



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

---

---

FACULTAD DE CIENCIAS

Modelación de Fideicomisos de Inversión y  
Bienes Raíces con modelos ARIMA (SARIMA)

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

Actuario

PRESENTA:

Guillermo González Orozco

TUTORA

Mat. Margarita Elvira Chávez Cano



Cd. Mx. Octubre, 2022



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



*Dedicado a  
mi familia*



# Resumen

En este trabajo se presenta una aplicación de los modelos SARIMA al sector financiero mexicano, particularmente, a ciertos activos relativamente nuevos pertenecientes al mercado bursátil conocidos como *FIBRAS*, Fideicomisos de Infraestructura y Bienes Raíces. Para lo anterior, se presentará una breve descripción del mercado bursátil mexicano, así como una explicación de qué son las *FIBRAS*, su historia e importancia. Después, se plantearán los conceptos necesarios para entender los modelos SARIMA y así, finalmente, aplicar la metodología de estos modelos a dichos instrumentos financieros.

El principal objetivo es mostrar un ejemplo claro y conciso de la metodología propuesta por *Box & Jenkins* y el modelo *SARIMA*, aplicados a las *FIBRAS*, las cuales hoy en día son un activo financiero de gran relevancia en México. Y con esto, hacer evidente que aún un modelo “clásico”, no siendo este el más sofisticado, puede dar muy buenos resultados siendo cuidadosos en la ejecución de la metodología y ajuste del modelo.



# Índice general

|  |            |
|--|------------|
| <b>Resumen</b>   | <b>III</b> |
| <b>Lista de figuras</b>  | <b>XI</b>  |
| <b>Lista de tablas</b>   | <b>XV</b>  |
| <b>1. Las FIBRAS y el Sector Financiero</b>                                | <b>1</b>   |
| 1.1. El mercado bursátil . . . . .   | 1          |
| 1.2. El mercado bursátil mexicano . . . . .                                | 2          |
| 1.3. Fideicomisos de Infraestructura y Bienes Raíces<br>(FIBRAS) . . . . . | 9          |
| 1.3.1. Los instrumentos financieros . . . . .                              | 11         |
| 1.3.2. Las FIBRAS en la actualidad . . . . .                               | 12         |
| 1.3.3. Metodología del índice S&P/BMV FIBRAS . . . . .                     | 18         |
| <b>2. Metodología Box &amp; Jenkins y los modelos SARIMA</b>               | <b>19</b>  |
| 2.1. Introducción . . . . .  | 19         |

|          |   |    |
|----------|---|----|
| 2.2.     | Procesos Estocásticos . . . . .                       | 21 |
| 2.2.1.   | Proceso Estocástico . . . . .                         | 21 |
| 2.2.2.   | Serie de Tiempo . . . . .                             | 21 |
| 2.2.3.   | Estacionariedad Estricta . . . . .                    | 21 |
| 2.2.4.   | Estacionariedad Débil . . . . .                       | 22 |
| 2.3.     | Medidas de dependencia . . . . .                      | 22 |
| 2.3.1.   | Función de medias . . . . .                           | 22 |
| 2.3.2.   | Función de autocovarianza . . . . .                   | 23 |
| 2.3.3.   | Función de autocorrelación . . . . .                  | 24 |
| 2.3.4.   | Función de autocorrelación parcial . . . . .          | 25 |
| 2.3.5.   | Medidas de dependencia muestrales . . . . .           | 27 |
| 2.3.5.1. | Función de medias muestral . . . . .                  | 28 |
| 2.3.5.2. | Función de Autocovarianza Muestral . . . . .          | 29 |
| 2.3.5.3. | Función de Autocorrelación Muestral . . . . .         | 30 |
| 2.3.5.4. | Función de Autocorrelación Parcial Muestral . . . . . | 30 |
| 2.4.     | Series de Tiempo Lineales . . . . .                   | 30 |
| 2.4.1.   | Proceso de Ruido Blanco . . . . .                     | 31 |
| 2.4.2.   | Modelos de Series de Tiempo Lineales . . . . .        | 32 |
| 2.4.3.   | Proceso de Medias Móviles, MA . . . . .               | 33 |
| 2.4.4.   | Proceso Autoregresivo, AR . . . . .                   | 35 |

|          |   |    |
|----------|---|----|
| 2.4.5.   | Proceso Autoregresivo de Medias Móviles, ARMA . . . . .           | 38 |
| 2.4.5.1. | Causalidad . . . . .  | 40 |
| 2.4.5.2. | Invertibilidad . . . . .  | 41 |
| 2.4.6.   | Proceso ARMA multiplicativo estacional, SARMA . . . . .           | 42 |
| 2.4.7.   | Distribuciones asintóticas de SACF y SPACF de un proceso lineal   | 43 |
| 2.5.     | Modelos de series de tiempo no estacionarias . . . . .            | 45 |
| 2.5.1.   | Modelos con tendencia determinista . . . . .                      | 45 |
| 2.5.2.   | Modelos con tendencia estocástica . . . . .                       | 46 |
| 2.5.3.   | Raíces unitarias . . . . .  | 47 |
| 2.5.4.   | Modelos con estacionalidad . . . . .                              | 50 |
| 2.6.     | Pronósticos . . . . .   | 52 |
| 2.6.1.   | Algoritmo de Inovaciones . . . . .                                | 56 |
| 2.6.1.1. | Cálculo recursivo de pronósticos a h-pasos . . . . .              | 60 |
| 2.6.2.   | Pronósticos de modelos ARMA . . . . .                             | 61 |
| 2.6.2.1. | Algoritmo de inovaciones para procesos ARMA . . . . .             | 64 |
| 2.6.3.   | Pronósticos de modelos ARIMA . . . . .                            | 67 |
| 2.6.4.   | Pronósticos de modelos SARIMA . . . . .                           | 70 |
| 2.6.5.   | Modelo de espacio de estados para pronósticos de series de tiempo | 71 |
| 2.6.6.   | Calidad de los pronósticos . . . . .                              | 76 |
| 2.7.     | Estimación de parámetros . . . . .                                | 78 |

|           |   |           |
|-----------|---|-----------|
| 2.8.      | Validación de supuestos . . . . .   | 80        |
| 2.8.1.    | Pruebas de Ljung-Box y Box-Pierce . . . . .   | 82        |
| 2.8.2.    | Prueba de McLeod-Li . . . . .   | 83        |
| 2.8.3.    | Prueba para media igual a cero . . . . .  | 85        |
| 2.8.4.    | Pruebas de normalidad . . . . .   | 86        |
| 2.8.4.1.  | Prueba de Anderson-Darling . . . . .  | 86        |
| 2.8.4.2.  | Prueba de Shapiro-Wilk . . . . .  | 87        |
| 2.8.4.3.  | Prueba de Jarque-Bera . . . . .   | 88        |
| <b>3.</b> | <b>Aplicación</b>   | <b>91</b> |
| 3.1.      | Introducción . . . . .  | 91        |
| 3.2.      | Análisis Descriptivo . . . . .  | 92        |
| 3.3.      | Índice S&P/BMV FIBRAS 2017 . . . . .  | 96        |
| 3.3.1.    | Introducción . . . . .  | 96        |
| 3.3.2.    | Análisis del índice S&P/BMV FIBRAS 2017 . . . . .   | 97        |
| 3.3.3.    | Especificación de los parámetros del modelo SARIMA para el<br>índice S&P/BMV FIBRAS . . . . . | 105       |
| 3.3.4.    | Validación de supuestos estadísticos . . . . .  | 107       |
| 3.3.4.1.  | Residuales no correlacionados . . . . .   | 107       |
| 3.3.4.2.  | Media cero de los residuales . . . . .  | 111       |
| 3.3.4.3.  | Varianza constante de los residuales . . . . .  | 112       |
| 3.3.4.4.  | Normalidad de los residuales . . . . .  | 115       |

|   |            |
|---|------------|
| <i>ÍNDICE GENERAL</i>                                   | IX         |
| 3.3.5. Pronósticos de los modelos . . . . .             | 121        |
| 3.3.6. Conclusiones . . . . .                           | 122        |
| <b>A. Anexo</b>   | <b>125</b> |
| A.1. Teorema de proyección . . . . .                    | 125        |
| A.2. Periodograma . . . . .                             | 126        |
| A.3. Criterios de información . . . . .                 | 130        |
| A.3.1. Criterio de información de Akaike, AIC . . . . . | 130        |
| A.3.2. Criterio de información Bayesiano, BIC . . . . . | 131        |
| <b>Bibliografía</b>                                     | <b>133</b> |



# Índice de figuras

|   |     |
|---|-----|
| 3.1. Serie Histórica S&P/BMV FIBRAS . . . . .                                 | 92  |
| 3.2. Series Anuales S&P/BMV FIBRAS . . . . .                                  | 93  |
| 3.3. Box-Plots Anuales S&P/BMV FIBRAS . . . . .                               | 94  |
| 3.4. Histogramas Anuales S&P/BMV FIBRAS . . . . .                             | 96  |
| 3.5. Serie S&P/BMV FIBRAS 2017 . . . . .                                      | 97  |
| 3.6. Descomposición Clásica de la Serie S&P/BMV FIBRAS 2017 . . . . .         | 98  |
| 3.7. Regresión de la Serie S&P/BMV FIBRAS 2017 . . . . .                      | 99  |
| 3.8. Índice S&P/BMV FIBRAS sin tendencia . . . . .                            | 101 |
| 3.9. log del Índice S&P/BMV FIBRAS . . . . .                                  | 101 |
| 3.10. Serie diferenciada del log(S&P/BMV FIBRAS) . . . . .                    | 102 |
| 3.11. Periodograma de la serie diferenciada del log(S&P/BMV FIBRAS) . . . . . | 103 |
| 3.12. log(S&P/BMV FIBRAS) sin tendencia y sin estacionalidad . . . . .        | 104 |
| 3.13. ACF de la serie diferenciada del log(S&P/BMV FIBRAS) . . . . .          | 105 |
| 3.14. PACF de la serie diferenciada del log(S&P/BMV FIBRAS) . . . . .         | 106 |
| 3.15. ACF de los residuales del modelo 2 . . . . .                            | 108 |

|   |     |
|---|-----|
| 3.16. ACF de los residuales del modelo 4 . . . . .                      | 108 |
| 3.17. ACF de los residuales del modelo 5 . . . . .                      | 109 |
| 3.18. ACF de los residuales del modelo 6 . . . . .                      | 109 |
| 3.19. Prueba de Ljung-Box para los residuales de los modelos . . . . .  | 110 |
| 3.20. Prueba de Box-Pierce para los residuales de los modelos . . . . . | 111 |
| 3.21. ACF del cuadrado de los residuales del modelo 2 . . . . .         | 112 |
| 3.22. ACF del cuadrado de los residuales del modelo 4 . . . . .         | 113 |
| 3.23. ACF del cuadrado de los residuales del modelo 5 . . . . .         | 113 |
| 3.24. ACF del cuadrado de los residuales del modelo 6 . . . . .         | 114 |
| 3.25. Prueba de McLeod-Li para los residuales. . . . .                  | 115 |
| 3.26. Histograma de los residuales del modelo 2 . . . . .               | 116 |
| 3.27. Histograma de los residuales del modelo 4 . . . . .               | 116 |
| 3.28. Histograma de los residuales del modelo 5 . . . . .               | 117 |
| 3.29. Histograma de los residuales del modelo 6 . . . . .               | 117 |
| 3.30. QQ-Plot de los residuales del modelo 2 . . . . .                  | 118 |
| 3.31. QQ-Plot de los residuales del modelo 4 . . . . .                  | 118 |
| 3.32. QQ-Plot de los residuales del modelo 5 . . . . .                  | 119 |
| 3.33. QQ-Plot de los residuales del modelo 6 . . . . .                  | 119 |
| 3.34. Normalidad de los residuales . . . . .                            | 121 |
| 3.35. Serie real vs serie ajustada del índice S&P/BMV FIBRAS . . . . .  | 123 |

3.36. Pronóstico índice S&P/BMV FIBRAS . . . . . 124



# Índice de cuadros

|   |     |
|---|-----|
| 3.1. Medianas de las tasas de crecimiento semanales para los años 2012-2017   | 94  |
| 3.2. C.V y Rango Intercuantil para los años 2012-2017. . . . .  | 95  |
| 3.3. Estimación Regresión Simple para el índice S&P/BMV FIBRAS . . . . .  | 99  |
| 3.4. Prueba de <i>Dicky-Fuller Aumentada</i> para indice S&P/BMV FIBRAS . . . . .   | 100 |
| 3.5. Estadísticas del periodograma . . . . .  | 103 |
| 3.6. Prueba de <i>Dicky-Fuller Aumentada</i> para la serie diferenciada del índice<br><i>S&amp;P/BMV FIBRAS</i> . . . . . | 104 |
| 3.7. Prueba $\mu = 0$ para los residuales de los modelos propuestos . . . . .   | 111 |
| 3.8. Curtosis empírica de los residuales de los modelos propuestos . . . . .  | 120 |
| 3.9. Pruebas de normalidad para los residuales de los modelos propuestos . . . . .  | 120 |
| 3.10. Errores de los pronósticos . . . . .  | 122 |
| 3.11. Bondad de Ajuste . . . . .  | 122 |



# Capítulo 1

## Las FIBRAS y el Sector Financiero

### 1.1. El mercado bursátil

Las *FIBRAS*, al ser un instrumento financiero, forman parte de un gran mercado conocido como *mercado bursátil*. Aunque su nombre suene un tanto impresionante, este, no es más que un conjunto conformado por instituciones, empresas e individuos que realizan transacciones de productos financieros en diferentes *Bolsas* alrededor del mundo. En el *mercado bursátil*, se realizan intercambios de productos o activos de naturaleza similar, como ocurre en la *Bolsa de Valores*, donde se realizan intercambios con acciones o bonos, aunque también existen *Bolsas* especializadas en otro tipo de productos o activos. La demanda y oferta de dichos productos y activos, es la fuerza que determina los precios, según los que se compran y venden.

De esta manera, este mercado permite a las empresas financiar sus proyectos y actividades a través de la venta de diferentes productos, activos o títulos, y por otro lado, a los inversionistas les permite tener una rentabilidad por sus ahorros. La evolución del mercado bursátil, se mide a través de diferentes índices que reflejan los movimientos en los precios de los diferentes productos, activos o títulos.

Como se dijo previamente, el mercado bursátil está conformado por distintos elementos, entre los cuales, los más importantes se listan a continuación:

- Bolsa de Valores<sup>1</sup> Es un mercado de capitales organizado, institucionalizado y re-

---

<sup>1</sup>La palabra *Bolsa*, tiene su origen en un edificio que perteneció a una familia noble, de apellido *Van*

gulado, donde se realizan operaciones de compra-venta de acciones u obligaciones.

- **Corredores de Bolsa:** También conocidos como *Brokers*, pueden ser personas físicas o jurídicas autorizadas para realizar actividades de compra-venta de valores en la *Bolsa de Valores* en favor de terceros.
- **Emisores:** Empresas que, cumpliendo con las diferentes normativas correspondientes, ofrecen al *Mercado de Valores*, los títulos representativos de su capital social (acciones) o valores en forma de crédito (obligaciones).
- **Inversionistas:** Es una persona física o jurídica que aporta sus recursos financieros propios, con el objetivo de obtener un beneficio futuro.
- **Instituciones reguladoras:** Son las diferentes instituciones que vigilan el correcto funcionamiento del mercado bursátil.

## 1.2. El mercado bursátil mexicano

En el caso mexicano, por más de 123 años, la *Bolsa Mexicana de Valores (BMV)*, ha sido el mercado donde se realiza la compra-venta de acciones u obligaciones.

La *BMV*, es una empresa pública con ingresos semestrales por aproximadamente 1,507 millones de pesos y tiene un crecimiento anual del 11.4%, así como utilidades netas de 583.1 millones de pesos<sup>2</sup>.

Hoy en día, el *Grupo BMV*, se conforma por siete empresas, las cuales se describen a continuación:

- **Mercado de Derivados (MexDer:)** Es la *Bolsa de Derivados de México*, la cual inició operaciones el 15 de diciembre de 1998, siendo constituida como una sociedad anónima de capital variable, autorizada por la *Secretaría de Hacienda y Crédito Público (SHCP)*. Este hecho, constituyó uno de los avances más significativos en el proceso de desarrollo e internacionalización del *Sistema Financiero Mexicano*.

---

*Der Buërse*, en la ciudad europea de Brujas, lugar donde se realizaban encuentros y reuniones de carácter mercantil. En aquellos tiempos, por el volumen de las negociaciones, la importancia de esta familia y las transacciones que en ese local se efectuaban, la gente le dio el nombre al sitio y a la función de *Buërse*, y por extensión en todo el mundo se siguió denominando *bolsa*

<sup>2</sup>Datos al cierre de junio de 2017

- **Contraparte Central de Valores (CCV):** Es una entidad que complementa los servicios de compensación y liquidación de operaciones de compraventa, en un mercado de activos financieros, como lo son las acciones, derivados, bonos, etc. Su función principal es convertirse en deudora y acreedora recíproca de las operaciones originalmente pactadas, es decir, se vuelve responsable de que el comprador obtenga los títulos que adquirió y que el vendedor obtenga el efectivo correspondiente.

La *CCV*, contribuye a crear un entorno financiero más seguro, en el cual, el riesgo sistémico y de contraparte sean mitigados; al responder de esta manera a las tendencias globales actuales de la industria en materia de administración de riesgos, compensación y liquidación de operaciones, hace posible la apertura del mercado mexicano a mercados y activos financieros más diversos.

- **Asigna, Compensación y Liquidación:** Es un fideicomiso de administración y pago. En los mercados de derivados listados o estandarizados, la función de la *Cámara de Compensación* es ser la contraparte, y por ende, el garante de todas las obligaciones financieras que se derivan de la operación de los contratos de derivados.

La misión de *Asigna*, es ofrecer servicios de compensación, liquidación y administración de riesgos para operaciones financieras derivadas, brindando a los participantes un mayor grado de seguridad, transparencia y calidad, con el objetivo de asegurar el desarrollo ordenado del mercado de derivados, y con ello, contribuir al fortalecimiento del sistema financiero mexicano.

- **Indeval:** Es una institución que cuenta con la autorización para operar como *Depósito Central de Valores*. En consecuencia, se encuentra catalogada como proveedora para el sistema financiero, de guarda, custodia, administración, compensación y liquidación de valores. Es decir, todos aquellos valores que se compran o venden en el mercado financiero, casas de bolsa, bancos, operadoras o distribuidoras de fondos, deben ser guardados y administrados únicamente por el *Indeval*.
- **SIF ICAP:** Es la empresa de corretaje financiero líder en México, la cual, ofrece a los usuarios la posibilidad de tener acceso a la operación de instrumentos globales, ofreciendo a los intermediarios mexicanos y extranjeros, el acceso a la liquidez en toda una diversidad de mercados.

*SIF ICAP*, nace de la fusión de dos instituciones, *Servicios de Integración Financiera (SIF)* constituida en agosto de 1998, la cual, tenía la intención de favorecer la igualdad en la negociación de títulos de deuda dentro del mercado. En el año 2000, *SIF* se fusionó con *Garban Intercapital*, hoy *ICAP*, empresa inglesa considerada la correduría más grande del mundo en cuanto a capital y posicionamiento.

- **Valuación Operativa y Referencia de Mercado (Valmer):** Es una empresa fundada en el 2000, dedicada a proporcionar diariamente precios actualizados para la valuación de instrumentos financieros, así como servicios integrales de cálculo, información, análisis y riesgos. Teniendo como objetivo generar una mayor transparencia en la valuación de activos financieros pertenecientes a las distintas instituciones financieras en México.
- **Bursatec:** Es una subsidiaria de la *BMV* y de *Indeval*, cuya meta es proporcionar tecnología de vanguardia al mercado nacional, y de esta manera, asegurar la continuidad operativa a través del uso de tecnología de punta, impulsando así, el crecimiento de los mercados financieros. Todo esto alrededor de un marco de competitividad internacional.

Debido a que la *BMV* cuenta con un gran volumen operativo, esta cuenta con un esquema de clasificación sectorial, cuya finalidad es poder tener un control operativo de calidad y con ello, poder ofrecer un servicio de primer mundo.

Dicha clasificación sectorial, fue desarrollada en forma conjunta con el *Comité Técnico de Metodología* de la *BMV* y el *Comité de Análisis* de la *Asociación Mexicana de Intermediarios Bursátiles (AMIB)*. Esta estructura, considera esquemas que marcan la pauta a nivel internacional, los cuales, son utilizados por otras Bolsas de Valores al rededor del mundo.

Esta estructura está compuesta por cuatro nivel de clasificación, incorporando, un total de 10 sectores, 24 subsectores, 78 ramos y 192 subramos. De esta manera, por ejemplo, el nivel de agrupamiento más alto está dado por el siguiente esquema:

- **SECTOR I:** Energía
- **SECTOR II:** Materiales
- **SECTOR III:** Industrial
- **SECTOR IV:** Servicios y bienes de consumo no básico
- **SECTOR V:** Productos de consumo frecuente
- **SECTOR VI:** Salud
- **SECTOR VII:** Servicios Financieros
- **SECTOR VIII:** Tecnología de la Información

- **SECTOR IX:** Servicios de telecomunicaciones
- **SECTOR X:** Servicios públicos

La *BMV*, por poco más de 42 años, había sido la única bolsa de valores en México, pues con la fusión de las *Bolsas de Guadalajara y Monterrey*, en 1975, se buscó unificar el mercado a escala nacional, y de esta manera, consolidar el sector bursátil. Sin embargo, a inicios del 2018, la *Secretaría de Hacienda y Crédito Público (SHCP)* entregó la concesión a una nueva bolsa de valores, la cual, lleva por nombre *Bolsa Institucional de Valores (BIVA)*, la cual, se piensa propiciará un mercado de competencia, y con esto, ayudará a la economía nacional.

La historia de *BIVA* comienza con la *Central de Corretajes (CENCOR)*, la cual, es una sociedad mexicana con más de 25 años de experiencia que administra un grupo de empresas dedicadas al desarrollo de infraestructura de mercados financieros en México, EUA y Latinoamérica.

Así como *Grupo BMV*, *CENCOR* también está compuesta por diversas empresas, las cuales se mencionan a continuación.

- **ENLACE INT:** Fundada en 1993, fue la primera empresa de corretaje interbancario en México, cuyo objetivo es difundir los precios de instrumentos del mercado de cambios, deuda y derivados entre sus diversos clientes, como lo son *Bancos y Casas de Bolsa*, nacionales y extranjeros, para facilitar sus operaciones y, de esta manera, generar mayor liquidez.
- **Proveedor Integral de Precios (PIP LATAM):** Fundada en el 2000, fue la primera sociedad autorizada por la *CNBV* para presentar el servicio de “fair value”, valuación a precios razonables. Por lo tanto, *PIP LATAM*, se encarga de la determinación y proveeduría de precios para la valuación diaria de valores, documentos e instrumentos financieros. *PIP LATAM* tiene presencia en México, Costa Rica, Panamá, Perú y Colombia. Todo lo anterior, hace de *PIP LATAM*, la empresa líder en precios de valuación en Latinoamérica, produciendo más de 100,000 precios diarios de valuación de instrumentos financieros.
- **Mercado Electrónico Institucional (MEI):** Fundada en 2005 para ofrecer el servicio de corretaje de bonos entre inversionistas institucionales e intermediarios financieros. *MEI* cuenta con plataformas para el préstamo de valores, subastas y operaciones entre fondos de inversión.

- **Capital Star Systems:** Provee soluciones de tecnología avanzada para los agentes del mercado interbancario e institucional. Cubre un extenso rango de productos en los mercados de dinero, derivados y cambios.
- **Bolsa Institucional de Valores (BIVA):** Inició operaciones el 25 de julio de 2018 al amparo de la Reforma Financiera. *BIVA* fue implementada con el objetivo de promover la competencia del sector y contribuir al crecimiento, modernización y transformación del mercado de valores.

En octubre de 2015, *CENCOR* solicitó formalmente una concesión para organizar y operar la *Bolsa Institucional de Valores*, la cual, fue otorgada en agosto de 2017.

A partir de que *BIVA* entró en operaciones, se generó una nueva estructura en el mercado de valores, ya que ahora, las empresas tendrán la opción de listar sus valores de deuda o capital en alguna de las dos bolsas y sus valores cotizarán en ambas. Mientras que la liquidación de valores se hará a través de la actual *CCV*. Las casas de bolsa dirigirán sus órdenes a cualquiera de las bolsas siguiendo los principios de *mejor ejecución*, en los que deberán considerar precio, volumen y probabilidad de ejecución.

Además, *BIVA* repercute de manera directa en la mejora de la oferta de servicios, promueve activamente al mercado de valores, fomenta la innovación continua en tecnología y modelos de operación, genera nuevos índices bursátiles modernos e incluyentes, pero sobre todo, representa una opción accesible y competitiva para que nuevas empresas puedan financiarse en el mercado público a través de deuda y/o capital.

Actualmente, en la *BMV*, cotizan al rededor de 146 empresas. Por su parte, *BIVA* buscará enlistar a más de 50 nuevas emisoras. De ser así, el número total de empresas en bolsa en México crecería un 33 %, un incremento significativo en pocos años.

A pesar de las ventajas mencionadas, algunos expertos consideran que el momento de la creación de una nueva bolsa no fue adecuado, pues México aún es un país en vías de desarrollo y pocas empresas acuden a la bolsa para financiarse. Para poner lo anterior en contexto, la *BMV* apenas representa el 42 % del PIB mexicano, según datos de la *Federación Mundial de Valores*, un porcentaje parecido al de economías como la brasileña. Sin embargo, en contraste con países desarrollados, este porcentaje del PIB está muy distanciado al porcentaje correspondiente a países como Japón o Estados Unidos, donde respectivamente se supera el 100 %.

En el caso de México, hay empresas que no quieren estar en la bolsa, debido a que listarse significa tener que hacer públicas sus cuentas, con la finalidad de que los inversionistas puedan tomar decisiones adecuadas con su dinero. Las compañías que están en los

mercados de valores deben de reportar cada trimestre sus resultados, los cuales, además deben ser auditados por un tercero para certificar su fiabilidad.

Sumado a lo anterior, se deben cumplir ciertos requisitos, como por ejemplo, tener una estructura de gobierno corporativo. Por lo tanto, debido a que muchas de las empresas del país son grupos familiares, las cuales desean retener el control de su empresa, la mayoría de estas empresas optan por no cotizar. Además, tienen otros caminos para obtener dinero sin necesidad de vender parte de su empresa, por ejemplo, a través de proveedores y/o bancos.

Aunque por poco más de 42 años, la *BMV* no había tenido competencia, la *BMV* debería estar preparada para competir con *BIVA*, pues es parte de un negocio global, y hoy en día, la *BMV* ya tiene estrategias para conseguir que las empresas formen parte de su mercado. Asimismo, al ser la bolsa otro mecanismo de financiamiento, la *BMV* ha competido durante años con otras instituciones que proveen este servicio, como lo son los bancos por ejemplo.

Para poder comparar diferentes bolsas de valores de una manera objetiva, se utilizan los *índices bursátiles*, los cuales representan de forma resumida los comportamientos de los diferentes activos financieros que se cotizan en dicha bolsa.

Cada bolsa de valores tiene su propio índice, sin embargo, una bolsa no necesariamente tiene un sólo índice. Por ejemplo, los Estados Unidos tienen el *índice DOW* y el *S&P 500*. En el caso de México, el índice bursátil de la *BMV* es el *S&P/BMW IPC*, mientras que el índice de *BIVA* es el *FTSE BIVA*. La *BMV* tiene una alianza con *S&P*, mientras que *BIVA* tiene una alianza con *FTSE Russell*, la cual, es una empresa del *London Stock Exchange Group (LSEG)*.

Aunque hoy en día, el índice de la *BMV*, lleva por nombre *S&P/BMW IPC*, no siempre a sido así. Esta alianza responde a la visión de combinar los recursos, presencia y alcance global de un proveedor de índices independiente, con la experiencia local de la *BMV*.

*S&P/Dow Jones Indices*, es el mayor centro mundial de conceptos, datos e investigación de índices. Además, es la cuna de emblemáticos indicadores del mercado financiero, como lo son el *S&P 500* y el *Dow Jones Industrial Average*. Hoy en día, existen más activos invertidos en productos basados en índices de *S&P*, que en los productos de cualquier otro proveedor en el mundo.

Por si no fuera poco, *S&P DJI* tiene más de 120 años de experiencia en el desarrollo de soluciones innovadoras y transparentes que se ajustan a las distintas necesidades de

los inversionistas a nivel global. Su acervo de índices es de más de un millón, los cuales abarcan un amplio rango de clases de activos en todo el mundo. Es claro que *S&P DJI* define la manera en que los inversionistas miden los mercados y participan en ellos.

Es por esto y más, que *S&P DJI* y la *BMV* consolidaron su alianza a fin de responder a las necesidades cambiantes de los mercados globales, aprovechando la solidez y posicionamiento de ambas instituciones para alcanzar una mayor presencia internacional. Actualmente dicha alianza cuenta, por ejemplo, con índices de renta variable, estrategias y gestión de riesgo. Los índices *S&P/BMV* cumplen con los más altos estándares internacionales mediante un renovado órgano de gobierno.

Por su parte, *BIVA* tiene una alianza con *FTSE Russell*, pionero y líder en la creación, cálculo y distribución de índices. Como se mencionó previamente, forma parte del *London Stock Exchange Group (LSEG)*. Cuenta con más de 30 años de experiencia y 15 trillones de dólares asociados a sus índices. Calculan más 400,000 índices diariamente en más de 70 países, cubriendo el 98 % de los activos invertibles a nivel mundial.

De esta manera, *BIVA*, en conjunto con *FTSE Russell*, buscan crear índices modernos e incluyentes que logren representar el comportamiento del mercado de valores mexicano y sus diferentes segmentos. La alianza *FTSE BIVA* trabaja con metodologías mundialmente probadas en diferentes mercados e industrias para diseñar, calcular, mantener y difundir índices en tiempo real.

Por lo tanto, el índice *FTSE BIVA*, está diseñado para reflejar el desempeño de las empresas mexicanas líquidas, que son componentes del índice *FTSE Mexico All Cap Index*. Todas las empresas que coticen en el mercado de valores en México, son susceptibles a ser parte del índice *FTSE BIVA*. No hay un número fijo de integrantes, por lo cual, las emisoras que cumplan con los criterios de selección, principalmente de liquidez, pueden ser incluidas.

Hasta aquí, queda muy claro cuales son las empresas en México que funcionan como la *Bolsa de Valores*, la cual forma parte del *mercado bursátil*. Otros participantes sumamente importantes del *mercado bursátil*, son las *entidades reguladoras*. En el caso de México, estas son representadas principalmente por la *Comisión Nacional Bancaria y de Valores (CNBV)*, por el *Banco de México (BANXICO)* y por la *Secretaría de Hacienda y Crédito Público (SHCP)*, las cuales, supervisan de manera directa tanto a la *BMV* como a *BIVA*.

Sin embargo, otro de los agentes fundamentales del *mercado bursátil* en México, es la *Asociación Mexicana de Intermediarios Bursátiles (AMIB)*, la cual, es el organismo que certifica a las personas que deseen participar como intermediarios bursátiles. La *AMIB*, es

un organismo autoregulatorio reconocido por la CNBV.

A estos *intermediarios bursátiles*, o *brokers*, los define la *BMV* como intermediarios financieros que actúan entre un comprador y vendedor, usualmente cobrando una comisión como honorario. Una característica interesante de los *brokers*, es que por lo general, estos no buscan a los compradores y/o a los vendedores, son los mismos compradores y/o vendedores quienes buscan a los *brokers*. Esto seguramente se debe a que *los brokers* son los únicos que están autorizados para realizar la compra/venta de acciones.

La misma *BMV* menciona a *Salomon Financial Services*, *Forexpros.es*, *VATRE Consultores*, *Servicios Financieros México* y *Sierra Agencia Aduanal*, como los principales *brokers* en México.

### 1.3. Fideicomisos de Infraestructura y Bienes Raíces (FIBRAS)

Las *FIBRAS* son definidas por la *BMV*, como vehículos destinados al financiamiento para la adquisición o construcción de bienes inmuebles que se destinen al arrendamiento o a la adquisición del derecho a percibir ingresos provenientes del arrendamientos de dichos bienes, así como otorgar financiamiento para estos fines. De tal manera que para el inversionista, se ofrecen pagos periódicos (producto del arrendamiento) y a su vez, tienen la posibilidad de ofrecer ganancias de capital (plusvalía). Mientras que para el que aporta en los bienes inmuebles, permiten el financiamiento de activos inmobiliarios mediante la emisión de certificados en el *Mercado de Valores*, a través de oferta pública al distribuirlos entre el gran público inversionista.

Por ley, estos fideicomisos distribuyen el 95% de las utilidades de las rentas entre los inversionistas tenedores de *FIBRAS*, por lo cual, esto hace de las *FIBRAS* un sistema mixto de inversión, mediante el cual, se perciben beneficios de tres maneras:

- **A través de Dividendos:** Es decir, gracias a las utilidades obtenidas de las rentas menos los gastos de operación, a esto se le conoce como *Renta Fija*.
- **A través de Rendimientos:** Mediante el desempeño de los *Certificados Bursátiles Fiduciarios Inmobiliarios (CBFI)* en la *BMV* y en la *BIVA*, a esto se le conoce como *Renta Variable*. El rendimiento puede generar utilidades si la plusvalía sube, y puede generar pérdidas si el mercado inmobiliarios decae.

- **A través de la Plusvalía:** En la mayoría de los casos, éste se incrementa al integrar inmuebles a una *FIBRA*, debido a la elevada ocupación del inmueble y el cobro adecuado de rentas.

Estos instrumentos pertenecen al *mercado de capitales*, el cual, es de gran importancia debido a su tamaño y cantidad de operaciones que se realizan día a día. Este mercado también es conocido como de renta variable, ya que los activos financieros que pertenecen a este mercado, cambian sus precios día a día, por lo cual, no se conoce la ganancia que se pueda obtener. Al pertenecer al *mercado de capitales*, estos instrumentos están bajo una normatividad semejante a cualquier acción que cotiza en bolsa.

A grandes rasgos, el funcionamiento de las *FIBRAS* es el siguiente:

1. Se establece un *Fideicomiso*, el cual, administra los bienes en renta
2. El dueño (o dueños) de los inmuebles, los aportan al *fideicomiso*, a cambio de una contraprestación económica. Durante la vida del *fideicomiso* se pueden realizar nuevas adquisiciones para ampliar el portafolio inmobiliario.
3. El *fideicomiso* coloca certificados de participación de los activos que componen el portafolio de inmuebles de una *FIBRA* en la bolsa de valores, para que el público inversionista pueda participar en él, y del dinero que se recauda en este paso, se paga a los dueños.

Estos certificados son los *CBFI*, cuyo rendimiento está vinculado a los ingresos de operación de dichas propiedades, o en su caso, al producto de su venta. Estos títulos son una mezcla de inversión de renta fija y renta variable. Los *CBFIs*, pueden comprarse o venderse con la misma facilidad que cualquier acción.

4. El *fideicomiso* es administrado por expertos en el mercado de bienes raíces, los cuales se encargan del mantenimiento, mercadotecnia y administración de los bienes inmuebles.
5. El 95% de la utilidad anual, obtenida de la renta de los bienes, se distribuye entre los tenedores de los *CBFIs*.
6. Cuando el inversionista requiere de liquidez, utiliza su casa de bolsa para vender su participación en el fideicomiso.

La gestión de las *FIBRAS* se ve reflejada en los *fees* (*comisiones*) que cobran. Los *fees* pueden cobrarse sobre la base del valor de los activos, sobre los ingresos por rentas, sobre

las ganancias de capital en el valor de los certificados en el mercado, o bien, por comprar inmuebles en el mercado.

### 1.3.1. Los instrumentos financieros

Las *FIBRAS* están inspiradas en los instrumentos financieros estadounidenses conocidos como *REITs*, *Real Estate Investment Trust* por sus siglas en inglés, los cuales se crearon en Estados Unidos en 1960, con el objetivo de que las inversiones a gran escala de bienes inmuebles, fueran accesibles también para los pequeños inversionistas.

La “fórmula” que escogieron, consistió en equiparar la inversión en productos inmobiliarios a la inversión en cualquier otra industria; es decir, a través de la compra de valores. Debido a esto, aunque no están obligados, la mayor parte de los *REITs* son sociedades que cotizan en bolsa de valores.

Por otro lado, en México, gracias al *Tratado de Libre Comercio de América del Norte (TLCAN)*, el cual se firmó en 1992, comenzó a cambiar sus estándares de desarrollo, ya que este cambio, se exigía de manera implícita por diversos sectores empresariales, donde por supuesto, estaba incluido el sector inmobiliario.

A partir de este momento, diversas empresas inmobiliarias extranjeras comienzan a instalarse en territorio nacional, las cuales, comienzan a comercializar y administrar la construcción de parques industriales de primer mundo en el territorio. Con esto, nace lo que se podría llamar “el sector inmobiliario profesional”, para después comenzar a expandirse a segmentos comerciales y corporativos.

Sin embargo, hasta antes del 2008, el sector inmobiliario mexicano tenía la desventaja de que era totalmente dependiente de la inversión extranjera, la cual, se realizaba a través de los fondos de capital privado del exterior, cuya fragilidad y dependencia se hizo ver en la crisis surgida en el 2008.

Cuatro años antes, el secretario de la *SHCP*, Francisco Gil, introdujo modificaciones a la *Ley del Impuesto sobre la Renta* para buscar democratizar la inversión en el sector inmobiliario. Estas modificaciones obedecen al hecho de que se comenzaba a buscar reproducir la manera en la que Estados Unidos utiliza los *REITs*. Específicamente, las modificaciones recayeron en los artículos 223 y 224; en el primero es donde se da origen a las *FIBRAS*, mientras que el segundo trata de su regulación. Ambas modificaciones aparecen en 2004.

Dichas reformas, fueron los cimientos de los instrumentos financieros que hoy se conocen como *FIBRAS*, sin embargo, este nuevo instrumento estuvo olvidado en la legislación por más de cuatro años, hasta que un grupo de asesores liderados por Pedro Aspe<sup>3</sup> y el equipo de *Protego-Evercore*<sup>4</sup>, liderado por el actual director de *Evercore*, el *Dr. Augusto Arellano*. Este equipo se da a la tarea de impulsar los cambios necesarios para culminar con la publicación del reglamento complementario en 2010. Hasta este punto, ya se tenía el marco legal para la estructuración de dichos fideicomisos.

La presentación oficial de las *FIBRAS* en México fue de la mano de *FIBRA UNO* en marzo de 2011, con una *Oferta Pública Inicial (OPI)* de más de 3,000 millones de pesos utilizados para la adquisición de un primer portafolio de inmuebles patrimoniales diversificados principalmente en los segmentos industrial, corporativo y comercial. En ese momento, el portafolio contaba con 16 inmuebles, hoy en día cuenta con 557 propiedades.

Con el nacimiento de este instrumento financiero, se abrió el mercado a los inversionistas pequeños en México, junto con la participación de las *Afores*, en el sector de inmuebles de arrendamiento.

### 1.3.2. Las FIBRAS en la actualidad

Los *Fideicomisos de Infraestructura y Bienes Raíces*, han tenido un éxito impresionante. Hasta mediados del 2013, se estimaba que más del 34 % de las emisiones de la *Bolsa Mexicana de Valores*, fueron a través de *FIBRAS*, lo cual representó aproximadamente 40,000 millones de pesos. Mientras que el valor de capitalización de estos instrumentos superaba los 100,000 millones de pesos.

Sin lugar a duda, la introducción y desarrollo del mercado de *FIBRAS*, es el cambio estructural más importante en el sector inmobiliario mexicano en los últimos años, ya que cambió la naturaleza poco líquida de un inmueble, por un instrumento líquido que se puede comprar y vender todos los días a través de las bolsas de valores.

Actualmente, este modelo de negocio suma más de 1,500 activos inmobiliarios, lo cual, permite una diversificación muy amplia entre varios tipos de *fideicomisos*, cada uno enfocado a un sector específico. El índice representativo de las *FIBRAS*, desde el año 2011, ha tenido un rendimiento acumulado de 144 %, aunque no hay que olvidar que el

---

<sup>3</sup>Pedro Carlos Aspe Armella, es un economista y político mexicano, principal representante del proyecto económico puesto en marcha durante el gobierno de Carlos Salinas de Gortari, el cual se desempeñó como *Secretario de Hacienda y Crédito Público*

<sup>4</sup>En agosto de 2006, Protego se fusionó con Evercore Partners

### 1.3. FIDEICOMISOS DE INFRAESTRUCTURA Y BIENES RAÍCES (FIBRAS) 13

año 2017 fue un año de minusvalías en la mayoría de las emisoras. Cabe destacar que estos instrumentos financieros son altamente regulados debido a que las *AFORES* son de las principales inversionistas en estos instrumentos.

Al cierre del primer trimestre del 2018, los rendimientos por dividendo son mucho más elevados que los que otorgan las empresas.

Actualmente existen 12 *FIBRAS*:

- **FIBRA Uno:** Es la primera más grande *FIBRA* en México, cuenta con 557 propiedades que suman aproximadamente 8.27 millones de metros cuadrados de *área bruta rentable*. Aproximadamente el 62 % de los inmuebles pertenecen al sector comercial, 21 % al sector industrial y 17 % a oficinas. Además, se convirtió en 2016, en el primer fideicomiso en estar presente en las 32 entidades de la República Mexicana.
- **FIBRA Inn:** Fideicomiso mexicano dedicado a la hotelería. Cuenta con 43 hoteles, los cuales operan bajo las marcas de *Holiday Inn Express*, *Hampton Inn*, *Crown Plaza*, *Fairfield Inn*, entre otras.
- **FIBRA Macquarie (MQ):** Listada en 2012; su cartera consta de 217 propiedades industriales y 17 locales divididos entre comerciales y oficinas. Dichas propiedades están ubicadas en 20 ciudades de 15 estados de la República Mexicana. Aproximadamente el 13.6 % de su portafolio está presente en Monterrey, el 11.9 % en Cd. Juárez y 11.4 % en la Zona Metropolitana de la Ciudad de México, siendo estas, las regiones con mayor concentración de dicho portafolio.
- **FIBRA Terrafina:** Listada desde 2013, es una *FIBRA* especializada en el sector industrial, cuyo portafolio está compuesto de almacenes y otras propiedades de manufactura ligera. Sus propiedades están ubicadas principalmente en el centro, bajío y norte de México. Cuenta con 284 propiedades industriales, de las cuales, 199 están en el norte del país, 55 en el bajío y 30 en el centro. Todas sus propiedades constituyen un *ARB* de 3.77 millones de metros cuadrados, y cuenta con una ocupación del 95.5 %; el 73.1 % de dicha *ARB*, es para distribución, mientras que el 26.9 % restantes se utiliza para manufactura.
- **FIBRA Prologis:** Fideicomiso constituido con el objetivo de adquirir y administrar inmuebles destinados a actividades industriales en México. Su portafolio consta de 178 propiedades dedicadas a logística y manufactura. El total de sus activos están valuados en 1.7 billones de dólares. El 31.3 % de su portafolio está concentrado en la Ciudad de México

- **FIBRA Shop:** *FIBRA* mexicana listada desde 2013, creada principalmente para adquirir, desarrollar y operar un amplio rango de centros comerciales en México. De hecho, fue la primer *FIBRA* en México, especializada exclusivamente en centros comerciales. La cartera de *FIBRA Shop*, está integrada por 18 centros comerciales y 2 más en construcción, los cuales, están ubicados en 13 estados de la República Mexicana. Además, su portafolio representa 787 mil 500 metros cuadrados de *ARB*. Algunos de los centros comerciales pertenecientes a *FIBRA Shop* son, *Plaza Cibeles*, *Urban Center Condesa*, *Kukulcán Plaza*, *Galerías Mall Sonora*, entre otras.
- **FIBRA Danhos:** *FIBRA* listada desde 2011, al cierre de 2017 cuenta con 765,000 metros cuadrados de *Área Rentable Bruta*. Desde su *oferta pública inicial (OPI)*, esta *FIBRA* ha añadido 498,000 metros cuadrados a su portafolio de inmuebles, lo cual, representa una tasa compuesta anual de crecimiento de 30%. Dicho portafolio está compuesto por 13 propiedades y 2 proyectos en desarrollo. Estas propiedades incluyen centros comerciales, oficinas y proyectos de usos mixtos. Algunas de sus propiedades más representativas son *Parque Delta*, *Parque Toreo*, *Parque Central*, *Parque Lindavista*, *Parque Tezontle*, *Parque Esmeralda (Corporativo Elektra)*, *Reforma 222*, etc.
- **FIBRA Monterrey (MTY):** Fideicomiso constituido en 2014. Cuenta con 43 propiedades distribuidas en un *ARB* de 504 mil 534 metros cuadrados, de las cuales, 12 propiedades son oficinas, 25 de giro industrial y 6 de giro comercial. Tiene presencia en los estados de Nuevo León, Jalisco, Sinaloa, Chihuahua, Coahuila, Guanajuato, Querétaro, San Luis Potosí y el Estado de México.
- **FIBRA Hotel:** Fideicomiso mexicano creado principalmente para desarrollar, adquirir, poseer y operar hoteles en México. Su portafolio consta de 87 hoteles, lo cual representa un total de 12 mil 723 habitaciones en 26 estados de la República Mexicana. Entre los hoteles que destacan se encuentran el *Hotel Marriot*, *Camino Real*, *Fiesta Inn* y *Fiesta Americana*. Esta *FIBRA* está listada desde noviembre de 2012.
- **FIBRA Plus (PL):** Fue la primer *FIBRA* en México destinada al desarrollo de bienes inmuebles destinados al arrendamiento. Listada desde 2014, cuenta con 914 inmuebles destinados al alquiler para empresas de logística y manufactura en México. Estos inmuebles constituyen un *ARB* de 185 mil 679 metros cuadrados, los cuales, están ubicados en la Ciudad de México, Guadalajara, Monterrey, Reynosa, Tijuana y Ciudad Juárez. Los proyectos con enfoque en el desarrollo comercial, representaron el 41 % del *ARB*, oficinas el 23 %, industrial el 22 % y por último, con enfoque de vivienda un 14 %.

