos inmediatamente anteriores, la condición de normalidad de este comportamiento se ve afectada cuando se presenta un cambio inesperado del clima.

Se investiga acerca del comportamiento climático del país en los meses de octubre y noviembre del año 2013 con los datos históricos del IDEAM (Instituto de Hidrología, Meteorología y Estudios Ambientales) y se encuentra que en promedio el mes de octubre tuvo un déficit de lluvias muy por debajo de la media histórica, y que para el mes de Noviembre las lluvias incrementaron por encima de la media histórica. Aunque ese dato no brinda mucha información para concluir el porqué de la falla de la estimación, si da una pista para profundizar acerca de los aportes hidrológicos y así estudiar si el cambio climático si afecta el pronóstico de manera significativa. Por tal motivo se ingresa a los datos históricos de aportes hidrológicos de XM, para estudiar que pasó en el país exactamente en los días de la semana en estudio. Se analizan los 3 ríos con mayores aportes hídricos en el país, ubicados en 3 zonas distintas y los resultados obtenidos se expresan en la siguiente gráfica de aportes.

Figura 4.4.: Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 4



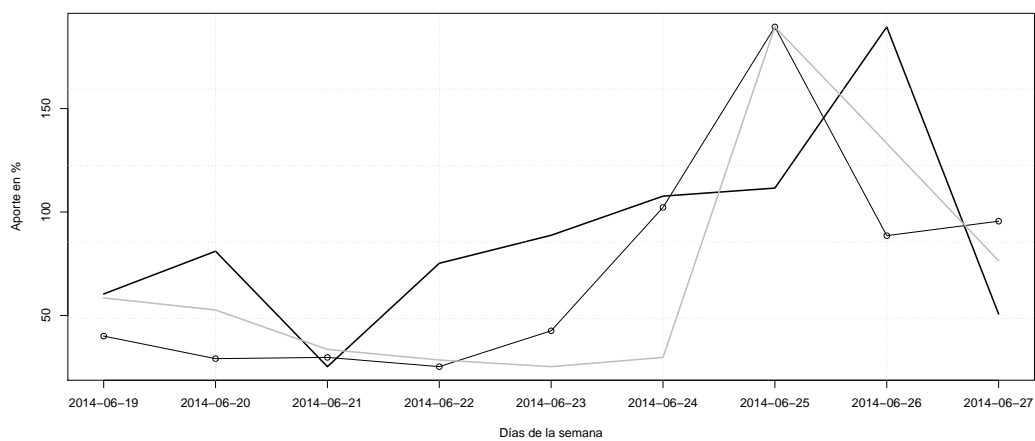
línea continua: Río Bogotá (centro)
 línea con puntos: Río Nare (Antioquia)
 línea gris: Río Altoanchicaya (Valle)

Fuente: (XM, 2014a)

Se estudian 3 ríos ubicados en 3 zonas de influencia de generación eléctrica en el país, en la región de Antioquia, en el centro y sur occidente. La gran mayoría de generación eléctrica hidráulica de Colombia se concentra en Antioquia y en el centro del país, el aporte de generación en el sur no es tan influyente como lo es en las regiones anteriormente mencionadas, por ende el precio de la energía está altamente influenciado por el comportamiento climático en las regiones de Antioquia y centro del país, así que el estudio se enfoca en la tendencia hidrológica de estas dos zonas. Como se ve claramente, la tendencia no es la misma durante la semana ni en las 48 horas siguientes a la semana de estudio. Se observa que en los ríos de las regiones de Antioquia y centro del país a partir del día 5 de la semana, hay un crecimiento en aportes hídricos bastante abrupto. Lo cual modifica el precio de la energía drásticamente, ya que al contar con gran capacidad de generación hidráulica, no es necesario encender todas las plantas y de prescindir del uso de la energía térmica lo que hace que la energía tenga un costo mucho más barato que en días de sequía y si se dirige todo ese análisis al error absoluto de la estimación, se puede deducir que por la tendencia que lleva la curva de precios de la energía los primeros 5 días, el modelo pronostica un precio de escases (alto), cuando en realidad por el cambio hidrológico el precio comienza a bajar y al comparar la estimación con el precio real hay una diferencia de 23.83 % reflejado en el inicio de las lluvias durante el transcurso de la semana en estudio.

Análogamente se analiza el comportamiento las ventanas de tiempo semanales 2, 6, y 10, comprendidas entre los días 19 y 25 de junio de 2014, 25 de noviembre y 1 de diciembre de 2012 y 26 de octubre y 1 de noviembre de 2013 respectivamente, sumando las 48 horas siguientes a la semana ensayada para analizar el comportamiento del pronóstico, las cuales reflejaron el siguiente comportamiento ilustrado en las siguientes gráficas

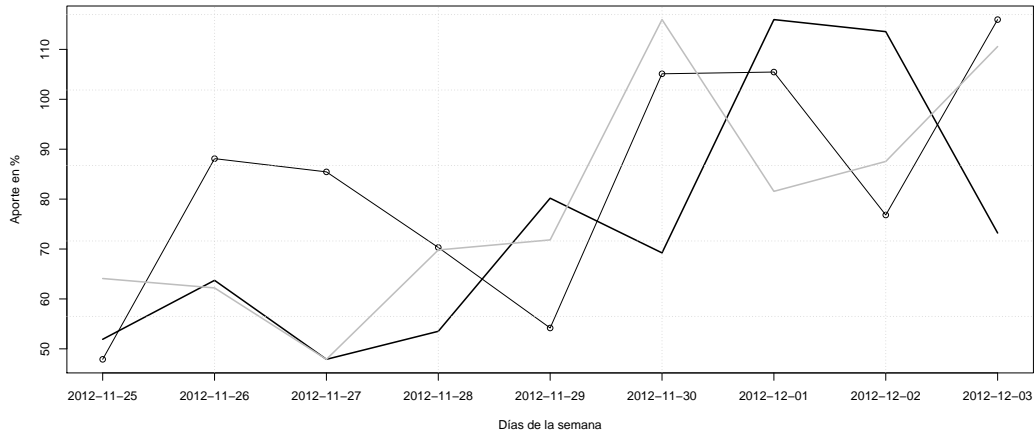
Figura 4.5.: Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 2



línea continua: Río Bogotá (centro)
 línea con puntos: Río Nare (Antioquia)
 línea gris: Río Altoanchicaya (Valle)

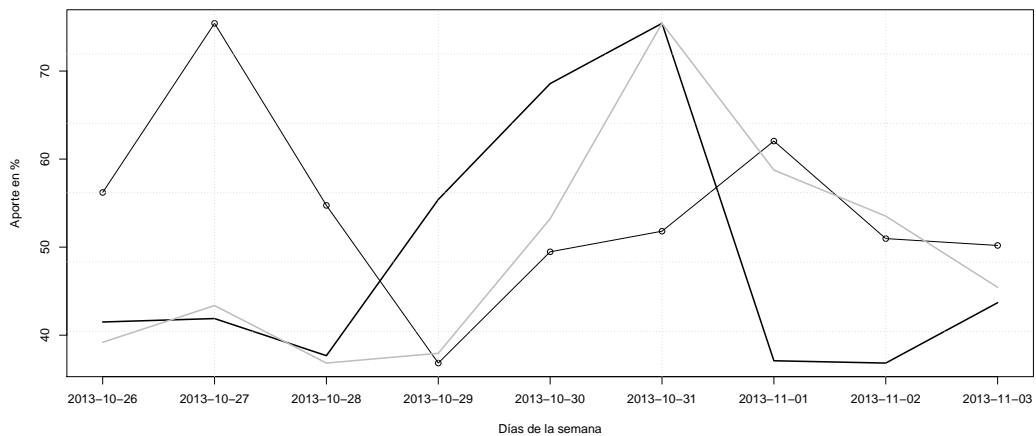
Fuente: (XM, 2014a)

Figura 4.6.: Comparación de aportes hidrológicos en 3 importantes zonas del país en las semana 6



línea continua: Río Bogotá (centro)
línea con puntos: Río Nare (Antioquia)
línea gris: Río Altoanchicaya (Valle)
Fuente: (XM, 2014a)

Figura 4.7.: Comparación de aportes hidrológicos en 3 importantes zonas del país en la semana 10



línea continua: Río Bogotá (centro)
línea con puntos: Río Nare (Antioquia)
línea gris: Río Altoanchicaya (Valle)
Fuente: (XM, 2014a)

Si se detalla el análisis de error absoluto para cada una de estas ventanas, se observa que el valor pronosticado en los tres casos es mayor que el valor real, lo cual permite concluir que el error en estas ventanas se origina por el caso estudiado en la ventana de semana 4, el cual por tendencia el programa pronostica un valor alto pero a causa repentina de lluvias y mantener tal pluviosidad, no es posible seguir una tendencia y hacer una buena estimación.

Vale aclarar que aunque en todos los casos estudiados el aumento de las lluvias desvió el pronóstico, no quiere decir que si llueve esa ventana de tiempo no es pronosticable, lo ideal es que en la ventana de tiempo a estimar, se mantenga una tendencia de comportamiento de precio de energía, ya sea subir y bajar diariamente, mantener un alza constante o una baja constante, sin importar cuál sea la dirección que tenga el precio pero que siga unos parámetros periódicos. Por tal razón no se analiza ninguna ventana de tiempo durante el año 2011, ya que si se observa el comportamiento general de las horas de este año, es un comportamiento muy plano que no fluctúa, por ende hacer un pronóstico en esta ventana de tiempo es innecesario ya que se indaga acerca del comportamiento climático para ese periodo y se encuentra que el país sufre para esa época el fenómeno del niño con más fuerza en la media histórica, donde se alteraron gran cantidad de factores meteorológicos que aumentan la temperatura del país, disminuyen los niveles de embalses, se hace necesario el uso de plantas de generación térmica por ende hay una tendencia en el precio de escasez. La fuerza del fenómeno fue tal, que se presentaron apagones por falta de capacidad, y aunque se resolvieron dentro de los estándares de calidad de generación de energía eléctrica, las autoridades gubernamentales encargadas del control, despacho y regulación toman acciones necesarias de inmediato y para el futuro, y así evitar de nuevo que en otra crisis exista otro apagón. El reflejo de estas decisiones se observan en el tercer trimestre del año 2015 donde la sequía obliga a los acueductos llevar un plan de racionamientos de suministro de agua potable pero la energía eléctrica sigue supliendo la demanda de manera constante y robusta.

CAPÍTULO 5

CONCLUSIONES

- Desde el punto de vista de la predicción de datos, se concluye que no es necesario una cantidad de datos cuantiosa para realizar modelos tipo ARIMA, ya que se comprobó que con una semana de datos, se puede realizar una predicción que se encuentre dentro de una franja de error menor al 5 %, lo cual logra una aproximación muy cercana a los valores de datos reales para dichas predicciones.
- En cuanto al uso de los modelos ARIMA y el resultado de las predicciones arrojadas por los mismos, se concluye, basándose en el resultado de los cálculos de errores en los costos por consumo de energía, que se aceptan los valores de predicción dados por éstos.
- Se pudo demostrar que los errores absolutos superiores al 10 %, los cuales no representan valores significativos, no se deben a fallas en el modelo matemático si no a los cambios abruptos que modifican de manera negativa, la condición de normalidad y la tendencia del comportamiento del precio de la energía por factores atmosféricos, que alteran la capacidad de producción de las hidroeléctricas y modifican la relación entre demanda y oferta lo cual obliga a realizar cambios en las condiciones de generación de energía, por ende varia la tendencia del precio. Estos cambios también provocan que cambie el comportamiento de las estrategias por parte de los agentes y por ende es de suma importancia la introducción de esta información para un pronóstico más preciso del precio de la energía eléctrica.
- Los modelos logran seguir el comportamiento de la variación típica de los costos de energía en un día normal, siguiendo la tendencia de picos y valles propios de la variación de los precios. Es esta cualidad la que permite que los errores en los costos sean tan

pequeños.

- Con este trabajo de grado se sienta un precedente para proyectos de grado futuros, en los cuales el análisis del precio de bolsa de energía debe analizarse a través de un modelo de recursos hídricos, es decir, realizar el proceso mostrado en este trabajo de grado para dichos recursos en los años con los que aquí se trabaja. Este comportamiento de la variable de recursos hídricos se debe tomar como una variable independiente al momento de realizar una correlación con el precio de bolsa, ya que, es el modelo de precio de bolsa una variable dependiente de la anterior. Se recomienda así mismo, analizar el comportamiento climático como una variable dicótoma cuando se realice el modelado del precio de bolsa de la energía.

BIBLIOGRAFÍA

- CASIMIRO, PILAR GONZÁLES. 2009. Análisis de series temporales: Modelos ARIMA. 77.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 1995a. Resolución CREG 024.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 1995b. Resolución CREG 025.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 1997. Resolución CREG 199.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 1998. Resolución CREG 131.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 2004. Sistema electrónico de contratos bilaterales -SEC-. 005, 2–8.
- COMISIÓN DE REGULACIÓN DE ENERGÍA Y GAS, CREG. 2007. Definición de la fórmula tarifaria de energía eléctrica, para el próximo periodo tarifario. 043, 4–11.
- CRYER, JONATHAN D., & CHAN, KUNG-SIK. 2008. *Time Series Analysis with applications in R*. Springer.
- FOUNDATION, THE R. 1993. *The R Project for Statistical Computing*. www.r-project.com.
- GUJARATI, DAMODAR., & PORTER, DAWN. 2010. *Econometría*. Mc Graw Hill.
- SERNA-SUAREZ, I.D., ORDONEZ-PLATA, G., PETIT-SUAREZ, J.F., & CAICEDO, G.CARRILLO. 2015 (Oct). Storage system scheduling effects on the life of lead-acid batteries. Pages 740–745 of: *Innovative Smart Grid Technologies Latin America (ISGT LATAM), 2015 IEEE PES*.

WALPOLE, RONALD E., MYERS, RAYMOND H., & MYERS, SHARON L. 1999. *Probabilidad y estadística para ingenieros*. Prentice Hall.

XM. 2010. *Precio de bolsa nacional*. <http://informacioninteligente10.xm.com.co/transacciones/>.

XM. 2012. *Precio de bolsa nacional*. <http://informacioninteligente10.xm.com.co/transacciones/>.

XM. 2013. *Precio de bolsa nacional*. <http://informacioninteligente10.xm.com.co/transacciones/>.

XM. 2014a (marzo). *Aportes hídricos diarios*. <http://informacioninteligente10.xm.com.co/transacciones/>.

XM. 2014b. *Precio de bolsa nacional*. <http://informacioninteligente10.xm.com.co/transacciones/>.

ANEXOS

ANEXO A

MERCADO DE ENERGÍA ELÉCTRICA

El marco teórico contiene toda la información necesaria para desarrollar un pronóstico a una serie de tiempo dada, teniendo en cuenta todas las características que involucran tal proceso. Por tal razón en los siguientes anexos se profundiza acerca del mercado de energía eléctrica y sus componentes dinámicas, como lo son el precio de la energía eléctrica en la bolsa, los contratos bilaterales, la regulación del mercado mayorista y la definición de la fórmula tarifaria. Todos los temas anteriormente mencionados tienen una estrecha relación con el arte de pronosticar, ya que un buen pronóstico significa un ahorro bastante significativo en las finanzas porque permite jugar con la variación del precio de la energía y hacer compras acertadas, o en su defecto, desconexiones de la red del sistema favorables para la economía del usuario. A continuación los anexos.

A.1 DESPACHOS ECONÓMICOS IDEALES Y PRECIO DE BOLSA

La Comisión de Regulación de Energía y Gas, CREG, adopta mediante las Resoluciones CREG-024 y CREG-025 de 1995 el esquema de subasta a aplicar diariamente para el despacho de generación del SIN, el cual tiene las siguientes características generales:

Los agentes generadores deben realizar diariamente para cada planta y/o unidad: i) oferta de precios¹, que corresponde a un precio para todas las horas del día, y ii) declaración de disponibilidad a nivel horario.

Con la información reportada por los agentes y la demanda esperada, se hace el programa de despacho mediante un proceso de optimización de 24 horas, donde se busca encontrar la solución más económica para atender la demanda del día, sujeto a las restricciones del sistema.

El precio de bolsa horario se determina como el mayor precio de oferta de las plantas y/o unidades. Este proceso se lleva a cabo mediante el despacho ideal, el cual se realiza considerando la oferta de precios por orden de méritos de menor a mayor, sin considerar las restricciones del sistema, hasta atender la demanda real.

Las diferencias entre el despacho programado y el ideal de las plantas y/o unidades generación se concilian mediante un proceso de reconciliaciones en el cual se evalúan las desviaciones para definir las compensaciones (positivas o negativas). (comisión de regulación de energía y gas, 1995a,b)

A.2 CONTRATOS BILATERALES DE ENERGÍA ELÉCTRICA

Los contratos bilaterales son compromisos adquiridos por generadores y comercializadores para vender y comprar energía a precios, cantidades y condiciones contractuales negociadas libremente entre las partes. La función de estos contratos es reducir la exposición a la volatilidad de precios en el mercado de corto plazo del generador y del usuario final. La entrega física de la energía contratada se efectúa a través de la bolsa por parte del generador que suscribió el contrato o por parte de otro generador según lo determine el despacho ideal.

No hay restricción alguna a la energía que un generador o un comercializador puede comprometer en contratos bilaterales ni al horizonte de tiempo que estos acuerdos deben cubrir. El único requisito es que el contrato especifique la cantidad que será utilizada en cada hora por el ASIC (Administrador del Sistema de Intercambios Comerciales) para la liquidación

Objetivos de los contratos de energía eléctrica Los objetivos buscados al presentar propuestas de mecanismos de transacción bilateral a través de un mercado de subastas, con estandarización de contratos y anonimato entre las partes, son brindar al mercado un mecanismo que genere la liquidez necesaria para el desarrollo futuro del mismo, dinamizar la comercialización, permitir avanzar al sector con la utilización de nuevas herramientas financieras.

Tipos de contratos de energía eléctrica Los contratos de energía eléctrica en Colombia son principalmente los siguientes:

Contrato pague lo contratado (cplc) y contrato pague lo demandado (cpld) Los CPLC y CPLD se pueden negociar incluyendo límites de precios, por encima o por debajo de un precio de referencia pactado. En los condicionamientos existentes pueden encontrarse entre

otros: precios atados a la bolsa, precios por submercados definidos por el destino de la energía final, y primas de ejecución de los mismos asociados con la ejecución de otros contratos.

Contrato pague lo generado De la misma forma, en los contratos Pague lo Generado - CPLG-, se da una transferencia de riesgo del vendedor hacia el comprador respecto a un CPLC, que en un ambiente de competencia se refleja en un menor precio. El valor de equilibrio de dicho riesgo está dado por la diferencia PPLC - PPLG. Por su parte, en los contratos condicionados, el movimiento del precio respecto al CPLC dependerá de la dirección de la transferencia de riesgo. (comisión de regulación de energía y gas, 2004)

A.3 REGULACIÓN DEL MERCADO ELÉCTRICO MAYORISTA

En la resolución CREG 131 de 1998, se modifica la resolución CREG 199 de 1997, expedida el 30 de septiembre de ese mismo año, donde se dictan ciertas disposiciones sobre el mercado competitivo de energía eléctrica en Colombia, resolviendo textualmente los siguientes 8 artículos.

Artículo 1º. Definiciones. Para efectos de la presente resolución se adoptan las siguientes definiciones:

Mercado competitivo. Es el conjunto de generadores y comercializadores en cuanto compran y venden energía eléctrica entre ellos. Forman parte de él, igualmente, los usuarios no regulados y quienes les proveen de energía eléctrica.

Usuario No Regulado. Para todos los efectos regulatorios, es una persona natural o jurídica con una demanda máxima superior a un valor en MW o a un consumo mensual mínimo de energía en MWh, definidos por la Comisión, por instalación legalizada, a partir de la cual no utiliza redes públicas de transporte de energía eléctrica y la utiliza en un mismo predio o en predios contiguos. Sus compras de electricidad se realizan a precios acordados libremente entre el comprador y el vendedor.

Artículo 2º. Límites para contratación en el mercado competitivo. A partir de la vigencia de la presente resolución, se establecen los siguientes límites de potencia o energía mensuales para que un usuario pueda contratar el suministro de energía en el mercado competitivo:

- Hasta el 31 de diciembre de 1999 0.5 MW o 270 MWh
- A partir del 1º de enero del 2000 0.1 MW o 55 MWh

Parágrafo. Para verificar las condiciones que deben cumplir los usuarios para comercializar en el mercado competitivo, se aplicará lo establecido en el Anexo No. 1 de la presente resolución.

Artículo 3º. Equipos de medición. Es requisito indispensable para acceder al mercado competitivo, que el usuario instale un equipo de medición con capacidad para efectuar telemetría, de modo que permita determinar la energía transada hora a hora, de acuerdo con los requisitos establecidos en el Código de Medida del Código de Redes y el Reglamento de Distribución.

Artículo 4º. Obligación de estar representado por un comercializador. Todo usuario no regulado debe estar representado por un comercializador ante el mercado mayorista.

Artículo 5º. Obligación de recaudar la contribución de solidaridad. En las facturas de los usuarios pertenecientes al mercado competitivo de energía que, de acuerdo con las normas legales que rigen la materia, están sujetos a la contribución de solidaridad de que trata la Ley 142 de 1994, los comercializadores deberán distinguir entre el valor que corresponde al servicio y dicha contribución, en las condiciones establecidas en la Resolución CREG-093 de 1994, o las normas que la modifiquen o sustituyan.

Artículo 6º. Neutralidad. Al vender electricidad en el mercado competitivo, los comercializadores no discriminarán entre personas o clases de personas, salvo que puedan demostrar que las diferencias en los precios reflejan diferencias en los costos por las circunstancias de dicha venta. A los comercializadores les está prohibido emplear prácticas para restringir, distorsionar o evitar la competencia en la generación o comercialización de la electricidad.

Artículo 7º. Reglas sobre comercializadores. Sin excepción, todos los comercializadores de energía en el Sistema Interconectado Nacional estarán sujetos a lo establecido en la Resolución CREG-054 de 1994, o las disposiciones que la modifiquen o sustituyan.

Artículo 8º. Vigencia. La presente resolución rige a partir de la fecha de su publicación en el Diario Oficial y deroga todas las disposiciones que le sean contrarias, en especial la Resolución CREG-199 de 1997, expedida por la comisión de regulación de energía y gas.(comisión de regulación de energía y gas, 1998, 1997)

A.4 DEFINICIÓN DE LA FÓRMULA TARIFARIA PARA USUARIOS REGULADOS

La definición de la fórmula tarifaria permitirá a los comercializadores establecer el Costo Unitario de Prestación del Servicio del servicio de energía eléctrica a usuarios regulados, en el marco del régimen de Libertad Regulada.

Es de indicar que esta fórmula cumple tanto con lo dispuesto en las Leyes 142 y 143 de 1994, como con los decretos del Ministerio de Minas y Energía 3734 de 2003, 3735 de 2003 y 387 de 2007, mediante los cuales se fijan las políticas y directrices que buscan fomentar la universalización del servicio.

