



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

UNA INTRODUCCIÓN AL ANÁLISIS DE SERIES DE
TIEMPO Y SUS APLICACIONES

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

ACTUARIO

P R E S E N T A :

GERARDO SISNIEGA LIRA

DR. RICARDO RAMÍREZ ALDANA

2015

Ciudad Universitaria, D. F.





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.

Hoja de Datos del Jurado

1. Datos del alumno

Sisniega

Lira

Gerardo

56 69 29 43

Universidad Nacional Autónoma de México

Facultad de Ciencias

Actuaría

303561493

2. Datos del tutor

Dr.

Ricardo

Ramírez

Aldana

3. Datos del sinodal 1

Dr.

Alberto

Contreras

Cristán

4. Datos del sinodal 2

Dra.

Ruth Selene

Fuentes

García

5. Datos del sinodal 3

Dr.

Yuri

Salazar

Flores

6. Datos del sinodal 4

M en C.

José Salvador

Zamora

Muñoz

7. Datos del trabajo escrito

Una introducción al análisis de series de tiempo y sus aplicaciones

300 p.

2015

Agradecimientos

Con el mayor de los agradecimientos a mi familia. En primer lugar a mis padres, Gerardo y Ma. del Carmen, por el constante ejemplo que han sido a lo largo de mi vida, por enseñarme el significado de ser una familia, por toda la paciencia que me han tenido, por todo el cariño que me han brindado, por ser aquellas personas que han influido enormemente en mi forma de ser y pensar, por ser simplemente, aquellas que han logrado sacar lo mejor de mí. En segundo lugar, a mi hermano Luis Gabriel, por ser mi motivación a ser un mejor hermano mayor y un buen ejemplo en su vida.

Al resto de mi familia, por todos sus consejos, paciencia, amor y grandes momentos de felicidad, por aquellas reuniones en Navidad y Año Nuevo donde juntos hacíamos valer el significado de la familia. A mis abuelos paternos y abuelo materno, por los momentos tan especiales que pasé mientras estuvieron con nosotros. A mi abuela materna, por continuar enviando sus bendiciones para hacer de mi vida mucho más amena.

A mi director de tesis, Ricardo, no sólo por tenerme paciencia y guiarme a lo largo de este extenso proyecto sino también por creer en mí en estos ya dos años de impartir clases juntos. Gracias a su dedicación, apoyo incondicional y críticas constructivas he logrado dar por finalizado este proyecto.

A mis sinodales, Dr. Alberto Contreras Cristán, Dra. Ruth Selene Fuentes García, Dr. Yuri Salazar Flores y M. en C. José Salvador Zamora Muñoz, por haber leído esta tesis y complementado la misma con consejos y sugerencias.

A mis mejores amigos (Cristina, Lourdes, Mónica, Ana, Carlos, Andrés, Alejandro), por estar a mi lado cuando muestro mi mejor versión y la peor, por enseñarme un sinfín de cosas, por las risas y locuras que hemos compartido, por haberme retado, por haberme hecho fuerte y valiente, por compartir sus vidas conmigo y dejarme ser una parte importante de ellas.

A mis alumnos, por todas aquellas preguntas y dudas que tuvieron y que a partir de ellas hicieron de mí una persona mucho más preparada.

Finalmente, a todos aquellos que se han cruzado en mi camino por la razón que haya sido, buena o mala, ya que gracias a ellos yo no sería lo que soy el día de hoy.

Contenido

1. Modelo de Regresión Lineal Simple	1
1.1. Introducción	1
1.2. Estimación de los parámetros	3
1.3. Propiedades de los estimadores	7
1.4. Estimación por máxima verosimilitud	9
1.5. Pruebas de hipótesis sobre α y β	11
1.6. Intervalos de confianza	14
1.7. Coeficiente de determinación	15
1.8. Verificación del ajuste del modelo	16
1.9. Modelos de regresión con errores autocorrelacionados	22
1.10. Conclusiones	24
2. Análisis Clásico de Series de Tiempo	29
2.1. Introducción	29
2.1.1. Notas previas y etapas a cubrir al analizar series de tiempo	30
2.2. Componentes de una serie de tiempo	32
2.2.1. Representación gráfica	33
2.2.2. Diferencias y cocientes estacionales	34
2.2.3. Gráfico media-desviación estándar	36
2.3. Análisis de Tendencia	37
2.3.1. Modelo de tendencia lineal	37
2.3.2. Modelo de tendencia polinómica	41
2.3.3. Modelo de tendencia exponencial	44
2.3.4. Método de promedios móviles	47
2.3.5. Suavizamiento exponencial simple	51
2.3.6. Método de Holt o doble exponencial	56
2.4. Análisis de Estacionalidad	61
2.4.1. Filtro de estacionalidad	62
2.4.2. Método de Holt-Winters	63
2.5. Descomposición clásica	68
2.6. Análisis de la componente irregular	72
2.7. Prueba de hipótesis para estacionalidad	73
3. Modelos para Series de Tiempo Univariadas.	75
3.1. Introducción	75
3.2. Procesos estacionarios	76
3.2.1. Ruido blanco	80
3.2.2. Transformada de Box-Cox	83
3.2.3. Eliminación de tendencia y estacionalidad	85
3.2.4. Ejemplos de procesos estacionarios	88

CONTENIDO

3.3. Modelos de promedios móviles (MA)	91
3.3.1. Modelo MA(1)	91
3.3.2. Modelo MA(2)	94
3.3.3. Modelo MA(q)	97
3.4. Modelos autorregresivos (AR)	99
3.4.1. Modelo AR(1)	99
3.4.2. Modelo AR(2)	101
3.4.3. Modelo AR(p)	104
3.5. Modelos autorregresivos de promedio móviles	106
3.5.1. Causalidad	110
3.5.2. Invertibilidad	121
3.5.3. Identificación del modelo a partir de la muestra	127
3.5.4. Estimaciones usando máxima verosimilitud	155
3.5.5. Diagnósticos, verificación y corrección de supuestos	162
3.5.6. Propiedades asintóticas de $\hat{\Phi}$ y $\hat{\Theta}$	170
3.5.7. Criterios de información	174
3.6. Análisis de procesos ARMA(p,q) en ITSM	178
3.7. Modelos ARIMA	195
3.8. Pronósticos con modelos ARIMA	208
3.9. Modelos SARIMA	212
3.10. Modelos GARCH	220
4. Modelos de Regresión Dinámica	223
4.1. Introducción	223
4.2. Inferencia Bayesiana	225
4.3. Modelos de espacio de estados	232
4.3.1. Modelos dinámicos lineales	232
4.4. Modelos de regresión dinámica	246
4.4.1. Estimación de parámetros	247
4.4.2. Selección del modelo y pronósticos	249
Conclusiones	267
A. Tablas Estadísticas	271

Índice de figuras

1.1.1. Figura que relaciona la presión barométrica con el punto de ebullición	2
1.8.1. Figura que muestra distintas versiones de la gráfica de probabilidad normal . .	17
1.8.2. Figura para corroborar el supuesto de homoscedasticidad	18
1.8.3. Figura que muestra a los residuos y gráfica de probabilidad normal	19
1.8.4. Figura que muestra el histograma y el diagrama de caja y bigotes de los residuos	20
1.8.5. Figura de residuos (estandarizados y estudentizados) contra valores ajustados .	20
1.10.1. Figura que muestra el financiamiento dado a través de Banca Privada y Mixta	24
1.10.2. Figura que muestra a los residuos y gráfica de probabilidad normal	26
1.10.3. Histograma y diagrama de caja y bigotes de los residuos (financiamiento) . . .	26
1.10.4. Figura de residuos (estandarizados y estudentizados) contra valores ajustados .	27
2.2.1. Figura que muestra el número de acciones de atención otorgadas por Condusef	32
2.2.2. Representación gráfica del número de acciones de atención	34
2.2.3. Figura que muestra la relación media-desviación estándar	36
2.3.1. Figura que muestra el PIB trimestral de Estados Unidos	38
2.3.2. Figura que muestra el PIB trimestral de México	40
2.3.3. Figura que muestra el total de importaciones de bienes y servicios de Australia	43
2.3.4. Figura que muestra el ajuste de la función polinómica al total de importaciones	43
2.3.5. Figura que muestra a los residuos del total de importaciones australianas . . .	44
2.3.6. Figura que muestra el gasto en servicios financieros en Australia	47
2.3.7. Suavizamiento por promedios móviles en el número de acciones de atención . .	50
2.3.8. Figura correspondiente al nivel (pies) del lago Hurón	50
2.3.9. Suavizamiento por promedios móviles del nivel (pies) del lago Hurón con $q=2$.	51
2.3.10. Suavizamiento por promedios móviles del nivel (pies) del lago Hurón con $q=8$.	51
2.3.11. Suavizamiento exponencial simple en el número de acciones de atención	53
2.3.12. Suavizamiento exponencial simple ($\alpha = 0.8$) del número de acciones de atención	55
2.3.13. Figura que muestra el total de importaciones de bienes y servicios de Australia	56
2.3.14. Suavizamiento exponencial simple del total de importaciones ($\alpha=0.3$)	56
2.3.15. Figura que muestra el método de Holt para el número de acciones de atención	59
2.3.16. Método de Holt ($\alpha = 0.8$ y $\beta = 0.2$) para el número de acciones de atención . .	59
2.3.17. Método de Holt ($\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$) para el número de acciones de atención . .	60
2.4.1. Método Holt-Winters aditivo para el número de acciones de atención	64
2.4.2. Figura que muestra las predicciones para el número de acciones de atención . .	65
2.4.3. Figura que muestra el número de pasajeros mensuales	66
2.4.4. Figura que muestra el número de pasajeros mensuales bajo Holt-Winters	66
2.4.5. Método de Holt-Winters multiplicativo para el número de acciones de atención	68
2.5.1. Figura que muestra las ventas trimestrales de cierta empresa mexicana	69
2.5.2. Figura que muestra las ventas trimestrales sin la presencia de estacionalidad . .	71
3.2.1. Figura que muestra la simulación de un proceso de ruido blanco Gaussiano . .	81
3.2.2. Figura que representa a un proceso estacionario	81

ÍNDICE DE FIGURAS

3.2.3. Figura que muestra un proceso con tendencia	82
3.2.4. Figura que muestra un proceso con heteroscedasticidad	82
3.2.5. Figura que muestra un proceso con estacionalidad	82
3.2.6. Figura de un proceso con heteroscedasticidad	84
3.2.7. Figura de un proceso con homoscedasticidad	84
3.3.1. Figura que muestra dos procesos MA(1) simulados con $\theta_1 = 0.7$ y $\theta_1 = -0.7$. .	93
3.3.2. Figura que muestra un proceso MA(2) simulado con $\theta_1 = 0.3$ y $\theta_2 = -0.9$	96
3.4.1. Figura que muestra dos procesos AR(1) simulados con $\phi_1 = 0.7$ y $\phi_1 = -0.7$. .	101
3.4.2. Figura que muestra un proceso AR(2) simulado con $\phi_1 = 0.7$ y $\phi_2 = -0.3$. . .	104
3.5.1. Figura que muestra dos procesos ARMA(1,1) simulados	109
3.5.2. Figura que muestra la serie de pesos $\{\psi_j\}$ para un proceso ARMA(1,1)	115
3.5.3. Figura que muestra la serie de pesos $\{\psi_j\}$ para un proceso ARMA(2,0)	116
3.5.4. Figura que representa las ACF muestrales de varios procesos simulados	131
3.5.5. Figura que representa las PACF muestrales de varios procesos simulados . . .	150
3.5.6. Figura que muestra la estimación de θ_1	161
3.5.7. Figura que muestra la existencia de datos discrepantes	168
3.6.1. Figura que muestra el número de muertes accidentales en E.U.A.	178
3.6.2. Figura que muestra el número de muertes accidentales sin tendencia	179
3.6.3. Figura que muestra el número de muertes accidentales sin estacionalidad . . .	179
3.6.4. Figura que muestra el número de muertes accidentales restando la media . . .	179
3.6.5. Figura ACF del número de muertes accidentales en E.U.A.	180
3.6.6. Figura PACF del número de muertes accidentales en E.U.A.	180
3.6.7. Pronóstico del número de muertes accidentales en E.U.A. para el año 1979 . .	181
3.6.8. Figura que muestra el histograma de los residuos	182
3.6.9. Figura ACF de los residuos	182
3.6.10. Figura que muestra el número de acciones de atención	183
3.6.11. Figura que muestra el número de acciones de atención sin tendencia	184
3.6.12. Figura que muestra el número de acciones de atención sin estacionalidad . . .	184
3.6.13. Figura que muestra el número de acciones de atención restando la media . . .	184
3.6.14. Figura ACF del número de acciones de atención	185
3.6.15. Figura PACF del número de acciones de atención	185
3.6.16. Pronóstico del número de acciones de atención para el año 2013	186
3.6.17. Figura que muestra el histograma de los residuos	186
3.6.18. Figura ACF de los residuos	186
3.6.19. Figura que muestra el número de ventas mensuales de vino tinto	188
3.6.20. Figura que muestra la descomposición clásica de las ventas de vino tinto	188
3.6.21. Figura que muestra ventas mensuales de vino sin tendencia ni estacionalidad .	188
3.6.22. Figura ACF del número de ventas mensuales de vino tinto	189
3.6.23. Figura PACF del número de ventas mensuales de vino tinto	189
3.6.24. Pronóstico del número de ventas mensuales de vino tinto para el año 1992 . .	190
3.6.25. Figura que muestra el histograma de los residuos	190
3.6.26. Figura ACF de los residuos	190
3.6.27. Figura que muestra el nivel del lago Hurón	191
3.6.28. Figura que muestra la descomposición clásica del nivel del lago Hurón	191
3.6.29. Figura que muestra el nivel del lago Hurón sin tendencia ni estacionalidad . . .	192
3.6.30. Figura ACF del nivel del lago Hurón	192
3.6.31. Figura PACF del nivel del lago Hurón	192
3.6.32. Pronóstico del nivel del lago Hurón para el año 1973	193
3.6.33. Figura que muestra el histograma de los residuos	193
3.6.34. Figura ACF de los residuos	194

3.7.1.	Figura que muestra el índice de precios al mayoreo en la Ciudad de México . . .	200
3.7.2.	Figura ACF del índice de precios al mayoreo	200
3.7.3.	Figura ACF del índice de precios al mayoreo diferenciado	201
3.7.4.	Figura PACF del índice de precios al mayoreo diferenciado	201
3.7.5.	Figura que muestra la verificación de los supuestos del modelo	203
3.7.6.	Figura que muestra a los residuos y su correspondiente histograma	204
3.7.7.	Pronóstico para el índice de precios al mayoreo en la Ciudad de México	205
3.7.8.	Figura que muestra los residuos y su correspondiente histograma (logaritmo) .	206
3.9.1.	Figura de un proceso con estacionalidad	213
3.9.2.	Figura que muestra el nivel de lluvia en Port-Harcourt	214
3.9.3.	Figura que muestra el nivel de lluvia sin la presencia de estacionalidad	215
3.9.4.	Figura ACF del nivel de lluvia en Port-Harcourt	216
3.9.5.	Predicción del nivel de lluvia en Port-Harcourt para los años 2009-2012	217
3.9.6.	Figura que muestra la verificación de los supuestos del modelo	218
3.9.7.	Figura que muestra los residuos e histograma	219
4.1.1.	Figura que muestra el número de pasajeros de cierta aerolínea internacional . .	223
4.1.2.	Figura que muestra el consumo de gas en el Reino Unido	224
4.1.3.	Figura que muestra el índice de utilidades del Dow Jones	224
4.1.4.	Figura que muestra el índice de precios al consumo (IPC)	224
4.2.1.	Figura que muestra la distribución <i>a priori</i> y <i>a posteriori</i> de los actuarios . . .	230
4.2.2.	Figura que muestra la distribución <i>a posteriori</i> del actuario A y B $\bar{y} = 770$. .	231
4.3.1.	Figura que muestra el flujo anual del río Nilo en los años 1871-1970	238
4.3.2.	Figura que muestra el filtro para el flujo anual del río Nilo	239
4.3.3.	Figura que muestra el suavizamiento para el flujo anual del río Nilo	240
4.3.4.	Figura que muestra el pronóstico a un paso para el flujo anual del río Nilo . . .	241
4.3.5.	Figura que muestra ambos filtros para el flujo anual del río Nilo	242
4.3.6.	Figura que muestra ambos pronósticos a un paso para el flujo anual del Nilo .	243
4.3.7.	Figura que muestra gráficos de residuos para el primer modelo	244
4.3.8.	Figura que muestra gráficos de residuos para el segundo modelo	245
4.4.1.	Figura que muestra los cambios porcentuales del ingreso y gasto en U.S.A. . .	249
4.4.2.	Figura que muestra los residuos del proceso AR(2)	251
4.4.3.	Figura que muestra la sucesión de autocorrelaciones muestrales (ACF)	252
4.4.4.	Figura que muestra la sucesión de autocorrelaciones parciales (PACF)	252
4.4.5.	Figura que muestra el pronóstico del ingreso para los próximos 3 años	254
4.4.6.	Figura que muestra el pronóstico para el PIB mexicano para los próx. 3 años .	259
4.4.7.	Figura que muestra el pronóstico para el PIB de U.S.A para años 1991-1994 .	261
4.4.8.	Figura que muestra el pronóstico para el IPC mexicano para los próx. 2 años .	265

Índice de tablas

1.1.1.	Tabla que relaciona el punto de ebullición con la presión barométrica	2
1.3.1.	Propiedades y resultados del ajuste por mínimos cuadrados ordinarios	9
1.8.1.	Tabla que muestra los residuos, residuos estandarizados y estudentizados . . .	19
1.10.1.	Financiamiento otorgado a través de la Banca Privada y Mixta	24
2.2.1.	Tabla que muestra el cálculo de diferencias y cocientes estacionales	35
2.2.2.	Tabla que muestra el cálculo de media-desv. estándar	36
2.3.1.	Tabla que muestra la estimación de tendencia para el PIB de Estados Unidos .	39
2.3.2.	Tabla que muestra la estimación de tendencia para el PIB de México	41
2.3.3.	Tabla que muestra la estimación de tendencia exponencial	46
2.3.4.	Tabla que muestra el suavizamiento exponencial simple en las acciones de atención	54
2.3.5.	Suavizamiento exponencial simple en las acciones de atención con $\alpha = 0.8$. . .	55
2.3.6.	Tabla que muestra el método de Holt para el número de acciones de atención .	58
2.3.7.	Método de Holt ($\alpha = 0.8$ y $\beta = 0.2$) para el número de acciones de atención . .	60
2.3.8.	Método de Holt ($\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$) para el número de acciones de atención . .	61
2.5.1.	Tabla que muestra las ventas trimestrales de cierta empresa mexicana	68
2.5.2.	Tabla que muestra el promedio de las estaciones (trimestres)	70
2.5.3.	Tabla que muestra los cálculos de ventas trimestrales de la empresa mexicana .	70
2.5.4.	Tabla que muestra la descomposición clásica mediante el uso del comando stl .	72
2.7.1.	Tabla que muestra los cálculos para la prueba Kruskal-Wallis	73
2.7.2.	Tabla que muestra la suma de los rangos para la prueba Kruskal-Wallis	74
3.2.1.	Tabla que muestra los valores aproximados para las transformaciones de Box-Cox	85
3.5.1.	Tabla que muestra la sucesión de autocorrelaciones muestrales para $\nabla^2 X_t$. . .	150
3.5.2.	Tabla que muestra la sucesión de autocorrelaciones parciales para $\nabla^2 X_t$	150
3.5.3.	Tabla que muestra la sucesión de autocorrelaciones muestrales para ∇X_t . . .	151
3.5.4.	Tabla que muestra la sucesión de autocorrelaciones parciales para ∇X_t	151
3.5.5.	Tabla que muestra la sucesión de autocorrelaciones de los residuos de $\{Z_t\}$. .	172
3.6.1.	Pronósticos para el número de acciones de atención para el año 2013	187
3.7.1.	Tabla que muestra las características teóricas de la ACF y PACF	196
3.7.2.	Tabla que muestra el índice de precios al mayoreo en la Ciudad de México . .	199

Resumen

Este proyecto se caracteriza por su extensión ya que se han querido abordar varios objetivos. El primero de ellos es el de servir como material de consulta en el tema de series de tiempo para futuras generaciones con preparación matemática moderada o elevada. Todos los temas elaborados en este escrito se han tratado con la mayor seriedad posible guiando al lector a través de ejemplos sencillos para la fácil comprensión del tema en cuestión, hasta proposiciones un tanto complicadas pero con inmensa utilidad en la práctica.

El segundo objetivo ha sido el de introducir al lector al uso de paquetes estadísticos, a saber, R e ITSM, como herramientas auxiliares en el análisis estadístico. Todos y cada uno de los temas abordados constan de ejemplos prácticos en dichos paquetes. Lo anterior con la finalidad de no quedarse con el aspecto teórico de cada tema sino mostrar una visión mucho más amplia de las aplicaciones de cada uno de ellos. Los códigos para la reproducción de todos los ejemplos prácticos se incluyen en este escrito.

El tercer objetivo ha sido el de brindar un método alternativo para el pronóstico del número de acciones de atención otorgadas por Condusef. Durante mi estancia en dicha Institución tuve la inquietud de proponer un método que pronosticase de manera eficiente el número de acciones de atención mensuales otorgadas por Condusef. Lo anterior con la finalidad de estar preparados en el futuro para atender las necesidades de los usuarios de servicios financieros.

El cuarto y último objetivo ha sido el de introducir un conjunto de modelos que permitan la inclusión de variables relevantes en el pronóstico de cierta información y no solamente la inclusión de observaciones pasadas. Por ejemplo, los efectos de días festivos, las actividades de los competidores en cierto mercado, cambios en las regulaciones, una expansión económica u otras variables externas que pueden llegar a explicar algunas de las variaciones históricas y permitir una mejor predicción en cierta variable respuesta.

El contenido del trabajo comienza con un capítulo de carácter introductorio, que nos recuerda nociones básicas del modelo de regresión lineal simple, estimación de parámetros, los supuestos del modelo y las consecuencias de que algunos de ellos no se lleguen a cumplir. Especial atención se pondrá en el supuesto de correlación entre las observaciones de las variables involucradas. A partir de la falla en este supuesto y otras en los demás supuestos se dará pie a utilizar series de tiempo como técnica alternativa en la previsión de fenómenos de cualquier naturaleza.

El capítulo 2 está dedicado completamente al análisis clásico de series de tiempo. En dicho capítulo se estudiarán las distintas componentes en que se presume dividen a una serie de tiempo, a saber: tendencia, estacionalidad y componente irregular. Para cada una de las componentes anteriores se realizará un análisis exhaustivo y se estudiarán varios métodos para su correspondiente estimación. La razón para realizar dicha descomposición no es otra que interpretar mejor los cambios que sufre una variable a lo largo del tiempo.

RESUMEN

Aún a pesar de que la metodología descrita en el capítulo anterior es bastante útil y arroja resultados aceptables, su empleo se ha visto limitado debido a su enfoque determinista. De tal forma que, en el capítulo 3, se tratará la metodología sugerida por Box & Jenkins. Especial interés se tendrá en los modelos autorregresivos $AR(p)$, de medias móviles $MA(q)$, autorregresivos de medias móviles $ARMA(p,q)$ y autorregresivos integrado de medias móviles $ARIMA(p,d,q)$. Al final de este capítulo se verán nociones básicas del modelo $SARIMA(p, d, q) \times (P, D, Q)$.

Finalmente, en el capítulo 4 se da una introducción a los modelos de regresión dinámica. Los modelos de series de tiempo nos permiten la inclusión de información de observaciones pasadas pero no nos permiten la inclusión de otra información que pueda llegar a ser relevante. Dicho capítulo servirá como introducción para aquellos que deseen tener un mayor conocimiento en cómo las variables pueden afectar a una serie de tiempo además del marco temporal.

Capítulo 1

Modelo de Regresión Lineal Simple

1.1. Introducción

El análisis de regresión es una de las técnicas más utilizadas al momento de investigar y modelar la relación que existe entre dos o más variables. Sus aplicaciones son vastas y las podemos encontrar en cualquier campo de estudio, por ejemplo, en las ciencias físicas, químicas, biológicas y sociales. Comencemos con un ejemplo que utilizaremos a lo largo de este capítulo para desglosar y comprender el uso del así llamado **análisis de regresión simple**.

Supongamos que un investigador sospecha que el punto de ebullición del agua ($^{\circ}\text{F}$) en cierta ciudad está relacionado con la presión barométrica (pulgadas de mercurio) ejercida en dicha ciudad. Para corroborar esta suposición, el investigador realiza un experimento obteniendo los datos mostrados en la **Tabla 1.1.1**. Dichos datos fueron tomados de Montgomery y Peck (1992).

Las 17 observaciones graficadas en el diagrama de dispersión que se muestra en la **Figura 1.1.1**, sugieren una fuerte relación entre la presión barométrica y el punto de ebullición del agua. De hecho, nuestra primera impresión es que los datos, en su mayoría, caen a lo largo de una línea recta. Sea y la variable que representa el punto de ebullición del agua y x la variable que representa la presión barométrica, entonces la ecuación que relaciona a estas dos variables está dada por:

$$y = \alpha + \beta x \tag{1.1.1}$$

Definamos los parámetros de la **ec. 1.1.1**. El parámetro β corresponde a la **pendiente** mientras que el parámetro α representa el **intercepto**. A estos parámetros, α y β , se les conoce como los **coeficientes de la regresión**. Ahora bien, observemos en la **Figura 1.1.1**, que no todos los puntos caen sobre una línea recta.

Debido a esto, nos vemos en la necesidad de añadir una nueva variable ϵ que considere la distancia que existe entre las observaciones y la línea recta. Es decir, ϵ , es una variable aleatoria que explica la falla o error del modelo de regresión al tratar de ajustar los datos. Nuestra nueva ecuación, tomando en consideración la variable ϵ , será conocida como **modelo de regresión lineal**. Dicha ecuación será de la forma:

$$y = \alpha + \beta x + \epsilon \tag{1.1.2}$$

Obs. i	Ebullición y_i	Presión x_i
1	199.5	20.79
2	199.3	20.79
3	197.9	22.40
4	198.4	22.67
5	199.4	23.15
6	199.9	23.35
7	200.9	23.89
8	201.1	23.99
9	201.9	24.02
10	201.3	24.01
11	203.6	25.14
12	204.6	26.57
13	209.5	28.49
14	208.6	27.76
15	210.7	29.64
16	211.9	29.88
17	212.2	30.06

Tabla 1.1.1: Tabla que relaciona el punto de ebullición con la presión barométrica

Llamaremos a x la **variable predictora** o **explicativa** y a y la **variable respuesta**. Claramente, la **ec. 1.1.2** sólo depende de una variable regresora, de aquí que dicha ecuación sea conocida como **modelo de regresión lineal simple**. Pensemos que podemos fijar el valor de la variable predictora, x , y observemos su correspondiente valor de respuesta y . Si el valor de x es fijo, la componente ϵ en la **ec. 1.1.2** determina las propiedades de y . Esto quiere decir que y es una variable aleatoria que depende a su vez de la componente ϵ . Supongamos que la media y la varianza de ϵ son 0 y σ^2 , respectivamente. Además, asumiremos que los errores no están correlacionados.

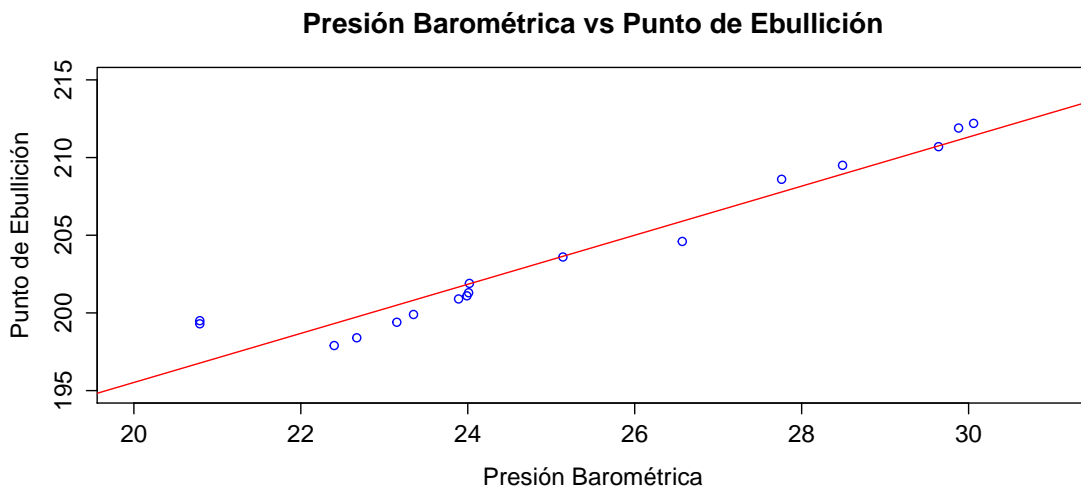


Figura 1.1.1: Figura que relaciona la presión barométrica con el punto de ebullición

La respuesta media para cualquier valor de la variable predictora, \mathbf{x} , es:

$$\mathbb{E}(y|x) = \mathbb{E}(\alpha + \beta x + \epsilon) = \alpha + \beta x \quad (1.1.3)$$

La varianza de \mathbf{y} dado cualquier valor de \mathbf{x} es:

$$\text{Var}(y|x) = \text{Var}(\alpha + \beta x + \epsilon) = \sigma^2 \quad (1.1.4)$$

La pendiente β puede ser interpretada como el cambio en la media de \mathbf{y} debido a un cambio unitario en \mathbf{x} . Más aún, la variabilidad de \mathbf{y} para algún valor particular de \mathbf{x} esta determinada por la varianza del error del modelo, σ^2 .

Uno de los objetivos principales del análisis de regresión es el de estimar los parámetros desconocidos en el modelo. Este proceso es conocido como **ajuste del modelo a los datos**. En este capítulo introductorio veremos el método de mínimos cuadrados ordinarios y la estimación por máxima verosimilitud, así como las diferencias que existen en ambos métodos. La siguiente etapa del análisis de regresión es el estudio de la **validez del modelo** y la **calidad del ajuste**. A través de este análisis exhaustivo se determina la utilidad del modelo. El resultado de dicho estudio puede indicar que el modelo es razonable para ajustarse a unos datos, o bien, que el modelo debe ser modificado. Por tanto, podemos decir que el análisis de regresión es un procedimiento iterativo, en el cual los datos llevan al ajuste de un modelo, después del cual el modelo es verificado, llevando a la modificación o adopción del mismo.

Antes de comenzar de lleno con el modelo de regresión lineal simple debemos hacer hincapié en lo siguiente. El modelo de regresión lineal no implica una relación de causa-efecto entre las variables. A pesar de que pueda existir una fuerte relación empírica entre dos o más variables, esto no puede ser considerado como evidencia para relacionar a las variables en una manera de causa-efecto. Para establecer **causalidad**, la relación entre variables explicativas y variable respuesta deben tener una base fuera de la muestra que se ha tomado (datos). El análisis de regresión puede ayudar a confirmar la relación causa-efecto pero no puede ser la única base para hacer dicha aseveración. El análisis de regresión debe de estar bien establecido por los expertos en el área de estudio en la cual se aplica el análisis.

1.2. Estimación de los parámetros

En este apartado nos dedicaremos a la estimación de los parámetros de la regresión haciendo uso del método de **mínimos cuadrados ordinarios**, esto es, estimaremos $\hat{\alpha}$ y $\hat{\beta}$ de tal manera que la suma de los cuadrados de las diferencias entre las observaciones y_i y la línea recta estimada, $\hat{\alpha} + \hat{\beta}x_i$, sea mínima. De aquí que el criterio de mínimos cuadrados sea:

$$S(\hat{\alpha}, \hat{\beta}) = \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)^2 \quad (1.2.1)$$

donde

$$y_i = \alpha + \beta x_i + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (1.2.2)$$

La ec. 1.2.2 es conocida como **modelo de regresión muestral** en términos de los n pares de datos (y_i, x_i) , $i = 1, 2, \dots, n$. Los estimadores por mínimos cuadrados de α y β , evaluados en $\hat{\alpha}$ y $\hat{\beta}$ respectivamente, deberán satisfacer:

$$\left. \frac{\partial S}{\partial \alpha} \right|_{\hat{\alpha}, \hat{\beta}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i) = 0 \quad (1.2.3)$$

$$\left. \frac{\partial S}{\partial \beta} \right|_{\hat{\alpha}, \hat{\beta}} = -2 \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta}x_i)x_i = 0 \quad (1.2.4)$$

Al simplificar, obtenemos las **ecuaciones normales por mínimos cuadrados**, a saber:

$$\sum_{i=1}^n y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^n x_i \quad (1.2.5)$$

$$\sum_{i=1}^n y_i x_i = \hat{\alpha} \sum_{i=1}^n x_i + \hat{\beta} \sum_{i=1}^n x_i^2 \quad (1.2.6)$$

Los estimadores por mínimos cuadrados de los parámetros del modelo, $\hat{\alpha}$ y $\hat{\beta}$, serán:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \quad (1.2.7)$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

donde

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (1.2.8)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$

y además, para el estimador $\hat{\beta}$ tenemos:

$$S_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \quad (1.2.9)$$

$$S_{xy} = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)$$

Además de obtener los estimadores $\hat{\alpha}$ y $\hat{\beta}$, necesitamos calcular un estimador de σ^2 para, más adelante, poder realizar pruebas de hipótesis y construir intervalos de confianza. El estimador para σ^2 se obtiene a partir de la **suma de los errores al cuadrado**.

Esta suma, denotada como SS_{Res} , esta dada por:

$$\begin{aligned}
 SS_{Res} &= \sum_{i=1}^n e_i^2 \\
 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &= \sum_{i=1}^n y_i^2 - n\bar{y}^2 - \hat{\beta}S_{xy} \\
 &= \sum_{i=1}^n (y_i - \bar{y})^2 - \hat{\beta}S_{xy} \\
 &= SS_T - \hat{\beta}S_{xy}
 \end{aligned} \tag{1.2.10}$$

donde SS_T es la llamada **suma de cuadrados total**. SS_{Res} , la suma de los errores al cuadrado, tiene asociada una **distribución ji-cuadrada**, con $n - 2$ grados de libertad debido a la estimación de $\hat{\alpha}$ y $\hat{\beta}$. Además, al asumir ciertos supuestos distribucionales (los cuales se verán en la **Sección 1.4**), se puede demostrar que $\mathbb{E}(SS_{Res}) = (n - 2)\sigma^2$ y de aquí concluimos que un estimador insesgado para σ^2 es el dado por:

$$\hat{\sigma}^2 = \frac{SS_{Res}}{n - 2} = MS_{Res} \tag{1.2.11}$$

Al valor MS_{Res} se le conocerá a partir de ahora como **error cuadrático medio**. La raíz cuadrada de $\hat{\sigma}^2$ se le conoce como **error estándar de la regresión** y tiene las mismas unidades que la **variable respuesta**.

Definamos además al i -ésimo residual como la diferencia entre el valor observado y_i y su correspondiente valor ajustado, \hat{y}_i . Es decir, $e_i = y_i - \hat{y}_i = y_i - (\hat{\alpha} + \hat{\beta}x_i)$, para $i = 1, 2, \dots, n$. Finalmente, debemos recalcar que el estimador $\hat{\sigma}^2$ es dependiente del modelo que hayamos ajustado ya que está relacionado con los residuos del modelo de regresión.

Retomemos por un momento el experimento realizado por el investigador referente al punto de ebullición del agua en cierta ciudad. No es difícil concluir, al observar el correspondiente diagrama de dispersión, que los grados a los que ebulle el agua están relacionados de manera positiva con la presión barométrica ejercida. Entre mayor presión barométrica exista en la ciudad a mayor temperatura ebullición del agua. Es razonable por tanto, pensar en un modelo de regresión lineal simple para ajustar los datos del investigador y modelar dicha relación.

Nuestra primer tarea será estimar los parámetros de la regresión. Pasaremos directamente al valor de los parámetros estimados, dados por $\hat{\alpha} = 163.9307$ y $\hat{\beta} = 1.5796$. El ajuste por mínimos cuadrados ordinarios será entonces $\hat{y} = 163.9307 + 1.5796x$

¿Cuál es el significado de los parámetros que acabamos de calcular? El parámetro $\hat{\alpha}$ representa la temperatura a la cual ebullición el agua si no existiese presión barométrica (163.9307°F), lo cual resulta inverosímil ya que la presión barométrica es la fuerza ejercida por la atmósfera en todos los objetos existentes. Aún así, ésta sería nuestra interpretación del parámetro $\hat{\alpha}$. Ahora bien, el parámetro $\hat{\beta}$ lo podemos interpretar como el aumento promedio en grados Fahrenheit debido al aumento unitario en la presión barométrica, es decir, ante un aumento unitario en la presión barométrica, se incrementa en promedio la temperatura en 1.5796°F .

Los comandos en R, pertinentes a la estimación de los parámetros $\hat{\alpha}$ y $\hat{\beta}$ para nuestro ejemplo son los siguientes:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 1/Ejercicios")
> PtoEbull<-read.table("PtoEbull.txt",header=T)
> attach(PtoEbull)
> Ajuste<-lm(Ebullición~Presión)
> summary(Ajuste)
```

Call:

```
lm(formula = Ebullición ~ Presión)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.41483	-0.91550	-0.05148	0.76941	2.72840

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	163.9307	2.6551	61.74	< 2e-16 ***
Presión	1.5796	0.1051	15.04	1.88e-10 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.291 on 15 degrees of freedom

Multiple R-squared: 0.9378, Adjusted R-squared: 0.9336

F-statistic: 226 on 1 and 15 DF, p-value: 1.879e-10

Para la estimación de $\hat{\sigma}^2$ tenemos el siguiente código:

```
> anova.lm(Ajuste)
```

Analysis of Variance Table

Response: Ebullición

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Presión	1	376.92	376.92	226.04	1.879e-10 ***
Residuals	15	25.01	1.67		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Del código anterior se desprende que $\hat{\sigma}^2 = MS_{Res} = 1.67$. Observemos que $\hat{\sigma} = 1.2922$, la cual se considera una desviación estándar pequeña tomando en cuenta los valores en grados Fahrenheit. Otra forma de ver la relación que existe entre la **variable predictora** y la **variable respuesta**, además del diagrama de dispersión, es a través de la correlación de Spearman. Realizaremos una prueba de hipótesis para demostrar que existe una correlación (positiva en este caso y con valor igual a 0.9711835) entre la presión barométrica y el punto de ebullición del agua. Al estar correlacionadas las variables de manera positiva implicará que, al aumentar una variable la otra también lo hará.

Rechazamos nuestra hipótesis nula ($H_0 : \rho = 0$ vs $H_1 : \rho \neq 0$, la correlación de Spearman entre las variables involucradas es cero contra que esta sea distinta de cero) ya que tenemos un p-value de $9.454e-11 \approx 0 < 0.05$.

El código en R para obtener la correlación de Spearman está dado por:

```
> cor.test(Presión,Ebullición,method="spearman",alternative="two.sided",conf.level=0.95)

Spearman's rank correlation rho

data:  Presión and Ebullición
S = 23.5143, p-value = 9.454e-11
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9711835
```

En la siguiente sección se anexa una tabla resumiendo los resultados y algunas propiedades del ajuste por mínimos cuadrados ordinarios.

1.3. Propiedades de los estimadores

Una de las propiedades más importantes de los estimadores por mínimos cuadrados ordinarios, $\hat{\alpha}$ y $\hat{\beta}$, es la de ser estimadores insesgados. Para demostrar que $\hat{\beta}$ es insesgado, recordemos que por hipótesis $\mathbb{E}(\epsilon_i) = 0$, y considerando tanto la linealidad del valor esperado como el hecho de que el estimador de β pueda escribirse como combinación lineal de la **variable respuesta** con coeficiente $c_i = \frac{x_i - \bar{x}}{S_{xx}}$ para $i = 1, 2, \dots, n$, tenemos:

$$\begin{aligned}
 \mathbb{E}(\hat{\beta}) &= \mathbb{E}\left(\sum_{i=1}^n c_i y_i\right) \\
 &= \sum_{i=1}^n c_i \mathbb{E}(y_i) \\
 &= \sum_{i=1}^n c_i \mathbb{E}(\alpha + \beta x_i + \epsilon_i) \\
 &= \sum_{i=1}^n c_i [\mathbb{E}(\alpha) + \mathbb{E}(\beta x_i) + \mathbb{E}(\epsilon_i)] \\
 &= \sum_{i=1}^n c_i (\alpha + \beta x_i) \\
 &= \alpha \sum_{i=1}^n c_i + \beta \sum_{i=1}^n c_i x_i \\
 &= \beta
 \end{aligned} \tag{1.3.1}$$

Es posible demostrar que:

$$\sum_{i=1}^n c_i = 0 \quad \text{y} \quad \sum_{i=1}^n c_i x_i = 1 \tag{1.3.2}$$

Concluimos que $\hat{\beta}$ es un estimador insesgado de β . Ahora bien, utilizando el hecho de que los errores no están correlacionados, la varianza de $\hat{\beta}$ será:

$$\begin{aligned}
 \text{Var}(\hat{\beta}) &= \text{Var}\left(\sum_{i=1}^n c_i y_i\right) \\
 &= \sum_{i=1}^n c_i^2 \text{Var}(y_i) \\
 &= \sigma^2 \sum_{i=1}^n c_i^2 \\
 &= \frac{\sigma^2}{S_{xx}} \sum_{i=1}^n (x_i - \bar{x})^2 \\
 &= \frac{\sigma^2}{S_{xx}}
 \end{aligned} \tag{1.3.3}$$

Para demostrar que $\hat{\alpha}$ es insesgado, tenemos el siguiente procedimiento:

$$\begin{aligned}
 \mathbb{E}(\hat{\alpha}) &= \mathbb{E}(\bar{y} - \hat{\beta}\bar{x}) \\
 &= \mathbb{E}(\bar{y}) - \mathbb{E}(\hat{\beta}\bar{x}) \\
 &= \mathbb{E}(\bar{y}) - \beta\mathbb{E}(\bar{x}) \\
 &= \alpha + \beta\mathbb{E}(\bar{x}) - \beta\mathbb{E}(\bar{x}) \\
 &= \alpha
 \end{aligned} \tag{1.3.4}$$

Debido a que $\text{Var}(y_i) = \sigma^2$, a que la covarianza entre \bar{y} y $\hat{\beta}$ puede probarse que es cero y a los resultados anteriores, la varianza de $\hat{\alpha}$ será:

$$\begin{aligned}
 \text{Var}(\hat{\alpha}) &= \text{Var}(\bar{y} - \hat{\beta}\bar{x}) \\
 &= \text{Var}(\bar{y}) + \bar{x}^2 \text{Var}(\hat{\beta}) \\
 &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)
 \end{aligned} \tag{1.3.5}$$

Importante será recordar el **Teorema de Gauss-Markov**, el cual nos habla acerca de la calidad de los estimadores por mínimos cuadrados ordinarios. Enunciamos a continuación dicho teorema. La demostración correspondiente se encuentra en Neter *et al.* (1990).