### 1.3. FIDEICOMISOS DE INFRAESTRUCTURA Y BIENES RAÍCES (FIBRAS) 15

- **FIBRA HD:** Esta *FIBRA* está enfocada en la adquisición y administración de inmuebles de alta calidad de tamaño medio, con inquilinos asociados con marcas nacionales e internacionales de renombre. Actualmente cuenta con 295,815 metros cuadrados de *ARB*; su portafolio cuenta con 37 propiedades en todo el país, tienen una ocupación del 96 % y 488 inquilinos. El 51 % de su portafolio pertenece al sector comercial, 30 % al sector industrial, 13 % al sector educativo y el 6 % restante al sector de oficinas. Algunos de sus inquilinos son *Aeromexico*, *Telmex*, *HSBC*, *UVM*, *Soriana*, *GMC*, *Famsa*, *Elektra*, etc.

Las *FIBRAS* que mejor pagan<sup>5</sup> a sus inversionistas son; *FIBRA Shop*, que otorga un rendimiento anualizado por *CBFI* de 8.8 %, le siguen *FIBRA Inn* y *FIBRA MTY*, con un rendimiento de 8.7 % cada una.

Una *FIBRA* que destaca, la cual, no fue enlistada intencionalmente, lleva por nombre *FIBRA E*. *FIBRA E* fue impulsada por el gobierno en 2015, derivado del débil crecimiento económico y el boquete presupuestas ocasionado por la caída del precio y producción de petróleo, intentando de esta manera, recaudar capital para proyectos de energía e infraestructura.

De la misma manera en que las *FIBRAS* simulan los instrumentos de inversión norteamericanos *REITs*. *FIBRA E* reproduce los vehículos de inversión, también norteamericanos, conocidos como *Master Limited Partnership*, conocidos en México como *Sociedades de Participación Limitada*, los cuales, han recaudado capital en Estados Unidos por más de 700,000 millones de dólares.

Debido a la gran similitud entre *FIBRA E* y los *MLP*, es necesario entender el funcionamiento de estos últimos para poder comprender el funcionamiento de *FIBRA E*. Los *MLP* son instrumentos financieros estructurados bajo las leyes de Estados Unidos, los cuales cotizan en la bolsa de valores, usados para recaudar capital requerido en el desarrollo de ciertos proyectos, principalmente aquellos relacionados con recursos naturales, como por ejemplo, el almacenamiento y transporte de petróleo y gas natural.

Lo que hace atractivos a estos instrumentos, es que combinan los beneficios fiscales de una *sociedad limitada*, y la liquidez de un instrumento que cotiza en la bolsa de valores. Para que una sociedad reciba la calificación de *MLP*, es necesario que esta, genere al menos el 90 % de sus ingresos por medio de recursos naturales.

Entonces, volviendo al caso mexicano, la reforma energética se desarrolló de manera que permitiera que este tipo de instrumentos financieros, fueran una opción para una

---

<sup>5</sup> cifras al último trimestre del 2017

inversión directa del gran público inversionista en activos energéticos, dando como resultado, el nacimiento de *FIBRA E*.

*FIBRA E* está compuesta por proyectos relacionados con actividades como:

- Tratamiento, refinación, enajenación, comercialización, transporte, almacenamiento de petróleo y transporte; almacenamiento, distribución y expendio de gas natural; y transporte por ducto y almacenamiento de petrolíferos.
- Generación, transmisión o distribución de energía eléctrica.
- Proyectos de inversión en infraestructura relacionados con caminos, carreteras, vías férreas y puentes, así como terminales marítimas e instalaciones portuarias, crecimiento en la red de telecomunicaciones, seguridad pública y readaptación social, agua potable, drenaje y alcantarillado.

Además, se goza de ciertos beneficios fiscales, los principales de estos beneficios son:

- No existe la obligación de hacer pagos provisionales por parte de las sociedades promovidas, ni por *FIBRA E*.
- La venta de los certificados emitidos por *FIBRA E* se encuentra exenta de impuestos.
- Las empresas promovidas no están obligadas a retener ISR del 10% en el pago de dividendos.
- Al menos el 70% del valor neto promovido anual de *FIBRA E* será invertido en acciones de empresas promovidas, y el 95% de sus resultados fiscales será distribuido anualmente a los tenedores de *CBFI*.
- *FIBRA E* es tratada como un fideicomiso empresarial para efectos fiscales.
- Las *AFORES* quedan exentas del pago de ISR.

Si bien *FIBRA E* fue impulsada por el gobierno, este sólo funge como intermediario entre el vehículo financiero, las empresas promovidas y los inversionistas, pero no tiene responsabilidad o injerencia en la operación y funcionamiento de la misma, sin embargo, sí puede ampliar la recaudación a través del ISR.

### 1.3. FIDEICOMISOS DE INFRAESTRUCTURA Y BIENES RAÍCES (FIBRAS) 17

Algunas restricciones que tiene *FIBRA E*, son por ejemplo, que el fiduciario debe ser un banco mexicano autorizado, o una casa de bolsa. Mientras que los accionistas de las empresas promovidas, deberán ser residentes mexicanos para efectos fiscales, por lo cual, los inversionistas extranjeros, están obligados a registrar ganancias en México, y después, tener que tomar los caminos de doble tributación para agilizar las cuestiones fiscales.

Hasta mayo de 2018, hay tres *FIBRAS E* listadas en la *Bolsa Mexicana de Valores*, dos de ellas están enfocadas al sector industrial, mientras que una está enfocada a energía. En total, estas recaudaron 52,223 millones de pesos.

Se esperaba que *Petróleos Mexicanos (PEMEX)* y la *Comisión Federal de Electricidad (CFE)*, fueran los primeros emisores de un fideicomiso de energía. Hasta ahora, la empresa eléctrica del Estado, ya hizo una colocación por 16,388 millones de pesos en febrero de 2017. Por otro lado, en el caso de *PEMEX*, algunos expertos aseguran que colocar en el mercado de valores una participación minoritaria podría llevar algunos años, debido a que antes, se necesita hacer una reforma constitucional.

Es muy clara la importancia que han tenido las *FIBRAS* en el sector financiero mexicano, sin embargo, esto aún no ha sido suficiente para que estos fideicomisos pertenezcan al *Índice de Precios y Cotizaciones*, de hecho, en el 2017 se realizó una consulta por *S&P Dow Jones Índices* entre inversionistas, en la cual, se determinó no incluir a las *FIBRAS* dentro de la canasta de índices accionarios de la *Bolsa Mexicana de Valores*, incluida, su principal muestra, el *S&P/BMV IPC*.

En los resultados de dicha consulta no se expresaron los motivos por los cuales se decidió no incluir a las *FIBRAS*, a pesar de que, de acuerdo con especialistas en temas bursátiles, es positivo para una emisora o empresa ser parte de un índice accionario, ya que esto permite aumentar la oportunidad de que más inversionistas, locales y/o extranjeros inviertan en ella y, al mismo tiempo, aumenta la negociación (compraventa) de sus acciones o títulos. En tanto, el índice representaría de una mejor manera al mercado y sería incluyente al integrar un nuevo instrumento financiero.

Sin embargo, gracias a *BIVA*, por primera vez en la historia del mercado mexicano, las *FIBRAS* podrán formar parte del índice insignia, ya que es claro que este instrumento representa un sector relevante y representativo de la economía. La muestra preliminar del índice *FTSE BIVA*, está integrada por 57 emisoras, cinco de ellas son fideicomisos inmobiliarios: *FIBRA Uno*, *FIBRA Danhos*, *FIBRA Macquarie*, *FIBRA Prologis* y *FIBRA Terrafina*. Debido a esto, se considera que finalmente el índice *FTSE BIVA*, en comparación con el índice *S&P/BMV IPC*, sea más referenciada por incluir a estos instrumentos.

### 1.3.3. Metodología del índice S&P/BMV FIBRAS

Se tiene un *Universo Elegible*, el cual, está compuesto por todos los valores del mercado de renta variable listados en la *BMV* y clasificados como *FIBRAS*. De este *Universo Elegible*, se debe seleccionar la muestra que representará al índice, lo cual, lleva al *Universo de Selección*. Este *Universo de Selección*, está compuesto por todas las series accionarias del universo elegible que cumplan con los criterios, enunciados a continuación, de la fecha de referencia de la reconstitución del índice conforme al *Universo de selección*:

- **Historial de Operación:** Las series accionarias elegibles no deberán sumar cinco o más días sin operación en los tres meses previos a la fecha de referencia de la reconstitución del índice.
- **Series Accionarias Múltiples:** Si una compañía, cuenta con más de una serie accionaria, aquella con el mayor importe operado en bolsa es elegida.

Para la **selección de componentes**, todas las acciones del *Universo de Selección*, son ordenadas según su importe operado en bolsa. El importe operado se define como la mediana de las medianas mensuales del último semestre calendario. La mediana del importe operado (*Median Daily Value Traded, MDVT*, por sus siglas en inglés) mensual, se define como la mediana del importe operado en bolsa por día para cada serie accionaria en cuestión. El importe operado es calculado multiplicando el número de acciones negociadas por su precio.

Las 20 acciones con los puntajes más altos en función de su importe operado, serán seleccionadas para constituir el índice. Si en el universo de selección, no hubiera al menos 20 *FIBRAS* para constituir el índice, todas las que cumplan con los criterios de elegibilidad serán incluidas en el.

Después, se realiza una **ponderación de los componentes** del índice. Esta ponderación se hace en función del importe operado en bolsa de cada uno de los componentes, sujeto a un límite de no más del 25 % en la ponderación de cada serie accionaria.

Por último, al índice se le aplica un **rebalanceo** de forma periódica, es decir, es reconstituido. Lo anterior se hace de manera bianual, con fecha efectiva después del cierre de mercado del tercer viernes de junio y diciembre. La fecha de referencia para cada reconstitución, es el último día hábil de abril y octubre. La cuenta de acciones del índice se calcula tomando los precios de cierre de los siete días hábiles previos a la fecha efectiva del rebalanceo.

## Capítulo 2

# Metodología Box & Jenkins y los modelos SARIMA

### 2.1. Introducción

El análisis estadístico de datos en forma de series de tiempo tiene como pionero a *George Udny Yule*, el cual, en los años veinte del siglo XX escribió tres importantes publicaciones sobre análisis de series de tiempo, los cuales fueron:

1. “*On the time correlation problem - 1921*”
2. “*An investigation of a form of spurious correlation - 1926*”
3. “*On a method of investigating periodicities in disturbed series, with special reference to Wolfer’s Sunspot Numbers*”

En el último artículo, *Yule* hizo uso de un modelo autoregresivo para modelar la serie de tiempo de manchas solares en lugar del método establecido del *Periodograma de Schuster*. No obstante, el origen de los modelos matemáticos para hacer predicciones se remontan mucho antes de *Yule*.

Como en muchas áreas de estadística, la aplicación de las series de tiempo es un concepto clave en el desarrollo de las metodologías de análisis. En los negocios y en la economía, el análisis de series de tiempo se utiliza entre otros motivos, para estudiar la

estructura dinámica de algún proceso; estudiar la relación dinámica entre variables; hacer ajustes de estacionalidad sobre datos económicos, como por ejemplo, sobre las tasas de desempleo; mejorar los análisis de regresión cuando los errores están correlacionados serialmente o, por último, dar predicciones puntuales y de intervalos sobre la serie de tiempo.

La importancia de la metodología de *Box & Jenkins* en 1970, mostrada en la publicación "*Time Series Analysis: Forecasting and Control*", radica en el hecho de que esta metodología, dio como resultado un procedimiento sistemático que permitió hacer uso de métodos de series de tiempo para lograr realizar predicciones. Dicha metodología, popularizó a los modelos *autoregresivos integrados de medias móviles*, *ARIMA* por sus siglas en inglés, al utilizar un procedimiento iterativo, el cual, consiste en **identificación, estimación y validación** del modelo.

El éxito de los modelos *ARIMA* generó una gran cantidad de líneas de investigación en el análisis de series de tiempo, sin embargo, la historia del análisis de series de tiempo no ha sido muy parsimoniosa. Originalmente el análisis de series de tiempo estaba dividido en el dominio "*frecuentista*" y en el dominio "*temporal*". Los defensores de ambos planteamientos no se veían con buenos ojos, de hecho, existían fuertes debates y críticas entre ellos.

Por un lado, el enfoque "*temporal*" hace uso de la función de autocorrelación de los datos y de modelos paramétricos, como es el caso del modelo *ARIMA*, para describir la dependencia dinámica de la serie. Mientras que el enfoque "*frecuentista*" se enfoca en el análisis espectral o las funciones potencia, para el análisis de series de tiempo.

Sin embargo, estas diferencias se han ido disipando con el tiempo, y hoy en día, el objetivo del análisis, así como la experiencia del analista, son los factores determinantes para decidir cual de los dos enfoques utilizar.

La metodología de *Box & Jenkins* y los modelos *ARIMA*, son apenas un pequeño ejemplo del vasto mundo de modelos y metodologías que comprende el análisis de series de tiempo. Otros casos importantes del análisis de series de tiempo son los modelos de *espacios de estados*, los modelos *ARCH* y *GARCH*, así como los modelos multivariados *VAR* y *VARMA*, siendo el último, objeto de investigación muy reciente, tanto desde el punto de vista teórico como técnico, para poder obtener una metodología similar a la dada por *Box & Jenkins*, así como para poder afrontar el problema de *identificabilidad* que se presenta en los modelos *VARMA*.

## 2.2. Procesos Estocásticos

### 2.2.1. Proceso Estocástico

En general, a una colección de variables aleatorias,  $X(w, t)$  indexadas a  $T$ , se le conoce como *proceso estocástico*. Donde  $w$  pertenece a un espacio muestral  $\Omega$  y  $t$ , pertenece al conjunto de índices  $T$ .

Generalmente el conjunto de índices  $T$ , es un subconjunto de  $\mathbb{R}$ , comúnmente este subconjunto es  $\mathbb{Z} \subset \mathbb{R}$ ,  $\mathbb{N} \subset \mathbb{R}$ ,  $[0, \infty) \subset \mathbb{R}$  ó  $(-\infty, \infty) \subseteq \mathbb{R}$ . Puede ocurrir que el conjunto de índices no sea un subconjunto de  $\mathbb{R}$ .

Por lo tanto, para cada  $t \in T$  fija,  $X(\cdot, t)$  es una variable aleatoria, mientras que para cada  $w \in \Omega$ ,  $X(w, \cdot)$  es una función de  $t$ , a la cual se le conoce como *función muestral*, *realización del proceso* o camino muestral.

### 2.2.2. Serie de Tiempo

El término de *Serie de Tiempo* se utiliza para referirse tanto al proceso estocástico que describe el fenómeno en estudio, o bien, a una realización de dicho proceso estocástico. Donde las funciones  $\{X(\cdot, w), w \in \Omega\}$  en  $T$ , son la realizaciones de la serie de tiempo  $\{X_t, t \in T\}$ . De aquí en adelante se comenzará a usar el término *Serie de Tiempo* para hacer referencia a un proceso estocástico, o a una realización del mismo, según sea el caso.

### 2.2.3. Estacionariedad Estricta

Se dice que una serie de tiempo es estrictamente estacionaria si las funciones de distribución conjunta de los vectores aleatorios  $(X_{t_1}, \dots, X_{t_k})$  y  $(X_{t_1+h}, \dots, X_{t_k+h})$  son iguales, es decir, si se cumple que:

$$F_{X_{t_1}, \dots, X_{t_k}}(X_{t_1}, \dots, X_{t_k}) = F_{X_{t_1+h}, \dots, X_{t_k+h}}(X_{t_1+h}, \dots, X_{t_k+h})$$

Para todo entero positivo  $k$ , para todos los enteros positivos  $t_1, \dots, t_k$ , así como pa-

ra cualquier entero  $h$ . Es decir, se requiere que la función de distribución conjunta del vector  $(X_{t_1}, \dots, X_{t_k})$  sea invariante ante traslaciones, por lo cual, las leyes de probabilidad que rigen el comportamiento del proceso estocástico no cambian por cambios en el “origen”, en este caso, se dice que la serie está en equilibrio estadístico. A las series de tiempo *estrictamente estacionarias*, también se les conoce como *fuertemente estacionarias* o *completamente estacionarias*.

### 2.2.4. Estacionariedad Débil

Se dice que una serie de tiempo es *débilmente estacionaria*, *estacionaria en covarianza*, *estacionaria de segundo orden* ó *estacionaria en el sentido amplio* si ocurre que:

$$i) \mu_t = \mathbb{E}(X_t) = c \quad \forall t \in T$$

$$ii) \gamma_x(s, t) = \mathbb{E}[(X_s - \mu_s)(X_t - \mu_t)] \text{ para toda } s \text{ y } t \text{ depende únicamente de } |s - t|$$

Por lo tanto, se pide que la media de la serie de tiempo  $\{X_t\}$ , así como su covarianza, sean invariantes ante el tiempo. Por lo tanto, débilmente estacionario implica que la serie fluctúe con una varianza constante alrededor de un nivel constante. Es importante notar que implícitamente se está pidiendo que los primeros dos momentos de  $\{X_t\}$  sean finitos.

## 2.3. Medidas de dependencia

De aquí en adelante se supondrá que  $T$ , el conjunto de índices, es igual a  $\mathbb{Z}$ , es decir, que  $T = \mathbb{Z}$ .

### 2.3.1. Función de medias

Se define como

$$\mu_t(x) = \mathbb{E}(X_t) = \int_{-\infty}^{\infty} x f_t(x) dx$$

Es importante notar que  $\mu_t$  es la media teórica de la serie de tiempo en el punto particular de tiempo  $t$ .

### 2.3.2. Función de autocovarianza

Si  $\{X_t\}$  es una serie de tiempo tal que  $Var(X_t) < \infty \forall t \in \mathbb{Z}$ , entonces, la función de autocovarianza  $\gamma_x(\cdot, \cdot)$  de  $\{X_t\}$  se define como

$$\gamma_x(s, r) = Cov(X_s, X_r) = \mathbb{E}[(X_s - \mu_s)(X_r - \mu_r)] \quad \text{con } s, r \in \mathbb{Z}$$

La autocovarianza mide la dependencia lineal entre dos puntos de la serie a diferentes tiempos. Si  $\gamma_x(s, r) = 0$ , entonces  $X_s$  y  $X_r$  no están linealmente relacionados, pero aún puede existir una estructura de dependencia entre ambas, a menos que la distribución conjunta de  $X_s$  y  $X_r$  siga una distribución normal bivariada, entonces  $\gamma_x(s, r) = 0$  asegura su independencia. Es fácil ver que  $\gamma_x(r, s) = \gamma_x(s, r)$ .

Si  $\{X_t, t \in \mathbb{Z}\}$  es estacionario, entonces  $\mathbb{E}(X_t) = \mu$  y la varianza  $Var(X_t) = \mathbb{E}(X_t - \mu)^2 = \sigma^2$ . Además la covarianza  $Cov(X_t, X_s)$ , depende únicamente de la diferencia dada por  $|t - s|$  y no de los valores  $t$  y  $s$  en concreto, pues

$$Cov(X_t, X_s) = Cov(X_{t-k}, X_{s-k}) \quad \forall s, t, k \in \mathbb{Z}$$

Haciendo  $k = s$  se obtiene

$$\gamma_x(t, s) = \gamma_x(t - k, s - k) = \gamma_x(t - s, s - s) = \gamma_x(t - s, 0)$$

Además, como la autocovarianza es simétrica, se cumple que  $\gamma_x(t - s, 0) = \gamma_x(0, t - s)$ . Por otro lado, si  $k = t$

$$\gamma_x(t, s) = \gamma_x(t - t, s - t) = \gamma_x(0, s - t) = \gamma_x(0, t - s)$$

$$\therefore \gamma_x(t, s) = \gamma_x(0, |t - s|)$$

Por lo tanto, es conveniente redefinir a la función de autocovarianza de un proceso estacionario como una función univariada. Sea  $h = |t - s|$ , entonces.

$$\gamma_x(h) := \gamma_x(h, 0) = \gamma_x(|t - s|, 0) = \text{Cov}(X_{t+|h|}, X_t) \quad \forall t, h \in \mathbb{Z}$$

Por lo tanto,  $\gamma_x(\cdot)$ , será la función de autocovarianza de  $\{X_t\}$  y  $\gamma_x(h)$  su imagen con rezago  $h$ .

Algunas propiedades básicas de esta función son:

- )  $\gamma_x(0) = \text{Var}(X_t)$
- )  $\gamma_x(h) = \gamma_x(-h)$
- )  $|\gamma_x(h)| \leq \gamma_x(0)$

### 2.3.3. Función de autocorrelación

Si  $\{X_t\}$  es una serie de tiempo tal que  $\text{Var}(X_t) < \infty \forall t \in \mathbb{Z}$ , entonces, la función de autocorrelación  $\rho(\cdot, \cdot)$  de  $\{X_t\}$  se define como

$$\rho(s, r) = \frac{\gamma_x(r, s)}{\sqrt{\gamma_x(r, r)}\sqrt{\gamma_x(s, s)}} \quad \text{con } s, r \in \mathbb{Z}$$

La autocorrelación mide la fuerza de la dependencia lineal entre  $X_r$  y  $X_s$ , es fácil ver que  $|\rho_x(r, s)| \leq 1$  y que  $\rho_x(r, s) = \rho_x(s, r)$ . Otra manera de interpretar a la autocorrelación, es viéndola como la predictibilidad lineal de la serie a tiempo  $t$  con la información a tiempo  $s$ , digamos de  $X_t$  usando únicamente a  $X_s$ . Si se puede predecir perfectamente a  $X_t$  usando  $X_s$  a través de una relación lineal  $X_t = \beta_0 + \beta_1 X_s$ , entonces la autocorrelación será 1 si  $\beta_1 > 0$  ó -1 si  $\beta_1 < 0$ .

Debido a que la autocorrelación está descrita por la autocovarianza, en el caso en que la serie de tiempo  $\{X_t\}$  es estacionaria, también es conveniente redefinir a la autocorrelación de este tipo de procesos. Por lo tanto, la autocorrelación de una serie de tiempo estacionaria es:

$$\rho_x(h) := \frac{\gamma_x(h)}{\gamma_x(0)}$$

Como función de  $h$ , a  $\rho_x(h)$  se le conoce como función de autocorrelación, *ACF* por sus siglas en inglés.

De las propiedades básicas de la función de autocovarianza enunciadas previamente, se derivan las siguientes propiedades de la *ACF*:

- )  $\rho_x(0) = 1$
- )  $\rho_x(h) = \rho_x(-h)$
- )  $|\rho_x(h)| \leq 1$

Debido a que la *ACF* es simétrica, usualmente se calculan únicamente los valores de  $h \in \mathbb{Z}^+$

### 2.3.4. Función de autocorrelación parcial

La función de autocorrelación parcial, *PACF* por sus siglas en inglés, así como la función de autocorrelación, provee información vital sobre la dependencia estructural de una serie de tiempo. En el caso de que la serie de tiempo  $\{X_t\}$  sea estacionaria, así como la *ACF*, la *PACF* depende únicamente de las propiedades de segundo orden de la serie de tiempo. Generalmente se denota a la correlación parcial con rezago  $h$  como  $\phi_{hh}$  o  $\alpha_x(h)$ . Para facilitar un poco la notación, se escribirá  $\gamma_x(h) = \gamma(h)$  y  $\rho_x(h) = \rho_h$  en el supuesto de que se entiende que son los referentes a la serie de tiempo  $\{X_t\}$

La correlación parcial se puede entender como la correlación entre dos variables después de ser ajustadas por un factor común que puede estar afectandolas. Por ejemplo, considérense 3 variables aleatorias  $X, Y, Z$ . Ahora, consideremos la regresión lineal simple de  $X$  en  $Z$  y de  $Y$  en  $Z$ .

$$\hat{X} = a_1 + b_1 Z \text{ donde } b_1 = \frac{\text{Cov}(Z, X)}{\text{Var}(Z)}$$

$$\hat{Y} = a_2 + b_2 Z \text{ donde } b_2 = \frac{\text{Cov}(Z, Y)}{\text{Var}(Z)}$$

$\therefore$  La correlación parcial entre  $X, Y$  es  $\text{Corr}(X - \hat{X}, Y - \hat{Y})$

Hay varias maneras de definir a la autocorrelación parcial. De las menos generales es, si  $\{X_t\}$  fuera un proceso *Gaussiano*, entonces  $\phi_{hh}$  se define como:

$$\phi_{hh} := \text{Corr}(X_t, X_{t-h} | X_{t-h+1}, \dots, X_{t-1})$$

Es decir,  $\phi_{hh}$  es la correlación en la distribución bivariada de  $X_t$  y  $X_{t-h}$  condicional a  $X_{t-h+1}, X_{t-h+2}, \dots, X_{t-1}$ . Sin embargo, la forma más general es considerando una serie de tiempo  $\{X_t\}$  y los pronósticos de  $X_t$  y  $X_{t-h}$  en función de las variables intermedias  $X_{t-h+1}, X_{t-h+2}, \dots, X_{t-1}$ . Es decir:

$$\begin{aligned} \hat{X}_t &= \beta_1 X_{t-h+1} + \dots + \beta_{h-1} X_{t-1} \\ \hat{X}_{t-h} &= \beta_1^* X_{t-h+1} + \dots + \beta_{h-1}^* X_{t-1} \end{aligned}$$

Donde  $\beta_1, \beta_1^*, \dots, \beta_{h-1}, \beta_{h-1}^*$  minimizan el *Error Cuadrático Medio* de sus respectivas ecuaciones. Por lo tanto, de esta manera  $\phi_{hh}$  se define como:

$$\phi_{hh} := \text{Corr}(X_t - \hat{X}_t, X_{t-h} - \hat{X}_{t-h})$$

Se prueba que si  $\{X_t\}$  es *Gaussiano*, ambas definiciones son equivalentes. Esta definición induce un sistema de ecuaciones dado por:

$$\rho_j = \sum_{i=1}^h \phi_{ih} \rho_{j-i} \quad j = 1, 2, \dots, h$$

$$\Rightarrow \begin{cases} \rho_1 = \phi_{1h} + \phi_{2h}\rho_1 + \cdots + \phi_{hh}\rho_{h-1} \\ \rho_2 = \phi_{1h}\rho_1 + \phi_{2h} + \cdots + \phi_{hh}\rho_{h-2} \\ \vdots \\ \rho_h = \phi_{1h}\rho_{h-1} + \phi_{2h}\rho_{h-2} + \cdots + \phi_{hh} \end{cases} \quad (2.1)$$

2.1 se puede reescribir como:

$$\underbrace{\begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{h-1} \\ \rho_1 & 1 & \cdots & \rho_{h-2} \\ \rho_2 & \rho_1 & \cdots & \rho_{h-3} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{h-1} & \rho_{h-2} & \cdots & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} \phi_{1h} \\ \phi_{2h} \\ \phi_{3h} \\ \vdots \\ \phi_{hh} \end{bmatrix}}_B = \underbrace{\begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_h \end{bmatrix}}_C$$

Sea  $A = P_h$ ,  $B = \Phi_h$  y  $C = \rho_h$ . Por lo tanto  $\Phi_h = P_h^{-1}\rho_h$  para cualquier  $h = 1, 2, \dots$ , el último coeficiente de  $\Phi_h$  es la correlación parcial con rezago  $h$ . Es claro que el obtener la correlación parcial con rezago  $h$ ,  $\phi_{hh}$  resolviendo el sistema 2.1 puede llegar a ser muy “complicado”. Sin embargo, *Levinson(1947) & Durbin(1960)* dieron un método eficiente para obtener la solución del sistema 2.1, tanto para el caso teórico como para el caso muestral. Ellos probaron que se puede resolver 2.1 de forma recursiva, obteniendo que:

$$\phi_{hh} = \frac{\rho_h - \sum_{j=1}^{h-1} \phi_{h-1,j}\rho_{h-j}}{1 - \sum_{j=1}^{h-1} \phi_{h-1,j}\rho_j}$$

Donde  $\phi_{h,j} = \phi_{h-1,j} - \phi_{hh}\phi_{h-1,h-j}$  para  $j = 1, 2, \dots, h-1$  y  $\phi_{00} = 1$

### 2.3.5. Medidas de dependencia muestrales

Una serie de tiempo estacionaria, es caracterizada por su media  $\mu$ , varianza  $\sigma^2$ , autocorrelación  $\rho_h$  y autocorrelación parcial  $\phi_{hh}$ . Debido a que en general no se posee el conjunto de todas las posibles realizaciones de la serie de tiempo, y de hecho, la mayoría

de las veces ni siquiera se tiene más de una realización, se deben tener en cuenta ciertas consideraciones de tal manera que se garanticen estimadores estadísticamente significativos y validos para las medidas de dependencia poblacionales.

### 2.3.5.1. Función de medias muestral

Un posible estimador para la función de medias  $\mu_t = \mathbb{E}(X_t)$  es la *media muestral*

$$\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t$$

pues al ser la serie de tiempo estacionaria, se cumple que  $\mu_t = \mu$ . En este caso, la media muestral es el promedio de las observaciones a lo largo del tiempo.

Claramente  $\bar{X}$  es un estimador insesgado de  $\mu$ , pues

$$\mathbb{E}(\bar{X}) = \frac{1}{n} \sum_{t=1}^n \mathbb{E}(X_t) = \frac{1}{n} (n\mu) = \mu$$

Además, se prueba que,

$$\text{Var}(\bar{X}) = \frac{\gamma_0}{n} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k$$

si se cumple que

$$\lim_{n \rightarrow \infty} \sum_{k=-(n-1)}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho_k < \infty$$

Entonces  $\text{Var}(\bar{X}) \rightarrow 0$  si  $n \rightarrow \infty$ , y por lo tanto,  $\bar{X}$  es un estimador consistente de  $\mu$  en media cuadrática.

Si lo anterior se cumple para la serie de tiempo  $\{X_t\}$ , se dice que la serie de tiempo es *ergódica para la media*. Una condición suficiente para que se cumpla lo anterior, es que  $\rho_k \rightarrow 0$  si  $k \rightarrow \infty$

### 2.3.5.2. Función de Autocovarianza Muestral

Al igual que en el caso de la función de medias, dado que generalmente se tiene una única realización de la serie de tiempo, para estimar a la función de autocorrelación parcial se proponen los siguientes dos estimadores

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (2.2)$$

$$\tilde{\gamma}_k = \frac{1}{n-k} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \quad (2.3)$$

La suma en ambos estimadores está restringida, pues  $X_{t+k}$  no está definida para  $t+k > n$

Se prueba que tanto  $\hat{\gamma}_k$  como  $\tilde{\gamma}_k$  son estimadores sesgados para  $\gamma_k$ . Si se cumple que la serie de tiempo sea *ergódica para la media*, entonces se prueba que ambos estimadores son asintóticamente insesgados.

De estos dos estimadores, se prefiere a  $\hat{\gamma}_k$  sobre  $\tilde{\gamma}_k$ , pues, en el caso en que la serie de tiempo  $\{X_t\}$  es *gaussiana*, se demuestra que  $Var(\tilde{\gamma}_k) \geq Var(\hat{\gamma}_k)$ . Además, de forma más general, el estimador  $\hat{\gamma}_k$  asegura que la matriz

$$\hat{\Gamma}_n := [\hat{\gamma}_{i-j}]_{i,j=1}^n$$

sea definida semipositiva, lo cual asegura que las varianzas muestrales de combinaciones lineales de  $X_t$ , siempre serán no negativas

Por otro lado, se prueba que la serie de tiempo es *ergódica para la ACF*, es decir, que en media cuadrática el

$$\lim_{n \rightarrow \infty} \hat{\gamma}_k = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \hat{X})(X_{t+k} - \hat{X}) = \gamma_k$$

La prueba de este resultado para nada es trivial, sin embargo, la idea general de la prueba radica en el hecho de notar que  $\hat{\gamma}_k$  es asintóticamente insesgado, por lo cual, una condición suficiente para que  $\hat{\gamma}_k$  sea consistente en media cuadrática, es que la autocovarianza sea absolutamente sumable, es decir, que  $\sum_{-\infty}^{\infty} |\gamma_i| < \infty$  y por lo tanto  $\lim_{n \rightarrow \infty} Var(\hat{\gamma}_k) = 0$ .

### 2.3.5.3. Función de Autocorrelación Muestral

La función de autocorrelación muestral se define en términos de la función de autocovarianza muestral. Si se usa como estimador de la autocovarianza muestral a  $\hat{\gamma}_k$  definida en 2.2. Entonces el estimador de  $\rho_k$  es

$$\hat{\rho}_k := \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \text{ con } |k| < n$$

Debido a que se usó a  $\hat{\gamma}_k$ , entonces la correspondiente matriz  $\hat{R}_n := [\hat{\rho}_{i-j}]_{i,j=1}^n$  también es una matriz definida semipositiva.

### 2.3.5.4. Función de Autocorrelación Parcial Muestral

La función de autocorrelación parcial muestral  $\phi_{hh}$ , se obtiene al sustituir  $\rho_i$  por su correspondiente estimador  $\hat{\rho}_i$

$$\hat{\phi}_{hh} = \frac{\hat{\rho}_h - \sum_{j=1}^{h-1} \hat{\phi}_{h-1,j} \hat{\rho}_{h-j}}{1 - \sum_{j=1}^h \hat{\phi}_{h-1,j} \hat{\rho}_j}$$

y  $\hat{\phi}_{h,j} = \hat{\phi}_{h-1,j} - \hat{\phi}_{hh} \hat{\phi}_{h-1,h-j}$

## 2.4. Series de Tiempo Lineales

Un supuesto importante que se hace en los modelos de *Box & Jenkins*, es el supuesto de linealidad. Un filtro lineal, es por ejemplo, un operador lineal de una serie de tiempo  $X_t$  en otra serie de tiempo  $Y_t$ . De tal manera que

$$Y_t = L(X_t) = \sum_{-\infty}^{\infty} \phi_i X_{t-i} \text{ con } t \in \mathbb{Z}$$

En este sentido, se puede interpretar al *filtro lineal* como un proceso que “transforma” a  $X_t$  en  $Y_t$ , y dicha conversión involucra todos los valores presentes, pasados y futuros, en

forma de una serie con diferentes pesos  $\{\phi_i\}$  en cada  $X_t$ . Dicho filtro tiene las siguientes propiedades:

1. **Invariante ante el tiempo**, ya que los coeficientes  $\{\phi_i\}$  no dependen de  $t$ .
2. **Físicamente realizable**, es decir, si  $\phi_i = 0$  para  $i < 0$ . Lo cual quiere decir, que  $Y_t$  es una función lineal de las observaciones actuales y pasadas de  $X_t$ .

$$Y_t = \phi_0 X_t + \phi_1 X_{t-1} \dots$$

3. **Estable**, si  $\sum_{-\infty}^{\infty} |\phi_i| < \infty$

### 2.4.1. Proceso de Ruido Blanco

La importancia del **Proceso de Ruido Blanco**<sup>1</sup> no radica en el hecho de que sea un modelo interesante por sí mismo, sino por el hecho de que muchos procesos se pueden construir a partir de un ruido blanco, además de sus aplicaciones en distintas disciplinas, por ejemplo, es ampliamente utilizado como modelo para el “ruido” en aplicaciones de ingeniería.

Un proceso  $\{a_t\}$  se dice que es de ruido blanco,  $\{a_t\} \sim WN(\mu_a, \sigma_a^2)$ , si es una sucesión de variables aleatorias no correlacionadas de una distribución fija con media constante  $\mathbb{E}(a_t) = \mu_a$ , usualmente se supone que  $\mu_a = 0$ , varianza constante  $Var(a_t) = \sigma_a^2$  y  $Cov(a_t, a_{t+k}) = 0 \quad \forall k \neq 0$ .

Por definición, se sigue inmediatamente que un proceso de ruido blanco,  $\{a_t\}$  es estacionario con función de autocovarianza

$$\begin{aligned} \gamma_k &= \begin{cases} \sigma_a^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \\ \implies \rho_k &= \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \\ \implies \phi_{kk} &= \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases} \end{aligned}$$

<sup>1</sup>El término “blanco” se originó por la analogía con la *luz blanca*, ya que todas las posibles oscilaciones periódicas están presentes con la misma “fuerza”

Si las variables aleatorias,  $a_t$ , además son independientes e idénticamente distribuidas, entonces se tiene un ruido blanco independiente  $\{a_t\} \sim IID(\mu_a, \sigma_a^2)$ . Un proceso de ruido blanco particularmente útil es aquel que además de ser *i.i.d.*, se cumple que las variables aleatorias siguen una distribución normal, a este proceso se le conoce *proceso de ruido blanco Gaussiano*,  $\{a_t\} \sim IIDN(\mu_a, \sigma_a^2)$ . De aquí en adelante se supondrá que  $\mu_a = 0$  para los procesos de ruido blanco.

### 2.4.2. Modelos de Series de Tiempo Lineales

La clase de modelos de series de tiempo lineales, las cuales incluyen a la clase de modelos *autoregresivos de medias móviles*, *ARMA*, proveen un marco general para estudiar procesos estacionarios. De hecho, cada proceso estacionario de segundo orden es un proceso lineal, o puede ser transformado en uno al abstraer una componente determinista. A este resultado se le conoce como *descomposición de Wold*.

Entonces, se dirá que una serie de tiempo  $X_t$ , es una serie lineal si tiene la siguiente representación

$$X_t = \mu + \sum_{-\infty}^{\infty} \psi_j a_{t-j} \quad \forall t \in \mathbb{Z}.$$

Donde  $\{a_t\} \sim WN(0, \sigma_a^2)$  y  $\{\psi_j\}$  es una sucesión de constantes tales que  $\sum_{-\infty}^{\infty} |\psi_j| < \infty$ .

Introduciendo el operador de rezago  $B^j X_t = X_{t-j}$ , y definiendo a  $\dot{X}_t = X_t - \mu$ , podemos reescribir a  $X_t$  como

$$\dot{X}_t = \Psi(B)a_t$$

donde  $\Psi(B) = \sum_{-\infty}^{\infty} \psi_j B^j$ . Pensando que  $\Psi(B)$  es un operador lineal, tiene la propiedad de ser realizable, por lo tanto

$$\dot{X}_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad \forall t \in \mathbb{Z}.$$

A un proceso lineal también se le conoce como, *representación de Wold* o como *proceso no determinista*. Se prueba que  $\{X_t\}$  es estacionario y que su función de autocovarianza

es

$$\begin{aligned}\gamma_X(k) &= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \psi_i \psi_j \gamma_a(i-j-k) \\ &= \sigma_a^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k} \\ \implies \rho_X(k) &= \frac{\gamma_X(k)}{\gamma_X(0)} = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+k}}{1 + \sum_{i=1}^{\infty} \psi_i^2} \quad \text{con } \psi_0 = 1\end{aligned}$$

A este proceso se le llama, *proceso de medias móviles infinito* y sirve como una clase general de modelos para cualquier serie de tiempo estacionaria. Lo anterior se cumple por el *Teorema de Wold*, el cual, básicamente establece que cualquier serie de tiempo no determinista y débilmente estacionaria, se puede representar como un *proceso de medias móviles infinito* donde  $\{\psi_i\}$  satisface que  $\sum_{i=0}^{\infty} \psi_i^2 < \infty$ . Una forma intuitiva para entender este teorema es que, una serie de tiempo estacionaria se puede expresar como la suma ponderada de sus *choques* aleatorios presentes y pasados.

### 2.4.3. Proceso de Medias Móviles, MA

Si sólo un número finito de pesos  $\psi$  son distintos de cero, es decir,  $\psi_1 = -\theta_1, \psi_2 = -\theta_2, \dots, \psi_q = -\theta_q$  y  $\psi_k = 0 \quad \forall k > q$ , entonces, el proceso resultante se dice que es un *proceso de medias móviles de orden q* y se denota como MA(q).

$$X_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (2.4)$$

Donde  $\{\varepsilon_t\}$  es un proceso de ruido blanco. Debido a que 2.4 es un caso especial de un  $MA(\infty)$ , con únicamente un número finito de pesos, un  $MA(q)$  siempre es estacionario sin importar el valor de sus pesos. En términos del operador de rezago y usando a  $\dot{X}_t = X_t - \mu$ , 2.4, se puede escribir como

$$\begin{aligned}
X_t &= \mu + (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \\
&= \mu + \sum_{i=0}^q \theta_i B^i \varepsilon_t \\
&= \mu + \Theta(B) \varepsilon_t \\
\dot{X}_t &= \Theta(B) \varepsilon_t \quad \text{donde} \quad \Theta(B) = \sum_{i=0}^q \theta_i B^i \quad \text{y} \quad \theta_0 = 1
\end{aligned}$$

Este tipo de procesos son útiles para describir fenómenos donde los eventos producen efectos inmediatos, los cuales duran un corto periodo de tiempo. La terminología de medias móviles surge del hecho que  $X_t$  se obtiene al aplicar los pesos  $1, -\theta_1, -\theta_2, \dots, -\theta_q$  a las variables  $\varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q}$  y luego, “se mueven los pesos” y se aplican a  $\varepsilon_{t+1}, \varepsilon_t, \dots, \varepsilon_{t-q+1}$  para obtener a  $X_{t+1}$ . Este tipo de procesos fueron inicialmente considerados por *Slutsky(1927)* y *Wold(1938)*.

Además, este proceso tiene una forma particular para su función de autocorrelación.

Considérese un proceso estocástico,  $\{X_t\}$ ,  $MA(q)$ , considérese el proceso  $\dot{X}_t = X_t - \mu$ , es importante notar que  $\mathbb{E}[\dot{X}_t] = 0$ . Entonces,  $\gamma_{\dot{X}}(h) = Cov(\dot{X}_t, \dot{X}_{t+h}) = \mathbb{E}[\dot{X}_t \dot{X}_{t+h}]$ .

Por otro lado,

$$\begin{aligned}
\dot{X}_t \cdot \dot{X}_{t+h} &= \left( \sum_{i=0}^q \theta_i \varepsilon_{t-i} \right) \left( \sum_{j=0}^q \theta_j \varepsilon_{t+h-j} \right) \\
&= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \varepsilon_{t-i} \cdot \varepsilon_{t+h-j} \\
\implies \mathbb{E}[\dot{X}_t \cdot \dot{X}_{t+h}] &= \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \mathbb{E}[\varepsilon_{t-i} \cdot \varepsilon_{t+h-j}]
\end{aligned}$$

Como  $\mathbb{E}[\varepsilon_t] = 0 \quad \forall t$ , se tiene que  $\gamma_{\varepsilon}(h) = Cov(\varepsilon_t, \varepsilon_{t+h}) = \mathbb{E}[\varepsilon_t \cdot \varepsilon_{t+h}]$ .

Por lo cual, se tiene que

$$\gamma_{\dot{X}}(h) = \sum_{i=0}^q \sum_{j=0}^q \theta_i \theta_j \gamma_{\varepsilon}(h+i-j)$$

Además, como  $\gamma_\varepsilon(h)$  es  $\sigma^2$  si  $h = 0$  y 0 e.o.c, en la ecuación anterior, es de interés los casos cuando  $h + i - j = 0$ , es decir, cuando  $j = h + i$ .