A.4. Definición de la fórmula tarifaria para usuarios regulados

La fórmula tarifaria es la siguiente.

$$CUv_{n,m} = G_m + T_m + D_{n,m} + Cv_m + PR_{n,m} + R_m \quad (A.1)$$

$$CUf_m = Cf_m \quad (A.2)$$

Donde:

n = Nivel de tensión de conexión del usuario

m = El mes para el cual se calcula el costo unitario de la prestación de servicio

CUV_{n,m} = Componente variable del costo unitario de prestación del servicio (\$/KWh)

G_m = Costo de compra de energía (\$/KWh)

R_m = Costo por restricciones

T_m = Costo por uso del sistema nacional de transmisión (\$/KWh)

D_{n,m} = Costo por uso de sistemas de distribución (\$/KWh)

CV_m = Margen de comercialización que incluye los costos variables de la actividad de la comercialización de la energía (\$/KWh)

PR_{n,m} = Costo de compra, transporte y reducción de pérdidas de energía (\$/KWh)

CUF_m = Componente fija de costo unitario de prestación del servicio (\$/factura)

CF_m = Costo base de comercialización (\$/factura)

(comisión de regulación de energía y gas, 2007)

ANEXO B

RESULTADOS DE MODELADO Y PREDICCIÓN.

En este anexo del capítulo 2, predicción del precio de bolsa, se muestran los resultados de las restantes cinco ventanas de tiempo: cinco años, un año, seis meses, tres meses y una semana; al ser analizadas paso a paso como se hace con la ventana de tiempo de un mes de datos en el capítulo anteriormente mencionado. También, se muestra un breve introducción a R, las librerías que se utiliza y el script con el que se trabaja, de manera particular, con la ventana de tiempo de un mes.

Es cuanto al proceso de análisis de datos para cada ventana de tiempo, es necesario dejar claro que es una secuencia de pasos repetitiva, donde solo varía la estimación del modelo que se toma para cada ventana de tiempo y por ende, también varían los resultados de predicción.

B.1 INTRODUCCIÓN AL USO DE R

La herramienta de software libre R es un poderoso lenguaje y entorno de programación para el manejo estadístico de bases de datos (foundation, 1993), teniendo una gran variedad de librerías que, para ventaja de este trabajo, tienen como propósito el análisis de las series de tiempo y los modelados como AR, MA, ARMA y ARIMA. Existen varios motivos por los cuales se utiliza R para el análisis de los datos, teniendo en cuenta que hay también otras herramientas software para tal fin como lo son Eviews y Stata, siendo estos muy utilizados para el análisis estadístico. El primero de estos motivos es que; R es un software libre, es decir, esta herramienta se puede adquirir de forma gratuita a diferencia de Eviews y Stata. El segundo motivo es la capacidad que tiene R para el procesamiento de datos, por ejemplo, para la ventana de tiempo de 5 años, que se analiza en la siguiente sección, se tienen 43824 datos los

cuales se utilizan de base para la especificación del modelo de 5 años, el software R los puede manejar sin inconvenientes, pero, si se quiere utilizar la herramienta Eviews esta solo nos deja analizar los primeros 400 datos, limitando la base de datos de una manera muy significativa. El tercer y último motivo para el uso de R en el análisis de los datos es la versatilidad y facilidad de comandos con los que se cuenta y la gran cantidad de librerías que contienen a éstos, pudiendo realizar el proceso con tan solo unos pasos en un script del programa.

Se debe aclarar que, para que R pueda leer los datos, éstos se deben tener en un archivo separado por comas (.csv) y que todos los puntos decimales se encuentren asignados mediante punto y no con coma, ya que R trabaja la separación decimal mediante el uso del punto. También, es necesario aclarar que, el encabezado del archivo no puede llevar espacio alguno.

Para comenzar a trabajar los datos en R es necesario primero acondicionarlo para tal objetivo, lo primero que se debe hacer es tener los datos de trabajo en el mismo directorio de trabajo de R, para esto se tienen dos opciones:

- La primera es llevar el archivo .csv de la base de datos hacia donde está el directorio de R, para lograrlo, primero se debe verificar dónde se encuentra el directorio, para lo cual se usará el comando `getwd()`. Ya teniendo clara la dirección de la carpeta de trabajo de R, se moverá el archivo a ese lugar y ya se podrá trabajar en R tranquilamente.
- La segunda opción y tal vez la más cómoda, ya que se trabaja en una carpeta propia, es cambiar el directorio de R, para tal fin vamos en la consola a `achivo`, `cambiar dir...` y se escoge la carpeta con la dirección de los archivos para analizar, con lo cual ya se tendría como carpeta de trabajo a la carpeta donde se encuentran todos los archivos ; terminado este paso ya es posible trabajar en R.

B.1.1 Librerías y comandos a usar Las librerías que se usan para el tratamiento y análisis de los datos son las siguientes:

- TSA
- Tseries
- forecast
- akima

Todas indispensables para el manejo que se le dan a los datos ya que, los comandos que traen son las principales herramientas de este trabajo. Los comandos a utilizar tienen un orden de importancia en el proceso; el primero de ellos es todo lo relacionado al manejo de los datos,

comando que ayudan a leer los archivos que se contienen en una base de datos en Excel y pasarlos a R, como lo es `read.csv()`, ya estando en R, se tiene un comando para revisar que los datos pasados sean los correctos y que todo esté en el orden deseado, éste es `View()`. En segundo orden se tienen todos los comandos que llevan a la especificación del modelo, comandos que se encuentran principalmente en la librería TSA, como los comandos `acf()` y `pacf()` que muestran los correlogramas de los datos a tratar. Si en algún momento del proceso es requerida una simulación de un ARIMA para hacer algún tipo de comparaciones, entonces se utiliza `arima.sim()`. En tercer orden se presenta lo relacionado con la estimación de parámetros, lo cual se realiza con el comando `arima()`. En cuarto lugar se muestra el diagnóstico del modelo que se realiza por medio de varias pruebas, entre ellas la prueba de cuantil-cuantil que se hace con los comandos `qqnorm()` y `qqline()` y también el test de Ljung-box que se analiza con el comando `tsdiag()`. En el siguiente paso se halla el pronóstico como tal, que presenta el modelo obtenido con el comando `predict()`. Entre comandos extras que ayudan a extraer los datos del R a un archivo de texto plano se tiene el comando `write.table()`.

B.1.2 Script a utilizar Al aclarar lo referente a las librerías y los comandos a utilizar, es más sencillo que en la plataforma de R se describa un script con todos los comandos necesarios, de esta forma, se deja claro que todo el código que se escriba en el script no va a correr automáticamente sino que irá corriendo paso a paso conforme se avanza y, su vez se activa de manera manual seleccionando la parte deseada y presionando las teclas `ctrl + R`. A continuación se presenta, a modo de ejemplo, un script a utilizar en la ventana de tiempo de un mes, cuyo análisis de datos se muestra en la siguiente sección, es necesario aclarar que, lo que cambia entre las distintas ventanas de tiempo son el nombre del archivo de la base de datos y la cantidad de rezagos a analizar:

```
#-----ESPECIFICACIÓN DEL MODELO-----  
# Introducción de los datos a R  
  datos<- read.csv("Preciohoralmes.csv",header=T)  
# Confirmar los datos que se introdujeron en R son los esperados  
  View(datos)  
# Dibujar los datos  
  plot(datos$Preciohora, type='l')  
# Analizar correlograma de autocorrelación simple  
  acf(datos, 730)  
# Analizar correlograma de autocorrelación parcial
```

```

pacf(datos, 730)
# Analizar correlograma de autocorrelación simple de la primera
# diferencia de los datos
acf(diff(datos$Preciohora), 730)
# Analizar correlograma de autocorrelación parcial de la primera
# diferencia de los datos
pacf(diff(datos$Preciohora), 730)
# Analizar correlograma de autocorrelación simple de la segunda
# diferencia de los datos
acf(diff(datos$Preciohora,2), 730)
# Analizar correlograma de autocorrelación parcial de la segunda
# diferencia de lo datos
pacf(diff(datos$Preciohora,2),730)
#-----ESTIMACIÓN DE PARÁMETROS-----
estimacion<-arima(log(datos), order=c(20,1,1))
#-----DIAGNÓSTICO DEL MODELO-----
# Gracias a la función arima() tenemos los valores de sigma^2 que
# debe ser lo más cercana a 0, el valor de los likelihoods que debe
# ser un valor positivo y el valor aic que debe darnos un valor
# menos a 2 y si es posible negativo para validar la eficacia del
# modelo. Aún así determinamos los siguientes análisis:
# Prueba quantil-quantil
qqnorm(residuals(estimacion)); qqline(residuals(estimacion))
# Test de Ljung-box
signif(acf(residuals(estimacion),plot=F)$acf[1:50],2)
signif(pacf(residuals(estimacion),plot=F)$acf[1:50],2)
tsdiag(estimacion ,gof=50,omit.initial=F)
#-----PRONÓSTICO DEL MODELO-----
# Pronosticar 48 datos futuros predict(estimacion , n.ahead=48)
#-----PASAR DATOS DE R A ARCHIVO DE TEXTO PLANO-----
write.table(exp(x),file="Precio hora futuro 1 mes.txt")

```

Con este script se realiza el análisis de los datos de las 6 ventanas de tiempo, cambiando, claro está, el nombre del archivo donde se encuentra la base de datos, la cantidad de rezagos

a analizar en los correlogramas y al final el nombre del archivo de texto plano donde se desea guardar la información de las predicciones halladas.

B.2 VENTANA DE TIEMPO DE CINCO AÑOS.

Es la ventana que comprende desde el inicio del 2010 hasta el final del 2014, es decir, la totalidad de la base de datos con la que se trabaja en este proyecto.

Especificación del modelo A continuación se muestran las gráficas de datos de la ventana de tiempo, las autocorrelaciones simple y parcial, en su estado normal, en primera y segunda diferencia

Figura B.1.: Análisis del comportamiento de cinco años de datos mediante el uso de R

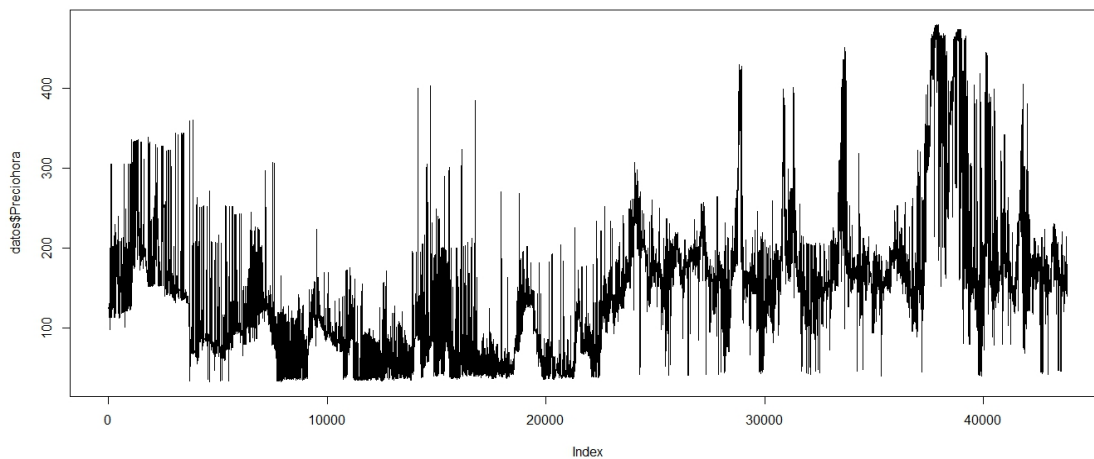


Figura B.2.: Gráfico de correlograma simple para cinco años de datos

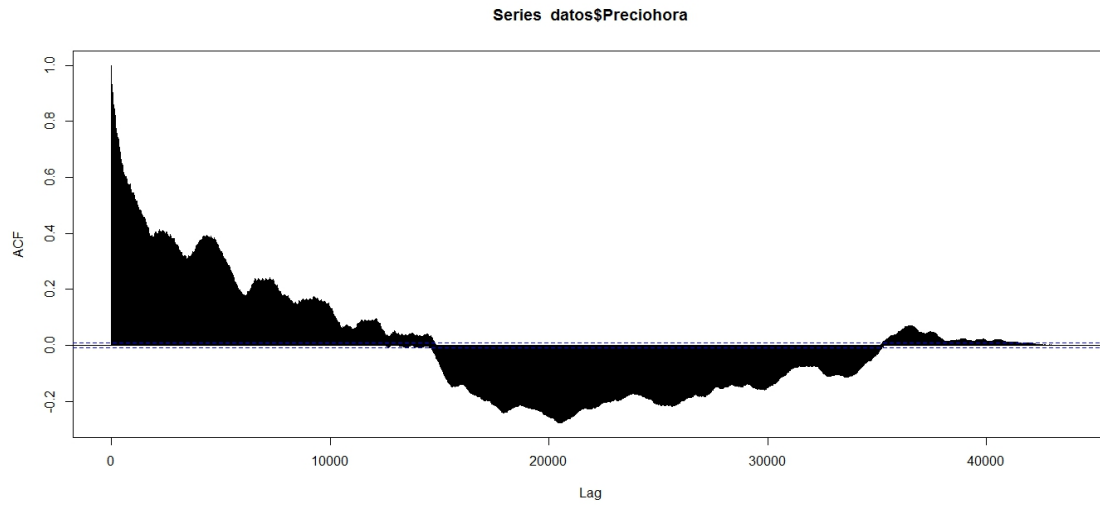


Figura B.3.: Gráfico de correlograma parcial para cinco años de datos

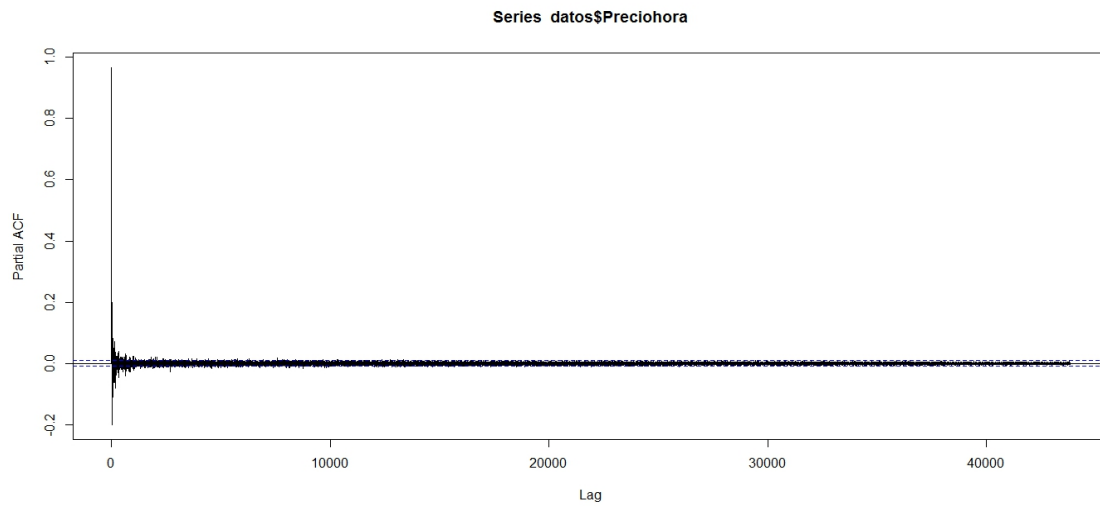


Figura B.4.: Gráfico de correlograma simple en primera diferencia para cinco años de datos.

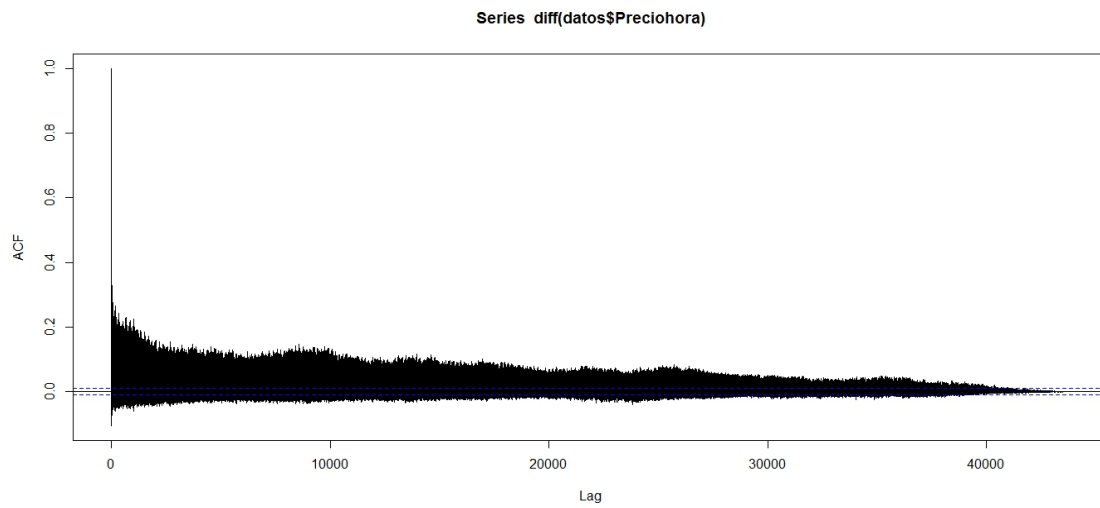


Figura B.5.: Gráfico de correlograma parcial en primera diferencia para cinco años de datos.

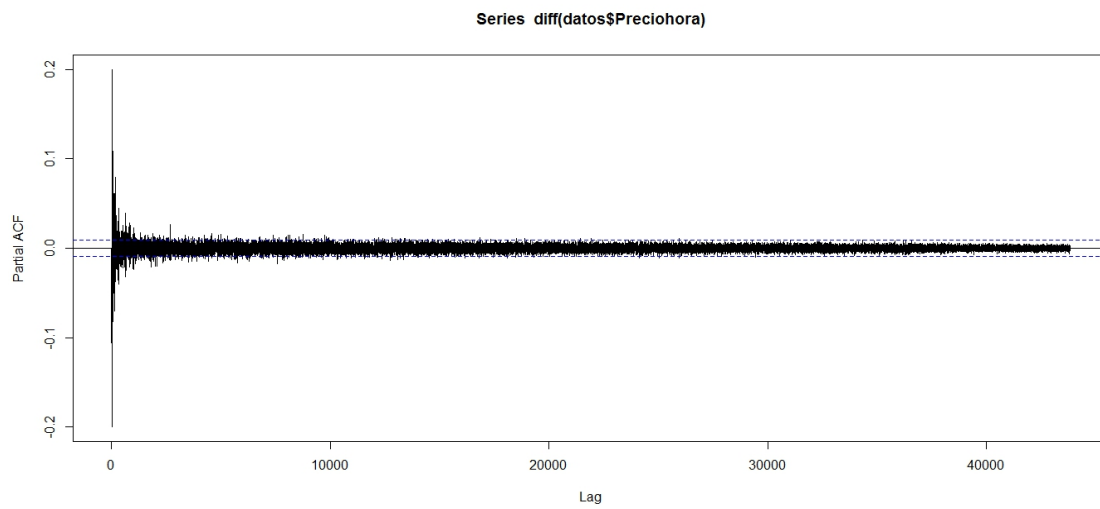


Figura B.6.: Gráfico de correlograma simple en segunda diferencia para cinco años de datos..

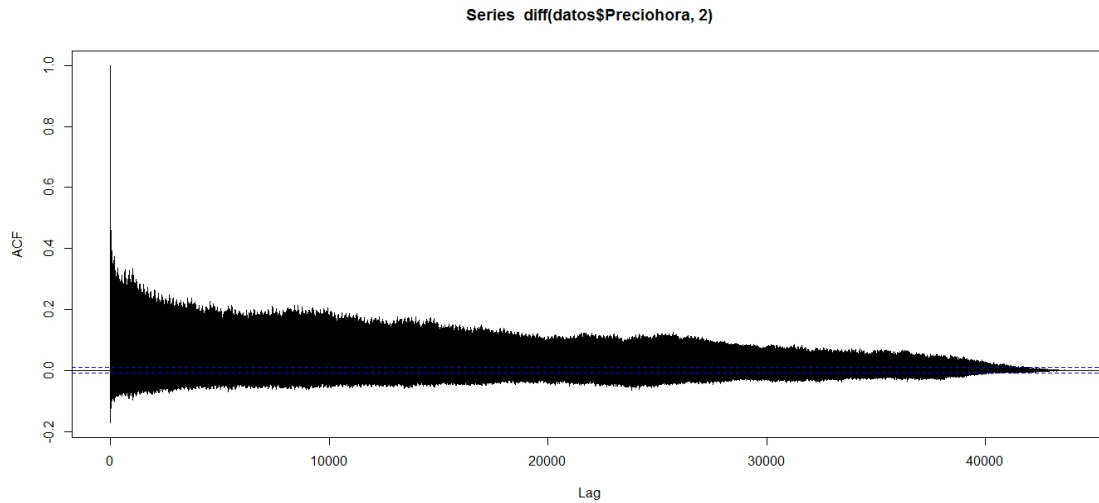
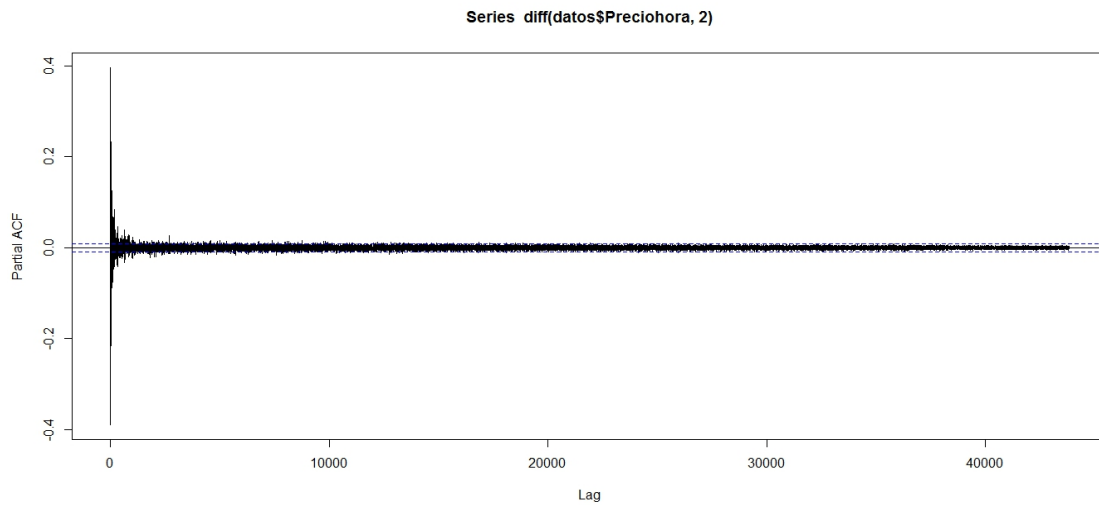


Figura B.7.: Gráfico de correlograma parcial en segunda diferencia para cinco años de datos.



Como resultado de la observación hecha a las gráficas se tiene como modelo un ARIMA (15, 1, 7)

Estimación de parámetros Se presenta a continuación una tabla con los datos arrojados por R y los resultados del test de significancia de los parámetros hallados, así también, la

ANEXO B. RESULTADOS DE MODELADO Y PREDICCIÓN.

fórmula final para el modelo.