Teorema 1.3.1 *Sea $y_i = \alpha + \beta x_i + \epsilon_i$, para $i = 1, 2, \dots, n$, con x_1, x_2, \dots, x_n valores fijos conocidos, α y β valores observados y $\epsilon_1, \epsilon_2, \dots, \epsilon_n$, variables aleatorias con $\mathbb{E}(\epsilon_i) = 0$, $\text{Var}(\epsilon_i) = \sigma^2 > 0 \forall i$ y $\text{Cov}(\epsilon_i, \epsilon_j) = 0 \forall i \neq j$. Bajo estas condiciones, los estimadores de α y β por mínimos cuadrados ordinarios, $\hat{\alpha}$ y $\hat{\beta}$ respectivamente, son, dentro de la clase de estimadores lineales insesgados, los mejores, en el sentido de que su varianza es mínima.*

La **Tabla 1.3.1** muestra algunas propiedades y resultados del ajuste por mínimos cuadrados ordinarios. Por ejemplo, una propiedad útil es el hecho de que la suma de los residuos, definidos anteriormente como $e_i = y_i - \hat{y}_i$, en cualquier modelo de regresión que contenga al intercepto, α , siempre será cero. Otra propiedad interesante es que la suma de los valores observados, y_i , es igual a la suma de los valores ajustados, \hat{y}_i . Por último, obtenemos que la suma de los valores ajustados ponderados por su correspondiente residual, es cercana a cero.

i	x_i	y_i	\hat{y}_i	e_i	$x_i e_i$
1	20.79	199.5	196.77	2.73	56.74
2	20.79	199.3	196.77	2.53	52.59
3	22.40	197.9	199.31	-1.41	-31.67
4	22.67	198.4	199.74	-1.34	-30.38
5	23.15	199.4	200.50	-1.10	-25.43
6	23.35	199.9	200.81	-0.91	-21.35
7	23.89	200.9	201.67	-0.77	-18.33
8	23.99	201.1	201.83	-0.73	-17.40
9	24.02	201.9	201.87	0.03	0.66
10	24.01	201.3	201.86	-0.56	-13.37
11	25.14	203.6	203.64	-0.04	-1.05
12	26.57	204.6	205.90	-1.30	-34.56
13	28.49	209.5	208.93	0.57	16.14
14	27.76	208.6	207.78	0.82	22.75
15	29.64	210.7	210.75	-0.05	-1.48
16	29.88	211.9	211.13	0.77	23.03
17	30.06	212.2	211.41	0.79	23.64

$\frac{1}{n} \sum_{i=1}^n x_i = 25.1$	$\sum_{i=1}^n y_i = 3460.7$	$\sum_{i=1}^n \hat{y}_i = 3460.7$	$\sum_{i=1}^n e_i = 0.0$	$\sum_{i=1}^n x_i e_i \approx 0$
---------------------------------------	-----------------------------	-----------------------------------	--------------------------	----------------------------------

Tabla 1.3.1: Propiedades y resultados del ajuste por mínimos cuadrados ordinarios

1.4. Estimación por máxima verosimilitud

A través de secciones anteriores, hemos utilizado el método de **mínimos cuadrados ordinarios** para la estimación de los parámetros del modelo de regresión lineal sin importarnos como se distribuyen los errores ϵ . Ahora bien, si la distribución de los errores ϵ fuese conocida, entonces un método alternativo para la estimación de los parámetros será el método de **máxima verosimilitud**.

Asumamos que los errores pertenecientes al modelo de regresión lineal simple son independientes, distribuidos normalmente con media 0 y varianza σ^2 , es decir, $\epsilon_i \sim NID(0, \sigma^2)$. La función de verosimilitud se encuentra mediante la distribución conjunta del total de observaciones. Dicha función será:

$$\begin{aligned}
 L(x_i, y_i, \alpha, \beta, \sigma^2) &= \prod_{i=1}^n \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \exp \left[-\frac{1}{2\sigma^2} (y_i - \alpha - \beta x_i)^2 \right] \\
 &= \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \right]
 \end{aligned}$$

Definimos la log-verosimilitud como:

$$\ln L(x_i, y_i, \alpha, \beta, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \quad (1.4.1)$$

Los estimadores por máxima verosimilitud, $\hat{\alpha}$, $\hat{\beta}$ y $\tilde{\sigma}^2$ deben satisfacer:

$$\left. \frac{\partial \ln L}{\partial \alpha} \right|_{\alpha, \beta, \sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) = 0 \quad (1.4.2)$$

$$\left. \frac{\partial \ln L}{\partial \beta} \right|_{\alpha, \beta, \sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (y_i - \alpha - \beta x_i) x_i = 0 \quad (1.4.3)$$

$$\left. \frac{\partial \ln L}{\partial \sigma^2} \right|_{\alpha, \beta, \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 = 0 \quad (1.4.4)$$

La solución a las **ecs. 1.4.2, 1.4.3 y 1.4.4** nos brinda los siguientes estimadores por máxima verosimilitud:

$$\begin{aligned} \hat{\alpha} &= \bar{y} - \hat{\beta} \bar{x} \\ \hat{\beta} &= \frac{S_{xy}}{S_{xx}} \\ \tilde{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 \end{aligned} \quad (1.4.5)$$

Nótese que los estimadores por máxima verosimilitud para el **intercepto** y la **pendiente** coinciden con los estimadores por mínimos cuadrados ordinarios. Además, $\tilde{\sigma}^2$ es un estimador sesgado relacionado con el estimador insesgado de $\hat{\sigma}^2$, en la forma, $\tilde{\sigma}^2 = \frac{(n-1)\hat{\sigma}^2}{n}$.

Propiedades estadísticas de los estimadores por máxima verosimilitud

1. Los estimadores de α y β son insesgados, mientras que el estimador de σ^2 es insesgado asintóticamente, es decir, es insesgado conforme n se hace grande.
2. Los estimadores tienen varianza mínima cuando son comparados con los demás estimadores insesgados. Esto gracias al **Teorema de Gauss-Markov** referido en la sección anterior.
3. Los estimadores son **consistentes**, es decir, los estimadores difieren de su valor verdadero en una cuantía muy pequeña conforme n se hace grande.
4. Los estimadores son un conjunto de **estadísticas suficientes**, es decir, los estimadores contienen toda la información de la muestra original de tamaño n con respecto a nuestro parámetro de interés.

1.5. Pruebas de hipótesis sobre α y β

Para realizar pruebas de hipótesis sobre la **pendiente** y el **intercepto**, así como para la construcción de intervalos de confianza, seguimos con el supuesto distribucional en los errores introducido en la **sección 1.4**, es decir, $\epsilon_i \sim NID(0, \sigma^2)$. Se desprende de manera inmediata que las observaciones $y_i \sim NID(\alpha + \beta x_i, \sigma^2)$. Por otro lado, utilizando los cálculos para las varianzas de los estimadores de la **sección 1.3**, tenemos que:

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]\right) \quad \text{y} \quad \hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{S_{xx}}\right) \quad (1.5.1)$$

Ahora bien, supongamos que deseamos probar las hipótesis de que el **intercepto** y la **pendiente** toman los valores $\alpha = \tilde{\alpha}$ y $\beta = \tilde{\beta}$. Si σ^2 fuese conocida, podríamos utilizar los estadísticos:

$$Z_\alpha = \frac{\hat{\alpha} - \tilde{\alpha}}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \quad (1.5.2)$$

y

$$Z_\beta = \frac{\hat{\beta} - \tilde{\beta}}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \quad (1.5.3)$$

cuyas distribuciones son $N(0, 1)$, para probar las hipótesis:

$$H_0 : \alpha = \tilde{\alpha} \quad \text{vs} \quad H_1 : \alpha \neq \tilde{\alpha} \quad (1.5.4)$$

y

$$H_0 : \beta = \tilde{\beta} \quad \text{vs} \quad H_1 : \beta \neq \tilde{\beta} \quad (1.5.5)$$

De no ser conocido σ^2 , se utilizan los estadísticos:

$$t_\alpha = \frac{\hat{\alpha} - \tilde{\alpha}}{\sqrt{MS_{Res} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \quad (1.5.6)$$

y

$$t_\beta = \frac{\hat{\beta} - \tilde{\beta}}{\sqrt{\frac{MS_{Res}}{S_{xx}}}} \quad (1.5.7)$$

cuyas distribuciones son t , comparando los valores obtenidos de t_α y t_β con $t_{1-\frac{\alpha}{2}, n-2}$, el cuantil de una **distribución t** con $n-2$ grados de libertad que acumula $1 - \frac{\alpha}{2}$ de probabilidad, y rechazando la hipótesis nula si $|t_\alpha|, |t_\beta| > t_{1-\frac{\alpha}{2}, n-2}$. Importante será la prueba de hipótesis:

$$H_0 : \beta = 0 \quad vs \quad H_1 : \beta \neq 0 \quad (1.5.8)$$

ya que representa la **significancia de la regresión**. Rechazar H_0 implica que la **variable predictora** es de gran valor al explicar la variabilidad de la **variable respuesta**. Fallar en rechazar H_0 implicaría que la variable \mathbf{x} es de poco valor al explicar a la variable \mathbf{y} , o bien, que la verdadera relación entre la variable \mathbf{x} y la variable \mathbf{y} dista de ser lineal.

Existe otro enfoque para obtener la **significancia de la regresión** y no es otro que el del **análisis de varianza (ANOVA)**. Dicho análisis comienza con la descomposición del total de la variabilidad de \mathbf{y} de la siguiente forma:

$$y_i - \bar{y} = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \quad (1.5.9)$$

Elevando al cuadrado ambos lados de la ecuación y sumando sobre todas las observaciones:

$$\begin{aligned} \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 + 2 \sum_{i=1}^n \hat{y}_i e_i - 2\bar{y} \sum_{i=1}^n e_i \\ &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned} \quad (1.5.10)$$

dado que la suma, sobre toda la muestra n , de los residuos siempre es cero, por propiedades de los estimadores por mínimos cuadrados ordinarios, y la suma de los residuos ponderados por su correspondiente valor ajustado también es igual a cero. Se desprenden algunas propiedades, a saber:

1. $\sum_{i=1}^n (y_i - \bar{y})^2$ es la suma corregida del cuadrado de las observaciones, la cual mide la variabilidad total en las observaciones (SS_T). SS_T tiene $n-1$ grados de libertad.
2. $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ es la variabilidad de las observaciones y_i acometidas por la línea de regresión (SS_R). SS_R tiene 1 grado de libertad.
3. $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ es la variación residual que no logró ser explicada por el modelo de regresión (SS_{Res}). SS_{Res} tiene $n-2$ grados de libertad.

A partir de lo anterior se construye la **prueba F** para probar la hipótesis:

$$H_0 : \beta = 0 \quad vs \quad H_1 : \beta \neq 0 \quad (1.5.11)$$

usando como estadístico:

$$F_0 = \frac{SS_R(n-2)}{SS_{Res}(1)} \quad (1.5.12)$$

el cual sigue una distribución $F_{1,n-2}$. Sea $F_{1-\alpha,1,n-2}$ el cuantil de una **distribución F** con 1 y $n-2$ grados de libertad que acumula $1-\alpha$ de probabilidad. Se rechaza la hipótesis nula, H_0 , si $F_0 > F_{1-\alpha,1,n-2}$. A continuación mostramos la tabla **ANOVA** desplegada por R para nuestro ejemplo del punto de ebullición:

```
> anova.lm(Ajuste)

Analysis of Variance Table

Response: Ebullición
      Df Sum Sq Mean Sq F value    Pr(>F)
Presión  1 376.92  376.92  226.04 1.879e-10 ***
Residuals 15  25.01    1.67
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

¿Qué podemos decir a partir de la tabla **ANOVA** mostrada anteriormente? Observamos que $F_0 = 226.04$ y dado $\alpha = 5\%$, comparamos F_0 con el cuantil $F_{1,15} = 4.54$, por lo que rechazamos H_0 y concluimos que la presión barométrica es de gran valor para explicar al punto de ebullición del agua. La significancia de la regresión en nuestro ejemplo puede obtenerse también a partir de los estadísticos t_α y t_β , o bien, a partir de sus correspondientes p-values, dados a continuación en código R:

```
> summary(Ajuste)

Call:
lm(formula = Ebullición ~ Presión)

Residuals:
      Min       1Q   Median       3Q      Max
-1.41483 -0.91550 -0.05148  0.76941  2.72840

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 163.9307     2.6551   61.74 < 2e-16 ***
Presión      1.5796     0.1051   15.04 1.88e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.291 on 15 degrees of freedom
Multiple R-squared:  0.9378, Adjusted R-squared:  0.9336
F-statistic:  226 on 1 and 15 DF, p-value: 1.879e-10
```

Dado $\alpha = 5\%$, observamos para el estimador $\hat{\alpha}$, un p-value de $2e - 16 \approx 0 < 0.05$ para probar la hipótesis $H_0 : \alpha = 0$ vs $H_1 : \alpha \neq 0$, por lo que se rechaza la hipótesis nula y se puede decir que este parámetro es significativamente distinto de cero. Análogamente se puede realizar el mismo procedimiento para $\hat{\beta} = 1.5796$, llegando a la misma conclusión, es decir, que dicho parámetro es significativamente distinto de cero.

1.6. Intervalos de confianza

Además de realizar pruebas de hipótesis sobre los parámetros, será de nuestro interés la construcción de intervalos de confianza para los parámetros α , β y σ^2 . Recordemos, por una parte, que los errores ϵ_i son independientes y siguen una **distribución Normal**, y por otra, que tanto:

$$\frac{\hat{\alpha} - \alpha}{\sqrt{MS_{Res} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}} \quad \text{y} \quad \frac{\hat{\beta} - \beta}{\sqrt{\frac{MS_{Res}}{S_{xx}}}}$$

siguen una **distribución t** con $n - 2$ grados de libertad. Utilizando la misma notación para los cuantiles ya introducida, el intervalo de confianza al $100(1 - \alpha)\%$ para α y β es:

$$\left(\hat{\alpha} - t_{1-\frac{\alpha}{2}, n-2} \sqrt{MS_{Res} \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]}, \hat{\alpha} + t_{1-\frac{\alpha}{2}, n-2} \sqrt{MS_{Res} \left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right]} \right) \quad (1.6.1)$$

y

$$\left(\hat{\beta} - t_{1-\frac{\alpha}{2}, n-2} \sqrt{\frac{MS_{Res}}{S_{xx}}}, \hat{\beta} + t_{1-\frac{\alpha}{2}, n-2} \sqrt{\frac{MS_{Res}}{S_{xx}}} \right) \quad (1.6.2)$$

Sólo nos resta la construcción del intervalo de confianza para σ^2 . Se puede demostrar que la distribución de $\frac{(n-2)(MS_{Res})}{\sigma^2}$ es ji-cuadrada con $n - 2$ grados de libertad. De aquí que el intervalo de confianza al $100(1 - \alpha)\%$ para σ^2 está dado por:

$$\left(\frac{(n-2)MS_{Res}}{\chi_{\frac{\alpha}{2}, n-2}^2}, \frac{(n-2)MS_{Res}}{\chi_{1-\frac{\alpha}{2}, n-2}^2} \right) \quad (1.6.3)$$

A continuación mostramos el intervalo de confianza al 95% y 90% para la **pendiente** y el **intercepto** para nuestro ejemplo en código R:

```
> confint(Ajuste)

                2.5 %      97.5 %
(Intercept) 158.271527 169.589940
Presión      1.355702   1.803592
```

```
> confint(Ajuste,level=0.90)

              5 %      95 %
(Intercept) 159.27621 168.585254
Presión     1.39546   1.763835
```

Observemos como en ambos intervalos de confianza no está incluido el cero. Este hecho es una forma adicional, más no suficiente, para corroborar el hecho de que tanto α como β son significativamente distintos a cero.

1.7. Coeficiente de determinación

Recordemos por un momento que SS_T nos mide la variabilidad de \mathbf{y} sin considerar el efecto de la variable \mathbf{x} y SS_{Res} nos mide la variabilidad de \mathbf{y} después de haber considerado a la variable \mathbf{x} . De aquí que el cociente:

$$R^2 = 1 - \frac{SS_{Res}}{SS_T} \quad (1.7.1)$$

sea conocido como **coeficiente de determinación**. Dicho coeficiente mide la variación de \mathbf{y} explicada por la variable explicativa \mathbf{x} . Debido a que $0 \leq SS_{Res} \leq SS_T$ se sigue que $0 \leq R^2 \leq 1$. Valores cercanos a 1 implican que la mayor parte de la variabilidad en \mathbf{y} es explicada por el modelo de regresión lineal. Valores cercanos a 0 implican que la variable explicativa \mathbf{x} no es de gran valor para explicar a la variable \mathbf{y} .

El **coeficiente de determinación** para nuestro ejemplo se puede obtener en R mediante el siguiente código, el cual ya se mostró en la **sección 1.2 y 1.5**:

```
> summary(Ajuste)

Call:
lm(formula = Ebullición ~ Presión)

Residuals:
    Min       1Q   Median       3Q      Max
-1.41483 -0.91550 -0.05148  0.76941  2.72840

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 163.9307     2.6551   61.74 < 2e-16 ***
Presión     1.5796     0.1051   15.04 1.88e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.291 on 15 degrees of freedom
Multiple R-squared:  0.9378, Adjusted R-squared:  0.9336
F-statistic: 226 on 1 and 15 DF, p-value: 1.879e-10
```

De aquí se concluye que $R^2 = 0.9378$, es decir, el 93.78 % de la variabilidad del punto de ebullición es explicada por la presión barométrica.

1.8. Verificación del ajuste del modelo

A lo largo de este capítulo hemos hecho varias hipótesis que deben cumplirse para la aplicación del modelo de regresión lineal simple. En particular, las hipótesis más importantes y las que nos servirán en esta sección para el análisis de los residuos y por tanto, para verificar la validez del modelo de regresión lineal, serán las siguientes:

1. La relación existente entre las variables explicativas y la variable respuesta, \mathbf{y} , es lineal.
2. Los errores, ϵ , tienen media 0 y varianza constante e igual a σ^2 .
3. Los errores, ϵ , son no-correlacionados.
4. Los errores, ϵ , se distribuyen normalmente. Junto con el supuesto anterior, esto implicaría que los errores, ϵ , son variables aleatorias independientes.

Estas serán nuestras hipótesis básicas y con base en ellas utilizaremos distintos diagnósticos para evitar consecuencias serias en la implementación de nuestro modelo. En la sección **1.2**, definimos a los **residuos** como $e_i = y_i - \hat{y}_i$, para $i = 1, 2, \dots, n$. Observemos que los residuos, e_i , pueden ser vistos como la desviación que existe entre la observación y_i y su correspondiente valor ajustado \hat{y}_i , es decir, los residuos, e_i , son una medida de la variabilidad en \mathbf{y} no explicada por la variable predictora \mathbf{x} . Dado que los errores verdaderos ϵ_i no son observables, podemos considerar a los residuos como la realización de estos errores. Así pues, cualquier propiedad asumida sobre los errores verdaderos, deberá verificarse a través de los residuos.

De gran importancia será para el estudio de nuestro modelo, identificar **valores extremos**, observaciones que se encuentran alejadas de manera importante del resto de nuestras observaciones. Para identificar dichos valores extremos, definimos los **residuos estandarizados** y los **residuos estudentizados** como:

$$d_i = \frac{e_i}{\sqrt{MS_{Res}}} \quad (1.8.1)$$

y

$$r_i = \frac{e_i}{\sqrt{MS_{Res} \left(1 - \left[\frac{1}{n} + \frac{(x_i - \bar{x})^2}{S_{xx}} \right] \right)}} \quad (1.8.2)$$

donde por una parte, $d_i \sim N(0, 1)$ aproximadamente (valores $d_i > 3$ en la **ec. 1.8.1** indican un posible valor extremo), y por otra, observamos en la **ec. 1.8.2** que, observaciones x_i cercanas al promedio \bar{x} hacen que $x_i - \bar{x}$ sea pequeña y por tanto r_i grande. De igual manera, observaciones x_i lejanas al promedio \bar{x} hacen que $x_i - \bar{x}$ sea grande y por tanto r_i pequeña.

Otra forma para investigar la validez del modelo es a través de la gráfica de residuos. Debemos resaltar la importancia de la distribución normal en los errores, ϵ , ya que a partir de esta hipótesis hemos realizado la construcción de los estadísticos **F**, **t**, intervalos de confianza, etc. Desviaciones potenciales de dicha distribución afectarán a nuestro modelo de manera significativa, llevándonos a modelos inestables, es decir, al tomar una muestra distinta del problema en cuestión, podríamos obtener un modelo completamente distinto con conclusiones contradictorias a las que habíamos adoptado en un principio.

Así pues, tomando siempre en consideración lo anterior, la primera gráfica que analizaremos será la **gráfica de probabilidad normal**. Sean $e_{[1]} < e_{[2]} < e_{[3]} < \dots < e_{[n]}$, los n residuos de la muestra, ordenados de manera ascendente. Sea $P_i = \frac{1}{n}(i - \frac{1}{2})$, $i = 1, 2, 3, \dots, n$, los valores esperados de las estadísticas de orden de una **distribución Normal estándar**. Al graficar el conjunto e_i ordenado contra su respectiva probabilidad acumulada, P_i , debiésemos obtener aproximadamente una línea recta. Desviaciones de dicha línea implicaría una distribución diferente a la normal de nuestros residuos, e_i .

En la **Figura 1.8.1** se muestran gráficas de probabilidad idóneas así como pequeñas desviaciones. El panel (a) y (b) de la **Figura 1.8.1** muestran aproximaciones ideales a la gráfica de probabilidad normal, mientras que el panel (c) representa la existencia de colas pesadas en la distribución de los residuos. Finalmente, el panel (d) muestra la presencia de sesgo negativo en los residuos.

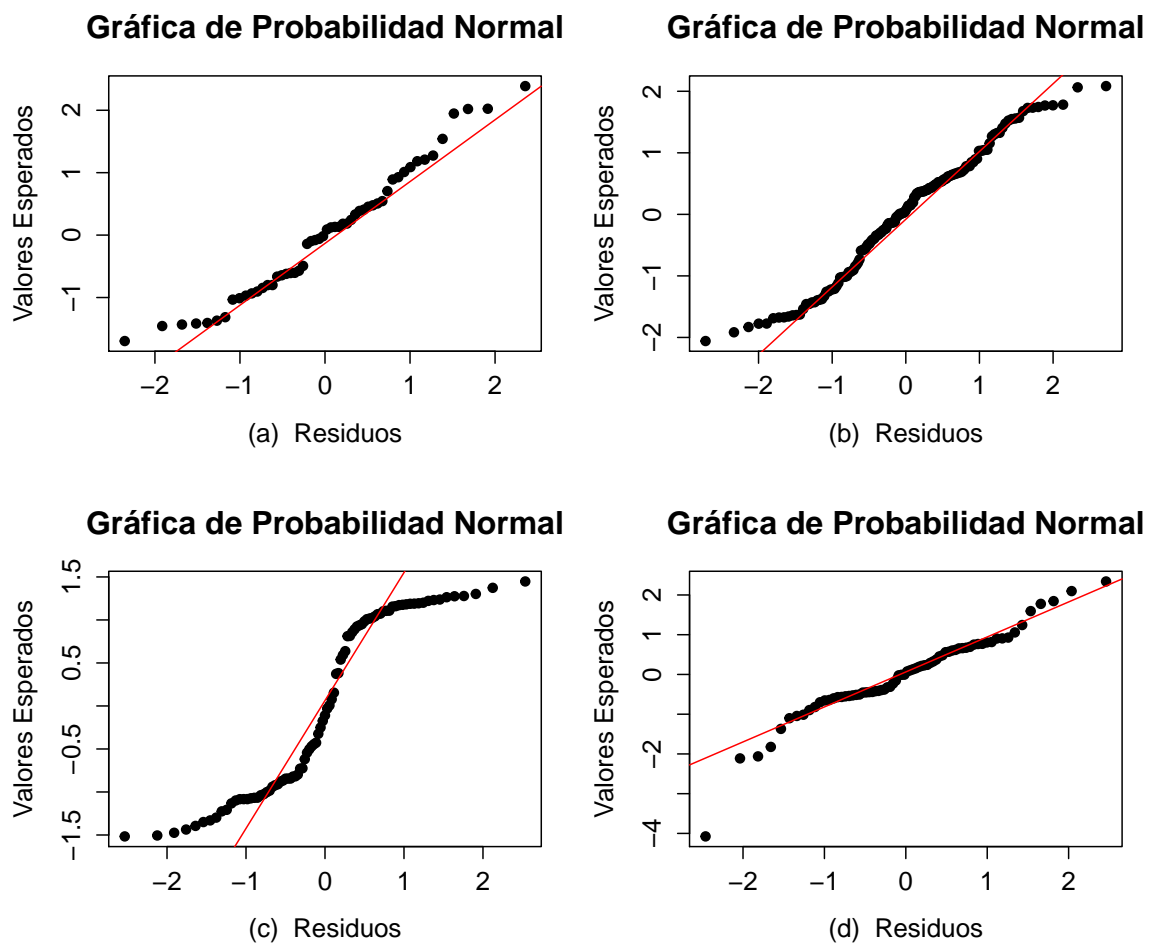


Figura 1.8.1: Figura que muestra distintas versiones de la gráfica de probabilidad normal

Nuestra siguiente gráfica a considerar será aquella que relacione los residuos, (e_i, d_i, r_i) , con su correspondiente valor ajustado \hat{y}_i . Realizar esta gráfica nos confirmará la hipótesis de **homocedasticidad** o varianza constante así como la existencia de posibles valores extremos.

Si la gráfica no tiene ningún patrón, es simétrica alrededor del cero y la mayoría de los puntos están comprendidos entre los valores -2 y 2 (el 95.5% de los residuos debe de encontrarse dentro de 2 desviaciones estándar mientras que el 99.7% de los mismos, en 3 desviaciones estándar) significa que no hay defectos obvios en el modelo y que podemos considerar que los errores tienen varianza constante. Ante las desviaciones de la gráfica idónea, por ejemplo, para **heteroscedasticidad**, podemos aplicar transformaciones a la variable predictora, respuesta o ambas, para así, estabilizar la varianza. Ante la no-linealidad debemos de transformar las variables del modelo.

El panel (a) de la **Figura 1.8.2** no muestra la existencia de algún patrón en la relación de los residuos y sus valores ajustados. Es en este caso donde se confirma la hipótesis de homoscedasticidad. Los paneles (b), (c) y (d) de la **Figura 1.8.2** muestran patrones en los residuos, explicando así la existencia de heteroscedasticidad.

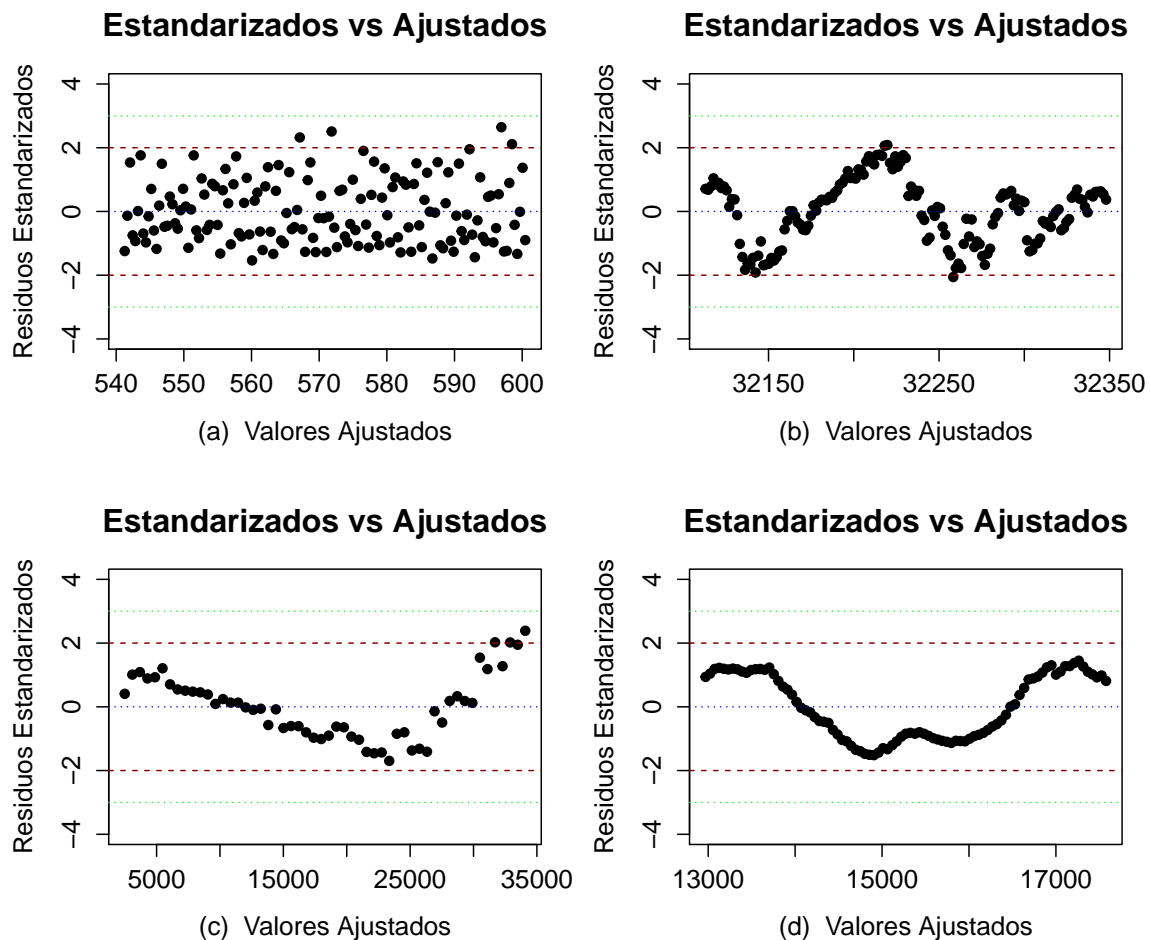


Figura 1.8.2: Figura para corroborar el supuesto de homoscedasticidad

Retomemos nuestro ejemplo. A continuación mostramos el código en R para el cálculo de los valores ajustados, residuos, residuos estandarizados y residuos estudentizados, cuyos valores se muestran en la **Tabla 1.8.1**. En la **Figura 1.8.3** se muestra la gráfica de probabilidad normal junto con una figura que representa a los residuos. Además, en la **Figura 1.8.4** se muestra el histograma junto con el diagrama de caja y bigotes (*Box-Plot*) de los residuos. La interpretación de estas figuras y resultados se explica más adelante. Anexamos también la curtosis y el sesgo:

```

> ygorro<-fitted(Ajuste)
> residuos<-resid(Ajuste)
> residuosestandarizados<-rstandard(Ajuste)
> residuosestudentizados<-rstudent(Ajuste)

> skewness(residuos)
[1] 0.9542871

> kurtosis(residuos)
[1] 3.004527

```

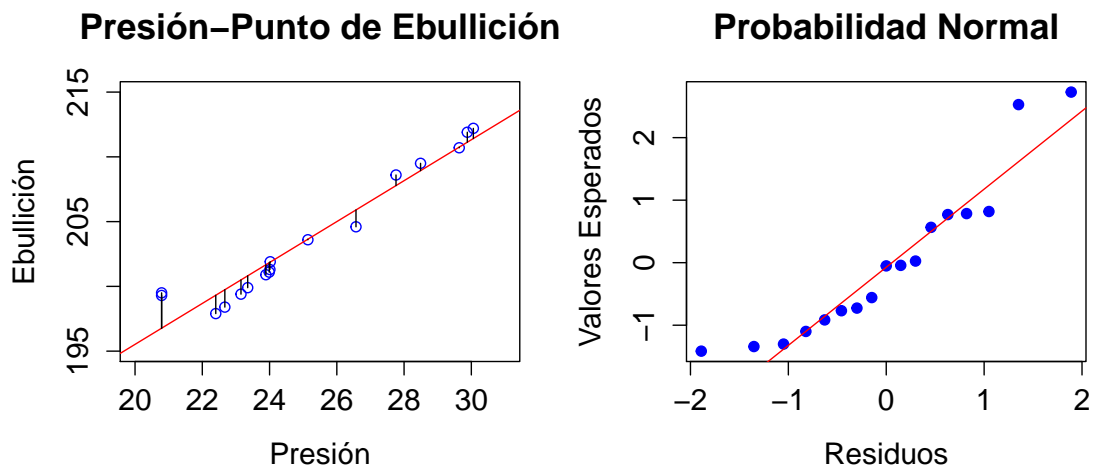


Figura 1.8.3: Figura que muestra a los residuos y gráfica de probabilidad normal

i	x_i	y_i	\hat{y}_i	e_i	d_i	r_i
1	20.79	199.5	196.7716	2.72839950	2.33537777	2.82820054
2	20.79	199.3	196.7716	2.52839950	2.16418746	2.52114061
3	22.40	197.9	199.3148	-1.41483262	-1.15935568	-1.17387171
4	22.67	198.4	199.7413	-1.34133739	-1.09354503	-1.10127482
5	23.15	199.4	200.4996	-1.09956808	-0.88962208	-0.88306841
6	23.35	199.9	200.8155	-0.91549754	-0.73873085	-0.72702956
7	23.89	200.9	201.6685	-0.76850707	-0.61660378	-0.60339192
8	23.99	201.1	201.8265	-0.72647180	-0.58239974	-0.56912308
9	24.02	201.9	201.8739	0.02613878	0.02095015	0.02024006
10	24.01	201.3	201.8581	-0.55806474	-0.44732131	-0.43506502
11	25.14	203.6	203.6431	-0.04306617	-0.03437731	-0.03321294
12	26.57	204.6	205.9020	-1.30196178	-1.04732960	-1.05097300
13	28.49	209.5	208.9349	0.56511544	0.47058600	0.45802283
14	27.76	208.6	207.7817	0.81825796	0.67012899	0.65732037
15	29.64	210.7	210.7515	-0.05147893	-0.04444974	-0.04294536
16	29.88	211.9	211.1306	0.76940572	0.67055609	0.65775225
17	30.06	212.2	211.4149	0.78506921	0.68929854	0.67672977

Tabla 1.8.1: Tabla que muestra los residuos, residuos estandarizados y estudentizados

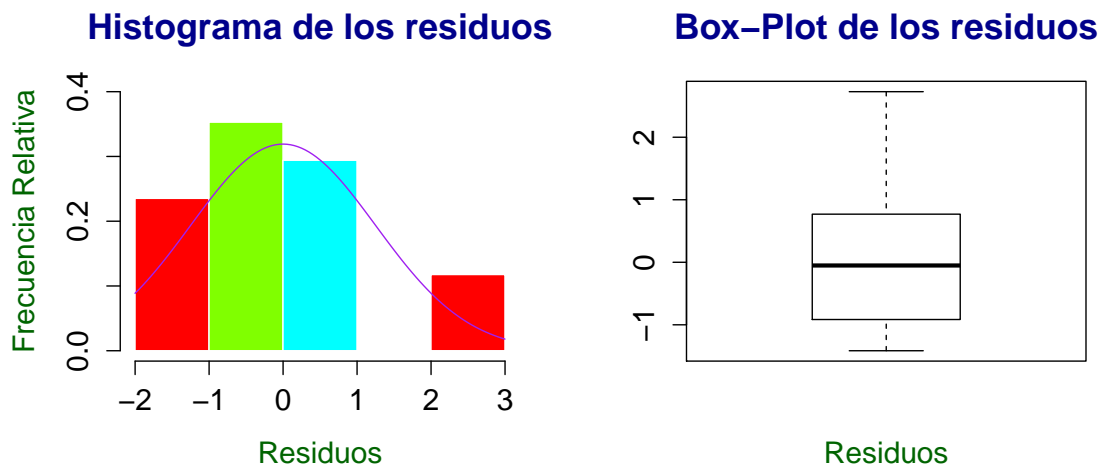


Figura 1.8.4: Figura que muestra el histograma y el diagrama de caja y bigotes de los residuos

Además agregamos la prueba Anderson-Darling para normalidad en los residuos:

```
> library(nortest)
> library(lmtest)
> library(moments)

> ad.test(rnorm(residuos,mean=0,sd=1))
```

Anderson-Darling normality test

```
data:  rnorm(residuos, mean = 0, sd = 1)
A = 0.4022, p-value = 0.3199
```

Por último agregamos la gráfica que nos confirmará la **homoscedasticidad**:

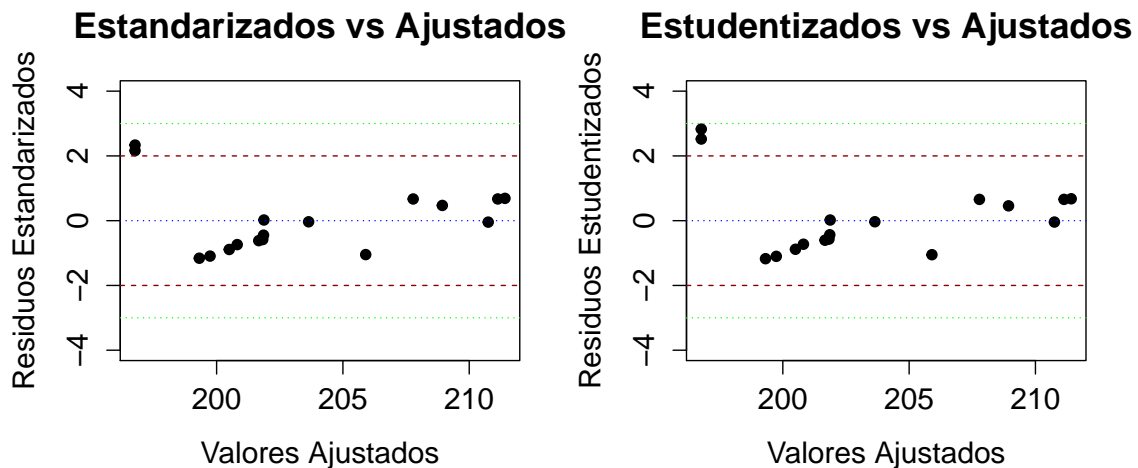


Figura 1.8.5: Figura de residuos (estandarizados y estudentizados) contra valores ajustados

¿Qué podemos decir de los resultados mostrados por el código? Observemos de la **Figura 1.8.3**, la gráfica de probabilidad normal. Si los residuos tuviesen distribución normal, los puntos caerían aproximadamente en una línea recta, pero esto no sucede en nuestro ejemplo. Por tanto, nuestra primer conclusión a partir de la gráfica de probabilidad normal es que los residuos no se distribuyen normalmente. Asimismo, esto se podría corroborar en la **Figura 1.8.4**, tanto con el histograma como con el diagrama de caja y bigotes, que en nuestro ejemplo nos muestra claramente que la distribución de los residuos dista de ser normal. Observamos que el diagrama de caja y bigotes no es simétrico debido a la presencia de sesgo positivo (0.9542871). Como bien sabemos, el sesgo mide la simetría de la distribución de una variable. Una distribución normal estándar tiene sesgo igual a 0. Un sesgo mayor a cero nos dice que hay más residuos negativos que positivos, lo cual corresponde a nuestro ejemplo. Además, el diagrama de caja y bigotes no muestra la existencia de valores extremos (también conocidos como *outliers*). La curtosis refleja el peso de las colas de la distribución en relación al valor central. Generalmente, la curtosis se compara con el valor 3 para medir la distancia respecto al valor normal. Un valor menor que 3 en la curtosis, es evidencia de colas más pesadas que la normal. Un valor mayor que 3 en la curtosis, es evidencia de colas más ligeras. En nuestro ejemplo, del código mostrado en R, la curtosis = 3.004527 \approx 3.

Pasemos a la interpretación de la prueba Anderson-Darling. Observamos del código en R, que el p-value para la prueba es igual a 0.3199. Nuestra prueba de hipótesis para la distribución de los residuos será:

$$H_0 : \text{Normalidad en los residuos} \quad vs \quad H_1 : \text{No normalidad en los residuos} \quad (1.8.3)$$

y con un nivel de significancia de $\alpha = 5\%$, entonces, no se rechaza H_0 , lo cual nos indica que las desviaciones que se ven en las gráficas respecto a la normal no son tan graves.

Por lo tanto, tomando en cuenta la gráfica de probabilidad normal, el histograma, el diagrama de caja y bigotes, así como la prueba Anderson-Darling podemos decir que los residuos tienen una distribución normal.

Finalmente, la **Figura 1.8.5** que muestra la relación existente entre los residuos, estandarizados y estudentizados, d_i y r_i , contra los correspondientes valores ajustados \hat{y}_i , nos será útil para detectar insuficiencias en el modelo. Si la gráfica no tiene ningún patrón, significa que no hay defectos obvios en el modelo y que podemos considerar que los errores tienen varianza constante. En nuestro ejemplo, no existe patrón alguno en la gráfica de residuos estandarizados y estudentizados (a pesar de que existen dos posibles valores extremos a examinar), por lo que, a partir de la gráfica, podríamos concluir que la varianza es constante. Sin embargo, podemos realizar una prueba numérica para corroborar este supuesto. Dicha prueba es la llamada **prueba de Breusch-Pagan**, la cual tiene como hipótesis H_0 : La varianza es constante *vs* H_1 : La varianza no es constante, y viene dada por el siguiente código en R:

```
> bptest(Ebullición~Presión,data=PtoEbull)

studentized Breusch-Pagan test

data:  Ebullición ~ Presión
BP = 6.1473, df = 1, p-value = 0.01316
```

\therefore Con un nivel de significancia de $\alpha = 5\%$, rechazamos la hipótesis nula (el p-value es menor al nivel de significancia) y concluimos que la varianza en los residuos no es constante.

1.9. Modelos de regresión con errores autocorrelacionados

En campos como la economía, finanzas, econometría, etc, encontramos modelos de regresión cuyas variables explicativas y respuesta están relacionadas de manera secuencial en intervalos de tiempo. Dichos modelos son conocidos como **datos de series de tiempo**. Modelos de regresión que usan dichos datos vuelven inapropiado el supuesto de no-correlación o independencia en los errores, ϵ_i , es decir, ahora, el valor de un error dependerá del valor de otro error. A partir de ahora, los errores ϵ_i serán **autocorrelacionados**. Se desprenden los siguientes efectos:

1. Los parámetros estimados de la regresión por mínimos cuadrados ordinarios siguen siendo insesgados pero dejan de tener varianza mínima al compararlos con otros estimadores.
2. Los intervalos de confianza y pruebas de hipótesis basadas en las distribuciones **t** y **F** dejan de ser apropiadas.
3. MS_{Res} subestima de manera considerable a σ^2 si los errores, ϵ_i , están autocorrelacionados de manera positiva. De manera inmediata, tanto:

$$\sqrt{\frac{MS_{Res}}{S_{xx}}} \quad \text{y} \quad \sqrt{MS_{Res} \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)}$$

serán muy pequeños y por tanto, los intervalos de confianza serán mucho más estrechos de lo que deberían ser en principio, dando la falsa impresión de buena precisión del modelo.

En esta sección utilizaremos la prueba de **Durbin-Watson** como medio para detectar autocorrelación. Dicha prueba supone a los errores, ϵ_i , como un proceso autorregresivo de primer orden, **AR(1)** (dicho proceso se explicará en el **capítulo 3**), es decir, supone:

$$\epsilon_t = \rho\epsilon_{t-1} + a_t \tag{1.9.1}$$

donde ϵ_t es el error del modelo a tiempo t , $\{a_t\}$ es un conjunto de variables aleatorias independientes, distribuidas normalmente con media cero y varianza σ_a^2 , es decir, $a_t \sim NID(0, \sigma_a^2)$ y $|\rho| < 1$ es el parámetro de autocorrelación. El modelo de regresión lineal simple con proceso autorregresivo AR(1) en los errores será:

$$\begin{aligned} y_t &= \alpha + \beta x_t + \epsilon_t \\ &= \alpha + \beta x_t + \rho\epsilon_{t-1} + a_t \\ &= \alpha + \beta x_t + \rho(\rho\epsilon_{t-2} + a_{t-1}) + a_t \end{aligned} \tag{1.9.2}$$

Observemos que:

$$\epsilon_t = \sum_{i=0}^{\infty} \rho^i a_{t-i} \tag{1.9.3}$$

es una combinación lineal de las observaciones $a_t \sim NID(0, \sigma_a^2)$ recientes y pasadas.

Se puede demostrar que para un modelo AR(1), se tiene que $\mathbb{E}(\epsilon_t) = 0$ y además:

$$\text{Var}(\epsilon_t) = \sigma_a^2 \frac{1}{1 - \rho^2} \quad (1.9.4)$$

$$\text{Cov}(\epsilon_t, \epsilon_{t+i}) = \rho^{|i|} \frac{\sigma_a^2}{1 - \rho^2}$$

Esto nos dice que los errores tienen media 0, varianza constante e igual a $\frac{\sigma_a^2}{1 - \rho^2}$, pero están autocorrelacionados a menos que $\rho = 0$. De este punto se desprende la prueba de **Durbin-Watson** para correlación positiva con hipótesis nula y alternativa:

$$H_0 : \rho = 0 \quad vs \quad H_1 : \rho > 0 \quad (1.9.5)$$

con estadístico:

$$d = \frac{\sum_{t=2}^n (\epsilon_t - \epsilon_{t-1})^2}{\sum_{t=1}^n \epsilon_t^2} \quad (1.9.6)$$

rechazando $H_0 : \rho = 0$ si $d < d_L$, no rechazando $H_0 : \rho = 0$ si $d > d_U$ y no se puede concluir si $d_L \leq d \leq d_U$. Los límites de d_L y d_U se muestran en el anexo de tablas. Ahora bien, si deseásemos una prueba para correlación negativa, tendríamos que usar la prueba de **Durbin-Watson** con hipótesis nula y alternativa:

$$H_0 : \rho = 0 \quad vs \quad H_1 : \rho < 0 \quad (1.9.7)$$

y estadístico:

$$D = 4 - d \quad (1.9.8)$$

tratando a D como si fuese el estadístico d para correlación positiva.

1.10. Conclusiones

En esta sección realizaremos un estudio similiar al hecho en la **sección 1.8** con datos del financiamiento (millones de pesos) otorgado a través de la Banca Privada y Mixta pero ahora usando al tiempo como variable explicativa y al financiamiento como nuestra variable respuesta. Los datos fueron tomados de Guerrero (2003). La **Tabla 1.10.1** muestra los datos a utilizar.

Año	1978	1979	1980	1981	1982
Enero	4,500	10094	12727	19662	37002
Febrero	8085	11464	12464	20496	41774
Marzo	9081	11501	12850	18214	38625
Abril	8703	12092	13990	19095	42909
Mayo	9486	11932	16023	19232	43123
Junio	11497	12134	16504	26218	45870
Julio	9585	12913	15628	25040	
Agosto	9385	10918	15722	29047	
Septiembre	9794	14005	14376	30374	
Octubre	10246	11631	14772	30223	
Noviembre	10732	12533	15493	30524	
Diciembre	10993	13112	14754	38205	

Tabla 1.10.1: Financiamiento otorgado a través de la Banca Privada y Mixta

Anexaremos la prueba de **Durbin-Watson** al estudio realizado, llegando así a la conclusión de que los errores están autocorrelacionados, haciendo inapropiado, por tanto, el uso del modelo de regresión lineal simple para modelar los datos, dando pie a la introducción de los modelos de **series de tiempo**. La **Figura 1.10.1** muestra la relación existente entre la variable explicativa (tiempo) y la variable respuesta (financiamiento otorgado por parte de la Banca Privada y Mixta). A continuación mostramos el código utilizado en R para realizar el análisis correspondiente. La interpretación de la **Figura 1.10.2**, la **Figura 1.10.3** y la **Figura 1.10.4**, código correspondiente y resultados se explica más adelante:

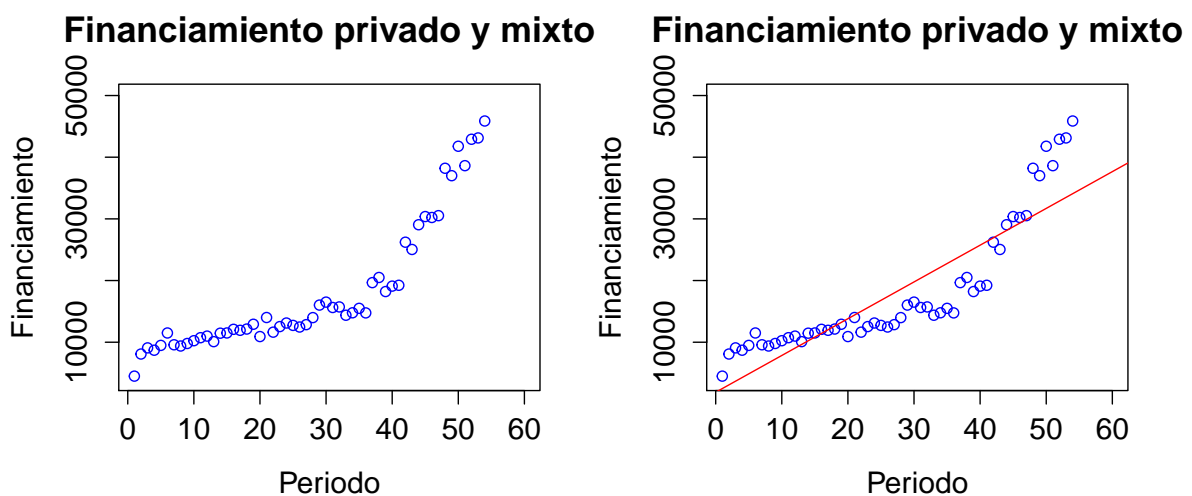


Figura 1.10.1: Figura que muestra el financiamiento dado a través de Banca Privada y Mixta

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 1/Ejercicios")
> Financiamiento<-read.table("Financiamiento.txt",header=T)
> attach(Financiamiento)
> Ajuste<-lm(Financ~Periodo)
> summary(Ajuste)
```

Call:

```
lm(formula = Financ ~ Periodo)
```

Residuals:

Min	1Q	Median	3Q	Max
-8597.3	-4062.1	180.2	2678.2	11788.6

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1891.07	1414.98	1.336	0.187
Periodo	596.12	44.76	13.317	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5127 on 52 degrees of freedom
 Multiple R-squared: 0.7733, Adjusted R-squared: 0.7689
 F-statistic: 177.3 on 1 and 52 DF, p-value: < 2.2e-16

```
> anova.lm(Ajuste)
```

Analysis of Variance Table

Response: Financ

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Periodo	1	4661382933	4661382933	177.34	< 2.2e-16 ***
Residuals	52	1366838336	26285353		

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> confint(Ajuste)
```

	2.5 %	97.5 %
(Intercept)	-948.2996	4730.4310
Periodo	506.2914	685.9435

```
> ygorro<-fitted(Ajuste)
> residuos<-resid(Ajuste)
> residuosestandarizados<-rstandard(Ajuste)
> residuosestudentizados<-rstudent(Ajuste)
> par(mfrow=c(1,2))
> plot(Periodo,Financ,pch=21,xlim=c(1,60),ylim=c(4000,50000),xlab="Periodo",
  ylab="Financiamiento",main="Financiamiento privado-mixto",col="blue")
> abline(Ajuste,col="red")
> segments(Periodo,fitted(Ajuste),Periodo,Financ,col="blue")
> qqnorm(residuos,xlab="Residuos",ylab="Valores Esperados",
  main="Probabilidad Normal",pch=19,col="blue")
> qqline(residuos,col=2)
```

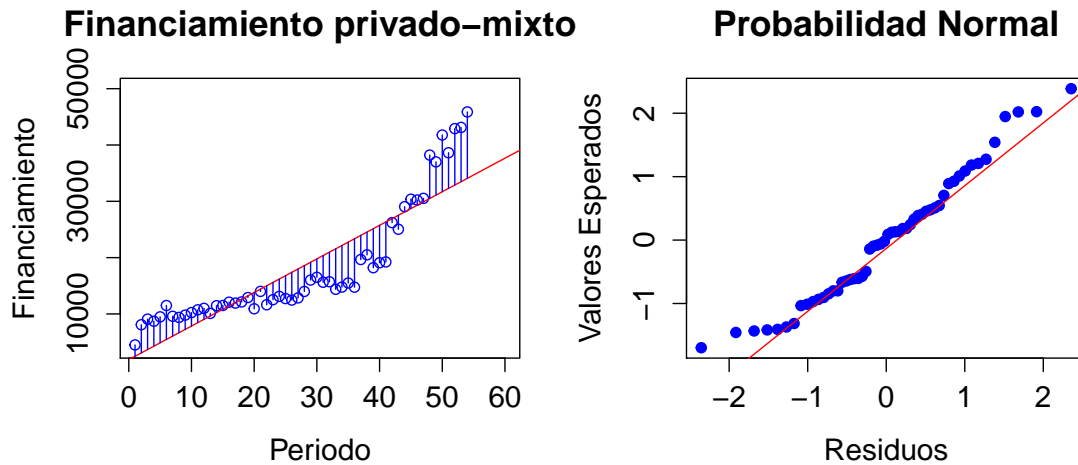


Figura 1.10.2: Figura que muestra a los residuos y gráfica de probabilidad normal

```
> hist(residuos,col=rainbow(4),main="Histograma de los residuos",xlab="Residuos",
  ylab="Frecuencia Absoluta",col.main="darkblue",col.axis="black",col.lab="darkgreen",
  nclass="scott",ylim=c(0,20))
> curve(dnorm(x,mean(residuos),sd(residuos)),add=T,col="gold")
> boxplot(residuos,xlab="Residuos",main="Box-Plot de los residuos",
  col.main="darkblue",col.lab="darkgreen")
```

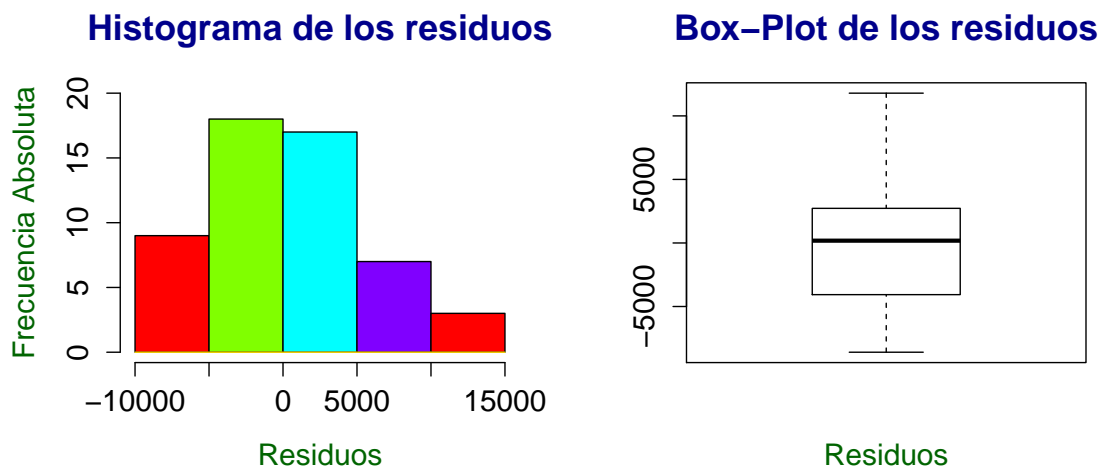


Figura 1.10.3: Histograma y diagrama de caja y bigotes de los residuos (financiamiento)

```

> plot(ygorro,residuosestandarizados,xlab="Valores Ajustados",
  ylab="Residuos Estandarizados",main="Estandarizados vs Ajustados",pch=19,ylim=c(-4,4))
> abline(h=0,col="blue",lty=3)
> abline(h=c(-2,2),col="darkred",lty=2)
> abline(h=c(-3,3),col="green",lty=3)
> plot(ygorro,residuosestudentizados,xlab="Valores Ajustados",
  ylab="Residuos Estudentizados",main="Estudentizados vs Ajustados",pch=19,ylim=c(-4,4))
> abline(h=0,col="blue",lty=3)
> abline(h=c(-2,2),col="darkred",lty=2)
> abline(h=c(-3,3),col="green",lty=3)

```

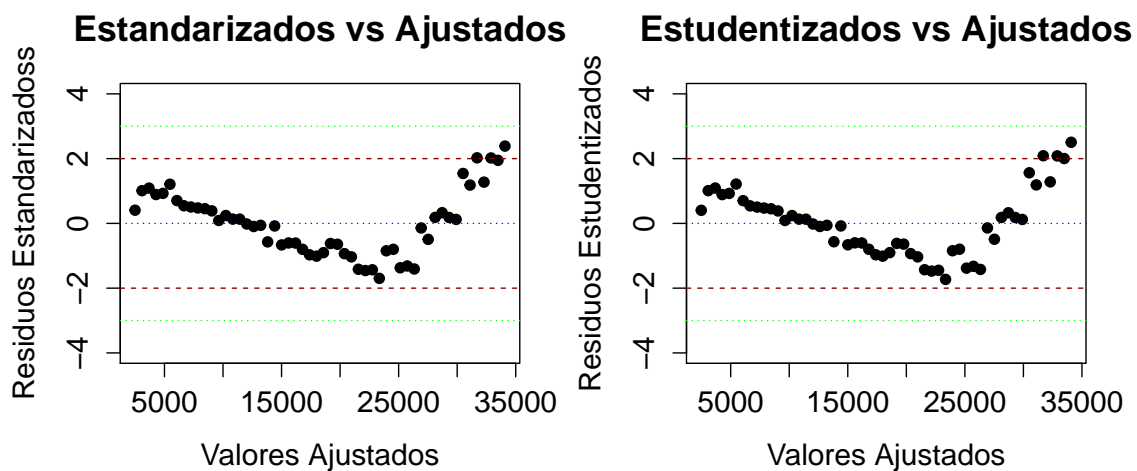


Figura 1.10.4: Figura de residuos (estandarizados y estudentizados) contra valores ajustados

```

> ad.test(rnorm(residuosestandarizados,mean=0,sd=1))

```

Anderson-Darling normality test

```

data:  rnorm(residuosestandarizados, mean = 0, sd = 1)
A = 0.7005, p-value = 0.06364

```

```

> skewness(residuos)

```

```

[1] 0.3638064

```

```

> kurtosis(residuos)

```

```

[1] 2.423226

```

```

> dwtest(Ajuste)

```

Durbin-Watson test

```

data:  Ajuste
DW = 0.1788, p-value < 2.2e-16
alternative hypothesis: true autocorrelation is greater than 0

```


¿Qué podemos decir de los resultados mostrados por el código anterior? Los estimadores para el intercepto y la pendiente, $\hat{\alpha}$ y $\hat{\beta}$, son 1891.07 y 596.12 respectivamente. Esto querrá decir que a tiempo cero, el financiamiento otorgado por la Banca Privada y Mixta fue de \$1,891,070,000 y que al aumentar una unidad de tiempo, el financiamiento aumenta, en promedio, \$596,120,000. Además, el 77.33 % de la variabilidad en el financiamiento es explicada por el modelo de regresión.

Observamos para el estimador $\hat{\alpha} = 1891.07$ un p-value de $0.187 > 0.05$, para probar la hipótesis $H_0 : \alpha = 0$ vs $H_1 : \alpha \neq 0$, por lo que no se rechaza la hipótesis nula y se puede decir que este parámetro puede ser cero. Esto se puede corroborar con el intervalo de confianza para el intercepto ya que este contiene al cero. Asimismo, observamos para el estimador $\hat{\beta} = 596.12$ un p-value de $2e - 16 \approx 0 < 0.05$ para probar la hipótesis $H_0 : \beta = 0$ vs $H_1 : \beta \neq 0$, por lo que se rechaza la hipótesis nula y se puede decir que este parámetro es significativamente distinto de cero.

Ahora bien, pasemos a la **Figura 1.10.2**. Observemos que los puntos, en la gráfica de probabilidad normal, caen en su mayoría en una línea recta por lo que podemos concluir que los residuos tienen una distribución aproximadamente normal. Esto igualmente lo podemos corroborar tanto con el histograma como con el diagrama de caja y bigotes, mostrados en la **Figura 1.10.3**. Para una conclusión más certera veamos la prueba Anderson-Darling. Observamos que el p-value para dicha prueba es de $0.0636 > 0.05$. Ya hemos visto anteriormente, que nuestra prueba de hipótesis para la distribución de los residuos es:

$$H_0 : \text{Normalidad en residuos} \quad vs \quad H_1 : \text{No normalidad en residuos} \quad (1.10.1)$$

y al tomar un nivel de significancia de $\alpha = 5\%$, entonces, no se rechaza H_0 , por lo que los residuos tienen una distribución normal.

Interesante será observar la **Figura 1.10.4**, donde se muestra la relación existente entre residuos (estandarizados y estudentizados) y sus correspondientes valores ajustados, ya que presentan un patrón usual de rectángulo inclinado, por lo que afirmamos la existencia de heteroscedasticidad. Para corroborar este supuesto tomamos la **prueba de Breusch-Pagan** y observamos que tiene un p-value cercano a cero, por lo que rechazamos la hipótesis nula y concluimos que la varianza en los residuos no es constante:

```
> bptest(Financ~Periodo,data=Financiamiento)
```

```
studentized Breusch-Pagan test
```

```
data: Financ ~ Periodo
BP = 15.6056, df = 1, p-value = 7.802e-05
```

Finalmente, pasemos a la prueba que nos concierne en esta sección, es decir, la prueba de **Durbin-Watson**, con estadístico $d = 0.1788$, $d_L = 1.50$ y $d_U = 1.59$, para 54 observaciones, con $\alpha = 5\%$ y una sola variable explicativa. Observamos que $d = 0.1788 < d_L = 1.50$, así que rechazamos $H_0 : \rho = 0$ vs $H_1 : \rho > 0$, por lo que existe correlación positiva entre los residuos y no se cumple el supuesto de que los residuos, ϵ_t , son independientes. Por lo anterior se usarán otras técnicas para ajustar modelos que involucren al tiempo.

Capítulo 2

Análisis Clásico de Series de Tiempo

2.1. Introducción

A lo largo del capítulo anterior hemos estudiado el uso del análisis de regresión como técnica habitual para explicar y modelar la relación que existe entre una o más variables. Al final de dicho capítulo hemos remarcado la existencia de datos con los cuales, la modelación de los mismos con una regresión se vuelve inapropiada, debido a la fuerte autocorrelación que existe entre los residuos. A partir de aquí, y siendo nuestra principal motivación, se introduce el uso de **series de tiempo** como método alternativo para la elaboración de predicciones que anticipen el comportamiento de algún fenómeno en particular.

Como actuario, palabras como riesgo o incertidumbre, lejos de ser desconocidas, aparecen de manera recurrente en nuestra vida cotidiana, siendo una obligación la implementación de técnicas que ayuden a reducir el grado de incertidumbre ante cierto fenómeno. El disponer, el día de hoy, de información acerca de cómo se desarrollará el futuro, aunque sea de manera aproximada, nos facilita la toma de decisiones en las que incurrimos en el presente.

Pensemos por un momento en un empresario que desea crear un plan estratégico de expansión para su empresa. Para lograrlo, necesita conocer a cuánto ascenderán sus ventas futuras. Su decisión no puede basarse solamente en los datos actuales disponibles sobre sus ventas, sino también requerirá apoyarse en sus valores históricos así como en otras técnicas de predicción. Veremos más adelante que los métodos de predicción pueden dividirse en:

1. Métodos de información subjetiva: opiniones, encuestas, etc.
2. Métodos de series de tiempo: promedios móviles, suavizamiento exponencial, método de Holt, método de Holt-Winters, metodología ARIMA, etc.
3. Métodos causales: modelos econométricos, indicadores, etc.

Es en este marco donde encuadraremos el estudio de series temporales como una de las técnicas más habituales que se utilizan en la predicción de fenómenos de cualquier naturaleza. Una serie de tiempo, es decir, un conjunto de datos pertenecientes a una variable tomados en periodos regulares de tiempo, será nuestro punto de partida para el estudio de cualquier fenómeno. Atención especial deberemos poner al realizar el análisis de los datos ya que cualquier equivocación nos llevará a una falsa interpretación de los mismos o bien, a decisiones equivocadas. Lo que buscamos es reconocer las propiedades pasadas del fenómeno y suponiendo que éstas se mantienen constantes con el paso del tiempo, hacer predicciones con respecto al futuro.

Comencemos dando una definición clásica de serie de tiempo. Dicha definición cambiará en el siguiente capítulo cuando manejemos a las series de tiempo como un proceso estocástico.

Definición 2.1.1 *Una serie de tiempo es una sucesión de observaciones correspondientes a una variable tomadas en intervalos regulares y de duración constante.*

Así pues, el número de pólizas anuales vendidas por una empresa aseguradora, las ventas semestrales de una empresa distribuidora de ropa deportiva, el número de casos mensuales de personas jubiladas en cierta empresa, el total de accidentes diarios de tránsito en el D.F., e inclusive la población que tiene año con año cada país constituyen, todos ellos, ejemplos de series de tiempo.

Antes de continuar hagamos una diferencia importante en lo que se refiere a los datos. Éstos pueden referirse a un momento en particular o bien a un intervalo de tiempo. Pensemos por un momento en el último censo del año 2010 en toda la **Republica Mexicana**. Al 31 de diciembre de 2010 se nos informó que la población total en todo México era de 112,336,538 habitantes. Este ejemplo hace referencia a **datos de tipo transversales** (*datos de tipo stock*), es decir, datos tomados en un instante particular de tiempo. Otro ejemplo puede ser el reporte final de ingresos y egresos de una compañía aseguradora al cierre del año.

Pasemos ahora a los datos recogidos en un intervalo de tiempo o **datos de tipo longitudinales** (*datos de tipo flujo*). Pensemos por un momento en la producción mensual de playeras de futbol del equipo **Real Madrid**[®] por parte de la marca deportiva **Adidas**[®]. La leyenda que identificaría a este tipo de dato sería la siguiente: “A lo largo del mes de Enero se confeccionaron y pusieron a la venta **X** cantidad de playeras”. Otro ejemplo sería el número de tarjetas de crédito emitidas por cierta Institución de Banca Multiple durante el mes de Noviembre.

Existen varias metodologías para tratar y analizar una serie de tiempo. Principalmente trabajaremos nosotros, con los enfoques deterministas y estocásticos. El **enfoque determinista** supone un patrón de comportamiento fijo de nuestra variable a estudiar, es decir, los datos oscilarán entorno a un comportamiento predecible. El **enfoque estocástico** surge con los trabajos de **Box & Jenkins** y la **metodología ARIMA**, basada en la teoría de los **procesos estocásticos**. Dicho enfoque asume una conducta completamente impredecible de nuestra variable a estudiar. En este capítulo nos enfocaremos de lleno al enfoque determinista o de **descomposición clásica** y en el siguiente capítulo discutiremos el enfoque estocástico.

Ahora bien, definamos el objetivo de este capítulo. ¿Qué es lo que buscamos al analizar una serie de tiempo? En resumen, al analizar una serie de tiempo buscamos reconocer y extraer todas aquellas características o regularidades que conforman al comportamiento pasado de nuestra variable a estudiar, obtener el procedimiento que las genera y haciendo el gran supuesto de que dichas características se mantienen constantes a lo largo del tiempo, predecir el comportamiento futuro de la variable.

2.1.1. Notas previas y etapas a cubrir al analizar series de tiempo

1. Como primer paso, y a nuestro parecer uno de los más importantes, debemos tomar en consideración el hecho de que los datos son la materia prima con la que iniciamos cualquier estudio. De aquí que debe ser de nuestro completo interés el preguntarnos de dónde provienen los datos, la metodología que se utilizó para obtenerlos, las modificaciones que han sufrido los mismos, la evolución que han tenido, etc, con la única finalidad de trabajar con datos homogéneos.

2. Tener presente que a pesar de que la variable tiempo se introduce de manera continua, en la vida real no se maneja así, sino de manera discreta. Este supuesto será primordial al momento de hacer conclusiones o aseveraciones respecto a la variable en cuestión.
3. Tener presente el concepto de **poder adquisitivo** al momento de trabajar con series de tiempo expresadas en términos monetarios. El poder adquisitivo se define como el valor de una unidad monetaria específica en términos de la cantidad de bienes y servicios que se pueden adquirir con ella.

Como bien sabemos, el valor del dinero cambia conforme pasa el tiempo por el **efecto inflacionario**. \$ 100 el día de hoy no valen lo mismo que hace 20 años, en el sentido de que no podemos obtener el mismo número de bienes con dicha cantidad en el presente como tal vez pudimos haberlo hecho hace 20 años.

Es aquí donde introducimos los conceptos de **precios corrientes** y **precios constantes**. Por ejemplo, cuando hablamos del **PIB real**, hablamos de las variaciones que experimenta la producción física de cierta economía entre periodos diferentes valorando todos los bienes producidos en los dos periodos a los mismos precios, es decir a precios constantes. Por otro lado, el **PIB nominal** mide el valor de la producción en un determinado periodo a los precios de ese periodo, es decir, a precios corrientes. Así pues, series como el nivel de salarios de todos los trabajadores de la **UNAM** deberán tomarse a precios constantes para volverlas así, comparables. Este procedimiento es conocido como **deflactación** y consiste en dividir la serie por un deflactor, logrando así eliminar la variación que han sufrido los precios con respecto al año de referencia. En México se utiliza el **deflactor del PIB** así como el índice nacional de precios al consumidor (INPC) para lograr este cometido.

4. Ahora bien, tal vez en algún momento nos preguntaremos qué es lo que sucede con la diferencia en el número de días que tienen los distintos meses al trabajar con una serie mensual. ¿Debemos hacer caso omiso a este suceso o tomar alguna medida al respecto? Debemos hacer algo al respecto. Nuestra tarea es y será siempre volver homogénea nuestra materia prima para así, al final, obtener resultados satisfactorios. Ahora bien, la manera de volver comparables nuestros datos, será escoger un mes referencia, observar el número de días que tiene (30 días se recomienda), para después expresar los datos de cada mes en términos diarios y después multiplicarlos por el número de días que tenga el mes de referencia. Este punto es mera sugerencia ya que en la práctica se ignora la diferencia de días por mes.
5. Después de haber trabajado exhaustivamente con los datos, lo más recomendable será realizar un gráfico del tiempo contra los valores de la serie para observar su comportamiento, características, evolución, métodos que pueden llegar a aplicarse, etc. En este apartado, al observar el gráfico podemos detectar lo que más adelante definiremos como tendencia, estacionalidad y componente irregular.
6. Nuestra siguiente tarea será encontrar un modelo que reproduzca las características más importantes de la serie. Lo que deseamos es encontrar el modelo que mejor se ajuste a nuestros datos. Al emplear la palabra “mejor” nos estamos refiriendo, no a encontrar un modelo sofisticado o que incluya una gran cantidad de variables para explicar el comportamiento de la serie, sino más bien, a encontrar un modelo sencillo que explique de manera eficiente la realidad y que pueda ser entendido en su totalidad por el analista.
7. Como último paso nos resta la importante tarea de validar nuestro modelo, es decir, nuestro deseo será saber si el modelo que hemos elegido es suficiente para describir el pasado de nuestros datos así como efectivo para realizar predicciones.

2.2. Componentes de una serie de tiempo

Observemos el siguiente gráfico donde hemos representado el número de acciones de atención realizadas por **Condusef**, desde el año 2007 hasta el año 2012. Por acción de atención entenderemos toda aquella labor de asesoría, protección y defensa de los intereses de los usuarios de servicios financieros, buscando en todo momento la sana interacción entre los agentes del Sistema Financiero Mexicano.

```
> library(stats)
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Condusef")
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> plot(c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="blue")
> legend(2007,100000,c("AccAtn"),lty=c(1),bty="n",lwd=c(2.5),col=c("blue"))
```

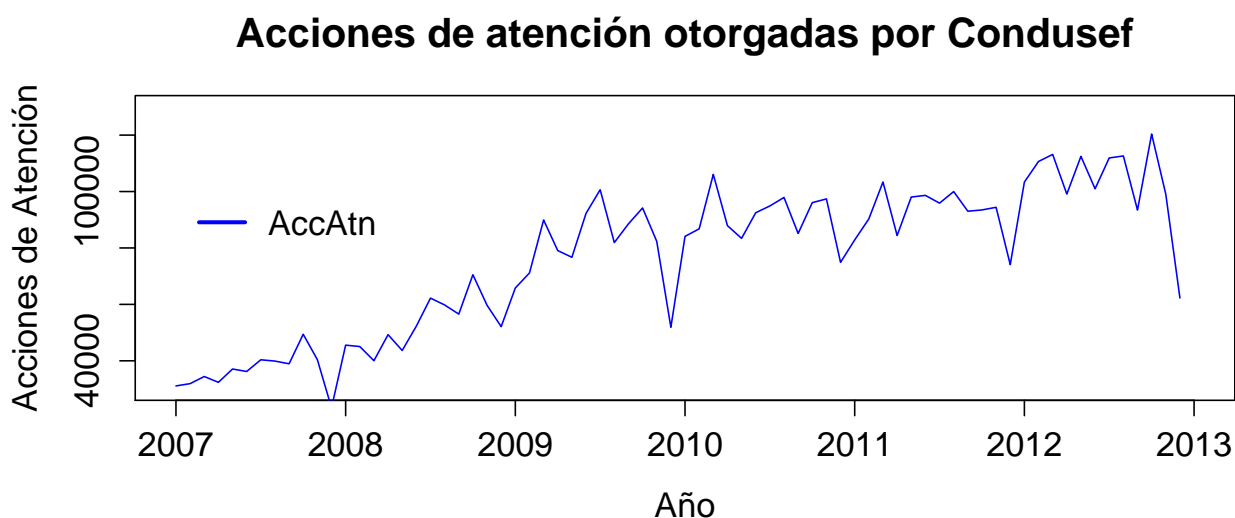


Figura 2.2.1: Figura que muestra el número de acciones de atención otorgadas por Condusef

Ahora bien, de la **Figura 2.2.1** observamos lo siguiente:

1. Un movimiento de continuo crecimiento en el número de acciones de atención otorgadas por Condusef, explicado principalmente por la fuerte campaña de innovación por parte de la Comisión por defender los intereses de los usuarios de servicios financieros. (tendencia)
2. Un comportamiento ascendente de la serie en el primer trimestre de cada año, coincidiendo con la época de mayores quejas y asesoramientos (después de época navideña). (estacionalidad)
3. Variaciones de crecimiento y decrecimiento rápido quizás debidas a una mejor difusión de información por parte de Condusef referente al uso y manejo de los instrumentos financieros que existen en el Sistema Financiero Mexicano. (componente irregular)

Concluimos que el enfoque clásico presume dividida a una serie de tiempo en 3 componentes fundamentales, a saber, tendencia, estacionalidad y componente irregular.

Así pues, podemos pensar a una serie de tiempo como una función que dependa de las componentes anteriores, es decir:

$$Y_t = f(T_t, E_t, \epsilon_t) \quad (2.2.1)$$

Grosso modo, el objetivo del análisis clásico será el de estimar por separado las componentes de una serie, para después englobarlas en uno de los 3 esquemas siguientes:

1. **Esquema aditivo:** Este esquema se usa cuando la variabilidad de la serie se mantiene constante a través del tiempo:

$$Y_t = T_t + E_t + \epsilon_t \quad (2.2.2)$$

2. **Esquema multiplicativo:** Este esquema se utiliza cuando la variabilidad en los datos es cada vez mayor o menor, es decir, no se mantiene constante con el tiempo:

$$Y_t = T_t * E_t * \epsilon_t \quad (2.2.3)$$

3. **Esquema mixto:** Este esquema se utiliza cuando las oscilaciones estacionales tienden a crecer con el tiempo, y la variabilidad de estas oscilaciones se mantiene constante. Este esquema sólo será enunciado más no utilizado debido a la complejidad que surge al analizar la serie:

$$Y_t = (T_t * E_t) + \epsilon_t \quad (2.2.4)$$

¿Pero cómo elegir el mejor esquema? Existen criterios para detectar el esquema que mejor se ajusta a la serie. A continuación veremos tres de ellos.

2.2.1. Representación gráfica

Este criterio consiste en graficar nuestra serie y con base en ella observar la variabilidad de los datos. Para elegir el esquema aditivo, las oscilaciones o variaciones de nuestros datos deben tener la misma amplitud entorno a la tendencia de la serie. En caso contrario, es decir, variaciones cada vez mayores o menores entorno a la tendencia, se utilizará el esquema multiplicativo. En la **Figura 2.2.2** se observa el número de acciones de atención otorgadas por Condusef, la cual nos servirá para elegir el esquema adecuado.

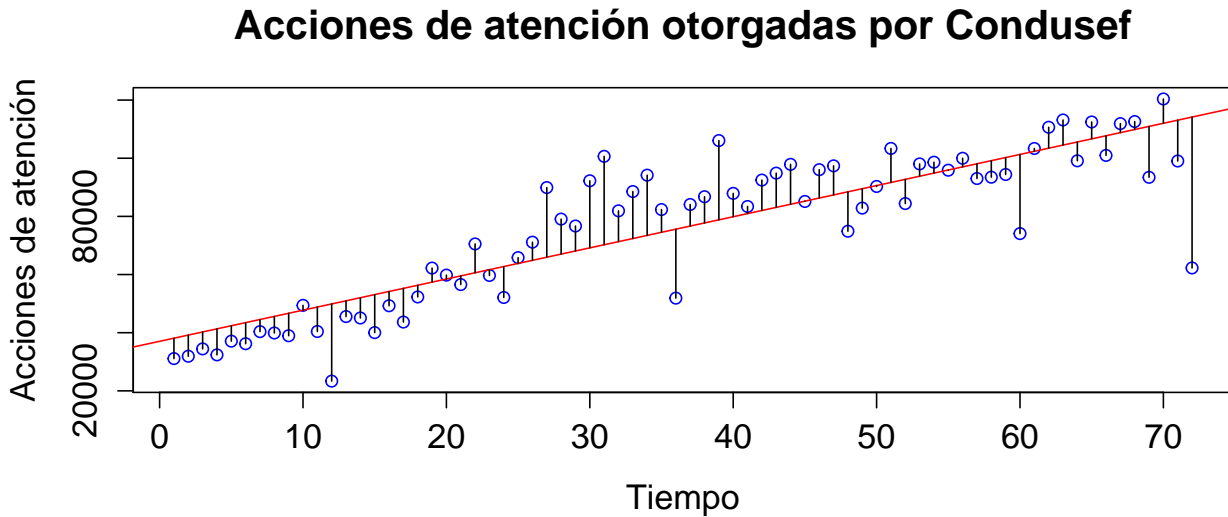


Figura 2.2.2: Representación gráfica del número de acciones de atención

A pesar de que las variaciones no crecen de manera sustancial a través del tiempo, pensamos que estas variaciones son suficientes para justificar el uso de un esquema multiplicativo. Dicha aseveración se comprobará con los siguientes criterios.

2.2.2. Diferencias y cocientes estacionales

Se denota $d_{t,i}$, diferencia estacional, a la resta entre dos datos de años consecutivos pero de la misma estación. Se denota $c_{t,i}$, cociente estacional, a la división entre dos datos de años consecutivos pero de la misma estación. Sea $Y_{t,i}$ el valor de la serie de tiempo en la estación i del periodo (o año) t , entonces:

$$d_{t,i} = Y_{t,i} - Y_{t-1,i} \tag{2.2.5}$$

$$c_{t,i} = \frac{Y_{t,i}}{Y_{t-1,i}}$$

A continuación enumeramos los pasos a seguir:

1. Calcular todas las diferencias y cocientes estacionales.
2. Obtener el coeficiente de variación (CV) para diferencias (d) y cocientes (c) estacionales:

$$CV(d) = \frac{\text{Desviación estándar (d)}}{\text{media (d)}} \tag{2.2.6}$$

$$CV(c) = \frac{\text{Desviación estándar (c)}}{\text{media (c)}} \tag{2.2.7}$$

3. Si $CV(c) > CV(d)$ elegiríamos el esquema aditivo. Si $CV(c) < CV(d)$ elegiríamos el esquema multiplicativo.

La obtención de diferencias estacionales equivale a tomar la serie de incrementos interanuales, mientras que la obtención de cocientes estacionales equivale a tomar la serie de tasas de crecimiento. Por tanto, si la tasa de crecimiento interanual tiene menor variabilidad que la serie de incrementos interanuales optaríamos por el modelo multiplicativo. En caso contrario, el aditivo. Tomemos nuevamente el número de acciones de atención y realicemos este criterio:

$$CV(d) = \frac{\text{Desviación estándar (d)}}{\text{media (d)}} = 0.9385 \quad (2.2.8)$$

$$CV(c) = \frac{\text{Desviación estándar (c)}}{\text{media (c)}} = 0.2303 \quad (2.2.9)$$

Como $CV(c) = 0.2303 < CV(d) = 0.9385$, el esquema óptimo es el multiplicativo. El número de acciones de atención y los cálculos parciales para la obtención de estos coeficientes se presentan en la **Tabla 2.2.1**.

t	AccAtn	Y_t	d_{ti}	c_{ti}	t	AccAtn	Y_t	d_{ti}	c_{ti}
1	31101	-	-	-	37	84094	18280	1.2778	
2	31909	-	-	-	38	86772	15635	1.2198	
3	34435	-	-	-	39	106082	16169	1.1798	
4	32380	-	-	-	40	87916	8843	1.1118	
5	37096	-	-	-	41	83399	6704	1.0874	
6	36216	-	-	-	42	92487	239	1.0026	
7	40367	-	-	-	43	94883	-5768	0.9427	
8	39902	-	-	-	44	97899	15989	1.1952	
9	38954	-	-	-	45	85130	-3415	0.9614	
10	49412	-	-	-	46	96060	1879	1.0200	
11	40412	-	-	-	47	97379	15029	1.1825	
12	23355	-	-	-	48	74898	23004	1.4433	
13	45553	14452	1.4647	49	82854	-1240	0.9853		
14	45057	13148	1.4120	50	90295	3523	1.0406		
15	40029	5594	1.1625	51	103375	-2707	0.9745		
16	49268	16888	1.5216	52	84414	-3502	0.9602		
17	43667	6571	1.1771	53	98046	14647	1.1756		
18	52294	16078	1.4439	54	98633	6146	1.0665		
19	62229	21862	1.5416	55	95882	999	1.0105		
20	59801	19899	1.4987	56	99991	2092	1.0214		
21	56553	17599	1.4518	57	93020	7890	1.0927		
22	70527	21115	1.4273	58	93480	-2580	0.9731		
23	59687	19275	1.4770	59	94409	-2970	0.9695		
24	52113	28758	2.2313	60	74088	-810	0.9892		
25	65814	20261	1.4448	61	103356	20502	1.2474		
26	71137	26080	1.5788	62	110649	20354	1.2254		
27	89913	49884	2.2462	63	113157	9782	1.0946		
28	79073	29805	1.6050	64	99107	14693	1.1741		
29	76695	33028	1.7564	65	112468	14422	1.1471		
30	92248	39954	1.7640	66	100971	2338	1.0237		
31	100651	38422	1.6174	67	111903	16021	1.1671		
32	81910	22109	1.3697	68	112614	12623	1.1262		
33	88545	31992	1.5657	69	93443	423	1.0045		
34	94181	23654	1.3354	70	120379	26899	1.2878		
35	82350	22663	1.3797	71	99046	4637	1.0491		
36	51894	-219	0.9958	72	62274	-11814	0.8405		

Tabla 2.2.1: Tabla que muestra el cálculo de diferencias y cocientes estacionales

2.2.3. Gráfico media-desviación estándar

Para este criterio usaremos primordialmente datos con periodicidad menor al año, por ejemplo, periodicidad diaria, mensual, semestral, trimestral, etc. Este criterio consiste en obtener la media y la desviación estándar de nuestros datos de manera anual, representarlos mediante un diagrama de dispersión, y con base en ello analizar el conjunto de puntos que se forman. Si el aspecto del conjunto de puntos corresponde con una línea más o menos creciente, elegiríamos el modelo multiplicativo. Si por el contrario, el conjunto de puntos se distribuye de manera aleatoria entonces elegiríamos el modelo aditivo.

Justificación: La media de los datos anuales corresponde al nivel de tendencia de cada año. La desviación estándar nos cuantifica la variabilidad que un conjunto de datos tiene entorno a su media. Dentro del año es perfectamente lícito pensar que la tendencia no cambiará de manera considerable (esta componente se manifiesta en el largo plazo) y por tanto, la desviación estándar no se verá muy influida por ella. Los cambios de valor de la serie estarán más bien causados por la estacionalidad. Por tanto, unos puntos de media y desviación estándar más o menos alineados en una recta (con pendiente positiva, o bien, negativa) significará que a medida que el nivel de tendencia crece (o bien decrece) también lo hará la dispersión debida a la estacionalidad y de aquí que se elija el esquema multiplicativo. Si el conjunto de puntos se distribuye de manera aleatoria entonces suponemos independencia entre la tendencia y la estacionalidad y por tanto, un esquema aditivo sería apropiado.

La **Tabla 2.2.2** muestra el número de acciones de atención así como el cálculo de la media y desviación estándar anual de los datos. Observemos la **Figura 2.2.3**. El conjunto de puntos corresponde con una línea creciente, de aquí que el esquema óptimo es el multiplicativo.

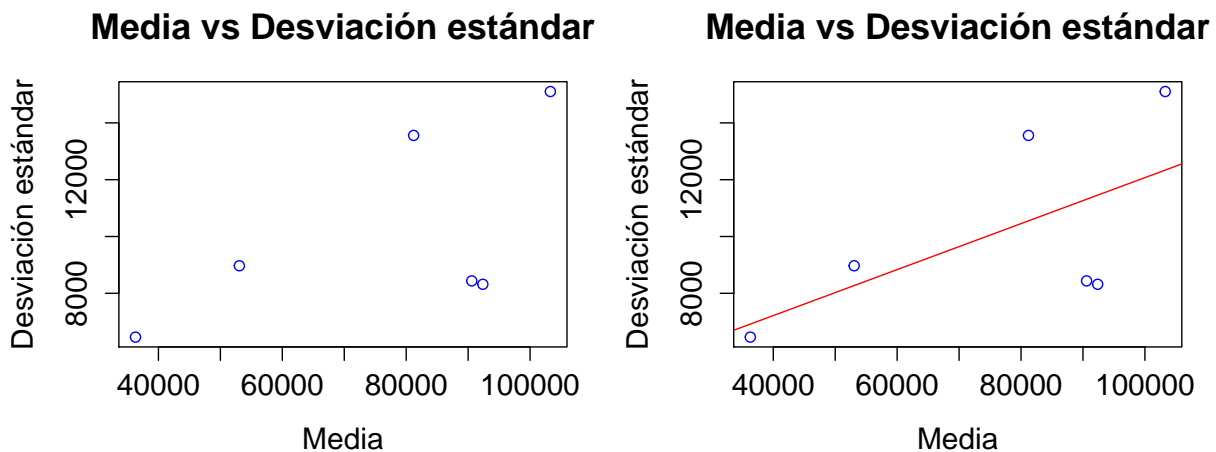


Figura 2.2.3: Figura que muestra la relación media-desviación estándar

Año	Y_{t_m}	Sd_t
2007	36295	6457.98
2008	53065	8969.78
2009	81201	13558.80
2010	90583	8437.74
2011	92374	8318.47
2012	103281	15102.89

Tabla 2.2.2: Tabla que muestra el cálculo de media-desv. estándar

2.3. Análisis de Tendencia

Definición 2.3.1 Llamaremos **tendencia** al movimiento suave y regular de la serie a largo plazo.

Debemos hacer hincapié que para observar un movimiento de continuo crecimiento o decrecimiento en nuestra serie es necesario disponer de información que abarque al menos 3-4 años. La tendencia será así la componente principal de una serie de tiempo. Nos indica la dirección en la cual se mueve dicha serie. Existen dos objetivos principales al trabajar con la tendencia; su estimación y su suavizamiento. Para estimarla se ajusta una función matemática que dependa del tiempo. Para suavizarla se emplea el método de promedios móviles, entre otros.

Ahora bien, para estimar la tendencia uno de los métodos que veremos en este capítulo será el **método del ajuste analítico**, el cual consiste en encontrar aquella función que mejor se ajuste al modelar la tendencia. Usualmente una función lineal bastará para ajustar nuestra tendencia aunque debe decirse que existen series, económicas sobre todo, cuya tendencia se ajusta mejor a una función exponencial o de grado mayor a 1.

La función que mejor ajuste la tendencia será aquella que cumpla la **ec. 2.3.1**, tomando en consideración que antes de aplicar este método deberemos eliminar la estacionalidad, ya que de lo contrario, las oscilaciones influirían en la identificación de la función matemática a ajustar.

$$\text{Ajuste de una función: } Y_t = f(t) + \epsilon_t \quad (2.3.1)$$

Entonces, para seleccionar la función matemática que mejor se ajusta a los datos, es menester realizar antes que nada el gráfico de la serie para después pasar a la estimación de los parámetros, que en concreto, en este trabajo, se utilizarán los estimadores por mínimos cuadrados ordinarios, que de acuerdo con el Teorema de Gauss-Markov visto en el capítulo anterior, son los mejores estimadores lineales insesgados.

El método de mínimos cuadrados ordinarios es empleado en el análisis de regresión lineal, he ahí una de las razones del porqué de nuestro capítulo anterior. Sabemos que el análisis de regresión lineal es usado para explicar la relación que existe entre variables explicativas y la variable respuesta. Sin embargo, al tratar una serie de tiempo, buscamos, no analizar la capacidad explicativa del tiempo sino más bien encontrar regularidades en el mismo.

A continuación, utilizaremos el esquema de regresión lineal simple, por lo cual el planteamiento que deseamos hacer es el siguiente: Sea $Y_t = T_t + \epsilon_t$ el modelo de partida, donde T_t es la tendencia y ϵ_t es la parte aleatoria del modelo. Suponemos ϵ_t variable aleatoria con media cero, varianza constante e igual a σ_ϵ^2 , no autocorrelacionada y con **distribución Normal**, es decir, $\epsilon_t \sim NID(0, \sigma_\epsilon^2)$.

2.3.1. Modelo de tendencia lineal

Suponemos $Y_t = T_t + \epsilon_t = a + \beta t + \epsilon_t$, donde el parámetro a representa el nivel de tendencia al comienzo del periodo de análisis, $t = 0$, y el coeficiente β representa el incremento o decremento de tendencia promedio cuando pasamos de un periodo a otro.

Imaginemos que tenemos la serie correspondiente al PIB trimestral (millones de dólares) de E.U.A. a partir del año 1947 hasta el año 1991. Los datos, el cálculo de la tendencia estimada denotada como T_t y los residuos correspondientes ϵ_t se muestran en la **Tabla 2.3.1**. A continuación haremos el ajuste de la tendencia con una función lineal.

El código en R será:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Ejercicios")
> Usgnp<-read.table("Usgnp.txt",header=T)
> attach(Usgnp)
> Ajuste<-lm(PIBtrim~t)
> summary(Ajuste)
```