Como  $j \leq q \implies h + i \leq q \implies i \leq q - h$  y como  $i \geq 0$  se tiene que  $0 \leq i \leq q - h$ .

$$\therefore \gamma_{\dot{X}}(h) = \begin{cases} \sum_{j=0}^{q-h} \theta_j \theta_{h+j} \sigma_\varepsilon^2 & \text{si } 0 \leq h \leq q \\ 0 & \text{si } h > q \end{cases}$$

Lo anterior implica que  $\text{Var}(\dot{X}) = \sigma_\varepsilon^2 \sum_{j=0}^q \theta_j^2$  y

$$\therefore \rho_{\dot{X}}(h) = \frac{\gamma_{\dot{X}}(h)}{\gamma_{\dot{X}}(0)} = \begin{cases} \frac{\sum_{j=0}^{q-h} \theta_j \theta_{h+j}}{\sum_{j=0}^q \theta_j^2} & \text{si } 0 \leq h \leq q \\ 0 & \text{si } h > q \end{cases}$$

Por ejemplo, un proceso de medias móviles de orden 1,  $MA(1)$ , tendrá como función de autocovarianza

$$\gamma_{\dot{X}}(h) = \sigma_\varepsilon^2 \sum_{j=0}^1 \theta_j^2 = \sigma_\varepsilon^2 (1 + \theta_1^2), \quad \text{para } 0 \leq h \leq 1$$

y como función de autocorrelación

$$\rho_{\dot{X}}(h) = \frac{\sum_{j=0}^{1-h} \theta_j \theta_{h+j}}{\sum_{j=0}^1 \theta_j^2} = \frac{\sum_{j=0}^1 \theta_j^2}{\sum_{j=0}^1 \theta_j^2} = 1, \quad \text{para } 0 \leq h \leq 1$$

#### 2.4.4. Proceso Autoregresivo, AR

Para tener clara la forma general de un proceso autoregresivo, primero se considerará el proceso autoregresivo más simple, el proceso autoregresivo de primer orden, el  $AR(1)$ .

Considérese el proceso dado por

$$X_t = \mu + \Psi(B)\varepsilon_t$$

donde  $\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$ , así como en el caso de los procesos *MA* finitos, también se supondrá que la contribución de los “choques” ocurridos en el pasado, deben ser menores en comparación con los ocurridos recientemente. Debido a que estos *choques*, son un proceso de ruido blanco, se puede suponer un conjunto infinito de pesos, pero decrecientes, que reflejen la disminución en la magnitud de la contribución del *choque*.

Un conjunto simple, e intuitivo, de este tipo de pesos, se puede generar a través de un decaimiento exponencial. Para esto, se define  $\psi = \phi^i$ , donde,  $|\phi| < 1$  para garantizar el decaimiento exponencial.

$$\begin{aligned} \text{Entonces } X_t &= \mu + \varepsilon_t + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \dots \\ &= \mu + \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i} \end{aligned}$$

Pero también se cumple que

$$X_{t-1} = \mu + \varepsilon_{t-1} + \phi\varepsilon_{t-2} + \phi^2\varepsilon_{t-3} + \dots$$

Por lo cual, multiplicando a  $X_{t-1}$  por  $\phi$  tenemos

$$\begin{aligned} \phi X_{t-1} &= \phi\mu + \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \dots \\ \implies \phi X_{t-1} - \phi\mu &= \phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \dots \end{aligned}$$

Esta expresión, se puede sustituir en la dada por  $X_t$ , pues

$$\begin{aligned} X_t &= \mu + \varepsilon_t + \underbrace{\phi\varepsilon_{t-1} + \phi^2\varepsilon_{t-2} + \phi^3\varepsilon_{t-3} + \dots}_{\phi X_{t-1} - \phi\mu} \\ \implies X_t &= \mu + \varepsilon_t + \phi X_{t-1} - \phi\mu \\ \implies X_t &= \delta + \phi X_{t-1} + \varepsilon_t \quad \text{donde } \delta = \mu - \phi\mu \end{aligned}$$

El proceso definido por esta ecuación es conocido como un proceso autoregresivo de primer orden,  $AR(1)$ , debido a que se puede ver como una regresión de  $X_t$  en  $X_{t-1}$ , y de aquí, el término *autoregresivo*. El supuesto  $|\phi| < 1$ , además de garantizar el decaimiento exponencial, también garantiza que  $\sum_{i=0}^{\infty} |\psi_i| < \infty$ . Por lo tanto, el  $AR(1)$  es un proceso estacionario si  $|\phi| < 1$ .

Se prueba que

$$\begin{aligned}\mathbb{E}(X_t) &= \mu = \frac{\delta}{1 - \phi} \\ \gamma_k &= \frac{\sigma^2 \phi^k}{1 - \sigma^2} \quad \text{para } k = 0, 1, 2, \dots, \\ \rho_k &= \phi^k \quad \text{para } k = 0, 1, 2, \dots,\end{aligned}$$

Por lo tanto, ya que ahora se tiene la idea de un modelo autoregresivo, es claro que los modelos autoregresivos se basan en que el valor actual de la serie  $X_t$  puede ser explicado como función de  $p$  valores pasados,  $X_{t-1} \dots X_{t-p}$ , donde  $p$  determina el número de “pasos” hacia el pasado necesarios para pronosticar el valor actual.

Por lo tanto, se define al proceso autoregresivo general de orden  $p$ ,  $AR(p)$  como

$$X_t = \delta + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (2.5)$$

donde  $\varepsilon_t$  es un proceso de ruido blanco, 2.5 se puede escribir como

$$\Phi(B)X_t = \delta + \varepsilon_t$$

donde  $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$

Es decir, el valor actual  $X_t$  es una combinación lineal de sus valores más recientes más una *innovación*<sup>2</sup>. Yule(1926) llevó a cabo los trabajos originales sobre este tipo de procesos.

El proceso  $AR(p)$  es estacionario, si las raíces de su polinomio característico en  $B$ ,

$$1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p = 0$$

---

<sup>2</sup>Se usa el término *innovación* debido a que este término incorpora todo lo nuevo en la serie a tiempo  $t$  que no es explicado por los valores pasados.

son mayores a 1 en valor absoluto. Además, bajo este supuesto, se dice que el proceso AR(p) tiene una representación MA( $\infty$ ) absolutamente sumable. Además, se prueba que

$$\mathbb{E}(X_t) = \frac{\delta}{1 - \phi_1 - \dots - \phi_p}$$

$$\gamma_k = \sum_{i=1}^p \phi_i \gamma_{k-i} + \begin{cases} \sigma^2 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

$$\rho_k = \sum_{i=1}^p \phi_i \rho_{k-i} \quad \text{para } k = 1, 2, \dots,$$

Las ecuaciones dadas por  $\rho_k$ , inducen a las ecuaciones de *Yule-Walker*, las cuales son muy usadas para la estimación de los parámetros del modelo.

### 2.4.5. Proceso Autoregresivo de Medias Móviles, ARMA

Una extensión natural de los proceso autoregresivos puros y de los procesos de medias móviles, es el proceso ARMA, el cual, combina las ideas de los modelos AR y MA de una forma compacta, de tal manera que el número de parámetros usados se mantenga “pequeño”. Este proceso contiene una basta clase de modelos de series de tiempo parsimoniosos muy útiles para describir a una amplia gama de series de tiempo en la práctica.

Entonces, un modelo ARMA(p,q) general es de la forma

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

$$X_t - \sum_{i=1}^p \phi_i X_{t-i} = \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i} \quad (2.6)$$

donde  $\{\varepsilon\}$  es un proceso de ruido blanco,  $\{X_t\}$  es un proceso estacionario,  $\phi_p \neq 0$ ,  $\theta_q \neq 0$ ,  $\sigma_\varepsilon^2 > 0$  y,  $p, q$  son enteros no negativos. Es claro que los modelos AR y MA son casos particulares del modelo ARMA(p,q). Usando el operador de rezago, se puede expresar al modelo 2.6 como

$$\phi(p)X_t = \theta(q)\varepsilon_t$$

En la ecuación anterior.  $\Phi(p) = 1 - \phi_1 B - \dots - \phi_p B^p$  es el *polinomio AR* del modelo, mientras que  $\Theta(q) = 1 - \theta_1 B - \dots - \theta_q B^q$  es el *polinomio MA*. Se requiere que no haya factores comunes entre los polinomios  $(1 - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p})$  y  $(1 - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q})$ , pues si esto no fuera así, el orden  $(p, q)$  del modelo ARMA podría reducirse.

En caso de que la media de  $\{X_t\}$  fuera distinta de cero, se dice que el proceso  $\{X_t\}$  es un proceso ARMA(p,q) con media  $\mu$ , si  $\{X_t - \mu\}$  es un proceso ARMA(p,q). En cuyo caso, 2.6 se puede escribir de la siguiente manera

$$X_t = \alpha + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{i=1}^q \theta_i \varepsilon_{t-i},$$

donde  $\alpha = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$

Como en el caso AR puro, se requiere que las soluciones del polinomio característico asociado al polinomio AR del modelo ARMA tengan su módulo mayor a uno, para que el proceso ARMA sea estacionario. Es necesario que  $\{X_t\}$  sea estacionario, pues, para que exista una única solución estacionaria de la ecuación 2.6, se debe cumplir que  $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0$  para toda  $z \in \mathbb{C}$  con  $|z| \leq 1$ . A la región definida por el conjunto de números complejos  $z$  tales que  $|z| = 1$ , se le conoce como círculo unitario.

Si  $\phi(z) \neq 0$  para toda  $z$  en el círculo unitario, entonces existe  $\delta > 0$  tal que

$$\frac{1}{\phi(z)} = \sum_{j=-\infty}^{\infty} \chi_j z^j \quad \text{para } 1 - \delta < |z| < 1 + \delta \quad \text{y} \quad \sum_{j=-\infty}^{\infty} |\chi_j| < \infty$$

Entonces, se puede definir a  $\frac{1}{\phi(B)}$  como el filtro lineal con coeficientes absolutamente sumables

$$\frac{1}{\phi(B)} = \sum_{j=-\infty}^{\infty} \chi_j B^j$$

Aplicando el operador  $\chi(B) := \frac{1}{\phi(B)}$  a ambos lados de 2.6 se obtiene

$$X_t = \chi(B)\phi(B)X_t = \chi(B)\theta(B)\varepsilon_t = \psi(B)\varepsilon_t = \sum_{j=-\infty}^{\infty} \psi_j \varepsilon_{t-j}$$

donde  $\psi(z) = \chi(z)\theta(z) = \sum_{j=-\infty}^{\infty} \psi_j z^j$ , se sigue que la única solución estacionaria de 2.6 es  $\psi(B)\varepsilon_t$

Un proceso ARMA(p,q) estacionario, tiene tres representaciones importantes las cuales tienen diferente utilidad, la primera representación es la que está dada tal cual por su definición

$$\phi(p)X_t = \theta(q)\varepsilon_t$$

Esta representación es compacta y útil para estimar los parámetros, así como para calcular, de forma recursiva, pronósticos de la serie  $X_t$  a múltiples “pasos”. Las otras dos representaciones se detallan a continuación

### 2.4.5.1. Causalidad

Se dice que un proceso ARMA(p,q) definido por la ecuación  $\phi(p)X_t = \theta(q)\varepsilon_t$ , es **causal**, o una función causal, si existe una sucesión de constantes  $\{\psi_j\}$  tales que  $\sum_{-\infty}^{\infty} |\psi_j| < \infty$  y si la serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  puede ser escrita como un proceso lineal  $MA(\infty)$

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} = \psi(B)\varepsilon_t, \quad \forall t \in \mathbb{Z}$$

Donde  $\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j$  y  $\psi_0 = 1$ . Es importante notar que la causalidad no es únicamente una propiedad del proceso  $\{X_t\}$ , sino de la relación entre los procesos,  $\{X_t\}$  y  $\{\varepsilon_t\}$ , presentes en la ecuación del modelo ARMA. Se puede decir que  $\{X_t\}$  es causal si se puede expresar en términos de  $\{\varepsilon_t\}$  al aplicar un *filtro lineal causal*.

Se prueba que si  $\{X_t\}$  es un proceso ARMA(p,q), para el cual, los polinomios  $\phi(\cdot)$  y  $\theta(\cdot)$  no tienen ceros en común. Entonces,  $\{X_t\}$  es causal sí y sólo sí  $\phi(z) \neq 0$ ,  $\forall z \in \mathbb{C}$  tal que  $|z| \leq 1$ . Los coeficientes  $\{\psi_j\}$  se determinan a través de la relación

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}, \quad |z| \leq 1.$$

Lo anterior se puede entender de la siguiente manera, un proceso ARMA es causal únicamente cuando la raíces del polinomio AR,  $\phi(z)$ , caen fuera del círculo unitario, es decir,  $\phi(z) = 0$  únicamente cuando  $|z| > 1$

La representación *causal* del proceso ARMA muestra explícitamente el impacto de los *choques* aleatorios pasados,  $\varepsilon_{t-i}$  tales que  $i > 0$ , en el valor actual de la serie  $X_t$ . Los

coeficientes  $\{\psi_j\}$  son conocidos como la *función impulso de respuesta* del modelo ARMA. Para una serie estacionaria, los coeficientes  $\psi_j$  decaen exponencialmente conforme  $i$  aumenta.

Esta representación es útil para calcular la varianza del error de pronóstico, de igual manera, provee de una prueba simple de la propiedad de tendencia a la media de una serie de tiempo estacionaria.

### 2.4.5.2. Invertibilidad

Se dice que un proceso ARMA(p,q) definido por la ecuación  $\phi(p)X_t = \theta(q)\varepsilon_t$  es invertible si existe una sucesión de constantes  $\{\pi_j\}$  tales que  $\sum_{j=0}^{\infty} |\pi_j| < \infty$  y además si el proceso  $\{\varepsilon_t \in \mathbb{Z}\}$  puede ser escrito como un proceso lineal

$$\varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} = \pi(B)X_t, \quad \forall t \in \mathbb{Z}$$

Donde  $\pi(B) = \sum_{j=0}^{\infty} \pi_j B^j$  y  $\pi_0 = 1$ . Así como la causalidad, la propiedad de ser invertible no es una propiedad de  $\{X_t\}$  por sí solo, sino de la relación entre los dos procesos  $\{X_t\}$  y  $\{\varepsilon_t\}$  presentes en la ecuación que define al modelo ARMA.

Se prueba que si  $\{X_t\}$  es un proceso ARMA(p,q), para el cual, los polinomios  $\phi(\cdot)$  y  $\theta(\cdot)$  no tienen ceros en común, entonces,  $\{X_t\}$  es **invertible** sí y sólo sí  $\theta(z) \neq 0, \quad \forall z \in \mathbb{C}$  tal que  $|z| \leq 1$ . Los coeficientes  $\{\pi_j\}$  se determinan por la relación

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}, \quad |z| \leq 1$$

Lo anterior se puede entender de la siguiente manera, un proceso ARMA es invertible únicamente cuando la raíces del polinomio MA,  $\theta(z)$ , caen fuera del círculo unitario, es decir,  $\theta(z) = 0$  únicamente cuando  $|z| > 1$

Por último, se cumplirá la propiedad de causalidad e invertibilidad para un proceso ARMA, si la serie de tiempo  $\{X_t : t \in \mathbb{Z}\}$  es una solución estacionaria de la ecuación

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

y si  $\phi(z), \theta(z) \neq 0$  para  $|z| \leq 1$ , entonces

$$X_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \quad \text{y} \quad \varepsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$$

donde  $\sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}$  y  $\sum_{j=0}^{\infty} \pi_j z^j = \frac{\phi(z)}{\theta(z)}$

### 2.4.6. Proceso ARMA multiplicativo estacional, SARMA

Diversas series de tiempo en áreas como Economía o Finanzas, muestran cierto comportamiento cíclico o periódico a lo largo del tiempo  $t$ . Al periodo de tiempo en el que ocurre este fenómeno repetitivo se le conoce como *periodo estacional*. Por lo cual, a las series de tiempo que tienen estos comportamientos, se les conoce como *series de tiempo estacionales*. Los fenómenos estacionales pueden estar causados por diversos factores, como por ejemplo, el clima, el cual afecta actividades económicas como el turismo.

El análisis de dichas series de tiempo, tiene una larga historia, sin embargo, en algunas aplicaciones, la estacionalidad es de poca importancia y simplemente es removida de los datos, resultando en una serie de tiempo ajustada por estacionalidad, la cual, es usada para hacer inferencia. No obstante, combinando las propiedades estacionales y no estacionales en los modelos ARMA, se pueden desarrollar modelos parsimoniosos que contengan la autocorrelación de los rezagos estacionales, y además, para rezagos “pequeños”, la correlación de los valores vecinos de la serie.

En general, se define un *ARMA multiplicativo estacional*, como un proceso  $ARMA(p, q) \times (P, Q)_s$ , con periodo estacional  $s$ , el cual, posee un polinomio característico AR,  $\phi(x)\Phi(x)$  y un polinomio característico MA,  $\theta(x)\Theta(x)$ , este modelo satisface la siguiente ecuación

$$\phi(B)\Phi(B)X_t = \theta(B)\Theta(B)\varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

Donde,

$$\begin{aligned} \phi(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi(B) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \\ &\text{y} \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \\ \Theta(B) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_Q B^{Qs} \end{aligned}$$

Es importante notar, que se tiene únicamente un caso especial de un modelo ARMA, cuya componente AR es de orden  $p + P * s$  y cuya componente MA es de orden  $q + Q * s$ , además, los coeficientes del modelo SARMA son únicamente determinados por  $p, P, q, Q$ , teniendo en total  $p + P + q + Q$  coeficientes.

### 2.4.7. Distribuciones asintóticas de SACF y SPACF de un proceso lineal

*SACF* y *SPACF*, se refieren a las funciones de autocorrelación y autocorrelación parcial muestrales, lo anterior por sus siglas en inglés. Intentar encontrar la distribución de  $\hat{\rho}_k$  y de  $\hat{\phi}_{kk}$  es sumamente complicado incluso para las series de tiempo más simples, sin embargo, supóngase que se tiene un proceso de la forma  $X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_t$ , donde  $\{a_t\}$  es un proceso de ruido blanco. Además, se supone que  $\sum_{j=0}^{\infty} |\psi_j| < \infty$  y  $\sum_{j=0}^{\infty} j \psi_j^2 < \infty$ , nótese que estas restricciones se satisfacen para cualquier proceso ARMA estacionario.

Entonces, para cualquier  $k$  fija, la distribución conjunta del vector

$$(\sqrt{n}(\hat{\rho}_1 - \rho_1), \sqrt{n}(\hat{\rho}_2 - \rho_2), \dots, \sqrt{n}(\hat{\rho}_k - \rho_k))^T$$

tiende a una distribución normal multivariada con media cero, varianza  $c_{jj}$ , y covarianza  $c_{ij}$ , si  $n \rightarrow \infty$ . Donde

$$c_{ij} = \sum_{-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2) \quad (2.7)$$

Lo anterior, se puede escribir de forma compacta de la siguiente manera

$$\hat{\rho}_k \approx N_k(\rho_k, n^{-1}C)$$

donde  $\rho_k = (\rho_1, \dots, \rho_k)^T$  y  $C$  es la matriz de varianzas y covarianzas, cuyas entradas  $(i, j)$  están dadas por la ecuación 2.7, a esta ecuación se le conoce como *ecuación de Bartlett*.

Por lo tanto, para  $n$  suficientemente grande,  $\hat{\rho}_k$  se distribuye aproximadamente normal, con media  $\rho_k$  y varianza  $\frac{c_{kk}}{n}$ . Además,  $\text{Corr}(\hat{\rho}_k, \hat{\rho}_j) \approx \frac{c_{kj}}{\sqrt{c_{kk}c_{jj}}}$ . Aunque claramente 2.7 es difícil de interpretar y calcular, se simplifica si  $\{X_t\}$  es un proceso de ruido blanco, pues

entonces, 2.7 se reduce considerablemente y se obtiene que  $Var(\hat{\rho}_k) \approx \frac{1}{n}$  y  $Corr(\hat{\rho}_k, \hat{\rho}_j) \approx 0$  para  $k \neq j$

Por ejemplo, si se tuviera un proceso  $\{X_t\} \sim IID(0, \sigma^2)$ , entonces  $\rho_k = 0 \quad \forall |k| > 0$

$$\implies c_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{e.o.c} \end{cases}$$

Lo cual implica que, para  $n$  suficientemente grande,  $\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_k$ , son variables aleatorias aproximadamente normales, independientes e idénticamente distribuidas con media 0 y varianza  $\frac{1}{n}$ . Lo anterior provee de un método para poder probar cuándo posibles “picos” en  $\hat{\rho}_k$  sean significativos, al determinar cuándo estos picos caigan fuera del intervalo dado por  $\pm \frac{2}{\sqrt{n}}$ . Es claro que para el ejemplo, aproximadamente el 95% de los valores de la ACF deben estar dentro de estos límites.

Para procesos en los cuales  $\rho_k = 0$  para  $k > m$ , la aproximación de *Bartlett* para  $Var(\hat{\rho}_k)$  es  $Var(\hat{\rho}_k) \simeq \frac{1}{n}(1 + 2\rho_1^2 + 2\rho_2^2 + \dots + 2\rho_m^2)$ . Como generalmente  $\rho_i$  es desconocido, se sustituye por su estimador  $\hat{\rho}_i$ , lo cual, induce el siguiente error estándar para  $\hat{\rho}_k$ ,

$$S_{\hat{\rho}_k} = \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \dots + 2\hat{\rho}_m^2)}$$

Lo descrito en el párrafo anterior se cumple para cualquier proceso MA(q), por lo tanto, se obtuvo un método para probar la hipótesis nula de  $H_0 : \rho_l = 0 \quad \text{vs} \quad H_a : \rho_l \neq 0$  para  $l > q$ . Es importante recordar que  $\hat{\rho}_l$ , es un estimador sesgado de  $\rho_l$ , cuyo sesgo es del orden de  $\frac{1}{n}$ , el cual, puede ser considerable si  $n$  es pequeño.

Por último, en el caso de la PACF, *Quenouille(1949)* probó que, bajo la hipótesis nula de que un modelo AR(p) es correcto para una serie de tiempo  $\{X_t\}$ , la PACF muestral con rezagos mayores a  $p$ , tiene distribución aproximadamente normal con media 0 y varianza  $\frac{1}{n}$ , es decir,

$$\sqrt{n}\hat{\phi}_{kk} \xrightarrow{d} N(0, 1) \quad \text{para } k > p$$

Por lo tanto, para  $k > p$ ,  $\pm \frac{2}{\sqrt{n}}$ , pueden usarse como límites críticos sobre  $\hat{\phi}_{kk}$  para probar la hipótesis nula de que es correcto modelar la serie  $\{X_t\}$  con un AR(p).

## 2.5. Modelos de series de tiempo no estacionarias

Con respecto a la clase de procesos estacionarios en covarianza, la no estacionariedad, puede ocurrir de diversas formas; por ejemplo, puede tener función de medias  $\mu_t$  no constante, segundos momentos dependientes del tiempo  $t$ , es decir,  $\sigma_t^2$  o covarianza dependientes del tiempo  $t$ .

Un proceso, que no es estacionario debido a que su función de medias no es constante con respecto al tiempo  $t$ , puede ser un problema difícil de solucionar si no se tienen suficientes realizaciones de dicho proceso para poder estimar la función de medias  $\mu_t$ . Sin embargo, afortunadamente existen modelos que permiten construir estimaciones para la función  $\mu_t$  a partir de una única realización del proceso.

En muchas situaciones, las series de tiempo no estacionarias, pueden verse como “algo” compuesto por dos componentes, una componente de tendencia no estacionaria y una componente estacionaria, es decir

$$X_t = \mu_t + Y_t$$

donde  $\mu_t$  es la función de medias que se busca estimar y  $Y_t$  es una serie de tiempo estacionaria. Dos de los modelos más comunes para modelar la función de medias  $\mu_t$  en procesos no estacionarios se describen a continuación.

### 2.5.1. Modelos con tendencia determinista

La función de medias de un proceso no estacionario puede representarse por una tendencia determinista en el tiempo. En tal caso, un modelo de regresión puede ser útil para describir el fenómeno. Es importante recalcar que la tendencia no necesariamente debe ser lineal. Por ejemplo, si la tendencia resulta ser lineal, puede modelarse a  $\mu_t$  mediante una regresión lineal, de tal manera que  $\mu_t = \alpha_0 + \alpha_1 t$ . Por lo tanto, en este caso la serie de tiempo  $X_t$  quedaría descrita por la siguiente ecuación

$$X_t = \alpha_0 + \alpha_1 t + Y_t$$

### 2.5.2. Modelos con tendencia estocástica

Aunque muchas series de tiempo no sean estacionarias, debido a ciertas fuerzas de equilibrio, distintas “partes” de estas serie tienen comportamientos similares excepto por ciertas diferencias en los niveles locales de la media. *Box & Jenkins* se refieren a este tipo de comportamiento no estacionario, como *no estacionariedad homogénea*. En términos de los modelos ARMA, el proceso no es estacionario si algunas raíces del polinomio AR no están fuera del círculo unitario. Debido a esto, en la literatura de series de tiempo, a las series no estacionarias, generalmente se les conoce como *series de tiempo no estacionarias con raíces unitarias*.

Formalmente se dirá, que una serie de tiempo  $X_t$ , es *homogénea no estacionaria*, si  $W_t = (1 - B)^d X_t$  es una serie de tiempo estacionaria.

Por lo cual, se define a  $\phi(B)$  como el operador autoregresivo que describe el comportamiento de la serie, entonces, se tiene que

$$\psi(B)(X_t + c) = \psi(B)X_t$$

para cualquier constante  $c$ . Esta ecuación, implica que la forma de  $\psi(B)$  debe ser  $\psi(B) = \phi(B)(1 - B)^d$ , para alguna  $d > 0$ , donde  $\phi(B)$  es un operador autoregresivo estacionario. Por lo tanto, una serie homogénea no estacionaria, puede reducirse a una serie estacionaria tomando una serie diferenciada apropiadamente. En otras palabras, la serie  $\{X_t\}$  no estacionaria, al ser diferenciada  $d$  veces, produce la serie  $\{(1 - B)^d X_t\}$ , la cual, para alguna  $d \geq 1$  es estacionaria.

Por ejemplo, si la serie diferenciada  $d$  veces, sigue un proceso de ruido blanco, entonces se tiene

$$(1 - B)^d X_t = a_t$$

Considerando a  $d = 1$ , se tiene

$$\begin{aligned} (1 - B)X_t &= a_t \\ \implies X_t &= X_{t-1} + a_t \end{aligned}$$

Entonces, dada la información pasada de  $X_t$ , es decir,  $X_{t-1}, X_{t-2}, \dots$ , el nivel de la serie al tiempo  $t$  es

$$\mu_t = X_{t-1}$$

el cual está sujeto al disturbio estocástico al tiempo  $t - 1$ . En otras palabras, el nivel de la media del proceso  $X_t$  en  $(1 - B)^d X_t$ , con  $d \geq 1$ , cambia a lo largo del tiempo de

forma estocástica, por lo cual, caracterizamos al proceso diciendo que tiene una tendencia estocástica. Es clara la diferencia con el caso determinista, donde la tendencia (nivel de la media), al tiempo  $t$  es una función puramente determinista de  $t$ .

El operador diferencia no es únicamente de utilidad para filtrar una tendencia posible-mente estocástica, en el caso en el que sea una tendencia determinista, la cual se puede modelar a través de un polinomio, puede ser filtrada con este operador. Por ejemplo, supongase que

$$X_t = \mu_t + Y_t$$

donde  $\mu_t = \beta_0 + \beta_1 t$  y  $Y_t$  es un proceso estacionario, al diferenciar  $X_t$  se obtiene una serie estacionaria, pues

$$\begin{aligned}\nabla X_t &= X_t - X_{t-1} \\ &= Y_t - Y_{t-1} + \beta_0 + \beta_1 t - \beta_0 - \beta_1(t-1) \\ &= \beta_1 + \nabla Y_t\end{aligned}$$

En el caso en que  $\mu_t$  es estocástica, supongase de la forma  $\mu_t = \mu_{t-1} + v_t$ , donde  $v_t$  es un proceso de ruido blanco, se tendría

$$\begin{aligned}\nabla X_t &= X_t - X_{t-1} \\ &= Y_t + \mu_t - Y_{t-1} - \mu_{t-1} \\ &= \nabla Y_t + \mu_{t-1} + v_t - \mu_{t-1} \\ &= \nabla Y_t + v_t \quad \text{la cual, es una serie estacionaria}\end{aligned}$$

### 2.5.3. Raíces unitarias

Previamente se mencionó, que a las series de tiempo que no son estacionarias, se les conoce como *serie de tiempo no estacionarias CON RAÍZ UNITARIA*. Esto tiene una razón de ser, pues, el problema de raíces unitarias en series de tiempo surge cuando el polinomio AR o MA, de un modelo ARMA, tiene una raíz en el círculo unitario o “cerca” de él. Por ejemplo, una raíz cerca del 1 en el polinomio autoregresivo, sugiere que los datos deberían ser diferenciados antes de ajustar el modelo ARMA, mientras que una raíz cerca del 1 en el polinomio MA, sugiere que los datos fueron sobrediferenciados.

El acercamiento para conocer de una forma objetiva cuántas veces se debería diferenciar una serie fue desarrollado por *Dickey & Fuller* en 1979. A continuación se mostrará el caso de raíces unitarias para un proceso AR(1), el cual, se generaliza al caso general AR(p).

Sea  $X_1, \dots, X_n$  observaciones de un modelo AR(1)

$$\begin{aligned} X_t - \mu &= \phi_1(X_{t-1} - \mu) + \varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2) \\ \implies X_t &= \phi_0 + \phi_1 X_{t-1} + \varepsilon_t, \quad \text{donde } \phi_0 = \mu(1 - \phi_1) \end{aligned} \quad (2.8)$$

Es importante notar que si  $\phi_0 = 0$  se tiene

$$X_t = \phi_1 X_{t-1} + \varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2) \quad (2.9)$$

Suponiendo 2.8 y que  $|\phi_1| < 1$  y  $\mu = \mathbb{E}(X_t)$ , para una  $n$  suficientemente grande, el estimador máximo verosímil,  $\hat{\phi}_1$  de  $\phi_1$  tiene una distribución aproximada  $N(\phi_1, \frac{1-\phi_1^2}{n})$ . Sin embargo, en el caso de raíz unitaria, esta aproximación normal no es aplicable, incluso asintóticamente, lo cual, impide su uso para realizar *la prueba de la raíz unitaria*  $H_0 : \phi_1 = 1$  v.s.  $H_1 : \phi_1 < 1$ . Para construir la prueba apropiada, se escribe el modelo de la siguiente manera

$$\begin{aligned} \nabla X_t &= X_t - X_{t-1} \\ &= \phi_0^* + \phi_1^* X_{t-1} + \varepsilon_t \\ &= X_t + \mu - \mu\phi_1 + \phi_1 X_{t-1} + \varepsilon_t - X_{t-1} \\ &= X_t + X_{t-1}(\phi_1 - 1) + \varepsilon_t + \mu(1 - \phi_1) \\ \text{donde } \phi_0^* &= \mu(1 - \phi_1) \quad \text{y} \quad \phi_1^* = (\phi_1 - 1) \\ \implies \nabla X_t &= \phi_0^* + \phi_1^* X_{t-1} + \varepsilon_t \end{aligned}$$

Sea  $\hat{\phi}_1^*$  el estimador obtenido por mínimos cuadrados ordinarios de  $\phi_1^*$ , obtenido al realizar la regresión de  $\nabla X_t$  sobre 1 y  $X_{t-1}$ . El estimador del error estándar de  $\hat{\phi}_1^*$  es

$$\begin{aligned} \hat{SE}(\hat{\phi}_1^*) &= S \left( \sum_{t=2}^n (X_{t-1} - \bar{X})^2 \right)^{\frac{1}{2}} \\ \text{Donde } S^2 &= \frac{\sum_{t=2}^n (\nabla X_t - \hat{\phi}_0^* - \hat{\phi}_1^* X_{t-1})^2}{n-3} \end{aligned}$$

y  $\bar{X}$  es la media muestral de  $X_1, \dots, X_n$

*Dickey & Fuller* obtuvieron la distribución limite, cuando  $n \rightarrow \infty$ , del cociente  $\tau$ , la cual, es una distribución asintótica no estándar.

$$\hat{\tau} := \frac{\hat{\phi}_1^*}{\hat{SE}(\hat{\phi}_1^*)}$$

Bajo la hipótesis nula de raíz unitaria  $\phi_1^* = 0$ , la cual, es equivalente a probar  $H_0 : \phi_1 = 1$ . Los cuantiles para  $\alpha = .01, .05, .1$ , de la distribución límite de  $\hat{\tau}$ , son  $-3.43, -2.86, -2.57$  respectivamente. Una observación importante es, si en lugar de considerar la ecuación 2.8, se considera 2.9, y se cumple que  $\{\varepsilon_t\}$  es un ruido blanco con momentos finitos de orden ligeramente mayor a 2, entonces se prueba que la estadística  $\hat{\tau}$ , converge a un *Movimiento Browniano Estándar* cuando  $n \rightarrow \infty$ . En cualquier caso, no importa de que ecuación 2.8, 2.9, se parta, se usan simulaciones para obtener los valores críticos necesarios para realizar la prueba estadística.

El procedimiento realizado para el caso AR(1), se extiende al caso más general, cuando  $\{X_t\}$  sigue un proceso AR(p) con media  $\mu$ , dicho proceso satisface la siguiente ecuación

$$X_t - \mu = \phi_1(X_t - \mu) + \dots + \phi_p(X_{t-p} - \mu) + \varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

Este modelo se puede reescribir como

$$\begin{aligned} \nabla X_t &= X_t - X_{t-1} \\ &= \phi_0^* + \phi_1^* X_{t-1} + \phi_2^* \nabla X_{t-2} + \dots + \nabla \phi_p^* X_{t-p+1} + \varepsilon_t \end{aligned}$$

Donde  $\phi_0^* = \mu(1 - \phi_1 - \phi_2 - \dots - \phi_p)$ ,

$$\phi_1^* = \left( \sum_{i=1}^p \phi_i - 1 \right)$$

y  $\phi_j^* = - \sum_{i=j}^p \phi_i$  para  $j = 2, \dots, p$

Si el polinomio autoregresivo tiene una raíz unitaria en 1, entonces,  $0 = \phi(1) = -\phi_1^*$  y la serie diferenciada  $\{\nabla X_t\}$  es un proceso AR(p-1). Consecuentemente, probar la hipótesis nula de raíz unitaria en 1 del polinomio autoregresivo, es equivalente a probar  $\phi_1^* = 0$ . Como en el caso AR(1),  $\phi_1^*$  puede estimarse como el coeficiente de  $X_{t-1}$  al realizar la regresión por mínimos cuadrados ordinarios de  $\nabla X_t$  en  $1, X_{t-1}, \nabla X_{t-1}, \dots, \nabla X_{t-p+1}$ . Para  $n$  suficientemente grande, el estimador del coeficiente  $\tau$

$$\hat{\tau} := \frac{\hat{\phi}_1^*}{\widehat{SE}(\hat{\phi}_1^*)}$$

tiene la misma distribución límite que en el caso AR(1).

### 2.5.4. Modelos con estacionalidad

Convencionalmente, se pensaba que las series de tiempo consistían de una mezcla de tres componentes, una componente de *tendencia*  $P_t$ , otra componente *estacional*  $S_t$  y una componente irregular  $e_t$ . Suponiendo que estas componentes son aditivas, entonces se puede escribir a la serie de tiempo  $X_t$  como

$$X_t = P_t + S_t + e_t$$

En la literatura existen muchos métodos de descomposición para estimar cada una de las componentes anteriores, entre los métodos más conocidos se encuentran:

1. Método de regresión
2. Método de medias móviles
3. Suavizamiento exponencial
4. Holt-Winters

Sin embargo, los métodos previos, y los métodos tradicionales en general, se basan en el supuesto de que las componentes de *tendencia* y *estacionalidad* son deterministas e independientes entre ellas. Previamente se mencionaron algunas maneras de modelar la tendencia, ya sea determinista o estocástica, en el caso de tendencia estocástica, se hizo uso del operado diferencia, por lo cual, no es de sorprender que una herramienta muy importante para modelar la estacionalidad estocástica es la *diferencia estacional*.

A la diferencia estacional de periodo  $s$  de la serie  $\{X_t\}$  se le denota como  $\nabla_s X_t$  y se define como

$$\nabla_s X_t = X_t - X_{t-s}$$

Por ejemplo, para series de tiempo mensuales, se pueden considerar los cambios ocurridos de enero a enero, febrero a febrero, y así sucesivamente para todos los años que componen a la serie en estudio. De esta manera, para una serie de longitud  $n$ , la diferencia estacional producirá una serie de longitud  $n - s$ ; es decir,  $s$  observaciones se pierden debido a la diferencia estacional.

Considere un proceso generado por la siguiente ecuación

$$X_t = S_t + e_t$$

$$\text{con } S_t = S_{t-s} + \varepsilon_t$$

donde  $\{e_t\}$  y  $\{\varepsilon_t\}$  son procesos de ruido blanco independientes. En este ejemplo,  $\{S_t\}$  es una “*caminata aleatoria estacional*”, y si  $\sigma_\varepsilon < \sigma_e$ ,  $\{S_t\}$  describirá una componente estacional que cambia suavemente. Debido a que  $\{S_t\}$  no es estacionaria,  $\{X_t\}$  tampoco lo es, sin embargo, si se realiza la diferencia estacional sobre  $\{X_t\}$ , se tiene

$$\begin{aligned} \nabla_s X_t &= S_t - S_{t-s} + e_t - e_{t-s} \\ &= S_{t-s} + \varepsilon_t - S_{t-s} + e_t - e_{t-s} \\ &= \varepsilon_t + e_t - e_{t-s} \end{aligned}$$

La cual, es fácil ver que es una serie estacionaria y cuya función de autocorrelación corresponde a la de un modelo  $MA(1)_s$ .

Ahora, si modificamos un poco la composición de  $X_t$ , suponiendo que tiene un cambio estocástico en su tendencia, se tiene

$$X_t = \mu_t + S_t + e_t$$

$$\text{con}$$

$$S_t = S_{t-s} + \varepsilon_t$$

$$\text{y } \mu_t = \mu_{t-1} + \xi_t$$

donde  $\{e_t\}$ ,  $\{\varepsilon_t\}$  y  $\{\xi_t\}$  son procesos de ruido blanco mutuamente independientes. En este caso, se realiza una diferencia estacional y una diferencia ordinaria no estacional sobre  $X_t$ , dando como resultado

$$\begin{aligned} \nabla \nabla_s X_t &= \nabla(\mu_t + S_t + e_t - \mu_{t-s} - S_{t-s} - e_{t-s}) \\ &= \nabla(\mu_t + S_{t-s} + \varepsilon_t + e_t - \mu_{t-s} - S_{t-s} - e_{t-s}) \\ &= \nabla(\mu_t - \mu_{t-s} + \varepsilon_t + e_t - e_{t-s}) \\ &= (\xi_t + \varepsilon_t + e_t) - (\varepsilon_{t-1} + e_{t-1}) - (\xi_{t-s} + e_{t-s}) + e_{t-s} - 1 \end{aligned}$$

Este proceso es estacionaria y tiene únicamente una autocorrelación distinta de cero en los rezagos 1 y  $|s+1|$ , cuya estructura de correlación corresponde a la de un  $ARMA(0, 1) \times (0, 1)_s$ .

Supóngase que no se sabe que la serie  $\{X_t\}$  contiene la variación dada por los periodos estacionales y se ajusta un modelo ARIMA no estacional a la serie, es decir,

$$\phi_p(B)(1-B)^d X_t = \theta_q(B)b_t$$

Claramente la serie  $\{b_t\}$  no será un ruido blanco, pues su correlación tiene información no explicada en los periodos estacionales, entonces, sea

$$\rho_{j(s)} = \frac{\mathbb{E}[(b_{t-js} - \mu_b)(b_t - \mu_b)]}{\sigma_b^2}, \quad j = 1, \dots$$

la función de autocorrelación de  $\{b_t\}$ , la cual, representa a la relación entre los periodos estacionales no explicada. Se puede ver que esta relación aún presente en  $\{b_t\}$ , se puede modelar a través de un modelo ARMA de la siguiente manera.

$$\Phi_P(B^s)(1-B^s)^D b_t = \Theta_Q(B^s)\varepsilon_t \quad (2.10)$$

donde

$$\begin{aligned} \Phi_P(B^s) &= 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps} \\ \Theta_Q(B^s) &= 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs} \end{aligned}$$

son polinomios en  $B^s$  sin raíces en común.

## 2.6. Pronósticos

Los pronósticos tienen como objetivo predecir valores futuros de una serie de tiempo,  $X_{n+m}$   $m = 1, 2, \dots$ , basados en las observaciones obtenidas hasta el presente,  $X = \{X_n, X_{n-1}, \dots, X_1\}$ . Se supondrá que  $X_t$  es un proceso estacionario y los parámetros del modelo son conocidos. Se define al *mínimo error cuadrático medio de pronóstico* de  $X_{n+m}$  como

$$X_{n+m}^n = \mathbb{E}(X_{n+m} | X_n, \dots, X_1)$$

ya que la esperanza condicional minimiza al *error cuadrático medio*

$$\mathbb{E} [(X_{n+m} - g(x))^2]$$

donde  $g(x)$  es una función de las observaciones de  $X$ .

En particular, se pondrá especial interés en los pronósticos que son funciones lineales de las observaciones, es decir, predictores de la forma

$$X_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k X_k \quad (2.11)$$

donde  $\{\alpha_n\}$  son números reales. A los predictores de la forma 2.11, que además, minimizan al error cuadrático medio, se les conoce como los *mejores predictores lineales*, *BLP* por sus siglas en inglés. La siguiente propiedad de los *BLP* se basa en el teorema de proyección (ver anexo).

Dadas  $X_1, \dots, X_n$  observaciones, el mejor predictor lineal,  $X_{n+m}^n = \alpha_0 + \sum_{k=1}^n \alpha_k X_k$  de  $X_{n+m}$ , para  $m \geq 1$  se encuentra al resolver

$$\mathbb{E} [(X_{n+m} - X_{n+m}^n) X_k] = 0, \quad k = 0, 1, \dots, n$$

donde  $X_0 = 1$

A las ecuaciones inducidas por la ecuación anterior, se les conoce como *ecuaciones de predicción*, y se usan para encontrar los valores de los coeficientes  $\{\alpha_i\}_{i=1}^n$ . Si  $\mathbb{E}(X_t) = \mu$ , la primera ecuación, con  $k = 0$ , implica que

$$\begin{aligned} \mathbb{E} [(X_{n+m} - X_{n+m}^n)] &= 0 \\ \iff \mathbb{E}(X_{n+m}) - \mathbb{E}(X_{n+m}^n) &= 0 \\ \iff \mathbb{E}(X_{n+m}) = \mu = \mathbb{E}(X_{n+m}^n) \end{aligned}$$

Por lo cual, al sacar la esperanza de 2.11 se tiene que

$$\begin{aligned}\mathbb{E}(X_{n+m}^n) &= \alpha_0 + \sum_{k=1}^n \alpha_k \mu \\ \implies \mu &= \alpha_0 + \sum_{k=1}^n \alpha_k \mu \\ \implies \alpha_0 &= \mu - \sum_{k=1}^n \alpha_k \mu \\ \therefore \alpha_0 &= \mu \left( 1 - \sum_{k=1}^n \alpha_k \right)\end{aligned}$$

Entonces, para una serie de tiempo  $\{X_t\}$ , con media  $\mu$ , se tiene que la forma del *BLP* es

$$X_{n+m}^n = \mu + \sum_{k=1}^n \alpha_k (X_k - \mu)$$

Primero se va a considerar el pronóstico a *un paso*, es decir, dados  $X_1, \dots, X_n$ , se desea pronosticar el valor de la serie al siguiente punto del tiempo,  $X_{n+1}$ . Supongamos que  $\mathbb{E}(X_t) = 0$ , si esto no fuera así, bastaría con sustituir a  $X_t$  por  $X_t - \mu$ . Por lo cual, el *BLP* de  $X_{n+1}$  es

$$X_{n+1}^n = \sum_{k=1}^n \alpha_k X_{n+1-j}$$

Entonces, se debe satisfacer que

$$\mathbb{E} \left[ \left( X_{n+1} - \sum_{k=1}^n \alpha_k X_{n+1-j} \right) X_{n+1-k} \right] = 0, \quad k = 1, 2, \dots, n$$

o bien, que

$$\sum_{k=1}^n \alpha_j \gamma_{k-j} = \gamma_k, \quad k = 1, 2, \dots, n$$

La ecuación anterior al dividirla entre  $\gamma_0$ , es idéntica a la ecuación que induce a las ecuaciones de *Yule-Walker*,  $\rho_k = \sum_{j=1}^n \phi_{jn} \rho_{k-j}$ ,  $k = 1, 2, \dots, n$ , por lo cual, usaremos la notación  $\phi_{jn}$  en lugar de  $\alpha_j$ . Por lo tanto, las ecuaciones de predicción, se pueden escribir como

$$\Gamma_n \phi_n = \gamma_n$$

donde  $\Gamma_n = \{\gamma_{k-j}\}_{j,k=1}^n$ , matriz de dimensión  $n \times n$ ,  $\phi_n = (\phi_{1n}, \phi_{2n}, \dots, \phi_{mn})'$ , vector de dimensión  $n \times 1$ , y  $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)'$ , vector de dimensión  $n \times 1$ . La matriz  $\Gamma_n$ , es una matriz definida semipositiva, es importante notar que si  $\Gamma_n$  es singular, entonces existen una infinidad de soluciones, sin embargo, por el *teorema de proyección*,  $X_{n+1}^n$  es único.