TABLA B.1.: Parámetros arrojados por R para cinco años de datos

AR	1	2	3	4	5	6	7	8	9
ϕ	-0.6867	-0.0982	0.4471	0.6489	0.3107	-0.3165	-0.8377	-0.1014	-0.0122
s.e.	0.0091	0.0071	0.0072	0.0065	0.0079	0.0080	0.0111	0.0081	0.0072
AR	10	11	12	13	14	15			
ϕ	0.0081	-0.0557	-0.1753	-0.2259	-0.2213	-0.1655			
s.e.	0.0069	0.0069	0.0063	0.0063	0.0061	0.0052			

MA	1	2	3	4	5	6	7
θ	0.4869	-0.2075	-0.7153	-0.7633	-0.2764	0.4272	0.9073
s.e.	0.0084	0.0038	0.0060	0.0035	0.0053	0.0052	0.0105

TABLA B.2.: Resultado de prueba de significancia para parámetros de cinco años de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	-0.6867	0.0091	75.4615	0.4869	0.0084	57.9643
2	-0.0982	0.0071	13.8310	-0.2075	0.0038	54.6053
3	0.4471	0.0072	62.0972	-0.7153	0.006	119.2167
4	0.6489	0.0065	99.8308	-0.7633	0.0035	218.0857
5	0.3107	0.0079	39.3291	-0.2764	0.0053	52.1509
6	-0.3165	0.008	39.5625	0.4272	0.0052	82.1538
7	-0.8377	0.0111	75.4685	0.9073	0.0105	86.4095
8	-0.1014	0.0081	12.5185			
9	-0.0122	0.0072	1.6944			
10	0.0081	0.0069	1.1739			
11	-0.0557	0.0069	8.0725			
12	-0.1753	0.0063	27.8254			
13	-0.2259	0.0063	35.8571			
14	-0.2213	0.0061	36.2787			
15	-0.1655	0.0052	31.8269			

Con los parámetros significativos se tiene la fórmula final que caracteriza a la ventana de 5 años:

$$y(t) = -0,6867y(t-1) - 0,0982y(t-2) + 0,4471y(t-3) + 0,6489y(t-4) + 0,3107y(t-5) - 0,3165y(t-6) - 0,8377y(t-7) - 0,1014y(t-8) - 0,0557y(t-11) - 0,1753y(t-12) - 0,2259y(t-13)$$

$$-0,2213y(t-14)-0,1655y(t-15)+0,4869e(t-1)-0,2075e(t-2)-0,7153e(t-3)-0,7633e(t-4) \\ -0,2764e(t-5)+0,4272e(t-6)+0,9073e(t-7)$$

Diagnóstico del modelo A continuación se muestran los resultados de σ^2 , log likelihood y aic que dio como resultado R para esta ventana de tiempo. De igual manera se presentan las gráficas de las pruebas de q-q y Ljung-box.

TABLA B.3.: Valores de σ^2 , log likelihood y aic arrojados por R para cinco años de datos

σ^2	Log likelihood	AIC
0.02018	23333.97	-46621.93

Figura B.8.: Resultado de prueba de q.q para cinco años de datos

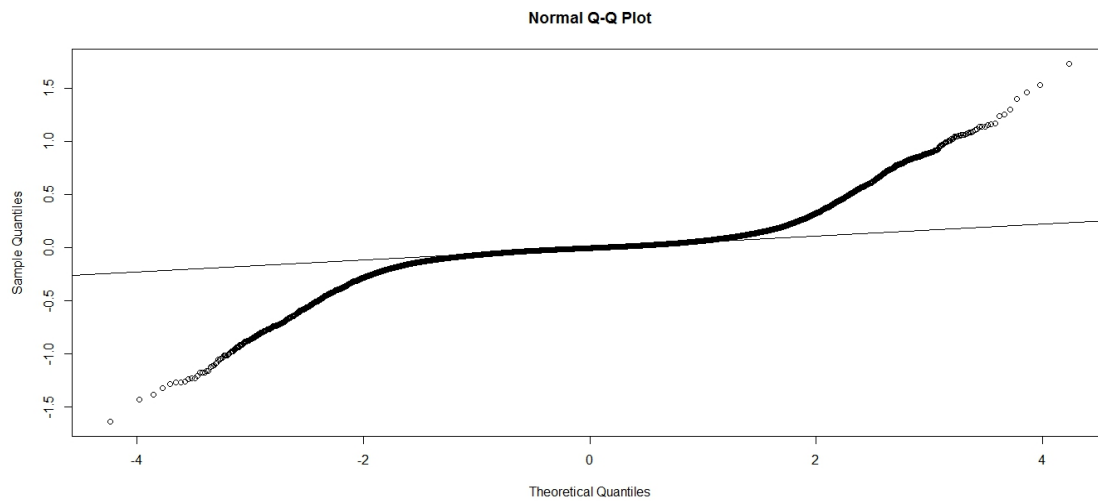
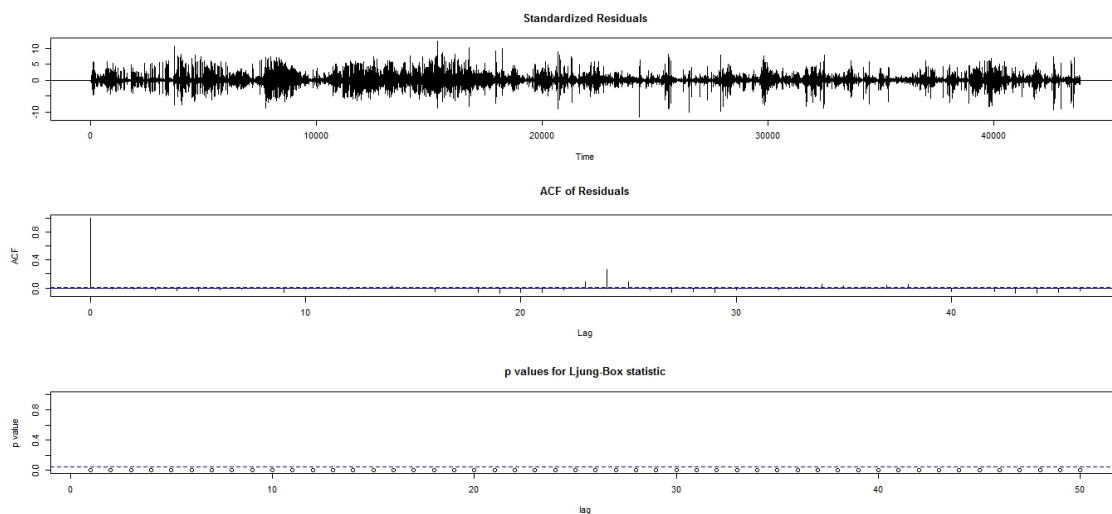


Figura B.9.: Resultado de prueba de Ljung-box para cinco años de datos



Predicción En esta parte se presentan los resultados de predicción de las 48 horas siguientes a la semana de tiempo con la que se trabaja, así mismo, los gráficos de franja de error de predicción y del comportamiento del precio de la predicción, como también el resultado arrojado como archivo de texto plano por R.

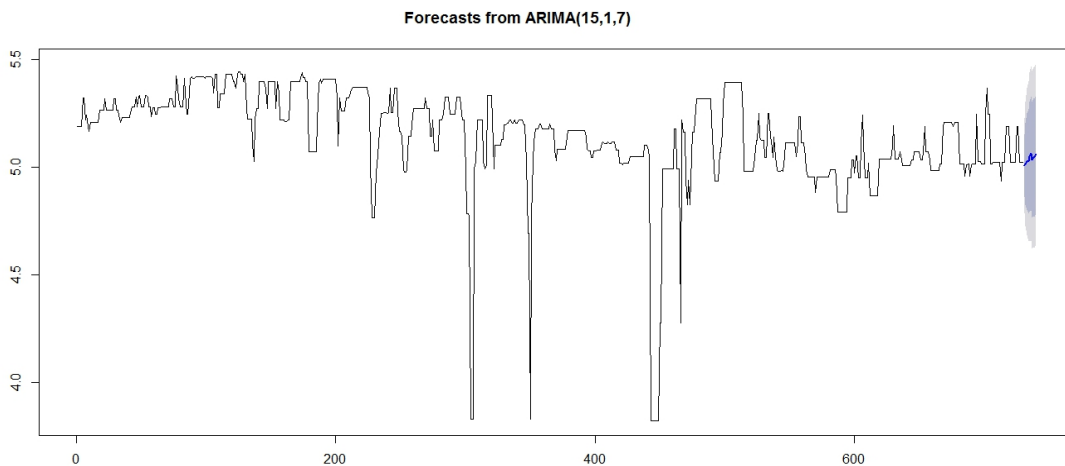
TABLA B.4.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para cinco años de datos.

[1 - 8]	4.9877	4.8612	4.9548	4.9678	4.9612	4.9637	4.9557	4.9718
[9 - 16]	4.9823	4.9840	5.0137	5.0387	5.0394	5.0682	5.0769	5.0939
[17 - 24]	5.0941	5.0818	5.0876	5.0851	5.0549	5.0485	5.0272	5.0180
[25 - 32]	4.9973	4.9705	4.9690	4.9668	4.9455	4.9542	4.9553	4.9727
[33 - 40]	4.9820	4.9870	5.0155	5.0403	5.0417	5.0652	5.0747	5.0916
[41 - 48]	5.0923	5.0818	5.0878	5.0853	5.0571	5.0487	5.0287	5.0185

[1 - 8]	0.1420	0.1819	0.2028	0.2148	0.2223	0.2267	0.2297	0.2326
[9 - 16]	0.2351	0.2398	0.2457	0.2516	0.2562	0.2600	0.2628	0.2648
[17 - 24]	0.2673	0.2695	0.2723	0.2760	0.2847	0.2856	0.2911	0.2965
[25 - 32]	0.3026	0.3082	0.3138	0.3193	0.3246	0.3298	0.3346	0.3386
[33 - 40]	0.3427	0.3460	0.3488	0.3513	0.3535	0.3557	0.3577	0.3595
[41 - 48]	0.3618	0.3640	0.3665	0.3693	0.3723	0.3759	0.3797	0.3836

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho

Figura B.10.: Gráfico de franja de error de predicción para cinco años de datos con acercamiento al último mes.



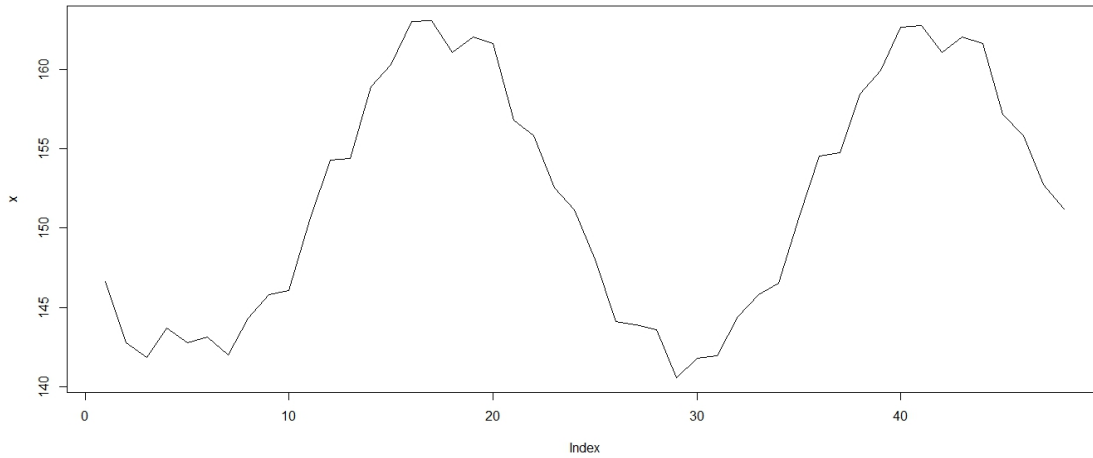
Los resultados dados en texto plano son:

ANEXO B. RESULTADOS DE MODELADO Y PREDICCIÓN.

TABLA B.5.: Resultados finales de datos de predicción.

1	146.6067746	17	163.0723553	33	145.7726185
2	142.7660105	18	161.0669311	34	146.5107786
3	141.8607581	19	162.0088559	35	150.7455023
4	143.7118136	20	161.5962596	36	154.5256342
5	142.7742912	21	156.7973175	37	154.7444428
6	143.1335342	22	155.8026278	38	158.4171906
7	141.9830948	23	152.5090594	39	159.9306046
8	144.293728	24	151.115735	40	162.6586771
9	145.8111075	25	148.0237894	41	162.7686716
10	146.0615342	26	144.1124648	42	161.0782062
11	150.471395	27	143.8888318	43	162.0474186
12	154.2807508	28	143.5681736	44	161.6408664
13	154.3796765	29	140.5540342	45	157.1415513
14	158.8912491	30	141.7785028	46	155.8327006
15	160.2886095	31	141.9319901	47	152.7421186
16	163.0278426	32	144.4262505	48	151.1884391

Figura B.11.: Gráfico de valores de predicción de precio de 48 horas futuras para cinco años de datos.



B.3 VENTANA DE TIEMPO DE UN AÑO

Es la ventana que comprende desde el inicio del 2014 hasta finales del mismo, es decir, el último año presente en la base de datos de cinco años.

Especificación del modelo A continuación se muestran las gráficas de datos de la ventana de tiempo, las autocorrelaciones simple y parcial, en su estado normal, en primera y segunda diferencia

Figura B.12.: Análisis del comportamiento de un año de datos mediante el uso de R

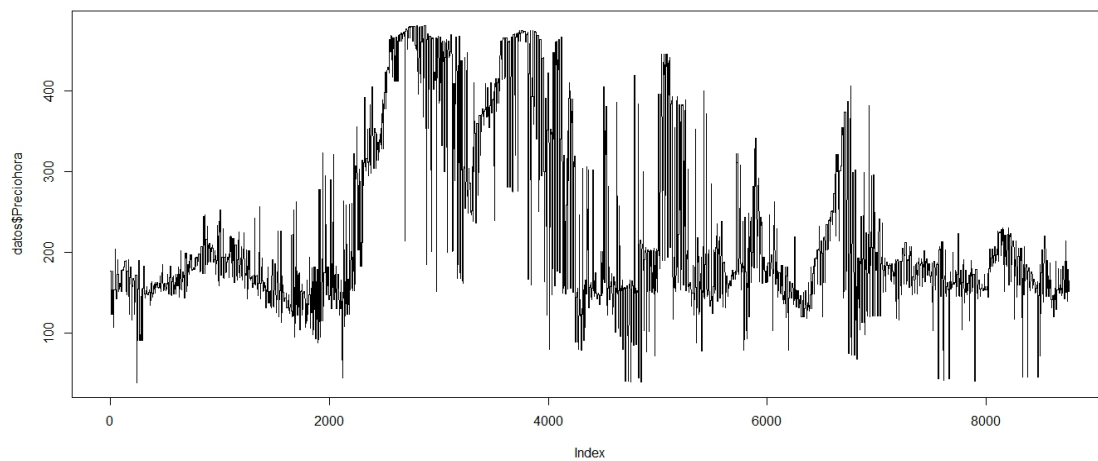


Figura B.13.: Gráfico de correlograma simple para un año de datos

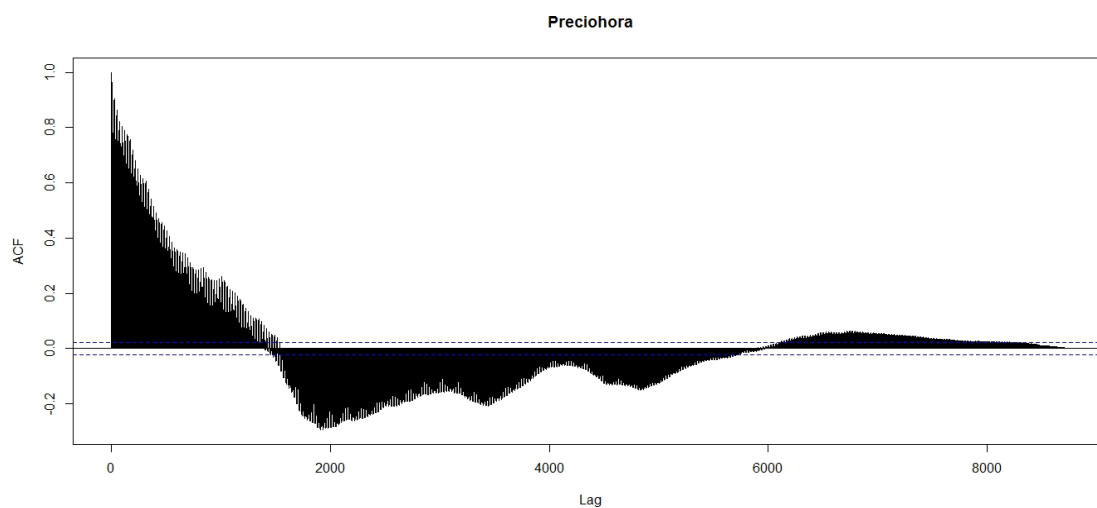


Figura B.14.: Gráfico de correlograma parcial para un año de datos

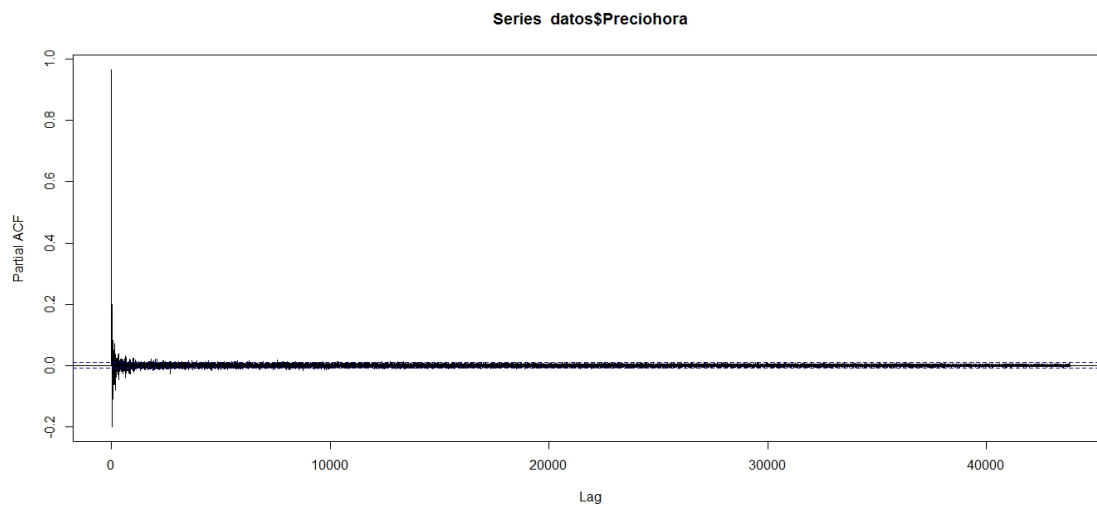


Figura B.15.: Gráfico de correlograma simple en primera diferencia para un año de datos.

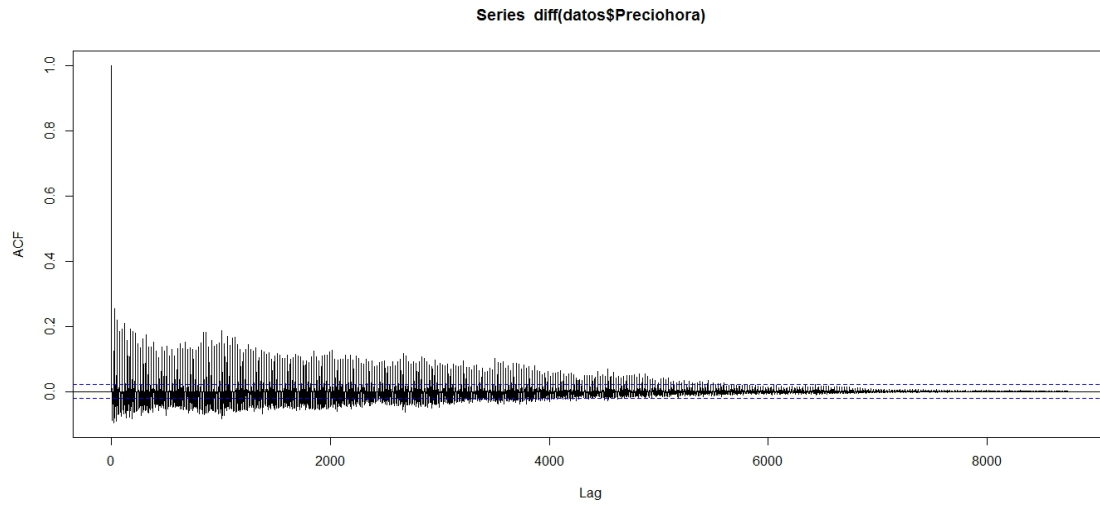


Figura B.16.: Gráfico de correlograma parcial en primera diferencia para un año de datos.

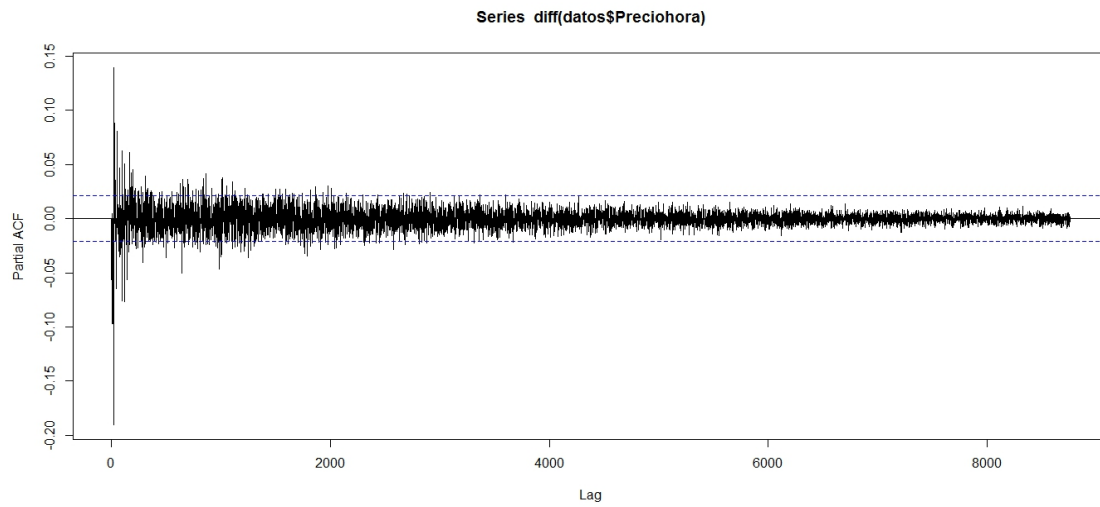


Figura B.17.: Gráfico de correlograma simple en segunda diferencia para un año de datos.