Call:

```
lm(formula = PIBtrim ~ t)
```

Residuals:

Min	1Q	Median	3Q	Max
-229.26	-100.14	-13.33	80.75	264.16

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	837.1148	18.2262	45.93	<2e-16 ***
t	17.7184	0.1776	99.77	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 120.7 on 175 degrees of freedom

Multiple R-squared: 0.9827, Adjusted R-squared: 0.9826

F-statistic: 9953 on 1 and 175 DF, p-value: < 2.2e-16

```
> library("graphics")
> par(mfrow=c(1,2))
> plot(t,PIBtrim,pch=21,xlim=c(0,180),main="PIB trimestral de Estados Unidos",
  col="blue")
> plot(t,PIBtrim,pch=21,xlim=c(0,180),main="PIB trimestral de Estados Unidos",
  col="blue")
> abline(Ajuste,col="red")
```

En la **Figura 2.3.1** se muestra la serie original junto con un ajuste de tendencia lineal en color rojo.

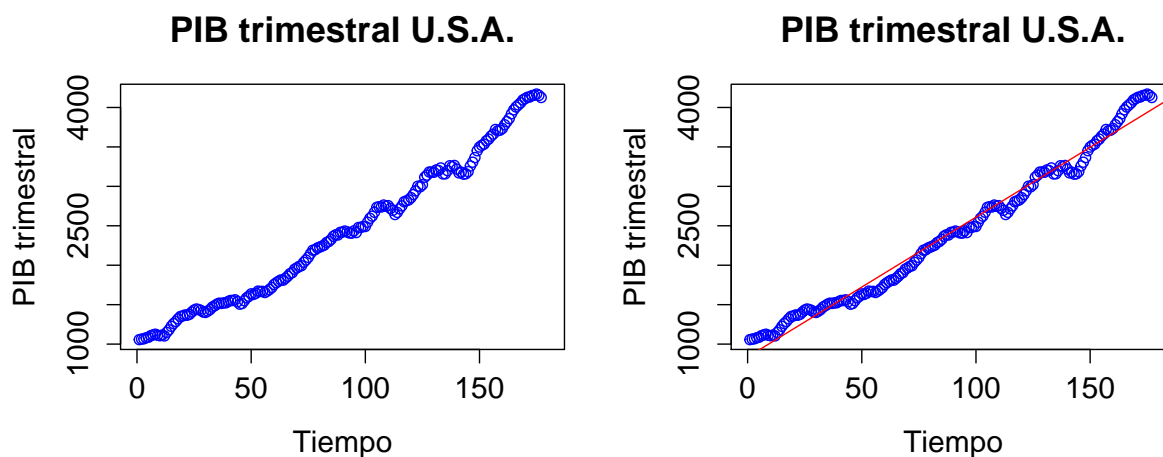


Figura 2.3.1: Figura que muestra el PIB trimestral de Estados Unidos

t	Y_t	T_t	ϵ_t	t	Y_t	T_t	ϵ_t	t	Y_t	T_t	ϵ_t
1	1056.5	854.8	201.7	60	1754.9	1900.2	-145.3	119	2828.6	2945.6	-117.0
2	1063.2	872.6	190.7	61	1777.9	1917.9	-140.0	120	2856.8	2963.3	-106.5
3	1067.1	890.3	176.8	62	1796.4	1935.7	-139.3	121	2896.0	2981.0	-85.0
4	1080.0	908.0	172.0	63	1813.1	1953.4	-140.3	122	2942.7	2998.8	-56.1
5	1086.8	925.7	161.1	64	1810.1	1971.1	-161.0	123	3001.8	3016.5	-14.7
6	1106.1	943.4	162.7	65	1834.6	1988.8	-154.2	124	2994.1	3034.2	-40.1
7	1116.3	961.1	155.2	66	1860.0	2006.5	-146.5	125	3020.5	3051.9	-31.4
8	1125.5	978.9	146.6	67	1892.5	2024.2	-131.8	126	3115.9	3069.6	46.3
9	1112.4	996.6	115.8	68	1906.1	2042.0	-135.9	127	3142.6	3087.4	55.3
10	1105.9	1014.3	91.6	69	1948.7	2059.7	-111.0	128	3181.6	3105.1	76.5
11	1114.3	1032.0	82.3	70	1965.4	2077.4	-112.0	129	3181.7	3122.8	58.9
12	1103.3	1049.7	53.6	71	1985.2	2095.1	-109.9	130	3178.7	3140.5	38.2
13	1148.2	1067.5	80.7	72	1993.7	2112.8	-119.1	131	3207.4	3158.2	49.2
14	1181.0	1085.2	95.8	73	2036.9	2130.6	-93.7	132	3201.3	3175.9	25.3
15	1225.3	1102.9	122.4	74	2066.4	2148.3	-81.9	133	3233.4	3193.7	39.7
16	1260.2	1120.6	139.6	75	2099.3	2166.0	-66.7	134	3157.0	3211.4	-54.4
17	1286.6	1138.3	148.3	76	2147.6	2183.7	-36.1	135	3159.1	3229.1	-70.0
18	1320.4	1156.0	164.4	77	2190.1	2201.4	-11.3	136	3199.2	3246.8	-47.6
19	1349.8	1173.8	176.0	78	2195.8	2219.2	-23.4	137	3261.1	3264.5	-3.4
20	1356.0	1191.5	164.5	79	2218.3	2236.9	-18.6	138	3250.2	3282.3	-32.1
21	1369.2	1209.2	160.0	80	2229.2	2254.6	-25.4	139	3264.6	3300.0	-35.4
22	1365.9	1226.9	139.0	81	2241.8	2272.3	-30.5	140	3219.0	3317.7	-98.7
23	1378.2	1244.6	133.6	82	2255.2	2290.0	-34.8	141	3170.4	3335.4	-165.0
24	1406.8	1262.4	144.4	83	2287.7	2307.7	-20.0	142	3179.9	3353.1	-173.2
25	1431.4	1280.1	151.3	84	2300.6	2325.5	-24.9	143	3154.5	3370.9	-216.3
26	1444.9	1297.8	147.1	85	2327.3	2343.2	-15.9	144	3159.3	3388.6	-229.3
27	1438.2	1315.5	122.7	86	2366.9	2360.9	6.0	145	3186.6	3406.3	-219.7
28	1426.6	1333.2	93.4	87	2385.3	2378.6	6.7	146	3258.3	3424.0	-165.7
29	1406.8	1350.9	55.8	88	2383.0	2396.3	-13.3	147	3306.4	3441.7	-135.3
30	1401.2	1368.7	32.5	89	2416.5	2414.1	2.5	148	3365.1	3459.4	-94.4
31	1418.0	1386.4	31.6	90	2419.8	2431.8	-12.0	149	3451.7	3477.2	-25.5
32	1438.8	1404.1	34.7	91	2433.2	2449.5	-16.3	150	3498.0	3494.9	3.1
33	1469.6	1421.8	47.8	92	2423.5	2467.2	-43.7	151	3520.6	3512.6	8.0
34	1485.7	1439.5	46.2	93	2408.6	2484.9	-76.3	152	3535.2	3530.3	4.9
35	1505.5	1457.3	48.2	94	2406.5	2502.6	-96.1	153	3577.5	3548.0	29.5
36	1518.7	1475.0	43.7	95	2435.8	2520.4	-84.6	154	3599.2	3565.8	33.5
37	1515.7	1492.7	23.0	96	2413.8	2538.1	-124.3	155	3635.8	3583.5	52.3
38	1522.6	1510.4	12.2	97	2478.6	2555.8	-77.2	156	3662.4	3601.2	61.2
39	1523.7	1528.1	-4.4	98	2478.4	2573.5	-95.1	157	3721.1	3618.9	102.2
40	1540.6	1545.9	-5.2	99	2491.1	2591.2	-100.1	158	3704.6	3636.6	68.0
41	1553.3	1563.6	-10.3	100	2491.0	2609.0	-118.0	159	3712.4	3654.3	58.0
42	1552.4	1581.3	-28.9	101	2545.6	2626.7	-81.1	160	3733.6	3672.1	61.5
43	1561.5	1599.0	-37.5	102	2595.1	2644.4	-49.3	161	3781.2	3689.8	91.4
44	1537.3	1616.7	-79.4	103	2622.1	2662.1	-40.0	162	3820.3	3707.5	112.8
45	1506.1	1634.4	-128.3	104	2671.3	2679.8	-8.5	163	3858.9	3725.2	133.7
46	1514.2	1652.2	-138.0	105	2734.0	2697.5	36.4	164	3920.7	3742.9	177.8
47	1550.0	1669.9	-119.9	106	2741.0	2715.3	25.7	165	3970.2	3760.7	209.5
48	1586.7	1687.6	-100.9	107	2738.3	2733.0	5.3	166	4005.8	3778.4	227.4
49	1606.4	1705.3	-98.9	108	2762.8	2750.7	12.1	167	4032.1	3796.1	236.0
50	1637.0	1723.0	-86.0	109	2747.4	2768.4	-21.0	168	4059.3	3813.8	245.5
51	1629.5	1740.8	-111.3	110	2755.2	2786.1	-30.9	169	4095.7	3831.5	264.2
52	1643.4	1758.5	-115.1	111	2719.3	2803.9	-84.6	170	4112.2	3849.2	262.9
53	1671.6	1776.2	-104.6	112	2695.4	2821.6	-126.2	171	4129.7	3867.0	262.7
54	1666.8	1793.9	-127.1	113	2642.7	2839.3	-196.6	172	4133.2	3884.7	248.5
55	1668.4	1811.6	-143.2	114	2669.6	2857.0	-187.4	173	4150.6	3902.4	248.2
56	1654.1	1829.3	-175.3	115	2714.9	2874.7	-159.8	174	4155.1	3920.1	235.0
57	1671.3	1847.1	-175.8	116	2752.7	2892.5	-139.7	175	4170.0	3937.8	232.2
58	1692.1	1864.8	-172.7	117	2804.4	2910.2	-105.8	176	4153.4	3955.6	197.8
59	1716.3	1882.5	-166.2	118	2816.9	2927.9	-111.0	177	4126.5	3973.3	153.2

Tabla 2.3.1: Tabla que muestra la estimación de tendencia para el PIB de Estados Unidos

Ahora veamos el comportamiento del PIB trimestral en México. Los valores de la serie están dados a precios constantes, con base al año 2008. Realizaremos un análisis similar al hecho con el PIB trimestral de Estados Unidos. Las cifras datan de Enero de 1993 al tercer trimestre del año 2013. Los datos, la tendencia estimada, T_t , y los residuos, ϵ_t , se muestran en la **Tabla 2.3.2**. Además, en la **Figura 2.3.2** se muestra la serie original correspondiente al PIB trimestral de México así como el ajuste de tendencia lineal correspondiente:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Ejercicios")
> PIBmex<-read.table("PIB a precios constantes.txt",header=T)
> attach(PIBmex)
> Ajuste<-lm(PIBmexicano~t)
> summary(Ajuste)
```

Call:

```
lm(formula = PIBmexicano ~ t)
```

Residuals:

```
      Min       1Q   Median       3Q      Max
-850816 -239881 -24363  232627 1094184
```

Coefficients:

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  452336      87044    5.197 1.49e-06 ***
t             184670       1800 102.584 < 2e-16 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 392900 on 81 degrees of freedom

Multiple R-squared: 0.9924, Adjusted R-squared: 0.9923

F-statistic: 1.052e+04 on 1 and 81 DF, p-value: < 2.2e-16

```
> plot(t,PIBmexicano,pch=21,xlim=c(0,85),main="PIB trimestral de México",
col="blue")
> plot(t,PIBmexicano,pch=21,xlim=c(0,85),main="PIB trimestral de México",
col="blue")
> abline(Ajuste,col="red")
```

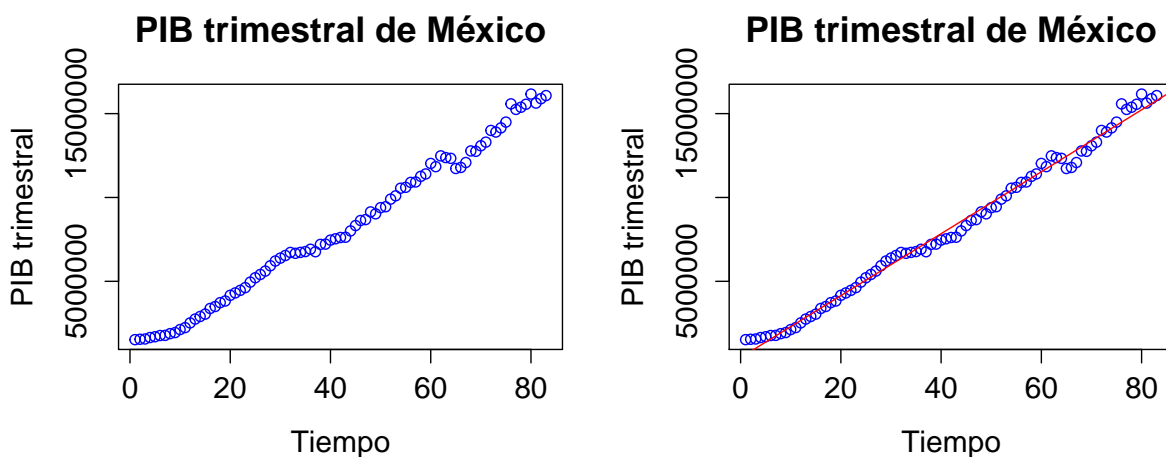


Figura 2.3.2: Figura que muestra el PIB trimestral de México

t	Y_t	T_t	ϵ_t	t	Y_t	T_t	ϵ_t
1	1522840.653	637006	885834.653	43	7628204.596	8393146	-764941.404
2	1546496.086	821676	724820.086	44	8004873.173	8577816	-572942.827
3	1562258.138	1006346	555912.138	45	8325115.117	8762486	-437370.883
4	1648991.525	1191016	457975.525	46	8628189.597	8947156	-318966.403
5	1695821.696	1375686	320135.696	47	8679278.536	9131826	-452547.464
6	1767910.187	1560356	207554.187	48	9140376.760	9316496	-176119.240
7	1775164.851	1745026	30138.851	49	9024572.973	9501166	-476593.027
8	1880117.299	1929696	-49578.701	50	9391883.853	9685836	-293952.147
9	1943934.288	2114366	-170431.712	51	9447420.289	9870506	-423085.711
10	2124121.794	2299036	-174914.206	52	9901523.442	10055176	-153652.558
11	2239478.929	2483706	-244227.071	53	10097726.589	10239846	-142119.411
12	2520238.798	2668376	-148137.202	54	10552369.212	10424516	127853.212
13	2749376.062	2853046	-103669.938	55	10603135.455	10609186	-6050.545
14	2903902.624	3037716	-133813.376	56	10899226.755	10793856	105370.755
15	3044154.698	3222386	-178231.302	57	10923201.557	10978526	-55324.443
16	3382698.666	3407056	-24357.334	58	11258204.123	11163196	95008.123
17	3498437.686	3591726	-93288.314	59	11402679.521	11347866	54813.521
18	3728952.344	3776396	-47443.656	60	12028967.970	11532536	496431.970
19	3827130.581	3961066	-133935.419	61	11836775.985	11717206	119569.985
20	4166567.200	4145736	20831.200	62	12480750.051	11901876	578874.051
21	4308085.303	4330406	-22320.697	63	12376956.511	12086546	290410.511
22	4460859.947	4515076	-54216.053	64	12332971.330	12271216	61755.330
23	4624146.663	4699746	-75599.337	65	11726944.569	12455886	-728941.431
24	4952370.733	4884416	67954.733	66	11789763.330	12640556	-850792.670
25	5212527.000	5069086	143441.000	67	12084463.505	12825226	-740762.495
26	5408820.010	5253756	155064.010	68	12774388.233	13009896	-235507.767
27	5610439.052	5438426	172013.052	69	12750648.740	13194566	-443917.260
28	5927679.283	5623096	304583.283	70	13082138.072	13379236	-297097.928
29	6211208.431	5807766	403442.431	71	13299168.637	13563906	-264737.363
30	6385478.202	5992436	393042.202	72	13996288.688	13748576	247712.688
31	6538881.963	6177106	361775.963	73	13901473.874	13933246	-31772.126
32	6721638.234	6361776	359862.234	74	14149231.456	14117916	31315.456
33	6670572.447	6546446	124126.447	75	14493670.424	14302586	191084.424
34	6724935.292	6731116	-6180.708	76	15581466.725	14487256	1094210.725
35	6773518.473	6915786	-142267.527	77	15249689.911	14671926	577763.911
36	6912567.691	7100456	-187888.309	78	15379454.110	14856596	522858.110
37	6763607.698	7285126	-521518.302	79	15557754.884	15041266	516488.884
38	7201786.460	7469796	-268009.540	80	16165492.995	15225936	939556.995
39	7218947.041	7654466	-435518.959	81	15628201.550	15410606	217595.550
40	7457653.418	7839136	-381482.582	82	15891371.809	15595276	296095.809
41	7535694.752	8023806	-488111.248	83	16074982.822	15779946	295036.822
42	7613721.754	8208476	-594754.246				

Tabla 2.3.2: Tabla que muestra la estimación de tendencia para el PIB de México

2.3.2. Modelo de tendencia polinómica

El modelo será de la forma:

$$Y_t = a_0 + a_1t + a_2t^2 + \dots + a_{k-1}t^{k-1} + \epsilon_t \quad (2.3.2)$$

con $k \in \mathbb{N}$ y a_i , $i = 0, \dots, k-1$, los coeficientes a estimar del modelo.

Como ejemplo, tomemos el modelo de tendencia cuadrática el cual estará dado por:

$$Y_t = a_0 + a_1t + a_2t^2 + \epsilon_t \quad (2.3.3)$$

Para detectar una tendencia cuadrática y no confundirla con una lineal podemos recurrir a la gráfica de las diferencias de la serie. La primera diferencia estará definida como:

$$\nabla Y_t = Y_t - Y_{t-1} \quad (2.3.4)$$

La segunda diferencia estará definida como:

$$\begin{aligned} \nabla^2 Y_t &= \nabla(\nabla Y_t) \\ &= \nabla(Y_t - Y_{t-1}) \\ &= \nabla Y_t - \nabla Y_{t-1} \end{aligned} \quad (2.3.5)$$

Para el caso de tendencia lineal, la primera diferencia oscila entorno a un valor medio constante, a saber:

$$\begin{aligned} \nabla Y_t &= Y_t - Y_{t-1} \\ &= [a + \beta t + \epsilon_t] - [a + \beta(t-1) + \epsilon_{t-1}] \\ &= \beta + (\epsilon_t - \epsilon_{t-1}) \end{aligned} \quad (2.3.6)$$

Para el caso de tendencia cuadrática, se necesitan dos diferencias para que se oscile entorno a un valor medio constante, a saber:

$$\begin{aligned} \nabla^2 Y_t &= \nabla(\nabla Y_t) \\ &= \nabla(Y_t - Y_{t-1}) \\ &= \nabla Y_t - \nabla Y_{t-1} \\ &= \nabla(a_0 + a_1t + a_2t^2 + \epsilon_t) - \nabla(a_0 + a_1(t-1) + a_2(t-1)^2 + \epsilon_{t-1}) \\ &= 2a_2 + \epsilon_t - 2\epsilon_{t-1} + \epsilon_{t-2} \end{aligned} \quad (2.3.7)$$

En general, el operador ∇^j extrae una tendencia de cualquier polinomio de orden j , es decir, $a_0 + a_1t + a_2t^2 + \dots + a_jt^j$. Esto se demostrará en el **capítulo 3**. Entonces, si la gráfica de la serie con una diferencia oscila entorno a una constante podemos hablar de una tendencia lineal. Si la gráfica con dos diferencias oscila entorno a una constante hablamos entonces de tendencia cuadrática y así sucesivamente. Es en este momento donde introducimos el paquete estadístico ITSM. Como bien sabemos, una técnica altamente utilizada para la estimación de la tendencia, T_t , es el método de mínimos cuadrados ordinarios. En este método buscamos ajustar una familia de funciones, a saber una de ellas la cuadrática, $T_t = a_0 + a_1t + a_2t^2$, a un conjunto de datos $\{x_1, x_2, \dots, x_n\}$, eligiendo los parámetros, a_0, a_1 y a_2 de tal forma que se minimice:

$$\sum_{t=1}^n (Y_t - T_t)^2$$

ITSM posee la opción “Regression” para realizar este método. Tomemos la serie correspondiente al total de importaciones de bienes y servicios de Australia en millones de dólares australianos a precios de 1984-1985. La serie comienza en septiembre de 1959 y termina en diciembre de 1990. Corremos ITSM y seleccionamos “File >Project >Open >Univariate” y elegimos el archivo “Imports.tsm”. La **Figura 2.3.3** muestra el total de importaciones.

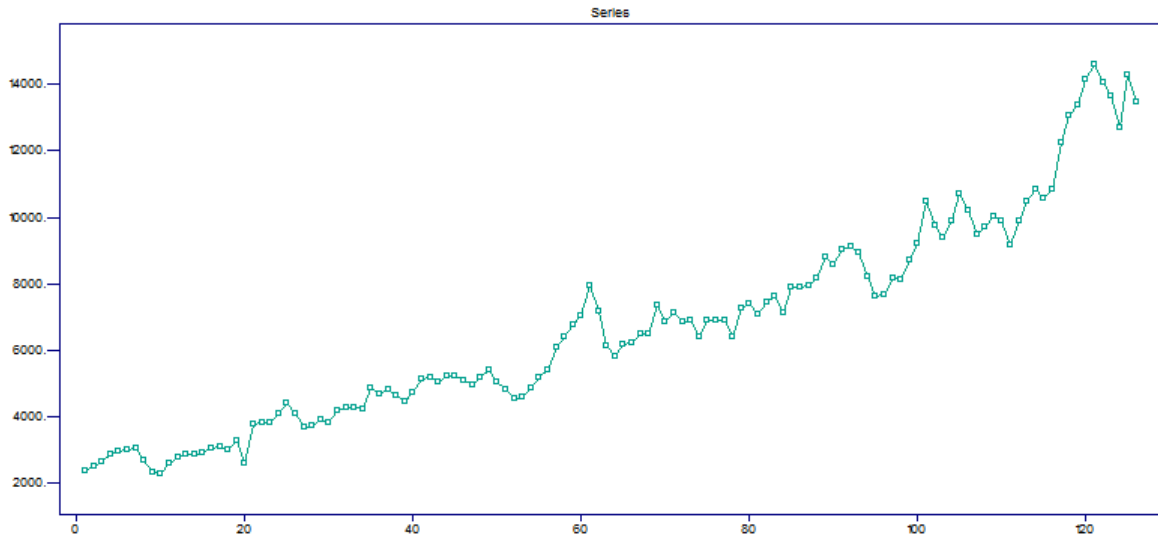


Figura 2.3.3: Figura que muestra el total de importaciones de bienes y servicios de Australia

A continuación seleccionamos “Regression >Especify >Polynomial Regression = 2”. Después seleccionamos “Regression >Estimation >Least Squares”. ITSM nos muestra los parámetros estimados, a saber, $a_0 = 2806.5676$, $a_1 = 21.79297867$ y $a_2 = 0.46695871$, y además “Regression >Show Fit”, nos muestra el ajuste de la función polinómica. La **Figura 2.3.4** muestra dicho ajuste.

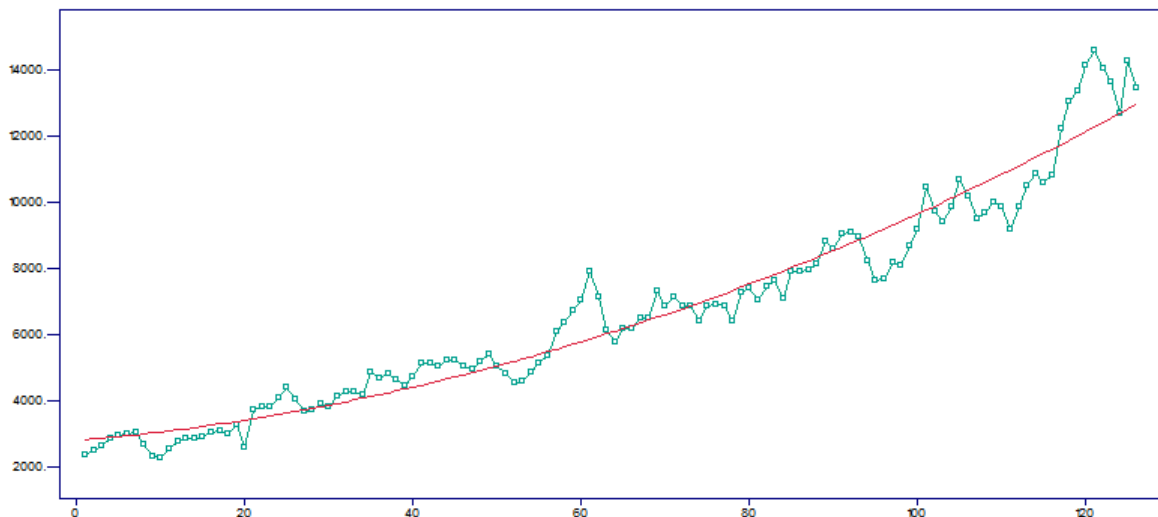


Figura 2.3.4: Figura que muestra el ajuste de la función polinómica al total de importaciones

De tal forma que $\hat{T}_t = 2806.5676 + 21.79297867t + 0.46695871t^2$ y ya que $Y_t = T_t + \epsilon_t$, podemos abstraer la tendencia y quedarnos con ϵ_t , el ruido blanco (definición que daremos en el **capítulo 3**), para después modelarlo con un proceso ARMA(p,q).

La **Figura 2.3.5** muestra los correspondientes residuos ϵ_t .

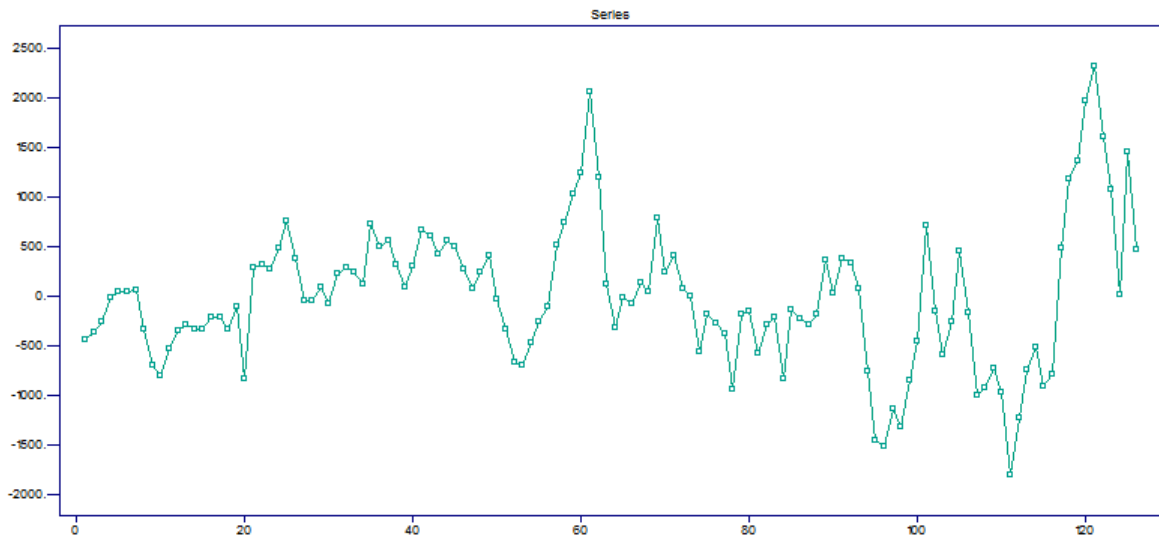


Figura 2.3.5: Figura que muestra a los residuos del total de importaciones australianas

2.3.3. Modelo de tendencia exponencial

El modelo será de la forma:

$$Y_t = e^{(a+\beta t)} \quad (2.3.8)$$

Dicho modelo es usado en series económicas principalmente. Una forma de saber si el modelo más adecuado es el exponencial es representar gráficamente la tasa de crecimiento de la variable y comprobar si más o menos oscila entorno a un valor constante. La tasa de crecimiento estará dada por:

$$TC_t = \frac{Y_t - Y_{t-1}}{Y_{t-1}} \quad (2.3.9)$$

Ya que hemos elegido el modelo de tendencia exponencial, aplicamos logaritmo a la serie, para así linealizar el modelo y poder utilizar la estimación por mínimos cuadrados ordinarios. Así, si $Y_t = e^{(a+\beta t)}$ entonces:

$$\begin{aligned} \ln(Y_t) &= \ln[e^{(a+\beta t)}] \\ &= a + \beta t \\ &= X_t \end{aligned} \quad (2.3.10)$$

y la estimación de los parámetros, \hat{a} y $\hat{\beta}$, por mínimos cuadrados, estarán dados por:

$$\hat{a} = \bar{X} - \hat{\beta} \bar{t} \quad (2.3.11)$$

En este caso, \bar{t} es el promedio de la variable explicativa y:

$$\hat{b} = \frac{\sum_{t=1}^T (t - \bar{t})(X_t - \bar{X})}{\sum_{t=1}^T (t - \bar{t})^2} \quad (2.3.12)$$

Como ejemplo tomemos la serie cuatrimestral de gastos en servicios financieros en Australia desde septiembre de 1969 hasta marzo de 1990. Una de nuestras opciones es ajustar la tendencia de la serie original Y_t con un modelo exponencial o bien, tomar el logaritmo natural de la serie original para después ajustar la tendencia de la nueva serie X_t con un modelo lineal. La serie original, el logaritmo natural de la serie original, la estimación de la tendencia T_t del logaritmo natural de la serie original y los residuos ϵ_t se muestran en la **Tabla 2.3.3**.

A continuación veamos el código R correspondiente a la estimación de la tendencia de la nueva serie $X_t = \ln(Y_t)$:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Ejercicios")
> Finserv<-read.table("Finserv.txt",header=T)
> attach(Finserv)
> Ajuste<-lm(lnFINSERV~t)
> summary(Ajuste)
```

Call:

```
lm(formula = lnFINSERV ~ t)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.15365	-0.06999	-0.02909	0.06746	0.28100

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.8918614	0.0222772	264.48	<2e-16 ***
t	0.0174458	0.0004448	39.22	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1024 on 84 degrees of freedom

Multiple R-squared: 0.9482, Adjusted R-squared: 0.9476

F-statistic: 1538 on 1 and 84 DF, p-value: < 2.2e-16