Por lo cual, si  $\Gamma_n$  es no singular, los elementos de  $\phi_n$  son únicos y están dados por

$$\phi_n = \Gamma_n^{-1} \gamma_n$$

Una pequeña observación es que para los procesos ARMA, el hecho de que  $\sigma_\varepsilon^2 > 0$  y que  $\gamma_n \rightarrow 0$  cuando  $n \rightarrow \infty$ , son condiciones suficientes para garantizar que  $\Gamma_n$  sea definida positiva.

Volviendo a los pronósticos, el pronóstico de  $X_{n+1}^n$  se puede escribir en notación vectorial de la siguiente manera

$$X_{n+1}^n = \phi_n' \underline{X}$$

donde  $\underline{X} = (X_n, X_{n-1}, \dots, X_1)'$

Por lo tanto, el error cuadrático medio de predicción a un paso es

$$\begin{aligned} P_{n+1}^n &= \mathbb{E} [(X_{n+1} - X_{n+1}^n)^2] \\ &= \mathbb{E} [(X_{n+1} - \phi_n' \underline{X})^2] \\ &= \mathbb{E} [(X_{n+1} - \gamma_n' \Gamma_n^{-1} \underline{X})^2] \\ &= \mathbb{E} (X_{n+1}^2 - 2\gamma_n' \Gamma_n^{-1} X_{n+1} + \gamma_n' \Gamma_n^{-1} \underline{X} \underline{X}' \Gamma_n^{-1} \gamma_n) \\ &= \gamma_0 - 2\gamma_n' \Gamma_n^{-1} \gamma_n + \gamma_n' \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_n \\ &= \gamma_0 - \gamma_n' \Gamma_n^{-1} \gamma_n \\ \therefore P_{n+1}^n &= \gamma_0 - \gamma_n' \Gamma_n^{-1} \gamma_n \end{aligned}$$

Una observación importante es que debido a que

$$\phi_n = \Gamma_n^{-1} \gamma_n$$

es necesario obtener la inversa de la matriz  $\Gamma_n$ , la cual, depende de  $n$ . Por lo que para  $n$  muy grande, el obtener la inversa es computacionalmente costoso, aún con métodos numéricos.

Sin embargo, existen soluciones iterativas que no requieren calcular la inversa de  $\Gamma_n$ , por ejemplo, la solución recursiva desarrollada por *Levinson(1947)* y *Durbin(1960)*. Este algoritmo está desarrollado en la subsección dedicada a la *Función de Autocorrelación Parcial*.

Otro algoritmo muy utilizado, por ciertas características estadísticas deseadas, es el *Algoritmo de Inovaciones*, el cual, se explicará con detalle a continuación.

### 2.6.1. Algoritmo de Inovaciones

Este algoritmo recursivo, así como el *Durbin-Levinson*, es aplicable a todas las series de tiempo con segundo momento finito, además, no se requiere que el proceso sea estacionario y en ciertos casos, como se verá más adelante, se puede simplificar para ciertos procesos.

Supóngase entonces que  $\{X_t\}$  es una serie de tiempo con  $\mathbb{E}(X_t) = 0$ ,  $\mathbb{E}(X_t^2) < \infty \quad \forall t$  y se denotará a  $\mathbb{E}(X_i X_j)$  como  $k(i, j)$ . Además, se denotará al mejor predictor a un paso y su error cuadrático medio como

$$\hat{X}_n = \begin{cases} 0 & \text{si } n = 1 \\ X_n^{n-1} & \text{si } n = 2, 3, \dots \end{cases} \quad (2.12)$$

y

$$V_n = \mathbb{E}[(X_{n+1} - X_n^n)^2]$$

Se define a las *inovaciones*, o errores de pronóstico a un paso, como

$$U_n = X_n - \hat{X}_n$$

En términos vectoriales

$$U = (U_1, U_2, \dots, U_n)' \quad \text{y} \quad \hat{X} = (\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n)'$$

entonces,  $U$  se puede escribir como

$$U = AX$$

donde  $X = (X_1, X_2, \dots, X_n)'$  y  $A$  es de la forma

$$A = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ a_{11} & 1 & 0 & \dots & 0 \\ a_{22} & a_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n-1,n-1} & a_{n-1,n-2} & a_{n-1,n-3} & \dots & 1 \end{bmatrix}$$

Como todos los elementos de la diagonal de  $A$  son distintos de cero, entonces,  $A$  es no singular, con inversa  $C$  de la forma

$$C = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ \theta_{11} & 1 & 0 & \dots & 0 \\ \theta_{22} & \theta_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \theta_{n-1,n-1} & \theta_{n-1,n-2} & \theta_{n-1,n-3} & \dots & 1 \end{bmatrix}$$

Ahora, el vector de predicción  $\hat{X}$  a un paso,  $\hat{X} := (0, P_1 X_2, \dots, P_{n-1} X_n)'$  se puede escribir como

$$\begin{aligned} \hat{X} &= X - U \\ &= CU - U \\ &= \Theta_n(X - \hat{X}) \end{aligned}$$

donde  $\Theta_n = C$  con su diagonal en cero y  $X$  satisfacen por sí mismos que  $X = C(X - \hat{X})$ , por lo tanto 2.12 se puede reescribir como

$$\hat{X}_{n+1} = \begin{cases} 0 & \text{si } n = 0 \\ \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) & \text{si } n = 1, 2, \dots \end{cases}$$

El predictor a un paso  $\hat{X}_1, \hat{X}_2, \dots$  se puede calcular de forma recursiva una vez que los coeficientes  $\theta_{ij}$  se han determinado. El algoritmo que se enuncia a continuación, genera estos coeficientes, así como el error cuadrático medio  $V_i = \mathbb{E} [(X_{i+1} - \hat{X}_{i+1})^2]$ , comenzando por la covarianza  $k(i, j)$ .

**Algoritmo de inovaciones**

Los coeficientes  $\theta_{n1}, \dots, \theta_{nm}$  se pueden calcular de forma recursiva a partir de las ecuaciones

$$V_0 = k(1, 1)$$

$$\theta_{n, n-k} = V_k^{-1} \left( k(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k, k-j} \theta_{n, n-j} V_j \right), \quad 0 \leq k < n$$

y

$$V_n = k(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n, n-j}^2 V_j$$

Por lo tanto, se resuelve primero para  $V_0$ , y después, de forma recursiva, para

$$\begin{aligned} &\theta_{1,1}, V_1; \\ &\theta_{2,2}, \theta_{2,1}, V_2; \\ &\theta_{3,3}, \theta_{3,2}, \theta_{3,1}, V_3; \dots \end{aligned}$$

Es importante notar que el algoritmo recursivo de *Durbin-Levinson* ayuda a obtener los coeficientes de  $X_1, \dots, X_n$  en la representación  $\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$ , mientras que el algoritmo recursivo de *inovaciones*, da los coeficientes de  $(X_1 - \hat{X}_1, \dots, X_n - \hat{X}_n)$ , en la representación  $\hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j})$ .

La última representación tiene varias ventajas sobre la primera, derivadas del hecho de que las inovaciones son no correlacionadas. Además, este algoritmo se simplifica mucho en el caso en que el proceso es un ARMA(p,q).

Con la finalidad de tener claro el algoritmo, considérese el caso en que  $\{X_t\}$  es una serie de tiempo definida como

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

es decir, es un proceso MA(1). De las secciones pasadas, se sabe entonces que  $k(i, j) = 0 \quad \forall |i - j| > 1$ ,  $k(i, i) = \sigma^2(1 + \theta^2)$ , y  $k(i, i + 1) = \theta\sigma^2$ . Entonces, al aplicar el algoritmo de inovaciones, se obtiene que

$$\begin{aligned} V_0 &= k(1, 1) = \sigma^2(1 + \theta^2), \\ \theta_{n1} &= V_{n-1}^{-1} \theta \sigma^2, \\ \theta_{nj} &= 0, \quad 2 \leq j \leq n, \\ V_n &= [1 + \theta^2 - V_{n-1}^{-1} \theta^2 \sigma^2] \sigma^2 \end{aligned}$$

De tal forma que en el caso particular en que

$$X_t = \varepsilon_t - 0.9\varepsilon_{t-1}, \quad \{\varepsilon_t\} \sim WN(0, 1)$$

el error cuadrático medio  $V_n$  de  $\hat{X}_{n+1}$  y los coeficientes  $\theta_{nj}$ ,  $1 \leq j \leq n$ , en la representación del algoritmo de inovaciones

$$\hat{X}_{n+1} = \sum_{j=1}^n \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}) = \theta_{n1}(X_n - \hat{X}_n)$$

se encuentran de la siguiente manera:

$$\begin{aligned} V_0 &= 1.8100, \\ \theta_{11} &= -.4972, \quad V_1 = 1.3625, \\ \theta_{21} &= -.6606, \quad \theta_{22} = 0, \quad V_2 = 1.2155, \\ \theta_{31} &= -.7404, \quad \theta_{32} = 0, \quad \theta_{33} = 0, \quad V_3 = 1.1436, \\ \theta_{41} &= -.7870, \quad \theta_{42} = 0, \quad \theta_{43} = 0, \quad \theta_{44} = 0, \quad V_4 = 1.1017 \end{aligned}$$

Ahora, si se aplica el algoritmo de *Durbin-Levinson* al mismo problema, se tiene que el error cuadrático medio  $V_n$  de  $\hat{X}_{n+1}$  y los coeficientes  $\phi_{nj}$ ,  $1 \leq j \leq n$ , en la representación

$$\hat{X}_{n+1} = \sum_{j=1}^n \phi_{nj} X_{n+1-j}$$

son los siguientes:

$$\begin{aligned}
V_0 &= 1.8100, \\
\theta_{11} &= -.4972, \quad V_1 = 1.3625, \\
\theta_{21} &= -.6606, \quad \theta_{22} = -.3285, \quad V_2 = 1.2155, \\
\theta_{31} &= -.7404, \quad \theta_{32} = -.4892, \quad \theta_{33} = -.2433, \quad V_3 = 1.1436, \\
\theta_{41} &= -.7870, \quad \theta_{42} = -.5828, \quad \theta_{43} = -.3850, \quad \theta_{44} = -.1914, \quad V_4 = 1.1017
\end{aligned}$$

Es importante hacer notar que conforme  $n$  aumenta,  $V_n$  se aproxima a la varianza del ruido blanco y  $\theta_{n1}$  se aproxima a  $\theta$ . Esto se cumple para cualquier proceso MA(1) con  $|\theta| < 1$ . El algoritmo de inovaciones es particularmente útil para pronosticar procesos MA(q), pues  $\theta_{nj} = 0 \quad \forall n - j > q$ . Para el caso AR(p), el algoritmo de *Durbin-Levinson* es, usualmente, más conveniente, pues  $\phi_{nj} = 0 \quad \forall n - j > p$ .

### 2.6.1.1. Cálculo recursivo de pronósticos a $h$ -pasos

Para realizar pronósticos a  $h$ -pasos, se usará el resultado

$$P_n(X_{n+k} - P_{n+k-1}X_{n+k}) = 0, \quad k \geq 1 \quad (2.13)$$

Lo anterior se sigue de las siguientes dos propiedades de  $P_n X_{n+h}$

1.  $\mathbb{E}[(\text{error} \times (\text{variable predicción}))] = 0$
2.  $\mathbb{E}[(X_{n+k} - P_{n+k-1}X_{n+k})X_{n+j-1}] = 0, \quad j = 1, \dots$

Por lo cual

$$\begin{aligned}
P_n X_{n+h} &= P_n P_{n+h-1} X_{n+h} \\
&= P_n \hat{X}_{n+h} \\
&= P_n \left( \sum_{j=1}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \right)
\end{aligned}$$

Aplicando 2.13 y usando la linealidad de  $P_n$ , se tiene que

$$P_n X_{n+h} = \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (2.14)$$

donde los coeficientes  $\theta_{nj}$  se determinan como en el caso a un paso por el algoritmo de inovaciones. Además, el error cuadrático medio se puede expresar como

$$\begin{aligned} \mathbb{E} \left[ (X_{n+h} - P_n X_{n+h})^2 \right] &= \\ &= \mathbb{E} (X_{n+h}^2) - [\mathbb{E} (P_n X_{n+h})]^2 \\ &= k(n+h, n+h) - \sum_{j=h}^{n+h-1} \theta_{n+h-1,j}^2 V_{n+h-j-1} \end{aligned}$$

### 2.6.2. Pronósticos de modelos ARMA

Hasta el momento, se ha visto como realizar pronósticos sobre cualquier serie de tiempo en general, tanto para las *ecuaciones generales de predicción*, como para el *algoritmo de inovaciones*. Sin embargo, por ejemplo, las *ecuaciones generales de predicción* proveen de poca información de los modelos ARMA generales. Existen diversas maneras de expresar estos pronósticos, y cada una de estas, permite entender la estructura especial de los pronósticos ARMA.

A partir de ahora, se supondrá que  $\{X_t\}$  es un proceso ARMA causal e invertible, con  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$ . Si  $\{X_t\}$  no tiene media cero, bastará con considerar en su lugar a  $\{X_t - \mu\}$ . Recordar que en el caso general, se definió a  $X_{n+m}^n$  como el pronóstico de  $X_{n+m}$  con error cuadrático medio más pequeño, basado en las observaciones  $\{X_1, X_2, \dots, X_n\}$ , es decir

$$X_{n+m}^n = \mathbb{E}(X_{n+m} | X_n, \dots, X_1)$$

Para los modelos ARMA es más sencillo calcular el pronóstico de  $X_{n+m}$  suponiendo que se tiene la historia completa del proceso  $\{X_t\}$ . Se denotará en este caso al pronóstico de  $X_{n+m}$  basado en su pasado “infinito” como

$$\tilde{X}_{n+m} = \mathbb{E}(X_{n+m} | X_n, X_{n-1}, \dots)$$

El punto central radica en que para muestras grandes,  $\tilde{X}_{n+m}$  será una buena aproximación de  $X_{n+m}^n$ . Expresando a  $X_{n+m}$  en su forma causal e invertible se tiene

$$X_{n+m} = \sum_{j=0}^{\infty} \psi_j \varepsilon_{n+m-j}, \quad \psi_0 = 1 \quad (2.15)$$

$$\varepsilon_{n+m} = \sum_{j=0}^{\infty} \pi_j X_{n+m-j}, \quad \pi_0 = 1 \quad (2.16)$$

Aplicando esperanza condicional a 2.15, se tiene

$$\begin{aligned} \tilde{X}_{n+m} &= \mathbb{E}(X_{n+m} | X_n, \dots, X_1) \\ &= \sum_{j=0}^{\infty} \psi_j \underbrace{\mathbb{E}(\varepsilon_{n+m-j} | X_n, \dots, X_1)}_{\text{No depende de } X_n, \dots, X_1} = \begin{cases} 0 & \text{si } t > n \\ \varepsilon_t & \text{si } t \leq n \end{cases} \\ &= \sum_{j=m}^{\infty} \psi_j \varepsilon_{n+m-j} \end{aligned} \quad (2.17)$$

De forma similar, al obtener la esperanza condicional de 2.16 se tiene

$$\begin{aligned} 0 &= \tilde{X}_{n+m} + \sum_{j=1}^{\infty} \pi_j \tilde{X}_{n+m-j} \\ \implies \tilde{X}_{n+m} &= - \sum_{j=1}^{\infty} \pi_j \tilde{X}_{n+m-j} \end{aligned}$$

Es importante notar que de 1 a  $m-1$  son “pronósticos” y de  $m$  a  $\infty$  son las observaciones

$$\therefore \tilde{X}_{n+m} = \sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j} - \sum_{j=m}^{\infty} \pi_j X_{n+m-j} \quad (2.18)$$

si  $m = 1$

$$\tilde{X}_{n+1} = - \sum_{j=1}^{\infty} \pi_j X_{n+1-j}$$

si  $m = 2$

$$\tilde{X}_{n+2} = -\pi_1 \tilde{X}_{n+1} - \sum_{j=2}^{\infty} \pi_j X_{n+2-j}$$

⋮

Por otro lado, de 2.17 se tiene

$$\begin{aligned}
 \tilde{X}_{n+m} &= \sum_{j=m}^{\infty} \psi_j \varepsilon_{n+m-j} \\
 \text{sumando } 0 &= \sum_{j=0}^{m-1} \psi_j \varepsilon_{n+m-j} - \sum_{j=0}^{m-1} \psi_j \varepsilon_{n+m-j} \\
 \implies \tilde{X}_{n+m} &= \underbrace{\sum_{j=0}^{\infty} \psi_j \varepsilon_{n+m-j}}_{X_{n+m}} - \sum_{j=0}^{m-1} \psi_j \varepsilon_{n+m-j} \\
 \implies X_{n+m} - \tilde{X}_{n+m} &= \sum_{j=0}^{m-1} \psi_j \varepsilon_{n+m-j}
 \end{aligned}$$

Por lo tanto, el error cuadrático medio de pronóstico, se puede escribir como

$$P_{n+m}^n = \mathbb{E} \left[ (X_{n+m} - \tilde{X}_{n+m})^2 \right] = \sigma_{\varepsilon}^2 \sum_{j=0}^{m-1} \psi_j^2 \quad (2.19)$$

Una observación importante, es que para un tamaño de muestra fijo  $n$ , los errores de pronóstico están correlacionados. Es decir, para  $k \geq 1$

$$\mathbb{E} \left[ (X_{n+m} - \tilde{X}_{n+m}) (X_{n+m+k} - \tilde{X}_{n+m+k}) \right] = \sigma_{\varepsilon}^2 \sum_{j=0}^{m-1} \psi_j \psi_{j+k}$$

Por lo tanto, se concluye que cuando  $n$  es pequeña, las ecuaciones generales de pronóstico se pueden usar fácilmente, en cambio, si  $n$  es grande, se usará 2.18 truncado, pues únicamente se tiene información de  $X_1, \dots, X_n$  disponible. Se puede truncar 2.18 al fijar

$$\sum_{j=n+m}^{\infty} \pi_j X_{n+m-j} = 0$$

Por lo tanto, el predictor truncado sería

$$\tilde{X}_{n+m}^n = \sum_{j=1}^{m-1} \pi_j \tilde{X}_{n+m-j}^n - \sum_{j=m}^{n+m-1} \pi_j X_{n+m-j}$$

el cual se calcula recursivamente, mientras que el error cuadrático medio de pronóstico se aproxima con 2.19.

Para poder tener una idea de la precisión de los pronósticos, típicamente se calculan intervalos de confianza a lo largo de los valores pronosticados. En general, los intervalos de pronóstico de  $(1 - \alpha) \times 100\%$  de confianza son de la forma

$$X_{n+m}^n \pm C_{\frac{\alpha}{2}} \sqrt{P_{n+m}^n}$$

donde  $C_{\frac{\alpha}{2}}$  se elige de tal manera que se tenga el grado de confianza deseado. De tal manera, que si por ejemplo, el proceso fuera gaussiano, entonces se escogería a  $C_{\frac{\alpha}{2}} = 2$  para tener intervalos aproximados del 95%.

### 2.6.2.1. Algoritmo de inovaciones para procesos ARMA

Además de los dos criterios previos para obtener pronósticos de un proceso ARMA, se puede hacer uso del algoritmo de inovaciones descrito previamente. Para el caso de un proceso ARMA causal

$$\phi(B)X_t = \theta(B)\varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2)$$

es posible simplificar dicho algoritmo de forma drástica. Lo anterior se logra al aplicar el algoritmo no sobre el proceso  $\{X_t\}$ , sino al proceso transformado

$$\begin{cases} W_t = \sigma^{-1}X_t, & t = 1, 2, \dots, m \\ W_t = \sigma^{-1}\phi(B)X_t, & t > m \end{cases} \quad (2.20)$$

donde  $m = \max(p, q)$ . Además, se define a  $\theta_0 := 1$  y  $\theta_j := 0$  para  $j > q$ , y se supone que  $p \geq 1$  y  $q \geq 1$ .

Se demuestra que la autocovarianza  $k(i, j) = \mathbb{E}(W_i W_j)$ ,  $m \leq i, j \leq m$  de  $\{W_t\}$  se puede expresar como

$$k(i, j) = \begin{cases} \sigma^{-2} \gamma_X(i-j), & 1 \leq i, j \leq m \\ \sigma^{-2} \left[ \gamma_X(i-j) - \sum_{r=1}^p \phi_r \gamma_X(r - |i-j|) \right], & \min(i, j) \leq m \leq \max(i, j) \leq 2m \\ \sum_{r=0}^q \theta_r \theta_{r+|i-j|}, & \min(i, j) > m \\ 0, & \text{e.o.c.} \end{cases} \quad (2.21)$$

Entonces, al aplicar el algoritmo de inovaciones sobre el proceso  $\{W_t\}$  se obtiene

$$\begin{cases} \hat{W}_{n+1} = \sum_{j=1}^n \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & 1 \leq n < m \\ \hat{W}_{n+1} = \sum_{j=1}^q \theta_{nj} (W_{n+1-j} - \hat{W}_{n+1-j}), & n \geq m \end{cases} \quad (2.22)$$

donde los coeficientes  $\theta_{nj}$ , así como el *error cuadrático medio de pronóstico*,  $r_n = \mathbb{E} \left[ (W_{n+1} - \hat{W}_{n+1})^2 \right]$ , se obtienen de formar recursiva a partir del algoritmo de inovaciones.

Lo que se obtiene a partir de los predictores  $\hat{W}_{n+1}$ , es la ausencia de  $\theta_{nj}$  cuando  $n \geq m$  y  $j > q$ . Esto es consecuencia del algoritmo de inovaciones, y del hecho de que  $k(r, s) = 0$  si  $r > m$  y  $|r - s| > q$ .

Es muy importante notar que el proceso transformado  $\{W_t\}$ , debido a como se definió, permite que cada  $X_n$ ,  $n \geq 1$ , se pueda escribir como combinación lineal de  $W_j$ ,  $1 \leq j \leq n$ , y que por consecuencia, cada  $W_n$ ,  $n \geq 1$ , se pueda escribir como combinación lineal de  $X_j$ ,  $1 \leq j \leq n$ . Esto implica que el mejor predictor lineal de cualquier variable aleatoria  $Y$ , en términos de  $\{1, X_1, \dots, X_n\}$ , es igual que el mejor predictor lineal de  $Y$ , en términos de  $\{1, W_1, \dots, W_n\}$ .

Se denotará al mejor predictor lineal de  $Y$ , como  $P_n Y$ , en particular, el predictor a un paso de  $W_{n+1}$  y  $X_{n+1}$  está dado por

$$\begin{aligned} \hat{W}_{n+1} &= P_n W_{n+1} \\ &\text{y} \\ \hat{X}_{n+1} &= P_n X_{n+1} \end{aligned}$$

Haciendo uso de la linealidad de  $P_n$ , y el como se definió al proceso  $\{W_t\}$ , se tiene

$$\begin{cases} \hat{W}_t = \sigma^{-1} \hat{X}_t, & t = 1, 2, \dots, m \\ \hat{W}_t = \sigma^{-1} (\hat{X}_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}), & t > m \end{cases}$$

lo cual prueba que  $X_t - \hat{X}_t = \sigma (W_t - \hat{W}_t) \quad \forall t \geq 1$

Ahora, al sustituir  $(W_j - \hat{W}_j)$  por  $\sigma^{-1}(X_j - \hat{X}_j)$  en 2.21, y sustituir lo obtenido en 2.22 se obtiene finalmente el pronóstico a un paso de  $X_{n+1}$

$$\hat{X}_{n+1} = \begin{cases} \sum_{j=1}^n \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \leq n < m \\ \sum_{j=1}^{p+1} \phi_j X_{n+1-j} + \sum_{j=1}^q \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j}), & n \geq m \end{cases} \quad (2.23)$$

y además, el error cuadrático medio de predicción asociado es  $\mathbb{E} \left[ (X_{n+1} - \hat{X}_{n+1})^2 \right] = \sigma^2 \mathbb{E} \left[ (W_{n+1} - \hat{W}_{n+1})^2 \right] = \sigma^2 r_n$ , donde  $\theta_{nj}$  y  $r_n$  se encuentran a través del algoritmo de inovaciones con  $k(i, j)$  como en 2.21. Además, la ecuación 2.23, permite obtener los pronósticos a un paso  $\hat{X}_2, \hat{X}_3, \dots$  de forma recursiva habiendo calculado  $\hat{X}_1$ .

Por último, se prueba que si  $\{X_t\}$  es un proceso invertible, entonces, cuando  $n \rightarrow \infty$  ocurre que

$$\begin{aligned} \mathbb{E} \left[ (X_n - \hat{X}_n - \varepsilon_n)^2 \right] &\rightarrow 0, \\ \theta_{nj} &\rightarrow \theta_j, \quad j = 1, \dots, q \\ r_n &\rightarrow 1 \end{aligned}$$

El trabajo desarrollado previamente, tuvo como objetivo exponer como obtener el pronóstico a un paso de  $X_{n+1}$  de un proceso ARMA a través del algoritmo de inovaciones, en caso de que se esté interesado en obtener de nuevo el pronóstico, pero ahora a  $h$  pasos, de un proceso ARMA, será necesario recordar de nuevo que  $P_n Y$ , el mejor predictor lineal de  $Y$ , en términos de  $X_1, \dots, X_n$ , es también el mejor predictor lineal de  $Y$ , pero en términos de  $W_1, \dots, W_n$ . Además, haciendo uso de la ecuación general para el pronóstico a  $h$  pasos 2.14, obtenida mediante el algoritmo de inovaciones, se tiene que

$$\begin{aligned} P_n W_{n+h} &= \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (W_{n+h-j} - \hat{W}_{n+h-j}) \\ &= \sigma^2 \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}) \end{aligned}$$

Usando este resultado y aplicando el operador  $P_n$ , a cada lado de la ecuación 2.20, se concluye que el pronóstico a  $h$  pasos  $P_n X_{n+h}$  satisface

$$P_n X_{n+h} = \begin{cases} \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), & 1 \leq h \leq m-n \\ \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^{n+h-1} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j}), & h > m-n \end{cases}$$

Si, como casi siempre ocurre,  $n > m = \max(p, q)$ , entonces  $\forall h \geq 1$

$$P_n X_{n+h} = \sum_{i=1}^p \phi_i P_n X_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1, j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (2.24)$$

Una vez obtenidos los pronósticos  $\hat{X}_1, \dots, \hat{X}_n$  a partir de 2.23, es inmediato obtener para  $n$  fija, los pronósticos  $P_n X_{n+1}, P_n X_{n+2}, \dots$

Además, se prueba que el error cuadrático medio de  $P_n X_{n+h}$ , está dado por la siguiente ecuación

$$\sigma_n^2(h) := \mathbb{E} \left[ (X_{n+h} - P_n X_{n+h})^2 \right] = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 V_{n+h-j-1} \quad (2.25)$$

donde los coeficientes  $\chi_j$ , se obtienen de forma recursiva, a través de las ecuaciones  $\chi_0 = 1$  y

$$\chi_j = \sum_{k=1}^{\min(p, j)} \phi_k, \quad 1, 2, \dots \quad (2.26)$$

### 2.6.3. Pronósticos de modelos ARIMA

Recordar que un proceso  $\{X_t\}$ , es un proceso  $ARIMA(p, d, q)$ , si satisface la siguiente ecuación

$$\phi^*(B)X_t \equiv \phi(B)(1-B)^d X_t = \theta(B)\varepsilon_t, \quad \{\varepsilon_t\} \sim WN(0, \sigma^2) \quad (2.27)$$

Si  $d \geq 1$ , el primero y segundo momento  $\mathbb{E}(X_t)$  y  $\mathbb{E}(X_{t+h}X_t)$ , no están determinados por la ecuación en diferencias 2.27, por lo cual, no se puede determinar el mejor predictor lineal de  $\{X_t\}$  sin mayores supuestos.

Para ilustrar la manera de obtener pronósticos de un modelo ARIMA, considérese un proceso  $\{X_t\}$  ARIMA, además, se supondrá que  $\{Y_t\}$  es un proceso ARMA(p,q) causal y  $X_0$  es cualquier variable aleatoria.

Se define a  $X_t = X_0 + \sum_{j=1}^t Y_j$ ,  $t = 1, 2, \dots$ , entonces,  $\{X_t\}$  es un proceso ARIMA(p,1,q), con media  $\mathbb{E}(X_t) = \mathbb{E}(X_0)$  y autocovarianza  $\mathbb{E}(X_{t+h}X_t) - [\mathbb{E}(X_0)]^2$ , la cual, depende de la  $Var(X_0)$  y de la  $Cov(X_0, Y_j)$ ,  $j = 1, 2, \dots$

El mejor predictor lineal de  $X_{n+1}$  basado en  $\{1, X_0, X_1, \dots, X_n\}$  es también el mejor predictor lineal, pero basado en el conjunto  $\{1, X_0, Y_1, \dots, Y_n\}$ , pues cada *combinación lineal* del segundo conjunto, es una *combinación lineal* del primero y viceversa. Por lo tanto, denotando al mejor predictor lineal en términos de cualquiera de los dos conjuntos como  $P_n$  y usando la linealidad de  $P_n$ , se puede escribir como

$$\begin{aligned} P_n X_{n+1} &= P_n \left( \underbrace{X_0 + Y_1 + \dots + Y_n}_{X_n} + Y_{n+1} \right) \\ &= P_n (X_n + Y_{n+1}) \\ &= X_n + P_n Y_{n+1} \end{aligned}$$

Ahora, para poder evaluar  $P_n Y_{n+1}$ , es necesario saber  $\mathbb{E}(X_0 Y_j)$ ,  $j = 1, 2, \dots, n+1$  y  $\mathbb{E}(X_0^2)$ . Sin embargo, si se añade el supuesto de que  $X_0$  es no correlacionado con  $\{Y_t\}$ , entonces  $P_n Y_{n+1}$  es el mismo que el mejor predictor lineal  $\hat{Y}_{n+1}$  de  $Y_{n+1}$  en términos de  $\{Y_1, \dots, Y_n\}$ , el cual, se puede calcular utilizando cualquiera de los métodos descritos para el caso ARMA. Por lo tanto, en este caso, el supuesto de que  $X_0$  no esté correlacionado con  $\{Y_t\}$ , es suficiente para determinar al mejor predictor lineal  $P_n X_{n+1}$ .

En el caso general, supóngase que el proceso observado  $\{X_t\}$ , satisface la ecuación en diferencias

$$(1 - B)^d X_t = Y_t, \quad t = 1, 2, \dots$$

donde  $\{Y_t\}$  es un proceso ARMA(p,q) causal, y el vector aleatorio  $(X_{1-d}, \dots, X_0)$  es no correlacionado con  $Y_t$ ,  $t > 0$ . Se demuestra que la ecuación en diferencias 2.27, se puede escribir como

$$X_t = Y_t - \sum_{j=1}^d \binom{d}{j} (-1)^j X_{t-j}, \quad t = 1, 2, \dots \quad (2.28)$$

Será conveniente reetiquetar el *eje-x*, si es necesario, para poder suponer que se observó a  $X_{1-d}, X_{2-d}, \dots, X_n$ , (entonces las observaciones de  $\{Y_t\}$  son  $Y_1, Y_2, \dots, Y_n$ ), como es usual, se denotará al mejor predictor lineal, en términos de las observaciones hasta tiempo  $n$ , como  $P_n$ , en este caso, en términos de  $1, X_{1-d}, \dots, X_n$ , o equivalentemente  $1, X_{1-d}, \dots, X_0, Y_1, \dots, Y_n$ .

El objetivo es determinar el mejor predictor lineal  $P_n X_{n+h}$ , para esto, al aplicar el operador  $P_n$  a la ecuación 2.28, se tiene

$$P_n X_{n+h} = P_n Y_{n+h} - \sum_{j=1}^d \binom{d}{j} (-1)^j P_n X_{n+h-j}, \quad t = 1, 2, \dots \quad (2.29)$$

Ahora, por el supuesto de que el vector aleatorio  $(X_{1-d}, \dots, X_0)$  es no correlacionado con  $Y_t$ ,  $t > 0$ , podemos obtener el mejor predictor lineal  $P_n Y_{n+h}$  de  $Y_{n+h}$  en términos de  $\{1, Y_1, \dots, Y_n\}$ , el cual, se calcula a partir de cualquiera de los métodos mostrados para el caso ARMA.

El pronóstico  $P_n X_{n+1}$ , se obtiene de forma directa haciendo uso de 2.29, al notar que  $P_n X_{n+1-j} = X_{n+1-j}$ ,  $\forall t \geq 1$ . Una vez obtenido  $P_n X_{n+1}$ , se puede obtener ahora a  $P_n X_{n+2}$  a partir de 2.29, de esta manera, los pronósticos  $P_n X_{n+3}, \dots$  se pueden obtener de forma recursiva.

Por otro lado, para poder obtener el error cuadrático medio de predicción, es conveniente expresar al pronóstico  $P_n Y_{n+h}$  en términos de  $\{X_t\}$ . Con este fin, para  $n \geq 0$ , se denotará al pronóstico a un paso como  $\hat{Y}_{n+1} = P_n Y_{n+1}$  y  $\hat{X}_{n+1} = P_n X_{n+1}$ . Entonces al restar de 2.28, 2.29, y haciendo uso de la observación de que para  $\hat{X}_{n+1}$ , se cumple que  $\hat{X}_{n+1-j} = X_{n+1-j}$ ,  $\forall t \geq 1$ , se tiene que

$$X_{n+1} - \hat{X}_{n+1} = Y_{n+1} - \hat{Y}_{n+1}$$

y por la ecuación 2.24, si  $n > m = \max(p, q)$  y  $h \geq 1$ , se puede expresar al pronóstico  $P_n Y_{n+h}$  como

$$P_n Y_{n+h} = \sum_{i=1}^p \phi_i P_n Y_{n+h-i} + \sum_{j=h}^q \theta_{n+h-1, j} (X_{n+h-j} - \hat{X}_{n+h-j}) \quad (2.30)$$

Sea  $\phi^*(z) = (1-z)^d \phi(z)$ , al sustituir 2.29 en 2.30, se obtiene que

$$P_n X_{n+h} = \sum_{j=1}^{p+d} \phi_j^* P_n X_{n+h-j} + \sum_{j=h}^q \theta_{n+h-1, j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

la cual, es una expresión equivalente a la ecuación 2.24, la cual corresponde a la ecuación de predicción a  $h$  pasos del proceso ARMA.

Por otro lado, como en el caso ARMA, el error cuadrático medio del predictor a  $h$  pasos es

$$\begin{aligned} \sigma^2(h) &= \mathbb{E} \left[ (X_{n+h} - P_n X_{n+h})^2 \right] \\ &= \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 V_{n+h-j-1} \end{aligned}$$

donde

$$\begin{aligned}\theta_{n,0} &= 1 \\ \chi(z) &= \sum_{r=0}^{\infty} \chi_r z^r = \left(1 - \phi_1^* z - \dots - \phi_{p+d}^* z^{p+d}\right)^{-1} \\ \text{y } V_{n+h-j-1} &= \mathbb{E} \left[ (X_{n+h-j} - \hat{X}_{n+h-j})^2 \right] = \mathbb{E} \left[ (Y_{n+h-j} - \hat{Y}_{n+h-j})^2 \right]\end{aligned}$$

Además, los coeficientes  $\chi_j$ , se pueden obtener de forma recursiva usando 2.26, usando  $\phi_j^*$  en lugar de  $\phi_j$ . Para  $n$  suficientemente grande, se puede aproximar a  $\sigma_n^2(h)$  mediante  $\sum_{j=0}^{h-1} \psi_j^2 \sigma^2$ , donde  $\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\psi(z)}$

#### 2.6.4. Pronósticos de modelos SARIMA

Pronosticar procesos SARIMA es completamente análogo al caso en el que se buscaba pronosticar procesos ARIMA, pues únicamente se necesita expandir el operador  $(1-B)^d(1-B^s)^D$  en potencias de  $B$ , después, se reagrupa la ecuación  $(1-B)^d(1-B^s)^D X_t = Y_t$  y fijando  $t = n+h$ , se obtiene la ecuación análoga a 2.28

$$X_{n+h} = Y_{n+h} - \sum_{j=1}^{d+Ds} \binom{d+Ds}{j} (-1)^j X_{n+h-j} \quad (2.31)$$

$$\text{Sea } a_j = - \binom{d+Ds}{j} (-1)^j$$

Además, al igual que en el caso ARIMA, si se supone que los primeros  $d+Ds$  observaciones diferenciadas,  $X_{-d-Ds+1}, \dots, X_0$ , no están correlacionadas con  $\{Y_t\}$ , se puede determinar al mejor predictor lineal  $P_n X_{n+h}$  de  $X_{n+h}$  basado en  $\{1, X_{-d-Ds+1}, \dots, X_n\}$ . Al aplicar el operador  $P_n$  a ambos lados de 2.31, se obtiene

$$P_n X_{n+h} = P_n Y_{n+h} + \sum_{j=1}^{d+Ds} a_j P_n X_{n+h-j} \quad (2.32)$$

El primer sumando del lado derecho de la ecuación 2.32, no es más que el mejor predictor lineal del proceso ARMA  $\{Y_t\}$  en términos de  $\{1, Y_1, \dots, Y_n\}$ , el cual, se puede

calcular con cualquiera de los métodos expuestos anteriormente. Por lo tanto, los predictores  $P_n X_{n+h}$  se pueden calcular de forma recursiva para  $h = 1, 2, \dots$  al tener la observación de que  $P_n X_{n+1-j} = X_{n+1-j} \quad \forall j \geq 1$ .

Por otro lado, el error cuadrático medio de predicción es

$$\sigma_n^2(h) = \mathbb{E} \left[ (X_{n+h} - P_n X_{n+h})^2 \right] = \sum_{j=0}^{h-1} \left( \sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 V_{n+h-j-1}$$

donde  $\theta_{nj}$  y  $V_n$  se obtienen a través del algoritmo de inovaciones, aplicado a la serie diferenciada  $\{Y_t\}$  y además

$$\chi(z) = \sum_{r=0}^{\infty} \chi_r z^r = \frac{1}{\phi(z)\Phi(z^s)(1-z)^d(1-z^s)^D}, \quad |z| < 1$$

Para  $n$  suficientemente grande, y si  $\theta(z)\Theta(z^s) \neq 0 \quad \forall |z| \leq 1$ , entonces se puede aproximar al *ECM* de predicción mediante

$$\sigma_n^2(h) = \sum_{j=0}^{h-1} \psi_j^2 \sigma^2$$

donde

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)\Theta(z^s)}{\phi(z)\Phi(z^s)(1-z)^d(1-z^s)^D}, \quad |z| < 1$$

En general, los softwares estadísticos utilizan la expresión anterior para aproximar el valor del *ECM* de predicción.

### 2.6.5. Modelo de espacio de estados para pronósticos de series de tiempo

La teoría desarrollada en las secciones anteriores, con los distintos criterios para lograr hacer pronósticos sobre series de tiempo observadas, no es usada, hoy en día, por todos los softwares estadísticos. Por ejemplo, en R a pesar de que está implementado el algoritmo de inovaciones, esta implementación no tiene la flexibilidad suficiente para aplicarse a un proceso SARIMA. De hecho, R utiliza por default, la representación de *espacio de*

*estados* de una serie de tiempo para realizar los pronósticos sobre esta, motivo por el cual, se desarrollará de forma breve los conceptos necesarios para tener un entendimiento general de los modelos de *espacio de estados*.

Desde no hace mucho tiempo, la representación de espacio de estados, así como las recursiones *Kalman* asociadas, han tenido un impacto profundo en el análisis de series de tiempo y sobre muchas otras áreas relacionadas. Esto se debe a que se trata de un modelo muy general que engloba a toda una clase de casos especiales de particular interés, de la misma manera que los modelos de regresión lineal.

El modelo de *espacio de estados*, o modelo *lineal dinámico*, fue introducido por Kalman(1960) y Kalman-Bucy(1961), cuyo desarrollo inicial fué como un modelo de uso preliminar para investigación sobre temas aeroespaciales, sin embargo, hoy en día ha sido ampliamente usado para analizar datos económicos, médicos, etc.

Se dice que una serie de tiempo se puede representar en su forma lineal de *espacio de estados*, si la serie  $\{Y_t\}$  satisface la ecuación

$$Y_t = G_t X_t + W_t, \quad t = 1, 2, \dots \quad (2.33)$$

a la ecuación 2.33 se le conoce como **ecuación de observaciones**, donde

$$X_{t+1} = F_t X_t + V_t, \quad t = 1, 2, \dots \quad (2.34)$$

a esta última ecuación se le conoce como **ecuación de estados**, es claro que la expresión anterior hace uso de un vector autoregresivo de orden uno.

La ecuación 2.34, se puede interpretar como la evaluación del estado  $X_t$  de un sistema al tiempo  $t$ , donde  $X_t$  es un vector de dimensión  $v \times 1$ , en términos de una sucesión de matrices  $F_1, F_2, \dots$  de dimensión  $v \times v$  y de una sucesión de vectores aleatorios  $X_1, V_1, V_2, \dots$  de dimensión  $v \times 1$ , además, se supone que el proceso comienza con un vector normal  $X_0$ , el cual tiene media  $\mu_0$  y matriz de varianzas y covarianzas  $\phi_0$  de dimensión  $v \times v$ .

Además, se supone que no se observa al vector de estados  $\mathbf{X}_t$  de forma directa, por lo cual, la ecuación 2.33, define una sucesión de observaciones  $Y_t$ , las cuales, son producto de aplicar una transformación lineal a  $X_t$  y añadirle un vector de *ruido aleatorio*,  $W_t$ ,  $t = 1, 2, \dots$  donde  $G_t$  es una matriz de dimensión  $w \times v$ , llamada *matriz de medición* o *matriz de observación*

Para poder estudiar ciertas propiedades de la *ecuación de estados* y de la *ecuación de observaciones*, se deben cumplir los siguientes supuestos:

1.  $F_1, F_2, \dots$ , es una sucesión de matrices especificadas
2.  $G_1, G_2, \dots$ , es una sucesión de matrices, de dimensión  $w \times v$ , especificadas
3.  $\left\{ X_t, \left( V_t', W_t' \right)', \quad t = 1, 2, \dots \right\}$  es una sucesión de vectores aleatorios ortogonales con segundo momento finito.

Recordar que dos vectores aleatorios  $X, Y$ , son ortogonales,  $X \perp Y$ , si la matriz  $\mathbb{E}(XY') = 0$

4.  $\mathbb{E}(V_t) = 0$  y  $\mathbb{E}(W_t) = 0$  para toda  $t$
5.  $\mathbb{E}(V_t V_t') = Q_t, \mathbb{E}(W_t W_t') = R_t$  y  $\mathbb{E}(V_t W_t') = S_t$  Donde  $\{Q_t\}, \{R_t\}$  y  $\{S_t\}$ , son sucesiones de matrices especificadas de orden  $v \times v, w \times w$  y  $v \times w$  respectivamente

Entre las grandes ventajas de esta formulación del modelo de espacio de estados, es que provee diversas formas especiales, las cuales, se pueden desarrollar a través de múltiples versiones de la matriz  $G_t$ , así como por el esquema de transición definido por la matriz  $F_t$ , permitiendo ajustar estructuras parsimoniosas, con menos parámetros que en el caso que se da con series de tiempo multivariadas.

Como se mencionó, es posible formular una gran variedad de series de tiempo, sin la necesidad de que ni  $\{Y_t\}$  o  $\{X_t\}$  sean necesariamente estacionarias. Sin embargo, como el objetivo es lograr modelar procesos ARMA, se deben considerar ecuaciones de estados y de observaciones que estén definidas  $\forall t \in \mathbb{Z}$ . Teniendo lo anterior en mente, considérense las ecuaciones de observaciones y estados

$$\begin{aligned} Y_t &= GX_t + W_t, \quad t \in \mathbb{Z} \\ X_{t+1} &= FX_t + V_t, \quad t \in \mathbb{Z} \end{aligned}$$

donde  $F$  y  $G$ , son matrices de dimensión  $v \times v$  y  $w \times w$  respectivamente,  $\{V_t\} \sim WN(0, Q)$ ,  $\{W_t\} \sim WN(0, R)$ ,  $\mathbb{E}(V_t W_t') = S, \forall t$  y  $V_s \perp W_t, \forall s \neq t$ .

Se dice que la ecuación 2.34 es estable o causal, si la matriz  $F$  tiene todos sus valores propios dentro del círculo unitario, o de forma equivalente, si  $\det(I - F_z) \neq 0, \forall z \in \mathbb{C}$  tal que  $|z| \leq 1$ . Bajo estas circunstancias, también se dice que la matriz  $F$  es estable.