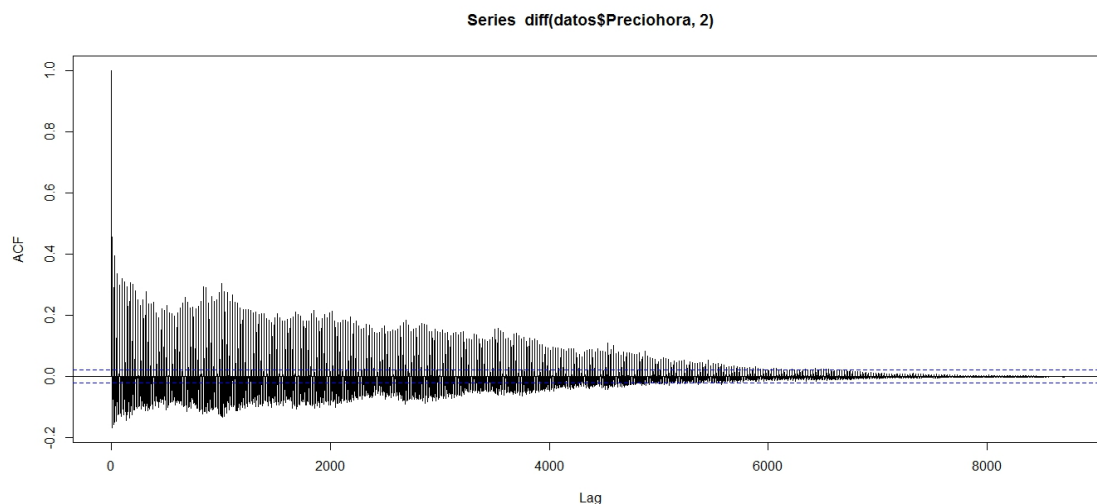
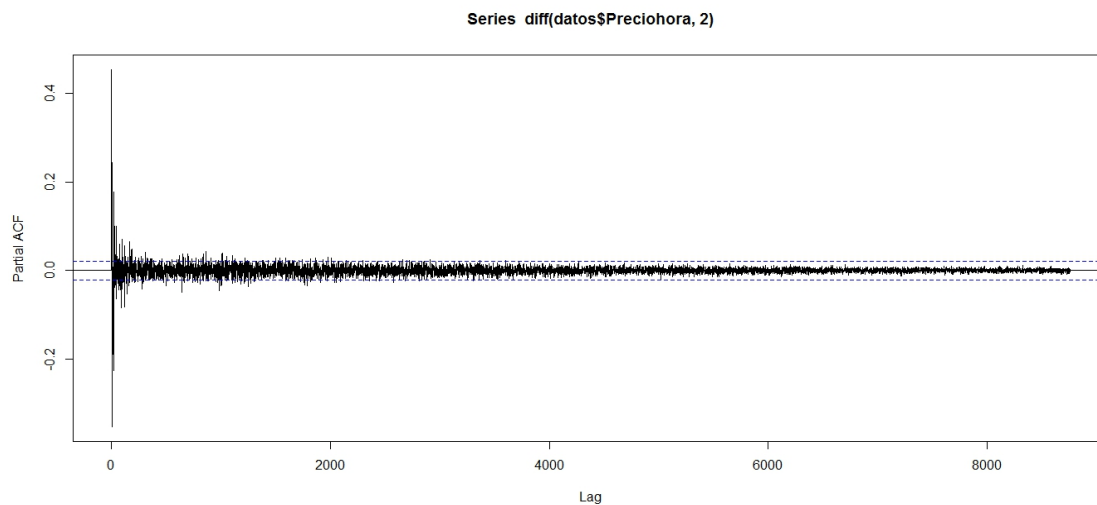


Figura B.18.: Gráfico de correlograma parcial en segunda diferencia para un año de datos.



Como resultado de la observación hecha a las gráficas se tiene como modelo un ARIMA (23, 1, 13)

Estimación de parámetros Se presenta a continuación una tabla con los datos arrojados por R y los resultados del test de significancia de los parámetros hallados, así también, la

fórmula final para el modelo.

TABLA B.6.: Parámetros arrojados por R para un año de datos

AR	1	2	3	4	5	6	7	8	9
ϕ	0.2155	-1.2586	0.2687	0.1413	-0.0356	-0.5875	-0.295	-0.4266	0.1065
s.e.	0.1816	0.1532	NAN	NAN	NAN	NAN	NAN	NAN	NAN
AR	10	11	12	13	14	15	16	17	18
ϕ	0.2799	-0.0756	1.00E-04	-0.5109	-0.1529	-0.117	-0.1079	-0.0817	-0.1409
s.e.	NAN	NAN	NAN	nan	NAN	NAN	0.0169	NAN	NAN
AR	19	20	21	22	23				
ϕ	-0.0913	-0.0644	-0.0364	-0.0575	0.0074				
s.e.	NAN	0.0097	NAN	NAN	0.0281				

MA	1	2	3	4	5	6	7	8	9
θ	-0.4585	0.1566	-0.4235	-0.144	0.0169	0.5722	0.1454	0.273	-0.3165
s.e.	0.1818	0.2014	NAN	NAN	NAN	NAN	NAN	NAN	NAN
MA	10	11	12	13					
θ	-0.3873	0.0412	-0.0624	0.4522					
s.e.	NAN	NAN	NAN	NAN					

TABLA B.7.: Resultado de prueba de significancia para parámetros de un año de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	0.2155	0.1816	1.1866	-0.4585	0.1818	2.5220022
2	-1.2586	0.1532	8.2154	0.1566	0.2014	0.7775571
3	0.2687	NAN	0.2687	-0.4235	NAN	-0.4235
4	0.1413	NAN	0.1413	-0.144	NAN	-0.144
5	-0.0356	NAN	-0.0356	0.0169	NAN	0.0169
6	-0.5875	NAN	-0.5875	0.5722	NAN	0.5722
7	-0.295	NAN	-0.295	0.1454	NAN	0.1454
8	-0.4266	NAN	-0.4266	0.273	NAN	0.273
9	0.1065	NAN	0.1065	-0.3165	NAN	-0.3165
10	0.2799	NAN	0.2799	-0.3873	NAN	-0.3873
11	-0.0756	NAN	-0.0756	0.0412	NAN	0.0412
12	1.00E-04	NAN	1.00E-04	-0.0624	NAN	-0.0624
13	-0.5109	NAN	-0.5109	0.4522	NAN	0.4522
14	-0.1529	NAN	-0.1529			
15	-0.117	NAN	-0.117			
16	-0.1079	0.0169	6.3846			
17	-0.0817	NAN	-0.0817			
18	-0.1409	NAN	-0.1409			
19	-0.0913	NAN	-0.0913			
20	-0.0644	0.0097	6.6391			
21	-0.0364	NAN	-0.0364			
22	-0.0575	NAN	-0.0575			
23	0.0074	0.0281	0.2633			

Con los parámetros significativos se tiene la fórmula final que caracteriza a la ventana de 1 año:

$$\begin{aligned}
 y(t) = & -1,2586y(t-2) + 0,2687y(t-3) + 0,1413y(t-4) - 0,5875y(t-6) - 0,295y(t-7) \\
 & -0,4266(t-8) + 0,1065y(t-9) - 0,2799y(t-10) - 0,5109y(t-13) - 0,1529y(t-14) - 0,117y(t-15) \\
 & -0,1079y(t-16) - 0,0644y(t-20) - 0,4585e(t-1) - 0,4235e(t-3) - 0,144e(t-4) + 0,5722e(t-6) \\
 & + 0,5722e(t-6) + 0,1454e(t-7) + 0,273e(t-8) - 0,3165e(t-9) - 0,3873e(t-10)
 \end{aligned}$$

Diagnóstico del modelo A continuación se muestran los resultados de σ^2 , log likelihood y aic que dio como resultado R para esta ventana de tiempo. De igual manera se presentan las gráficas de las pruebas de q-q y Ljung-box.

TABLA B.8.: Valores de σ^2 , log likelihood y aic arrojados por R para un año de datos.

σ^2	Log likelihood
0.01475	6038.45

Figura B.19.: Resultado de prueba de q,q para un año de datos.

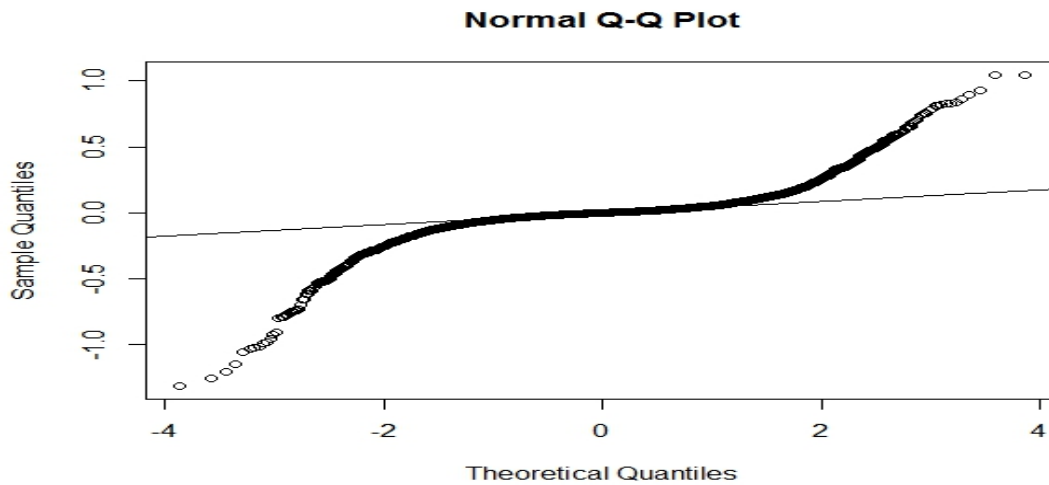
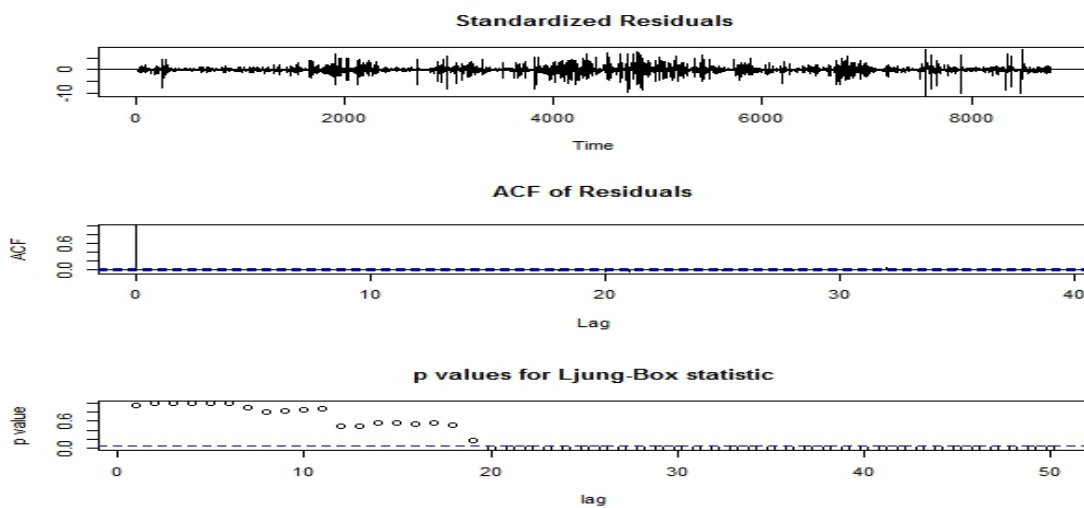


Figura B.20.: Resultado de prueba de Ljung-box para un año de datos



Predicción En esta parte se presentan los resultados de predicción de las 48 horas siguientes a la semana de tiempo con la que se trabaja, así mismo, los gráficos de franja de error de predicción y del comportamiento del precio de la predicción, como también el resultado arrojado como archivo de texto plano por R.

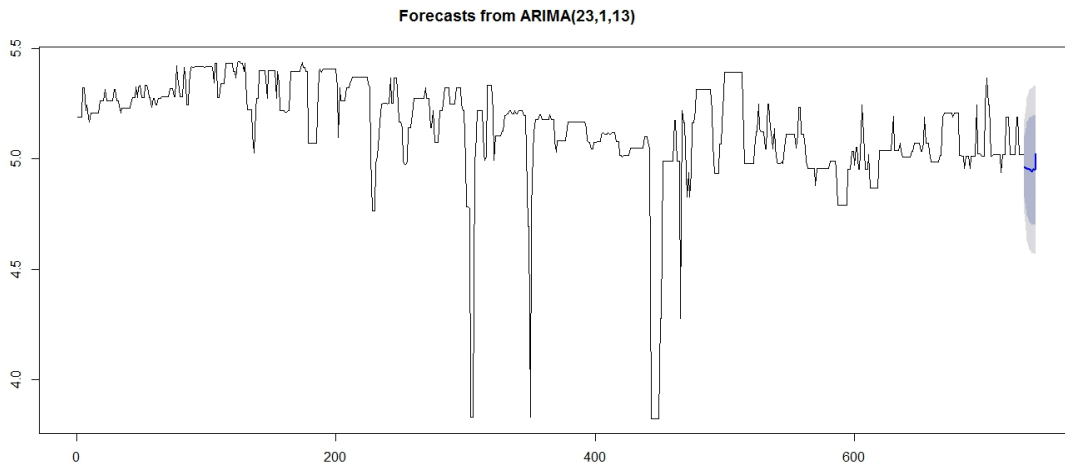
TABLA B.9.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para un año de datos.

[1 - 8]	4.9925	4.9654	4.9456	4.9399	4.9514	4.9649	4.9849	5.006
[9 - 16]	5.0252	5.0507	5.0603	5.0623	5.0409	5.0311	5.0455	5.0475
[17 - 24]	5.0373	5.0393	5.0698	5.0845	5.0743	5.061	5.0421	5.0256
[25 - 32]	4.9994	4.9705	4.9482	4.9388	4.9543	4.9741	4.9924	5.0071
[33 - 40]	5.0285	5.0548	5.0599	5.0499	5.0356	5.0334	5.0386	5.036
[41 - 48]	5.0352	5.0415	5.061	5.0764	5.0739	5.0601	5.0405	5.0258

[1 - 8]	0.12147	0.1523	0.1689	0.1786	0.1858	0.1906	0.1937	0.1962
[9 - 16]	0.1985	0.2008	0.2027	0.2046	0.2061	0.2072	0.2082	0.2092
[17 - 24]	0.21	0.211	0.2118	0.2129	0.2949	0.2178	0.2208	0.2251
[25 - 32]	0.2316	0.2386	0.2453	0.2514	0.257	0.2617	0.2651	0.2677
[33 - 40]	0.2694	0.2709	0.2722	0.2735	0.2746	0.2756	0.2766	0.2777
[41 - 48]	0.2786	0.2794	0.2801	0.2811	0.2825	0.2844	0.2868	0.29

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho

Figura B.21.: Gráfico de franja de error de predicción para un año de datos con acercamiento al último mes.

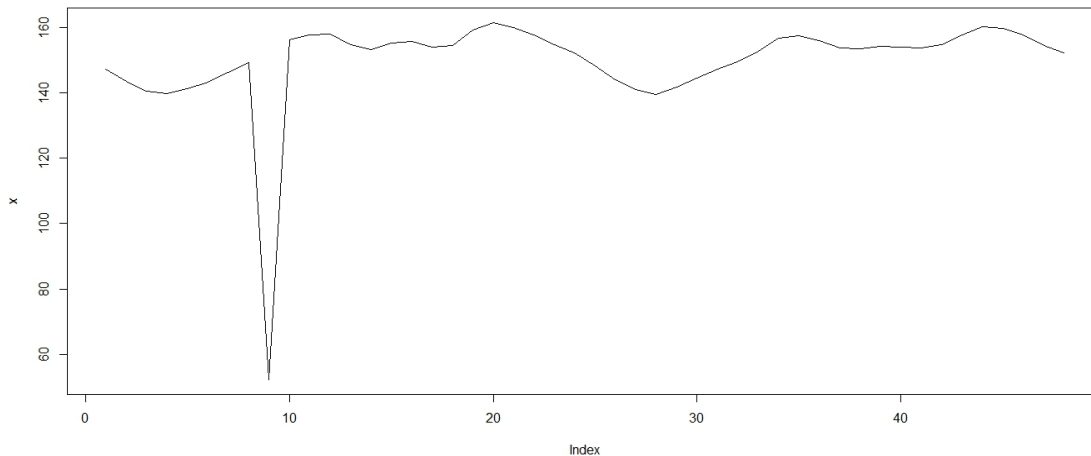


Los resultados dados en texto plano son:

TABLA B.10.: Resultados finales de datos de predicción.

1	147.308496	17	154.0590545	33	152.7141694
2	143.3753476	18	154.3722664	34	156.7758377
3	140.5661224	19	159.1507713	35	157.5753884
4	139.7645191	20	161.5148356	36	156.0124794
5	141.3776954	21	159.8610499	37	153.7979918
6	143.2998087	22	157.7481857	38	153.4632309
7	146.1901249	23	154.8003156	39	154.2681003
8	149.3081065	24	152.2749853	40	153.8592156
9	152.2136308	25	148.3250279	41	153.7427883
10	156.1445218	26	144.1084297	42	154.716746
11	157.6401652	27	140.9267145	43	157.7636457
12	157.9608156	28	139.6041615	44	160.2019169
13	154.6109559	29	141.7902709	45	159.8041495
14	153.1030157	30	144.6197668	46	157.6102164
15	155.3337457	31	147.3030457	47	154.5485057
16	155.6456579	32	149.4845436	48	152.2937162

Figura B.22.: Gráfico de valores de predicción de precio de 48 horas futuras para un año de datos.



B.4 VENTANA DE TIEMPO DE 6 MESES

Es la ventana que comprende desde mitad del 2014 hasta finales del mismo, es decir, el último año presente en la base de datos de cinco años.

Especificación del modelo A continuación se muestran las gráficas de datos de la ventana de tiempo, las autocorrelaciones simple y parcial, en su estado normal, en primera y segunda diferencia

Figura B.23.: Análisis del comportamiento de seis meses de datos mediante el uso de R

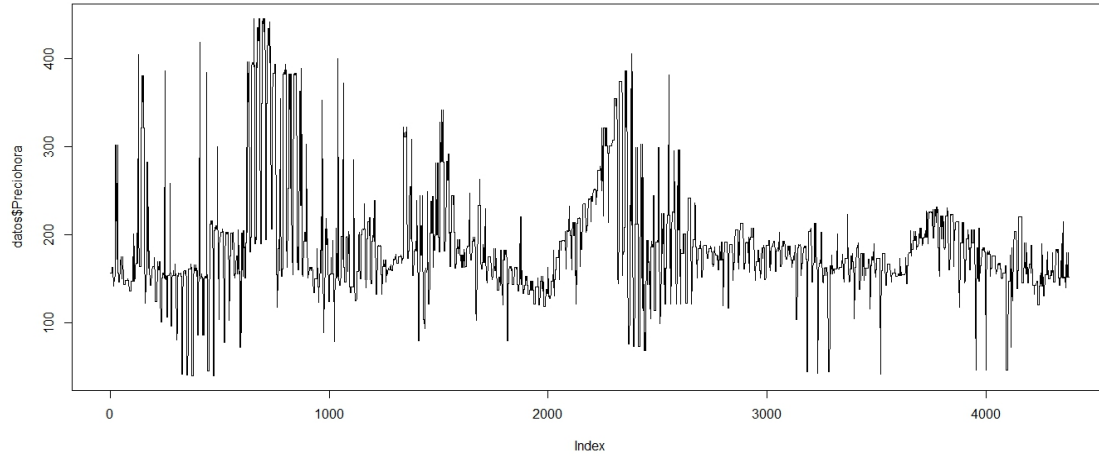


Figura B.24.: Gráfico de correlograma simple para seis meses de datos

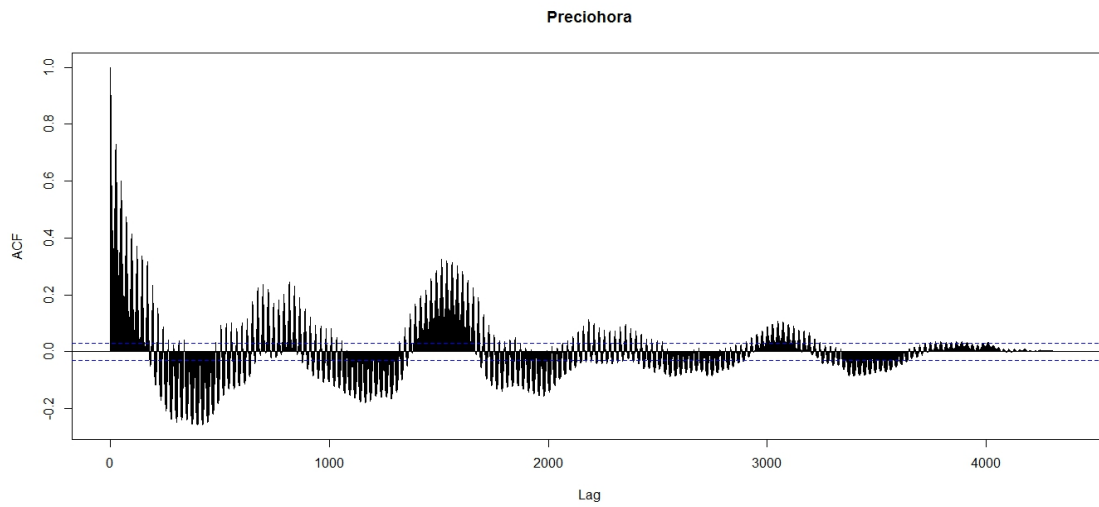


Figura B.25.: Gráfico de correlograma parcial para seis meses de datos.

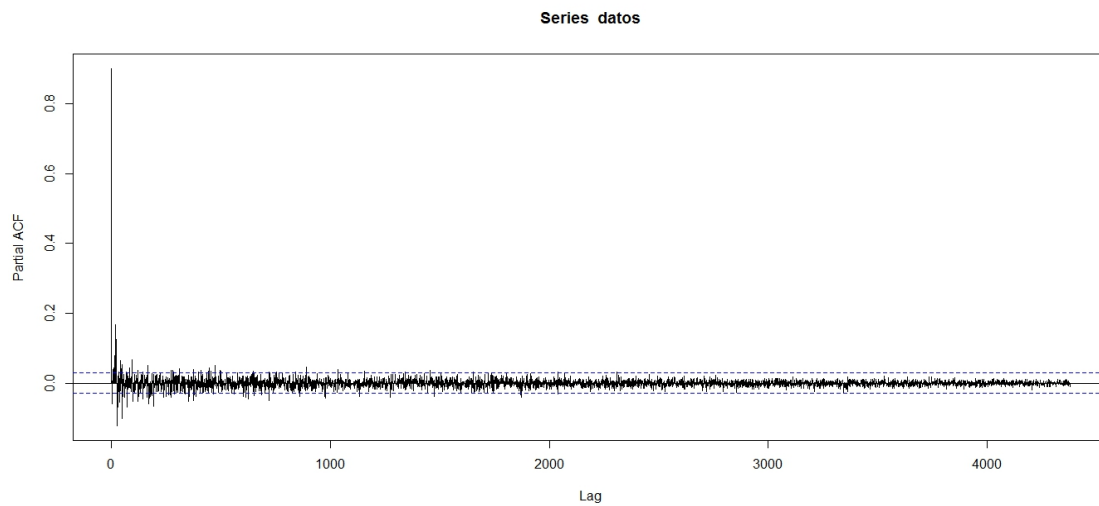


Figura B.26.: Gráfico de correlograma simple en primera diferencia para seis meses de datos.