```
> library("graphics")
> par(mfrow=c(1,2))
> plot(t,lnFINSERV,pch=21,xlim=c(0,90),main="Servicios Financieros",col="blue")
> plot(t,lnFINSERV,pch=21,xlim=c(0,90),main="Servicios Financieros",col="blue")
> abline(Ajuste,col="red")
```

La **Figura 2.3.6** muestra el logaritmo natural del gasto en servicios financieros australianos así como el ajuste de tendencia lineal.

t	Y_t	X_t	T_t	ϵ_t	t	Y_t	X_t	T_t	ϵ_t
1	393	5.9738	5.909307	0.06449286	44	738	6.6039	6.659475	-0.05557521
2	444	6.0958	5.926753	0.16904709	45	742	6.6093	6.676921	-0.06762098
3	403	5.9989	5.944199	0.05470133	46	774	6.6516	6.694367	-0.04276675
4	403	5.9989	5.961644	0.03725556	47	770	6.6464	6.711813	-0.06541252
5	430	6.0638	5.979090	0.08470979	48	792	6.6746	6.729258	-0.05465829
6	453	6.1159	5.996536	0.11936402	49	801	6.6859	6.746704	-0.06080406
7	413	6.0234	6.013982	0.00941825	50	805	6.6908	6.764150	-0.07334983
8	424	6.0497	6.031428	0.01827248	51	804	6.6896	6.781596	-0.09199559
9	432	6.0684	6.048873	0.01952671	52	836	6.7286	6.799041	-0.07044136
10	440	6.0868	6.066319	0.02048094	53	841	6.7346	6.816487	-0.08188713
11	445	6.0981	6.083765	0.01433517	54	851	6.7464	6.833933	-0.08753290
12	447	6.1026	6.101211	0.00138940	55	860	6.7569	6.851379	-0.09447867
13	503	6.2206	6.118656	0.10194363	56	871	6.7696	6.868824	-0.09922444
14	497	6.2086	6.136102	0.07249786	57	888	6.7890	6.886270	-0.09727021
15	498	6.2106	6.153548	0.05705210	58	878	6.7776	6.903716	-0.12611598
16	505	6.2246	6.170994	0.05360633	59	898	6.8002	6.921162	-0.12096175
17	569	6.3439	6.188439	0.15546056	60	897	6.7991	6.938608	-0.13950752
18	560	6.3279	6.205885	0.12201479	61	900	6.8024	6.956053	-0.15365329
19	556	6.3208	6.223331	0.09746902	62	921	6.8255	6.973499	-0.14799906
20	560	6.3279	6.240777	0.08712325	63	951	6.8575	6.990945	-0.13344482
21	531	6.2748	6.258223	0.01657748	64	957	6.8638	7.008391	-0.14459059
22	546	6.3026	6.275668	0.02693171	65	999	6.9068	7.025836	-0.11903636
23	544	6.2989	6.293114	0.00578594	66	1009	6.9167	7.043282	-0.12658213
24	547	6.3044	6.310560	-0.00615983	67	1021	6.9285	7.060728	-0.13222790
25	544	6.2989	6.328006	-0.02910560	68	1068	6.9735	7.078174	-0.10467367
26	559	6.3261	6.345451	-0.01935137	69	1083	6.9875	7.095619	-0.10811944
27	548	6.3063	6.362897	-0.05659714	70	1118	7.0193	7.113065	-0.09376521
28	551	6.3117	6.380343	-0.06864290	71	1185	7.0775	7.130511	-0.05301098
29	579	6.3613	6.397789	-0.03648867	72	1222	7.1082	7.147957	-0.03975675
30	584	6.3699	6.415234	-0.04533444	73	1349	7.2071	7.165403	0.04169749
31	604	6.4036	6.432680	-0.02908021	74	1410	7.2513	7.182848	0.06845172
32	602	6.4003	6.450126	-0.04982598	75	1445	7.2759	7.200294	0.07560595
33	591	6.3818	6.467572	-0.08577175	76	1491	7.3072	7.217740	0.08946018
34	629	6.4441	6.485018	-0.04091752	77	1534	7.3356	7.235186	0.10041441
35	645	6.4693	6.502463	-0.03316329	78	1574	7.3614	7.252631	0.10876864
36	667	6.5028	6.519909	-0.01710906	79	1640	7.4025	7.270077	0.13242287
37	667	6.5028	6.537355	-0.03455483	80	1710	7.4442	7.287523	0.15667710
38	667	6.5028	6.554801	-0.05200060	81	1809	7.5005	7.304969	0.19553133
39	696	6.5453	6.572246	-0.02694636	82	1850	7.5229	7.322414	0.20048556
40	708	6.5624	6.589692	-0.02729213	83	1907	7.5533	7.339860	0.21343979
41	718	6.5765	6.607138	-0.03063790	84	1972	7.5868	7.357306	0.22949402
42	728	6.5903	6.624584	-0.03428367	85	2063	7.6319	7.374752	0.25714826
43	729	6.5917	6.642029	-0.05032944	86	2150	7.6732	7.392198	0.28100249

Tabla 2.3.3: Tabla que muestra la estimación de tendencia exponencial

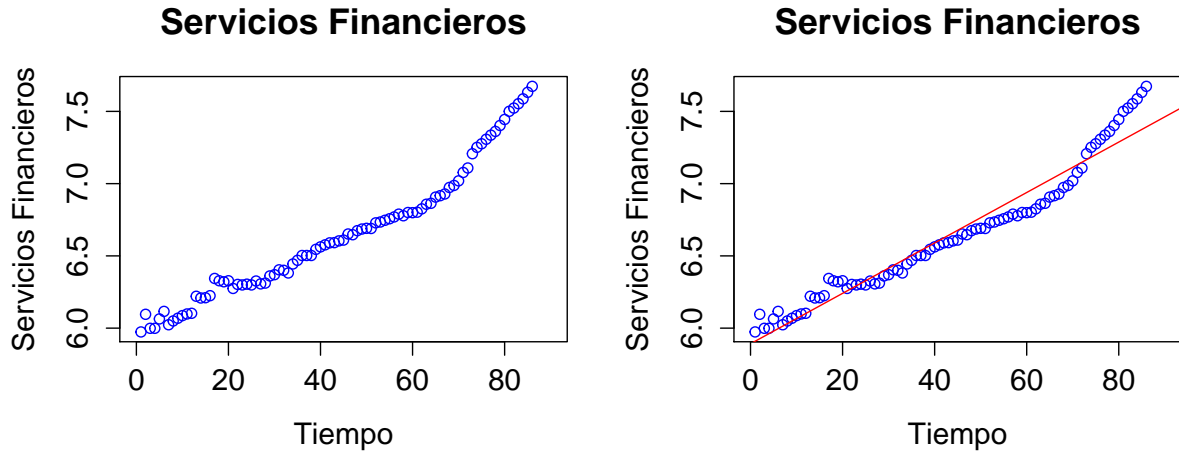


Figura 2.3.6: Figura que muestra el gasto en servicios financieros en Australia

2.3.4. Método de promedios móviles

Otro método para modelar tendencia es con un filtro de **promedios móviles**, el cual es uno de los métodos de suavizamiento que estudiaremos en este capítulo. Sea $q > 0$ con $q \in \mathbb{Z}$. Aplicar un filtro de promedios móviles a una serie Y_t consiste en obtener:

$$W_t = \frac{1}{2q+1} \sum_{j=-q}^q Y_{t+j} \quad q+1 \leq t \leq n-q \quad (2.3.13)$$

Por ejemplo si $q = 1$ tenemos:

$$W_t = \frac{1}{3} \sum_{j=-1}^1 Y_{t+j} \quad (2.3.14)$$

es decir, el promedio del valor de Y posterior a t , el anterior y el actual. Hemos supuesto desde un principio que $Y_t = T_t + \epsilon_t$. Ahora bien, suponemos:

1. T_t es aproximadamente lineal sobre el intervalo $[t-q, t+q]$
2. $\frac{1}{2q+1} \sum_{j=-q}^q \epsilon_{t+j} \approx 0$

Entonces, al aplicar el filtro de promedios móviles a Y_t obtenemos, usando $Y_t = T_t + \epsilon_t$:

$$\begin{aligned} W_t &= \frac{1}{2q+1} \sum_{j=-q}^q T_{t+j} + \frac{1}{2q+1} \sum_{j=-q}^q \epsilon_{t+j} \approx \frac{1}{2q+1} \sum_{j=-q}^q (a + \beta(t+j)) \\ &= \frac{1}{2q+1} \{a(2q+1)\} + \frac{1}{2q+1} \left\{ \beta t(2q+1) + \beta \sum_{j=-q}^q j \right\} \\ &= a + \beta t \end{aligned} \quad (2.3.15)$$

Es decir que, al aplicar el filtro obtenemos la tendencia $T_t = a + \beta t$, aunque en realidad, la estimamos como \hat{T}_t . Es de remarcar que el filtro de promedios móviles sólo sirve para los puntos $t \in [q + 1, n - q]$. Entonces, ¿qué hacer para los otros puntos? Se puede tomar $Y_t = Y_1$ para $t < q + 1$, y $Y_t = Y_{n-q}$ para $t > n - q$.

Hemos visto el filtro de promedio móviles para series con tendencia únicamente, pero si en algún momento llegásemos a tener una serie explicada por las tres componentes antes vistas, y usando el esquema aditivo, es decir, $Y_t = T_t + E_t + I_t$, la elección correcta del parámetro q nos identificaría de manera automática la tendencia de la serie, ya que al promediar varios valores de una serie obtenemos una nueva serie mucho más suavizada e identificada con la tendencia de la serie original. Además, al elegir una q adecuada, las variaciones causadas por la estacionalidad y componente irregular tenderán a anularse en promedio, reflejando así, las principales características de la serie.

Pero, ¿cómo elegir el valor de q adecuadamente?, o bien, ¿de qué depende la elección de una q pequeña o grande? Mientras mayor sea el valor de q , mayores datos intervendrán en el cálculo de W_t , por tanto, mayor será el suavizamiento conseguido con el filtro de promedios móviles y mejor se eliminarán las irregularidades de la serie. Por el contrario, cuando intervienen pocos datos en el filtro, es decir, el valor de q es pequeño, W_t refleja con mayor rapidez las variaciones de la serie a lo largo del tiempo. Pero cuidado, el parámetro q no debe ser muy grande ya que para q grande este filtro deja pasar funciones T_t lineales.

Ejemplo 2.3.1 Si $\{Z_t\}$ con $t = 0, \pm 1, \pm 2, \dots$, es un conjunto de variables aleatorias independientes con media $\mu = 0$ y varianza σ_Z^2 , demostrar que el promedio móvil:

$$A_t = \sum_{j=-q}^q a_t Z_{t-j}$$

con $a_t = \frac{1}{2q+1}$, es pequeño para q grande, en el sentido de que $\mathbb{E}(A_t) = 0$ y $\text{Var}(A_t) = \frac{\sigma_Z^2}{2q+1}$. Comencemos con la esperanza. Tenemos:

$$\mathbb{E}(A_t) = \mathbb{E} \left(\sum_{j=-q}^q a_t Z_{t-j} \right) = \mathbb{E} \left(\frac{1}{2q+1} \sum_{j=-q}^q Z_{t-j} \right) = \frac{1}{2q+1} \sum_{j=-q}^q \mathbb{E}(Z_{t-j})$$

y por tanto $\mathbb{E}(A_t) = 0$ ya que el proceso $\{Z_t\}$ es independiente y con media $\mu = 0$. Para la varianza tenemos:

$$\text{Var}(A_t) = \text{Var} \left(\sum_{j=-q}^q a_t Z_{t-j} \right) = \sum_{j=-q}^q a_t^2 \text{Var}(Z_{t-j}) = \frac{1}{(2q+1)^2} \sum_{j=-q}^q \sigma_Z^2 = \frac{\sigma_Z^2}{2q+1}$$

ya que el proceso $\{Z_t\}$ es independiente y absolutamente sumable.

Pasemos al software estadístico. Realicemos el método de promedios móviles, con $q = 1$ y $q = 2$, en el número de acciones de atención. En R se usa la función `filter`, instrucción después de la cual damos los pesos del filtro; para promedios móviles es $\frac{1}{2q+1}$. Veamos el código en R:

```

> library(stats)
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Condusef")
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))

      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2007 31101 31909 34435 32380 37096 36216 40367 39902 38954 49412 40412 23355
2008 45553 45057 40029 49268 43667 52294 62229 59801 56553 70527 59687 52113
2009 65814 71137 89913 79073 76695 92248 100651 81910 88545 94181 82350 51894
2010 84094 86772 106082 87916 83399 92487 94883 97899 85130 96060 97379 74898
2011 82854 90295 103375 84414 98046 98633 95882 99991 93020 93480 94409 74088
2012 103356 110649 113157 99107 112468 100971 111903 112614 93443 120379 99046 62274

> Promediosmov <- filter(Serie,filter=rep(1/3,3)) # Promedios móviles con q=1

      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2007  NA  32481.67 32908.00 34637.00 35230.67 37893.00 38828.33 39741.00 42756.00 42926.00 37726.33 36440.00
2008 37988.33 43546.33 44784.67 44321.33 48409.67 52730.00 58108.00 59527.67 62293.67 62255.67 60775.67 59204.67
2009 63021.33 75621.33 80041.00 81893.67 82672.00 89864.67 91603.00 90368.67 88212.00 88358.67 76141.67 72779.33
2010 74253.33 92316.00 93590.00 92465.67 87934.00 90256.33 95089.67 92637.33 93029.67 92856.33 89445.67 85043.67
2011 82682.33 92174.67 92694.67 95278.33 93697.67 97520.33 98168.67 96297.67 95497.00 93636.33 87325.67 90617.67
2012 96031.00 109054.00 107637.67 108244.00 104182.00 108447.33 108496.00 105986.67 108812.00 104289.33 93899.67  NA

> Promediosmov2 <- filter(Serie,filter=rep(1/5,5)) # Promedios móviles con q=2

      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
2007  NA  NA  33384.2 34407.2 36098.8 37192.2 38507.0 40970.2 41809.4 38407.0 39537.2 40757.8
2008 38881.2 40652.4 44714.8 46063.0 49497.4 53451.8 54908.8 60280.8 61759.4 59736.2 60938.8 63855.6
2009 67732.8 71610.0 76526.4 81813.2 87716.0 86115.4 88009.8 91507.0 89527.4 79776.0 80212.8 79858.2
2010 82238.4 83351.6 89652.6 91331.2 92953.4 91316.8 90759.6 93291.8 94270.2 90273.2 87264.2 88297.2
2011 89760.2 87167.2 91796.8 94952.6 96070.0 95393.2 97114.4 96201.2 95356.4 90997.6 91670.6 95196.4
2012 99131.8 100071.4 107747.4 107270.4 107521.2 107412.6 106279.8 107862.0 107477.0 97551.2  NA  NA

> plot (c(2007,2013),c(30000,130000),type="n",main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black")
> lines(Promediosmov,col="red")
> lines(Promediosmov2,col="blue")
> legend(2007,130000,c("AccAtn","Móviles q=1","Móviles q=2"),lty=c(1,1,1),bty="n",lwd=c(2.5,2.5,2.5),col=c("black","red","blue"))

```

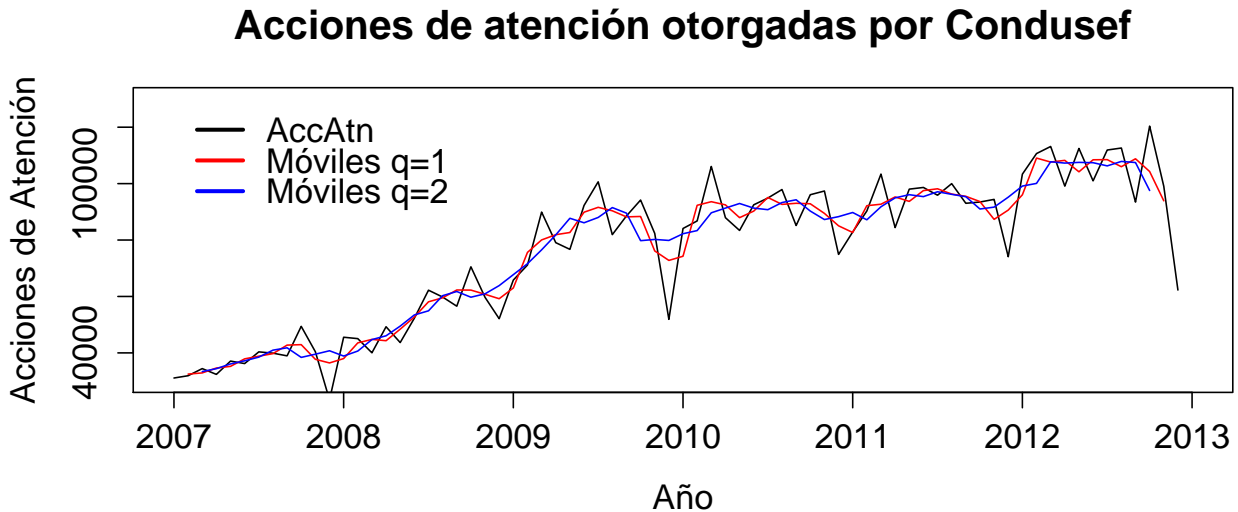


Figura 2.3.7: Suavizamiento por promedios móviles en el número de acciones de atención

En el paquete estadístico ITSM tenemos también disponible el método de promedios móviles. Tomemos la serie correspondiente al nivel (pies) del lago Hurón en los años 1875-1972. Ahora bien, seleccionamos “File >Project >Open >Univariate” y elegimos el archivo llamado “Lake.tsm”. La gráfica correspondiente se muestra en la **Figura 2.3.8**.

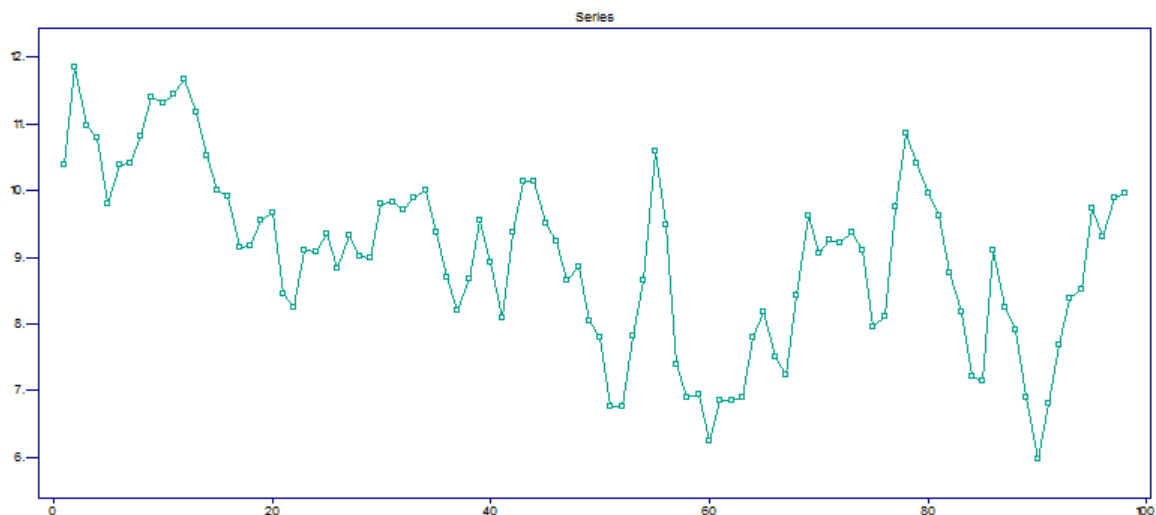


Figura 2.3.8: Figura correspondiente al nivel (pies) del lago Hurón

A continuación seleccionamos “Smooth >Moving Average >Enter Filter Order = 2” con “Theta(0) = Theta(1) = Theta(2) = 1 >Ok”. La **Figura 2.3.9** muestra el suavizamiento por promedios móviles. Recordemos que entre mayor sea el orden del filtro mayor será el suavizamiento de la serie, y entre menor sea el orden del filtro con mayor rapidez se reflejarán las oscilaciones de la serie. Por ejemplo, si tomamos “Smooth >Moving Average >Enter Filter Order = 8 >Theta(0) = Theta(1) = Theta(2) = ... = Theta(8) = 1 >Ok”, obtenemos la **Figura 2.3.10**, la cual claramente nos da un mayor suavizamiento que con un filtro igual a 2.

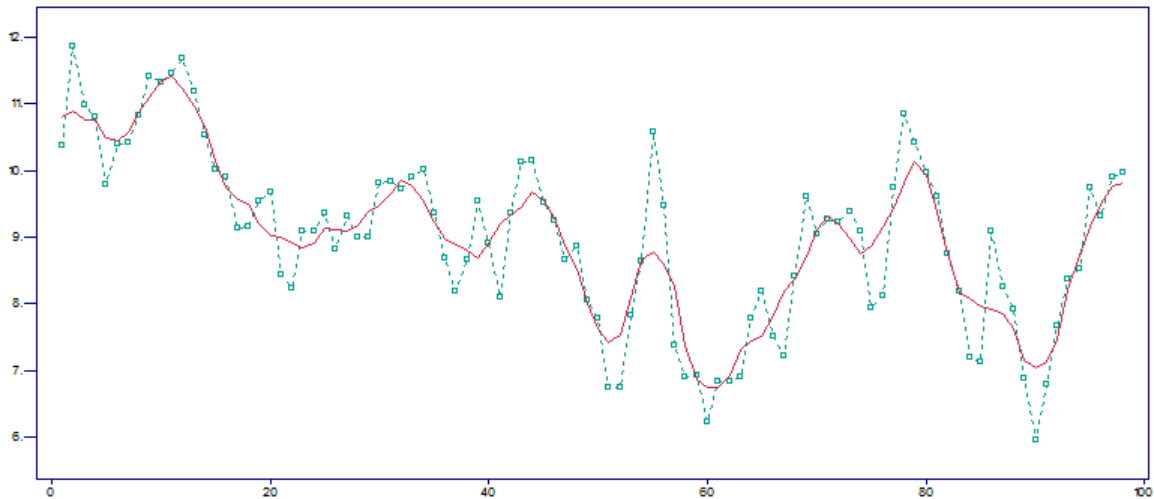


Figura 2.3.9: Suavizamiento por promedios móviles del nivel (pies) del lago Hurón con $q=2$

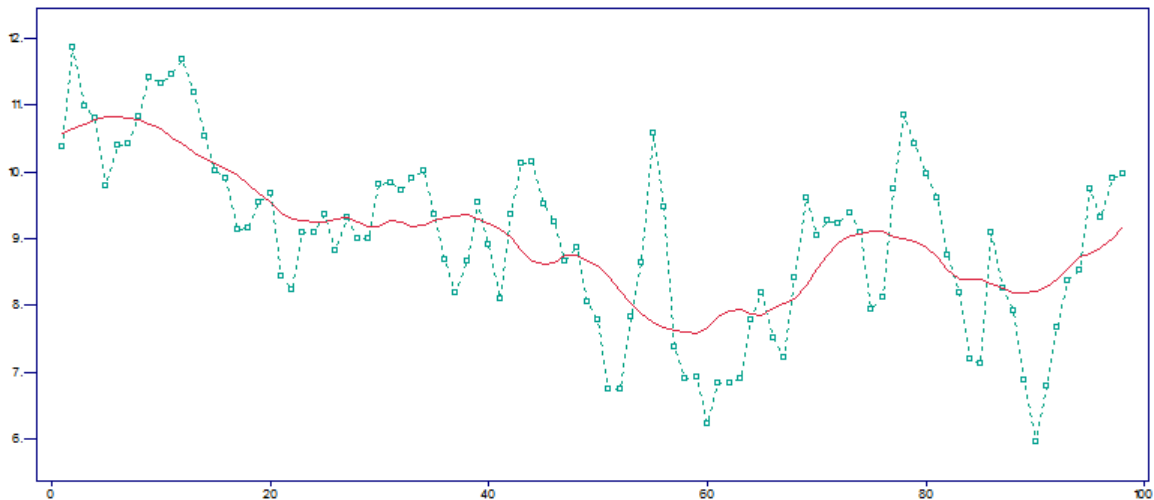


Figura 2.3.10: Suavizamiento por promedios móviles del nivel (pies) del lago Hurón con $q=8$

2.3.5. Suavizamiento exponencial simple

La principal limitación que tiene el procedimiento de promedios móviles es que pondera de la misma manera los valores que integran el promedio.

Pero quizás, cuando se trate de realizar una predicción, un pronóstico coherente de una observación Y_{t+1} sería el de asignar un promedio ponderado de los valores anteriores a este tiempo, es decir:

$$\hat{Y}_{t+1} = \sum_{i=0}^{\infty} w_i Y_{t-i} \quad \text{con} \quad \sum_{i=0}^{\infty} w_i = 1 \quad (2.3.16)$$

Como las observaciones recientes deberían de ser las más relevantes para nuestro pronóstico, éstas deberían tener mucho más peso que las observaciones correspondientes al pasado.

Una manera de hacerlo es asignando pesos que decrecen geométricamente:

$$w_i = \alpha(1 - \alpha)^i \quad \text{con} \quad 0 < \alpha < 1 \quad \text{e} \quad i = 0, 1, 2, \dots \quad (2.3.17)$$

Por ejemplo, si $\alpha = 0.5$ entonces:

$$\begin{aligned} w_0 &= 0.5 \\ w_1 &= 0.25 \\ w_2 &= 0.125 \end{aligned} \quad (2.3.18)$$

y así sucesivamente. Entonces:

$$\begin{aligned} \hat{Y}_{t+1} &= \alpha Y_t + \alpha(1 - \alpha)Y_{t-1} + \alpha(1 - \alpha)^2 Y_{t-2} + \dots \\ &= \alpha Y_t + (1 - \alpha)[\alpha Y_{t-1} + \alpha(1 - \alpha)Y_{t-2} + \dots] \\ &= \alpha Y_t + (1 - \alpha)\hat{Y}_t \end{aligned} \quad (2.3.19)$$

Entonces, el pronóstico en $t+1$ depende de la observación anterior, del pronóstico dado a esa misma observación y de un valor de α dado, $0 < \alpha < 1$. Un valor de α cercano a cero querría decir que le damos mayor importancia a la parte de suavizamiento. Valores de α más próximos a 1, dan más pesos a las observaciones recientes y nos proporcionará una serie de pronósticos con menor suavizamiento. No obstante, existe un criterio que permite seleccionar el valor de α más adecuado y es el de seleccionar el valor de α que minimice el **error cuadrático medio**, es decir, que minimice:

$$\frac{1}{N} \sum_{t=1}^N \eta_t^2 \quad (2.3.20)$$

donde N es el número de observaciones de la serie y $\eta_t = Y_t - \hat{Y}_t$. Para iniciar el algoritmo se puede tomar $\hat{Y}_1 = Y_1$ y empezar los ajustes desde \hat{Y}_2 con el suavizamiento exponencial simple. En todos los casos se puede obtener el error de la observación como:

$$\eta_t = Y_t - \hat{Y}_t \quad (2.3.21)$$

Entonces, bajo el suavizamiento exponencial simple:

$$\begin{aligned} \hat{Y}_{t+1} &= \alpha Y_t + (1 - \alpha)\hat{Y}_t \\ &= \alpha(Y_t - \hat{Y}_t) + \hat{Y}_t \\ &= \alpha\eta_t + \hat{Y}_t \end{aligned} \quad (2.3.22)$$

Como ejemplo, tomemos nuevamente el número de acciones de atención brindadas por Conducef y realicemos, en R, el suavizamiento exponencial simple.

El suavizamiento exponencial simple, método de Holt y método de Holt- Winters (multiplicativo y aditivo), los cuales se verán en las **Secciones 2.3.6** y **2.4.2**, se basan en la misma función `HoltWinters` ofrecida por R, la cual depende de 3 parámetros, α , β y γ . Si $\gamma = \text{FALSE}$, el suavizamiento será el de Holt. Si $\beta = \gamma = \text{FALSE}$, el suavizamiento será exponencial simple. Si no se especifican los parámetros, estos se determinan de forma automática de acuerdo con el error cuadrático medio. El código permitiendo a R la elección del parámetro α óptimo, será:

```
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> Exponencialsimple<- HoltWinters(Serie, gamma = FALSE, beta = FALSE)
```

Holt-Winters exponential smoothing without trend and without seasonal component.

Call:

```
HoltWinters(x = Serie, beta = FALSE, gamma = FALSE)
```

Smoothing parameters:

```
alpha: 0.397987
beta : FALSE
gamma: FALSE
```

Coefficients:

```
 [,1]
a 88370.7
```

```
> ValoresAjust<-fitted(Exponencialsimple)
> plot(c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black"); lines(fitted(Exponencialsimple)[,1], col = "red")
> legend(2007,120000,c("AccAtn","Exponencial simple"),bty="n",col=c("black","red"))
```

La **Tabla 2.3.4** muestra el número de acciones de atención otorgadas por Condusef, Y_t , desde enero de 2007 hasta diciembre de 2012, así como el valor ajustado por el método de suavizamiento exponencial simple, \hat{Y}_t . La **Figura 2.3.11** muestra el suavizamiento exponencial simple realizado a las acciones de atención con parámetro α óptimo e igual a 0.397987.

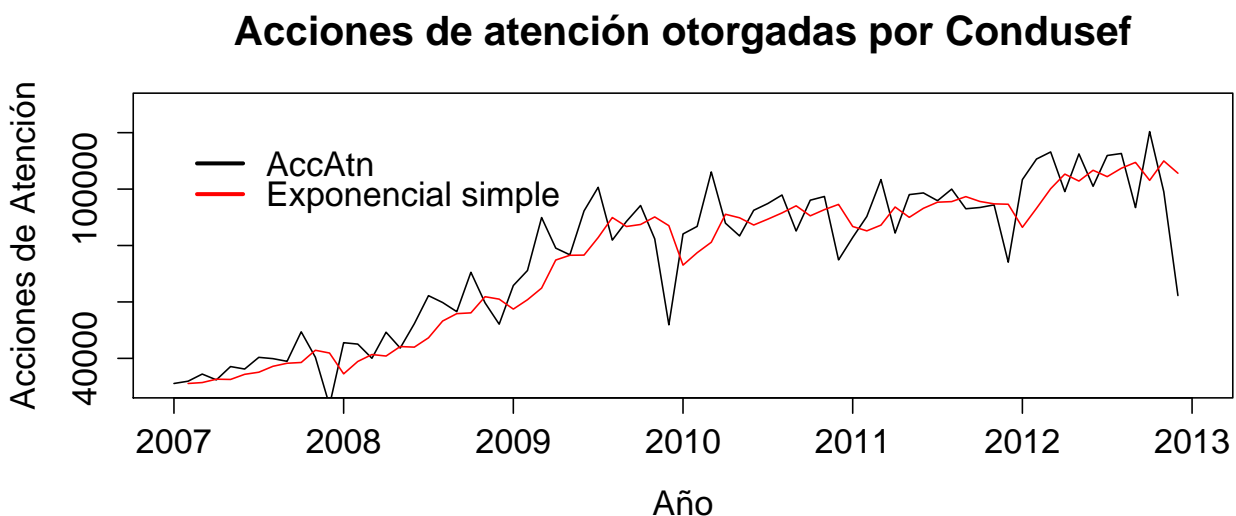


Figura 2.3.11: Suavizamiento exponencial simple en el número de acciones de atención

Mes	Año	Y_t	\hat{Y}_t	Mes	Año	Y_t	\hat{Y}_t
Ene	2007	31101	31101	Ene	2010	84094	73049
Feb	2007	31909	31101	Feb	2010	86772	77445
Mar	2007	34435	31423	Mar	2010	106082	81157
Abr	2007	32380	32621	Abr	2010	87916	91077
May	2007	37096	32525	May	2010	83399	89819
Jun	2007	36216	34344	Jun	2010	92487	87264
Jul	2007	40367	35089	Jul	2010	94883	89343
Ago	2007	39902	37190	Ago	2010	97899	91548
Sep	2007	38954	38269	Sep	2010	85130	94075
Oct	2007	49412	38542	Oct	2010	96060	90515
Nov	2007	40412	42868	Nov	2010	97379	92722
Dic	2007	23355	41891	Dic	2010	74898	94575
Ene	2008	45553	34514	Ene	2011	82854	86744
Feb	2008	45057	38907	Feb	2011	90295	85196
Mar	2008	40029	41355	Mar	2011	103375	87225
Abr	2008	49268	40827	Abr	2011	84414	93653
May	2008	43667	44186	May	2011	98046	89976
Jun	2008	52294	43980	Jun	2011	98633	93188
Jul	2008	62229	47289	Jul	2011	95882	95355
Ago	2008	59801	53235	Ago	2011	99991	95565
Sep	2008	56553	55848	Sep	2011	93020	97326
Oct	2008	70527	56129	Oct	2011	93480	95612
Nov	2008	59687	61859	Nov	2011	94409	94764
Dic	2008	52113	60995	Dic	2011	74088	94623
Ene	2009	65814	57460	Ene	2012	103356	86450
Feb	2009	71137	60785	Feb	2012	110649	93178
Mar	2009	89913	64905	Mar	2012	113157	100131
Abr	2009	79073	74858	Abr	2012	99107	105315
May	2009	76695	76535	May	2012	112468	102845
Jun	2009	92248	76599	Jun	2012	100971	106675
Jul	2009	100651	82827	Jul	2012	111903	104405
Ago	2009	81910	89921	Ago	2012	112614	107389
Sep	2009	88545	86733	Sep	2012	93443	109468
Oct	2009	94181	87454	Oct	2012	120379	103091
Nov	2009	82350	90131	Nov	2012	99046	109971
Dic	2009	51894	87034	Dic	2012	62274	105623

Tabla 2.3.4: Tabla que muestra el suavizamiento exponencial simple en las acciones de atención

Si nosotros quisieramos dar el parámetro α a R, por ejemplo $\alpha = 0.8$, el código será:

```
> Exponencialsimple2 <- HoltWinters(Serie, alpha=0.8, gamma = FALSE, beta = FALSE)
> ValoresAjust2<-fitted(Exponencialsimple2)
> plot(c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black"); lines(fitted(Exponencialsimple2)[,1],col = "green")
> legend(2007,130000,c("AccAtn","Exponencial simple(0.8)"),col=c("black","green"))
```

La **Tabla 2.3.5** muestra nuevamente el número de acciones de atención otorgadas por Condusef, Y_t , así como el valor ajustado por el método de suavizamiento exponencial simple, \hat{Y}_t , con $\alpha = 0.8$. La **Figura 2.3.12** muestra el suavizamiento exponencial simple realizado.

Mes	Año	Y_t	\hat{Y}_t	Mes	Año	Y_t	\hat{Y}_t
Ene	2007	31101	31101	Ene	2010	84094	58408
Feb	2007	31909	31101	Feb	2010	86772	78957
Mar	2007	34435	31747	Mar	2010	106082	85209
Abr	2007	32380	33897	Abr	2010	87916	101907
May	2007	37096	32683	May	2010	83399	90714
Jun	2007	36216	36213	Jun	2010	92487	84862
Jul	2007	40367	36215	Jul	2010	94883	90962
Ago	2007	39902	39537	Ago	2010	97899	94099
Sep	2007	38954	39829	Sep	2010	85130	97139
Oct	2007	49412	39129	Oct	2010	96060	87532
Nov	2007	40412	47355	Nov	2010	97379	94354
Dic	2007	23355	41801	Dic	2010	74898	96774
Ene	2008	45553	27044	Ene	2011	82854	79273
Feb	2008	45057	41851	Feb	2011	90295	82138
Mar	2008	40029	44416	Mar	2011	103375	88664
Abr	2008	49268	40906	Abr	2011	84414	100433
May	2008	43667	47596	May	2011	98046	87618
Jun	2008	52294	44453	Jun	2011	98633	95960
Jul	2008	62229	50726	Jul	2011	95882	98098
Ago	2008	59801	59928	Ago	2011	99991	96325
Sep	2008	56553	59826	Sep	2011	93020	99258
Oct	2008	70527	57208	Oct	2011	93480	94268
Nov	2008	59687	67863	Nov	2011	94409	93638
Dic	2008	52113	61322	Dic	2011	74088	94255
Ene	2009	65814	53955	Ene	2012	103356	78121
Feb	2009	71137	63442	Feb	2012	110649	98309
Mar	2009	89913	69598	Mar	2012	113157	108181
Abr	2009	79073	85850	Abr	2012	99107	112162
May	2009	76695	80428	May	2012	112468	101718
Jun	2009	92248	77442	Jun	2012	100971	110318
Jul	2009	100651	89287	Jul	2012	111903	102840
Ago	2009	81910	98378	Ago	2012	112614	110090
Sep	2009	88545	85204	Sep	2012	93443	112109
Oct	2009	94181	87877	Oct	2012	120379	97176
Nov	2009	82350	92920	Nov	2012	99046	115738
Dic	2009	51894	84464	Dic	2012	62274	102384

Tabla 2.3.5: Suavizamiento exponencial simple en las acciones de atención con $\alpha = 0.8$

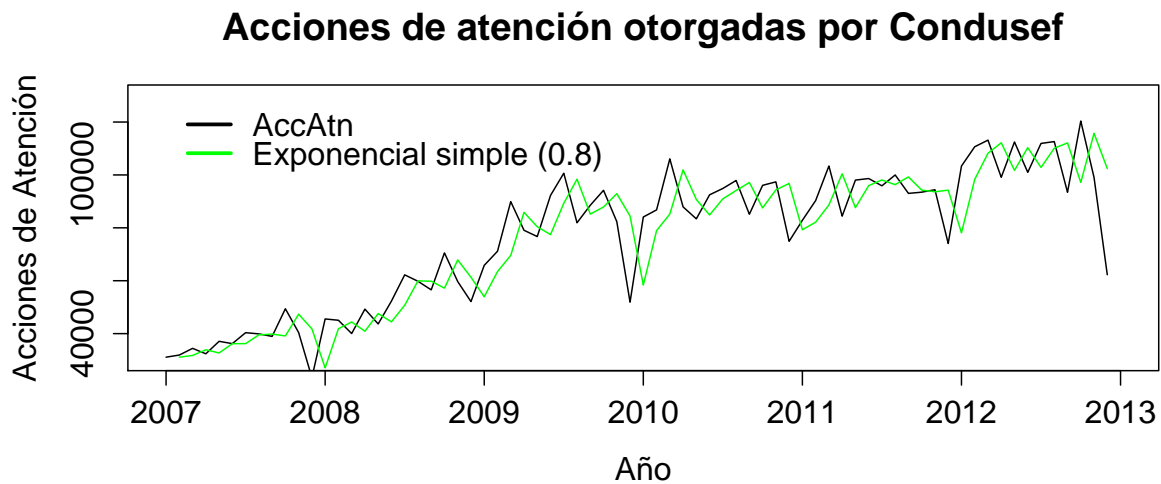


Figura 2.3.12: Suavizamiento exponencial simple ($\alpha = 0.8$) del número de acciones de atención

Nuevamente, ITSM nos da la opción de realizar el método de suavizamiento exponencial simple. Tomemos de nuevo la serie correspondiente al total de importaciones de bienes y servicios de Australia en millones de dólares australianos a precios de 1984-1985. Seleccionamos “File >Project >Open >Univariate” y elegimos el archivo “Imports.tsm”. La **Figura 2.3.13** muestra el total de importaciones de bienes y servicios australianos. A continuación seleccionamos “Smooth >Exponential >Alpha = 0.30 >Ok”. ITSM nos muestra el ajuste del suavizamiento exponencial simple. La **Figura 2.3.14** muestra dicho suavizamiento. Es de nuestra elección el parámetro α pero ITSM nos da la opción de elegir el parámetro óptimo, aquel que minimiza el error cuadrático medio, al escribir -1 en el parámetro α . En este caso el valor óptimo es $\alpha = 0.99$.

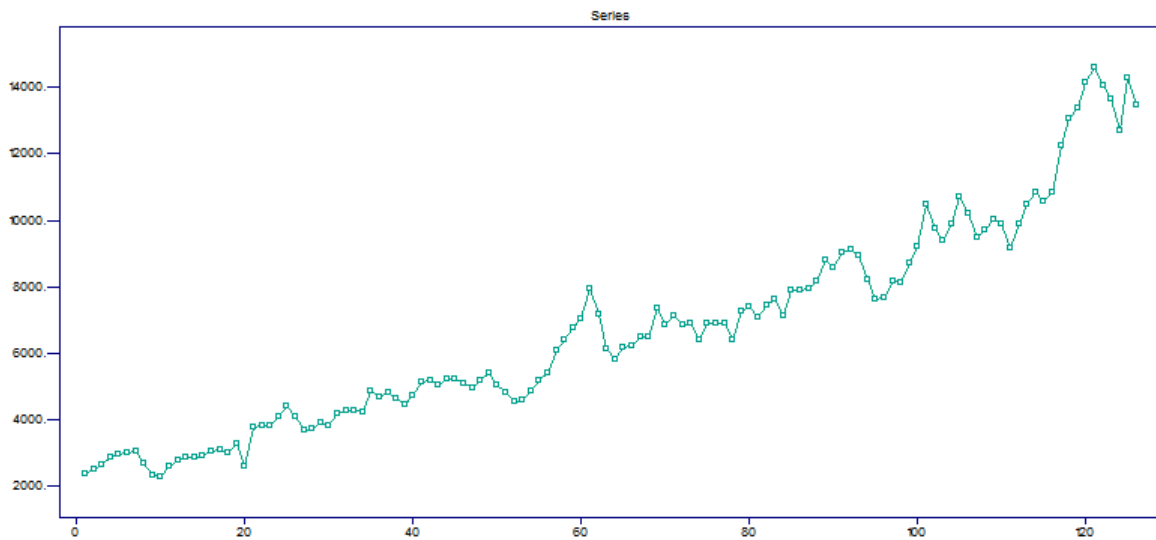


Figura 2.3.13: Figura que muestra el total de importaciones de bienes y servicios de Australia

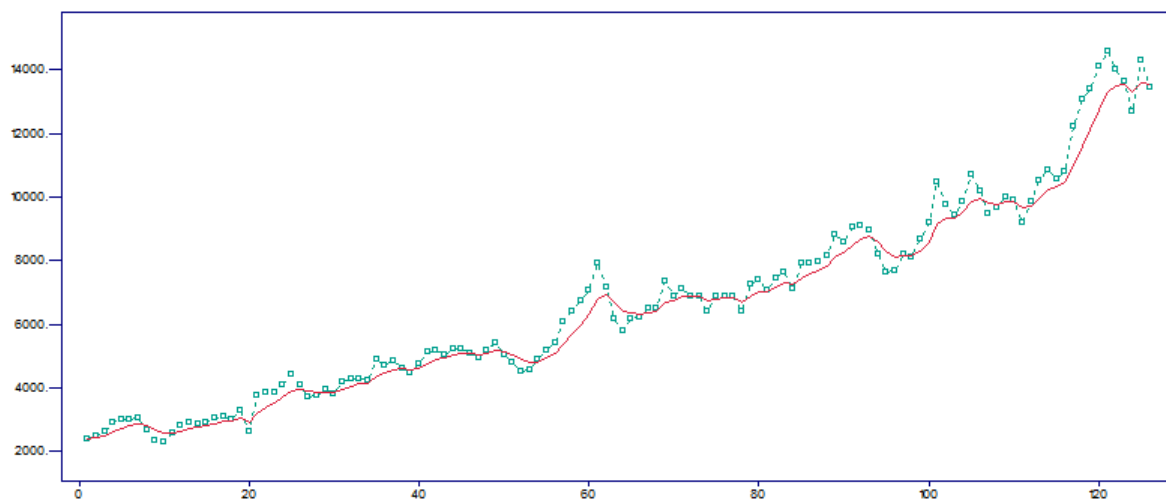


Figura 2.3.14: Suavizamiento exponencial simple del total de importaciones ($\alpha=0.3$)

2.3.6. Método de Holt o doble exponencial

Bajo este método, la estimación depende de una tendencia, así que se toma en cuenta una componente correspondiente a la pendiente en el método exponencial.

Se considera entonces un nivel L_t y una pendiente b_t y se realiza un suavizamiento exponencial simple en cada uno de los términos, por tanto se requieren 2 constantes de suavizamiento, a saber α para el nivel y β para la pendiente. Suponemos además $0 < \alpha, \beta < 1$. L_t , el nivel, se calcula de la misma forma que la estimación de \hat{Y}_{t+1} del suavizamiento exponencial simple, es decir:

$$\begin{aligned}\hat{L}_t &= \alpha Y_t + (1 - \alpha) \hat{Y}_t \\ &= \alpha Y_t + (1 - \alpha) [\hat{L}_{t-1} + \hat{b}_{t-1}] \\ &= \alpha Y_t + (1 - \alpha) \hat{L}_{t-1} + (1 - \alpha) \hat{b}_{t-1}\end{aligned}\tag{2.3.23}$$

Similarmente, b_t , la pendiente, se calcula como:

$$\begin{aligned}\hat{b}_t &= \beta(L_t - L_{t-1}) + (1 - \beta)\hat{b}_{t-1} \\ &= \beta(\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta)\hat{b}_{t-1} \\ &= \beta\hat{L}_t - \beta\hat{L}_{t-1} + (1 - \beta)\hat{b}_{t-1}\end{aligned}\tag{2.3.24}$$

Estas ecuaciones se pueden contemplar como fórmulas de actualización mediante las cuales las estimaciones son modificadas a medida que se dispone de nuevas informaciones. Como ya se mencionó anteriormente, existe software que calcula los valores α y β para el método doble exponencial (o para los métodos de suavizamiento en general) que minimizan el error cuadrático medio, es decir, que minimizan la **ec. 2.3.20**. Para inicializar el método podemos considerar $\hat{L}_1 = Y_1$, $\hat{b}_1 = Y_2 - Y_1$ o bien, $\hat{b}_1 = 0$. El pronóstico será $\hat{Y}_{t+1} = \hat{L}_t + \hat{b}_t$. Una vez más tomemos el número de acciones de atención y realicemos en R el análisis correspondiente. Dejamos a R escoger los parámetros óptimos:

```
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> Dobleexponencial <- HoltWinters(Serie, gamma = FALSE)
```

Holt-Winters exponential smoothing with trend and without seasonal component.

Call:

```
HoltWinters(x = Serie, gamma = FALSE)
```

Smoothing parameters:

```
alpha: 0.2816785
beta : 0
gamma: FALSE
```

Coefficients:

```
      [,1]
a 95584.62
b   808.00
```

```
> ValoresAjust3<-fitted(Dobleexponencial)
> plot(c(2007,2013),main="Acciones de atención otorgadas por Conducef")
> lines(Serie,col="black"); lines(fitted(Dobleexponencial)[,1], col = "blue")
> legend(2007,130000,c("AccAtn","Holt"),lty=c(1,1),bty="n",col=c("black","blue"))
```

La **Tabla 2.3.6** muestra los resultados obtenidos en R. Uno de ellos es el valor ajustado, \hat{Y}_t . También se muestran los estimadores para el nivel y la pendiente, \hat{L}_t y \hat{b}_t . La **Figura 2.3.15** muestra el suavizamiento del número de acciones de atención hecho por el método de Holt.

Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t	Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t
Mar	2007	32717	31909	808	Feb	2010	81611	80803	808
Abr	2007	34009	33201	808	Mar	2010	83872	83064	808
May	2007	34358	33550	808	Abr	2010	90936	90128	808
Jun	2007	35937	35129	808	May	2010	90894	90086	808
Jul	2007	36824	36016	808	Jun	2010	89591	88783	808
Ago	2007	38630	37822	808	Jul	2010	91214	90406	808
Sep	2007	39796	38988	808	Ago	2010	93056	92248	808
Oct	2007	40367	39559	808	Sep	2010	95228	94420	808
Nov	2007	43723	42915	808	Oct	2010	93192	92384	808
Dic	2007	43598	42790	808	Nov	2010	94808	94000	808
Ene	2008	38704	37896	808	Dic	2010	96340	95532	808
Feb	2008	41441	40633	808	Ene	2011	91108	90300	808
Mar	2008	43268	42460	808	Feb	2011	89591	88783	808
Abr	2008	43163	42355	808	Mar	2011	90597	89789	808
May	2008	45691	44883	808	Abr	2011	95005	94197	808
Jun	2008	45929	45121	808	May	2011	92829	92021	808
Jul	2008	48530	47722	808	Jun	2011	95107	94299	808
Ago	2008	53197	52389	808	Jul	2011	96908	96100	808
Sep	2008	55865	55057	808	Ago	2011	97427	96619	808
Oct	2008	56867	56059	808	Sep	2011	98957	98149	808
Nov	2008	61523	60715	808	Oct	2011	98093	97285	808
Dic	2008	61814	61006	808	Nov	2011	97602	96794	808
Ene	2009	59889	59081	808	Dic	2011	97510	96702	808
Feb	2009	62366	61558	808	Ene	2012	91721	90913	808
Mar	2009	65645	64837	808	Feb	2012	95806	94998	808
Abr	2009	73288	72480	808	Mar	2012	100795	99987	808
May	2009	75726	74918	808	Abr	2012	105085	104277	808
Jun	2009	76807	75999	808	May	2012	104209	103401	808
Jul	2009	81964	81156	808	Jun	2012	107344	106536	808
Ago	2009	88036	87228	808	Jul	2012	106357	105549	808
Sep	2009	87118	86310	808	Ago	2012	108727	107919	808
Oct	2009	88328	87520	808	Sep	2012	110630	109822	808
Nov	2009	90785	89977	808	Oct	2012	106597	105789	808
Dic	2009	89217	88409	808	Nov	2012	111287	110479	808
Ene	2010	79512	78704	808	Dic	2012	108647	107839	808

Tabla 2.3.6: Tabla que muestra el método de Holt para el número de acciones de atención

Si quisiéramos decirle a R el valor de los parámetros, α y β , tenemos el siguiente resultado:

```
> Dobleexponencial2<-HoltWinters(Serie, alpha=0.8, beta=0.2, gamma = FALSE)
> ValoresAjust4<-fitted(Dobleexponencial2)
> plot(c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black")
> lines(fitted(Dobleexponencial2)[,1], col = "purple")
> legend(2007,130000,c("AccAtn","Holt (0.8,0.2)"),bty="n",col=c("black","purple"))
```

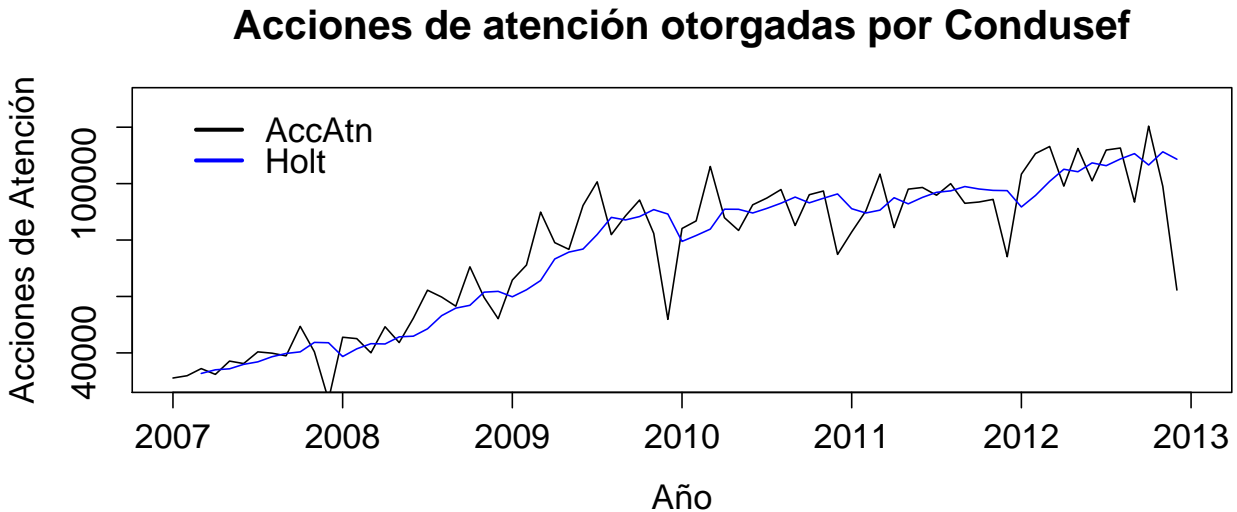


Figura 2.3.15: Figura que muestra el método de Holt para el número de acciones de atención

La **Figura 2.3.16** muestra el método de Holt aplicado al número de acciones de atención con parámetros $\alpha = 0.8$ y $\beta = 0.2$. Los parámetros estimados para el nivel y la pendiente, \hat{L}_t y \hat{b}_t , así como el valor ajustado, \hat{Y}_t , se muestran en la **Tabla 2.3.7**.

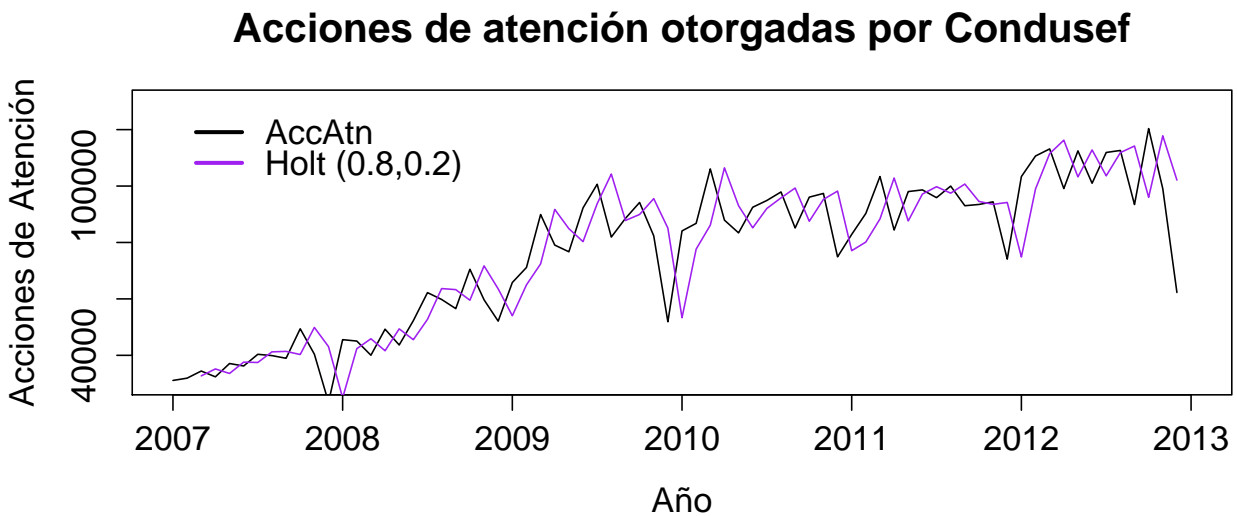


Figura 2.3.16: Método de Holt ($\alpha = 0.8$ y $\beta = 0.2$) para el número de acciones de atención

El código en R con valores iniciales $\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$ está dado por:

```
> Dobleexponencial3<-HoltWinters(Serie, alpha=0.8, beta=0.2, gamma = FALSE,
  l.start=Serie[1],b.start=0); ValoresAjust5<-fitted(Dobleexponencial3)
> plot(c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black"); lines(fitted(Dobleexponencial3)[,1], col = "orange")
> legend(2007,130000,c("AccAtn","Holt valores iniciales"),col=c("black","orange"))
```

La **Tabla 2.3.8** muestra los valores ajustados, los estimadores para la pendiente y el nivel pero ahora con valores iniciales $\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$. La **Figura 2.3.17** muestra dicho ajuste.

Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t	Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t
Mar	2007	32717	31909	808	Feb	2010	77654	77940	-286
Abr	2007	35174	34091	1083	Mar	2010	86121	84948	1173
May	2007	33575	32939	636	Abr	2010	106457	102090	4367
Jun	2007	37591	36392	1199	May	2010	93024	91624	1400
Jul	2007	37470	36491	979	Jun	2010	85184	85324	-140
Ago	2007	41230	39788	1443	Jul	2010	92055	91026	1029
Sep	2007	41398	40168	1230	Ago	2010	95798	94317	1481
Oct	2007	40282	39443	839	Sep	2010	99296	97479	1817
Nov	2007	49886	47586	2300	Oct	2010	87514	87963	-449
Dic	2007	43091	42307	784	Nov	2010	95269	94351	918
Ene	2008	24929	27302	-2374	Dic	2010	98213	96957	1256
Feb	2008	42354	41428	926	Ene	2011	77086	79561	-2475
Mar	2008	45875	44516	1359	Feb	2011	80149	81700	-1552
Abr	2008	41622	41198	423	Mar	2011	88337	88266	72
May	2008	49385	47739	1647	Abr	2011	102845	100367	2478
Jun	2008	45542	44811	732	May	2011	87629	88100	-471
Jul	2008	52756	50944	1812	Jun	2011	97158	95963	1195
Ago	2008	63662	60334	3328	Jul	2011	99769	98338	1431
Sep	2008	63283	60573	2710	Ago	2011	97469	96659	809
Oct	2008	59532	57899	1633	Sep	2011	100700	99487	1213
Nov	2008	71720	68328	3392	Oct	2011	94540	94556	-16
Dic	2008	63561	62094	1467	Nov	2011	93507	93692	-185
Ene	2009	54038	54403	-365	Dic	2011	94188	94229	-41
Feb	2009	64978	63459	1520	Ene	2012	74851	78108	-3257
Mar	2009	72410	69905	2505	Feb	2012	98959	97655	1304
Abr	2009	91718	86412	5305	Mar	2012	111485	108311	3174
May	2009	84884	81602	3282	Abr	2012	116264	112823	3442
Jun	2009	80305	78333	1972	May	2012	103235	102538	697
Jul	2009	93742	89859	3883	Jun	2012	112795	110621	2174
Ago	2009	104258	99269	4988	Jul	2012	103618	103336	282
Sep	2009	87792	86380	1413	Ago	2012	111854	110246	1608
Oct	2009	89928	88394	1533	Sep	2012	114191	112462	1729
Nov	2009	95544	93330	2214	Oct	2012	96002	97593	-1590
Dic	2009	85091	84989	103	Nov	2012	117813	115504	2310
Ene	2010	53325	58533	-5209	Dic	2012	102107	102800	-693

Tabla 2.3.7: Método de Holt ($\alpha = 0.8$ y $\beta = 0.2$) para el número de acciones de atención

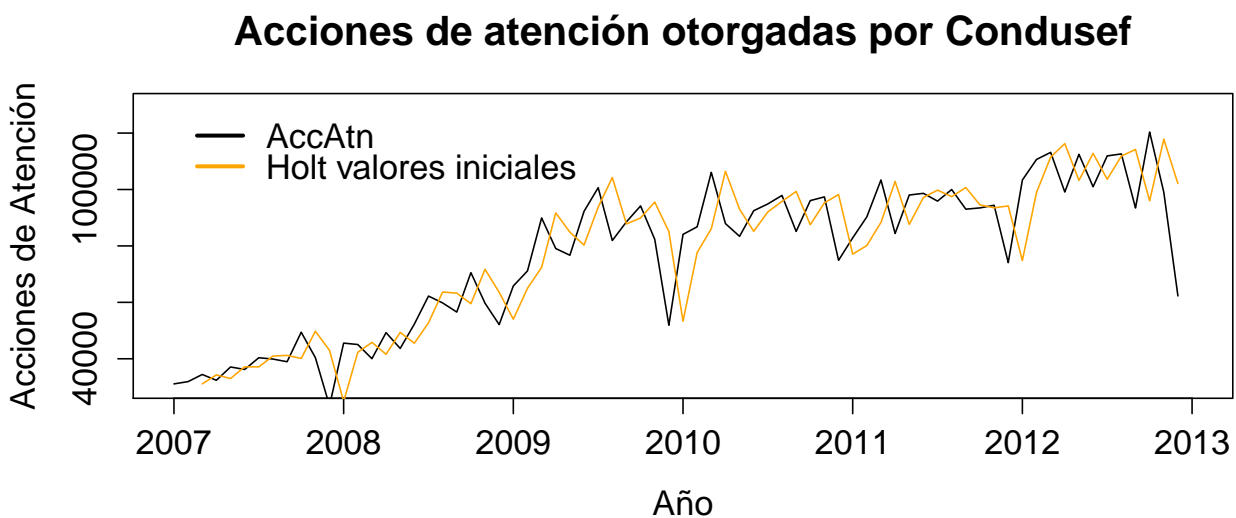


Figura 2.3.17: Método de Holt ($\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$) para el número de acciones de atención

Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t	Mes	Año	\hat{Y}_t	\hat{L}_t	\hat{b}_t
Mar	2007	31101	31101	0	Feb	2010	77654	77940	-286
Abr	2007	34302	33768	533	Mar	2010	86121	84948	1173
May	2007	32990	32764	226	Abr	2010	106456	102090	4367
Jun	2007	37158	36275	883	May	2010	93024	91624	1400
Jul	2007	37137	36404	732	Jun	2010	85184	85324	-140
Ago	2007	40970	39721	1249	Jul	2010	92055	91026	1029
Sep	2007	41194	40116	1078	Ago	2010	95798	94317	1481
Oct	2007	40122	39402	720	Sep	2010	99296	97479	1817
Nov	2007	49760	47554	2206	Oct	2010	87514	87963	-449
Dic	2007	42992	42282	711	Nov	2010	95269	94351	918
Ene	2008	24851	27282	-2431	Dic	2010	98213	96957	1256
Feb	2008	42294	41413	881	Ene	2011	77086	79561	-2475
Mar	2008	45827	44504	1323	Feb	2011	80149	81700	-1552
Abr	2008	41584	41189	395	Mar	2011	88337	88266	72
May	2008	49356	47731	1625	Abr	2011	102845	100367	2478
Jun	2008	45519	44805	715	May	2011	87629	88100	-471
Jul	2008	52738	50939	1798	Jun	2011	97158	95963	1195
Ago	2008	63648	60331	3317	Jul	2011	99769	98338	1431
Sep	2008	63272	60570	2702	Ago	2011	97469	96659	809
Oct	2008	59523	57897	1627	Sep	2011	100700	99487	1213
Nov	2008	71713	68326	3387	Oct	2011	94540	94556	-16
Dic	2008	63555	62092	1463	Nov	2011	93507	93692	-185
Ene	2009	54034	54401	-368	Dic	2011	94188	94229	-41
Feb	2009	64975	63458	1517	Ene	2012	74851	78108	-3257
Mar	2009	72408	69905	2503	Feb	2012	98959	97655	1304
Abr	2009	91716	86412	5304	Mar	2012	111485	108311	3174
May	2009	84883	81602	3281	Abr	2012	116264	112823	3442
Jun	2009	80304	78333	1971	May	2012	103235	102538	697
Jul	2009	93741	89859	3882	Jun	2012	112795	110621	2174
Ago	2009	104257	99269	4988	Jul	2012	103618	103336	282
Sep	2009	87792	86379	1412	Ago	2012	111854	110246	1608
Oct	2009	89927	88394	1533	Sep	2012	114191	112462	1729
Nov	2009	95544	93330	2213	Oct	2012	96002	97593	-1590
Dic	2009	85091	84989	102	Nov	2012	117813	115504	2310
Ene	2010	53324	58533	-5209	Dic	2012	102107	102800	-693

Tabla 2.3.8: Método de Holt ($\hat{L}_1 = Y_1$ y $\hat{b}_1 = 0$) para el número de acciones de atención

2.4. Análisis de Estacionalidad

Definición 2.4.1 *Llamaremos estacionalidad a los movimientos regulares de la serie que tienen una periodicidad inferior al año, es decir, la estacionalidad explica las oscilaciones que año con año se repiten en una serie de tiempo de manera periódica.*

Para no limitarnos a los años, usaremos dos índices, uno para el periodo y otro para la estación. Este hecho se explicará con más detalle en la siguiente subsección. La mayoría de las veces la estacionalidad obedece a factores institucionales o climatológicos.

Por ejemplo, la venta de ropa abrigadora (suéteres, chamarras, bufandas, etc.) en época de verano disminuye de manera sobresaliente en comparación a la vendida en época de invierno. He aquí un ejemplo de estacionalidad que obedece a factores climatológicos. Ahora bien, para la estacionalidad que obedece a factores institucionales podemos dar como ejemplo aquellas series relacionadas con la producción de cualquier bien en un país, ya que interviene el hecho de que las vacaciones laborales se concentran en la época de verano.

Debemos hacer hincapié en que el patrón sistemático que sigue la estacionalidad no tiene por que repetirse cada año. Por ejemplo, la serie que contempla el número de automóviles que transita a lo largo de la Avenida Insurgentes, varía cada hora, por lo que la estacionalidad se repite día con día. Otro ejemplo podría ser la serie que mide la afluencia en el transporte colectivo metro del D.F.

Con la componente estacional, nuestro propósito será eliminar toda aquella oscilación de carácter periódico que se presente en la serie a lo largo del año, para así poder comparar datos pertenecientes a distintas estaciones. Para esto introducimos el **filtro de estacionalidad**, explicado a continuación.

2.4.1. Filtro de estacionalidad

Si la magnitud de la tendencia no tiene cambios abruptos $Y_t = T_t + E_t + \epsilon_t$ donde $\mathbb{E}(\epsilon_t) = 0$, $E_{t+s} = E_t$ y además:

$$\sum_{j=1}^s E_j = 0 \quad (2.4.1)$$

La última condición indica que para cada estación, en total, la parte cíclica es cero. En realidad, usaremos en lugar de t un doble índice, uno indicando el número de estaciones en los datos e indexado por j , $j=1, \dots, e$ (e representa el número de periodos) y el otro el mes de la estación indexado por k , $k=1, \dots, s$ (para el año tenemos $s = 12$). De aquí que:

$$\begin{aligned} Y_{jk} &= T_{jk} + E_{jk} + \epsilon_{jk} \\ \sum_{k=1}^s Y_{jk} &= \sum_{k=1}^s T_{jk} + \sum_{k=1}^s E_{jk} + \sum_{k=1}^s \epsilon_{jk} \end{aligned}$$

Si suponemos que en promedio por estación, el error es cero, y además se cumple la **ec. 2.4.1**, es decir, para cada estación la parte cíclica es cero, por tanto:

$$\frac{1}{s} \sum_{k=1}^s Y_{jk} = \frac{1}{s} \sum_{k=1}^s T_{jk} \quad (2.4.2)$$

Como estamos considerando que no hay cambios abruptos en la tendencia entonces $T_{jk} = T_j$, $\forall j = 1, \dots, e$, de aquí que un estimador para T_j puede ser:

$$\hat{T}_j = \frac{1}{s} \sum_{k=1}^s Y_{jk}, \quad j = 1, \dots, e \quad (2.4.3)$$

Y para estimar la estacionalidad, tenemos:

$$\hat{E}_k = \frac{1}{e} \sum_{j=1}^e (Y_{jk} - \hat{T}_j), \quad k = 1, \dots, s \quad (2.4.4)$$

Ya que la estacionalidad es la misma para todos los periodos. Además:

$$\sum_{k=1}^s \hat{E}_k = \frac{1}{e} \sum_{k=1}^s \sum_{j=1}^e (Y_{jk} - \hat{T}_j) = \frac{1}{e} \sum_{j=1}^e \left(\sum_{k=1}^s Y_{jk} - s\hat{T}_j \right) = 0 \quad (2.4.5)$$

2.4.2. Método de Holt-Winters

El método de Holt-Winters es una extensión del método de Holt, utilizando ahora como esquema teórico de referencia un modelo que incluye la estacionalidad. Las series susceptibles de ser tratadas mediante la técnica de Holt-Winters son, por tanto, aquellas que contienen una componente de tendencia aproximadamente lineal y otra de carácter estacional. Este método tiene 2 versiones:

Esquema aditivo

$Y_{t+1} = L_t + b_t + C_{t-s+1}$ donde

$$\begin{aligned} L_t &= \text{Nivel al tiempo } t \\ b_t &= \text{Tendencia al tiempo } t \\ C_{t-s+1} &= \text{Estacionalidad al tiempo } t-s+1 \end{aligned} \quad (2.4.6)$$

Lo anterior significa que si, por ejemplo, $s = 12$ y $t = 13$ entonces $t - s + 1 = 2$, así que la estimación de la serie en el periodo $t + 1 = 14$ depende de la componente del periodo 2, C_2 , que es el mes que le corresponde al periodo 14.

Bajo este método:

$$\begin{aligned} \hat{L}_t &= \alpha(Y_t - \hat{C}_{t-s}) + (1 - \alpha)(\hat{L}_{t-1} + \hat{b}_{t-1}) \\ \hat{b}_t &= \beta(\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta)\hat{b}_{t-1} \\ \hat{C}_t &= \gamma(Y_t - \hat{L}_t) + (1 - \gamma)\hat{C}_{t-s} \end{aligned} \quad (2.4.7)$$

con $0 < \alpha, \beta, \gamma < 1$. Los valores iniciales de los que se puede partir son:

$$\hat{L}_0 = \frac{1}{s} \sum_{t=1}^s Y_t; \quad \hat{b}_0 = \frac{1}{s} \left[\frac{1}{s} \sum_{t=1}^s Y_t - \frac{1}{s} \sum_{t=s+1}^{2s} Y_t \right] \quad y \quad \hat{C}_0 = Y_k - \hat{L}_0 + \frac{(k-1)\hat{b}_0}{2} \quad (2.4.8)$$

donde L_0 es el promedio de la primer temporada, y existen s valores de \hat{C}_0 para $k = 1, \dots, s$. Cabe recalcar que también se puede usar $\hat{b}_0 = 0$.

El código en R para el esquema aditivo para el número de acciones de atención otorgadas por Condusef está dado por:

```
> library(stats)
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 2/Condusef")
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> HWaditivo <- HoltWinters(Serie, seasonal = "additive")
```

Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:

```
HoltWinters(x = Serie, seasonal = "additive")
```

Smoothing parameters:

```
alpha: 0.3127162
beta : 0.06479356
gamma: 0.5342237
```

```
> plot (c(2007,2013),main="Acciones de atención otorgadas por Condusef")
> lines(Serie,col="black")
> lines(fitted(HWaditivo)[,1], col = "red")
> legend(2007,130000,c("AccAtn","HW aditivo"),lty=c(1,1),bty="n",
  col=c("black","red"))
```

La **Figura 2.4.1** muestra el suavizamiento realizado por el método de Holt-Winters en el número de acciones de atención otorgadas por Condusef.

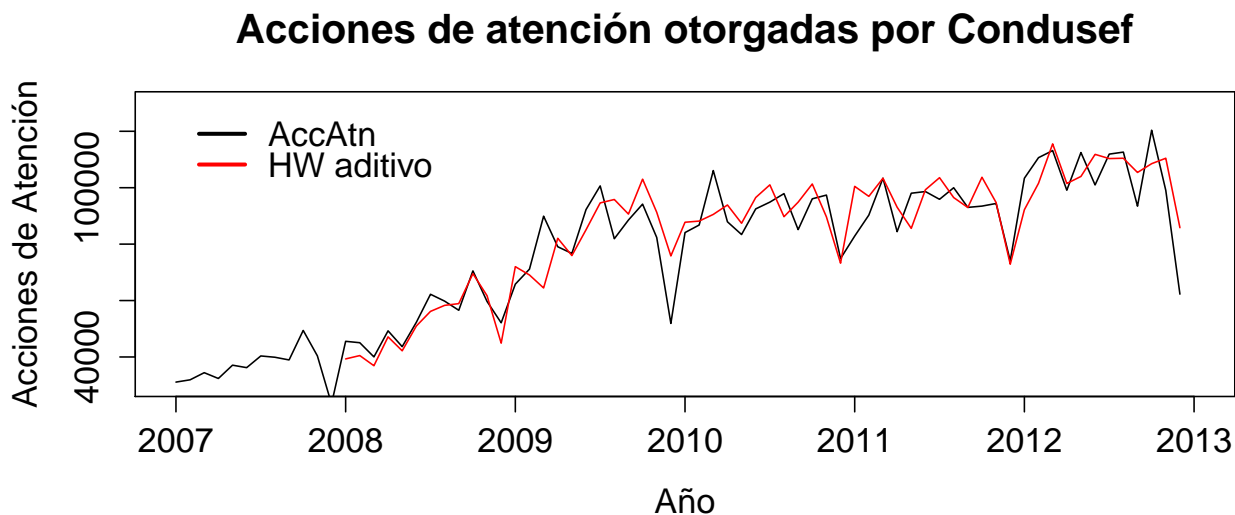


Figura 2.4.1: Método Holt-Winters aditivo para el número de acciones de atención

Una de nuestras motivaciones es y será siempre, la predicción de observaciones futuras. En particular estaremos interesados en conocer el pronóstico del número de asesorías que otorgará Condusef a los usuarios de servicios financieros del Sistema Financiero Mexicano, para así tener la debida preparación ante semejante suceso y tomar las decisiones correspondientes que ayudarán y beneficiarán en un futuro a la sociedad mexicana.

De aquí que, nuestra predicción en código R para el número de acciones de atención que otorgará Condusef para el año 2013 será:

```
> Predic<- predict(HWaditivo, 12, prediction.interval = TRUE)
```

	fit	upr	lwr
Jan 2013	100681.86	117998.0	83365.74
Feb 2013	105204.82	123455.7	86953.97
Mar 2013	111353.94	130602.5	92105.35
Apr 2013	97315.56	117621.7	77009.45
May 2013	103867.16	125287.5	82446.85
Jun 2013	101096.53	123684.8	78508.28
Jul 2013	106919.44	130726.6	83112.25
Aug 2013	106096.81	131171.4	81022.20
Sep 2013	94476.54	120864.7	68088.38
Oct 2013	109608.90	137354.7	81863.13
Nov 2013	98578.74	127724.2	69433.27
Dec 2013	72583.34	103168.9	41997.81

```
> plot(HWaditivo,Predic,main="Predicción para el número de acciones de atención")
> legend(2008,135000,c("AccAtn","Predicción"),lty=c(1,1),bty="n",
col=c("black","red"))
```

La **Figura 2.4.2** muestra las predicciones para el año 2013 del número de acciones de atención que otorgará Condusef, donde las líneas rojas representan la predicción mientras que las azules representan las bandas de confianza.

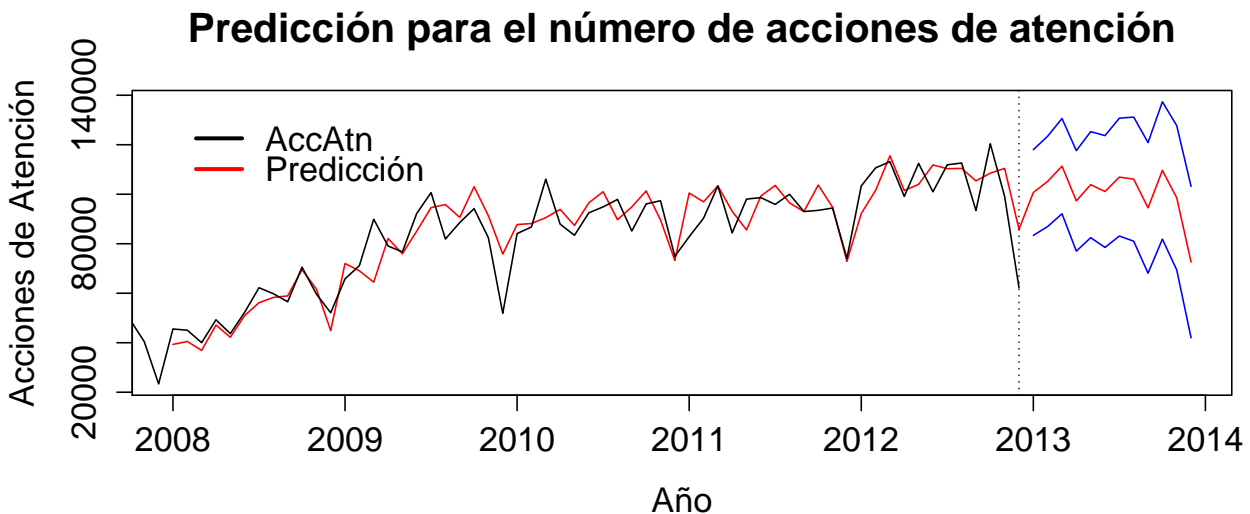


Figura 2.4.2: Figura que muestra las predicciones para el número de acciones de atención

En ITSM, tenemos disponible el método de Holt-Winters pero solamente bajo el esquema aditivo. Tomemos la serie del número de pasajeros mensuales de cierta aerolínea internacional desde el mes de Enero de 1949 hasta el mes de Diciembre de 1960. Seleccionamos ‘File >Project >Open >Univariate’ y elegimos el archivo ‘Airpass.tsm’. La gráfica correspondiente se muestra en la **Figura 2.4.3**.

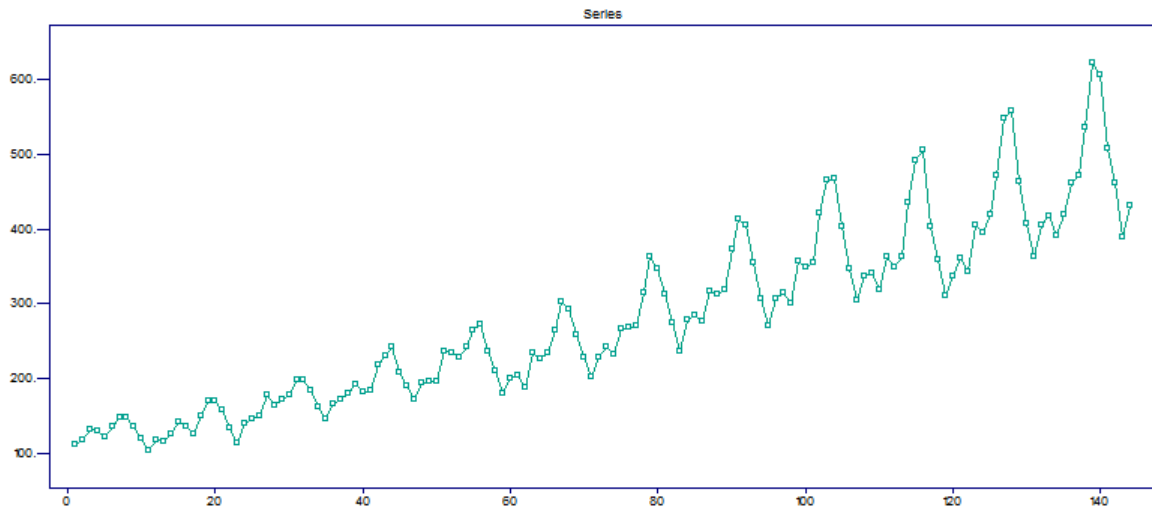


Figura 2.4.3: Figura que muestra el número de pasajeros mensuales

Ahora, nos vamos a “Forecasting >Seasonal Holt-Winters” y seleccionamos como 12 el número de predicciones, es decir, queremos predecir el siguiente año completo. Dejamos que optimice los valores de α , β y γ , obteniendo el resultado mostrado en la **Figura 2.4.4**.

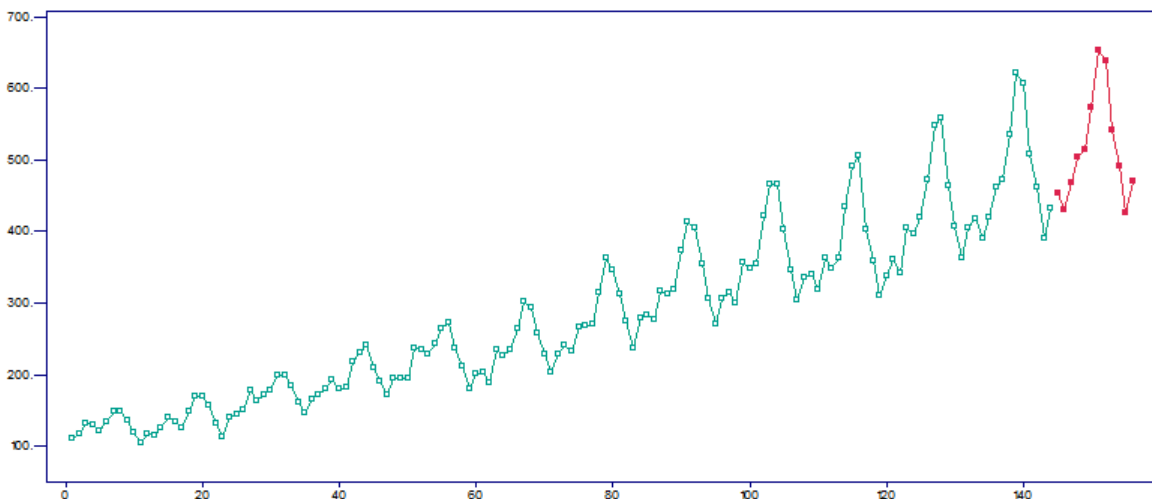


Figura 2.4.4: Figura que muestra el número de pasajeros mensuales bajo Holt-Winters

Desafortunadamente, ITSM, en este método, no nos muestra las bandas de confianza, por lo que, para utilizar el método de Holt-Winters es mejor utilizar R. Aún así se muestra el procedimiento disponible en ITSM.

Esquema multiplicativo

$Y_{t+1} = (L_t + b_t) * C_{t-s+1}$ donde nuevamente

$$\begin{aligned}
 L_t &= \text{Nivel al tiempo } t \\
 b_t &= \text{Tendencia al tiempo } t \\
 C_{t-s+1} &= \text{Estacionalidad al tiempo } t-s+1
 \end{aligned}
 \tag{2.4.9}$$

Bajo este método:

$$\begin{aligned}\hat{L}_t &= \alpha \frac{Y_t}{\hat{C}_{t-s}} + (1 - \alpha)(\hat{L}_{t-1} + \hat{b}_{t-1}) \\ \hat{b}_t &= \beta(\hat{L}_t - \hat{L}_{t-1}) + (1 - \beta)\hat{b}_{t-1} \\ \hat{C}_t &= \gamma \frac{Y_t}{\hat{L}_t} + (1 - \gamma)\hat{C}_{t-s}\end{aligned}\quad (2.4.10)$$

con $0 < \alpha, \beta, \gamma < 1$. Los valores iniciales con los que se puede iniciar son:

$$\hat{L}_0 = \frac{1}{s} \sum_{t=1}^s Y_t; \quad \hat{b}_0 = \frac{1}{s} \left[\frac{1}{s} \sum_{t=1}^s Y_t - \frac{1}{s} \sum_{t=s+1}^{2s} Y_t \right] \quad \text{y} \quad \hat{C}_0 = \frac{1}{\hat{L}_0} \left[Y_k - \frac{(k-1)\hat{b}_0}{2} \right] \quad (2.4.11)$$

El código en R para el esquema multiplicativo, junto con la predicción correspondiente será el siguiente:

```
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> Hwmultiplicativo<-HoltWinters(Serie, seasonal = "multiplicative")
```

Holt-Winters exponential smoothing with trend and multiplicative seasonal component.

Call:

```
HoltWinters(x = Serie, seasonal = "multiplicative")
```

Smoothing parameters:

```
alpha: 0.2215445
beta : 0.09581145
gamma: 0.615166
```

```
> plot(c(2007,2013),main="Acciones de atención otorgadas por Conducef")
> lines(Serie,col="black"); lines(fitted(Hwmultiplicativo)[,1], col = "green")
> legend(2007,130000,c("AccAtn","HW multiplicativo"),bty="n",
  col=c("black","green"))
> Predic2<-predict(Hwmultiplicativo, 12, prediction.interval = TRUE)
```

	fit	upr	lwr
Jan 2013	101067.27	114220.60	87913.95
Feb 2013	105879.60	119872.34	91886.85
Mar 2013	112390.10	127447.78	97332.42
Apr 2013	96889.46	112227.68	81551.24
May 2013	104760.73	121554.24	87967.22
Jun 2013	100684.30	118257.74	83110.86
Jul 2013	107285.65	126628.94	87942.35
Aug 2013	107282.04	127877.91	86686.17
Sep 2013	93938.07	114146.10	73730.03
Oct 2013	110870.79	134821.81	86919.78
Nov 2013	98541.51	121926.99	75156.02
Dec 2013	68154.17	82680.93	53627.42

La **Figura 2.4.5** muestra el suavizamiento del número de acciones de atención por el método de Holt-Winters.

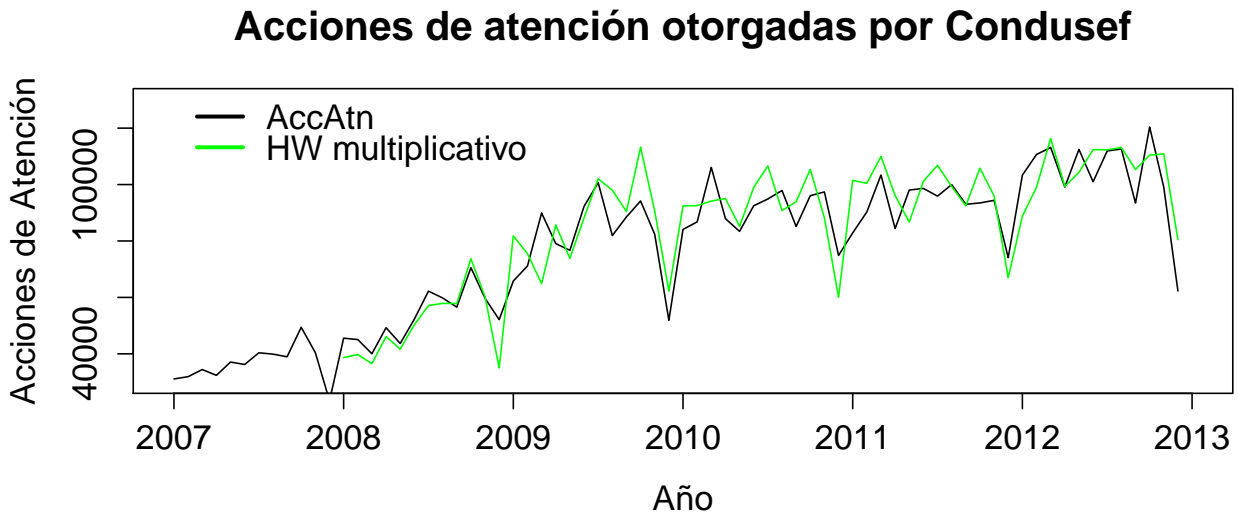


Figura 2.4.5: Método de Holt-Winters multiplicativo para el número de acciones de atención

2.5. Descomposición clásica

Hemos visto a lo largo de este capítulo, los distintos métodos para la obtención de cada una de las componentes en que se presume dividida una serie de tiempo. A continuación realizaremos un ejemplo que tome en cuenta la descomposición clásica para el esquema aditivo. La descomposición clásica para el esquema multiplicativo dista poco de aquel que utiliza el esquema aditivo. La **Tabla 2.5.1** muestra las ventas trimestrales (millones de pesos) de ropa deportiva de cierta empresa mexicana desde el año 2005 y hasta el año 2008. La **Figura 2.5.1** muestra dicha serie.

Año	Tiempo	Trimestre	Ventas $\{X_t\}$
2005	1	1	516.0
2005	2	2	546.8
2005	3	3	561.9
2005	4	4	565.7
2006	5	1	545.9
2006	6	2	571.3
2006	7	3	586.6
2006	8	4	584.2
2007	9	1	549.2
2007	10	2	583.2
2007	11	3	589.6
2007	12	4	584.3
2008	13	1	576.5
2008	14	2	607.0
2008	15	3	616.3
2008	16	4	610.8

Tabla 2.5.1: Tabla que muestra las ventas trimestrales de cierta empresa mexicana

Número de ventas en ropa deportiva

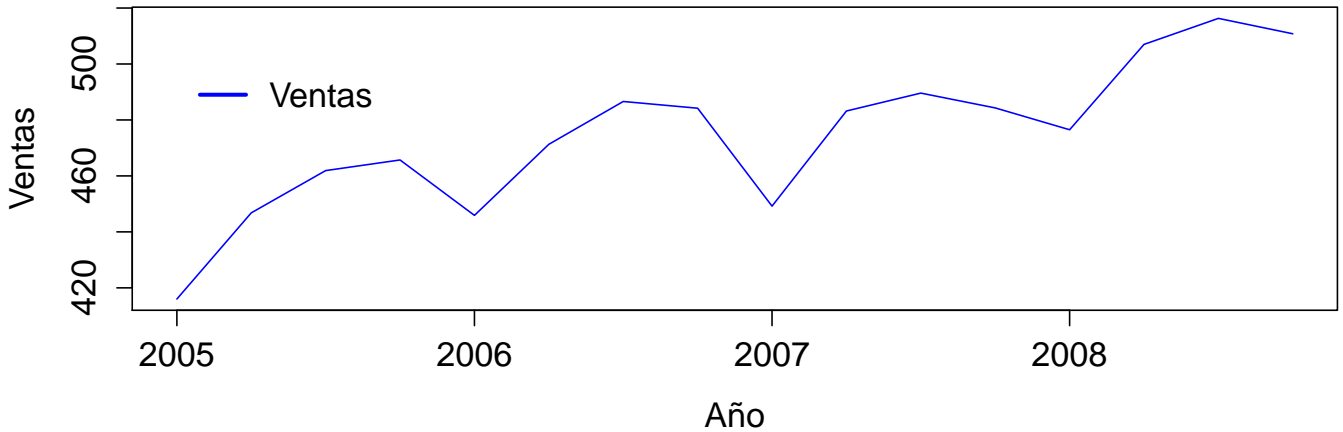


Figura 2.5.1: Figura que muestra las ventas trimestrales de cierta empresa mexicana

Claramente, la serie posee tanto tendencia como estacionalidad. Observamos una caída cada primer trimestre y un pico cada tercer trimestre. Vamos a tomar un promedio móvil de tamaño H , donde H es el número de estaciones en un año, es decir, en este ejemplo $H = 4$. Si H es un número par, debemos obtener un promedio que corresponda a los periodos de tiempo de los datos originales. Para lograr esto, un nuevo promedio móvil que abarque dos periodos y el anterior suavizamiento por promedios móviles debe realizarse. Por ejemplo, tenemos el primer promedio móvil dado por:

$$\begin{aligned} \text{Primer}_{MA} &= \frac{516 + 546.8 + 561.9 + 565.7}{4} = 547.60 \\ \text{Segundo}_{MA} &= \frac{546.8 + 561.9 + 565.7 + 545.9}{4} = 555.08 \\ \text{Tercer}_{MA} &= \frac{561.9 + 565.7 + 545.9 + 571.3}{4} = 561.20 \\ \text{Cuarto}_{MA} &= \frac{565.7 + 545.9 + 571.3 + 586.6}{4} = 567.38 \end{aligned}$$

y así sucesivamente. Pasamos ahora a calcular el nuevo promedio móvil centrado:

$$\begin{aligned} \text{Centrado}_{MA_3} &= \frac{547.60 + 555.08}{2} = 551.34 \\ \text{Centrado}_{MA_4} &= \frac{555.08 + 561.20}{2} = 558.14 \\ \text{Centrado}_{MA_5} &= \frac{561.20 + 567.38}{2} = 564.29 \end{aligned}$$

y así sucesivamente. Restamos el promedio móvil centrado de los datos originales, denotando a esta nueva serie como $\{E_t + \epsilon_t\}$; por ejemplo, tenemos:

$$\begin{aligned} E_3 + \epsilon_3 &= 561.90 - 551.34 = 10.56 \\ E_4 + \epsilon_4 &= 565.70 - 558.14 = 7.56 \end{aligned}$$

y así sucesivamente. Eliminamos la componente aleatoria $\{\epsilon_t\}$, calculando el promedio de cada una de las 4 estaciones. La **Tabla 2.5.2** muestra dicho cálculo.

1° Trimestre	2° Trimestre	3° Trimestre	4° Trimestre
-18.39	1.61	10.56	7.56
-26.98	6.64	14.19	9.89
-16.19	7.66	9.61	-2.08
-61.55	15.91	34.36	15.38
$\frac{-18.39-26.98-16.19}{3} = -20.52$	$\frac{1.61+6.64+7.66}{3} = 5.30$	$\frac{10.56+14.19+9.61}{3} = 11.45$	$\frac{7.56+9.89-2.08}{3} = 5.13$

Tabla 2.5.2: Tabla que muestra el promedio de las estaciones (trimestres)

Estos promedios estacionales deben sumar cero, de lo contrario debemos normalizarlos para que al final, al restarlos, éstos consistan en un valor contante. Usando los datos de la **Tabla 2.5.2** calculamos:

$$L = \frac{-20.52+5.30+11.45+5.13}{4} = 0.34$$

Los estimadores estacionales serán finalmente:

$$E_1 = -20.52-0.34 = -20.86$$

$$E_2 = 5.30-0.34 = 4.96$$

$$E_3 = 11.45-0.34 = 11.11$$

$$E_4 = 5.13-0.34 = 4.79$$

Pasamos a eliminar la estacionalidad de los datos restando los estimadores estacionales de la serie original y así obtener una nueva serie $\{X_t - E_t\}$, la cual solamente debería poseer tendencia (al menos a la vista, la serie carece de estacionalidad; en la siguiente sección veremos una prueba formal para demostrar lo anterior) y ésta se puede estimar mediante el análisis de regresión lineal. La **Tabla 2.5.3** muestra todo los cálculos realizados para este ejemplo. La **Figura 2.5.2** muestra la nueva serie.

Año	Tiempo	Trimestre	Ventas $\{X_t\}$	MA	C_{MA}	$\{E_t + \epsilon_t\}$	$\{E_t\}$	$\{X_t - E_t\}$
2005	1	1	516.0				-20.86	536.86
2005	2	2	546.8				4.96	541.84
2005	3	3	561.9	547.60	551.34	10.56	11.11	550.79
2005	4	4	565.7	555.08	558.14	7.56	4.78	560.92
2006	5	1	545.9	561.20	564.29	-18.39	-20.86	566.76
2006	6	2	571.3	567.38	569.69	1.61	4.96	566.34
2006	7	3	586.6	572.00	572.41	14.19	11.11	575.49
2006	8	4	584.2	572.83	574.31	9.89	4.78	579.42
2007	9	1	549.2	575.80	576.18	-26.98	-20.86	570.06
2007	10	2	583.2	576.55	576.56	6.64	4.96	578.24
2007	11	3	589.6	576.58	579.99	9.61	11.11	578.49
2007	12	4	584.3	583.40	586.38	-2.08	4.78	579.52
2008	13	1	576.5	589.35	592.69	-16.19	-20.86	597.36
2008	14	2	607.0	596.03	599.34	7.66	4.96	602.04
2008	15	3	616.3	602.65			11.11	605.19
2008	16	4	610.8				4.78	606.02

Tabla 2.5.3: Tabla que muestra los cálculos de ventas trimestrales de la empresa mexicana

Número de ventas en ropa deportiva

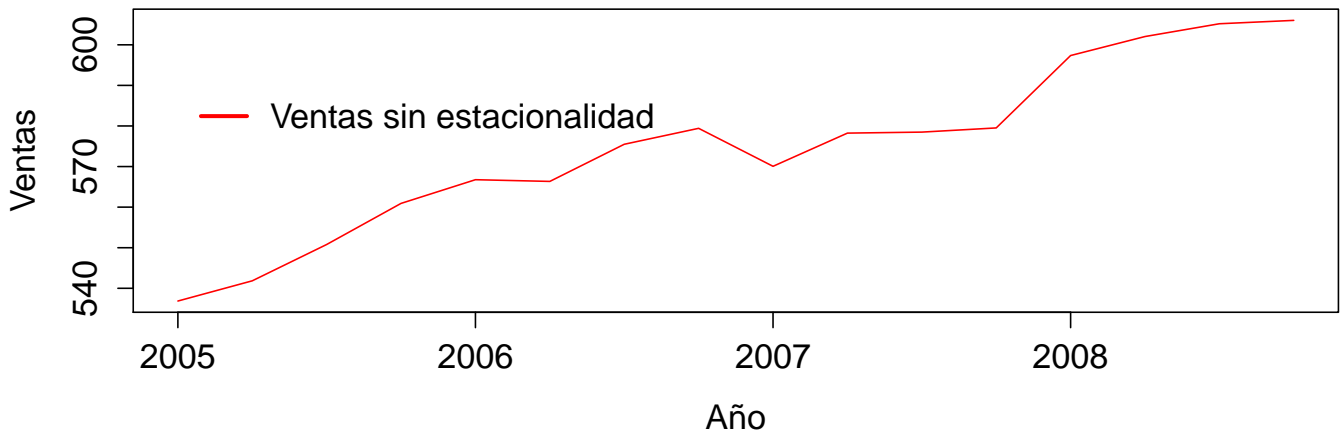


Figura 2.5.2: Figura que muestra las ventas trimestrales sin la presencia de estacionalidad

El modelo apropiado para la serie que carece de estacionalidad es un modelo de tendencia lineal con ecuación $T_t = 538.4385 + 4.2671t$. El código correspondiente a la estimación del modelo de tendencia lineal en R está dado por:

```
> Ventasreg<-read.table('Ventas.txt',header=T); attach(Ventasreg)
> Ajuste<-lm(Ventasmex~t); summary(Ajuste)
```

Call:

```
lm(formula = Ventasmex ~ t)
```

Residuals:

Min	1Q	Median	3Q	Max
-10.1236	-5.3109	0.9246	4.2500	7.1819

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	538.4385	3.0754	175.08	< 2e-16 ***
t	4.2671	0.3181	13.42	2.21e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.865 on 14 degrees of freedom

Multiple R-squared: 0.9278, Adjusted R-squared: 0.9227

F-statistic: 180 on 1 and 14 DF, p-value: 2.206e-09

La descomposición clásica con estacionalidad vista en este capítulo puede obtenerse en R con la función llamada `stl` que hace una descomposición que considera estacionalidad. A continuación mostramos el código para el número de acciones de atención:

```
> Acciones<-read.table('ACCAT.txt',header=T)
> Serie<-ts(Acciones, frequency = 12, start = c(2007, 1))
> stl(Serie[,1],s.window="periodic")
```

La **Tabla 2.5.4** muestra la estimación de las componentes en que se supone dividida una serie de tiempo con el comando `stl` en R.

Mes	Año	Estacionalidad	Tendencia	Residuos	Mes	Año	Estacionalidad	Tendencia	Residuos
Ene	2007	-2449	29808	3742	Ene	2010	-2449	86644	-101
Feb	2007	251	30934	724	Feb	2010	251	86971	-450
Mar	2007	7639	32060	-5264	Mar	2010	7639	87297	11146
Abr	2007	-2290	33196	1474	Abr	2010	-2290	87937	2269
May	2007	122	34333	2642	May	2010	122	88578	-5300
Jun	2007	3028	35461	-2273	Jun	2010	3028	89268	192
Jul	2007	7865	36590	-4088	Jul	2010	7865	89957	-2940
Ago	2007	4788	37699	-2585	Ago	2010	4788	90287	2824
Sep	2007	-2068	38807	2215	Sep	2010	-2068	90617	-3419
Oct	2007	8413	39918	1081	Oct	2010	8413	90890	-3244
Nov	2007	-963	41029	347	Nov	2010	-963	91163	7179
Dic	2007	-24335	42224	5466	Dic	2010	-24335	91584	7649
Ene	2008	-2449	43419	4583	Ene	2011	-2449	92005	-6702
Feb	2008	251	44869	-63	Feb	2011	251	92246	-2202
Mar	2008	7639	46318	-13928	Mar	2011	7639	92488	3249
Abr	2008	-2290	47945	3613	Abr	2011	-2290	92555	-5850
May	2008	122	49571	-6026	May	2011	122	92622	5303
Jun	2008	3028	51648	-2381	Jun	2011	3028	93047	2558
Jul	2008	7865	53724	640	Jul	2011	7865	93473	-5456
Ago	2008	4788	56401	-1388	Ago	2011	4788	94544	659
Sep	2008	-2068	59077	-456	Sep	2011	-2068	95614	-526
Oct	2008	8413	62083	31	Oct	2011	8413	96822	-11755
Nov	2008	-963	65089	-4439	Nov	2011	-963	98029	-2657
Dic	2008	-24335	68052	8396	Dic	2011	-24335	99143	-720
Ene	2009	-2449	71016	-2752	Ene	2012	-2449	100257	5548
Feb	2009	251	73634	-2748	Feb	2012	251	101308	9090
Mar	2009	7639	76252	6022	Mar	2012	7639	102358	3160
Abr	2009	-2290	78231	3132	Abr	2012	-2290	102456	-1059
May	2009	122	80210	-3637	May	2012	122	102554	9792
Jun	2009	3028	81446	7774	Jun	2012	3028	102322	-4379
Jul	2009	7865	82683	10103	Jul	2012	7865	102091	1947
Ago	2009	4788	83643	-6521	Ago	2012	4788	101730	6096
Sep	2009	-2068	84603	6010	Sep	2012	-2068	101369	-5858
Oct	2009	8413	85264	504	Oct	2012	8413	100896	11069
Nov	2009	-963	85925	-2611	Nov	2012	-963	100424	-415
Dic	2009	-24335	86284	-10056	Dic	2012	-24335	99857	-13248

Tabla 2.5.4: Tabla que muestra la descomposición clásica mediante el uso del comando `stl`

2.6. Análisis de la componente irregular

Definición 2.6.1 *La componente irregular es aquella variable que no responde a comportamientos sistemáticos o regulares. Dicha componente deberá ser incluida siempre que trabajemos con series estocásticas.*

Para la identificación y estimación de la componente irregular, primero identificamos cada una de las componentes restantes que explican a una serie de tiempo, a saber, tendencia y estacionalidad para después eliminarlas (su eliminación dependerá del esquema que se haya elegido para modelar la serie) y quedarse así con la estimación de la componente irregular.

2.7. Prueba de hipótesis para estacionalidad

Un problema común en estadística es decidir si un conjunto de muestras provienen de la misma población. Casi siempre las muestras difieren pero nuestra pregunta consistirá en ver si estas diferencias son significativas o bien simplemente se deban a variaciones esperadas entre muestras aleatorias de la misma población.

La técnica usual para resolver estos problemas es el análisis de varianza con un solo criterio de clasificación. La variación entre las medias muestrales, \bar{x}_i , es usada para estimar la variación entre individuos, asumiendo que las varianzas muestrales reflejan solamente el muestreo aleatorio de una población en la cual los individuos varían y en el hecho de que las varianzas muestrales de una muestra aleatoria de tamaño n_i es $\frac{\sigma^2}{n_i}$ donde σ^2 es la varianza poblacional. Esta estimación de σ^2 basada en la variación entre medias muestrales es entonces comparada con otra estimación basada solamente en la variación entre muestras.

A lo largo de este capítulo hemos confirmado la presencia de estacionalidad al inspeccionar la gráfica de unos datos o bien con conocimiento previo del comportamiento de una serie. A pesar de esto, algunas veces, la presencia de estacionalidad puede no parecer tan obvia. Es en estos casos que necesitamos algo más que una ayuda visual.

Un método que podemos tomar a consideración es el análisis de varianza con un solo criterio de clasificación dado por la prueba Kruskal-Wallis, sin empates, aplicada a la serie $\{E_t + \epsilon_t\}$ vista en la sección anterior. Esta serie se supone que contiene solamente la componente estacional y la componente aleatoria. Si no existe algún componente estacional específico, la serie $\{E_t + \epsilon_t\}$ deberá consistir solamente de errores aleatorios y por tanto, su distribución deberá ser la misma para todas las estaciones.

Esto significa que si esta serie es ordenada y las ordenaciones son agrupadas por estación, entonces el rango promedio para cada estación debería ser estadísticamente igual al rango promedio de cualquier otra estación. La prueba Kruskal-Wallis, determina si la suma de los rangos (y por tanto de las medianas) son distintos o no, entre varias grupos (estaciones).

Retomemos el ejemplo del número de ventas de ropa deportiva de cierta empresa mexicana visto en la sección anterior. Tomemos los datos que necesitamos para aplicar la prueba Kruskal-Wallis. La **Tabla 2.7.1** muestra la serie $\{E_t + \epsilon_t\}$ así como sus correspondientes rangos. La **Tabla 2.7.2** muestra la agrupación por estación (trimestre) así como la suma de sus respectivos rangos.

<i>Tiempo</i>	<i>Trimestre</i>	$\{E_t + \epsilon_t\}$	<i>Rango</i>
3	3	10.56	11
4	4	7.56	7
5	1	-18.39	2
6	2	1.61	5
7	3	14.19	12
8	4	9.89	10
9	1	-26.98	1
10	2	6.64	6
11	3	9.61	9
12	4	-2.08	4
13	1	-16.19	3
14	2	7.66	8

Tabla 2.7.1: Tabla que muestra los cálculos para la prueba Kruskal-Wallis

Suma rangos			
1	2	3	4
2	5	11	7
1	6	12	10
3	8	9	4
6	19	32	21

Tabla 2.7.2: Tabla que muestra la suma de los rangos para la prueba Kruskal-Wallis

La prueba Kruskal-Wallis tiene hipótesis nula y alternativa:

$$H_0 : E_1 = E_2 = E_3 = E_4 = 0 \quad vs \quad H_1 : E_n \neq 0 \quad (\text{para algunas estaciones}) \quad (2.7.1)$$

o bien, en palabras:

$$H_0 : \text{No existe estacionalidad} \quad vs \quad H_1 : \text{Existe estacionalidad en los datos} \quad (2.7.2)$$

con estadístico:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^C \frac{R_i^2}{n_i} - 3(N+1) \quad (2.7.3)$$

donde

- N = Número total de rangos
- R_i = Suma de los rangos de una estación específica
- n_i = Número de rangos en una estación específica
- C = Número de muestras

La distribución de H es ji-cuadrada con $C - 1$ grados de libertad, rechazando la hipótesis nula H_0 si $H > \chi_{\alpha, C-1}^2$, el cuantil de una distribución ji-cuadrada con $C - 1$ grados de libertad. Para nuestro ejemplo:

$$H = \frac{12}{12(13)} \left[\frac{6^2}{3} + \frac{19^2}{3} + \frac{32^2}{3} + \frac{21^2}{3} \right] - (3)(13) = 8.74 \quad (2.7.4)$$

y como $H = 8.74 > 7.81 = \chi_{0.05, 3}^2$, rechazamos H_0 y por tanto aún existe estacionalidad en la serie $\{E_t + \epsilon_t\}$. Es hasta este punto donde podemos concluir que la estacionalidad del ejemplo de las ventas trimestrales de ropa deportiva no se logró eliminar por completo. Algún otro método se deberá utilizar (algunos de ellos se han visto en este capítulo y otros se verán en el siguiente capítulo) para eliminar por completo la estacionalidad.

La prueba Kruskal-Wallis se realizó en la descomposición clásica del esquema aditivo pero igualmente se puede aplicar a cualquier otro método que involucre la estimación y eliminación de las componentes de una serie de tiempo. Lo que haremos será estimar la estacionalidad de la serie en cuestión, $\{E_t\}$, para después obtener la serie $\{X_t - E_t\}$ y entonces aplicar la prueba Kruskal-Wallis a una de estas dos series, dependiendo del método que se haya utilizado para estimar la estacionalidad.

Capítulo 3

Modelos para Series de Tiempo Univariadas.

3.1. Introducción

El impacto de una serie de tiempo es de tal magnitud, que suele abarcar un sinnúmero de campos científicos. Por ejemplo, muchos modelos económicos pueden ser reproducidos mediante el uso de una serie de tiempo ya que constantemente se trabaja con variables económicas dadas con periodicidad diaria, mensual, trimestral, semestral, anual, etc. Dentro del campo social, demógrafos trabajan con series de tiempo poblacionales que involucran variables como el nacimiento, defunción, índice de alistamiento educacional, índice de desempleo, etc, de cierta cohorte. En medicina, un modelo de series de tiempo puede ser utilizado para dar seguimiento al nivel de presión sanguínea de un conjunto de pacientes que sufran hipertensión. Resonancias magnéticas de ondas cerebrales pueden ser tratadas como una serie de tiempo, con el objetivo de estudiar la reacción del cerebro ante la presencia de ciertos estímulos.

En este capítulo dejaremos de lado la vinculación de una serie de tiempo con el modelo clásico y pasaremos a identificarla con un proceso estocástico. Consideremos $\{X(t) : t \in T\}$ un conjunto de variables aleatorias definidas en el mismo conjunto (Ω, F, \mathbb{P}) donde $T \neq \emptyset$ es un conjunto de índices. Estudiaremos T a lo más numerable. Usaremos además como notación $X(t) = X_t$.

Como ya fue mencionado en el **capítulo 2**, el enfoque estocástico surge con los trabajos de Box & Jenkins y la metodología ARIMA, basada en la teoría de los procesos estocásticos. Estaremos interesados en aquellos procesos que resulten ser estacionarios (en el sentido débil). Intuitivamente, una serie de tiempo es estacionaria si sus propiedades estadísticas se mantienen constantes a lo largo del tiempo.

Entonces, para aplicar la metodología de Box & Jenkins debemos primero, determinar si la serie de tiempo en cuestión es estacionaria o no. Si no lo es, debemos transformarla en una serie de tiempo que sí lo sea. A partir de la estacionariedad de una serie de tiempo, nuestro problema de interés en este capítulo será tanto el de replicar la realidad como el de pronosticar o estimar valores futuros del proceso $\{X_t\}$.

Comenzaremos con una definición formal de estacionariedad en una serie de tiempo para después desarrollar de manera precisa la metodología de Box & Jenkins. Además, dentro de este capítulo, al igual que en los capítulos anteriores, introduciremos paquetes estadísticos que nos ayudarán a alcanzar nuestro cometido, a saber, R e ITSM.

3.2. Procesos estacionarios

Definición 3.2.1 $\{X_t\}_{t \in T}$ con T , el conjunto de índices, a lo más numerable, es un proceso estocástico de 2º orden si $\mathbb{E}(X_t^2) < \infty$, $\forall t \in T$.

De aquí en adelante denotaremos a $\{X_t\}_{t \in T}$ simplemente como $\{X_t\}$, recordando siempre que el conjunto de índices es a lo más numerable.

Definición 3.2.2 Un proceso estocástico $\{X_t\}$ es completamente estacionario (fuertemente estacionario) si $\forall n \geq 1$, para cualesquiera $t_1, t_2, \dots, t_n \in T$ y $\forall \tau \in \mathbb{R}$ tal que $t_{1+\tau}, t_{2+\tau}, \dots, t_{n+\tau}$ están en T , la función de distribución conjunta de $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ es igual a la función de distribución conjunta de $X_{t_{1+\tau}}, X_{t_{2+\tau}}, \dots, X_{t_{n+\tau}}$. Es decir, si denotamos a $\mathbb{P}(X_t \leq a)$ como $F_t(a)$, tenemos:

$$\begin{aligned} F_{t_1, t_2, \dots, t_n}(a_1, a_2, \dots, a_n) &= \mathbb{P}(X_{t_1} \leq a_1, X_{t_2} \leq a_2, \dots, X_{t_n} \leq a_n) \\ &= \mathbb{P}(X_{t_{1+\tau}} \leq a_1, X_{t_{2+\tau}} \leq a_2, \dots, X_{t_{n+\tau}} \leq a_n) \\ &= F_{t_{1+\tau}, t_{2+\tau}, \dots, t_{n+\tau}}(a_1, a_2, \dots, a_n) \end{aligned} \quad (3.2.1)$$

Definición 3.2.3 Un proceso estocástico $\{X_t\}$ se dice completamente estacionario (débilmente estacionario) si $\forall n \geq 1$, para cualesquiera $t_1, t_2, \dots, t_n \in T$ y $\forall \tau \in \mathbb{R}$ tal que $t_{1+\tau}, t_{2+\tau}, \dots, t_{n+\tau}$ están en T , todos los momentos de orden 1 y 2 de $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ existen y son iguales a los correspondientes momentos de $X_{t_{1+\tau}}, X_{t_{2+\tau}}, \dots, X_{t_{n+\tau}}$.

En otras palabras, la media, $\mathbb{E}(X_t)$, la varianza, $V(X_t) = \mathbb{E}(X_t^2) - \mathbb{E}^2(X_t)$, y la covarianza, $Cov(X_t X_s) = \mathbb{E}(X_t X_s) - \mathbb{E}(X_t)\mathbb{E}(X_s)$, son las mismas, a pesar de movernos en τ unidades de tiempo. Finalmente, cuando hablemos de estacionario en general nos referimos a estacionario en el sentido débil.

Definición 3.2.4 Si $\{X_t\}$ es un proceso estocástico a tiempo discreto de 2º orden, entonces, $C(t_1, t_2) = \mathbb{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})]$ con $t_1, t_2 \in T$, es la función de autocovarianza del proceso.

La función de autocovarianza está bien definida ya que, si tomamos $X = X_{t_1} - \mu_{t_1}$, e $Y = X_{t_2} - \mu_{t_2}$, entonces $\mathbb{E}(X) = \mathbb{E}(Y) = 0$. Usando la desigualdad de Cauchy tenemos:

$$|C(t_1, t_2)| = |\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$$

pero por ser de 2º orden, se concluye que $\mathbb{E}(X^2), \mathbb{E}(Y^2) < \infty$.

Definición 3.2.5 La función de autocorrelación se define como:

$$\rho(t_1, t_2) = \frac{Cov(t_1, t_2)}{\sqrt{Var(X_{t_1})}\sqrt{Var(X_{t_2})}} = \frac{Cov(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}} \quad (3.2.2)$$

denotando a $Var(X_t)$ como σ_t^2 .

Utilizando la **Definición 3.2.4** tenemos:

$$|\rho(t_1, t_2)| = \frac{|C(t_1, t_2)|}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} \leq \frac{\sqrt{\mathbb{E}(X^2)} \sqrt{\mathbb{E}(Y^2)}}{\sqrt{\sigma_{t_1}^2} \sqrt{\sigma_{t_2}^2}} = 1 \quad (3.2.3)$$

es decir, $|\rho(t_1, t_2)| \leq 1$.

Definición 3.2.6 Sea $\{X_t\}$ un proceso estacionario en el sentido fuerte y de 2º orden, entonces $\{X_t\}$ es un proceso estacionario de 2º orden en el sentido débil.

Demostración

Primero demostraremos que $\mathbb{E}(X_t) = \mathbb{E}(X_{t+\tau})$. Tenemos que:

$$\mathbb{E}(X_t) = \int_{-\infty}^{\infty} x f_{X_t}(x) dx = \int_{-\infty}^{\infty} x f_{X_{t+\tau}}(x) dx = \mathbb{E}(X_{t+\tau}) \quad (3.2.4)$$

ya que $F_{X_t}(x) = F_{X_{t+\tau}}(x)$ o bien, $f_{X_t}(x) = f_{X_{t+\tau}}(x)$ por ser un proceso estacionario en el sentido fuerte. Ahora sigue demostrar que $Cov(t_1, t_2) = Cov(t_1+\tau, t_2+\tau)$. Sabemos, a partir de la definición de covarianza, que, $Cov(t_1, t_2) = \mathbb{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})] = \mathbb{E}[(X_{t_1} - \mu)(X_{t_2} - \mu)]$, con $\mathbb{E}(X_t) = \mathbb{E}(X_{t+\tau}) = \mu$. Entonces:

$$\begin{aligned} Cov(t_1, t_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu)(y - \mu) f_{X_{t_1}, X_{t_2}}(x, y) dx dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu)(y - \mu) f_{X_{t_1+\tau}, X_{t_2+\tau}}(x, y) dx dy \end{aligned}$$

de donde $Cov(t_1, t_2) = Cov(t_1 + \tau, t_2 + \tau)$ y además $\mathbb{E}(X_{t_1} X_{t_2}) = \mathbb{E}(X_{t_1+\tau} X_{t_2+\tau})$.

Tomando $t_1 = t_2$ en $Cov(t_1, t_2)$ tenemos, $Cov(t_1, t_2) = Var(X_{t_1})$ y de aquí se puede ver que $\mathbb{E}(X_{t_1}^2) = \mathbb{E}(X_{t_1+\tau}^2)$, así que los primeros dos momentos son iguales sin importar si se suma τ y entonces existe estacionalidad débil.

Ejemplo 3.2.1 Dar dos contraejemplos para el hecho de que si una serie es débilmente estacionaria esto no implica que sea fuertemente estacionaria.

Para nuestro primer contraejemplo tomemos una secuencia de variables aleatorias independientes $\{X_t\}$ tal que:

$$X_t \sim \begin{cases} \exp(1) & \text{si } t \text{ es par} \\ N(1, 1) & \text{si } t \text{ es impar} \end{cases}$$

Observamos que $\mathbb{E}(X_t) = 1 \forall t$, ya que si t es par, tenemos que $X_t \sim \exp(1)$ con esperanza igual a 1, y si t es impar, tenemos que $X_t \sim N(1, 1)$ con esperanza igual a 1. El mismo argumento se usa para ver que $Var(X_t) = 1 \forall t$.

Ahora bien, nosotros sabemos que $\text{Var}(X_t) = \mathbb{E}(X_t^2) - \mathbb{E}^2(X_t)$, por lo que al despejar $\mathbb{E}(X_t^2)$ obtenemos $\mathbb{E}(X_t^2) = \text{Var}(X_t) + \mathbb{E}^2(X_t) = 1 + 1 = 2 \forall t$. Es decir:

$$\begin{aligned}\mathbb{E}(X_t) &= 1 = \mathbb{E}(X_{t+\tau}) \\ \mathbb{E}(X_t^2) &= 2 = \mathbb{E}(X_{t+\tau}^2)\end{aligned}$$

Además, $\text{Cov}(X_t, X_{t+\tau}) = 0$, por ser una secuencia de variables aleatorias independientes.

$\therefore 0 = \mathbb{E}(X_t X_{t+\tau}) - (1)(1)$, por lo que $\mathbb{E}(X_t X_{t+\tau}) = 1$.

Podemos concluir que $\{X_t\}$ es débilmente estacionario. Sin embargo, si tomamos, X_1, X_2 con $X_1 \sim N(1, 1)$ y $X_2 \sim \exp(1)$, tenemos que, para $t_1 = 1$ y $\tau = 1$, $X_{t_1} = X_1$ y $X_{t_1+\tau} = X_2$, no tienen la misma distribución y entonces $\{X_t\}$ no puede ser fuertemente estacionaria.

Ahora bien, para nuestro segundo contraejemplo, sea W una variable aleatoria con distribución uniforme en el intervalo $(0, 2\pi]$. Sea $X_t = \cos(tW)$ para $t = 1, \dots, n$. Al aplicar el operador esperanza obtenemos:

$$\mathbb{E}(X_t) = \int_0^{2\pi} \frac{\cos(tW)}{2\pi} dW = \frac{1}{2\pi t} \text{sen}(tW) \Big|_0^{2\pi} = 0$$

Se concluye que $\mathbb{E}(X_t)$ no depende del tiempo. Ahora bien, para obtener la covarianza del proceso $\{X_t\}$ tenemos:

$$\begin{aligned}\text{Cov}(X_t, X_s) &= \mathbb{E}[\cos(tW)\cos(sW)] \\ &= \frac{1}{2} \mathbb{E}[\cos(tW - sW) + \cos(tW + sW)]\end{aligned}$$

Ahora bien, sea $a = t - s$ y $b = t + s$; entonces, la covarianza del proceso resulta ser:

$$\text{Cov}(X_t, X_s) = \frac{1}{2} \mathbb{E}[\cos(aW) + \cos(bW)] = \frac{1}{2} \mathbb{E}[\cos(aW)] + \frac{1}{2} \mathbb{E}[\cos(bW)]$$

$\therefore \text{Cov}(X_t, X_s) = 0, \forall t \neq s$. Finalmente para el segundo momento utilizaremos la propiedad $\cos^2(tW) = 1 - \text{sen}^2(tW)$, obteniendo:

$$\begin{aligned}\mathbb{E}(X_t^2) &= \int_0^{2\pi} \frac{1 - \text{sen}^2(tW)}{2\pi} dW \\ &= \int_0^{2\pi} \frac{1}{2\pi} dW - \int_0^{2\pi} \frac{\text{sen}^2(tW)}{2\pi} dW \\ &= 1 - \int_0^{2\pi} \frac{\text{sen}^2(tW)}{2\pi} dW\end{aligned}$$

pero $\int_0^{2\pi} \frac{\text{sen}^2(tW)}{2\pi} dW = \frac{1}{2\pi} \int_0^{2\pi} \frac{1}{2} [1 - \cos(2tW)] dW = 0$, por lo que $\mathbb{E}(X_t^2) = 1$, que no depende del tiempo.

El proceso $\{X_t\}$ es débilmente estacionario. Por otra parte, el proceso no es fuertemente estacionario ya que si tomamos $tW = 2\pi$ y $tW = \frac{\pi}{4}$, ambos en el intervalo $(0, 2\pi]$, la distribución conjunta depende de su lugar en la secuencia, es decir, en algunas ocasiones el coseno toma un valor y en otras ocasiones otro valor de manera cíclica y esto afecta por tanto la distribución.

Ahora bien, si $\{X_t\}$ es un proceso estacionario de 2º orden entonces la función de autocovarianza es una función que en realidad depende de la distancia entre los tiempos $|t_2 - t_1|$. Tomemos por un momento $\tau = -t_1$, entonces:

$$\begin{aligned} \text{Cov}(t_1, t_2) &= \text{Cov}(t_1 - t_1, t_2 - t_1) \\ &= \text{Cov}(0, t_2 - t_1) \end{aligned}$$

Análogamente, si $\tau = -t_2$ tenemos:

$$\begin{aligned} \text{Cov}(t_1, t_2) &= \text{Cov}(t_1 - t_2, t_2 - t_2) \\ &= \text{Cov}(t_1 - t_2, 0) \end{aligned}$$

Así que, dicha función depende de $|t_1 - t_2|$. Por tanto, obtenemos una sucesión $\gamma_X(\tau)$ tal que $C(t_1, t_2) = \gamma_X(|t_1 - t_2|) = \gamma_X(\tau)$, que es la sucesión de autocovarianzas teórica $\{\gamma_X(\tau)\}_\tau$.

La sucesión de autocovarianzas teórica de un proceso estacionario posee 3 propiedades deseables. La primera de ellas se deriva al tomar $\tau = 0$, con lo cual obtenemos:

$$\gamma_X(0) = \mathbb{E}[(X_t - \mu)^2] \quad (3.2.5)$$

que es claro, la varianza del proceso $\{X_t\}$. A partir de la desigualdad de Cauchy se puede obtener la segunda propiedad, a saber:

$$|\gamma_X(\tau)| \leq \gamma_X(0) \quad (3.2.6)$$

La última propiedad nos dice que la sucesión de autocovarianzas teórica de un proceso estacionario es simétrica respecto al origen, es decir, $\gamma_X(\tau) = \gamma_X(-\tau)$. Para demostrar lo anterior partimos de:

$$\begin{aligned} \gamma_X(\tau) &= \gamma_X(t + \tau - t) \\ &= \mathbb{E}[(X_{t+\tau} - \mu)(X_t - \mu)] \\ &= \mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)] \\ &= \gamma_X(t - (t + \tau)) \\ &= \gamma_X(-\tau) \end{aligned} \quad (3.2.7)$$

Similarmente, se puede obtener la sucesión de autocorrelaciones teórica (ACF), la cual está dada por:

$$\rho_X(\tau) = \frac{\gamma_X(\tau)}{\gamma_X(0)} \quad (3.2.8)$$

con $\gamma_X(0) = Cov(t, t) = Var(t)$, $\forall t$.

Debemos decir, a manera de observación, que usaremos de manera indistinta, por ejemplo para $\tau = 1$, y así $\forall \tau$, la notación $\rho_X(1)$ o bien, ρ_1 .

3.2.1. Ruido blanco

Definición 3.2.7 *Un ruido blanco es una sucesión de variables aleatorias $\{X_t\}$ no correlacionadas tales que $\mathbb{E}(X_t) = \mu$ y $Var(X_t) = \sigma_X^2 \forall t$. Dicho proceso lo denotaremos como $\{X_t\} \sim WN(0, \sigma_X^2)$.*

La sucesión de autocovarianzas teórica de un ruido blanco es:

$$Cov(X_t, X_{t+\tau}) = \gamma_X(\tau) = \begin{cases} Cov(X_t, X_t) = Var(X_t) = \sigma_X^2 & \text{si } \tau = 0 \\ Cov(X_0, X_\tau) = 0 & \text{si } \tau \neq 0 \end{cases}$$

$Cov(X_0, X_\tau) = 0$ ya que las variables aleatorias no están correlacionadas. La sucesión de autocorrelaciones teórica (ACF) está dada por:

$$\rho_X(\tau) = \begin{cases} 1 & \text{si } \tau = 0 \\ 0 & \text{si } \tau \neq 0 \end{cases}$$

Claramente, el ruido blanco es un proceso estacionario ya que $\mathbb{E}(X_t) = \mu$, $Var(X_t) = \sigma_X^2$ y $Cov(X_t, X_{t+\tau})$ no dependen del tiempo.

En algún momento vamos a requerir que el proceso de ruido blanco sea un conjunto de variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_X^2 . Distinguiremos dicho proceso con la notación $\{X_t\} \sim IID(0, \sigma_X^2)$. Como ejercicio mental, claramente, un proceso $\{X_t\} \sim IID(0, \sigma_X^2)$ es también $\{X_t\} \sim WN(0, \sigma_X^2)$ pero no al revés.

Ejemplo 3.2.2 *Realicemos a continuación una simulación de 350 observaciones normales de un proceso de ruido blanco con media cero y varianza igual a 1. El código en R correspondiente a dicha simulación será entonces el siguiente:*