En el caso estable, la ecuación 2.34 tiene una única solución estacionaria dada por

$$X_t = \sum_{j=0}^{\infty} F^j V_{t-j-1},$$

por lo cual, en este caso, la correspondiente ecuación de observaciones también es estacionaria

$$Y_t = W_t + \sum_{j=0}^{\infty} GF^j V_{t-j-1}$$

Con la finalidad de tener un panorama más claro de cómo es la representación de espacio de estados de los modelos de series de tiempo expuestos con anterioridad, se desarrollará el siguiente ejemplo.

**Ejemplo.- Representación de espacio de estados de un proceso AR(p) causal**

Se sabe que un proceso  $\{X_t\}$  es un proceso AR(p), si satisface la ecuación

$$X_t = \phi_1 X_t + \phi_2 X_{t-1} + \dots + \phi_p X_{t-p+1} + \varepsilon_t$$

donde  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$  y  $\phi(z) := 1 - \phi_1 z - \dots - \phi_p z^p \neq 0$  para  $|z| \leq 1$ . Para expresar a  $\{X_t\}$  en su representación de espacio de estados, se define al vector de estados

$$\mathbf{X}_t = \begin{bmatrix} X_{t-p+1} \\ X_{t-p+2} \\ \vdots \\ X_t \end{bmatrix}, \quad t = 0, \pm 1, \dots$$

Si al tiempo  $t$ , se observa que  $\mathbf{Y}_t = \mathbf{X}_t$ , entonces se obtiene a la ecuación de observaciones,

$$\mathbf{Y}_t = [0 \ 0 \ 0 \ \dots \ 1] \mathbf{X}_t$$

y la ecuación de estados,

$$\mathbf{X}_{t+1} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ \phi_p & \phi_{p-1} & \phi_{p-2} & \dots & \phi_1 \end{bmatrix} \mathbf{X}_t + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} Z_{t+1}$$

La representación de espacio de estados tiene varias ventajas; por decir algunas, su generalidad, su facilidad de análisis, y la disponibilidad de las recursiones *Kalman*, las cuales, permiten obtener fácilmente pronósticos por mínimos cuadrados y de igual manera, facilita la estimación de los parámetros asociados. De la misma manera, en el caso de datos faltantes, sistemas multivariados complejos, efectos mixtos, así como en ciertos tipos de no estacionariedad, es más sencillo trabajar en el esquema de modelos de espacio de estados.

Desde un punto de vista práctico, los primeros intentos de cualquier análisis que involucre al modelo de espacio de estados definido por 2.33 y 2.34, deben ser lograr producir estimadores para la señal subyacente no observada  $X_t$ , dadas las observaciones  $\mathbf{Y}_s = \{Y_1, \dots, Y_s\}$  a tiempo  $s$ . En el caso en que  $s < t$ , el problema anterior lleva por nombre **pronóstico** o *predicción*, cuando  $s = t$ , el problema se conoce como **filtración**, mientras que cuando  $s > t$ , se le llama **suavizamiento**. Además de la estimación, también es de particular interés medir la precisión de dichos estimadores.

La solución a los problemas planteados en el párrafo anterior, se obtiene a través del *filtro de Kalman*, en particular, se tendrá un especial interés en el caso en que se obtiene la solución al problema de pronóstico. Para lograr este cometido, se hará uso de la siguiente notación

$$\mathbf{X}_t^s = \mathbb{E}(\mathbf{X}_t | \mathbf{Y}_s)$$

y

$$\mathbf{P}_{t_1, t_2}^s = \mathbb{E}[(\mathbf{X}_{t_1} - \mathbf{X}_{t_1}^s)(\mathbf{X}_{t_2} - \mathbf{X}_{t_2}^s)]$$

en el caso en que  $t_1 = t_2$ , se escribirá  $\mathbf{P}_t^s$  por simplicidad.

El término de *filtro*, proviene del hecho de que  $\mathbf{X}_t^t$  es un filtro lineal de las observaciones  $\{\mathbf{Y}_1, \dots, \mathbf{Y}_t\}$ ; es decir,  $\mathbf{X}_t^t = \sum_{s=1}^t B_s \mathbf{Y}_s$ , para ciertas matrices adecuadas  $B_s$ . Una de las grandes características del *filtro de Kalman*, es que especifica como actualizar el filtro a partir de  $\mathbf{X}_{t-1}^{t-1}$  a  $\mathbf{X}_t^t$  una vez que una nueva observación  $\mathbf{Y}_t$  es obtenida, sin la necesidad de volver a procesar el conjunto de datos  $\mathbf{Y}_1, \dots, \mathbf{Y}_t$ .

Se demuestra que para el modelo de espacio de estados definido por las ecuaciones 2.33 y 2.34, con condiciones iniciales  $\mathbf{X}_0^0 = \boldsymbol{\mu}_0$  y  $\mathbf{P}_0^0$ , para  $t = 1, \dots, n$  se tiene que

$$\mathbf{X}_t^{t-1} = F\mathbf{X}_{t-1}^{t-1} \quad (2.35)$$

$$\mathbf{P}_t^{t-1} = F\mathbf{P}_{t-1}^{t-1}F' + Q \quad (2.36)$$

con

$$\mathbf{X}_t^t = \mathbf{X}_t^{t-1} F + K_t (\mathbf{Y}_t - G_t \mathbf{X}_t^{t-1}) \quad (2.37)$$

$$\mathbf{P}_t^t = [\mathbf{I} - K_t G_t] \mathbf{P}_t^{t-1} \quad (2.38)$$

donde

$$K_t = \mathbf{P}_t^{t-1} G_t' \left[ G_t \mathbf{P}_t^{t-1} G_t' + R \right]^{-1} \quad (2.39)$$

a la ecuación 2.39, se le conoce como *ganancia de Kalman*.

Los pronósticos para  $t > n$  se obtienen a través de 2.35 y 2.36 con condiciones iniciales  $\mathbf{X}_n^n$  y  $\mathbf{P}_n^n$ .

### 2.6.6. Calidad de los pronósticos

Existen muchas medidas estadísticas que sirven para describir qué tan bien se ajusta un modelo dada una muestra de datos, a dicha tarea se le conoce como *Bondad de Ajuste*, la cual, usualmente hace uso de los residuales para evaluar la calidad del modelo, sin embargo los residuales no necesariamente reflejan la capacidad del modelo para poder pronosticar observaciones futuras de forma satisfactoria.

Generalmente, quien realiza pronósticos, está preocupado sobre la precisión de los pronósticos con respecto a observaciones futuras, no necesariamente está preocupado sobre la bondad de ajuste del modelo, por lo cual, es importante lograr tener una forma objetiva para poder evaluar este aspecto.

En la literatura, a la precisión del pronósticos se le llama *error de pronósticos “fuera de la muestra”*, esto con la finalidad de poder distinguirlo de los residuales, errores, que surgen del proceso de ajustar un modelo.

Sin importar si se tiene un interés particular en ajustar un modelo de series de tiempo para obtener pronósticos de observaciones futuras, es recomendable hacer uso de las medidas que evalúan la precisión de los pronósticos como parte del esfuerzo realizado en la etapa de *validación del modelo*.

La forma usual para medir la precisión de los pronósticos es a través de los *errores a un paso* del pronóstico, definidos como

$$\varepsilon_t(1) = Y_t - \hat{Y}_t(t-1),$$

donde  $\hat{Y}_t(t-1)$ , es el pronóstico de  $Y_t$ , realizado un periodo antes, es decir, en  $t-1$ . Supóngase que se tienen  $n$  observaciones, para las cuales, se han calculado sus respectivos pronósticos, y por ende, se tienen  $n$  errores de pronósticos a un paso,  $\varepsilon_t(1)$ ,  $t = 1, 2, \dots, n$ . Entonces, a partir de estos errores se definirán las estadísticas más comunes que sirven para cuantificar la calidad de los pronósticos.

Las primeras tres estadísticas estandar de precisión de pronósticos son, el *error promedio*, o *error medio*, *ME* por sus siglas en inglés,

$$ME = \frac{1}{n} \sum_{t=1}^n \varepsilon_t(1),$$

la *desviación absoluta media*, o *error absoluto medio*, *MAD* por sus siglas en inglés,

$$MAD = \frac{1}{n} \sum_{t=1}^n |\varepsilon_t(1)|,$$

y por último, el *error cuadrático medio*, *MSE* por sus siglas en inglés,

$$MSE = \frac{1}{n} \sum_{t=1}^n [\varepsilon_t(1)]^2.$$

La estadística *ME*, es un estimador del valor esperado del error de pronóstico a un paso, el cual, se espera que sea igual a cero, es decir, si el valor del *ME* es considerablemente distinto de cero, es un indicador de cierto sesgo en el pronóstico. En el caso en el que ocurra que el *ME* del modelo ajustado, después de cierto tiempo, comience a alejarse del cero, es un indicador de que la serie subyacente ha cambiado de alguna manera, de tal manera que el modelo no pudo identificar dicho cambio, y por ende, se comenzaron a generar pronósticos sesgados, siendo que antes, no era así.

Por otro lado, tanto el *MAD*, como el *MSE*, miden la variabilidad del error de pronóstico. Es lógico que sea deseable que la variabilidad en los errores de pronóstico sea pequeña. El *MSE*, es un estimador directo de la varianza del error de pronóstico a un paso, es decir,

$$\hat{\sigma}_{\varepsilon(1)}^2 = MSE = \frac{1}{n} \sum_{t=1}^n [\varepsilon_t(1)]^2,$$

en el caso en que los errores de pronóstico a un paso, estén normalmente distribuidos, el *MAD* está relacionado con la desviación estandar del error de pronóstico a un paso de la siguiente manera,

$$\hat{\sigma}_{\varepsilon(1)} = \sqrt{\frac{\pi}{2}} MAD \cong 1.25MAD.$$

El error de pronóstico a un paso, tiene escala en términos de las unidades de medida de la serie de tiempo subyacente, y por ende, lo mismo ocurre con las estadísticas  $ME$ ,  $MAD$  y  $MSE$ , es decir, son medidas dependientes de escala. Las medidas de precisión que son dependientes de escala, son difíciles de interpretar, pues no es claro en ocasiones si el valor de la medida es grande o pequeño, además de que no facilita la comparación de un método de pronóstico a lo largo de distintas series, o distintos periodos de tiempo.

Es por lo anterior, que para lograr solventar estos inconvenientes, se hace uso de medidas relativas del error de los pronósticos. Se define al *error de pronóstico relativo*, en porcentaje, como

$$r\epsilon_t(1) = \left( \frac{Y_t - \hat{Y}_t(t-1)}{Y_t} \right) \cdot 100 = \left( \frac{\epsilon_t(1)}{Y_t} \right) \cdot 100$$

A este error, en la literatura se le conoce como *error de pronóstico porcentual*. A partir de estos errores, se definen las siguientes dos estadísticas, el *error de pronóstico porcentual medio*,  $MPE$  por sus siglas en inglés

$$MPE = \frac{1}{n} \sum_{t=1}^n r\epsilon_t(1),$$

y el *error de pronóstico porcentual absoluto medio*,  $MAPE$  por sus siglas en inglés

$$MAPE = \frac{1}{n} \sum_{t=1}^n |r\epsilon_t(1)|.$$

De esta manera, por ejemplo, el saber que el  $MAPE$  del modelo es de 3%, puede ser mucho más ilustrativo que saber que el  $MAD$  es de 5 unidades(kg,m,in,etc).

## 2.7. Estimación de parámetros

Para realizar la estimación de los parámetros asociados al modelo SARIMA, se utilizará la representación de espacio de estados de una serie de tiempo, se expondrá el caso general del modelo de espacio de estados definido por las ecuaciones 2.33 y 2.34.

Se define al vector de parámetros  $\Theta$  como  $\Theta = \{\mu_0, \Sigma_0, \Phi, Q, R\}$ , el cual, contiene a la media y a la matriz de covarianzas iniciales,  $\mu_0$  y  $\Sigma_0$ , a la matriz de transición  $\Phi$ , y a las matrices de covarianzas de los *estados* y de las *observaciones*.

Se utilizará el método de máxima verosimilitud, bajo el supuesto de que los estados iniciales, siguen una distribución normal multivariada, es decir,  $\mathbf{X}_0 \sim N_p(\mu_0, \Sigma_0)$ . También se hará el supuesto de que los vectores de errores  $W$  y  $V$  son no correlacionados y que tienen una distribución conjunta normal multivariada.

La verosimilitud, se calcula haciendo uso de las inovaciones  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ , las cuales están definidas como

$$\varepsilon_t = \mathbf{Y}_t - A_t \mathbf{X}_t^{t-1}$$

La expresión de la función de verosimilitud que se obtiene a través del algoritmo de inovaciones, fué obtenida por *Schweppe(1965)*. Dicha expresión se obtiene al notar que las inovaciones son vectores aleatorios gaussianos independientes, con vector de medias  $\mathbf{0}$  y, con matriz de covarianzas

$$\Sigma_t = A_t \mathbf{P}_t^{t-1} A_t' + R,$$

por lo tanto, la log-verosimilitud  $L_Y(\theta)$  es

$$-\ln(L_Y(\theta)) = \frac{1}{2} \sum_{t=1}^n \ln |\Sigma_t(\theta)| + \frac{1}{2} \sum_{t=1}^n \varepsilon_t'(\theta) \Sigma_t^{-1}(\theta) \varepsilon_t(\theta) \quad (2.40)$$

donde se enfatiza la dependencia de las inovaciones en los parámetros  $\theta$ . Claramente 2.40, es una función no lineal, sumamente complicada, de los parámetros desconocidos.

El proceso usual para encontrar los estimadores es fijar  $\mathbf{X}_0$  y desarrollar un conjunto de recursiones sobre la función de log-verosimilitud, así como sobre sus primeras dos derivadas. Después, se puede usar un algoritmo como el de *Newton-Raphson* de forma sucesiva para actualizar los valores de los parámetros hasta que se logre minimizar la menos log-verosimilitud.

El proceso anterior fue adoptado por *Jones(1980)* para desarrollar la estimación de los parámetros de un proceso ARMA en su representación de *espacio de estados*.

Los pasos involucrados para obtener estimaciones de los parámetros a través del método de Newton-Raphson son:

1. Elegir valores iniciales para los parámetros, se denotarán a estos valores como  $\Theta^{(0)}$
2. Aplicar el *filtro de Kalman*, usando los parámetros iniciales  $\Theta^{(0)}$ , y de esta manera, se obtiene un conjunto de *inovaciones*, así como las covarianzas de los errores,

denotándose como,

$$\begin{aligned} \{\varepsilon_t^{(0)}; \quad t = 1, 2, \dots, n\} \\ \{\Sigma_t^{(0)}; \quad t = 1, 2, \dots, n\} \end{aligned}$$

3. Aplicar una iteración del algoritmo de *Newton-Raphson*, con  $-\ln(L_y(\Theta))$  como la función de criterio para obtener un nuevo conjunto de estimadores, denotados como  $\Theta^{(1)}$
4. En la iteración  $j$ , con  $j = 1, 2, \dots$ , repetir el paso 2, usando  $\Theta^{(j)}$  en lugar de  $\Theta^{(j-1)}$ , para obtener un nuevo conjunto de innovaciones y de covarianzas de los errores,

$$\begin{aligned} \{\varepsilon_t^{(j)}; \quad t = 1, 2, \dots, n\} \\ \{\Sigma_t^{(j)}; \quad t = 1, 2, \dots, n\} \end{aligned}$$

Después, repetir 3 para obtener una nueva estimación  $\Theta^{(j+1)}$ . Este procedimiento se detendrá cuando se estabilicen los estimadores o la verosimilitud.

Cabe destacar, que en el caso univariado del proceso ARMA en su representación de *espacio de estados*, 2.40, es idéntica a la que se obtiene al desarrollar la verosimilitud de un proceso ARMA. Esta verosimilitud está dada por,

$$L(\beta, \sigma_w^2) = (2\phi\sigma_w^2)^{-\frac{n}{2}} [r_1^0(\beta) \cdot r_2^1(\beta) \cdots r_n^{n-1}(\beta)]^{\frac{1}{2}} \cdot \exp\left[-\frac{S(\beta)}{2\sigma_w^2}\right] \quad (2.41)$$

donde

$$S(\beta) = \sum_{t=1}^n \left[ \frac{(X_t - X_t^{t-1}(\beta))^2}{r_t^{t-1}(\beta)} \right]$$

y

$$r_t^{t-1} = \frac{1}{\sigma_w^2} P_t^{t-1}$$

## 2.8. Validación de supuestos

Típicamente, la bondad de ajuste de un modelo estadístico, basado en un conjunto de datos, es realizado al comparar los valores observados con sus correspondientes valores

pronosticados, los cuales, son obtenidos a través del modelo ajustado. En el caso en que el modelo ajustado es el apropiado, entonces, los residuales deberán comportarse de una manera que sea consistente con el modelo.

Cuando se ajusta un modelo  $ARMA(p, q)$  a una serie de datos dada, se determinan los estimadores máximo verosímiles  $\hat{\phi}, \hat{\theta}$  y  $\hat{\sigma}^2$  de los parámetros  $\phi, \theta$  y  $\sigma^2$ . A lo largo de este procedimiento, los pronósticos a un paso  $\hat{X}_t^{t-1}$  de  $X_t$ , basados en  $X_1, \dots, X_{t-1}$ , son ajustados a través del modelo ajustado. Por lo tanto, los residuales quedan definidos como

$$\hat{\varepsilon}_t = \frac{(X_t - \hat{X}_t^{t-1})}{\sqrt{\hat{P}_t^{t-1}}}, \quad t = 1, \dots, n,$$

donde, de nuevo,  $\hat{X}_t^{t-1}$  es el pronóstico a un paso de  $X_t$  y  $\hat{P}_t^{t-1}$ , es el estimador de la varianza del error de pronóstico a un paso.

Si se supone que el modelo  $ARMA(p, q)$  obtenido por máxima verosimilitud, es el verdadero proceso generador de  $\{X_t\}$ , entonces, se podría decir que  $\{\hat{\varepsilon}_t\} \sim WN(0, \hat{\sigma}^2)$ . Sin embargo, para validar qué tan apropiado es el modelo  $ARMA(p, q)$  para modelar los datos, únicamente se puede suponer que el proceso  $\{X_t\}$ , es generado por un proceso  $ARMA(p, q)$  con parámetros desconocidos  $\phi, \theta$  y  $\sigma^2$ , cuyos estimadores máximo verosímiles son  $\hat{\phi}, \hat{\theta}$  y  $\hat{\sigma}^2$  respectivamente. Entonces, no es cierto que  $\{\hat{\varepsilon}_t\}$  sea un proceso de ruido blanco. Sin embargo,  $\{\hat{\varepsilon}_t\}$  debería tener propiedades que sean similares a las de un proceso de ruido blanco, de hecho,  $\hat{\varepsilon}_t$ , se aproxima al proceso de ruido blanco en el sentido en que  $\mathbf{E}[(\hat{\varepsilon}_t - \varepsilon_t)^2] \rightarrow 0$  cuando  $t \rightarrow \infty$ . Por consecuencia, las propiedades de  $\{\hat{\varepsilon}_t\}$ , deberían reflejar las propiedades del proceso de ruido blanco  $\{\varepsilon_t\}$  que genera al modelo  $ARMA(p, q)$ .

En particular, la sucesión  $\{\hat{\varepsilon}_t\}$ , debería ser aproximadamente:

- *no correlacionada* si  $\{\varepsilon_t\} \sim WN(0, \sigma^2)$
- *independiente* si  $\{\varepsilon_t\} \sim IID(0, \sigma^2)$
- *normalmente distribuida* si  $\varepsilon_t \sim N(0, \sigma^2)$

Por lo tanto, las siguientes validaciones de supuestos que se presentan, están basadas en las propiedades esperadas de los residuales  $\{\hat{\varepsilon}_t\}$ , bajo el supuesto de que el modelo ajustado es correcto y que  $\{\varepsilon_t\} \sim IID(0, \sigma^2)$ .

### 2.8.1. Pruebas de Ljung-Box y Box-Pierce

La función de autocorrelación de los residuales, ayuda a visualizar y/o probar, si los residuales individuales a distintos rezagos, siguen un proceso de ruido blanco, de hecho, idealmente los residuales del modelo ajustado, deberían seguir un proceso de ruido blanco gaussiano. Si esto ocurre, entonces, la distribución de los coeficientes de la función de autocorrelación de los residuales, en el rezago  $k$ , para muestras grandes, es aproximadamente normal con media cero y varianza  $\frac{1}{T}$ , donde  $T$  el tamaño de muestra, esto es

$$r_k \sim N\left(0, \frac{1}{T}\right)$$

Por lo cual, se puede probar la hipótesis nula  $H_0 : \rho_k = 0$ , usando la estadística de prueba

$$Z_0 = \frac{r_k}{\sqrt{\frac{1}{T}}} = r_k \sqrt{T}$$

Grandes valores de  $Z_0$ , en el sentido en que  $|Z_0| > Z_{\frac{\alpha}{2}}$ , donde  $Z_{\frac{\alpha}{2}}$ , es el cuantil  $\frac{\alpha}{2}$  de una distribución normal estándar, sería un indicador de que el coeficiente de la función de autocorrelación,  $r_k$ , no es igual a cero. Este procedimiento es realizado “uno a la vez”, es decir, el nivel de significancia es aplicado a la correlación de los residuales en rezagos individuales. Sin embargo, es de particular interés, el tener una manera de evaluar la magnitud de la correlación de los residuales como grupo.

Por ejemplo, podría ocurrir que la mayoría de las autocorrelaciones de los residuales fueran “moderadas”, incluso, que algunas estuvieran cerca de sus correspondientes valores críticos, pero, que al considerar todos los residuales, su autocorrelación fuera excesiva. Por lo tanto, cuando se tiene el interés de evaluar al mismo tiempo un conjunto de autocorrelaciones para determinar si el proceso en estudio sigue un proceso de ruido blanco, se puede utilizar el proceso sugerido por *Box & Pierce (1970)*, el cual, consiste en considerar el cuadrado de la estadística  $Z_0$ .

La distribución de  $Z_0^2 = r_k^2 T$  es aproximadamente ji-cuadrada con un grado de libertad. Por lo cual, la estadística de *Box-Pierce* se define como

$$Q_{BP} = T \sum_{i=1}^k r_i^2,$$

la cual, se distribuye aproximadamente ji-cuadrada con  $k$  g.l., bajo la hipótesis nula de que la serie de tiempo es un proceso de ruido blanco. Por lo tanto, si  $Q_{BP} > \chi_{\alpha, k}^2$ , se

rechazaría la hipótesis nula y se consideraría que la serie de tiempo no es un proceso de ruido blanco, ya que algunas de las autocorrelaciones serían significativamente distintas de cero.

Cuando este procedimiento es aplicado a un conjunto de autocorrelaciones de residuales, la estadística  $Q_{BP} \sim \chi_{\alpha, k-p}^2$ , donde  $p$  es el número de parámetros en el modelo, por lo cual, el número de g.l. en la distribución ji-cuadrada es  $k - p$ . *Box & Pierce* llamaron a este procedimiento *Portmanteau* o *estadística general de bondad de ajuste*, pues, está probando la bondad de ajuste de la función de autocorrelación en estudio con la función de autocorrelación de un ruido blanco.

En particular, si se estima un modelo  $ARMA(p, q)$  adecuado, para una  $n$  suficientemente grande,  $Q_{BP}$  tiene una distribución aproximada ji-cuadrada con  $k - p - q$  grados de libertad. De tal manera, que en el caso de ajustar un modelo erróneo, provocaría que se infle el valor de la estadística  $Q_{BP}$ . Es importante resaltar que el valor máximo de rezago  $k$ , es escogido de forma arbitraria, pero se busca que sea lo suficientemente grande para que los pesos  $\phi$ , sean insignificantes para  $j > k$ . Además, la distribución aproximada ji-cuadrada de  $Q_{BP}$ , está basada en un teorema límite cuando  $n \rightarrow \infty$ .

Sin embargo, *Ljung-Box (1978)*, demostraron que incluso para  $n = 100$ , la aproximación no es satisfactoria. Es por esto, que hicieron una pequeña modificación sobre la estadística  $Q_{BP}$ , y de esta manera, definieron otra estadística de prueba, cuya distribución bajo  $H_0$ , es mucho más cercana, a una ji-cuadrada para tamaños de muestra más comunes. Por lo tanto, se define a la estadística de *Ljung-Box*, como

$$Q_{LB} = T(T+2) \sum_{i=1}^k \left( \frac{1}{T-i} \right) r_i^2$$

Una observación importante es que debido a que  $\frac{T+2}{T-k} > 1 \quad \forall k \geq 1$ , se cumple que  $Q_{LB} > Q_{BP}$ , por lo cual, esto explica de manera parcial que la estadística  $Q_{BP}$ , tienda a pasar por alto modelos inadecuados. Además, para valores grandes de  $T$ , los pesos  $\frac{T+2}{T-k}$ , serán aproximadamente 1, y por ende,  $Q_{LB}$  y  $Q_{BP}$  serán muy similares.

### 2.8.2. Prueba de McLeod-Li

Los modelos  $ARMA$ , incluidos los modelos  $SARIMA$ , operan bajo el supuesto, entre otros, de que la varianza del proceso de ruido blanco asociado es constante. En el caso en que se cumpliera lo anterior, la varianza de los residuales del modelo ajustado, debería

tener un comportamiento constante a lo largo del tiempo, sin embargo, puede ocurrir que la varianza de los residuales siga un comportamiento no lineal, y por ende, el modelo no sea el adecuado. Al fenómeno anterior, se le conoce como *heterocedasticidad condicional autoregresiva*, i.e., *efecto ARCH*, por sus siglas en inglés (*Autoregressive Conditional Heteroskedasticity*).

Examinar el cuadrado de los residuales, en ocasiones, puede mostrar desviaciones de los valores ajustados por el modelo en contraste con los valores observados, dichas desviaciones, podrían no detectarse si se estudiaran simplemente los residuales de forma directa. De hecho, *Granger & Anderson*, encontraron ejemplos donde los residuales no estaban correlacionados, pero el cuadrado de estos, sí lo estaban, debido a esto, ellos sugirieron que la función de autocorrelación del cuadrado de los residuales, de hecho sobre cualquier serie de tiempo, puede ser útil para identificar comportamientos no lineales. Por otro lado, *Bollerslev (1986)*, sugirió, que la función de autocorrelación, así como la función de autocorrelación parcial del cuadrado de los residuales, o de nuevo, de cualquier proceso, son útiles para identificar y validar un posible *efecto ARCH*.

Afortunadamente, se puede probar la correlación del cuadrado de los residuales de la misma manera en que se prueba para los residuales mismos. Lo cual implica que tanto la gráfica de la *ACF* como de la *PACF*, son útiles para identificar de forma visual, indicios de que la varianza de los residuales, es condicional de sus valores pasados, es decir, que exhiben un *efecto ARCH*. Más allá del análisis de las gráficas de la *ACF* y/o de la *PACF*, existen métodos formales para probar si el proceso en estudio tiene un *efecto ARCH*, como por ejemplo, la *prueba de McLeod-Li*, la *prueba de multiplicadores de lagrange de Engle*, la *prueba BDS*, entre otras.

En particular, *McLeod-Li (1983)* se basaron en la prueba de *Ljung-Box* para probar la existencia de algún *efecto ARCH* en los residuales, es por esto que la prueba hace uso de la función de autocorrelación del cuadrado de los residuales, y prueba cuando las primeras  $h$  autocorrelaciones del cuadrado de los residuales, son en conjunto, pequeñas en magnitud.

De esta manera, se define a la función de autocorrelación parcial del cuadrado de los residuales como

$$\hat{\rho}_{ww}(h) = \sum_{t=1}^{n-h} \frac{(\hat{W}_t^2 - \bar{W}^2)(\hat{W}_{t+h}^2 - \bar{W}^2)}{(\hat{W}_t^2 - \bar{W}^2)}, \quad h \geq 1,$$

donde  $\bar{W}^2 = \frac{\sum_{t=1}^n \hat{W}_t^2}{n}$

*McLeod-Li*, probaron que  $Q_{WW} = n(n+2) \sum_{j=1}^h \frac{\hat{\rho}_{WW}^2(j)}{n-j}$  tiene una distribución aproximada  $\chi^2(h)$ , bajo el supuesto de que el modelo es adecuado, es decir, que no hay *efecto ARCH*, consecuentemente, la adecuación del modelo se rechaza a un nivel de significancia  $\alpha$ , si  $Q_{WW} > \chi_{1-\alpha}^2(h)$ .

Las pruebas de *Box-Pierce*, *Ljung-Box* y *McLeod-Li*, son pruebas *Portmanteau*, las cuales, tienen la ventaja de reunir información de las correlaciones  $\hat{\rho}(i)$ ,  $i = 1, \dots, h$  de los errores, a distintos rezagos. Sin embargo, una desventaja es que frecuentemente fallan en rechazar modelos con un ajuste pobre. Es por esto que en la práctica, las prueba de tipo *Portmanteau*, son más útiles para descartar modelos insatisfactorios que para seleccionar al mejor modelo ajustado, sobre modelos candidatos cercanamente competitivos.

### 2.8.3. Prueba para media igual a cero

Los residuales del modelo *ARMA*, no están sujetos, en principio, a la restricción de que

$$\sum_{t=1}^T \hat{\varepsilon}_t = 0.$$

Esta condición es impuesta únicamente en el caso en que se esté estimando un modelo *AR*, y que sus parámetros sean estimados a través de mínimos cuadrados, sin embargo, en la estimación a través de máxima verosimilitud, esta restricción no es necesaria.

Para probar la hipótesis nula, en el caso general, de que la media de los residuales del modelo es igual a cero, se utiliza la estadística de prueba

$$t = \frac{\bar{X}}{\frac{S}{\sqrt{n^*}}},$$

donde  $\bar{X}$  corresponde a la media muestral de los residuales, por su parte,  $S$  corresponde a la desviación estándar muestral de los residuales, y  $n^*$ , es el tamaño de muestra de los residuales. Por otro lado, los grados de libertad asociados son  $n^* - k - 1$ , donde  $k$  es el número de parámetros en el modelo.

Por lo tanto, se rechaza la hipótesis nula si el valor de  $t$  es significativamente grande, es decir, si ocurre que

$$|t| > t_{n^*-k-1}^{1-\frac{\alpha}{2}}$$

a un nivel de significancia  $\alpha$ . Es importante mencionar, que esta prueba debe ser aplicada después de haber validado que los residuales no están correlacionados, esto con la finalidad de poder asegurar que  $\hat{\sigma}^2 = S^2$ , es un estimador razonable para la varianza.

## 2.8.4. Pruebas de normalidad

Aunque en teoría se puede usar cualquier tipo de prueba de normalidad, para probar la hipótesis nula de normalidad en los residuales, como por ejemplo la prueba *Lilliefors* o la prueba de *Shapiro-Francia*. El artículo *B. W. Yap & C. H. Sim (2011) Comparisons of various types of normality tests, Journal of Statistical Computation and Simulation, 81:12, 2141-2155*, recomienda que, para distribuciones empíricas simétricas con coeficiente de curtosis mayor que el coeficiente de una distribución, las pruebas de normalidad con mayor potencia, es decir, las que tienen mayor probabilidad de no cometer un *error tipo II*, son las pruebas de *Anderson-Darling*, *Jarque-Bera* y *Shapiro-Wilk*. Los resultados de dichas pruebas se muestran a continuación, así como los coeficientes de curtosis empíricos.

### 2.8.4.1. Prueba de Anderson-Darling

La prueba de *Anderson-Darling*, es una alternativa a la prueba de *Kolmogorov-Smirnov (K-S)* y a la prueba de la *ji-cuadrada*, las cuales, son comúnmente usadas para *bondad de ajuste*. De hecho, la prueba de *Anderson-Darling*, es una modificación de la prueba K-S, la cual, busca darle más peso a las colas, en comparación con la prueba K-S.

A grandes rasgos, esta prueba hace uso de la distribución que se desea probar, en el sentido en que los valores críticos son calculados a través de la distribución específica. Esto le da la característica a la prueba, de ser más sensitiva, y la clara desventaja, de que los valores críticos se deben calcular dependiendo de la distribución a probar.

Por lo tanto, la prueba se define de la siguiente manera,

$H_0$  : Los datos siguen una distribución especificada

$H_a$  : Los datos no siguen una distribución especificada

La estadística de prueba se define como,

$$A^2 = -N - S,$$

donde  $S = \sum_{i=1}^N \frac{2i-1}{N} [\ln(F(Y_i)) + \ln(1 - F(Y_{N+1-i}))]$  y  $F$  es la distribución especificada. Además, nótese que las  $Y'_i$ s, son las observaciones ordenadas, mientras que  $N$  es el tamaño de muestra.

Como se mencionó previamente, los valores críticos de la prueba de *Anderson-Darling*, dependen de la distribución especificada, la cual, es la que se está probando. Valores tabulados, así como ecuaciones, han sido publicados, *Stephen 1974, 1976, 1977, 1979*), para algunas distribuciones, como lo son, *normal, log-normal, exponencial, weibull y logística*. La prueba es de una cola, y la hipótesis nula es rechazada, si la estadística es significativamente mayor que los valores críticos correspondientes.

#### 2.8.4.2. Prueba de Shapiro-Wilk

La prueba de *Shapiro-Wilk*, propuesta en 1965, busca probar si una muestra aleatoria  $X_1, X_2, \dots, X_n$ , proviene de una distribución normal, para esto, se calcula una estadística  $W$ , la cual, si su valor es pequeño, provee de evidencia estadística de que la muestra proviene de una distribución normal. Valores críticos de  $W$  son obtenidos a través de *simulación Monte Carlo*, dichos valores críticos fueron reproducidos por *Pearson & Hartley* en 1972. De esta manera, la estadística de prueba  $W$  se define como

$$W = \frac{\left( \sum_{i=1}^n a_i X_{(i)} \right)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

donde las  $X'_{(i)}$ s, corresponden a los valores muestrales ordenados, mientras que las  $a'_i$ s son constantes generadas por medio de las medias, varianzas y covarianzas de las estadísticas de orden de una muestra de tamaño  $n$  de una distribución normal.

Específicamente,

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

$$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{C}$$

donde  $C$  es la norma de un vector dada por

$$C = \|V^{-1}m\| = (m^T V^{-1} V^{-1} m)^{\frac{1}{2}}$$

$$m = (m_1, \dots, m_n)^T$$

siendo  $m_1, \dots, m_n$  los valores medios de la estadística de orden de variables aleatorias independientes e idénticamente distribuidas, muestreadas de distribuciones normales. Por último,  $V$  es la matriz de varianzas y covarianzas de esa estadística de orden.

### 2.8.4.3. Prueba de Jarque-Bera

Esta prueba de bondad de ajuste busca probar cuando datos muestrales tienen la asimetría y curtosis de una distribución normal. La estadística de prueba siempre es no negativa, si está muy alejada de cero, es un indicador de que los datos no provienen de una distribución normal. La estadística de prueba  $JB$  está definida como

$$JB = \frac{n-k+1}{6} \left( S^2 + \frac{1}{4} (C-3)^2 \right)$$

donde  $n$  corresponde al número de observaciones;  $S$  corresponde al coeficiente de asimetría muestral, mientras que  $C$  corresponde a la curtosis muestral y por último,  $k$  corresponde al número de regresores, siendo  $k = 1$  fuera del contexto de regresión.

El coeficiente de asimetría muestral, así como la curtosis muestral, se definen como

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^{\frac{3}{2}}}$$

$$C = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left( \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right)^2}$$

respectivamente, donde  $\hat{\mu}_3$  y  $\hat{\mu}_4$ , son los estimadores, respectivamente, del tercero y cuarto momento central;  $\bar{X}$  es la media muestral; y  $\hat{\sigma}^2$  es el estimador del segundo momento central. Por lo cual, si los datos, en efecto provienen de una distribución normal, entonces, la estadística  $JB$ , tendrá una distribución asintótica, ji-cuadrada con 2 *g.l.*, por lo cual, la estadística  $JB$  es usada para probar la hipótesis de que los datos provienen de una distribución normal.

La hipótesis nula, es una hipótesis conjunta de la asimetría siendo igual a cero, así como del exceso de curtosis, Curtosis  $- 3$ , pues muestras que provengan de una distribución normal, tendrán un coeficiente de asimetría esperado igual a 0, y un exceso de curtosis, también, igual a 0. Como muestra la estadística  $JB$ , cualquier desviación de estos valores, se verá reflejado aumentando el valor de la estadística de prueba.



# Capítulo 3

## Aplicación

### 3.1. Introducción

Se buscará proponer un modelo capaz de realizar pronósticos confiables, que sirvan de guía para el público inversionista en la toma de decisiones respecto al comportamiento de los rendimientos otorgados por los Fideicomisos de Infraestructura y Bienes Raíces. Es importante aclarar que dicho modelo buscará explicar el comportamiento de todos los valores del mercado de renta variable listados como FIBRAS en la BMV de forma conjunta.

Dicho modelo intentará reproducir el comportamiento del índice S&P/BMV FIBRAS, el cual, está compuesto por una muestra de FIBRAS, las cuales, se eligieron de acuerdo a la metodología descrita en el capítulo uno.

El índice S&P/BMV FIBRAS, es un indicador de la tasa de retorno de las FIBRAS, el cómo está constituido y el cómo es analizado se explicó con detalle en el **Capítulo 1**.

De las 11 FIBRAS listadas en la BMV, únicamente 10 de estas componen actualmente el índice, esta información está al 31 de Agosto de 2018, justo después del rebalanceo bianual mencionado en la metodología del índice del capítulo 1. Además, cabe aclarar que aunque se dijo que actualmente hay 12 FIBRAS listadas, FIBRA E no es propiamente una FIBRA, por lo cual, tampoco está presente en el índice.

La fecha de lanzamiento del índice fue el 10 de octubre de 2011, toda la información presentada antes de la fecha de lanzamiento se realizó con datos históricos. Los cálculos de los valores históricos se basan en la misma metodología que estaba vigente cuando se lanzó oficialmente el índice. Aunque se tenga la información histórica previa al lanzamiento del índice, para no generar confusión se muestra a continuación la serie a comienzos del 2012 hasta finales de 2017. Es importante aclarar que para el análisis se utilizará únicamente la información diaria del año 2017.

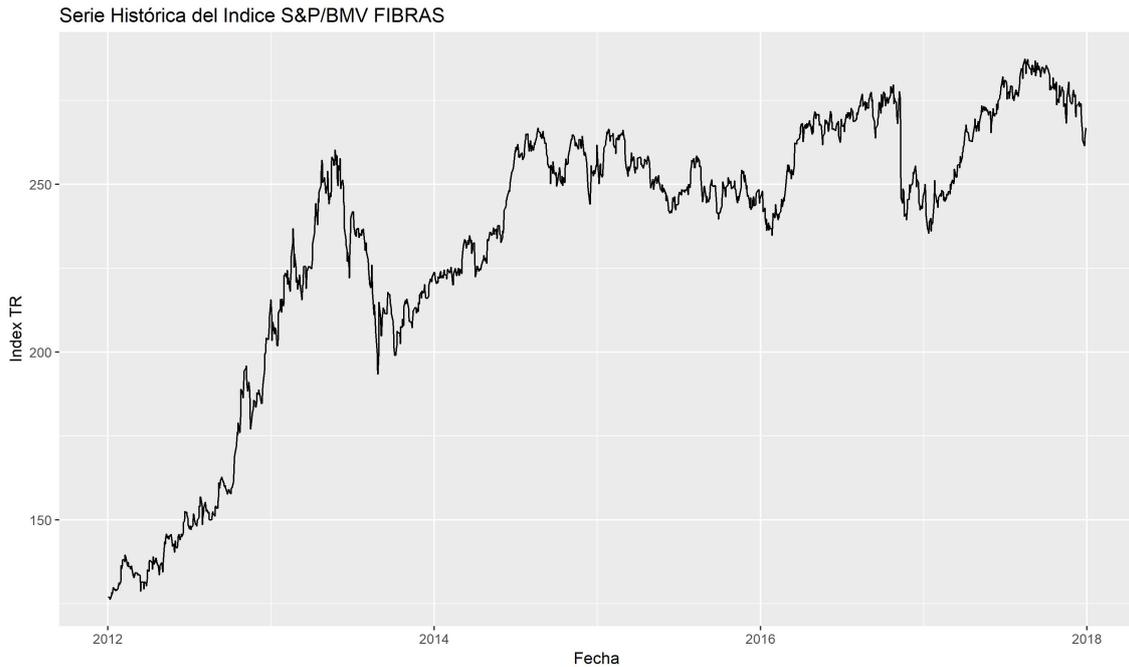


Figura 3.1: Serie Histórica S&P/BMV FIBRAS

## 3.2. Análisis Descriptivo

Para tener una idea general del comportamiento anual de la serie *S&P/BMV FIBRAS*, se muestran, a continuación, todas las series anuales en una misma gráfica. Lo primero que salta a la vista es que el nivel inicial del índice, es mayor o igual que el nivel inicial del año previo, exceptuando el año 2016, ya que a finales del 2015 el índice tuvo una fuerte caída. Sin embargo, el comportamiento previamente mencionado vuelve a estar presente a inicios del 2017.

El año 2012 fue muy particular pues hubo un claro crecimiento sostenido en el índice,

lo cual no es de sorprenderse debido a que las FIBRAS, seguían siendo un instrumento financiero relativamente nuevo.

Para los años posteriores, se aprecia en general, un comportamiento general. El cual, queda descrito con un crecimiento a inicios de año, para después comenzar a estabilizarse y fluctuar alrededor de cierto valor y por último, tiene un decaimiento. Este comportamiento ocurre en menor o mayor medida en los años 2013, 2014, 2016 y 2017.

El año 2015, al igual que el 2012, tiene un comportamiento totalmente distinto al resto de los años, pues en este año el índice tuvo un comportamiento decreciente.

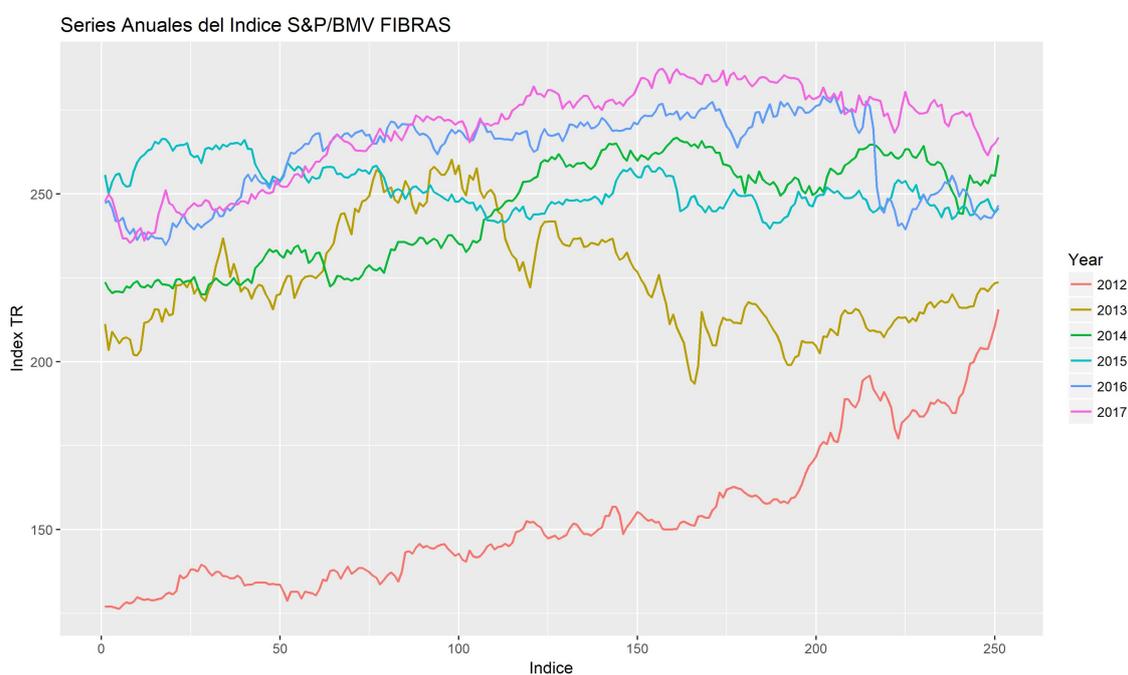


Figura 3.2: Series Anuales S&P/BMV FIBRAS

Otro evento muy evidente, es que a finales de 2016 hubo una caída muy grande en el índice. Históricamente, el índice nunca había tenido una caída de esta magnitud de un día para otro, lo cual, hace suponer que esta caída fue provocada por algún fenómeno externo que no es explicado por la naturaleza del instrumento.