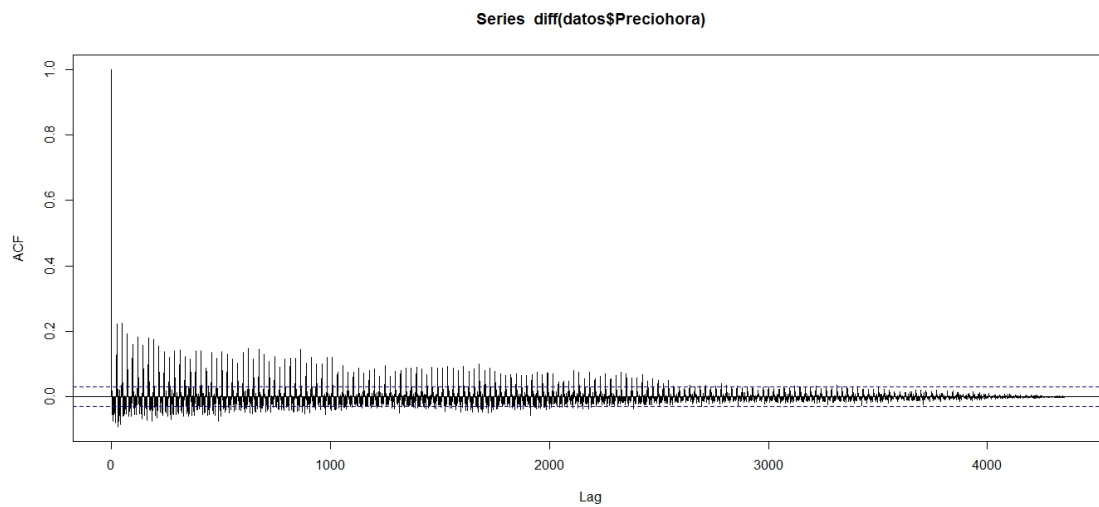


Figura B.27.: Gráfico de correlograma parcial en primera diferencia para seis meses de datos.

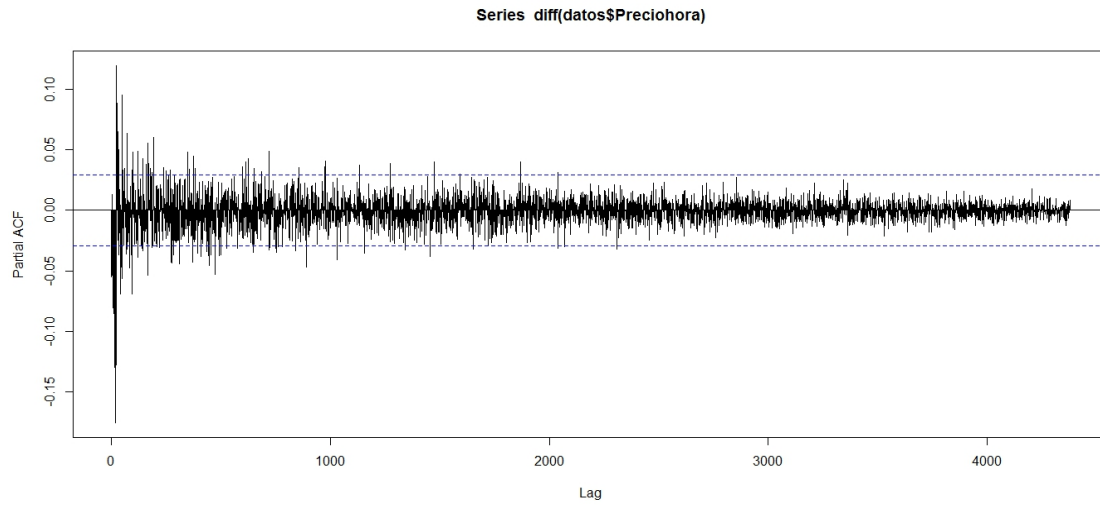


Figura B.28.: Gráfico de correlograma simple en segunda diferencia para seis meses de datos.

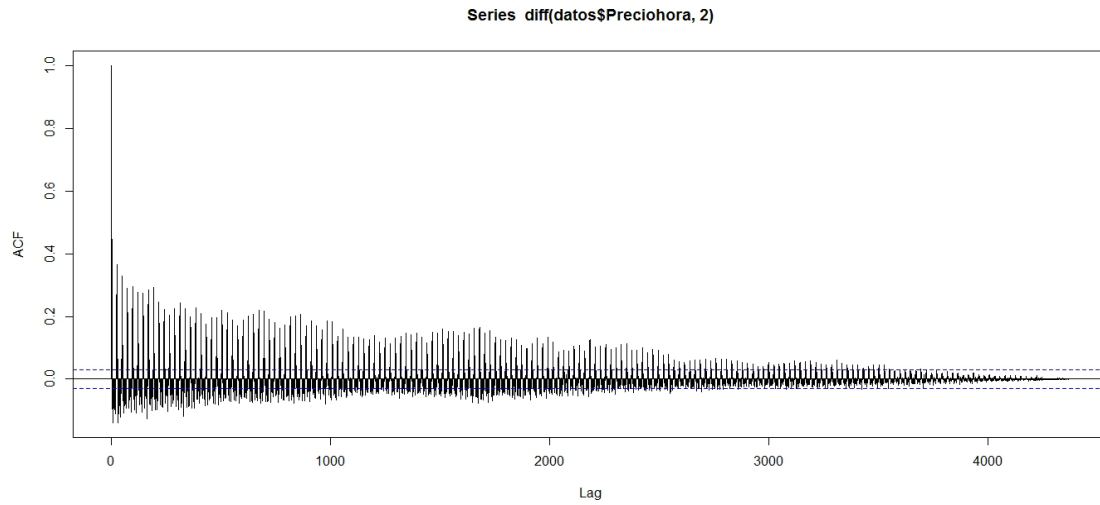
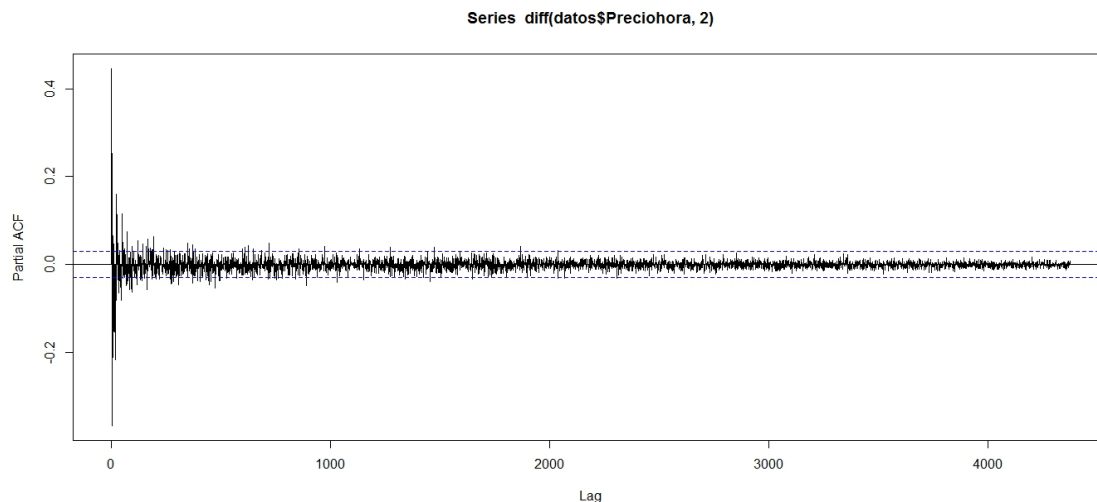


Figura B.29.: Gráfico de correlograma parcial en segunda diferencia para seis meses de datos.



Como resultado de la observación hecha a las gráficas se tiene como modelo un ARIMA (5, 1, 3)

Estimación de parámetros Se presenta a continuación una tabla con los datos arrojados por R y los resultados del test de significancia de los parámetros hallados, así también, la fórmula final para el modelo.

TABLA B.11.: Parámetros arrojados por R para seis meses de datos

AR	1	2	3	4	5
ϕ	1.3275	-0.0939	-0.3106	0.0308	-0.0608
s.e.	0.2133	0.37	0.1666	0.0253	0.0171

MA	1	2	3
θ	-1.444	0.1278	0.3433
s.e.	0.2132	0.3944	0.1853

TABLA B.12.: Resultado de prueba de significancia para parámetros de seis meses de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	1.3275	0.2133	6.2236	-1.444	0.2132	6.7730
2	-0.0939	0.37	0.2538	0.1278	0.3944	0.3240
3	-0.3106	0.1666	1.8643	0.3433	0.1853	1.8527
4	0.0308	0.0253	1.2174			
5	-0.0608	0.0171	3.5556			

Con los parámetros significativos se tiene la fórmula final que caracteriza a la ventana de 6 meses años:

$$y(t) = 1,3275y(t - 1) - 0,0608y(t - 5) - 1,4440e(t - 1)$$

Diagnóstico del modelo A continuación se muestran los resultados de σ^2 , log likelihood y aic que dio como resultado R para esta ventana de tiempo. De igual manera se presentan las gráficas de las pruebas de q-q y Ljung-box.

TABLA B.13.: Valores de σ^2 , log likelihood y aic arrojados por R para seis meses de datos

σ^2	Log likelihood	AIC
0.01961	2393.87	-4769.74

Figura B.30.: Resultado de prueba de q.q para seis meses de datos

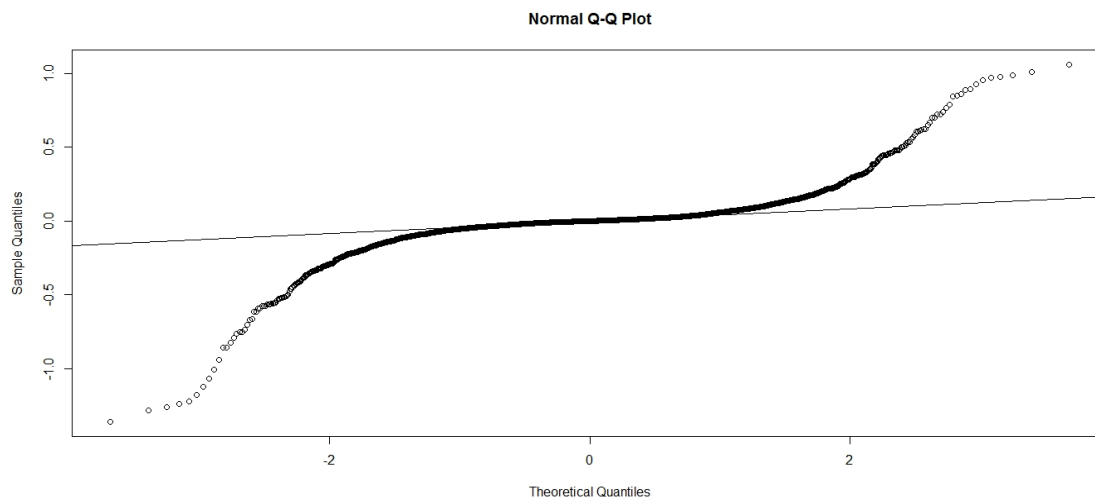
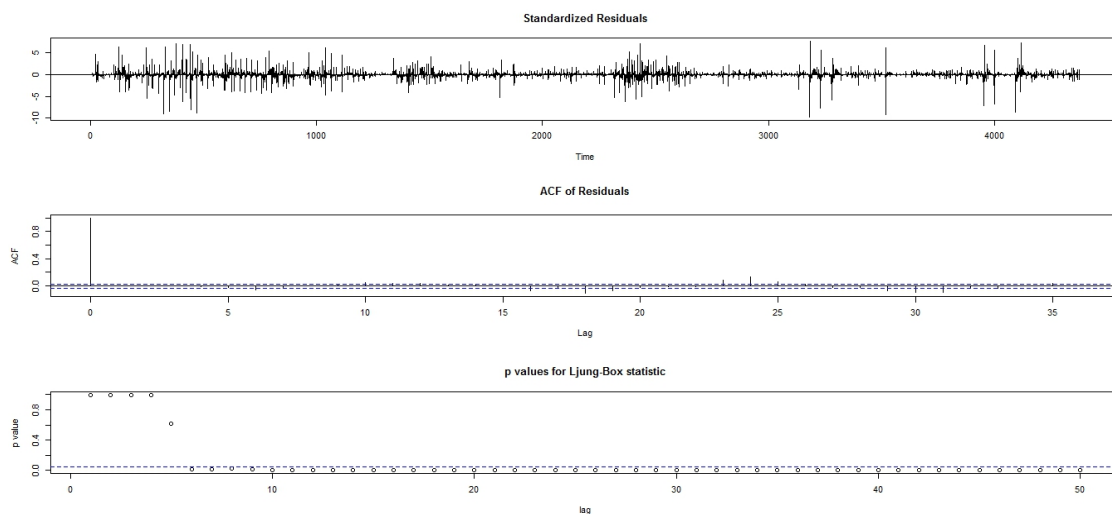


Figura B.31.: Resultado de prueba de Ljung-box para seis meses de datos.



Predicción En esta parte se presentan los resultados de predicción de las 48 horas siguientes a la semana de tiempo con la que se trabaja, así mismo, los gráficos de franja de error de predicción y del comportamiento del precio de la predicción, como también el resultado arrojado como archivo de texto plano por R.

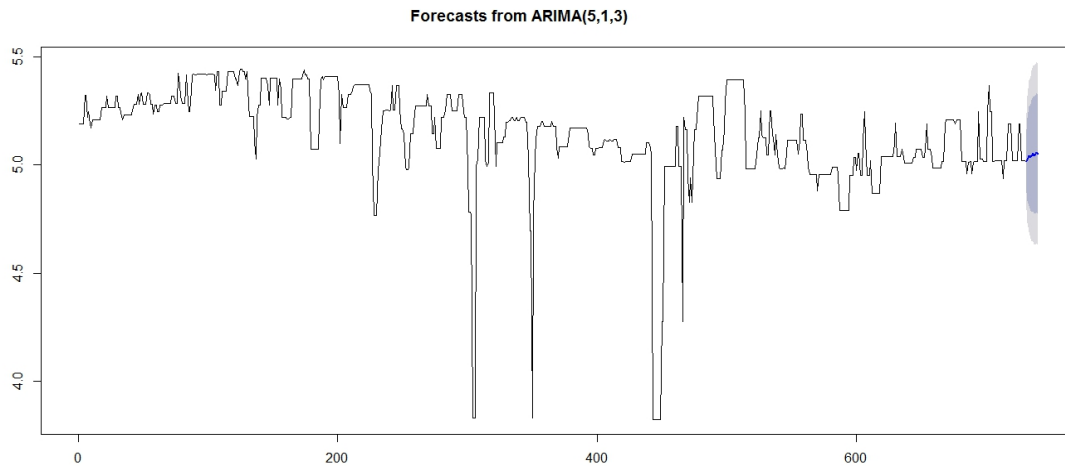
TABLA B.14.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para seis meses de datos.

[1 - 8]	5.0197	5.0244	5.0293	5.0356	5.0421	5.0488	5.0549	5.0603
[9 - 16]	5.0647	5.0679	5.0698	5.0705	5.07	5.0686	5.0664	5.0636
[17 - 24]	5.0606	5.0575	5.0545	5.0518	5.0469	5.0479	5.0468	5.0463
[25 - 32]	5.0463	5.0468	5.0478	5.049	5.0504	5.0518	5.0533	5.0546
[33 - 40]	5.0557	5.0566	5.0572	5.0575	5.0576	5.0574	5.057	5.0565
[41 - 48]	5.0559	5.0552	5.0545	5.0539	5.0533	5.0529	5.0525	5.0523

[1 - 8]	0.1400	0.1868	0.2152	0.2334	0.2467	0.2547	0.2591	0.2612
[9 - 16]	0.0262	0.2622	0.2622	0.2622	0.2623	0.2623	0.2623	0.2623
[17 - 24]	0.2626	0.2634	0.2648	0.267	0.2701	0.2739	0.2784	0.2833
[25 - 32]	0.2883	0.2933	0.2979	0.3021	0.3057	0.3089	0.3115	0.3137
[33 - 40]	0.3156	0.3171	0.3185	0.3197	0.3208	0.3219	0.3231	0.3243
[41 - 48]	0.3256	0.3271	0.3287	0.3305	0.3324	0.3344	0.3366	0.3388

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho

Figura B.32.: Gráfico de franja de error de predicción para seis meses de datos con acercamiento al último mes.

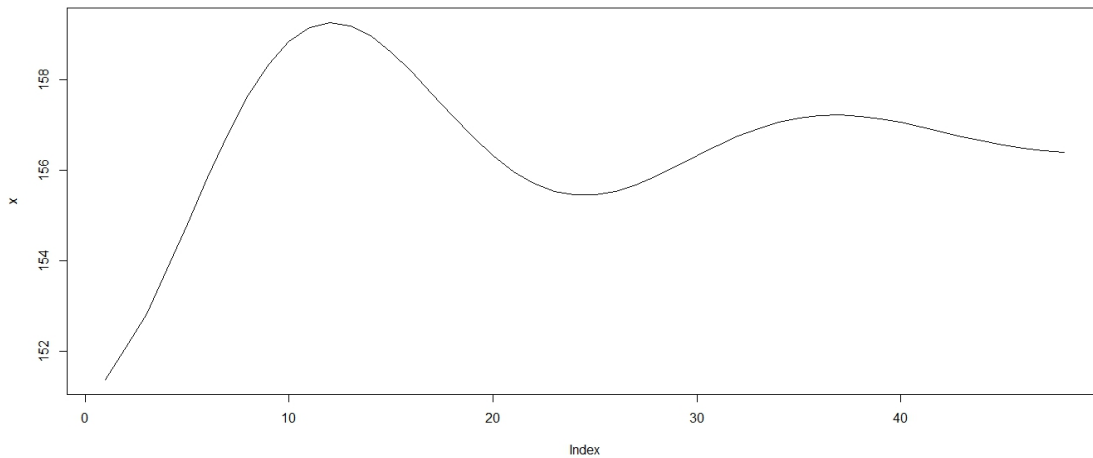


Los resultados dados en texto plano son:

TABLA B.15.: Resultados finales de datos de predicción.

1	151.3686118	17	157.6907758	33	156.9224917
2	152.0881064	18	157.1989185	34	157.0609584
3	152.8314989	19	156.7300659	35	157.1571091
4	153.8042977	20	156.3113732	36	157.2096084
5	154.8012444	21	155.9636551	37	157.2201418
6	155.837999	22	155.7006105	38	157.192945
7	156.7985719	23	155.5289674	39	157.1340087
8	157.6510427	24	155.4490461	40	157.0509068
9	158.3402185	25	155.4551087	41	156.951682
10	158.8451773	26	155.5368996	42	156.8446774
11	159.1499755	27	155.6806821	43	156.7375891
12	159.2588714	28	155.8699491	44	156.6374658
13	159.1861067	29	156.0876955	45	156.5496168
14	158.9581564	30	156.3165316	46	156.4785594
15	158.6078812	31	156.5408503	47	156.4267736
16	158.172625	32	156.7465234	48	156.395335

Figura B.33.: Gráfico de valores de predicción de precio de 48 horas futuras para seis meses de datos.



B.5 VENTANA DE TIEMPO DE TRES MESES

Es la ventana que comprende desde el inicio de Octubre de 2014 hasta el final del mismo año, es decir, los últimos tres meses de la base de datos.

Especificación del modelo A continuación se muestran las gráficas de datos de la ventana de tiempo, las autocorrelaciones simple y parcial, en su estado normal, en primera y segunda diferencia

Figura B.34.: Análisis del comportamiento de tres meses de datos mediante el uso de R

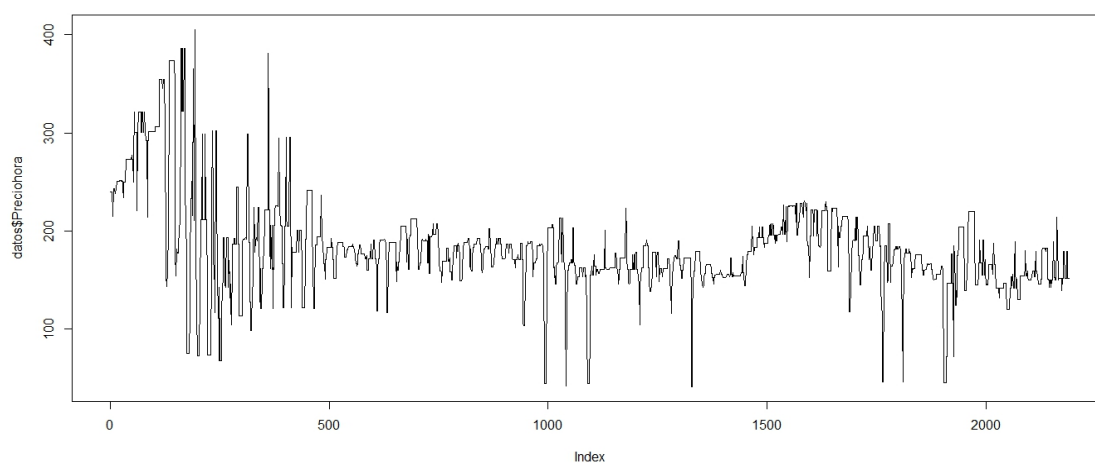


Figura B.35.: Gráfico de correlograma simple para tres meses de datos

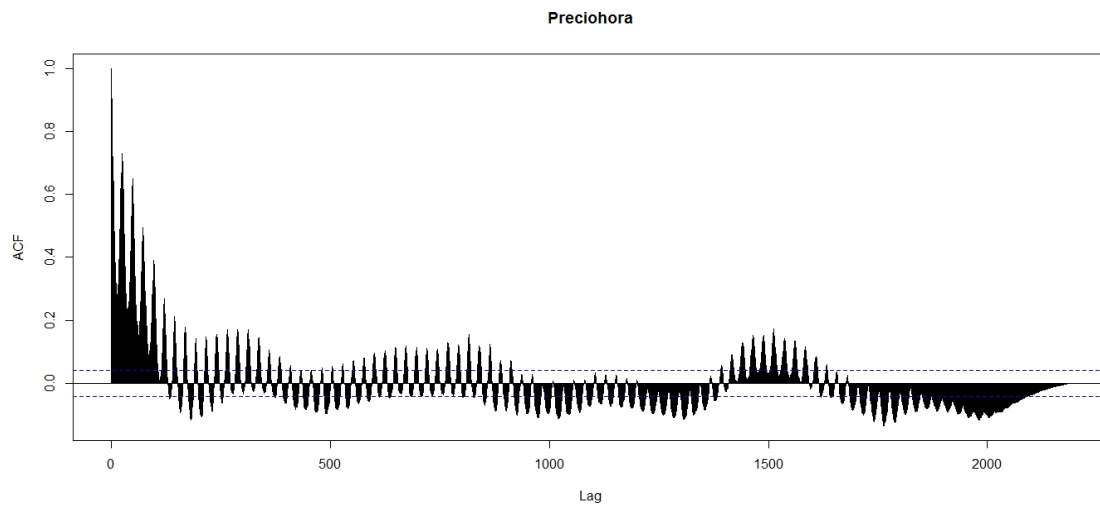


Figura B.36.: Gráfico de correlograma parcial para tres meses de datos

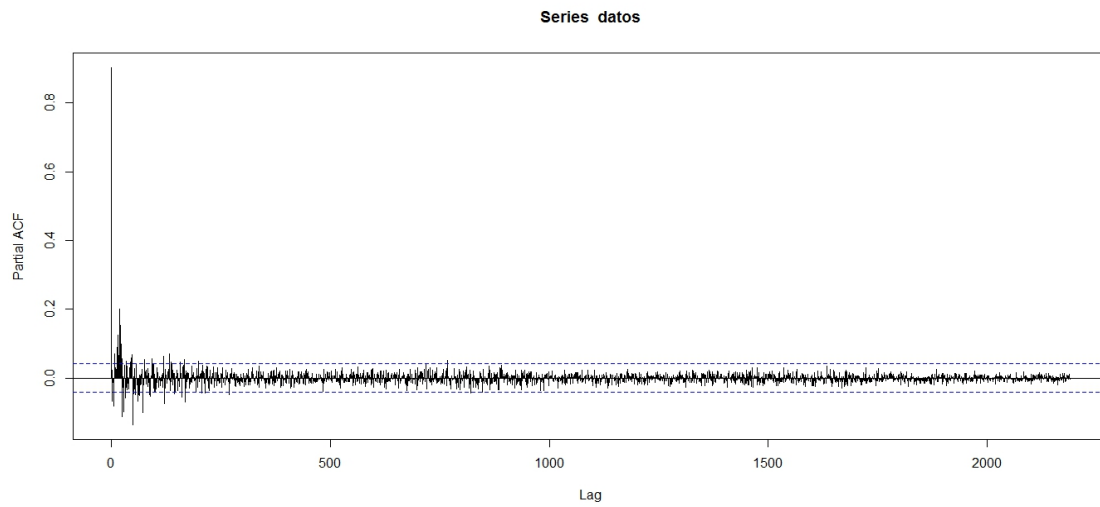


Figura B.37.: Gráfico de correlograma simple en primera diferencia para tres meses de datos.

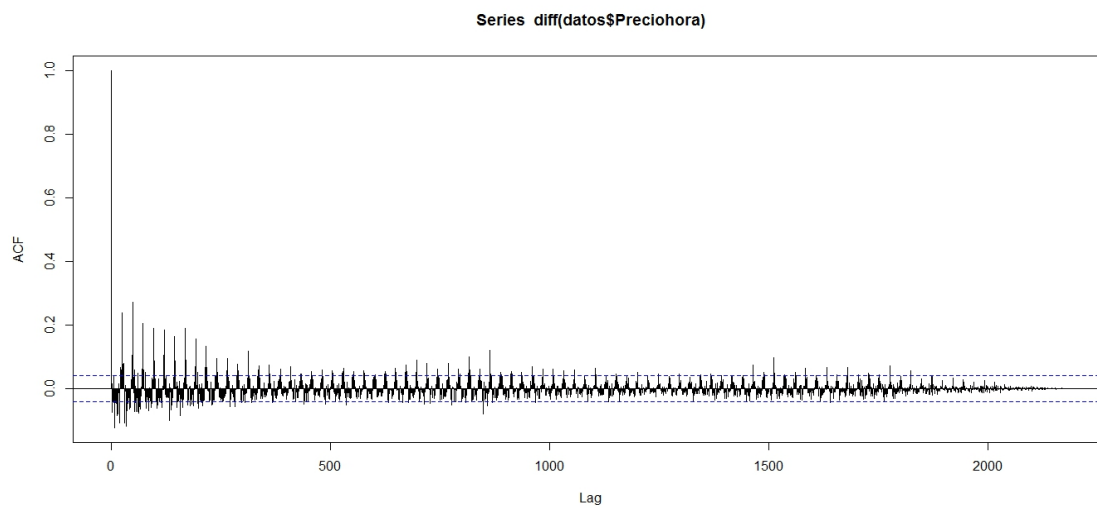


Figura B.38.: Gráfico de correlograma parcial en primera diferencia para tres meses de datos.

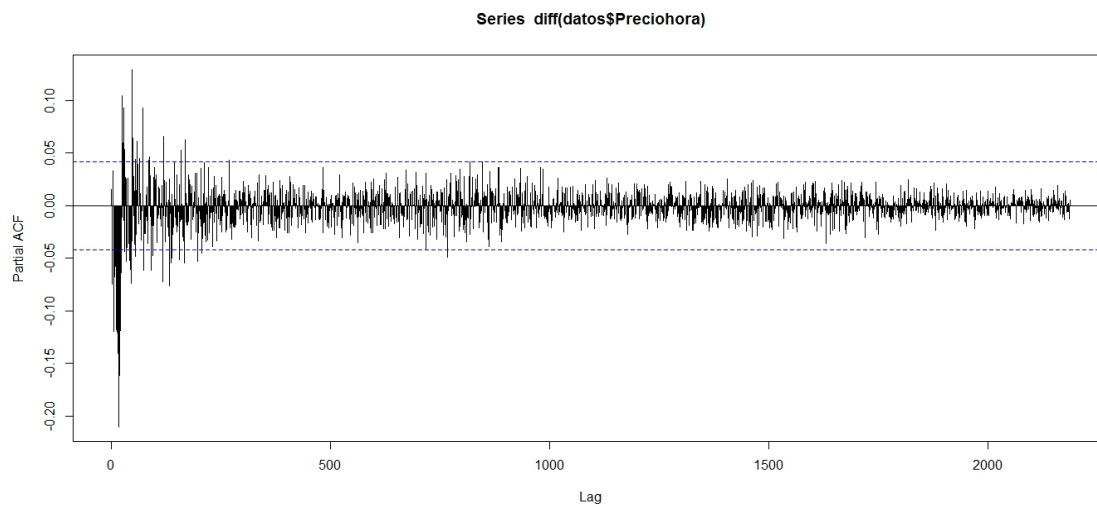


Figura B.39.: Gráfico de correlograma simple en segunda diferencia para tres meses de datos.

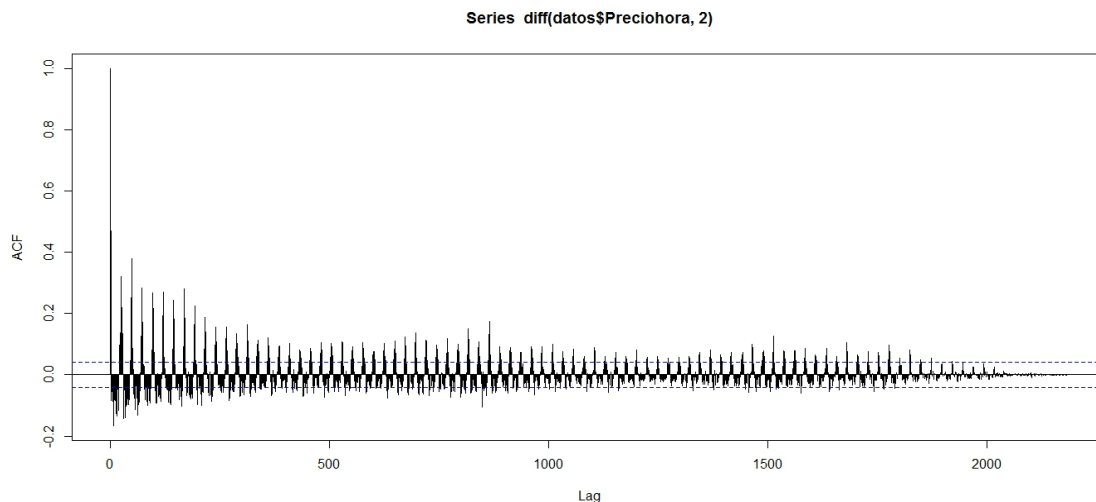
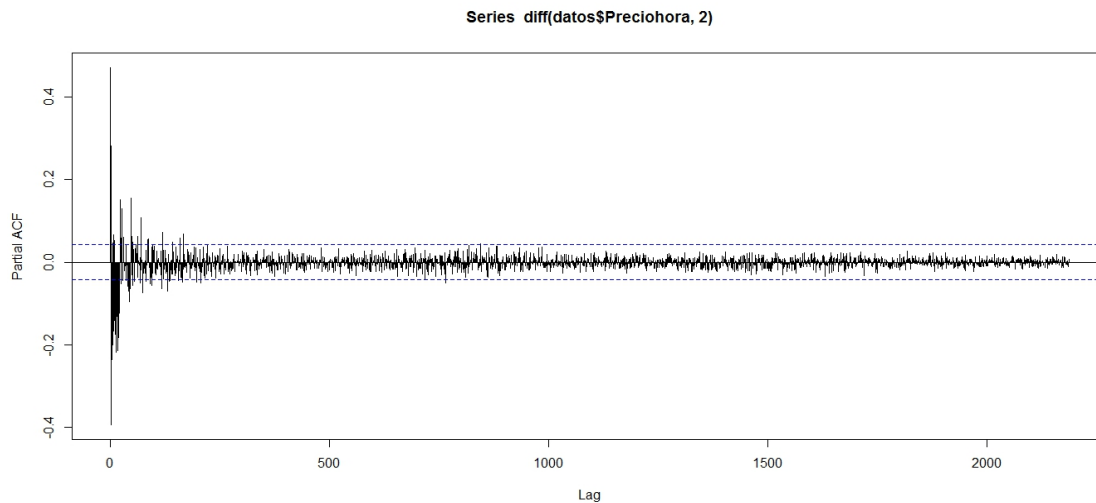


Figura B.40.: Gráfico de correlograma parcial en segunda diferencia para tres meses de datos.



Como resultado de la observación hecha a las gráficas se tiene como modelo un ARIMA (18, 1, 1)

Estimación de parámetros Se presenta a continuación una tabla con los datos arrojados por R y los resultados del test de significancia de los parámetros hallados, así también, la

fórmula final para el modelo.

TABLA B.16.: Parámetros arrojados por R para tres meses de datos

AR	1	2	3	4	5	6	7	8	9
ϕ	0.5846	-0.0515	-0.0912	-0.0052	-0.0554	-0.1555	-0.0213	-0.0766	-0.046
s.e.	0.0397	0.0247	0.0248	0.0251	0.0247	0.0248	0.0256	0.0252	0.0255
AR	10	11	12	13	14	15	16	17	18
ϕ	-0.031	-0.0352	-0.0924	-0.0494	-0.0587	-0.0567	-0.416	-0.015	-0.1409
s.e.	0.0253	0.025	0.025	0.0249	0.025	0.205	0.0249	0.0248	0.0236

MA	1
θ	-0.7063
s.e.	0.0353

TABLA B.17.: Resultado de prueba de significancia para parámetros de tres meses de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	0.5846	0.0397	14.7254	-0.7063	0.0353	20.0085
2	-0.0515	0.0247	2.0850			
3	-0.0912	0.0248	3.6774			
4	-0.0052	0.0251	0.2072			
5	-0.0554	0.0247	2.2429			
6	-0.1555	0.0248	6.2702			
7	-0.0213	0.0256	0.8320			
8	-0.0766	0.0252	3.0397			
9	-0.046	0.0255	1.8039			
10	-0.031	0.0253	1.2253			
11	-0.0352	0.025	1.4080			
12	-0.0924	0.025	3.6960			
13	-0.0494	0.0249	1.9839			
14	-0.0587	0.025	2.3480			
15	-0.0567	0.205	0.2766			
16	-0.416	0.0249	16.7068			
17	-0.015	0.0248	0.6048			
18	-0.1409	0.0236	5.9703			

Con los parámetros significativos se tiene la fórmula final que caracteriza a la ventana de 5 años:

$$y(t) = 0,5846y(t - 1) - 0,0515y(t - 2) - 0,0912y(t - 3) - 0,0554y(t - 5) - 0,1555y(t - 6)$$

$$-0,0766y(t-8)-0,0924y(t-12)-0,0587y(t-14)-0,416y(t-16)-0,1409y(t-18)-0,7063e(t-1)$$

Diagnóstico del modelo A continuación se muestran los resultados de σ^2 , log likelihood y aic que dio como resultado R para esta ventana de tiempo. De igual manera se presentan las gráficas de las pruebas de q-q y Ljung-box.

TABLA B.18.: Valores de σ^2 , log likelihood y aic arrojados por R para tres meses de datos

σ^2	Log likelihood	AIC
0.01571	1438.61	-2837.21

Figura B.41.: Resultado de prueba de q.q para tres meses de datos

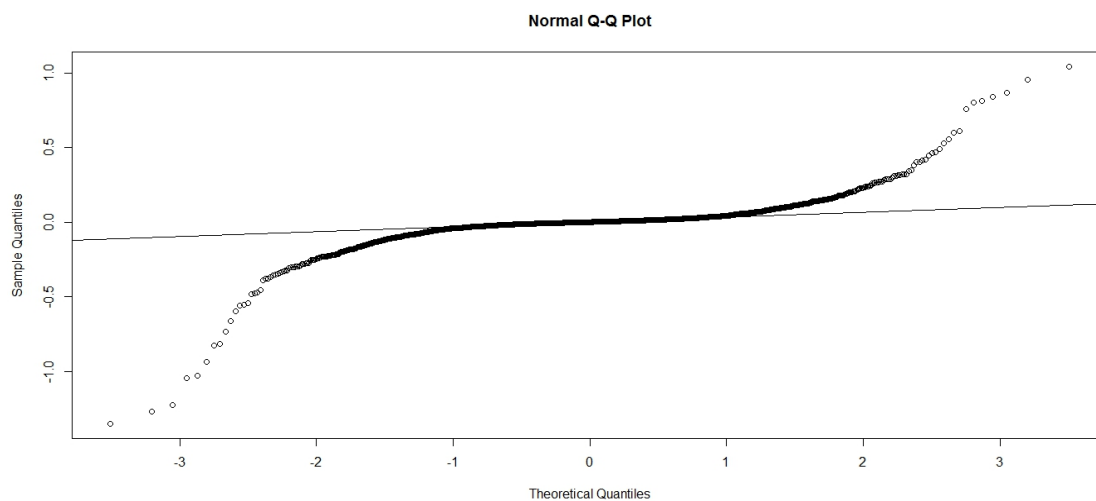
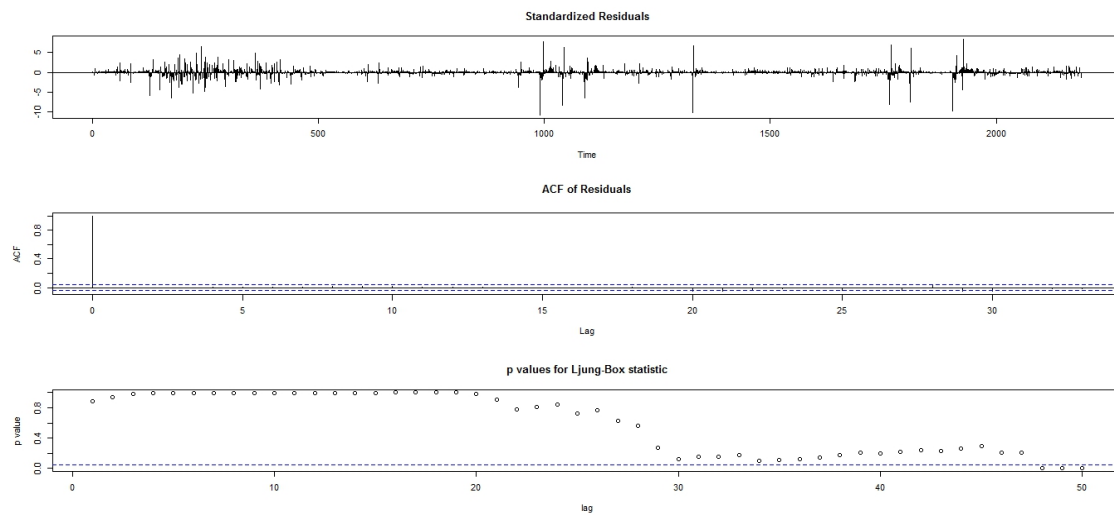


Figura B.42.: Resultado de prueba de Ljung-box para tres meses de datos



Predicción En esta parte se presentan los resultados de predicción de las 48 horas siguientes a la semana de tiempo con la que se trabaja, así mismo, los gráficos de franja de error de predicción y del comportamiento del precio de la predicción, como también el resultado arrojado como archivo de texto plano por R.

ANEXO B. RESULTADOS DE MODELADO Y PREDICCIÓN.

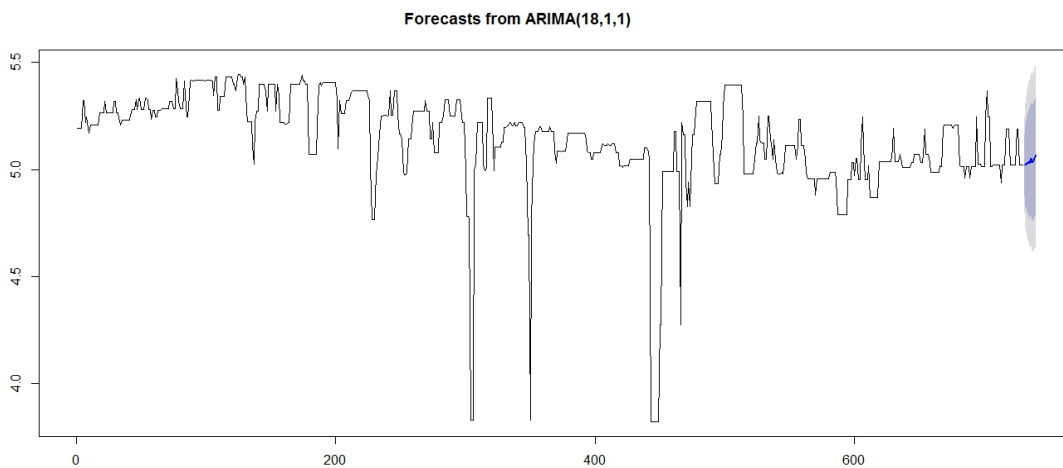
TABLA B.19.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para tres meses de datos.

[1 - 8]	5.0044	4.9948	5.0046	5.0155	5.0159	5.0244	5.0236	5.0452
[9 - 16]	5.0621	5.0692	5.0729	5.0783	5.0636	5.0529	5.067	5.0735
[17 - 24]	5.0741	5.0695	5.068	5.0653	5.0579	5.049	5.0411	5.034
[25 - 32]	5.0308	5.028	5.0263	5.0269	5.0298	5.0334	5.039	5.0454
[33 - 40]	5.0491	5.0517	5.0541	5.0571	5.0593	5.0606	5.0618	5.0627
[41 - 48]	5.0628	5.0621	5.0602	5.0573	5.054	5.0504	5.0469	5.0436

[1 - 8]	0.1253	0.1668	0.1918	0.206	0.216	0.2229	0.2252	0.2267
[9 - 16]	0.2266	0.2267	0.2268	0.2271	0.2272	0.2273	0.2273	0.2274
[17 - 24]	0.2274	0.2276	0.2276	0.2276	0.2279	0.2289	0.2311	0.2344
[25 - 32]	0.2391	0.2446	0.2503	0.2556	0.26	0.2633	0.2658	0.2677
[33 - 40]	0.2691	0.2701	0.2709	0.2714	0.2719	0.2722	0.2726	0.2728
[41 - 48]	0.273	0.2732	0.27358	0.274	0.2746	0.2756	0.2769	0.2787

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho

Figura B.43.: Gráfico de franja de error de predicción para tres meses de datos con acercamiento al último mes.

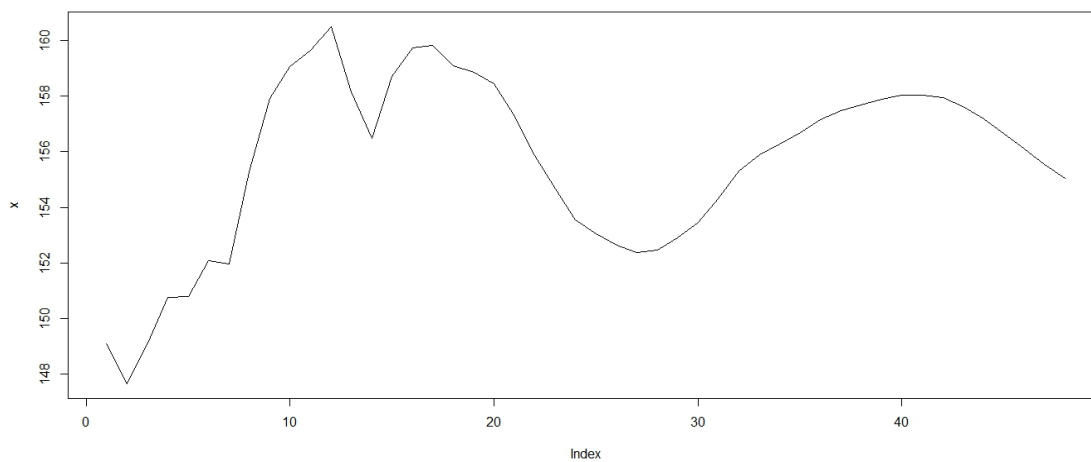


Los resultados dados en texto plano son:

TABLA B.20.: Resultados finales de datos de predicción.

1	149.074473	17	154.0590545	33	152.7141694
2	147.6478431	18	154.3722664	34	156.7758377
3	149.1056328	19	159.1507713	35	157.5753884
4	150.7459546	20	161.5148356	36	156.0124794
5	150.7978201	21	159.8610499	37	153.7979918
6	152.0861292	22	157.7481857	38	153.4632309
7	151.9576707	23	154.8003156	39	154.2681003
8	155.2798542	24	152.2749853	40	153.8592156
9	157.9274894	25	148.3250279	41	153.7427883
10	159.0565819	26	144.1084297	42	154.716746
11	159.6374011	27	140.9267145	43	157.7636457
12	160.5083554	28	139.6041615	44	160.2019169
13	158.1620278	29	141.7902709	45	159.8041495
14	156.4823149	30	144.6197668	46	157.6102164
15	158.6981547	31	147.3030457	47	154.5485057
16	159.735129	32	149.4845436	48	152.2937162

Figura B.44.: Gráfico de valores de predicción de precio de 48 horas futuras para tres meses de datos.



B.6 VENTANA DE TIEMPO DE UNA SEMANA

Es la ventana que comprende desde la última semana del año 2014, es decir, las últimas 168 horas de la base de datos.