Aunque el crecimiento del índice a través de los años no sea muy claro debido a sus constantes altas y bajas, se puede resumir su comportamiento general obteniendo la mediana de sus tasas de incremento (o decremento semanales), las cuales se observan a continuación.

| Año  | Tasa Crecimiento S&P/BMV FIBRAS |
|------|---------------------------------|
| 2012 | .6 %                            |
| 2013 | -.315 %                         |
| 2014 | .372 %                          |
| 2015 | -.031 %                         |
| 2016 | .161 %                          |
| 2017 | .0167 %                         |

Cuadro 3.1: Medianas de las tasas de crecimiento semanales para los años 2012-2017

Por lo tanto, la tabla resume bien el hecho de que en el año 2013 y 2015 fue cuando hubo una tendencia negativa en el índice S&P/BMV FIBRAS, esto se ve reflejado con las tasas negativas. De la misma manera, se ve reflejado el hecho de que en el año 2012 fue el año con mayor incremento en el índice, debido a esto, en este año se tiene la mediana más alta en las tasas de incremento semanales.

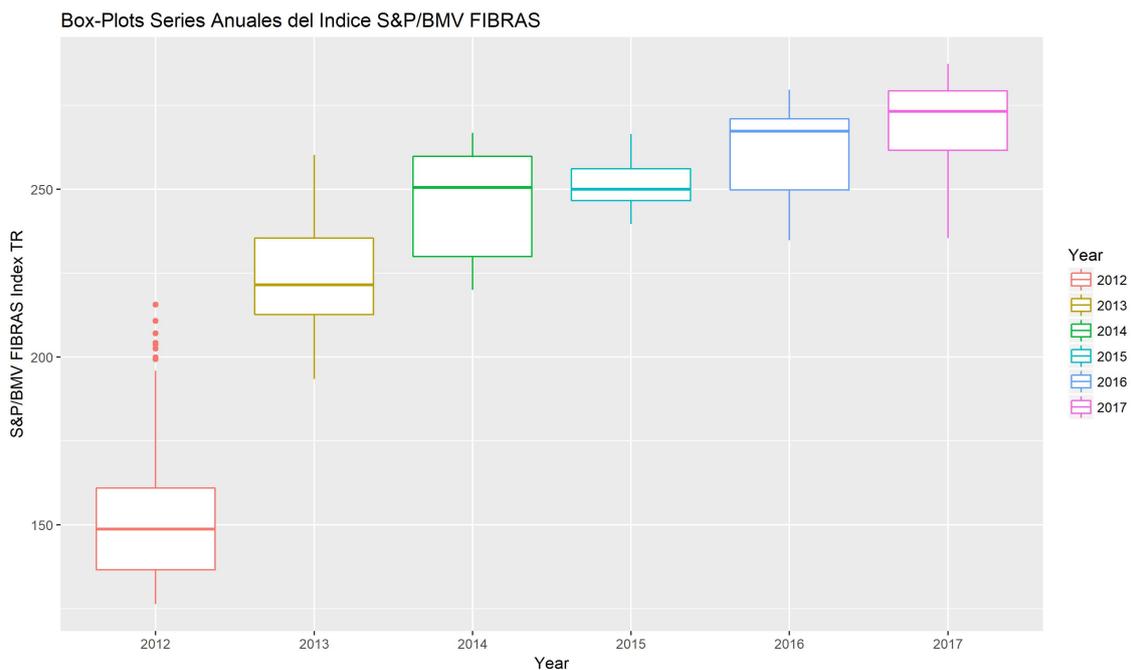


Figura 3.3: Box-Plots Anuales S&P/BMV FIBRAS

Hasta ahora, se ha puesto especial énfasis en conocer el comportamiento de la serie a través del tiempo, poniendo principalmente atención a cambios que se puedan presentar año con año. Sin embargo, también es de interés conocer la distribución del índice en cada

año.

Se puede extraer parcialmente esa información del Box-Plot anterior, del cual, destacan los años, 2012 y 2015. En el año 2012, es el único año donde hay presencia de *outliers*, esto está en línea con el hecho de que a lo largo de este año, el índice tuviera un crecimiento sostenido, no obstante, el índice en este año pareciera no tener algún sesgo hacia cierto lado.

Por otro lado, en el año 2015, también pareciera que la distribución del índice en ese año no tuviera algún sesgo, además, que este año fue el que presentó menor variabilidad.

En contraste, el índice en el resto de los años parece que tiene algún tipo de sesgo, aunque en general, aparentemente la misma variabilidad. Lo anterior lo podemos confirmar con la siguiente tabla, la cual está compuesta con el coeficiente de variación y rango intercuantil de cada año.

| <b>Año</b> | <b>Coeficiente de Variación</b> | <b>Rango Intercuartil</b> |
|------------|---------------------------------|---------------------------|
| 2012       | 13.6 %                          | 24.6                      |
| 2013       | 7.04 %                          | 22.7                      |
| 2014       | 6.35 %                          | 29.9                      |
| 2015       | 2.58 %                          | 9.4                       |
| 2016       | 4.81 %                          | 21.1                      |
| 2017       | 5.17 %                          | 17.8                      |

Cuadro 3.2: C.V y Rango Intercuartil para los años 2012-2017.

A pesar de que decir que el rango intercuantil es pequeño es totalmente subjetivo, aunque se puede observar que no hay años en los que haya un valor extremadamente diferente en comparación con el resto. Destaca por supuesto que el año 2012 tiene un coeficiente de variación más grande que el resto, lo cual se explica por el hecho de haber tenido un crecimiento sostenido, lo cual, por ende, el índice tomó valores en un rango mucho más amplio.

Los histogramas dan mayor información acerca de la distribución del índice en cada uno de los años, ya que con los *Box-Plots*, no es posible observar que la mayoría de los histogramas, parecieran ser bimodales. Lo anterior se explica debido al comportamiento de las series descrito con anterioridad, para esto se debe recordar que las series parecían tener dos comportamientos muy marcado a lo largo del año, uno de ellos era en forma de crecimiento, o decrecimiento, mientras que el otro, era una cierta fluctuación alrededor de cierto nivel.

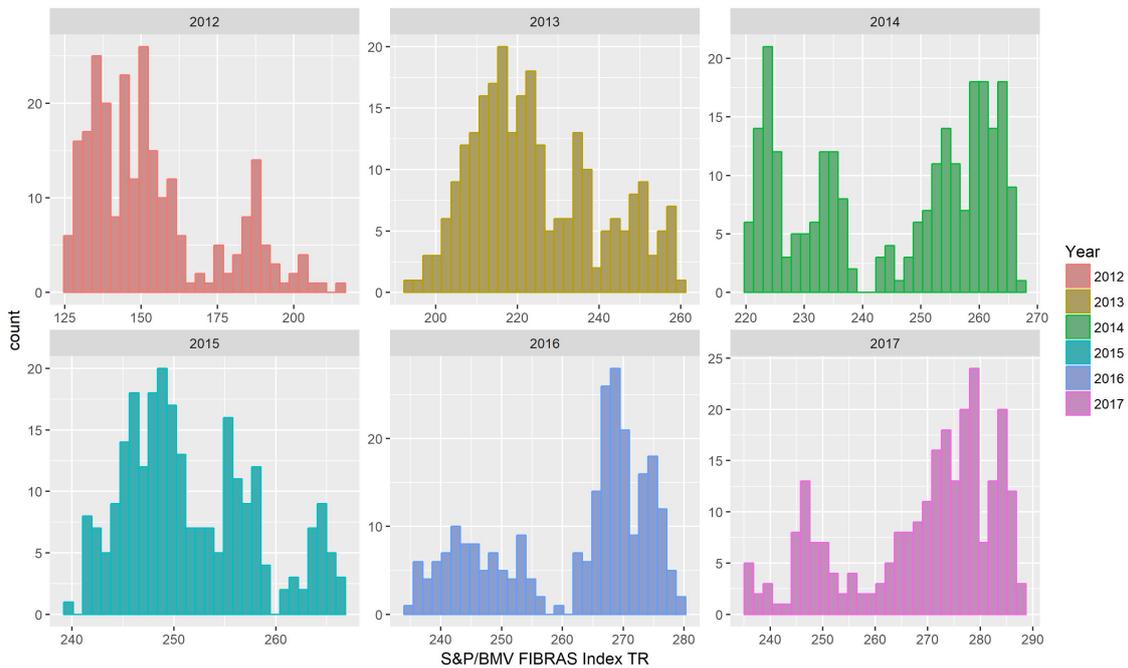


Figura 3.4: Histogramas Anuales S&P/BMV FIBRAS

### 3.3. Índice S&P/BMV FIBRAS 2017

#### 3.3.1. Introducción

El modelo que se buscará proponer, tiene como finalidad el poder proveer pronósticos confiables para la toma de decisiones a corto plazo. Es decir, se buscará que el modelo no tenga una dependencia temporal excesivamente larga.

Debido a lo anterior, el modelo se alimentará únicamente de la información del índice diario del año 2017 y se probará su capacidad predictiva contra las observaciones correspondientes del año 2018.

Un modelo más robusto para futuros trabajos sería el utilizar por ejemplo, la información de dos años, cuyo conjunto de datos tendría dos fuentes estacionales. Una que provenga de la frecuencia semanal, la cual, de forma eurística está presente en los datos diarios, así como de una fuente de estacionalidad anual. La cual, se observó con el análisis descriptivo sobre las series anuales del índice S&P/BMV FIBRAS.

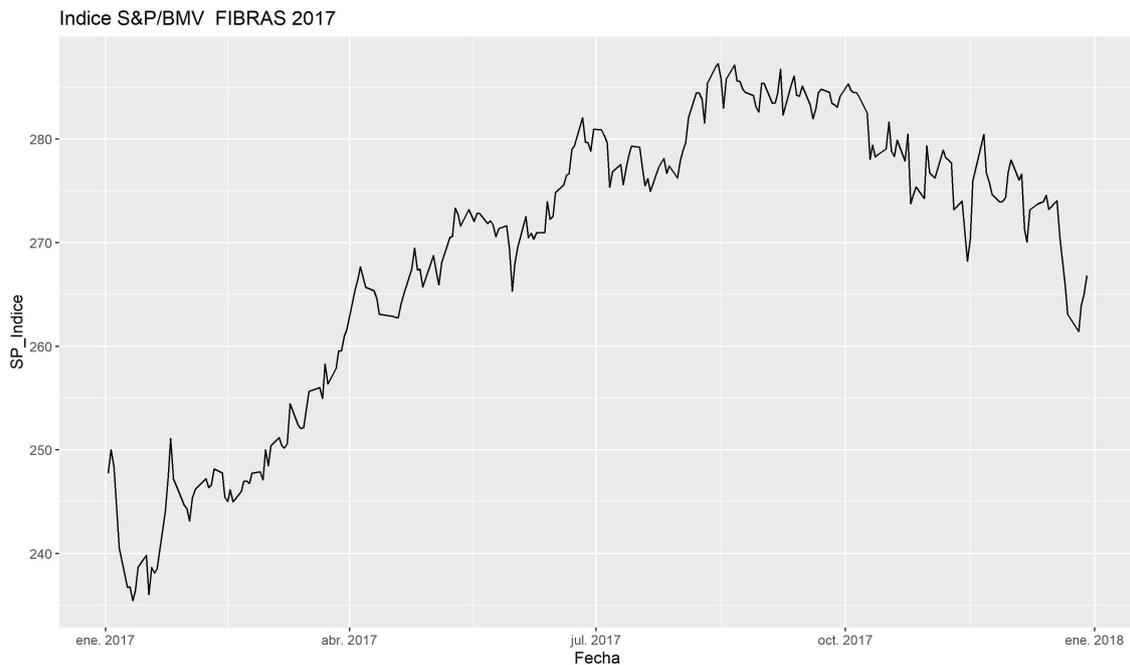


Figura 3.5: Serie S&P/BMV FIBRAS 2017

La serie del año 2017 del índice S&P/BMV FIBRAS, claramente tiene una componente de tendencia. Muy probablemente tenga también una componente estacional que esté “escondida”. Un detalle que no se mencionó previamente, el cual es importante tomar en consideración, es que, debido a que el índice es una serie financiera, no hay operaciones los fines de semana ni días festivo. Esto implica que en un año haya menos de 365 datos.

### 3.3.2. Análisis del índice S&P/BMV FIBRAS 2017

Aunque el modelo que se buscará ajustar es un modelo SARIMA, se hará uso de la descomposición clásica de una serie de tiempo, suponiendo un modelo aditivo. El cual se define como

$$Y_t = T_t + E_t + e_t$$

Donde,  $Y_t$ , es la serie de tiempo,  $T_t$ , es una componente de tendencia,  $E_t$  es una componente de estacionalidad y  $e_t$  es la componente aleatoria, la cual, no está explicada por

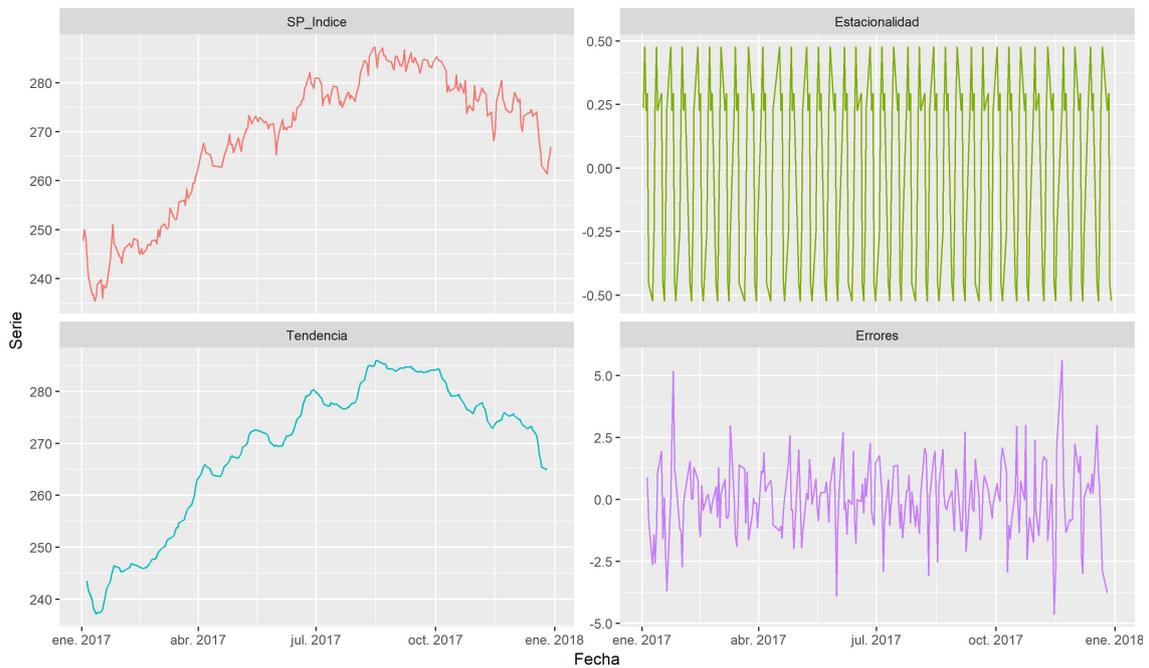


Figura 3.6: Descomposición Clásica de la Serie S&P/BMV FIBRAS 2017

las otras dos componentes. Esta descomposición se realizó con la finalidad de mostrar de forma más clara las componentes estacionales y de tendencia presentes en la serie.

La tendencia en la descomposición clásica suele ser estimada a través de promedios móviles, otra manera usual de realizar la estimación para esta componente es a través de una regresión polinomial, el polinomio más simple que se puede ajustar es a través de un modelo de regresión lineal simple. Dicho ajuste se muestra a continuación, de nuevo, es importante aclarar que estos ajustes tienen una finalidad descriptiva de la serie en estudio.

Además, con esta estimación simple de la tendencia, se puede dar una primera certeza estadística de la existencia de la componente de tendencia en la serie. Si analizamos la siguiente tabla que muestra el ajuste de la regresión lineal sobre la serie del *Índice S&P/BMV FIBRAS*, se puede observar que ambos parámetros son significativos a cualquier nivel de significancia usual, es decir para  $\alpha = .01, .05, .1$ . Además, como el coeficiente  $\hat{\beta}_1 > 0$ , es el coeficiente asociado al tiempo, quiere decir que la tendencia es positiva. Estas conclusiones por supuesto, únicamente nos indica que la regresión es significativa.

El tener claro que la serie en estudio tiene una componente de tendencia, permite tener la certeza de que finalmente, tendremos que modelar de alguna forma esta componente.

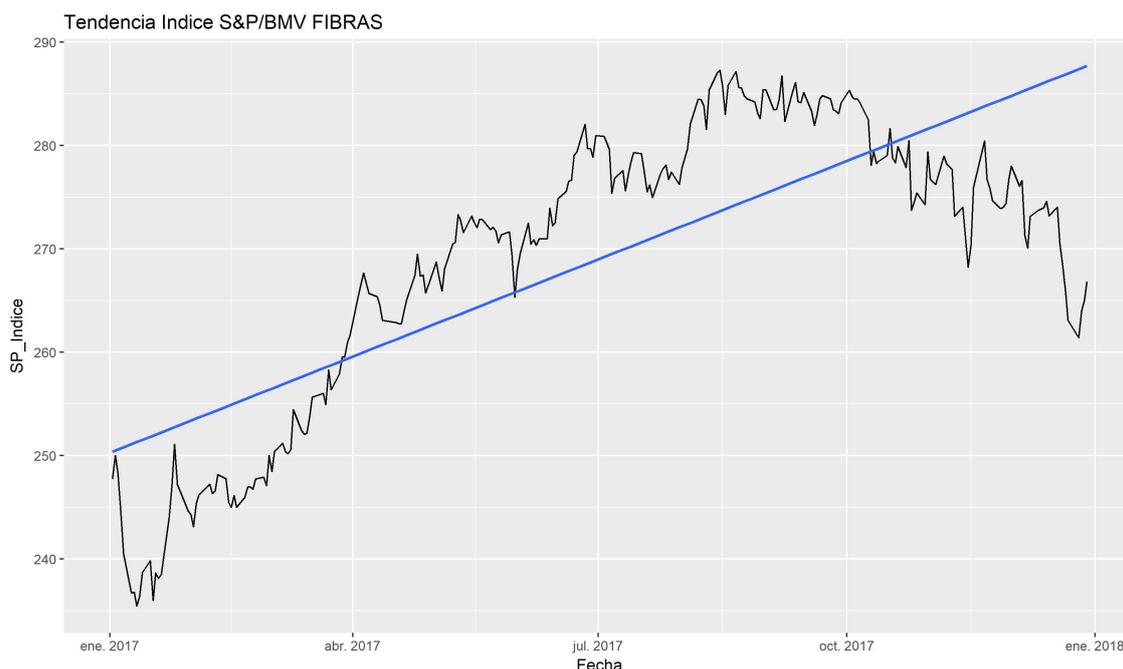


Figura 3.7: Regresión de la Serie S&amp;P/BMV FIBRAS 2017

| <b>Coefficiente</b> | <b>Estimación</b> | <b>Pr(&gt;  t )</b> |
|---------------------|-------------------|---------------------|
| <b>Intersección</b> | 250.317           | $< 2e - 16$         |
| <b>Tiempo</b>       | 0.1482            | $< 2e - 16$         |

Cuadro 3.3: Estimación Regresión Simple para el índice S&amp;P/BMV FIBRAS

Debido a que se requiere que la serie sea estacionaria. Existen muchas maneras de estimar esta tendencia, de hecho, mediante los últimos dos gráficos se han mostrado dos estimaciones de la tendencia de la serie, sin embargo, esta dos no son las únicas maneras de hacerlo.

Por ejemplo, ya se mostró el ajuste a través de una regresión lineal, la cual, será un buen estimador si la tendencia en la serie es lineal. Aunque claramente, la tendencia no tiene que ser necesariamente lineal, por lo cual, otros métodos como el de una regresión polinomial, suavizamiento exponencial o promedios móviles sería más adecuado. Además de los métodos anteriores, también se podría pensar en la tendencia de forma estocástica, para lo cual, un filtro de utilidad sería aplicar el *operador diferencia* en la serie.

En la serie del índice *S&P/BMV FIBRAS*, es claro que la regresión simple aunque sea significativa, no captura totalmente la tendencia que se observa, ya que la tendencia que

se busca modelar tiene una tendencia que no es lineal, la cual, es positiva desde inicios de año hasta mas o menos finales de septiembre, después la tendencia se invierte y a partir de ese momento es negativa hasta finales de año.

Un modelo de regresión polinomial o de medias móviles es más adecuado para modelar la tendencia en este caso, sin embargo, utilizaremos el operador diferencia para filtrar dicha tendencia de la serie.

Es importante recordar, que si la serie tiene una componente estacional así como una tendencia, la serie no será estacionaria. Entonces, debido a las gráficas anteriores, la serie de tiempo descrita por el índice S&P/BMV FIBRAS no debería ser estacionaria. Esta hipótesis la podemos probar mediante la prueba de *Dicky-Fuller Aumentada*. Hay que recordar que dicha prueba tiene como hipótesis nula, que la serie tiene raíces unitarias, es decir, que la serie no es estacionaria.

El resultado de dicha prueba sobre la serie estudiada se presenta a continuación.

| Estadística | Pr( $>  t $ ) |
|-------------|---------------|
| 0.011164    | 0.99          |

Cuadro 3.4: Prueba de *Dicky-Fuller Aumentada* para índice S&P/BMV FIBRAS

Es claro que no se rechaza la hipótesis nula de que la serie de tiempo no es estacionaria a ningún nivel de significancia habitual  $\alpha = .01, .05, .1$ . Por lo tanto, se tiene evidencia estadística para suponer que la serie NO es estacionaria.

Por lo tanto, al diferenciar la serie para filtrar la tendencia, se obtiene la siguiente serie diferenciada, la cual, claramente ya no tiene una tendencia.

Se podría seguir modelando esta serie diferenciada, sin embargo, al notar que la amplitud de los valores de la serie para los periodos a finales del 2017 es mucho más grande que la amplitud a inicios de año, es un indicio de que la serie no tiene varianza constante. Se hará uso de las transformaciones de Box-Cox, las cuales, son usadas frecuentemente para resolver este problema.

Se hará uso del conocimiento empírico que se tiene sobre series de naturaleza financiera, esto con la finalidad de elegir la *lambda* de la transformación. Dicho conocimiento empírico sugiere sacar el logaritmo de la serie, esta transformación se logra utilizando el valor de  $\lambda = 0$ .

Entonces, el logaritmo del *Índice S&P/BMV FIBRAS*, se muestra en la fig. 3.9.

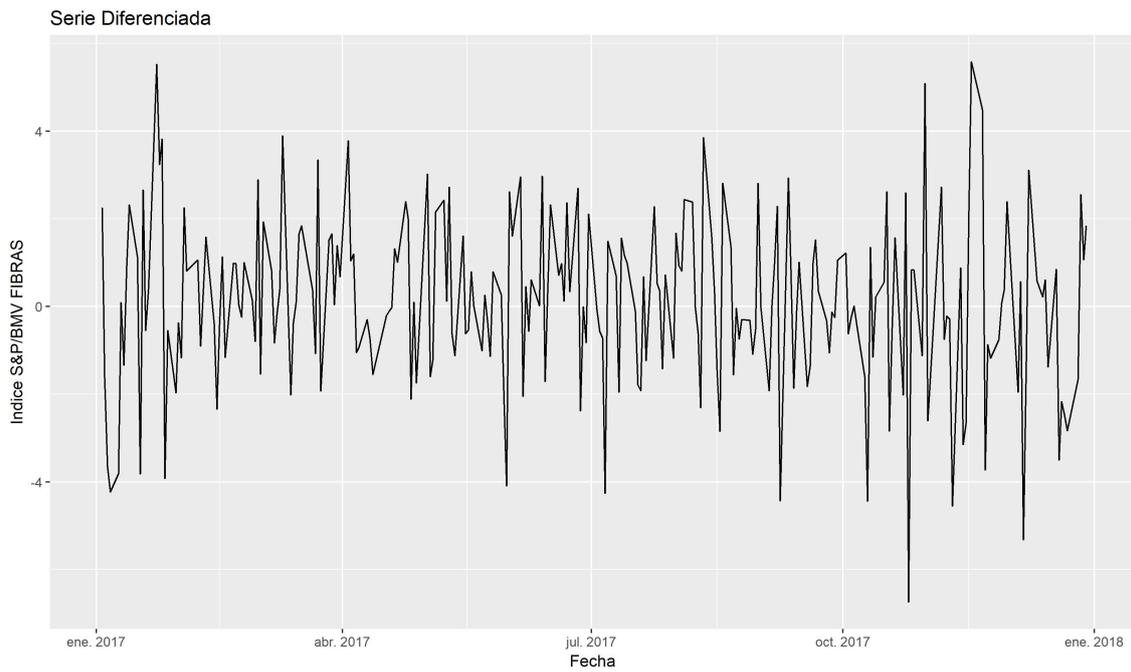


Figura 3.8: Índice S&P/BMV FIBRAS sin tendencia

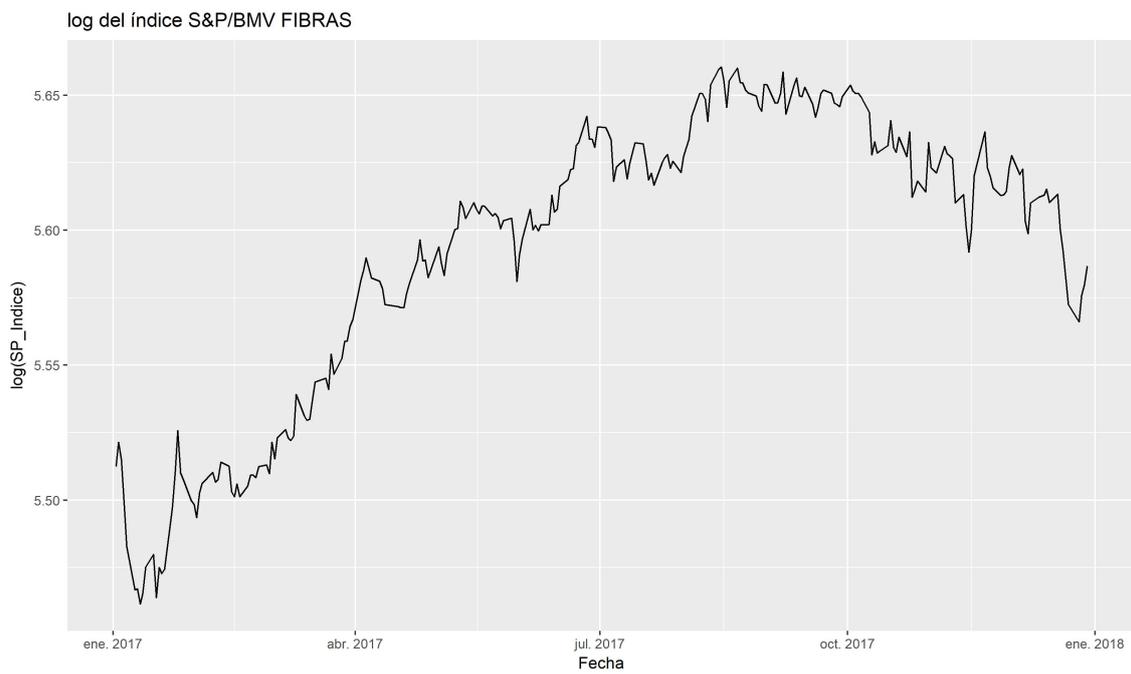


Figura 3.9: log del Índice S&P/BMV FIBRAS

No es de sorprender que la serie tenga la misma forma, únicamente está en una escala diferente. Al observar la serie diferenciada del logaritmo del índice *S&P/BMV FIBRAS* (fig. 3.10), aunque tengan también la misma forma, la amplitud del rango de valores de la serie diferenciada es menos grande. Por lo tanto, se utilizó esta serie para continuar con el modelado.

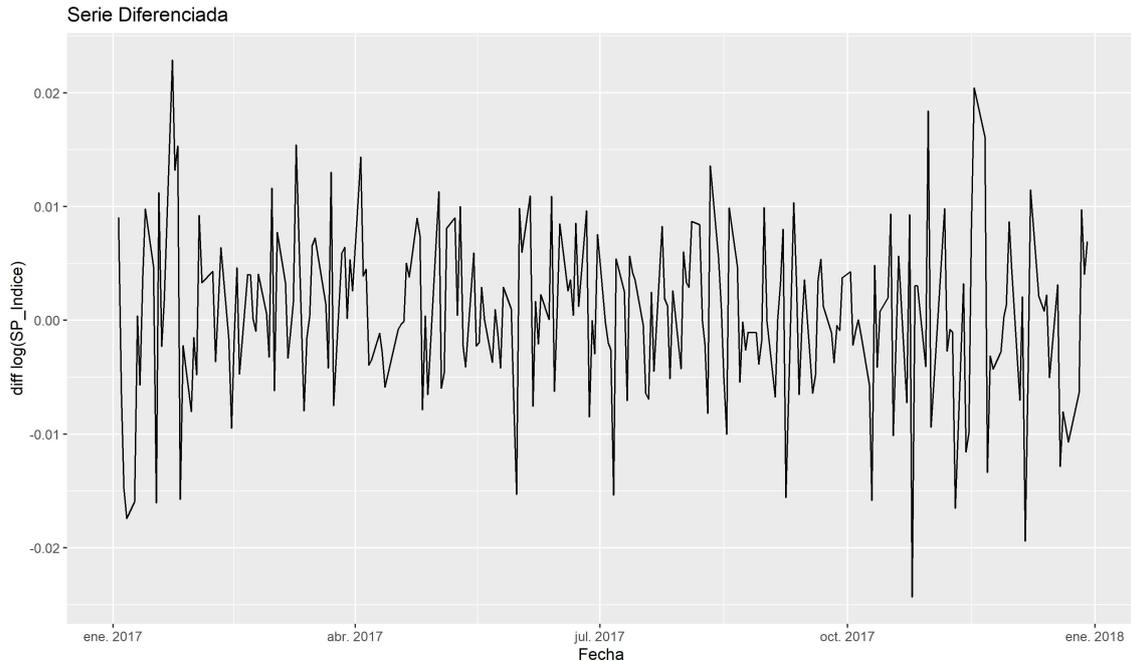


Figura 3.10: Serie diferenciada del log(S&P/BMV FIBRAS)

Hasta este punto, sólo se ha trabajado en quitar la componente estacional a través de alguno de los métodos o filtros conocidos, en este caso, se filtró usando el operador diferencia. Sin embargo, aún no se ha intentado verificar que la serie no tenga alguna componente estacional, aunque dicha componente no se vea de manera clara en la gráfica de la serie.

De forma eurística, se sabe que las series que se miden de manera diaria, generalmente presentan una componente estacional con periodicidad de 7 días, es decir, la estacionalidad es semanal.

Aunque este tipo de componente se estudia con detalle a través del análisis espectral de la serie, se utilizará el periodograma para poder cuantificar la estacionalidad de la serie en caso de tenerla. El periodograma es de las herramientas más simples utilizadas en el análisis espectral para proporcionar un primer acercamiento para conocer la naturaleza de la componente estacional.

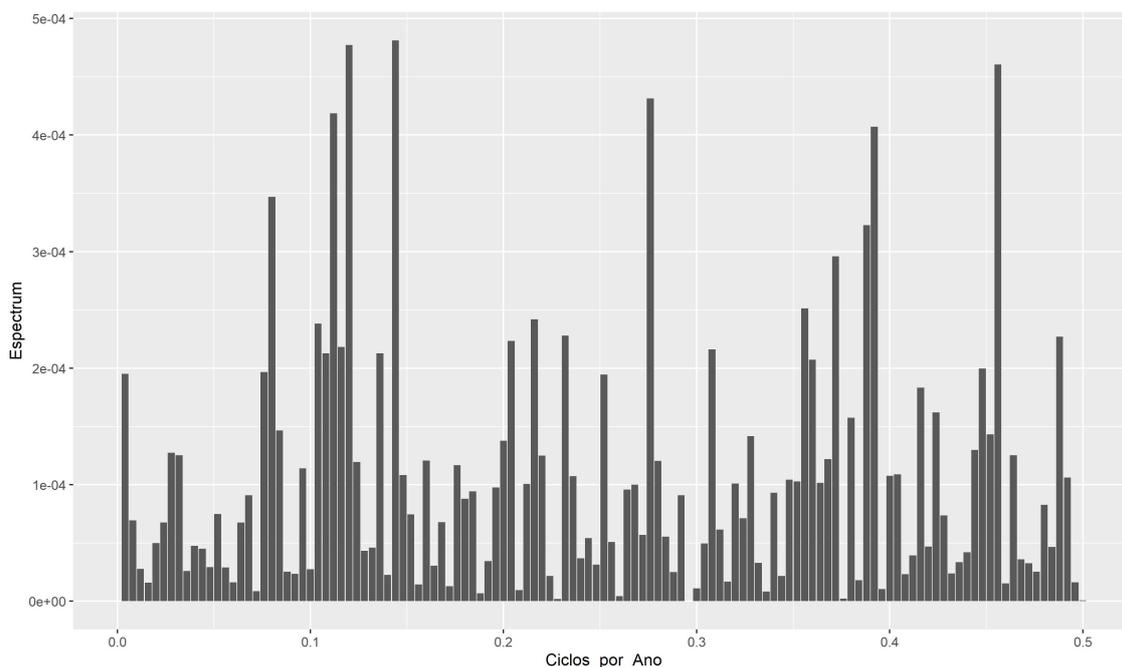


Figura 3.11: Periodograma de la serie diferenciada del log(S&P/BMV FIBRAS)

El periodograma se calculó sobre la serie sin tendencia, debido a que en teoría, de estar presente, aún tiene esta el efecto estacional. De la gráfica del periodograma se observa que hay varios ciclos que tienen una frecuencia alta, estos son candidatos para aproximar la frecuencia de la serie, por eso, se identificó cuál de estos ciclos es que tuvo mayor frecuencia.

| Ciclos | Frecuencia | Ciclos por Año | Frecuencia de la Serie |
|--------|------------|----------------|------------------------|
| 0.144  | .0004812   | 52.56          | 6.94                   |
| 0.120  | .0004773   | 52.56          | 8.33                   |
| 0.456  | .0004604   | 52.56          | 2.19                   |
| 0.276  | .0004313   | 52.56          | 3.63                   |
| 0.112  | .0004186   | 52.56          | 8.93                   |

Cuadro 3.5: Estadísticas del periodograma

La tabla anterior, muestra los 5 ciclos con mayor frecuencia, ordenados de mayor a menor de acuerdo a la frecuencia del ciclo. La primera columna indica de qué tamaño es el ciclo con respecto a un año, por ende, al sacar el número de ciclos presentes dentro de la serie se pudo calcular la frecuencia de la serie.

Por lo tanto el ciclo con mayor frecuencia es aquel que induce una periodicidad de

6.944  $\sim$  7. Lo cual, aporta una mayor certeza de que la serie tiene una componente estacional con periodicidad de 7 días, por lo tanto, se volvió a diferenciar la serie, sin embargo, esta vez con un *rezago* de 7 periodos.

De esta manera, se tiene la siguiente serie de tiempo, la cual, proviene de la serie del índice *S&P/BMV FIBRAS*. Esta última se diferenció una vez con un *rezago* = 1 para filtrar la componente de tendencia y se diferenció una vez con un *rezago* = 7 para filtrar la estacionalidad.

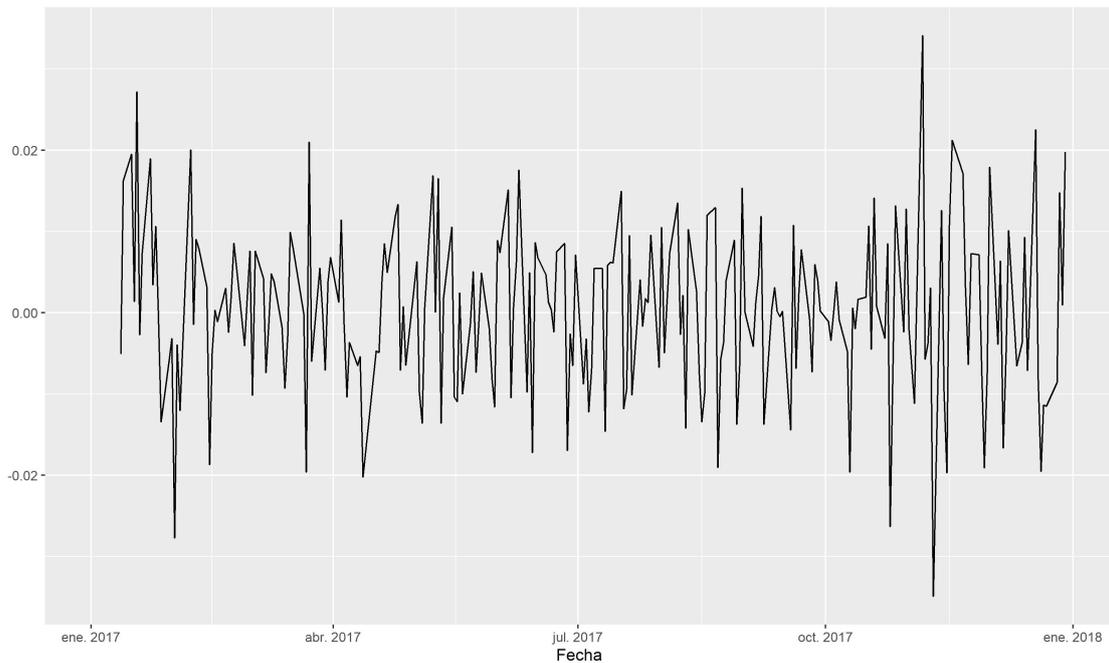


Figura 3.12:  $\log(\text{S\&P/BMV FIBRAS})$  sin tendencia y sin estacionalidad

La serie anterior, es la candidata a ser modelada, sin embargo, primero se debe validar que la serie sea estacionaria. Por lo cual, aplicando la prueba de *Dickey-Fuller Aumentada* a la serie, se tienen los siguientes resultados.

| Estadística | $\Pr(>  t )$            |
|-------------|-------------------------|
| -10.755     | $p\text{-value} < 0.01$ |

Cuadro 3.6: Prueba de *Dickey-Fuller Aumentada* para la serie diferenciada del índice *S&P/BMV FIBRAS*

Como el  $p\text{-value}$  es menor que .01, dicho  $p\text{-value}$ , es menor que cualquier nivel de significancia habitual. Por lo tanto, se rechaza la hipótesis nula en favor de la alternativa,

de que la serie no tiene presencia de raíces unitarias y por lo tanto, se puede suponer que la serie es estacionaria.

### 3.3.3. Especificación de los parámetros del modelo SARIMA para el índice S&P/BMV FIBRAS

Debido a que ya se obtuvo una serie de tiempo estacionaria a partir de la serie en estudio. Es sobre la serie estacionaria que se realizó el análisis de las gráficas de autocorrelación y autocorrelación parcial, con la finalidad de poder tener varios modelos candidatos, los cuales se pusieron a prueba de acuerdo a los supuestos estadísticos que deberían cumplir, así como la calidad de sus pronósticos.

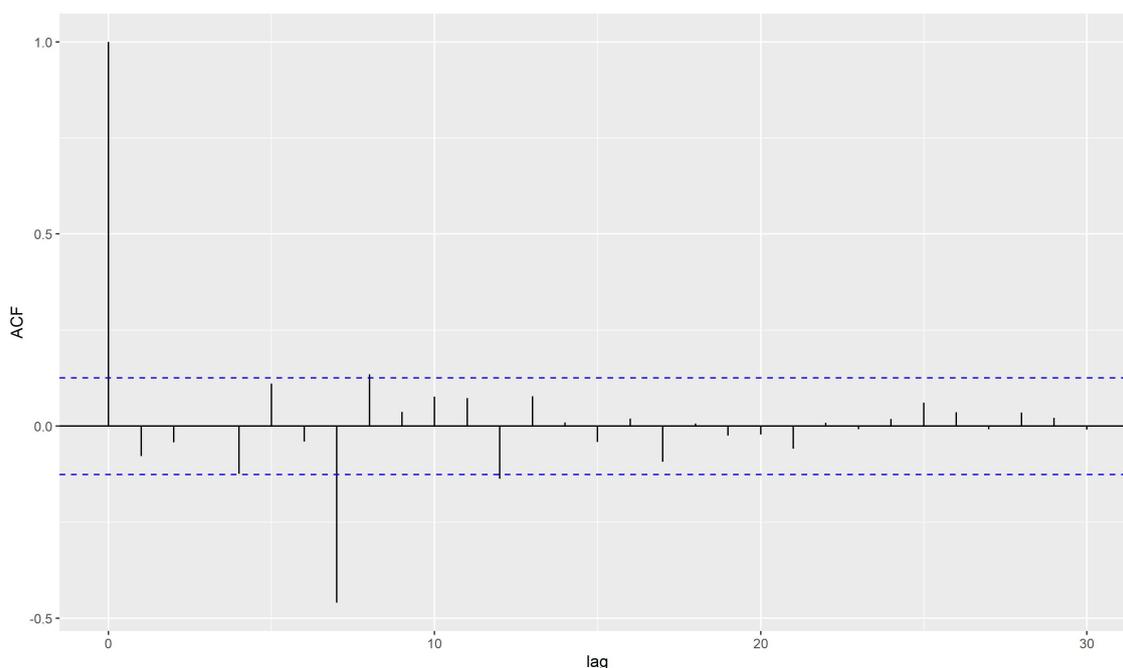


Figura 3.13: ACF de la serie diferenciada del log(S&P/BMV FIBRAS)

Aparentemene, hay presencia de autocorrelación con un retraso de 4 periodos o 8 periodos, aunque sea por muy poco que se salen de las bandas de confianza. Además, en el lag 7, hay una autocorrelación estacional. Por lo cual, el parámetro  $q$  de la parte  $MA$ , del modelo  $ARMA$  podría ser  $q = 4$  o  $q = 8$  y  $Q = 1$ .

Por otro lado, analizando la función de autocorrelación parcial, se observa que posibles valores para el parámetro  $p$ , de la parte  $AR$  del modelo  $ARMA$  podría ser  $p = 4$  o

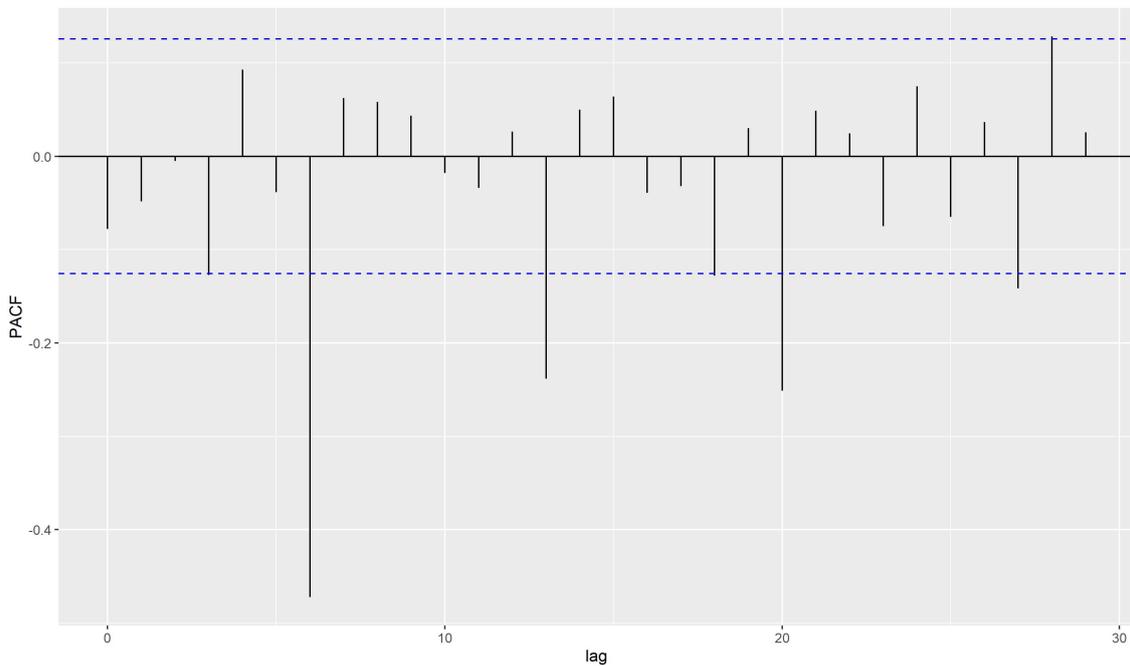


Figura 3.14: PACF de la serie diferenciada del log(S&P/BMV FIBRAS)

$p = 0$ . Además, en este caso hay presencia fuerte estacional hasta 4 periodos estacionales atrás, aunque se podrían considerar únicamente tres, de esta forma  $P$  podría ser  $P = 4$  o  $P = 3$ .