Especificación del modelo A continuación se muestran las gráficas de datos de la ventana de tiempo, las autocorrelaciones simple y parcial, en su estado normal, en primera y segunda diferencia

Figura B.45.: Análisis del comportamiento de una semana de datos mediante el uso de R

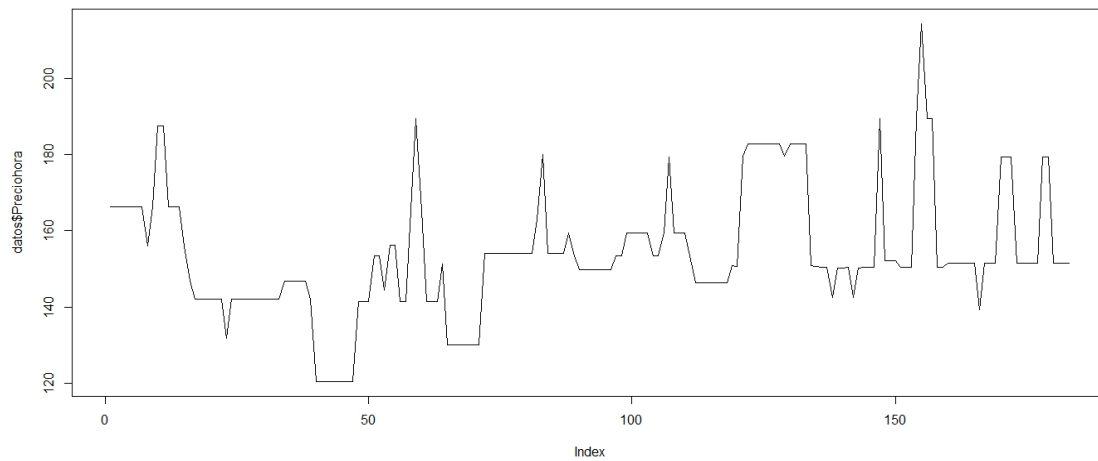


Figura B.46.: Gráfico de correlograma simple para una semana de datos

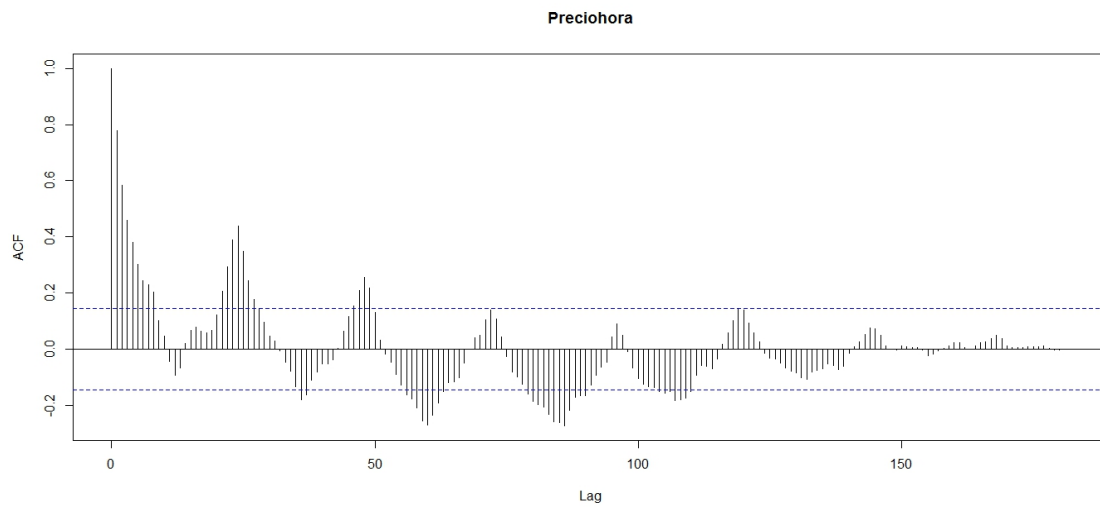


Figura B.47.: Gráfico de correlograma parcial para una semana de datos

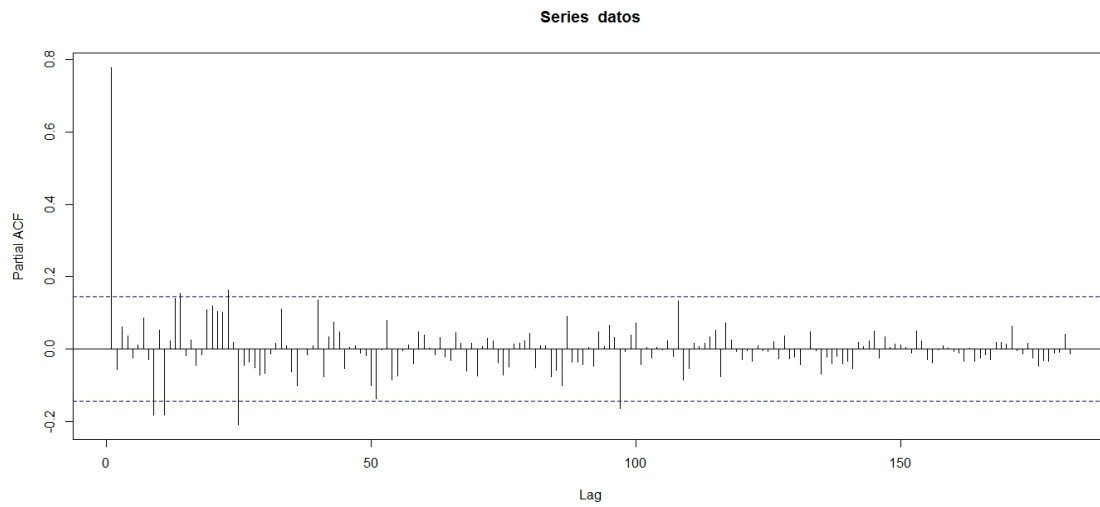


Figura B.48.: Gráfico de correlograma simple en primera diferencia para una semana de datos.

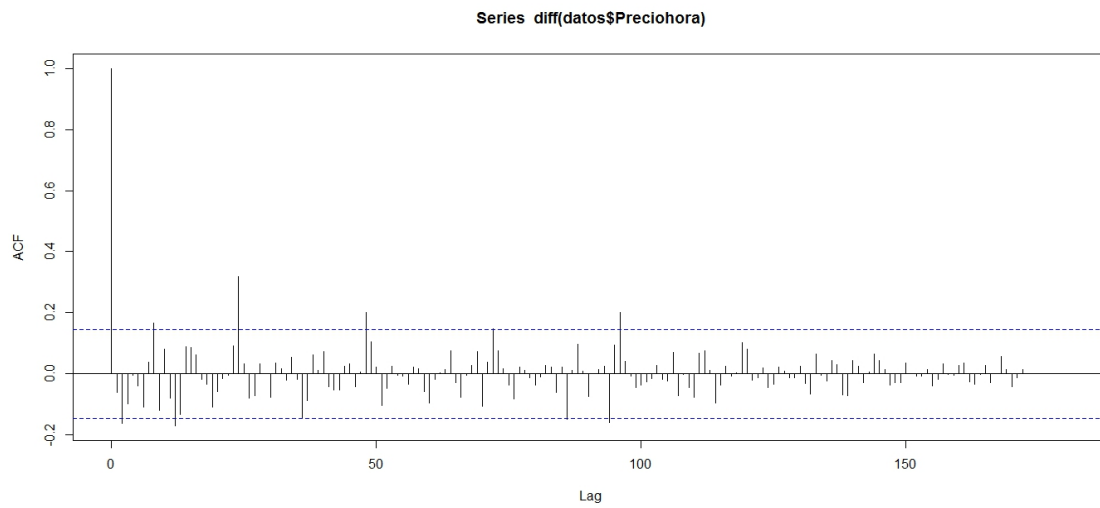


Figura B.49.: Gráfico de correlograma parcial en primera diferencia para una semana de datos.

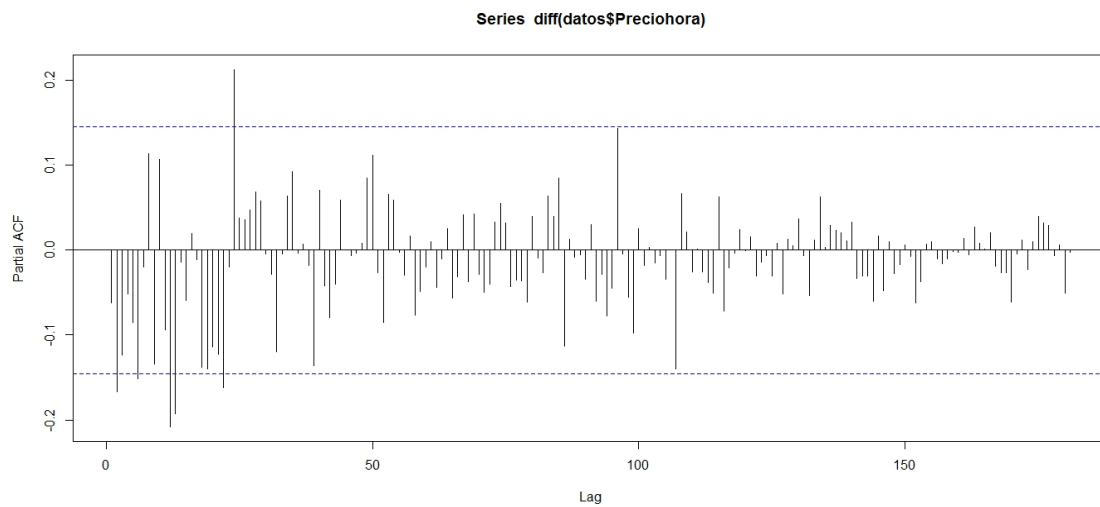


Figura B.50.: Gráfico de correlograma simple en segunda diferencia para una semana de datos.

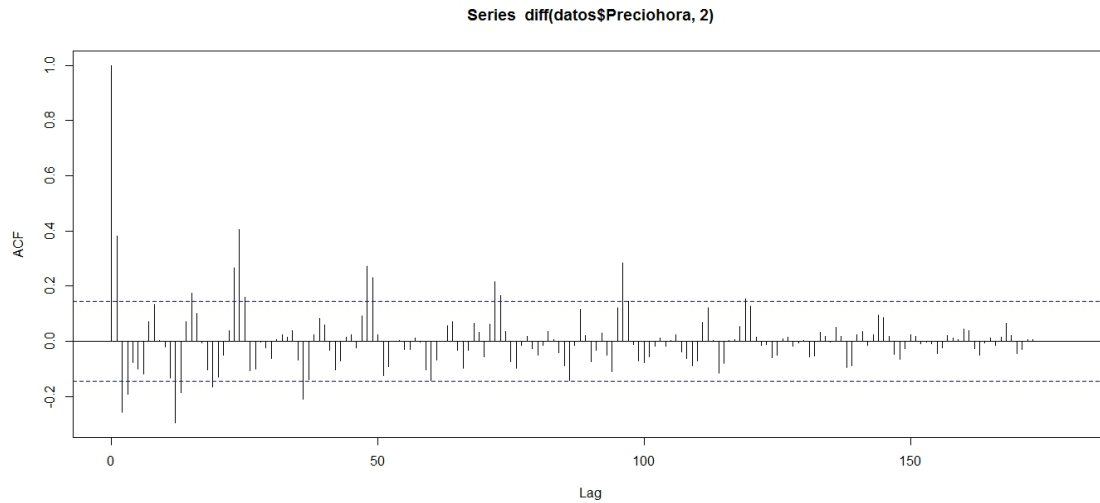
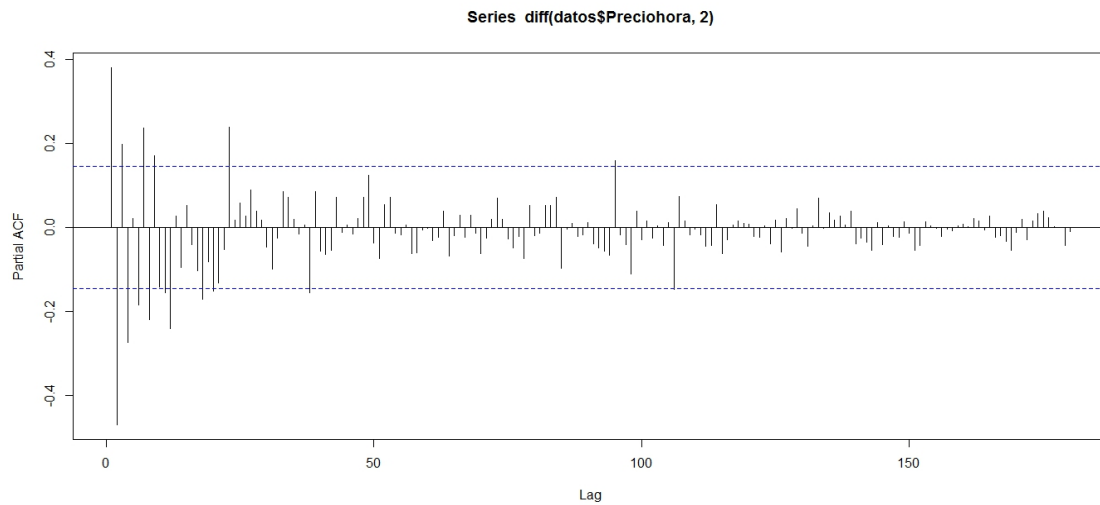


Figura B.51.: Gráfico de correlograma parcial en segunda diferencia para una semana de datos.



Como resultado de la observación hecha a las gráficas se tiene como modelo un ARIMA (24, 1, 25)

Estimación de parámetros Se presenta a continuación una tabla con los datos arrojados por R y los resultados del test de significancia de los parámetros hallados, así también, la

fórmula final para el modelo.

TABLA B.21.: Parámetros arrojados por R para una semana de datos

AR	1	2	3	4	5	6	7	8	9
ϕ	-0.1596	-0.0127	-0.1585	0.1082	0.077	-0.0846	-0.0613	-0.2208	0.0658
s.e.	0.0815	0.0357	0.1049	0.0644	0.0634	0.0497	0.077	0.0444	0.0155
AR	10	11	12	13	14	15	16	17	18
ϕ	-0.1788	0.1569	-0.1832	-0.0816	0.0222	-0.1432	0.175	-0.1456	-0.1079
s.e.	NAN	0.068	0.08	0.0184	0.055	0.0975	0.0597	NAN	0.0753
AR	19	20	21	22	23	24			
ϕ	-0.1576	-0.0574	0.1491	-0.1033	0.1742	0.7122			
s.e.	0.0512	NAN	0.0359	0.0348	0.0331	0.0659			
MA	1	2	3	4	5	6	7	8	9
θ	0.1128	-0.2589	-0.0245	-0.1742	-0.0555	0.0485	0.0442	0.4052	-0.3928
s.e.	0.1095	0.0365	0.1235	NAN	0.0701	0.0809	NAN	NAN	NAN
MA	10	11	12	13	14	15	16	17	18
θ	0.1861	-0.3111	0.0671	-0.0558	0.0111	0.2888	-0.1431	0.1749	0.0672
s.e.	0.0567	0.0932	NAN	NAN	0.1164	0.1122	0.00391	NAN	0.0744
MA	19	20	21	22	23	24	25		
θ	-0.1834	-0.1305	-0.3576	0.3526	-0.1254	-0.6168	0.3001		
s.e.	NAN	NAN	NAN	NAN	0.071	NAN	0.1039		

TABLA B.22.: Resultado de prueba de significancia para parámetros de una semana de datos

No.	AR	Error Ar (s.e.)	Significativos	MA	Error MA (s.e.)	Significativos
1	-0.1596	0.0815	1.9583	0.1128	0.1095	1.0301
2	-0.0127	0.0357	0.3557	-0.2589	0.0365	7.0932
3	-0.1585	0.1049	1.5110	-0.0245	0.1235	0.1984
4	0.1082	0.0644	1.6801	-0.1742	NAN	-0.1742
5	0.077	0.0634	1.2145	-0.0555	0.0701	0.7917
6	-0.0846	0.0497	1.7022	0.0485	0.0809	0.5995
7	-0.0613	0.077	0.7961	0.0442	NAN	0.0442
8	-0.2208	0.0444	4.9730	0.4052	NAN	0.4052
9	0.0658	0.0155	4.2452	-0.3928	NAN	-0.3928
10	-0.1788	NAN	-0.1788	0.1861	0.0567	3.2822
11	0.1569	0.068	2.3074	-0.3111	0.0932	3.3380
12	-0.1832	0.08	2.2900	0.0671	NAN	0.0671
13	-0.0816	0.0184	4.4348	-0.0558	NAN	-0.0558
14	0.0222	0.055	0.4036	0.0111	0.1164	0.0954
15	-0.1432	0.0975	1.4687	0.2888	0.1122	2.5740
16	0.175	0.0597	2.9313	-0.1431	0.00391	36.5985
17	-0.1456	NAN	-0.1456	0.1749	NAN	0.1749
18	-0.1079	0.0753	1.4329	0.0672	0.0744	0.9032
19	-0.1576	0.0512	3.0781	-0.1834	NAN	-0.1834
20	-0.0574	NAN	-0.0574	-0.1305	NAN	-0.1305
21	0.1491	0.0359	4.1532	-0.3576	NAN	-0.3576
22	-0.1033	0.0348	2.9684	0.3526	NAN	0.3526
23	0.1742	0.0331	5.2628	-0.1254	0.071	1.7662
24	0.7122	0.0659	10.8073	-0.6168	NAN	-0.6168
25	-	-	-	0.3001	0.1039	2.8884

Con los parámetros significativos se tiene la fórmula final que caracteriza a la ventana de 1 semana:

$$\begin{aligned}
 y(t) = & -0,2208y(t-8) - 0,0658y(t-9) + 0,1569y(t-11) - 0,1832y(t-12) - 0,0816y(t-13) \\
 & + 0,175y(t-16) - 0,1576y(t-19) + 0,1491y(t-21) - 0,1033y(t-22) + 0,1742y(t-23) + 0,7122y(t-1) \\
 & - 0,2589e(t-2) + 0,1861e(t-10) - 0,3111e(t-11) + 0,2888e(t-15) - 0,1431e(t-16) + 0,3001e(t-25)
 \end{aligned}$$

Diagnóstico del modelo A continuación se muestran los resultados de σ^2 , log likelihood y aic que dio como resultado R para esta ventana de tiempo. De igual manera se presentan las gráficas de las pruebas de q-q y Ljung-box.

TABLA B.23.: Valores de σ^2 , log likelihood y aic arrojados por R para una semana de datos

σ^2	Log likelihood	AIC
0.001973	288.36	-476.71

Figura B.52.: Resultado de prueba de q.q para una semana de datos

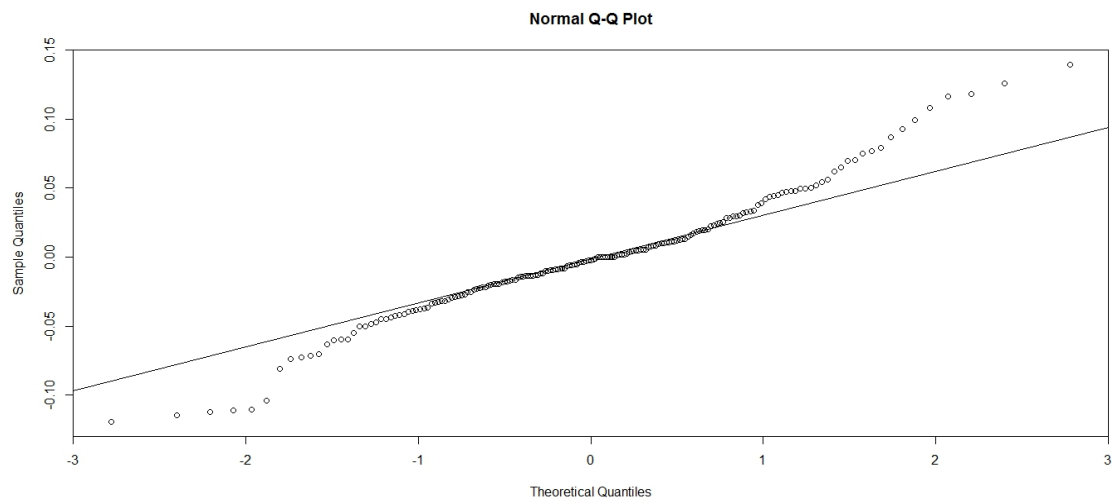
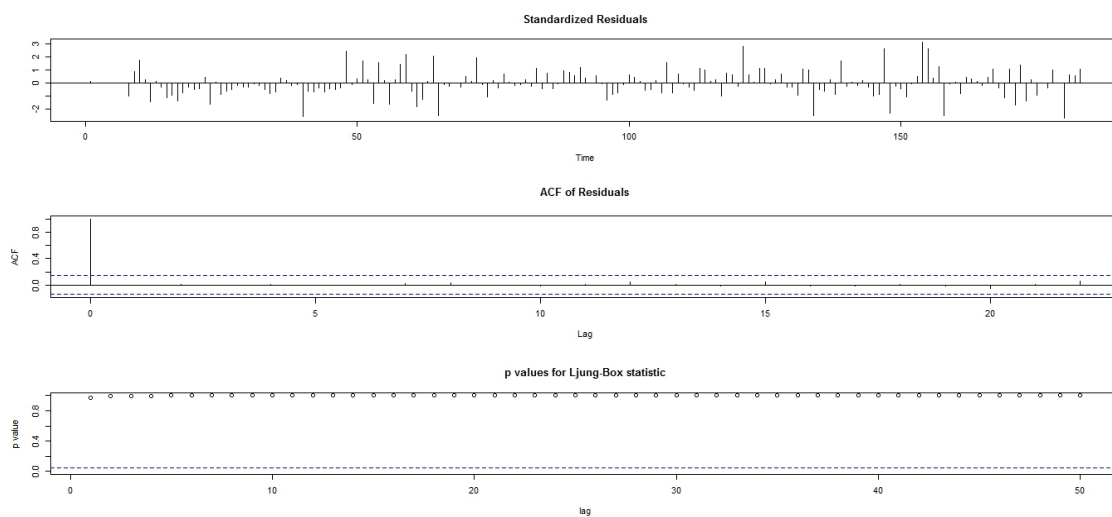


Figura B.53.: Resultado de prueba de Ljung-box para una semana de datos



Predicción En esta parte se presentan los resultados de predicción de las 48 horas siguientes a la semana de tiempo con la que se trabaja, así mismo, los gráficos de franja de error de predicción y del comportamiento del precio de la predicción, como también el resultado arrojado como archivo de texto plano por R.

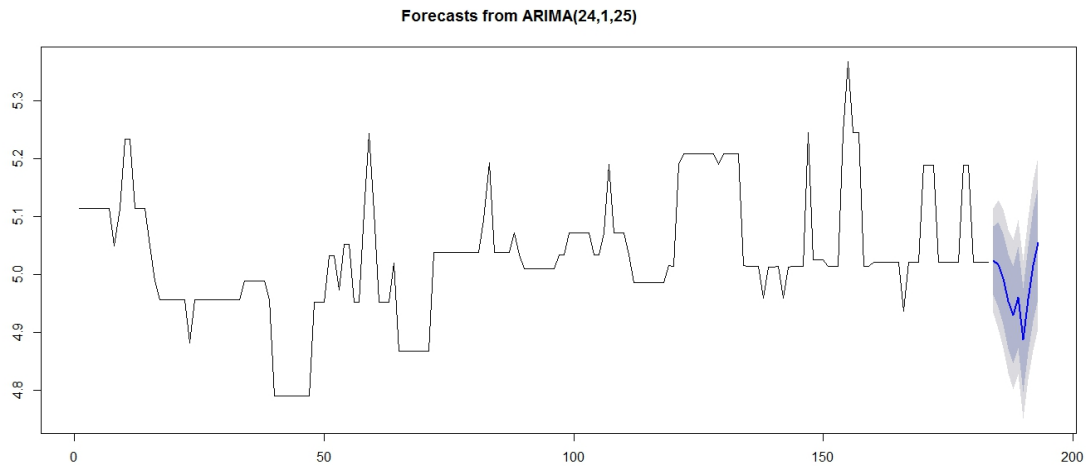
TABLA B.24.: Datos de predicción y el error de cada valor de 48 horas futuras dadas por R para una semana de datos.