Por lo tanto, se proponen los siguientes modelos para modelar la serie. Dado que se usó una vez el operador diferencia con un *rezago* = 1 para quitar tendencia y se usó de nuevo el operador diferencia pero con *rezago* = 7 para quitar estacionalidad, realmente se está haciendo uso de un modelo SARIMA. Los modelos propuestos se presentan a continuación, después se evaluará si los modelos cumplen los supuestos estadísticos, y para los que sean candidatos, se usarán criterios de evaluación para definir al modelo elegido.

1.  $SARIMA(4, 1, 4)(4, 1, 1)_7$
2.  $SARIMA(0, 1, 4)(4, 1, 1)_7$
3.  $SARIMA(4, 1, 8)(4, 1, 1)_7$
4.  $SARIMA(4, 1, 4)(3, 1, 1)_7$
5.  $SARIMA(0, 1, 4)(3, 1, 1)_7$

6.  $SARIMA(4, 1, 8)(3, 1, 1)_7$ 

De los modelos anteriores, el **modelo 1** y el **modelo 3**, se descartaron debido a que al realizar el ajuste, ninguno de los dos modelos resultantes era estacionario.

### 3.3.4. Validación de supuestos estadísticos

Como se mencionó previamente, a partir de los modelos propuestos, se intentará descartar los modelos que no cumplan alguno o varios de los supuestos necesarios para que el modelo ajustado tenga validez. Primero se validará que los residuales correspondientes a cada modelo, no estén correlacionados.

#### 3.3.4.1. Residuales no correlacionados

A continuación se muestran las funciones de autocorrelación de los residuales de los modelos 2, 4, 5 y 6. Es importante recordar por qué es de interés analizar las funciones de autocorrelación. De los primeros supuestos que se siguen en los modelos ARIMA es que los residuales son un *ruido blanco*, es decir, que el proceso estocástico relacionado con estos tiene media cero y además, las variables aleatorias a cualquier tiempo son no correlacionadas. Por lo tanto, la función de autocorrelación debe ser cercana a cero a cualquier lag. En particular, si además se impone el supuesto de que es un proceso gaussiano, se pueden contruir las bandas de confianza, para poder dar un grado de confiabilidad en las autocorrelaciones de los residuales.

Por lo tanto, de forma gráfica, al observar las figuras 3.15, 3.16, 3.17 y 3.18, es claro que ninguno de los residuales asociados a los modelos aparentan tener correlación serial, sin embargo, los residuales de los modelos 2 y 6 (figuras 3.15 y 3.18 respectivamente) son los que están más alejados de las bandas de confianza. En particular, los residuales del modelo 6, figura 3.18, son los que están totalmente acotados por las bandas de confianza.

El hecho de que en general, los residuales de los modelos tengan un comportamiento muy parecido a la función de autocorrelación parcial de un ruido blanco es un buen indicio. Sin embargo, es posible proveer de un dato más duro para poder suponer que en efecto, los residuales de los modelos sean no correlacionados.

Para lo anterior, se hizo uso de la prueba de *Ljung-Box* y de la prueba de *Box-Pierce*,

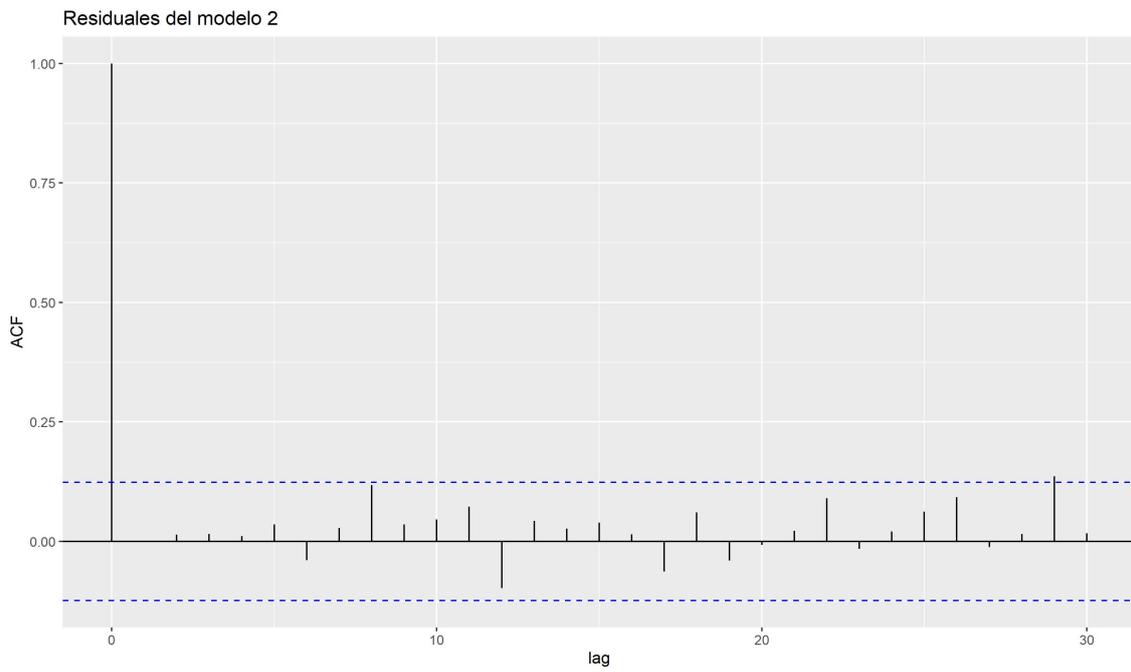


Figura 3.15: ACF de los residuales del modelo 2

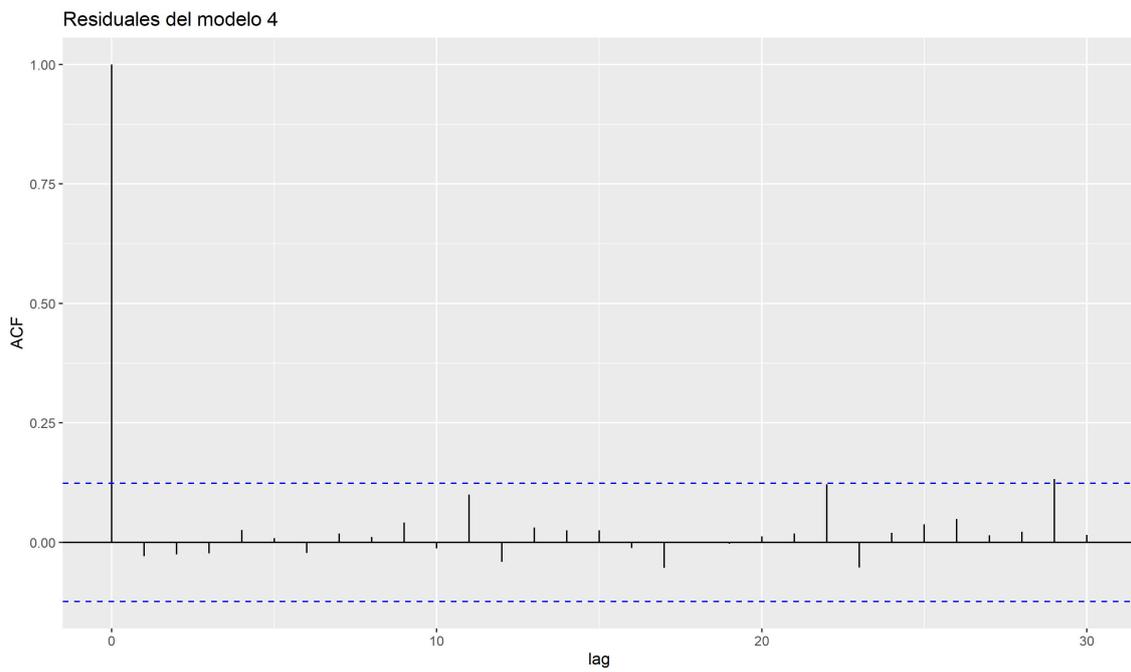


Figura 3.16: ACF de los residuales del modelo 4

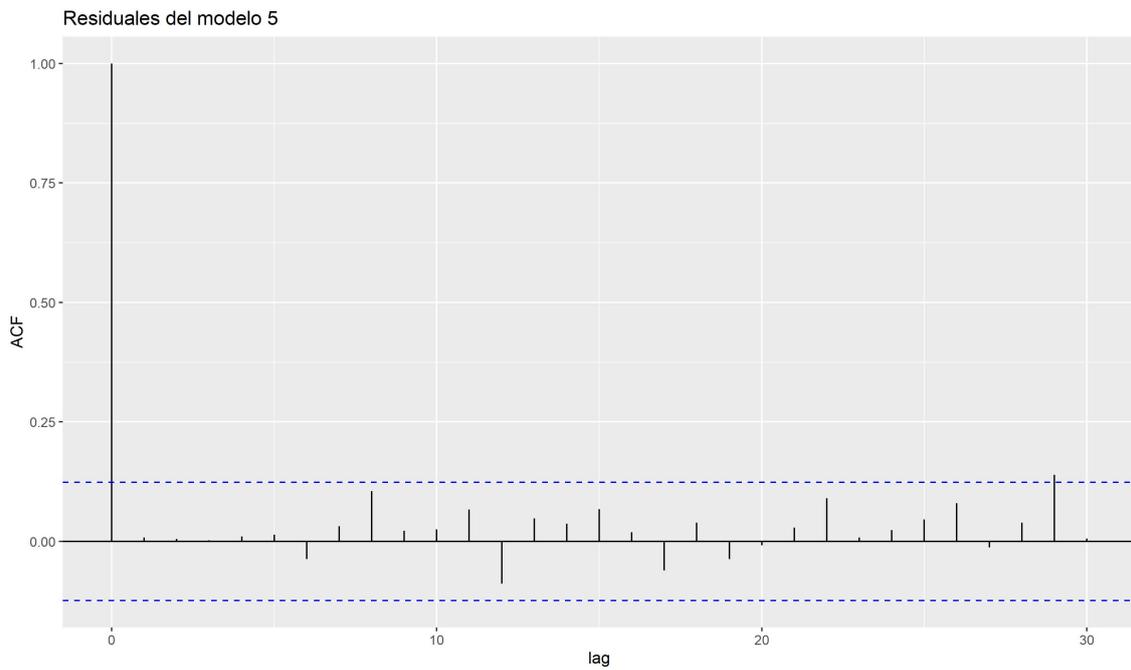


Figura 3.17: ACF de los residuales del modelo 5

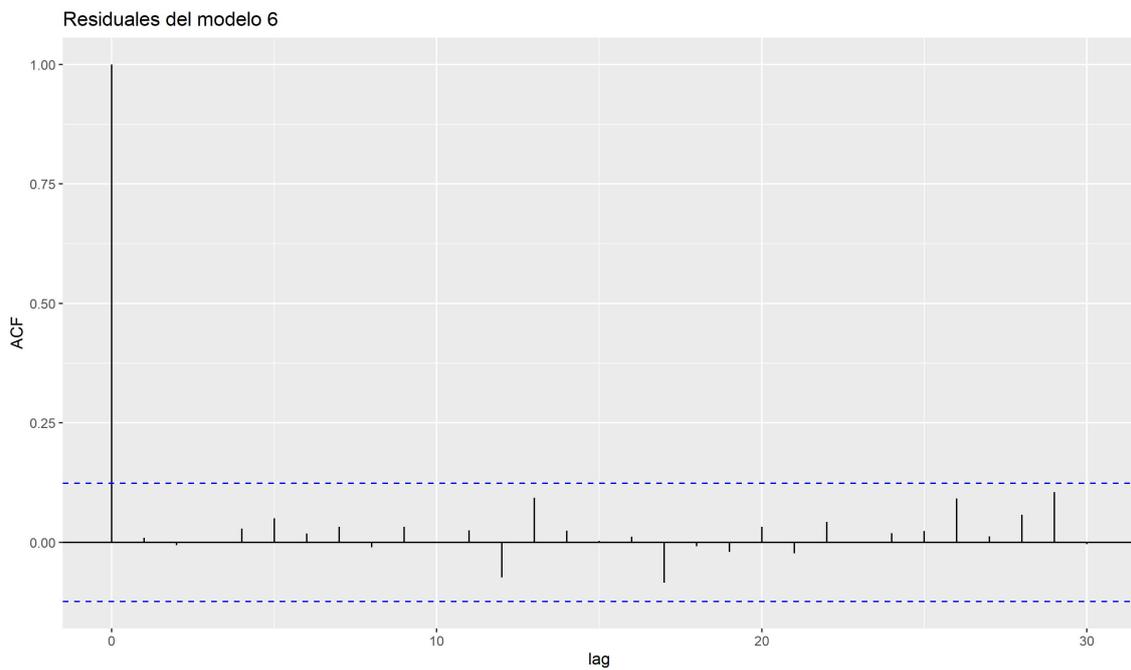


Figura 3.18: ACF de los residuales del modelo 6

ambas pruebas tienen como hipótesis nula que la correlación es igual a cero para cualquier lag. En el **Anexo** se puede consultar el detalle de las tablas que resumen los resultados de dichas pruebas. Las gráficas que se muestran a continuación, resumen la información de dichas tablas.

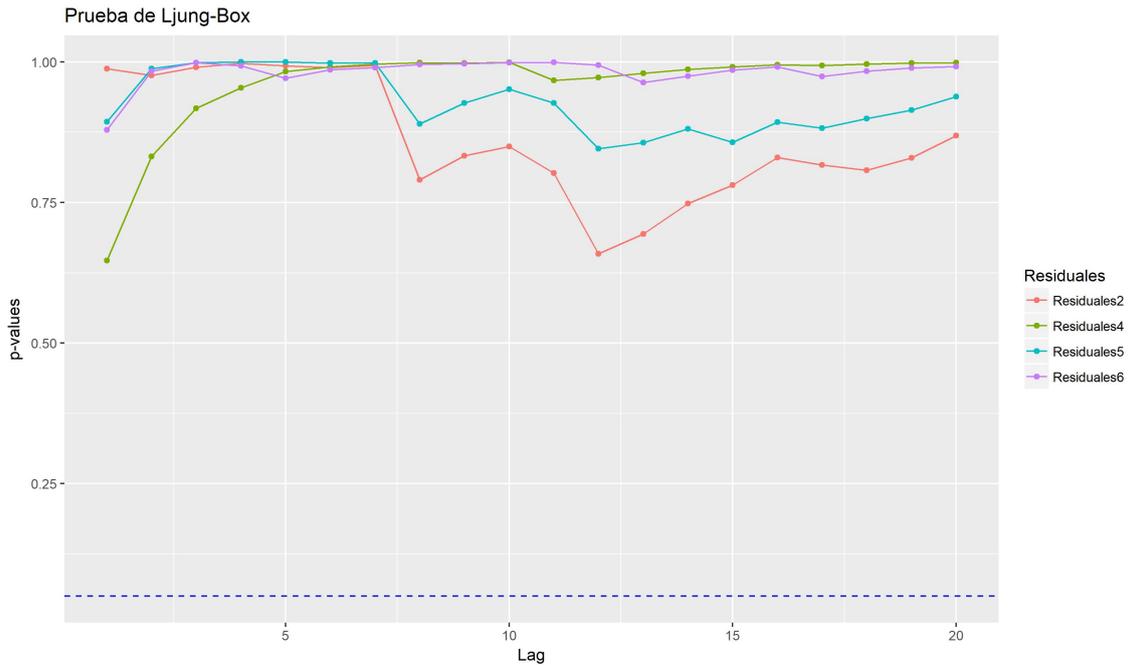


Figura 3.19: Prueba de Ljung-Box para los residuales de los modelos

Para ambas pruebas se evaluó la correlación de los residuales hasta un lag de 20. En las gráficas de ambas pruebas se comparan los  $p$ -valores obtenidos a distintos *rezagos* vs el nivel de significancia  $\alpha = .05$ .

Con lo cual, se concluye que de acuerdo a los resultados vistos en ambas pruebas, no hay evidencia estadística para suponer que los errores tienen correlación serial. Por lo cual, se puede suponer que los errores no están correlacionados.

Además, aunque los residuales de los cuatro modelos pasan las pruebas de Ljung-Box y de Box-Pierce, es interesante que los residuales del *modelo 6*, son lo que tiene un comportamiento más consistente considerando a sus  $p$ -values a lo largo de los retrasos de 1 a 20 en ambas pruebas.

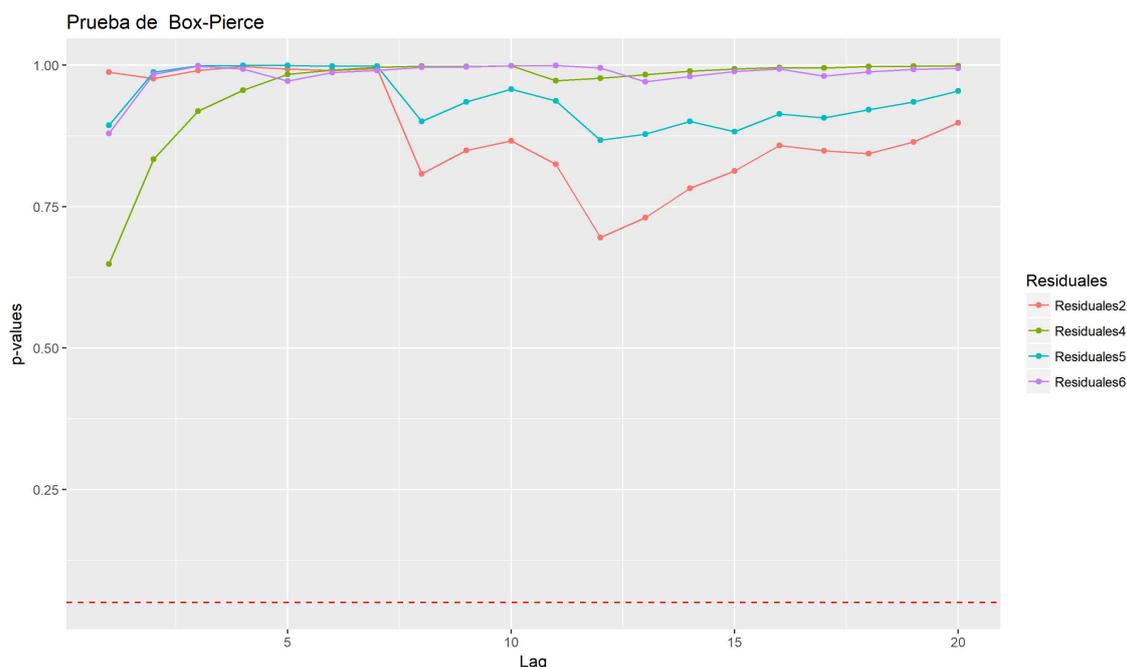


Figura 3.20: Prueba de Box-Pierce para los residuales de los modelos

3.3.4.2. Media cero de los residuales

Por construcción, se debería de cumplir que la media de los residuales para todos los modelos sea igual a cero. Por lo tanto, se usará una prueba *t*, la cual, tiene como hipótesis nula de que los residuales tienen media  $\mu = 0$ . A continuación, se presenta una tabla que resume los resultados de dicha prueba.

| p-value  |          |          |          |
|----------|----------|----------|----------|
| Modelo 2 | Modelo 4 | Modelo 5 | Modelo 6 |
| 0.8705   | 0.924    | 0.875    | 0.999    |

Cuadro 3.7: Prueba  $\mu = 0$  para los residuales de los modelos propuestos

Como el *p-value* de cada uno de los modelos es mayor que cualquier nivel de significancia habitual, es decir  $\alpha = .01, .05$ . Hay evidencia estadística suficiente para poder suponer que la media de los residuales es igual a cero para cualquiera de los cuatro modelos. Adicionalmente, destaca el *p-value* del modelo 6, ya que es el que “mejor” pasa la prueba.

### 3.3.4.3. Varianza constante de los residuales

Recordando lo que se especificó con detalle en el *Capítulo 2*. Se sabe por trabajos de *Box & Jenkins*, que aunque los residuales no sean correlacionados en el tiempo, el cuadrado de los residuales puede estar correlacionado, por lo tanto, se estudia la función de autocorrelación del cuadrado de los residuales, ya que esta puede ayudar a identificar series de tiempo no lineales. Además, tanto las funciones de autocorrelación como de autocorrelación parcial del cuadrado de los residuales son útiles para identificar algún efecto *GARCH*.

Debido a lo anterior, se muestran a continuación las funciones de autocorrelación del cuadrado de los residuales de los cuatro modelos propuestos.

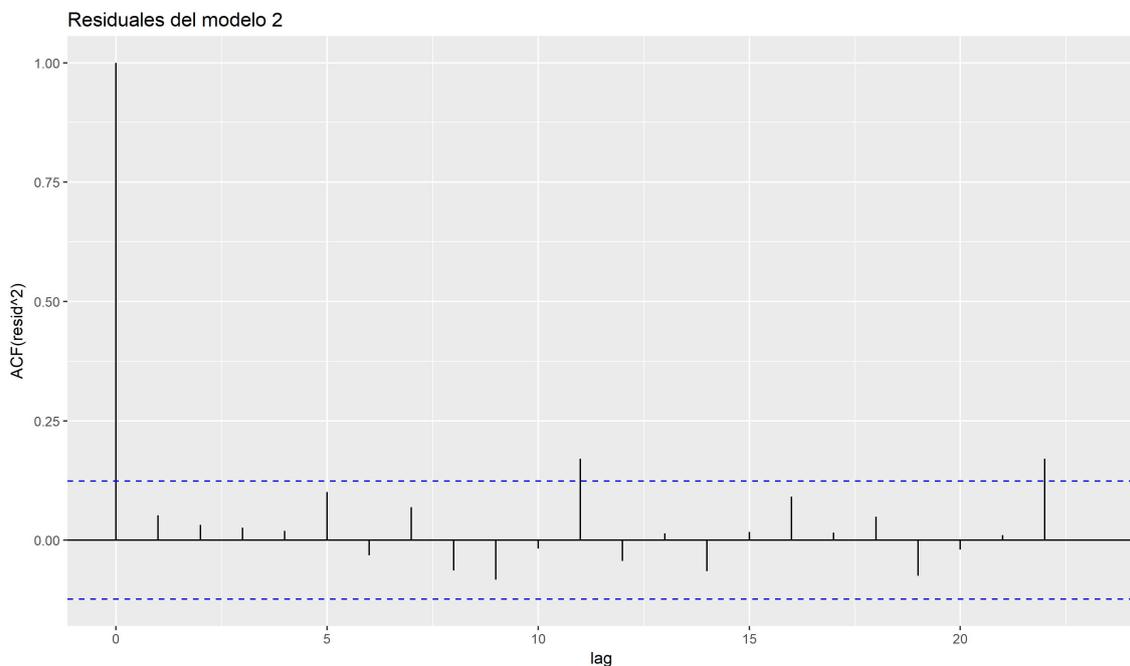


Figura 3.21: ACF del cuadrado de los residuales del modelo 2

Los modelos 2 y 5 son los que aparentemente tienen un cierto grado de autocorrelación en el cuadrado de sus residuales, en particular el modelo 2 es el que tiene una autocorrelación más marcada en los retrasos a tiempo 11 y 22. Por otro lado, para los modelos 4 y 6, la función de autocorrelación, para algunos retrasos, apenas queda dentro de las bandas de confianza. Como no es claro el suponer o no, la presencia de un efecto *ARCH*, se hará uso de pruebas formales para evaluar la hipótesis de que la serie no tiene efectos *ARCH*, es decir, que la varianza de los residuales es incondicional del tiempo y

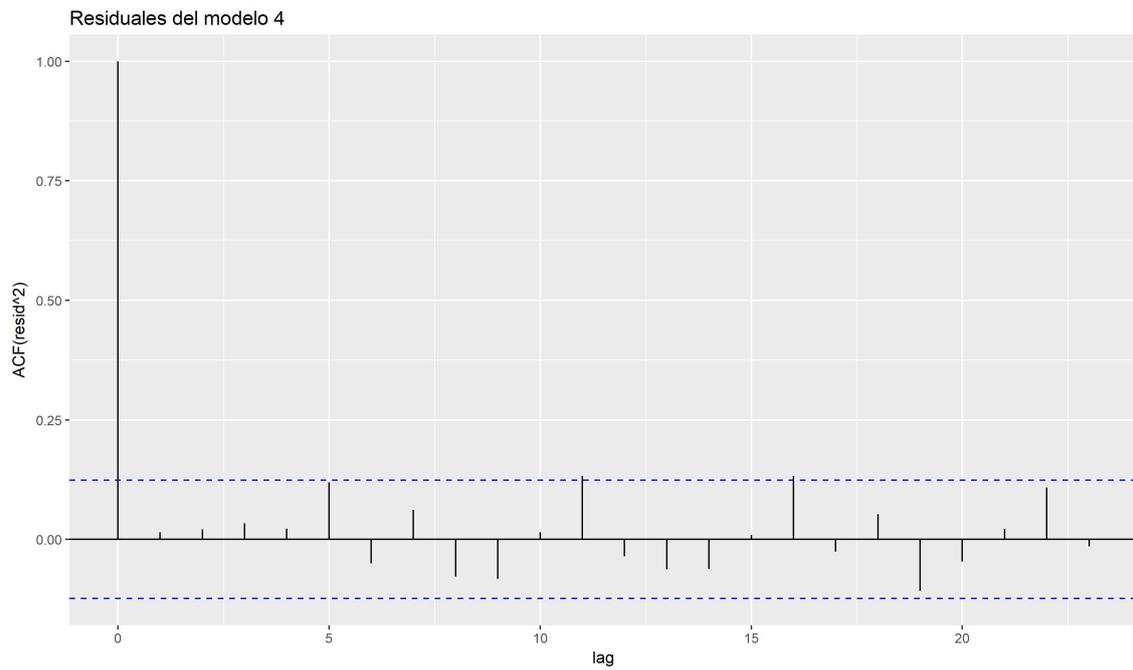


Figura 3.22: ACF del cuadrado de los residuales del modelo 4

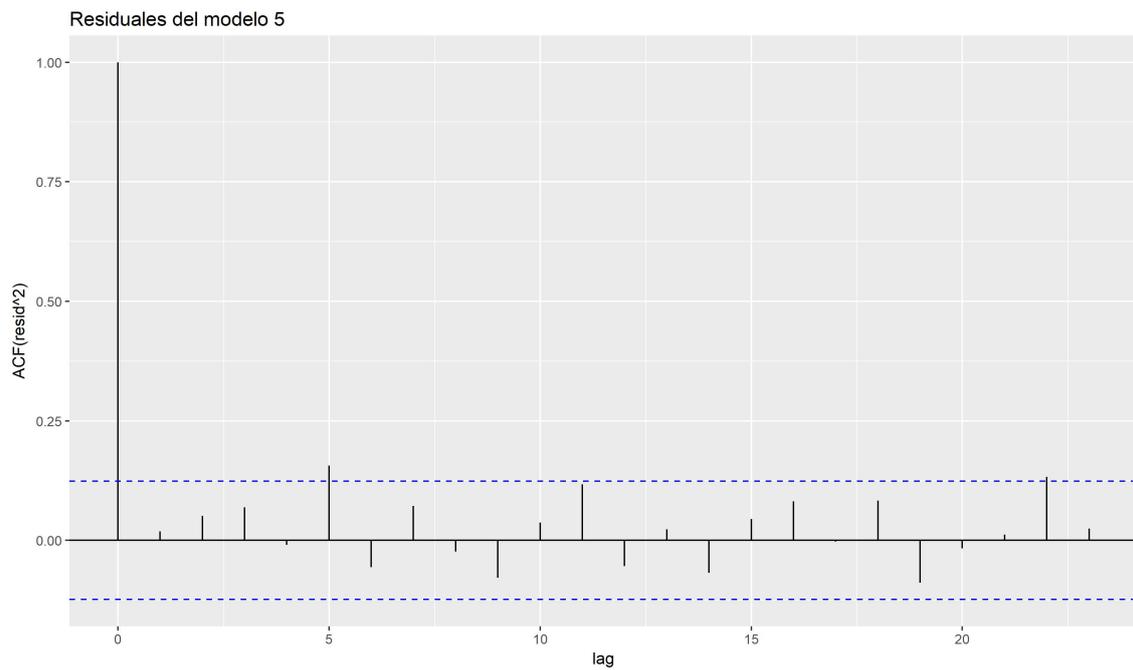


Figura 3.23: ACF del cuadrado de los residuales del modelo 5

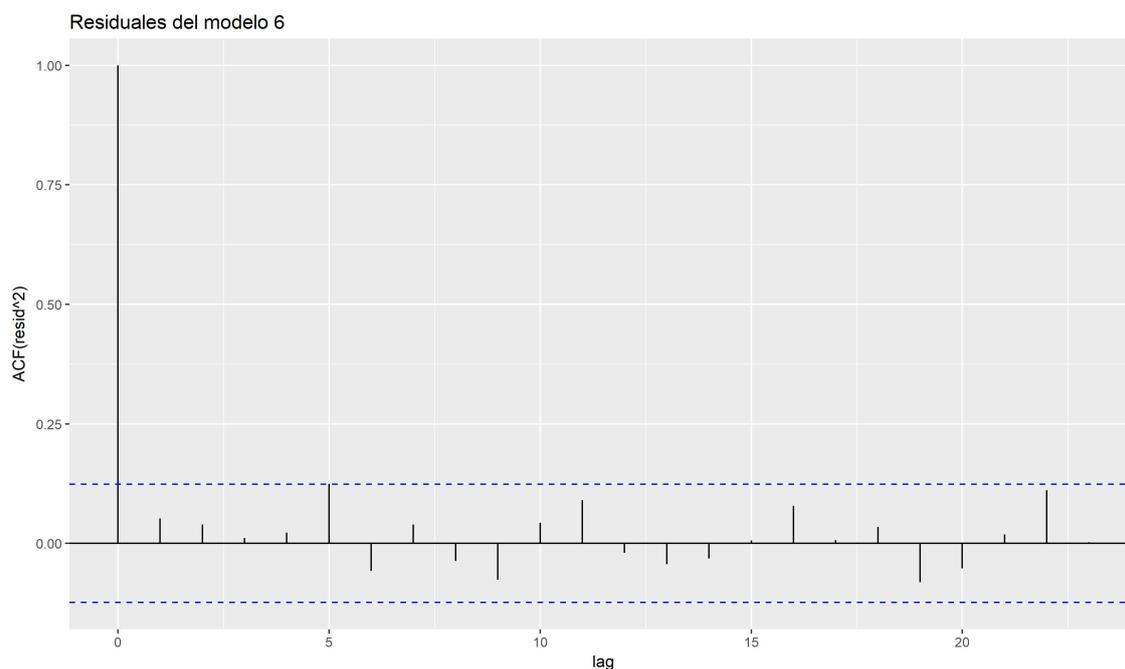


Figura 3.24: ACF del cuadrado de los residuales del modelo 6

por tanto, la varianza es constante.

A continuación, se presenta una gráfica que resume la tabla con los resultados de la prueba de *McLeod-Li*, cuya hipótesis nula es que los datos no tienen algún efecto *ARCH*. La tabla que se menciona se puede consultar en el *ANEXO*.

La prueba sobre los residuales de los modelos propuestos a cualquier *lag* entre 1 y 20, llevan a concluir que no hay evidencia estadística suficiente para suponer que hay presencia de efectos *GARCH*. Sin embargo, se observa que para el *rezago* = 22, el *p-value* de la prueba para los residuales de los modelos 2, 4 y 5, por poco es mayor que el nivel de significancia  $\alpha = .10$ . Esto está en línea con las gráficas de la función de autocorrelación del cuadrado de los residuales de los modelos 2 y 5, sin embargo, como se dijo previamente, no hay evidencia estadística suficiente para suponer que la varianza no es constante.

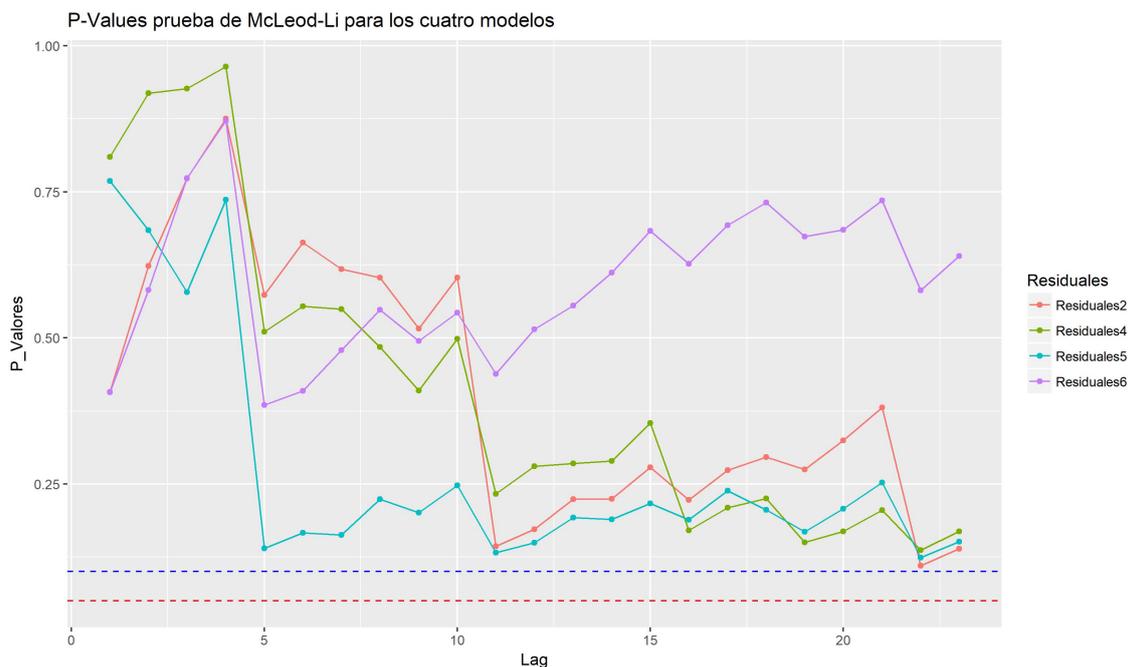


Figura 3.25: Prueba de McLeod-Li para los residuales.

#### 3.3.4.4. Normalidad de los residuales

Para validar que los residuales sigan una distribución normal, se analizaron las figuras 3.26, 3.27, 3.28 y 3.29, las cuales, corresponden a los histogramas de los residuales.

Es claro que los histogramas de los residuales se aproximan bastante a la función de densidad de una distribución normal. De los cuatro histogramas, el histograma de los residuales asociados al modelo 6, figura 3.29, es el que tiene una “mayor” diferencia, de igual manera, el histograma de los residuales asociados al modelo 2, figura 3.26, también tiene discrepancias notorias. Para tener una comparación gráfica un poco menos subjetiva, se analizaron las figuras 3.30, 3.31, 3.32 y 3.33, las cuales, corresponden a los *qqplot*'s de los residuales.

De la misma manera, los *qqplot*'s de los residuales son muy cercanos a la recta que se debería de ver en caso de una coincidencia perfecta entre los cuantiles teóricos y los cuantiles muestrales. Además, estos siguen el mismo comportamiento que los histogramas en el sentido de que los que tienen un “peor” comportamiento, son los *qqplot*'s relativos a los modelos 2 y 6, figuras 3.30 y 3.33 respectivamente.

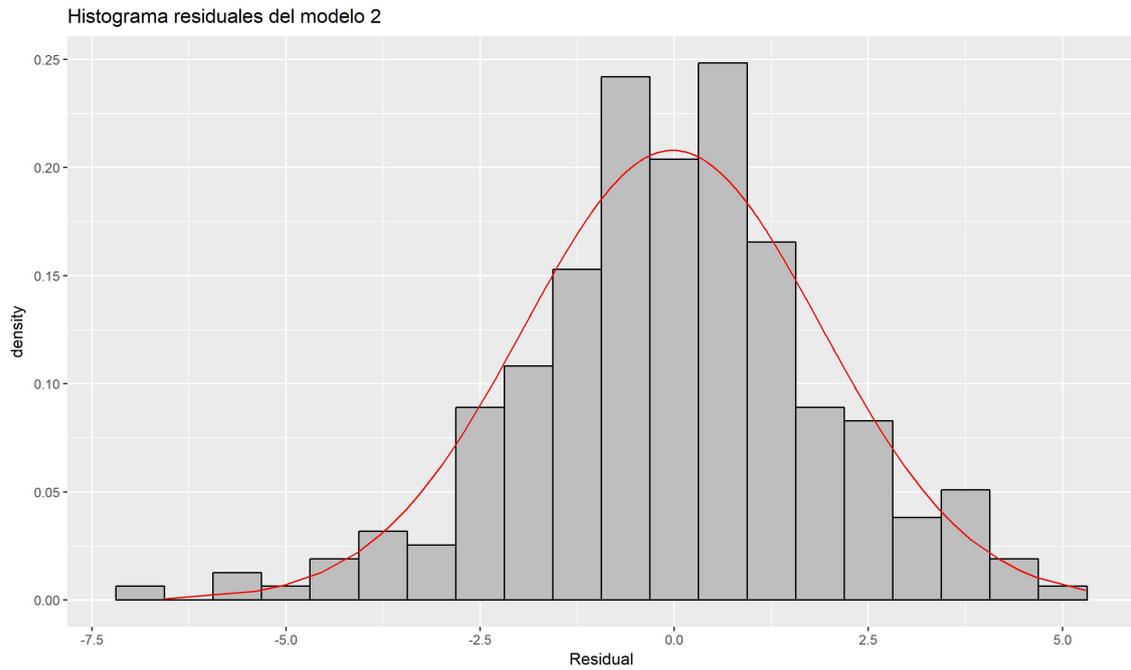


Figura 3.26: Histograma de los residuales del modelo 2

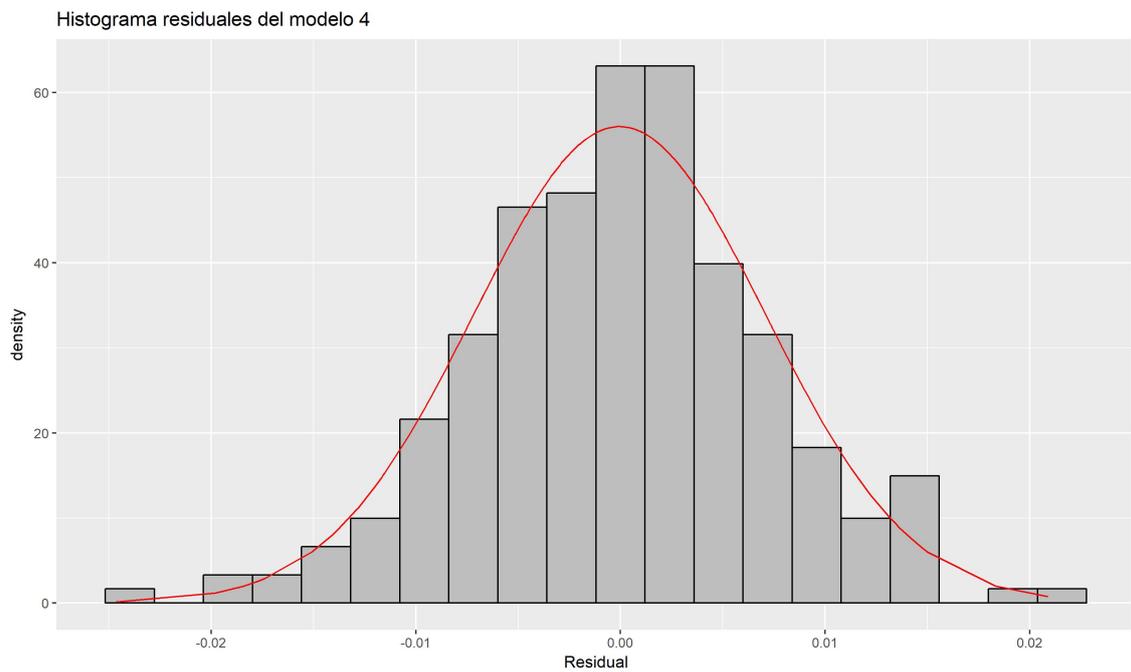


Figura 3.27: Histograma de los residuales del modelo 4

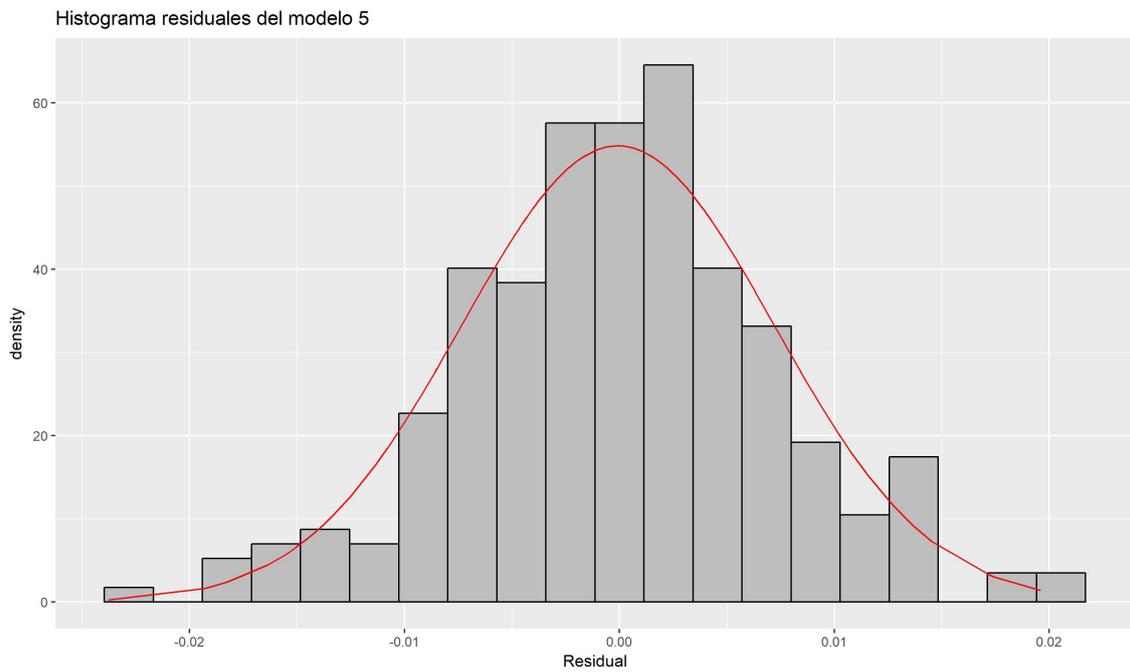


Figura 3.28: Histograma de los residuales del modelo 5

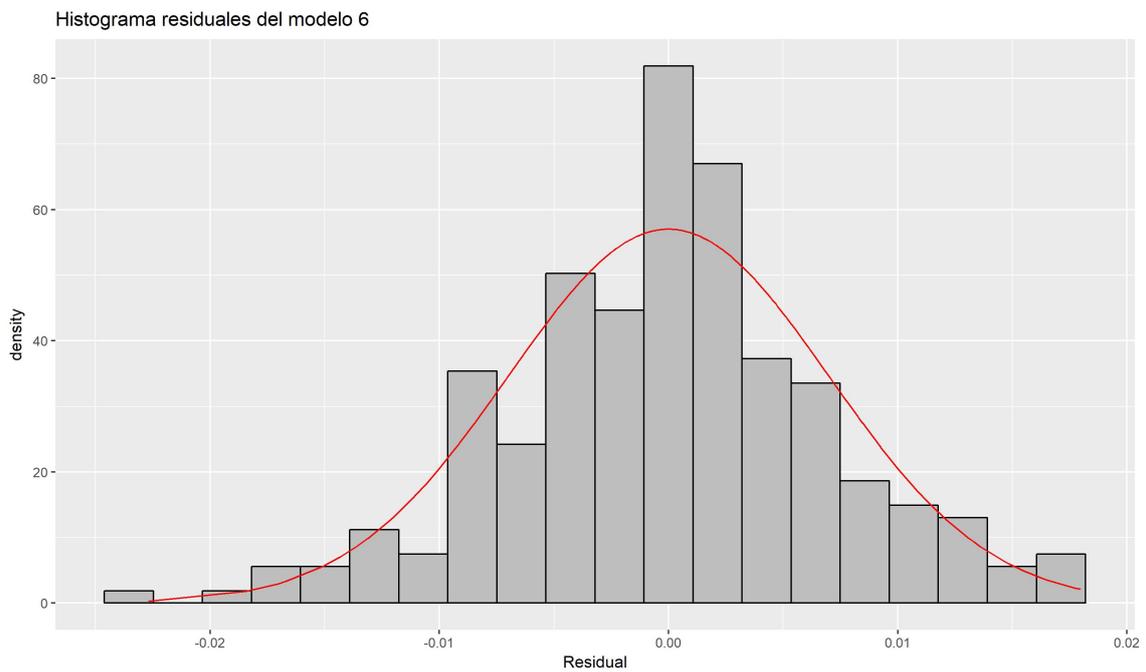


Figura 3.29: Histograma de los residuales del modelo 6

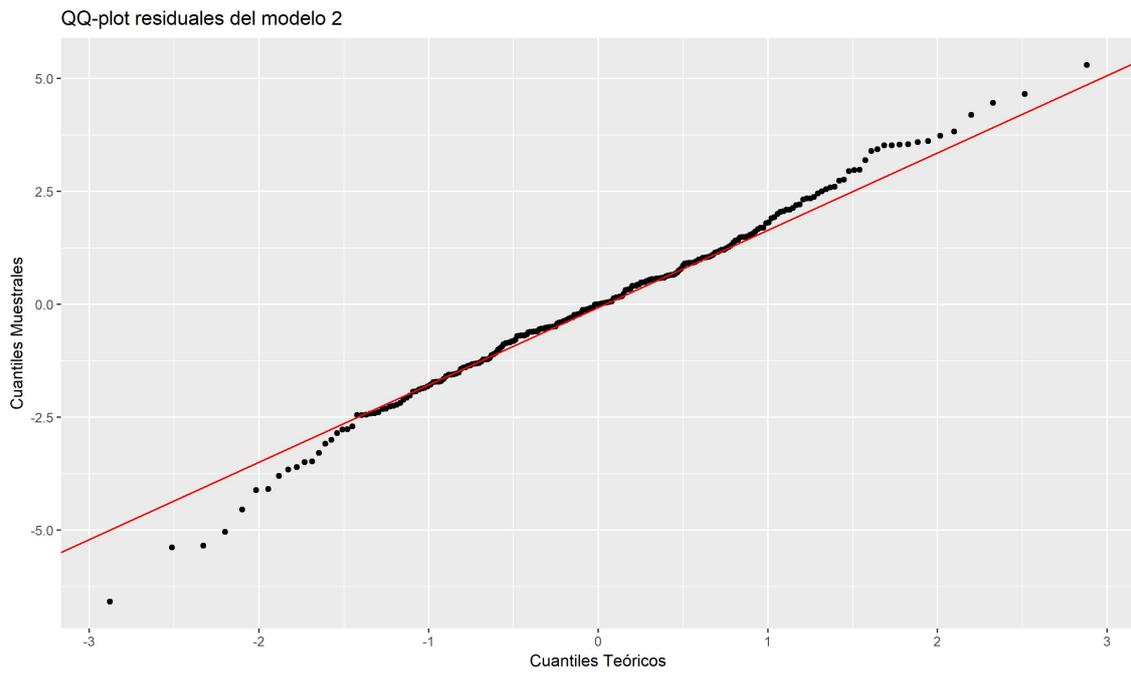


Figura 3.30: QQ-Plot de los residuales del modelo 2

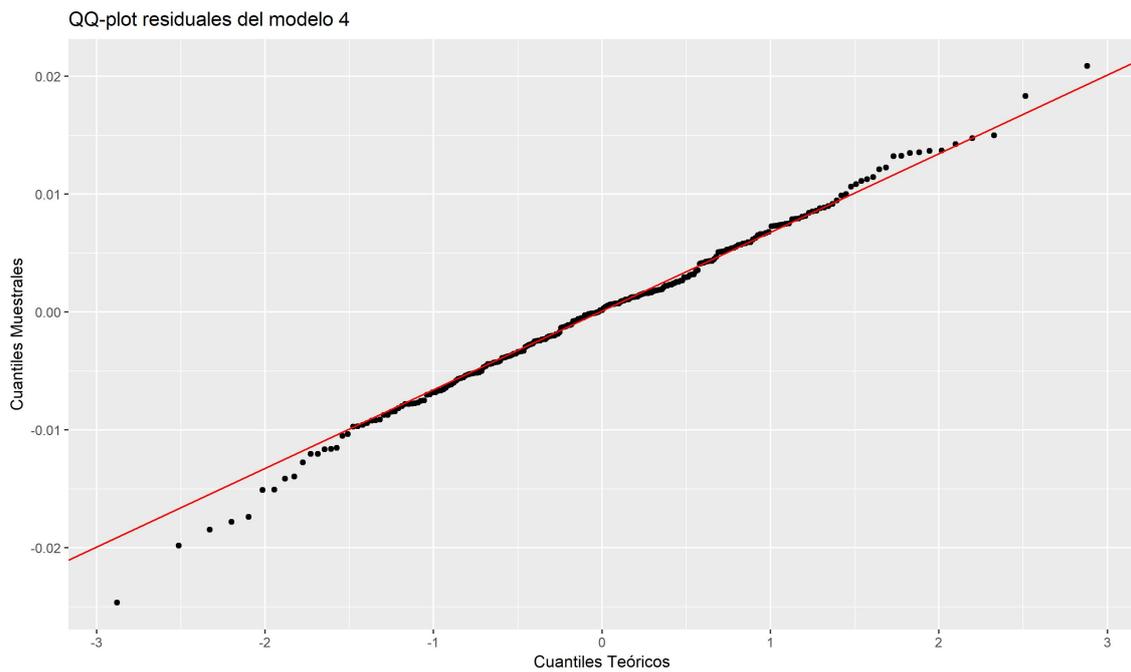


Figura 3.31: QQ-Plot de los residuales del modelo 4

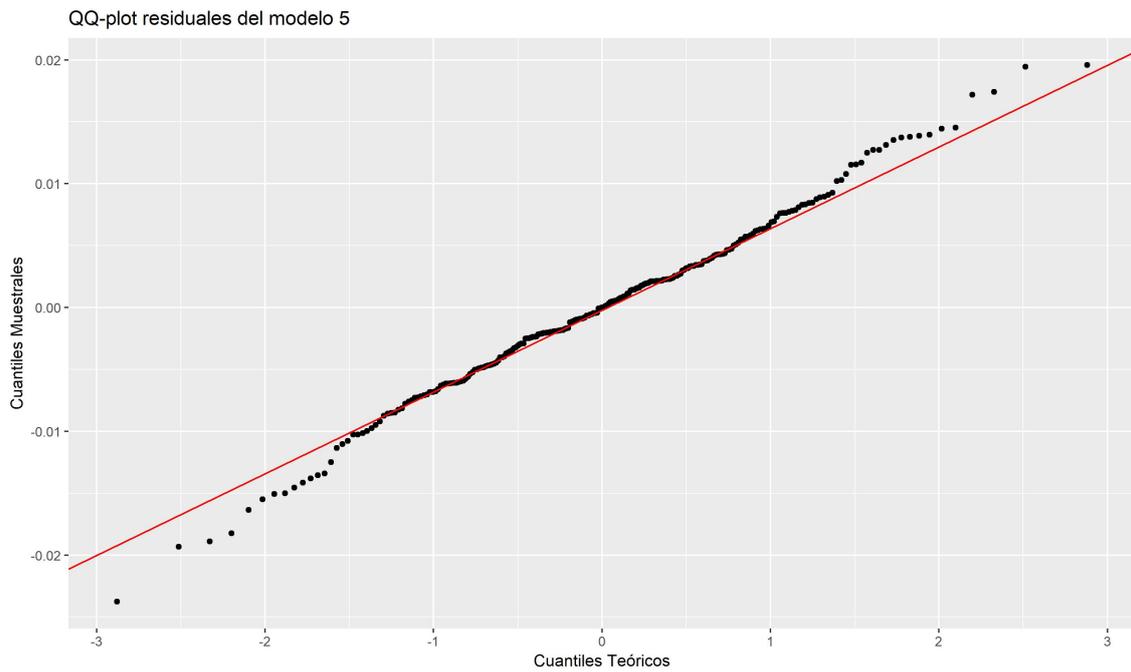


Figura 3.32: QQ-Plot de los residuales del modelo 5

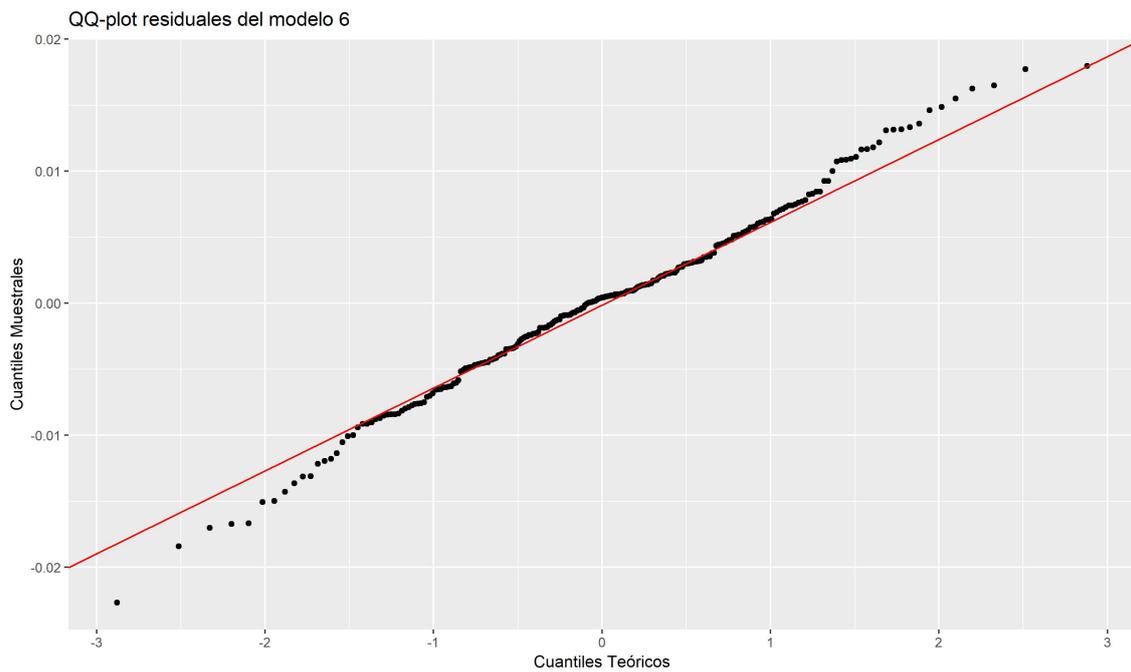


Figura 3.33: QQ-Plot de los residuales del modelo 6

Por último, debido a que aunque tanto los histogramas como los *qqplot's* tienen un comportamiento muy similar al comportamiento teórico, existen ciertas discrepancias que son difíciles de comparar llevando a conclusiones subjetivas. Por lo tanto, se realizaron las pruebas de *Anderson-Darling*, *Jarque-Bera* y *Shapiro-Wilk* sobre los residuales de los modelos propuestos, para poder concluir de una manera más veraz si los residuales siguen o no una distribución normal.

Debido que el análisis visual de los histogramas, puede llevar a conclusiones subjetivas, se evaluará la hipótesis de normalidad en los residuales con las pruebas de *Anderson-Darling*, *Jarque-Bera* y *Shapiro-Wilk*. Los resultados de dichas pruebas se muestran a continuación, así como los coeficientes de curtosis empíricos.

| Residuales      | Curtosis empírica |
|-----------------|-------------------|
| <b>Modelo 2</b> | 3.4852            |
| <b>Modelo 4</b> | 3.4125            |
| <b>Modelo 5</b> | 3.3898            |
| <b>Modelo 6</b> | 3.3181            |

Cuadro 3.8: Curtosis empírica de los residuales de los modelos propuestos

|                         | p-value residuales |          |          |          |
|-------------------------|--------------------|----------|----------|----------|
|                         | Modelo 2           | Modelo 4 | Modelo 5 | Modelo 6 |
| <b>Anderson-Darling</b> | 0.2269             | 0.5558   | 0.193    | 0.0945   |
| <b>Jarque-Bera</b>      | 0.1075             | 0.1895   | 0.319    | 0.4235   |
| <b>Shapiro-Wilk</b>     | 0.2808             | 0.6242   | 0.3487   | 0.2578   |

Cuadro 3.9: Pruebas de normalidad para los residuales de los modelos propuestos

Lo primero que se puede observar, es que en general los resultados son consistentes con las observaciones hechas con los histogramas, así como con los *qqplot's*. Ya que los *p-values*, de los modelos 2 y 6, son los que están más cerca de los niveles de significancia  $\alpha = .05$  y  $\alpha = .10$ .

Además, es interesante que los valores de los *p-values*, para los modelos 2 y 5, son los que tienen menor variabilidad, , en contraste con los *p-values* de los modelos 4 y 6, que tienen mayor variabilidad.

Por lo tanto, para todos los modelos, no se rechaza la hipótesis nula de normalidad de los residuos a cualquier nivel de significancia habitual, es decir, a un nivel de significancia  $\alpha = .01$  y  $\alpha = .05$ .

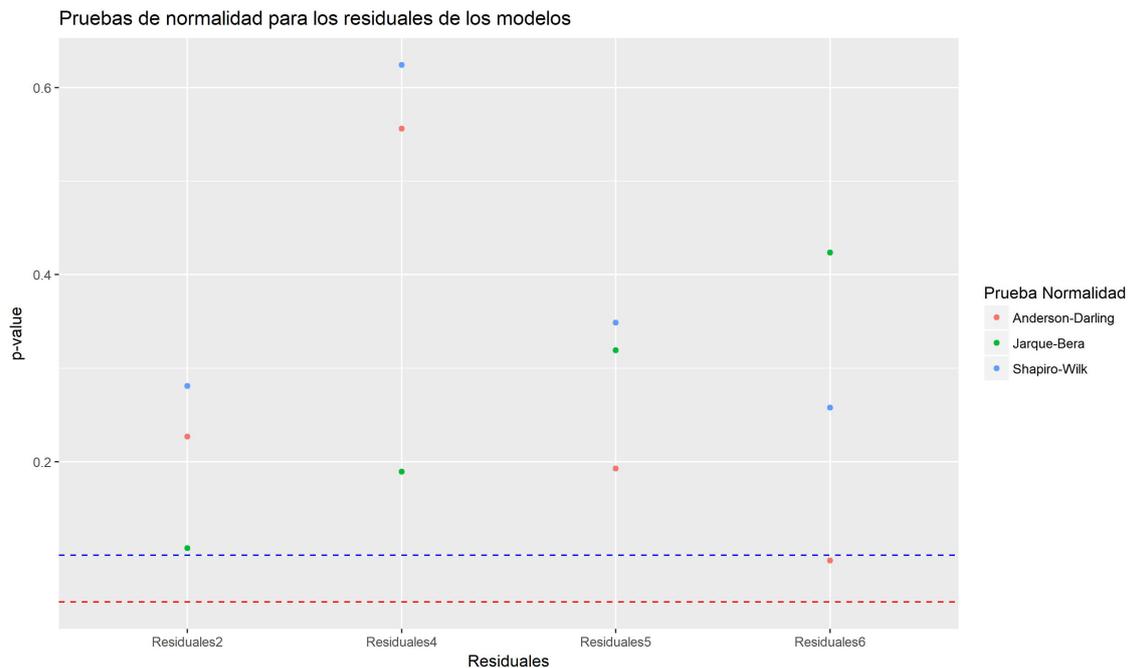


Figura 3.34: Normalidad de los residuales

Por lo tanto, por los resultados expuestos de las evaluaciones de los supuestos estadísticos que se realizaron sobre los residuales de los modelos, se concluye que los cuatro modelos son factibles. Por lo tanto, se utilizarán los criterios de información, así como la calidad de los pronósticos de los modelos para decidir cual modelo se elige.

### 3.3.5. Pronósticos de los modelos

Debido a que no se descartaron modelos por no cumplir alguno de los supuestos estadísticos necesarios, se utilizaron diversas estadísticas sobre los errores de los pronósticos de los modelos para poder decidir cuál es el modelo elegido.

Aunque hay diversas maneras de cuantificar el error de los pronósticos de un modelo de series de tiempo, en este caso, se compararon los pronósticos utilizando las estadísticas *MAE* (Mean Absolute Error), *RMSE* (Root Mean Square Error) y *MAPE* (Mean Absolute Percentage Error), las cuales se muestran a continuación.

| <b>Medidas de Error</b> |            |             |             |
|-------------------------|------------|-------------|-------------|
|                         | <b>MAE</b> | <b>RMSE</b> | <b>MAPE</b> |
| <b>Modelo 2</b>         | 3.0129     | 2.5388      | %0.9529     |
| <b>Modelo 4</b>         | 3.2371     | 2.5260      | %0.9443     |
| <b>Modelo 5</b>         | 2.8997     | 2.3595      | %0.8846     |
| <b>Modelo 6</b>         | 2.5273     | 1.9899      | %0.7425     |

Cuadro 3.10: Errores de los pronósticos

| <b>Criterios de Información</b> |            |            |
|---------------------------------|------------|------------|
|                                 | <b>AIC</b> | <b>BIC</b> |
| <b>Modelo 2</b>                 | 1050       | 1085       |
| <b>Modelo 4</b>                 | -1666      | -1620      |
| <b>Modelo 5</b>                 | -1666      | -1635      |
| <b>Modelo 6</b>                 | -1660      | -1601      |

Cuadro 3.11: Bondad de Ajuste

De acuerdo a la tabla con las estadísticas de los errores de los pronósticos de los modelos, el modelo con menor *MAE*, *RMSE* y *MAPE* es el *Modelo 6*, seguido del *Modelo 5*, del *Modelo 2* y por último del *Modelo 4*.

Por lo tanto, el modelo seleccionado, el cual, cumple con todos los supuestos estadísticos y además tiene errores de pronósticos mínimos en comparación con los otros 3 modelos es el modelo descrito por un proceso *SARIMA*(4, 1, 8)(3, 1, 1)<sub>7</sub>. Sin embargo, este modelo no es el que tiene menor *AIC* o *BIC*, cabe destacar que el segundo modelo con mejor desempeño en las medidas de pronóstico, que es el modelo 5, sí es el que tiene un mejor desempeño al considerar los *Criterios de Información*, por lo cual, este también podría ser una alternativa al modelo con mejor desempeño bajo las medidas de error de pronóstico.

### 3.3.6. Conclusiones

Es de vital importancia recordar, que la motivación para encontrar un modelo que describiera el índice *S&P/BMV FIBRAS*, era tener un modelo capaz de proveer pronósticos confiables de dicho índice para el público inversionista para la correcta toma de decisiones.

Teniendo lo anterior en mente, la gráfica que se muestra a continuación, muestra los pronósticos (línea amarilla) de los días del mes de enero, así como los valores reales

del índice en esos mismos días (línea azul), además de las correspondientes bandas de confianza de los pronósticos.

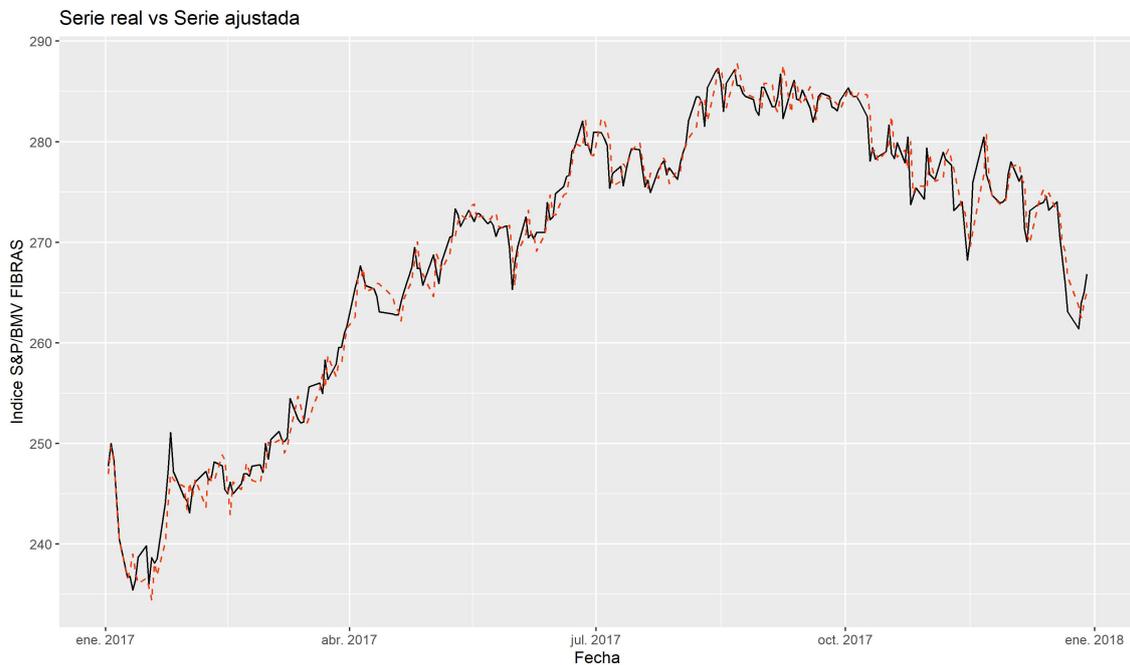


Figura 3.35: Serie real vs serie ajustada del índice S&P/BMV FIBRAS

A pesar de que el modelo algunas veces subestima o sobreestima el índice, el modelo siempre describe el comportamiento de la serie, esto es, los pronósticos del modelo siguen el comportamiento real de la serie. Cuando el modelo sugiere que el índice *S&P/BMV FIBRAS* aumentará, este en efecto aumenta, lo mismo ocurre cuando el modelo sugiere que el índice bajará, en efecto el índice baja.

Estos resultados son muy buenos, debido a que si en su momento se hubiera apostado en que el índice iba a subir, hubiera sido bueno comprar participaciones de las FIBRAS antes de estos aumentos, para que cuando el modelo sugiriera que iba a comenzar a bajar el rendimiento, en ese momento se pudieron vender las participaciones para tener una mayor ganancia por su valor, para después volver a comprar cuando el modelo sugiriera un valor más bajo.

De esta manera, si lo dicho anteriormente hubiera sido una estrategia formulada por el modelo, de hecho, en efecto se hubiera tenido ganancias debido a que justo ese fue el comportamiento de la serie. Además, para los inversionistas, el tener una idea, gracias a las bandas de confianza, de cuánto podrían perder o ganar como máximo, provee de mayor información para saber qué tanto conviene dicha estrategia.

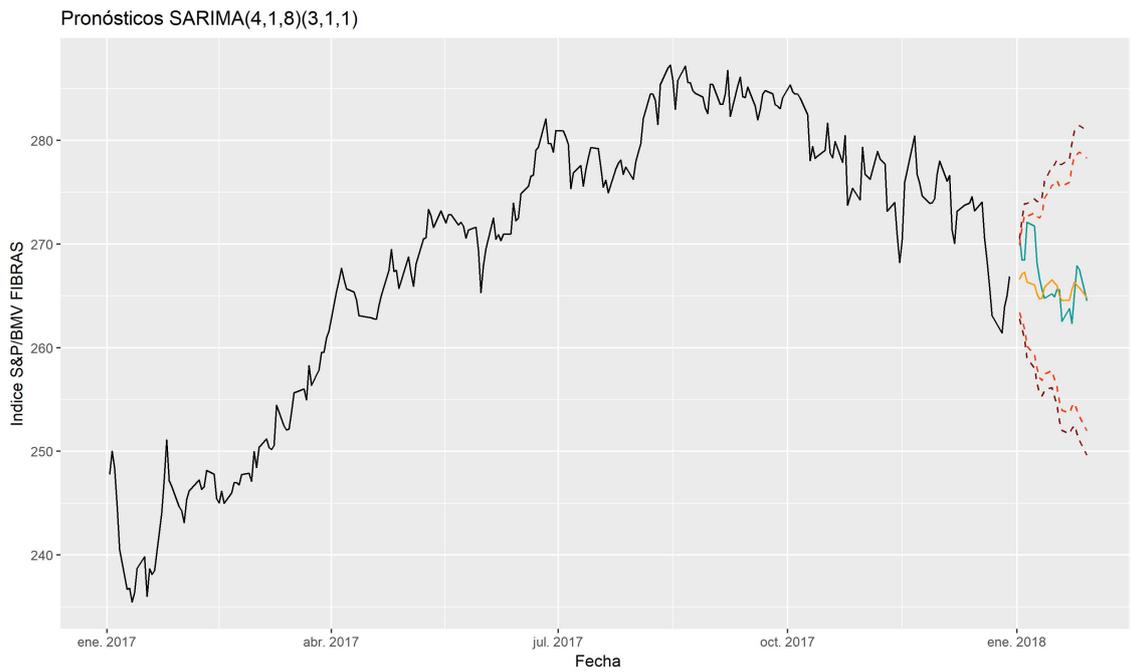


Figura 3.36: Pronóstico índice S&P/BMV FIBRAS

Por lo tanto, se tiene la certeza y la confianza de que se generó un modelo robusto, capaz de proveer información confiable al público inversionista para la correcta toma de decisiones.

# Apéndice A

## Anexo

### A.1. Teorema de proyección

**Teorema 1** [1] Sea  $M$  un subespacio del espacio de Hilbert,  $H$ , y sea  $y$  un elemento en  $H$ . Entonces,  $y$  puede representarse de forma única como

$$y = \hat{y} + z, \quad (\text{A.1})$$

donde  $\hat{y}$  pertenece a  $M$ , y  $z$  es ortogonal a  $M$ ; es decir,  $\langle z, w \rangle = 0 \quad \forall w \in M$ . Además,  $\hat{y}$  es el más cercano al elemento  $y$ , en el sentido en que, para cualquier  $w \in M$ ,  $\|y - w\| \geq \|y - \hat{y}\|$ , donde la igualdad se cumple sí y sólo sí  $w = \hat{y}$ .

Recordar que un *espacio de Hilbert*,  $H$ , se define como un espacio con un producto escalar, el cual, posee la *propiedad de Cauchy*. En otras palabras,  $H$  es un espacio completo con producto escalar. Esto quiere decir que toda sucesión de *Cauchy* converge en norma; esto es,  $x_n \rightarrow x \in H$  sí y sólo sí  $\|x_n - x_m\| \rightarrow 0$ , cuando  $m, n \rightarrow \infty$ . Esto es, la completitud de  $L^2$ .

Usando la notación del **Teorema 1**, definimos a la proyección de  $H$  en  $M$ , como  $P_M y = \hat{y}$  para  $y \in H$ . Además, el subespacio generado por el conjunto finito  $\{x_1, \dots, x_n\}$  de elementos de un *espacio de Hilbert*,  $H$ , se define como el conjunto de todas las combinaciones lineales  $w = a_1 x_1 + \dots + a_n x_n$ , donde  $a_1, \dots, a_n$  son escalares. Este subespacio de  $H$  se denota por  $M = \bar{s}p\{x_1, \dots, x_n\}$ . Por el teorema de proyección, la proyección de  $y \in H$  en  $M = \bar{s}p\{x_1, \dots, x_n\}$  es única y está dada por

$$P_M y = a_1 x_1 + \dots + a_n x_n,$$

donde los escalares  $a_1, \dots, a_n$  se encuentran utilizando el *principio de ortogonalidad*

$$\langle y - P_M y, x_j \rangle = 0, \quad \text{para } j = 1, \dots, n.$$

Específicamente,  $a_1, \dots, a_n$  se pueden obtener resolviendo

$$\sum_{i=1}^n a_i \langle x_i, x_j \rangle = \langle y, x_j \rangle, \quad \text{para } j = 1, \dots, n.$$

En particular, cuando los elementos de  $H$  son vectores, este problema se convierte en el problema de *regresión lineal*.

## A.2. Periodograma

Históricamente, el análisis espectral, comenzó con la búsqueda de *periodicidades escondidas* en los datos de series de tiempo. En toda la teoría desarrollada previamente, el entendimiento, así como la modelación de las series de tiempo, estuvo enfocada en analizar y entender las propiedades de correlación de las observaciones de la serie de tiempo subyacente. Dicho análisis se conoce como “*análisis sobre el dominio del tiempo*”. Por otro lado, cuando se analizan las propiedades de las frecuencias de las observaciones de las series de tiempo en general, se dice que se está trabajando en el “*dominio frecuentista*”. Dicho “*dominio frecuentista*”, es mejor conocido como *análisis espectral*, el cual, ha sido particularmente útil en áreas como *ingeniería en comunicaciones, acústica, ciencias geofísicas, ciencias biomédicas, etc.*

El objetivo de lo que se desarrollará a continuación sobre el análisis espectral, no es adentrarse en profundidad sobre los temas de esta extensa área de conocimiento, únicamente se busca que queden claras algunas nociones básicas utilizadas en la modelación previa.

Para lograr este objetivo, considérese la curva descrita por

$$R \cos(2\pi f t + \Phi) \tag{A.2}$$

donde  $R > 0$ , describe la amplitud,  $f$  la frecuencia, y  $\Phi$  la fase de la curva. Debido a que la curva se repite a sí misma cada  $\frac{1}{f}$  unidades de tiempo, a  $\frac{1}{f}$  se le conoce *el periodo de onda*. Ahora, aunque la *periodicidad* de la serie, generalmente está “*escondida*”, i.e., es

desconocida, el análisis espectral provee de herramientas para “descubrir” y estimar esta periodicidad.

Claramente, la ecuación dada por A.2, no es conveniente para estimar, pues los parámetros  $R$  y  $\Phi$ , no entran en la ecuación de forma lineal. En su lugar, se hace uso de una identidad trigonométrica, con la finalidad de reparametrizar a A.2, quedando como

$$R\cos(2\pi ft + \Phi) = A\cos(2\pi ft) + B\sin(2\pi ft) \quad (\text{A.3})$$

donde  $R = \sqrt{A^2 + B^2}$ ,  $\Phi = \arctan(-\frac{B}{A})$ , y consecuentemente,  $A = R\cos(\Phi)$ ,  $B = -R\sin(\Phi)$ .

Por lo tanto, para una frecuencia “ $f$ ” fija, se pueden usar, tanto a  $\cos(2\pi ft)$ , así como a  $\sin(2\pi ft)$ , como variables predictoras, y de esta manera, ajustar los valores de  $A$  y  $B$  a partir de los datos por medio de una regresión por mínimos cuadrados.

Una combinación lineal general de  $m$  curvas con amplitudes, frecuencias y fases arbitrarias, se puede escribir de la siguiente manera

$$Y_t = A_0 + \sum_{j=1}^m [A_j\cos(2\pi f_j t) + B_j\sin(2\pi f_j t)] \quad (\text{A.4})$$

Sobre esta ecuación, puede ajustarse un modelo de regresión por mínimos cuadrados ordinarios para estimar los valores de las  $A$ 's y  $B$ 's, de hecho, cuando las frecuencias son de una forma particular, esta regresión es particularmente sencilla. Supóngase que  $n$  es impar, por ende,  $n$  se puede expresar como  $n = 2k + 1$ , donde  $k \in \mathbf{N}$ , entonces, todas las frecuencias de la forma  $\frac{1}{n}, \frac{2}{n}, \dots, \frac{k}{n}$ , donde  $\frac{k}{n} = \frac{n-1}{2} = \frac{n-2}{2n} = \frac{1}{2} - \frac{1}{2n}$ , llevan por nombre *frecuencias de Fourier*.

Las *frecuencias de Fourier*, son de relevancia, debido a que las variables predictoras, seno y coseno, en estas frecuencias, así como en  $f = 0$ , son ortogonales, y por ende, los estimadores por mínimos cuadrados ordinarios, son simplemente

$$\hat{A}_0 = \bar{Y} \quad (\text{A.5})$$

$$\hat{A}_j = \frac{2}{n} \sum_{t=1}^n Y_t \cos\left(2\pi t \frac{j}{n}\right) \quad (\text{A.6})$$

$$\hat{B}_j = \frac{2}{n} \sum_{t=1}^n Y_t \sin\left(2\pi t \frac{j}{n}\right) \quad (\text{A.7})$$

Por otro lado, si el tamaño de muestra es par, es decir,  $n$  se puede escribir de la forma  $n = 2k$ , donde  $k \in \mathbf{N}$ , las ecuaciones A.6 y A.7, siguen siendo válidas para  $j = 1, 2, \dots, k-1$ , pero para  $j = k$ , se tiene

$$\hat{A}_k = \frac{1}{n} \sum_{t=1}^n (-1)^t Y_t \quad (\text{A.8})$$

$$\hat{B}_k = 0 \quad (\text{A.9})$$

pues, en el caso  $j = k$ ,  $f_k = \frac{k}{n} = \frac{1}{2}$ .

Una observación importantes, es que cualquier serie de tiempo de tamaño  $n$ , determinista o estocástica, con o sin periodicidad, puede ajustarse perfectamente por el modelo descrito por la ecuación A.4 al elegir a  $m$  igual a  $m = \frac{n}{2}$ , en el caso en que  $n$  sea par; o  $m = \frac{n-1}{2}$ , en el caso en que  $n$  sea impar. Por lo tanto, existirán  $n$  parámetros por estimar si se desea ajustar de forma perfecta a la serie de longitud  $n$ .

De esta manera, para muestras impares con  $n = 2k + 1$ , el *periodograma*  $I$ , con frecuencia fija  $f = \frac{j}{n}$  para  $j = 1, 2, \dots, k$ , se define como

$$I\left(\frac{j}{n}\right) = \frac{n}{2} (\hat{A}_j^2 + \hat{B}_j^2) \quad (\text{A.10})$$

Por otro lado, si la muestra es par y  $n = 2k$ , las ecuaciones A.6 y A.7, aún son usadas para estimar los coeficientes  $A_j$ 's y  $B_j$ 's, mientras que la ecuación A.10, aún define al *periodograma*, sin embargo, en la frecuencia  $f = \frac{k}{n} = \frac{1}{2}$ , se hace uso de la ecuación A.8, de tal manera que

$$I\left(\frac{1}{2}\right) = n (\hat{A}_k)^2$$

Por su definición, el *periodograma* es proporcional a la suma del cuadrado de los coeficientes de regresión asociados a la frecuencia  $\frac{j}{n}$ , por lo cual, la "altura" del *periodograma* muestra la "fuerza" relativa a la pareja *coseno, seno*, sobre distintas frecuencias, a lo largo del comportamiento completo de la serie.

Otra interpretación del *periodograma*, es en términos del análisis de varianza, pues, bajo esta perspectiva, el *periodograma*  $I\left(\frac{j}{n}\right)$  es la suma de cuadrados con  $2g.l.$ , asociada con la pareja de coeficientes  $(A_j, B_j)$  en la frecuencia  $\frac{j}{n}$ , esto para cuando  $n$  es impar, es decir  $n = 2k + 1$ . Entonces, se tiene que

$$\sum_{j=1}^n (Y_j - \bar{Y})^2 = \sum_{j=1}^k I\left(\frac{j}{n}\right)$$

Un resultado similar se sostiene cuando  $n$  es para, pero en este caso, existe un término adicional en la suma,  $I\left(\frac{1}{2}\right)$ , lo cual, también agrega un  $g.l.$ .

Es claro que para series grandes, el cálculo de un gran número de coeficientes de regresión puede ser muy intensivo. Afortunadamente, existen métodos numéricos muy efectivos basados en la transformación *FFT, Fast Fourier Transform*, los cuales, permiten realiza los cálculos de una manera eficaz.

Es importante resaltar, que aunque se definió el *periodograma* considerando únicamente las *frecuencias de Fourier*, se extiende la definición dada anteriormente del *periodograma*, a todas las frecuencias en el intervalo  $\left[0, \frac{1}{2}\right]$ , por lo cual, se tiene que para  $0 \leq f \leq \frac{1}{2}$

$$I(f) = \frac{n}{2} (\hat{A}_f^2 + \hat{B}_f^2) \quad (\text{A.11})$$

donde

$$\hat{A}_f = \frac{2}{n} \sum_{t=1}^n Y_t \cos(2\pi t f)$$

$$\hat{B}_f = \frac{2}{n} \sum_{t=1}^n Y_t \sen(2\pi t f)$$

Se consideran únicamente frecuencias positivas debido a la naturaleza par e impar de las funciones seno y coseno, pues, cualquier función seno y/o coseno con frecuencia negativa, dígase  $-f$ , puede ser expresada con una función seno y/o coseno con frecuencia  $+f$ , por lo cual, no se pierde generalidad al considerar únicamente frecuencias positivas. Además, se registren las frecuencias en el intervalo  $\left[0, \frac{1}{2}\right]$ , pues, si por ejemplo, se tiene una función coseno con frecuencia  $\frac{1}{4}$  y otra con frecuencia  $\frac{3}{4}$ , y se observan estas dos funciones en puntos discretos  $0, 1, 2, 3, \dots$ , se observará que las dos curvas son idénticas.

Debido a lo anterior, en el análisis de series de tiempo a *tiempo discreto*, nunca se podrían distinguir curvas como en el ejemplo anterior. De hecho, se dice que las dos frecuencias,  $\frac{1}{4}$  y  $\frac{3}{4}$  están *traslapadas* una con otra. En general, cada frecuencia  $f$ , contenida en el intervalo  $\left[0, \frac{1}{2}\right]$ , estará traslapada con cada frecuencia de la forma  $f + k\left(\frac{1}{2}\right)$ , para cualquier entero positivo  $k$ , por lo cual, es suficiente con centrarse en las frecuencias contenidas en el intervalo  $\left[0, \frac{1}{2}\right]$ .

### A.3. Criterios de información

Una vez que los datos correspondientes a una serie de tiempo observada han sido transformados, por ejemplo, a través de alguna combinación de *transformaciones de Box-Cox* y/o por medio de diferenciación, con la finalidad de remover tendencias o estacionalidades en la serie, así como para corregir cierta inestabilidad en la varianza, hasta el punto en que la serie transformada  $\{X_t\}$  puede ser potencialmente ajustada por un modelo *ARMA* de media cero. Es en este punto, cuando se presenta ahora el problema de seleccionar de manera adecuada, valores para los órdenes  $p$  y  $q$  del modelo, pues no es conveniente desde un punto de vista de pronóstico, escoger valores arbitrariamente grandes para los valores de  $p$  y  $q$ , debido a que ajustar un modelo de alto grado, en general, resultará en una varianza estimada, correspondiente al ruido blanco, muy pequeña, sin embargo, cuando el modelo es usado para pronosticar, el *error cuadrático medio* del pronóstico no depende únicamente de la varianza del ruido blanco, sino también de los errores que surgen de la estimación de los parámetros del modelo, lo cual, será más grande para modelos de un orden mayor.

Es por esta razón que es necesario considerar un factor que penalice el ajustar un modelo con muchos parámetros. Muchos criterios basados en factores de penalización han sido propuestos en la literatura, debido a que el problema de selección de modelos se presenta mucho en estadística, particularmente en el análisis de regresión. En particular, se presentará, en este trabajo, un interés particular en los criterios de *Akaike*, *AIC* y *BIC*, así como la versión del criterio de *AIC* que corrige cierto sesgo, conocido como *AICC*.

#### A.3.1. Criterio de información de Akaike, AIC

El criterio de información de *Akaike*(1975), conocido como *AIC*, fue diseñado con la finalidad de ser un estimador insesgado del índice *Kullback-Leibler* del modelo ajustado, relativo al verdadero modelo, posteriormente, se desarrolla una versión corregida por sesgo de este criterio, conocido como *AICC*, dicha corrección fue propuesta por *Hurvich & Tsai* (1989).

El modelo de *AIC*, buscar seleccionar el modelo que minimice

$$AIC = -2\log(\text{verosimilitud maximizada}) + 2(m + 1)$$

donde  $m$  es el número de parámetros en el modelo y el 1, considera que el modelo contiene un intercepto o constante, en caso contrario este 1 es omitido. De tal manera, que el

término  $2(m+1)$  o  $2m$ , sirve como una “función de penalización”, con la finalidad de ayudar a asegurar la selección de un modelo parsimonioso, pues si el modelo tiene muchos parámetros, el valor de  $2(m+1)$  o  $2m$  será grande.

Como se mencionó previamente, el *AIC* es un estimador del valor esperado de la *divergencia de Kullback-Leibler* del modelo estimado con respecto al modelo real. Esto quiere decir que si se considera a  $p(Y_1, \dots, Y_n)$ , la función de probabilidad del vector aleatorio  $(Y_1, \dots, Y_n)$  y a  $q_\theta(Y_1, \dots, Y_n)$ , la correspondiente función de probabilidad bajo el modelo con parámetros  $\theta$ , la *divergencia Kullback-Leibler* de  $q_\theta$  con respecto a  $p$ , queda definida por la ecuación

$$D(p, q_\theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(Y_1, \dots, Y_n) \log \left( \frac{p(Y_1, \dots, Y_n)}{q_\theta(Y_1, \dots, Y_n)} \right) dY_1, \dots, dY_n \quad (\text{A.12})$$

De tal forma, que el *AIC* estima  $\mathbf{E} [D(p, q_{\hat{\theta}})]$ , donde  $\hat{\theta}$ , es el estimador obtenido por máxima verosimilitud del vector de parámetros  $\theta$ , sin embargo, el estimador *AIC* es un estimador sesgado para el valor esperado de la *divergencia Kullback-Leibler*, y dicho sesgo puede ser considerablemente grande. De hecho, simulaciones montecarlo, han demostrado que cuando se busca elegir modelos *AR* a través del criterio *AIC*, el *AIC* tiende a sobrestimar el valor de  $p$ .

Debido a lo anterior, *Hervich & Tsai(1989)*, probaron que el sesgo puede eliminarse de manera aproximada, al añadir al *AIC*, un término adicional de penalización no estocástico, dando como resultado el “*AIC corregido*”, definido por la ecuación

$$AICC = AIC + \frac{2(k+1)(k+2)}{n-k-2},$$

donde  $k = m+1$  o  $k = m$ , según sea el caso. En esta expresión,  $n$  corresponde al tamaño de muestra efectivo, y de nuevo  $m$  es el número de parámetros en el modelo, excluyendo a la varianza del ruido blanco.

Los factores de penalización  $\frac{2(k+1)(k+2)}{n-k-2}$  y  $2(k+1)$ , son asintóticamente equivalentes cuando  $n \rightarrow \infty$ . La estadística *AICC* tiene una mayor penalización para modelos de orden mayor, lo cual, contraresta la tendencia de *overfitting* del *AIC*.

### A.3.2. Criterio de información Bayesiano, BIC

Otro criterio para determinar el orden de un modelo *ARMA*, que busca corregir la naturaleza de *overfitting* del criterio *AIC*, es seleccionar aquel modelo que minimice el

criterio de información Bayesiano (*BIC*), el cual está definido como

$$BIC = -2\log(\text{verosimilitud maximizada}) + k\log(n)$$

En el caso en que el proceso verdadero sigue un modelo  $ARMA(p, q)$ , entonces, se sabe que los órdenes especificados al minimizar el *BIC* son consistentes en el sentido de que si las observaciones  $\{X_1, \dots, X_n\}$  son en realidad, observaciones de un proceso  $ARMA(p, q)$ , y si  $\hat{p}$  y  $\hat{q}$  son el orden estimado al minimizar el *BIC*, entonces  $\hat{p} \rightarrow p$  y  $\hat{q} \rightarrow q$  con probabilidad 1 cuando  $n \rightarrow \infty$ .

Esta propiedad no se cumple para los criterios *AIC* y *AICC*, sin embargo, si el proceso verdadero no es un proceso  $ARMA$  finito, entonces, al minimizar el *AIC* o *AICC* sobre una clase grande y creciente de modelos, tiene la propiedad de que tenderá a un modelo  $ARMA$  óptimo, el cual, estará más cercano al proceso verdadero sobre las clases de modelos en estudio, es decir, los criterios *AIC* y *AICC* son asintóticamente eficientes, mientras que el criterio *BIC* no lo es. La cercanía es medida en términos de la *divergencia Kullback-Leibler*, la cual, es una de muchas medidas de disparidad de modelos.

Sin importar que criterio de información se use, *AIC*, *BIC*, *AICC*, tienen la desventaja de requerir la estimación de la función de verosimilitud, la cual, en el caso de los modelos  $ARMA$ , es difícil de estimar, y tiende a tener problemas numéricos debido a la multimodalidad de la función de verosimilitud, así como al problema de sobreparametrización, que ocurre cuando los órdenes de las partes *AR* y *MA* exceden los valores reales. Es por esto, que es útil inspeccionar otros métodos propuestos como el de *Hannan & Rissanen(1982)*, el cual busca solventar el problema de sobreparametrización.

# Bibliografía

- [1] D. S. S. Robert H. Schumway, *Time Series Analysis and Its Applications*. Springer, third ed., 2010.
- [2] K.-S. C. Jonathan D. Cryer, *Time Series Analysis: With Applications in R*. Springer, second ed., 2010.
- [3] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*. Pearson Addison Wesley, second ed., 2006.
- [4] R. S. Tsay, *Analysis of Financial Time Series*. Wiley, second ed., 2010.
- [5] R. A. D. Peter J. Brockwell, *Introduction to Time Series and Forecasting*. Springer, second ed., 2010.
- [6] R. A. D. Peter J. Brockwell, *Time Series: Theory and Methods*. Springer, second ed., 2009.
- [7] R. J. H. Spyros Makridakis, Steven C. Whellwright, *Forecasting, Methods and Applications*. Wiley, third ed., 1997.
- [8] M. K. Douglas C. Montgomery, Cheryl L. Jennings, *Introduction to Time Series Analysis and Forecasting*. Series in Probability and Statistics, Wiley, 2008.