[1 - 8]	5.0233	5.0165	4.9926	4.9533	4.9299	4.9607	4.8868	4.9573
[9 - 16]	5.0153	5.0549	5.1491	5.156	5.0881	4.9807	5.0211	5.0858
[17 - 24]	5.1092	5.0875	5.1219	5.1313	5.0706	5.0772	5.0191	5.0254
[25 - 32]	5.0151	5.0243	4.9973	4.9261	4.9256	4.9138	4.9207	4.9832
[33 - 40]	5.0501	5.0494	5.1132	5.1257	5.0608	5.0111	5.0313	5.1222
[41 - 48]	5.1109	5.0836	5.1011	5.0985	5.0754	5.075	5.0356	5.0255

[1 - 8]	0.046	0.0567	0.061	0.0632	0.0655	0.068	0.0696	0.071
[9 - 16]	0.0747	0.0756	0.0773	0.0779	0.0783	0.0783	0.0787	0.0801
[17 - 24]	0.0818	0.0836	0.0845	0.0848	0.0848	0.0848	0.0852	0.0856
[25 - 32]	0.0869	0.0877	0.0882	0.0882	0.0886	0.089	0.0891	0.0895
[33 - 40]	0.09	0.0902	0.0902	0.0902	0.0902	0.0902	0.0903	0.0908
[41 - 48]	0.0914	0.0916	0.0917	0.0917	0.0917	0.0917	0.0918	0.0919

Los intervalos en la parte izquierda de la tabla indican el orden en el que son arrojados los datos por R, siendo por ejemplo, el intervalo [1-8], la forma de indicar que los datos que le preceden son los datos del uno al ocho

Figura B.54.: Gráfico de franja de error de predicción para una semana de datos.

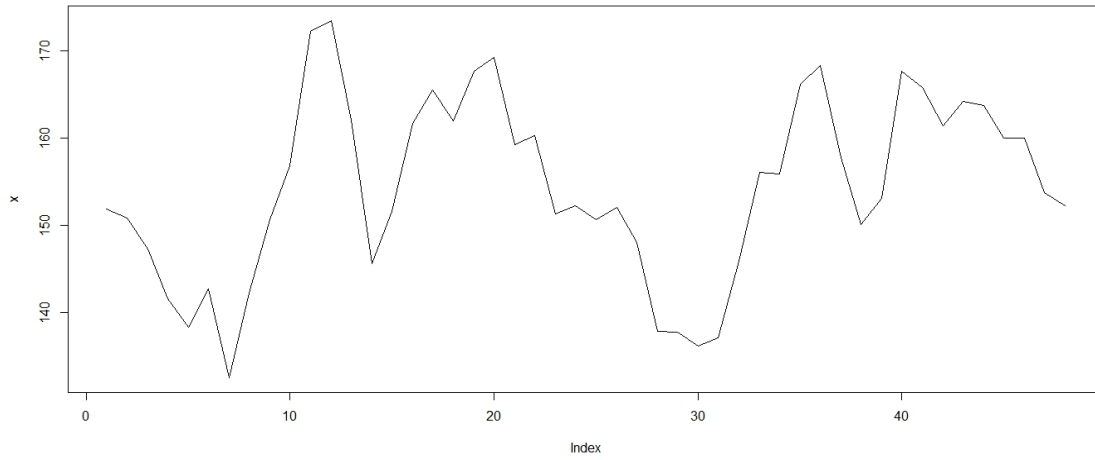


Los resultados dados en texto plano son:

TABLA B.25.: Resultados finales de datos de predicción.

1	151.9175592	17	165.5459831	33	156.0507071
2	150.8930036	18	161.9950858	34	155.9376113
3	147.322049	19	167.654112	35	166.2136485
4	141.6474183	20	169.2498457	36	168.2984683
5	138.3689959	21	159.2856291	37	157.7286261
6	142.7066322	22	160.3334966	38	150.0758751
7	132.5411284	23	151.2852307	39	153.1411431
8	142.2138627	24	152.2329632	40	167.7097824
9	150.702546	25	150.6713538	41	165.8200228
10	156.7902618	26	152.0660552	42	161.3696991
11	172.2774055	27	148.0132801	43	164.2170453
12	173.4801181	28	137.8523245	44	163.7889977
13	162.0889081	29	137.7854824	45	160.0494772
14	145.5828004	30	136.1664453	46	159.978751
15	151.5871944	31	137.1025245	47	153.7921476
16	161.7212218	32	145.9461911	48	152.2561043

Figura B.55.: Gráfico de valores de predicción de precio de 48 horas futuras para una semana de datos.



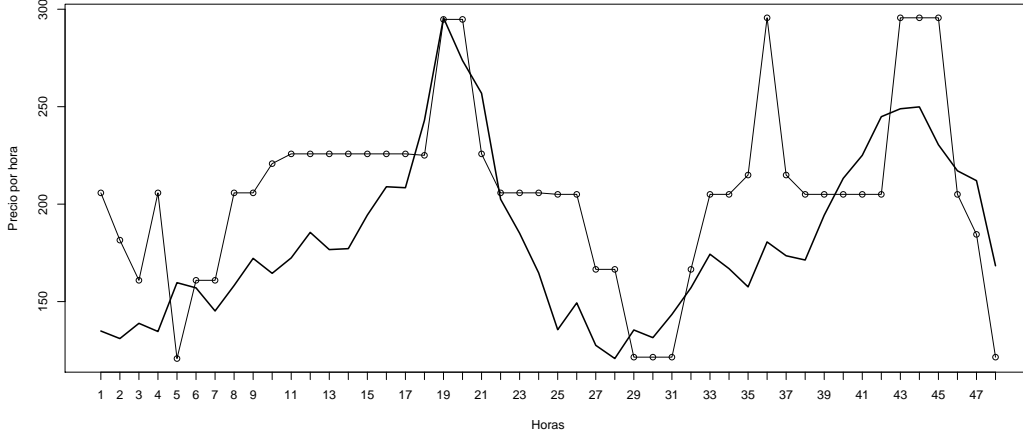
ANEXO C

GRÁFICOS COMPARATIVOS

Como se explicó en el capítulo de análisis de resultados, el modelo matemático es acertado a la hora de pronosticar precios de energía eléctrica, pero hay factores ajenos al modelo que afectan las predicciones que provocan que matemáticamente se arroje un resultado errado. Para entender de una manera más clara este fenómeno, se ilustran los comportamientos de los pronósticos y los precios reales, para ver gráficamente las fluctuaciones y los errores que arrojan las predicciones.

Para comenzar se ilustran los comportamientos de las semanas 8, 9, 5, 7 y 3 que presentaron errores del 0.06 %, 0.44 %, 2.52 %, 2.86 % y 4.47 % respectivamente. Es decir, errores menores al 5 % que son considerados valores muy bajos.

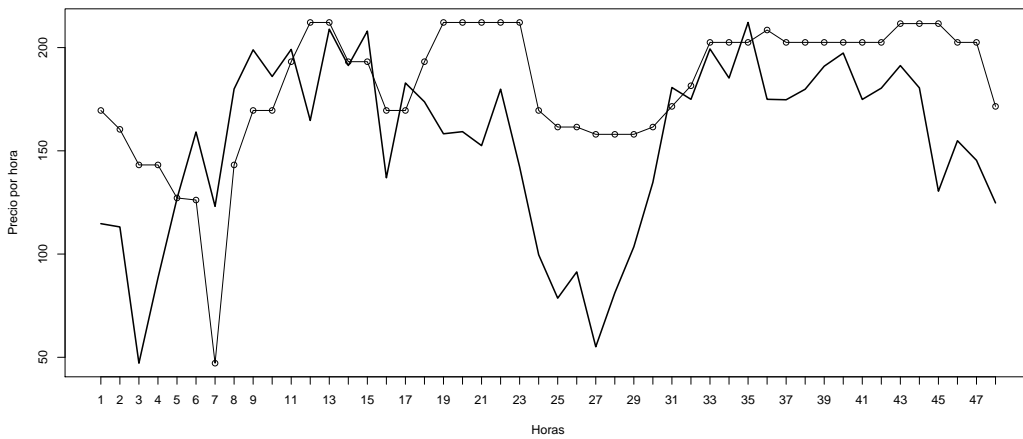
Figura C.1.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 8



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: (XM, 2014b)

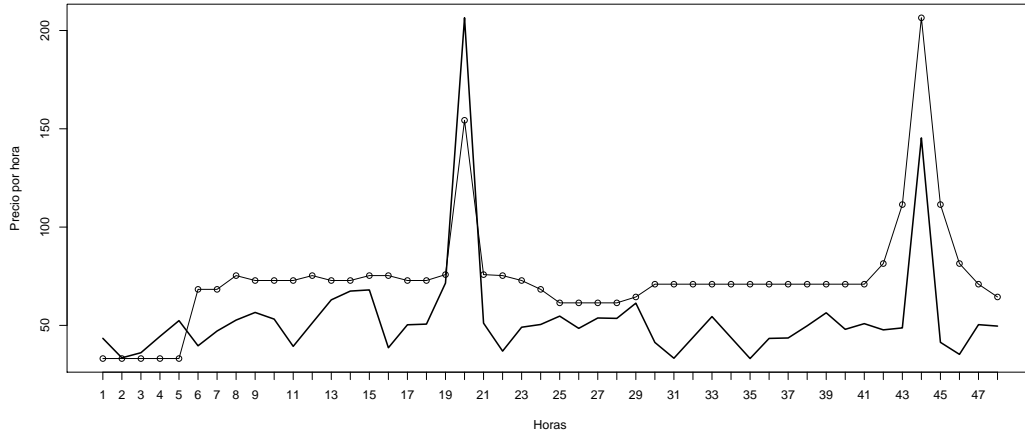
Figura C.2.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 9



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: (XM, 2013)

Figura C.3.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 5

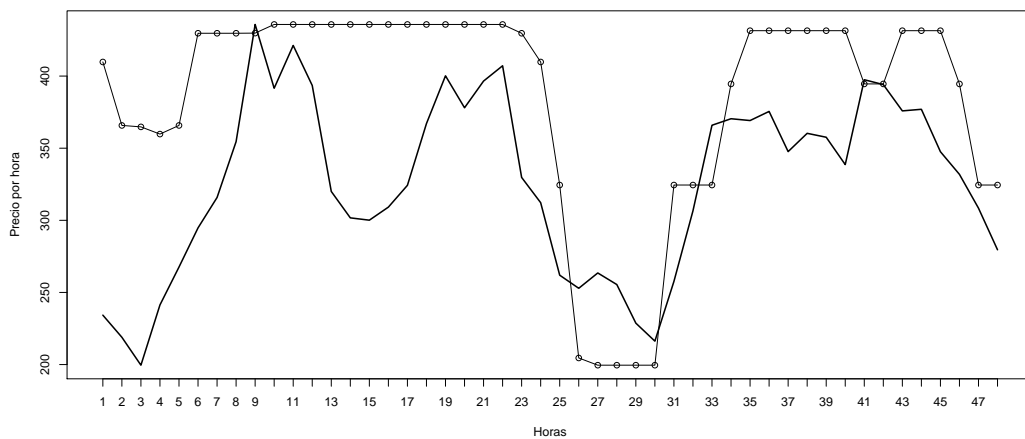


línea continua: datos de predicción

línea con puntos: datos reales

Fuente datos reales: (XM, 2010)

Figura C.4.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 7

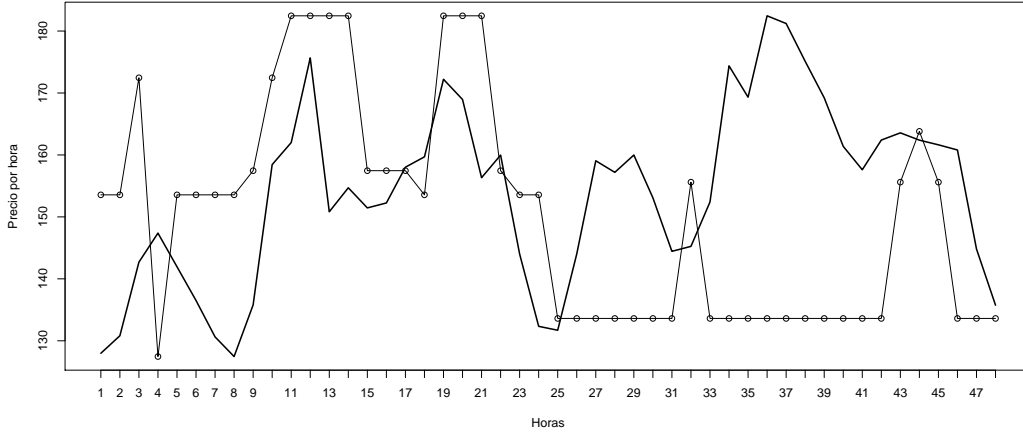


línea continua: datos de predicción

línea con puntos: datos reales

Fuente datos reales: (XM, 2013)

Figura C.5.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 3



línea continua: datos de predicción

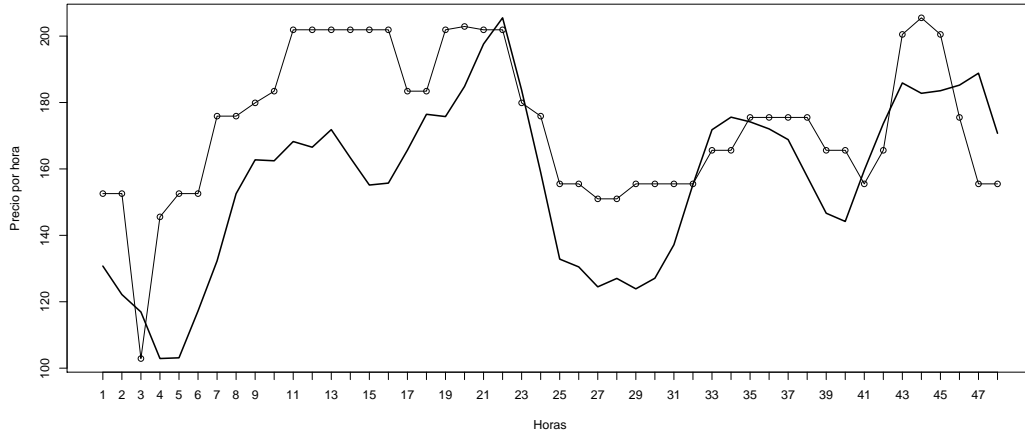
línea con puntos: datos reales

Fuente datos reales: (XM, 2013)

Como se observa en las gráficas comparativas, el pronóstico trata de seguir el comportamiento del precio real, sin importar que existan variaciones drásticas en el precio, lo que importa es que existe una tendencia y el modelo es capaz de predecir siempre y cuando las condiciones ajenas no afecten abruptamente el comportamiento del precio.

El caso siguiente a analizar es la ventana de tiempo semana 1, la cual presenta un error de 5.23 %, es decir, entre 5 % y 10 %, que son errores considerados aceptables. El comportamiento se representa en la siguiente figura.

Figura C.6.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 1



línea continua: datos de predicción

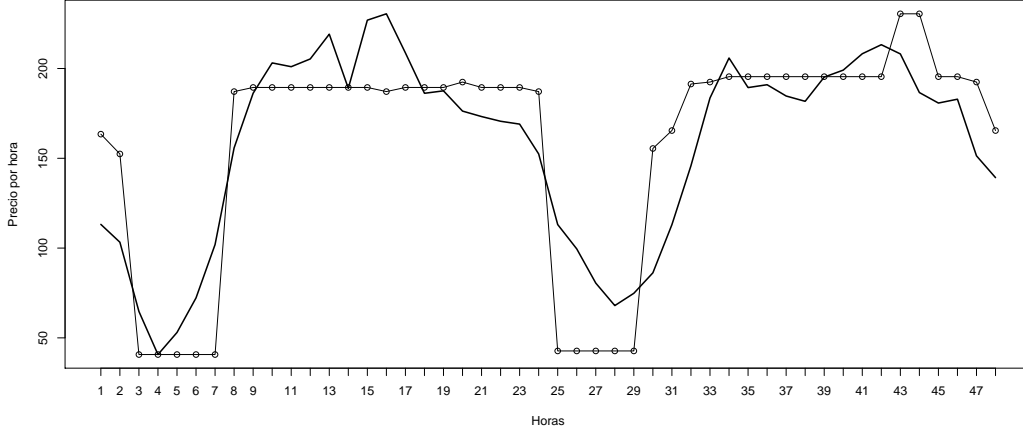
línea con puntos: datos reales

Fuente datos reales: (XM, 2014b)

Como se observa claramente el pronóstico trata de seguir el comportamiento del valor real, pero ese descenso abrupto en el precio de la energía presentado en las primeras 5 horas de predicción, (muy seguramente por cuestiones meteorológicas o por cuestiones de demanda y oferta) se aleja de la tendencia que debería llevar el precio en condiciones normales, es por eso que ahí el modelo no acierta y se presenta un error que aunque es moderado, podría ser más amortizado si ese descenso abrupto no se hubiera presentado.

Y para concluir estas comparaciones se continua con la ilustración de las semanas 6, 4, 10 y 2 cuyos errores fueron 18.10 %, 23.83 %, 54.56 % y 58.68 % respectivamente, lo que representa una estimación completamente errónea pero no a causa de fallas en el modelo, si no por comportamientos anormales que se desvían de la tendencia y hacen que el pronóstico no sea el óptimo.

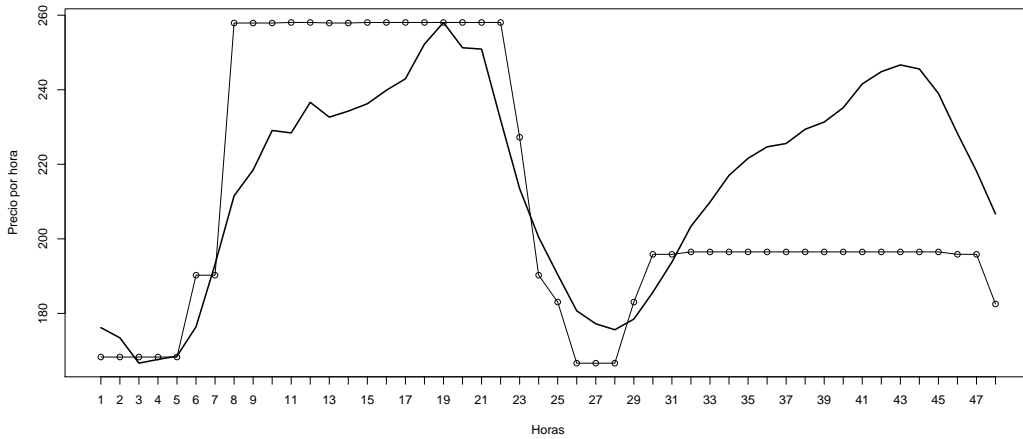
Figura C.7.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 6



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: XM (2012)

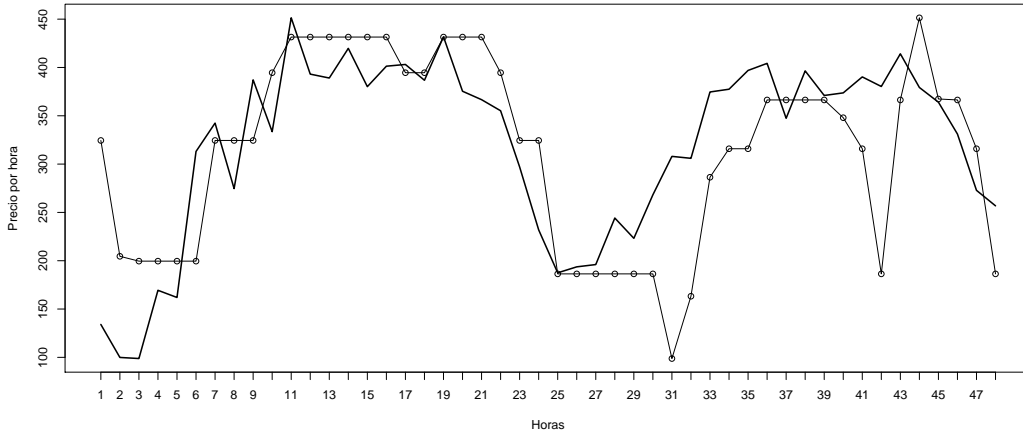
Figura C.8.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 4



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: (XM, 2013)

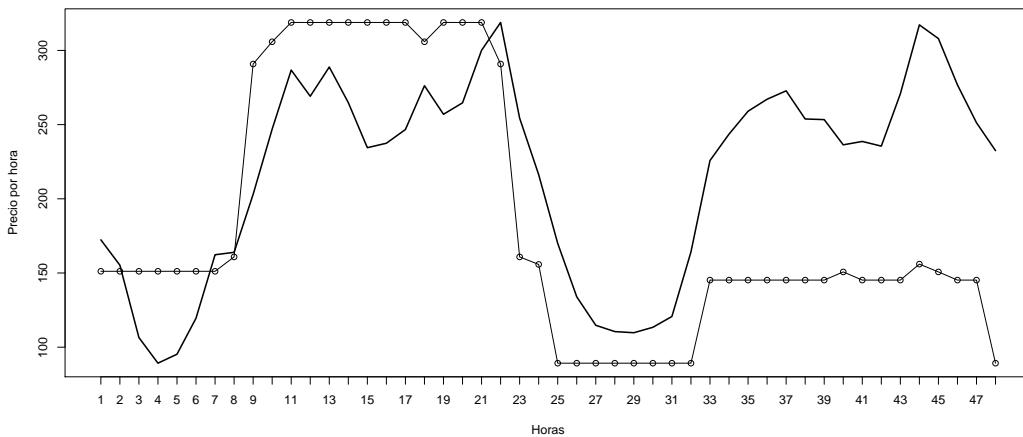
Figura C.9.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 10



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: (XM, 2013)

Figura C.10.: Gráfico comparativo entre valores pronosticados y valores reales en la semana 2



línea continua: datos de predicción
línea con puntos: datos reales

Fuente datos reales: (XM, 2014b)

Como se explica en la sección de análisis de errores en el capítulo de análisis de resultados,

aunque el modelo matemático está bien planteado, el comportamiento de estas ventanas no es el adecuado para hacer una predicción, ya que su tendencia se ve alterada por lluvias fuertes dentro de los últimos días de la semana lo cual hace que el precio baje pero como la tendencia era a mantener precios altos, el modelo arroja valores erróneos que no corresponden al a realidad del comportamiento del precio de la energía, es por eso que a la hora de hacer un pronóstico se debe tener muy en cuenta el comportamiento del precio dentro de esa semana en análisis, para que los factores ajenos al modelo no alteren el resultado pronosticado.