```
> WN=arima.sim(list(order=c(0,0,0)), n=350)
> plot(WN,xlab="",ylab="X(t)",main=expression(Proceso~WN[G](0,1)~simulado),col="blue")
```

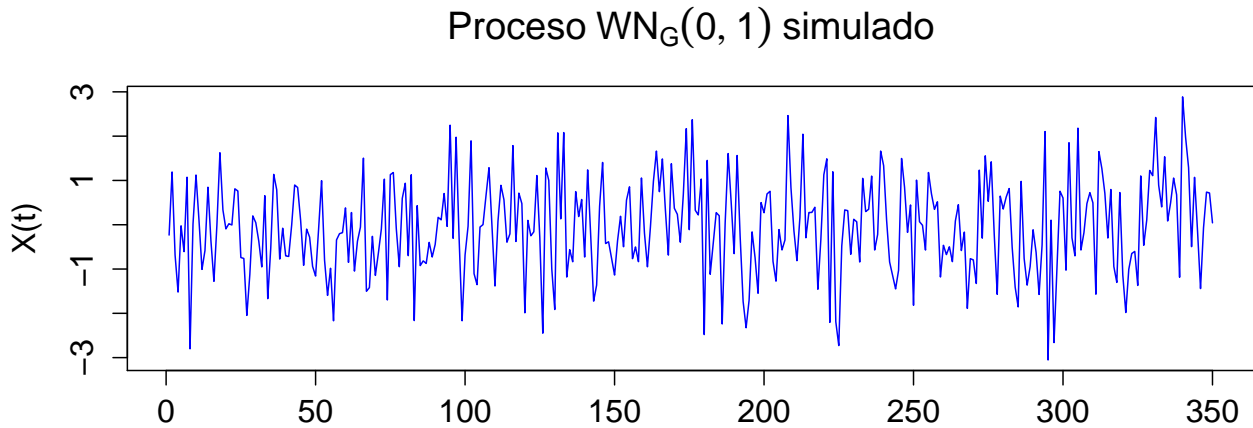


Figura 3.2.1: Figura que muestra la simulación de un proceso de ruido blanco Gaussiano

La **Figura 3.2.1** muestra la simulación de un ruido blanco, $\{X_t\} \sim WN_G(0,1)$, donde $WN_G(0,1)$ corresponde a un ruido blanco con distribución Gaussiana, media cero y varianza igual a uno.

Como observación final, debemos recalcar que en el ejemplo anterior, la distribución del ruido blanco es Gaussiana. Pero cuidado, no por este ejemplo debemos asumir que todo ruido blanco involucra una distribución Gaussiana o Normal. Este ejemplo es un proceso de ruido blanco particular, el cual involucra un proceso $\{X_t\}$ con variables aleatorias independientes con distribución Normal, media 0 y varianza igual a σ_X^2 . Recordemos que hemos denotado al ruido blanco como $\{X_t\} \sim WN(0, \sigma_X^2)$. Se hará una distinción clara en el texto cuando la distribución del ruido blanco sea Gaussiana, a saber con la notación, $\{X_t\} \sim WN_G(0, \sigma_X^2)$.

Ahora bien, como en la definición de los procesos ARMA(p,q), los cuales se verán en la **Sección 3.5**, se requiere que el proceso sea estacionario, debe ocurrir que el proceso tenga la misma media a través del tiempo y que la varianza sea constante en el tiempo. Gráficamente el proceso estacionario ideal debe ser algo como lo que se muestra en la **Figura 3.2.2**.

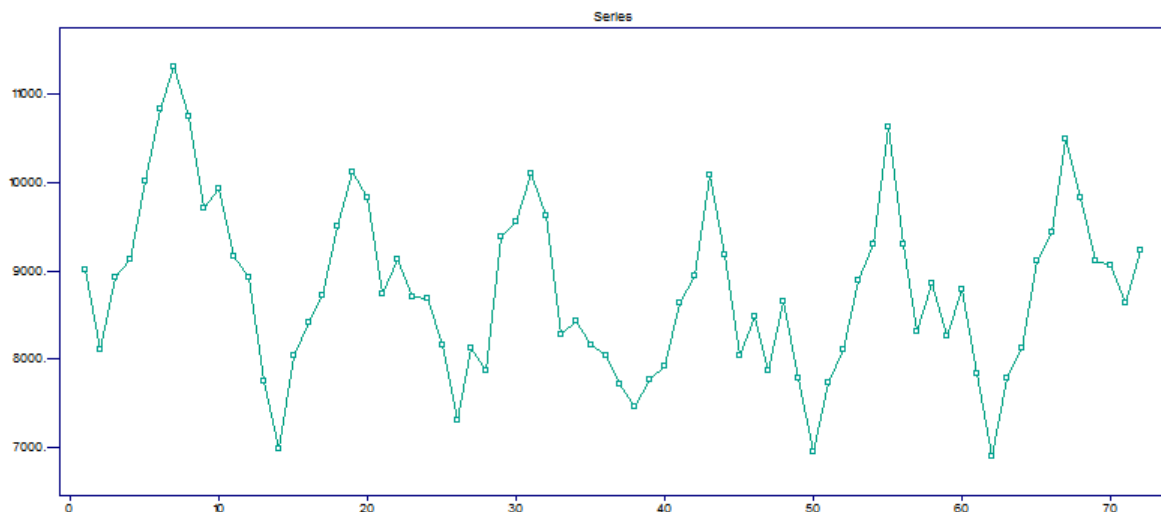


Figura 3.2.2: Figura que representa a un proceso estacionario

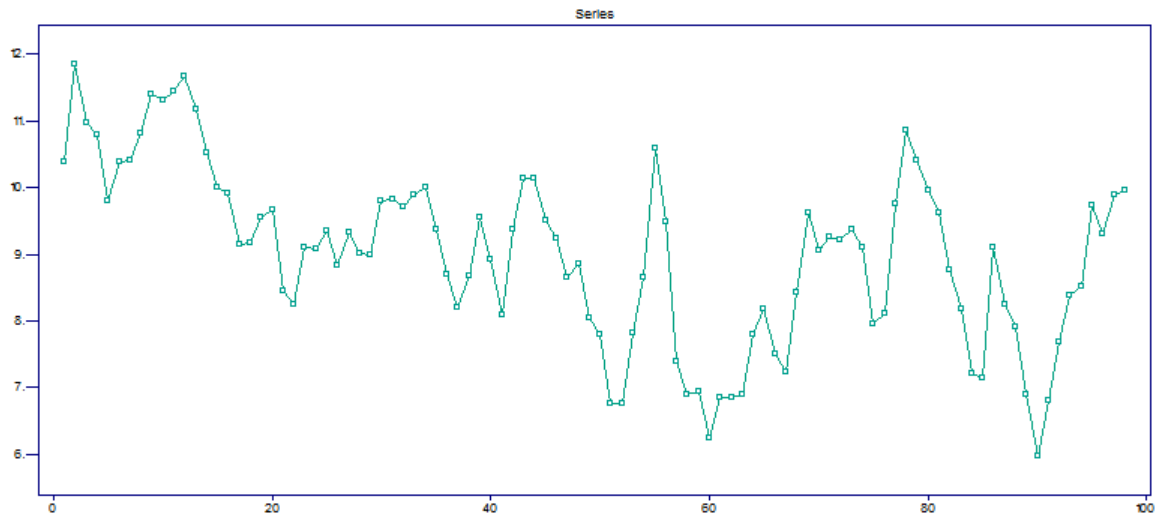


Figura 3.2.3: Figura que muestra un proceso con tendencia

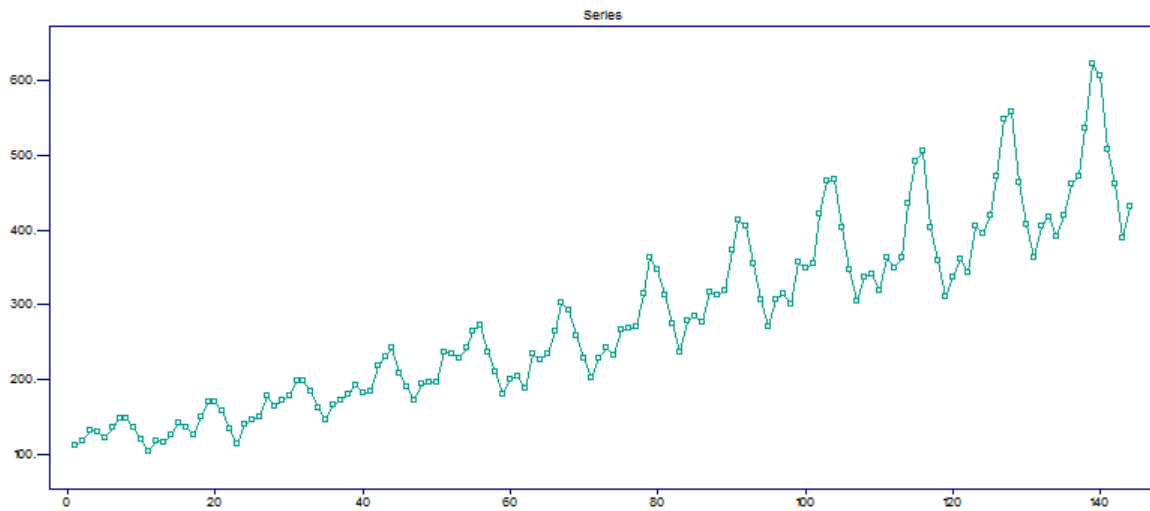


Figura 3.2.4: Figura que muestra un proceso con heteroscedasticidad

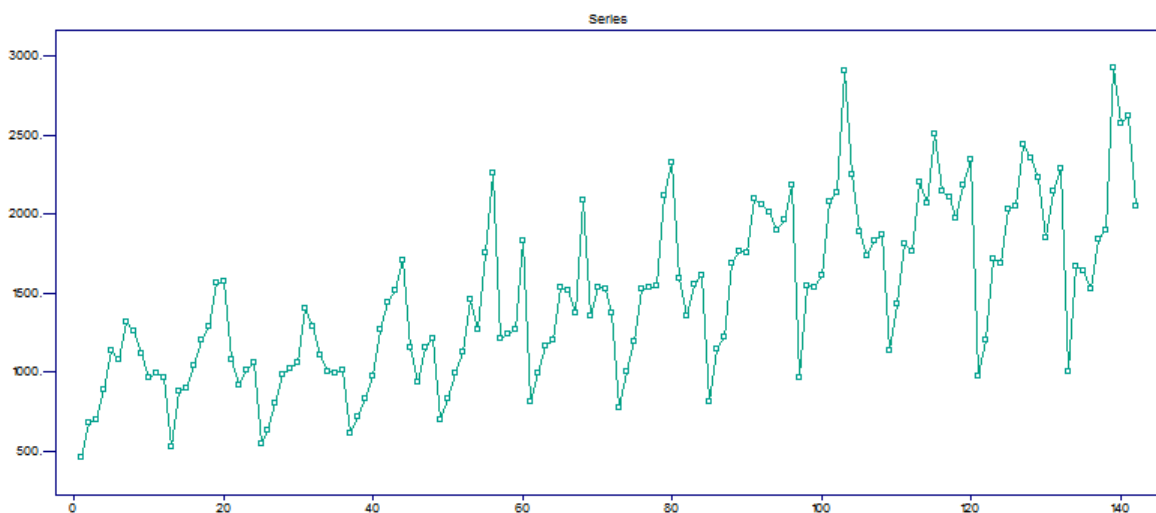


Figura 3.2.5: Figura que muestra un proceso con estacionalidad

El proceso no será estacionario cuando encontremos una sucesión con media no constante o que presente tendencia como la que se muestra en la **Figura 3.2.3**. La serie tampoco será estacionaria si la variabilidad alrededor de la media no es constante como la que se muestra en la **Figura 3.2.4** o finalmente para un proceso que presente estacionalidad como el que se muestra en la **Figura 3.2.5**.

3.2.2. Transformada de Box-Cox

Antes de eliminar tendencia y estacionalidad podemos aplicar transformaciones que ayuden a mejorar el supuesto de varianza constante en los datos. Existen muchos casos de series de tiempo que son estacionarias con respecto a su media pero que no lo son en su varianza.

Es muy común que la varianza de un proceso no estacionario, cambie conforme cambia el nivel de la serie. Es decir, $Var(X_t) = cf(\mu_t)$ para c constante positiva y f una función. Nuestra tarea será la de encontrar una función T tal que la serie $\{T(X_t)\}$ tenga varianza constante. Sea

$$T(X_t) = T(\mu_t) + T'(\mu_t)(X_t - \mu) \quad (3.2.9)$$

la expansión de $\{T(X_t)\}$ en series de Taylor alrededor de μ , donde $T'(\mu_t)$ es la primer derivada de la función $T(X_t)$ evaluada en μ . Al tomar la varianza de la **ec. 3.2.9** obtenemos:

$$\begin{aligned} Var[T(X_t)] &= Var[T(\mu_t) + T'(\mu_t)(X_t - \mu)] \\ &= [T'(\mu_t)]^2 Var(X_t) \\ &= c[T'(\mu_t)]^2 f(\mu_t) \end{aligned} \quad (3.2.10)$$

Para que la serie $\{T(X_t)\}$ tenga varianza constante, $T(X_t)$ debe escogerse de tal forma que:

$$T'(\mu_t) = \frac{1}{\sqrt{f(\mu_t)}} \quad (3.2.11)$$

La **ec. 3.2.11** implica entonces:

$$T(\mu_t) = \int \frac{1}{\sqrt{f(\mu_t)}} d\mu_t \quad (3.2.12)$$

Por ejemplo, si la desviación estándar de una serie de tiempo es proporcional a su nivel, de tal forma que $Var(X_t) = c^2\mu_t^2$ entonces, el logaritmo de la serie $\{X_t\}$, es la transformación adecuada para estabilizar la varianza:

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^2}} d\mu_t = \ln(\mu_t)$$

Si la varianza de la serie es proporcional al nivel de tal forma que $Var(X_t) = c\mu_t$ entonces, la raíz cuadrada de la serie $\{X_t\}$, es la transformación óptima para estabilizar la varianza:

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t}} d\mu_t = 2\sqrt{\mu_t}$$

Finalmente y como último ejemplo, si la desviación estándar de la serie de tiempo es proporcional al nivel a la cuarta potencia de tal forma que $Var(X_t) = c^2\mu_t^4$, entonces el recíproco de la serie $\{X_t\}$ es la transformación adecuada para estabilizar la varianza.

$$T(\mu_t) = \int \frac{1}{\sqrt{\mu_t^4}} d\mu_t = -\frac{1}{\mu_t}$$

Por ejemplo, si tuvieramos una serie de tiempo como la que se muestra en la **Figura 3.2.6**.

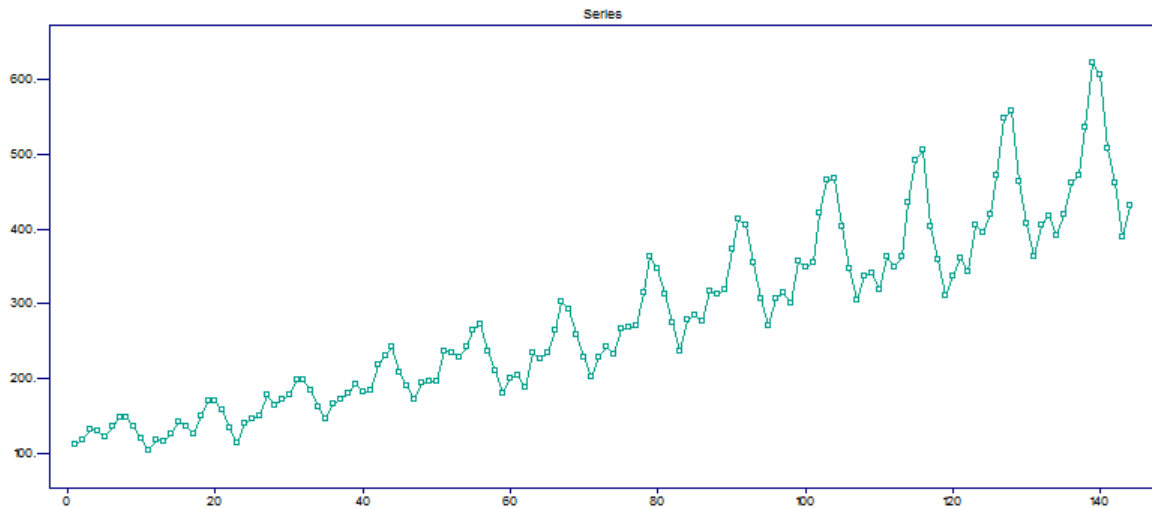


Figura 3.2.6: Figura de un proceso con heteroscedasticidad

veríamos que aparte de estacionalidad y tendencia los valores de la serie se incrementan al pasar el tiempo. Al aplicar logaritmo a toda la serie obtenemos la **Figura 3.2.7**.

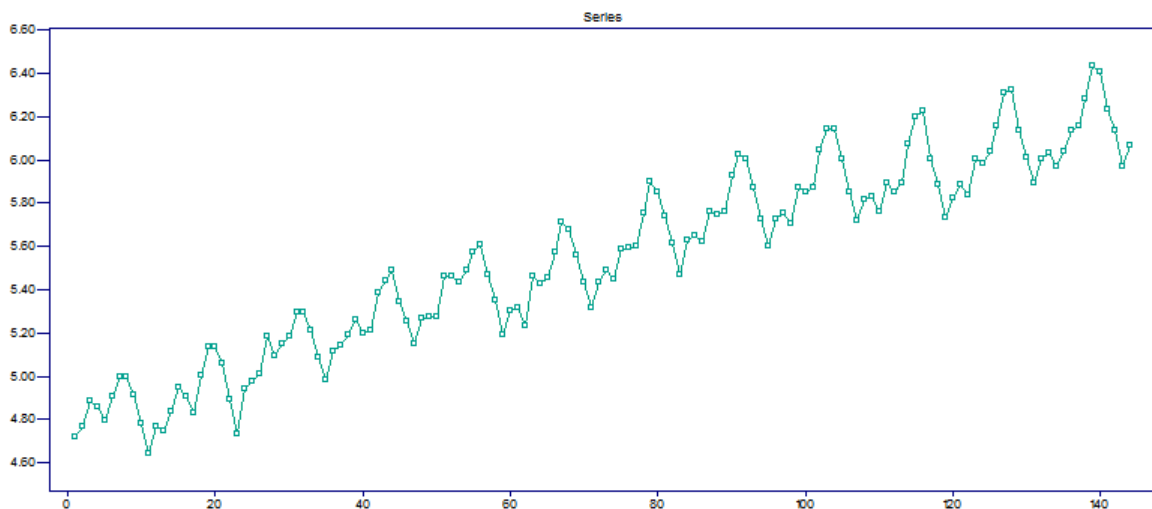


Figura 3.2.7: Figura de un proceso con homoscedasticidad

Si bien la estacionalidad y tendencia permanecen, ya no hay tanto problema de que al aumentar el tiempo, los valores de la serie se incrementen y de aquí que la varianza sea más constante. De manera más general, para estabilizar la varianza utilizamos las llamadas transformaciones de Box-Cox dadas por la función:

$$f_{\lambda}(X_t) = \begin{cases} \lambda^{-1}(X_t^{\lambda} - 1) & \text{si } \lambda \geq 0, X_t \geq 0 \\ \ln(X_t) & \text{si } \lambda = 0, X_t > 0 \end{cases}$$

La **Tabla 3.2.1** muestra algunas transformaciones para ciertos valores de λ .

Valores para λ	Transformación
-1	$\frac{1}{X_t}$
-0.5	$\frac{1}{\sqrt{X_t}}$
0	$\ln(X_t)$
0.5	$\sqrt{X_t}$
1	X_t

Tabla 3.2.1: Tabla que muestra los valores aproximados para las transformaciones de Box-Cox

Como comentarios finales debemos recordar que las transformaciones de Box-Cox aplican solamente para series $\{X_t\}$ positivas. Además, una transformación de Box-Cox, de ser necesaria, debe aplicarse antes de realizar cualquier otro procedimiento. Finalmente, el aplicar una transformación de Box-Cox no sólo estabiliza la varianza sino que también mejora la aproximación de los residuos a una distribución Normal.

3.2.3. Eliminación de tendencia y estacionalidad

Nuestra siguiente pregunta será, ¿cómo eliminar de una serie de tiempo la tendencia y estacionalidad para lograr obtener una serie estacionaria que se pueda modelar con un proceso ARMA(p,q)? Es decir, dada una serie de tiempo $\{Y_t\}$ no necesariamente estacionaria, queremos expresarla como:

$$Y_t = T_t + E_t + X_t \tag{3.2.13}$$

con T_t la tendencia al tiempo t, E_t la estacionalidad al tiempo t y X_t un proceso estacionario al tiempo t. El proceso $\{X_t\}$ es el que se modelaría con un ARMA(p,q).

En el caso simple, donde sólo exista tendencia, ¿cómo estimarla? Nuestras opciones, vistas en el capítulo anterior, son, el modelo de regresión lineal simple, visto en la **Sección 2.3**, el método de Holt, visto en la **Sección 2.3.6**, y el filtro de promedios móviles, visto en la **Sección 2.3.4**. Los residuos $\hat{e}_t = Y_t - \hat{Y}_t$ deberían ser un proceso estacionario, el cual podemos modelar con un proceso ARMA(p,q).

Sin embargo, otro método muy usual para obtener una serie en la cual se elimina la tendencia consiste en aplicar diferencias. Para ello, para un proceso $\{X_t\}$ se definen los operadores B y ∇ , tales que $BX_t = X_{t-1}$, $\nabla X_t = X_t - X_{t-1}$.

Para $j > 1$ tenemos $B^j X_t = X_{t-j}$ y como relación especial tenemos:

$$\nabla^j X_t = (1 - B)^j X_t = \sum_{k=0}^j \binom{j}{k} (-B)^k X_t \quad (3.2.14)$$

¿Qué pasa si aplicamos el operador ∇ a una tendencia lineal, es decir, a $T_t = a + \beta t$?

$$\nabla T_t = [a + \beta t] - [a + \beta(t-1)] = a + \beta t - a - \beta t + \beta = \beta$$

es decir, se elimina la tendencia lineal y si tenemos $Y_t = T_t + X_t$, quedaría como:

$$\nabla Y_t = \nabla T_t + \nabla X_t = \beta + \nabla X_t \quad (3.2.15)$$

proceso en el cual no existe tendencia. Ahora bien, para darnos una idea del operador ∇^j cuando $j = 2$, supongamos la componente de tendencia igual a $T_t = a_0 + a_1 t + a_2 t^2$. Al aplicar el operador ∇^j con $j = 1$ obtenemos:

$$\begin{aligned} \nabla T_t &= [a_0 + a_1 t + a_2 t^2] - [a_0 + a_1(t-1) + a_2(t-1)^2] \\ &= a_0 + a_1 t + a_2 t^2 - a_0 - a_1 t + a_1 - a_2 t^2 + 2a_2 t - a_2 \\ &= 2a_2 t + a_1 - a_2 \end{aligned}$$

Para dicha j , no logramos eliminar la tendencia, ya que aún sobrevive el término $2a_2 t$. Para eliminar dicho término pasamos, nuevamente, a aplicar el operador ∇ de la siguiente forma:

$$\begin{aligned} \nabla \nabla T_t &= \nabla(2a_2 t + a_1 - a_2) \\ &= [2a_2 t + a_1 - a_2] - [2a_2(t-1) + a_1 - a_2] \\ &= 2a_2 \end{aligned}$$

Finalmente logramos eliminar la tendencia y si nuestra serie estuviese dada por $Y_t = T_t + X_t$ tendríamos, después de aplicar el operador ∇^2 :

$$\nabla^2 Y_t = \nabla^2 T_t + \nabla^2 X_t = 2a_2 + \nabla^2 X_t \quad (3.2.16)$$

Un filtro es una sucesión de números $\{a_k\}$ diseñados para extraer información de una serie de tiempo. El operador ∇^j es un filtro ya que se tienen $a_k = \frac{(-B)^k j!}{k!(j-k)!}$ números, que escritos como combinación lineal extraen la tendencia de las series.

En general, el operador ∇^j extrae una tendencia de un polinomio de orden j , es decir, el diferenciar j veces elimina la tendencia correspondiente a un polinomio $a_0 + a_1 t + a_2 t^2 + \dots + a_j t^j$ (sólo queda la constante). La aseveración anterior será demostrada formalmente con el siguiente teorema.

Teorema 3.2.1 Para cada $k \geq 1$, sea $T_t^{k+1} = \sum_{l=0}^{k+1} a_l t^l$, polinomio de orden $k+1$, entonces:

$$\nabla^k T_t^{k+1} = c(\{a_1, a_2, \dots, a_{k+1}\}) + (k+1)!a_{k+1}t \quad (3.2.17)$$

Es decir, al diferenciar k veces queda una tendencia lineal. La demostración se hará por inducción, considerando que el operador ∇^j es lineal, lo cual fácilmente puede comprobar el lector. Dicha aseveración se usará a lo largo de la demostración.

Demostración

Si $k = 1$

$$\begin{aligned} \nabla T_t^2 &= \nabla(a_0 + a_1 t + a_2 t^2) \\ &= a_0 + a_1 t + a_2 t^2 - a_0 - a_1 t + a_1 - a_2 t^2 + 2a_2 t - a_2 \\ &= c(\{a_1, a_2\}) + 2!a_2 t \end{aligned} \quad (3.2.18)$$

Supongamos válido para $k = n$, es decir, $\nabla^n T_t^{n+1} = c(\{a_1, a_2, \dots, a_{n+1}\}) + (n+1)!a_{n+1}t$. Por demostrar válido para $k = n+1$. Como $T_t^{n+2} = T_t^{n+1} + a_{n+2}t^{n+2}$ tenemos:

$$\begin{aligned} \nabla^{n+1} T_t^{n+2} &= \nabla^{n+1} (T_t^{n+1} + a_{n+2}t^{n+2}) \\ &= \nabla^{n+1} T_t^{n+1} + \nabla^{n+1} a_{n+2}t^{n+2} \\ &= (n+1)!a_{n+1} + \nabla^n [a_{n+2}t^{n+2} - a_{n+2}(t-1)^{n+2}] \\ &= (n+1)!a_{n+1} + \nabla^n \left[a_{n+2}t^{n+2} - a_{n+2} \sum_{k=0}^{n+2} \binom{n+2}{k} t^k (-1)^{n+2-k} \right] \\ &= (n+1)!a_{n+1} + \nabla^n \left[a_{n+2}t^{n+2} - a_{n+2} \left(t^{n+2} - \binom{n+2}{n+1} t^{n+1} + \sum_{k=0}^n \binom{n+2}{k} t^k (-1)^{n+2-k} \right) \right] \\ &= (n+1)!a_{n+1} + \nabla^n \left[a_{n+2}(n+2)t^{n+1} + \sum_{k=0}^n \binom{n+2}{k} t^k (-1)^{n+3-k} a_{n+2} \right] \\ &= (n+1)!a_{n+1} + c(\{a'_1, a'_2, \dots, a'_{n+1}\}) + (n+1)!a'_{n+1}t \\ &= c(\{a_1, a_2, \dots, a_{n+2}\}) + (n+2)(n+1)!a_{n+2}t \\ &= c(\{a_1, a_2, \dots, a_{n+2}\}) + ((n+1)+1)!a_{(n+1)+1}t \end{aligned}$$

con $a'_k = \binom{n+2}{k} (-1)^{n+3-k} a_{n+2}$ y $a'_{n+1} = a_{n+2}(n+2)$. Así, utilizando este teorema, vemos que la serie original $Y_t = T_t^k + X_t$, la cual presenta una tendencia correspondiente a un polinomio de orden k , $T_t^k = a_0 + a_1 t + a_2 t^2 + \dots + a_k t^k$, se convierte en:

$$\begin{aligned} \nabla^k Y_t &= \nabla^k T_t^k + \nabla^k X_t \\ &= \nabla \nabla^{k-1} T_t^k + \nabla^k X_t \\ &= \nabla [c(\{a_1, \dots, a_k\}) + k!a_k t] + \nabla^k X_t \\ &= \nabla(a + \beta t) + \nabla^k X_t \\ &= k!a_k + \nabla^k X_t \end{aligned} \quad (3.2.19)$$

Recordemos que buscamos un proceso estacionario que se pueda modelar, si lo logramos después de las diferencias entonces estaremos satisfechos. ¿Cuántas veces diferenciar? En la práctica no deben ser demasiadas porque interpretar la serie de diferencias puede no tener sentido y regresar a los valores originales puede costar mucho trabajo.

Como observación final, si $\{Y_t\}$ es estacionario entonces el proceso $\{\nabla^k Y_t\}$ es estacionario, es decir, si los datos originales dependían de una tendencia y un proceso estacionario $\{X_t\}$ entonces al diferenciar se elimina la tendencia y queda otro proceso otra vez estacionario. Si $\{Y_t\}$ no es estacionario de 2° orden entonces no hay garantía de que $\{\nabla^k Y_t\}$ sea estacionario.

Ya vimos como se puede modelar y obtener una nueva serie donde se elimine la tendencia, pero, ¿qué ocurre con la estacionalidad? Ya vimos que el método de Holt-Winters nos sirve para modelar series que incluyen tendencia y estacionalidad, obteniendo estimadores \hat{Y}_t bajo el modelo. El error correspondiente $\hat{e}_t = Y_t - \hat{Y}_t$ puede modelarse con un proceso ARMA(p,q). Otra forma para eliminar estacionalidad y tendencia es con el filtro de estacionalidad visto en el capítulo anterior en la **Sección 2.4.1**. Otra manera de eliminar estacionalidad en series de tiempo es usando diferencias como en el caso de la tendencia, pero diferenciamos de acuerdo con un “lag” o retraso “s” que depende de la estacionalidad. Es decir, obtenemos:

$$X_{j,k} = Y_{j,k} - Y_{j-1,k} \quad (3.2.20)$$

con $k = 1, \dots, s$ el número de observaciones en el periodo y $j = 2, \dots, e$ el número de periodos. Lo que se está haciendo es aplicar el operador ∇_s el cual está definido por:

$$\nabla_s X_t = X_t - X_{t-s} \quad (3.2.21)$$

Si $Y_t = T_t + E_t + X_t$ entonces:

$$\begin{aligned} \nabla_s Y_t &= \nabla_s(T_t + E_t + X_t) \\ &= (T_t + E_t + X_t) - (T_{t-s} + E_{t-s} + X_{t-s}) \\ &= T_t - T_{t-s} + X_t - X_{t-s} \end{aligned} \quad (3.2.22)$$

ya que $E_t = E_{t-s}$. Si esta serie sigue teniendo tendencia podríamos usar el operador ∇^j .

3.2.4. Ejemplos de procesos estacionarios

Ejemplo 3.2.3 Sea $\{X_t\} = a + b\epsilon_t + c\epsilon_{t-2}$, con a, b y $c \in \mathbb{R}$ tales que $a, b, c < \infty$ y $\{\epsilon_t\}$ es un proceso de ruido blanco tal que $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. ¿Es $\{X_t\}$ un proceso estacionario?

El primer momento viene dado por:

$$\begin{aligned} \mathbb{E}(X_t) &= \mathbb{E}(a + b\epsilon_t + c\epsilon_{t-2}) \\ &= a + b\mathbb{E}(\epsilon_t) + c\mathbb{E}(\epsilon_{t-2}) \\ &= a < \infty \end{aligned}$$

ya que $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$.

Análogamente, el segundo momento será:

$$\begin{aligned}
 \mathbb{E}(X_t^2) &= \text{Var}(X_t) + \mathbb{E}^2(X_t) \\
 &= \text{Var}(a + b\epsilon_t + c\epsilon_{t-2}) + a^2 \\
 &= \text{Var}(b\epsilon_t + c\epsilon_{t-2}) + a^2 \\
 &= b^2\text{Var}(\epsilon_t) + c^2\text{Var}(\epsilon_{t-2}) + a^2 \\
 &= \sigma_\epsilon^2(b^2 + c^2) + a^2 < \infty
 \end{aligned}$$

La covarianza entre ϵ_t y ϵ_{t-2} es cero ya que al ser ruido blanco, las variables aleatorias no están correlacionadas. La sucesión de autocovarianzas viene dada por:

$$\begin{aligned}
 \gamma_X(\tau) &= \text{Cov}(X_t, X_{t+\tau}) \\
 &= \text{Cov}(a + b\epsilon_t + c\epsilon_{t-2}, a + b\epsilon_{t+\tau} + c\epsilon_{t+\tau-2}) \\
 &= b^2\text{Cov}(\epsilon_{t+\tau}, \epsilon_t) + bc\text{Cov}(\epsilon_{t+\tau}, \epsilon_{t-2}) + bc\text{Cov}(\epsilon_{t+\tau-2}, \epsilon_t) + c^2\text{Cov}(\epsilon_{t+\tau-2}, \epsilon_{t-2})
 \end{aligned}$$

Es claro que solamente para $\tau = 0, \pm 2$ la sucesión de autocovarianzas toma valores distintos de cero. Tenemos por tanto:

$$\gamma_X(\tau) = \begin{cases} \sigma_\epsilon^2(b^2 + c^2) & \text{si } \tau = 0 \\ bc\sigma_\epsilon^2 & \text{si } \tau = \pm 2 \\ 0 & \text{si } \tau \neq 0, \pm 2 \end{cases}$$

al ser $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. Podemos concluir que $\{X_t\}$ es un proceso estacionario ya que $\mathbb{E}(X_t)$, $\mathbb{E}(X_t^2)$ y $\gamma_X(\tau)$ son finitos y no dependen del tiempo.

Ejemplo 3.2.4 Sea $\{X_t\} = Y_1 + Y_2 + \dots + Y_t$, donde $\{Y_t\}$ son variables aleatorias independientes e idénticamente distribuidas con media 0 y varianza σ_Y^2 . ¿Es $\{X_t\}$ un proceso estacionario? Obtener la sucesión de autocovarianzas asociada al proceso $\{X_t\}$.

El primer momento será:

$$\begin{aligned}
 \mathbb{E}(X_t) &= \mathbb{E}(Y_1 + Y_2 + \dots + Y_t) \\
 &= \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \dots + \mathbb{E}(Y_t) \\
 &= 0
 \end{aligned}$$

mientras que el segundo momento será:

$$\begin{aligned}
 \mathbb{E}(X_t^2) &= \text{Var}(X_t) + \mathbb{E}^2(X_t) \\
 &= \text{Var}(X_t) \\
 &= \text{Var}(Y_1 + Y_2 + \dots + Y_t) \\
 &= \text{Var}(Y_1) + \text{Var}(Y_2) + \dots + \text{Var}(Y_t) && (\text{al ser } \{Y_t\} \text{ v.a.i.i.d}) \\
 &= \sigma_Y^2 t
 \end{aligned}$$

La sucesión de autocovarianzas está dada por:

$$\begin{aligned}
\gamma_X(\tau) &= \text{Cov}(X_t, X_{t+\tau}) \\
&= \text{Cov}(X_t, X_t + Y_{t+1} + Y_{t+2} + \dots + Y_{t+\tau}) \\
&= \text{Cov}(X_t, X_t) + \text{Cov}(X_t, Y_{t+1}) + \text{Cov}(X_t, Y_{t+2}) + \dots + \text{Cov}(X_t, Y_{t+\tau}) \\
&= \text{Cov}(X_t, X_t) \quad (\text{al ser } \{Y_t\} \text{ v.a.i.i.d}) \\
&= \text{Var}(X_t) \\
&= \sigma_Y^2 t
\end{aligned}$$

$\therefore \{X_t\}$ no es un proceso estacionario, ya que su varianza y covarianza depende del tiempo.

Ejemplo 3.2.5 Sean $\{X_t\}$ e $\{Y_t\}$ sucesiones no correlacionadas con medias μ_X y μ_Y , varianzas σ_X^2 y σ_Y^2 y sucesión de autocovarianzas $\gamma_X(\tau)$ y $\gamma_Y(\tau)$ respectivamente. Probar que $\{X_t + Y_t\}$ es estacionaria con sucesión de autocovarianzas igual a la suma de las sucesiones de autocovarianzas $\gamma_X(\tau) + \gamma_Y(\tau)$.

Sea $\{W_t\} = \{X_t + Y_t\}$. El primer momento de $\{W_t\}$ es:

$$\begin{aligned}
\mathbb{E}(W_t) &= \mathbb{E}(X_t + Y_t) \\
&= \mu_X + \mu_Y < \infty
\end{aligned}$$

mientras que el segundo momento viene dado por:

$$\begin{aligned}
\mathbb{E}(W_t^2) &= \text{Var}(W_t) + \mathbb{E}^2(W_t) \\
&= \text{Var}(X_t + Y_t) + (\mu_X + \mu_Y)^2 \\
&= \text{Var}(X_t) + \text{Var}(Y_t) + (\mu_X + \mu_Y)^2 \\
&= \sigma_X^2 + \sigma_Y^2 + (\mu_X + \mu_Y)^2 < \infty
\end{aligned}$$

Finalmente, la sucesión de autocovarianzas está dada por:

$$\begin{aligned}
\gamma_W(\tau) &= \text{Cov}(W_t, W_{t+\tau}) \\
&= \text{Cov}(X_t + Y_t, X_{t+\tau} + Y_{t+\tau}) \\
&= \text{Cov}(X_t, X_{t+\tau}) + \text{Cov}(Y_t, X_{t+\tau}) + \text{Cov}(X_t, Y_{t+\tau}) + \text{Cov}(Y_t, Y_{t+\tau}) \\
&= \text{Cov}(X_t, X_{t+\tau}) + \text{Cov}(Y_t, Y_{t+\tau}) \quad (\text{al ser } \{X_t\} \text{ e } \{Y_t\} \text{ no correlacionadas}) \\
&= \gamma_X(\tau) + \gamma_Y(\tau)
\end{aligned}$$

$\therefore \{W_t\}$ es estacionario ya que $\mathbb{E}(W_t)$, $\mathbb{E}(W_t^2)$ y $\gamma_W(\tau)$ son finitos y no dependen del tiempo. Además, $\gamma_W(\tau)$ es igual a la suma de la sucesión de autocovarianzas de X_t e Y_t .

Ejemplo 3.2.6 Sea $\{Y_t\}$ un proceso estacionario con media 0, varianza σ_Y^2 y variables aleatorias no correlacionadas. Sean a y b constantes. Sea $X_t = a + bt + E_t + Y_t$, donde E_t es una componente estacional con periodo 12. Probar que $\nabla \nabla_{12} X_t = (1 - B)(1 - B^{12})X_t$ es un proceso estacionario y expresar su sucesión de autocovarianzas en términos de $\{Y_t\}$.

Comencemos desarrollando $\nabla\nabla_{12}X_t$ de la siguiente forma:

$$\begin{aligned}
 (1 - B^{12} - B + B^{13})X_t &= X_t + X_{t-13} - X_{t-12} - X_{t-1} \\
 &= [a + bt + E_t + Y_t] + [a + b(t-13) + E_{t-13} + Y_{t-13}] \\
 &\quad - [a + b(t-12) + E_{t-12} + Y_{t-12}] - [a + b(t-1) + E_{t-1} + Y_{t-1}] \\
 &= Y_t + Y_{t-13} - Y_{t-12} - Y_{t-1}
 \end{aligned}$$

Sea $W_t = Y_t + Y_{t-13} - Y_{t-12} - Y_{t-1}$, entonces el primer momento del proceso $\{W_t\}$ será:

$$\begin{aligned}
 \mathbb{E}(W_t) &= \mathbb{E}(Y_t + Y_{t-13} - Y_{t-12} - Y_{t-1}) \\
 &= \mathbb{E}(Y_t) + \mathbb{E}(Y_{t-13}) - \mathbb{E}(Y_{t-12}) - \mathbb{E}(Y_{t-1}) \\
 &= 0
 \end{aligned}$$

mientras que la varianza del proceso viene dada por:

$$\begin{aligned}
 \text{Var}(W_t) &= \text{Var}(Y_t + Y_{t-13} - Y_{t-12} - Y_{t-1}) \\
 &= \text{Var}(Y_t) + \text{Var}(Y_{t-13}) - \text{Var}(Y_{t-12}) - \text{Var}(Y_{t-1}) \\
 &= 4\sigma_Y^2
 \end{aligned}$$

y sucesión de autocovarianzas:

$$\gamma_W(\tau) = \begin{cases} 4\sigma_Y^2 & \text{si } \tau = 0 \\ -2\sigma_Y^2 & \text{si } \tau = \pm 1, \pm 12 \\ \sigma_Y^2 & \text{si } \tau = \pm 13 \end{cases}$$

$\therefore \{W_t\}$ es estacionario ya que $\mathbb{E}(W_t)$, $\mathbb{E}(W_t^2)$ y $\gamma_W(\tau)$ son finitos y no dependen del tiempo.

3.3. Modelos de promedios móviles (MA)

3.3.1. Modelo MA(1)

Un proceso MA(1) está definido por:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} \quad (3.3.1)$$

donde μ y $\theta_1 \in \mathbb{R}$ y el proceso $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. Tomemos el proceso centrado dado por $\{Y_t\} = \{X_t - \mu\}$. El primer momento será:

$$\begin{aligned}
 \mathbb{E}(Y_t) &= \mathbb{E}(\epsilon_t + \theta_1 \epsilon_{t-1}) \\
 &= \mathbb{E}(\epsilon_t) + \theta_1 \mathbb{E}(\epsilon_{t-1}) \\
 &= 0
 \end{aligned} \quad (3.3.2)$$

El segundo momento está dado por:

$$\begin{aligned}
 \mathbb{E}(Y_t^2) &= \text{Var}(Y_t) + \mathbb{E}^2(Y_t) \\
 &= \text{Var}(\epsilon_t + \theta_1 \epsilon_{t-1}) \\
 &= \text{Var}(\epsilon_t) + \theta_1^2 \text{Var}(\epsilon_{t-1}) \\
 &= \sigma_\epsilon^2 + \theta_1^2 \sigma_\epsilon^2 \\
 &= \sigma_\epsilon^2 (1 + \theta_1^2) < \infty
 \end{aligned} \tag{3.3.3}$$

mientras que para obtener la sucesión de autocovarianzas teórica partimos de:

$$\begin{aligned}
 \gamma_Y(\tau) &= \text{Cov}(Y_t, Y_{t+\tau}) \\
 &= \text{Cov}(\epsilon_t + \theta_1 \epsilon_{t-1}, \epsilon_{t+\tau} + \theta_1 \epsilon_{t+\tau-1}) \\
 &= \text{Cov}(\epsilon_t, \epsilon_{t+\tau}) + \theta_1 \text{Cov}(\epsilon_{t+\tau}, \epsilon_{t-1}) + \theta_1 \text{Cov}(\epsilon_{t+\tau-1}, \epsilon_t) + \theta_1^2 \text{Cov}(\epsilon_{t+\tau-1}, \epsilon_{t-1})
 \end{aligned} \tag{3.3.4}$$

La sucesión de autocovarianzas teórica es por tanto:

$$\gamma_Y(\tau) = \begin{cases} \sigma_\epsilon^2 (1 + \theta_1^2) & \text{si } \tau = 0 \\ \sigma_\epsilon^2 \theta_1 & \text{si } |\tau| = 1 \\ 0 & \text{si } |\tau| > 1 \end{cases}$$

mientras que la sucesión de autocorrelaciones teórica (ACF) está dada por:

$$\rho_Y(\tau) = \begin{cases} 1 & \text{si } \tau = 0 \\ \frac{\theta_1}{1 + \theta_1^2} & \text{si } |\tau| = 1 \\ 0 & \text{si } |\tau| > 1 \end{cases}$$

Podemos concluir que el proceso $\{Y_t\}$ es un proceso estacionario ya que $\mathbb{E}(Y_t)$, $\mathbb{E}(Y_t^2)$ y $\gamma_Y(\tau)$ son finitos y no dependen del tiempo.

Ejemplo 3.3.1 Demuestre que para un proceso $MA(1)$, $-\frac{1}{2} \leq \rho_1 \leq \frac{1}{2}$.

Tomemos a partir de la sucesión de autocorrelaciones teórica (ACF) del proceso $MA(1)$ para $|\tau|$ a ρ_1 , la cual se puede expresar como $\rho_1(1 + \theta_1^2) - \theta_1 = 0$, con raíces:

$$\theta_1 = \frac{1 \pm \sqrt{1 - 4\rho_1^2}}{2\rho_1}$$

dadas por la fórmula general para resolver ecuaciones de segundo grado. Para que las raíces sean reales, tomamos en cuenta el discriminante, es decir, $1 - 4\rho_1^2 \geq 0 \Leftrightarrow \rho_1^2 \leq \frac{1}{4} \Leftrightarrow |\rho_1| \leq \frac{1}{2}$.

$\therefore -\frac{1}{2} \leq \rho_1 \leq \frac{1}{2}$, que es lo que queríamos demostrar.

Con este ejemplo hemos querido ilustrar lo siguiente. Aún cuando para $|\tau| = 1$, la autocorrelación es distinta de cero, dicha autocorrelación no puede ser muy elevada, ya que eso implicaría una fuerte dependencia de la observación actual con la observación anterior, y de aquí que sería más adecuado pensar en un proceso autorregresivo (el cual veremos a detalle en la **Sección 3.4**) para dicha situación.

Ejemplo 3.3.2 La **Figura 3.3.1** muestra la simulación de dos procesos $MA(1)$ de 200 observaciones con parámetros $\theta_1 = 0.7$ y $\theta_1 = -0.7$ respectivamente. Notemos que la observación X_t está correlacionada con la observación X_{t-1} pero no lo está con las observaciones $X_{t-2}, X_{t-3}, X_{t-4}, \dots$. En la figura, ACF denota la sucesión de autocorrelaciones $\rho_X(\tau)$ (eje Y) para distintos valores de τ (eje X). El caso contrario se dará cuando veamos, en la **Sección 3.4**, a los procesos autorregresivos $AR(p)$ y en particular al modelo $AR(1)$, en el cual, la correlación que existe entre X_t y $X_{t-\tau}$ nunca será cero.

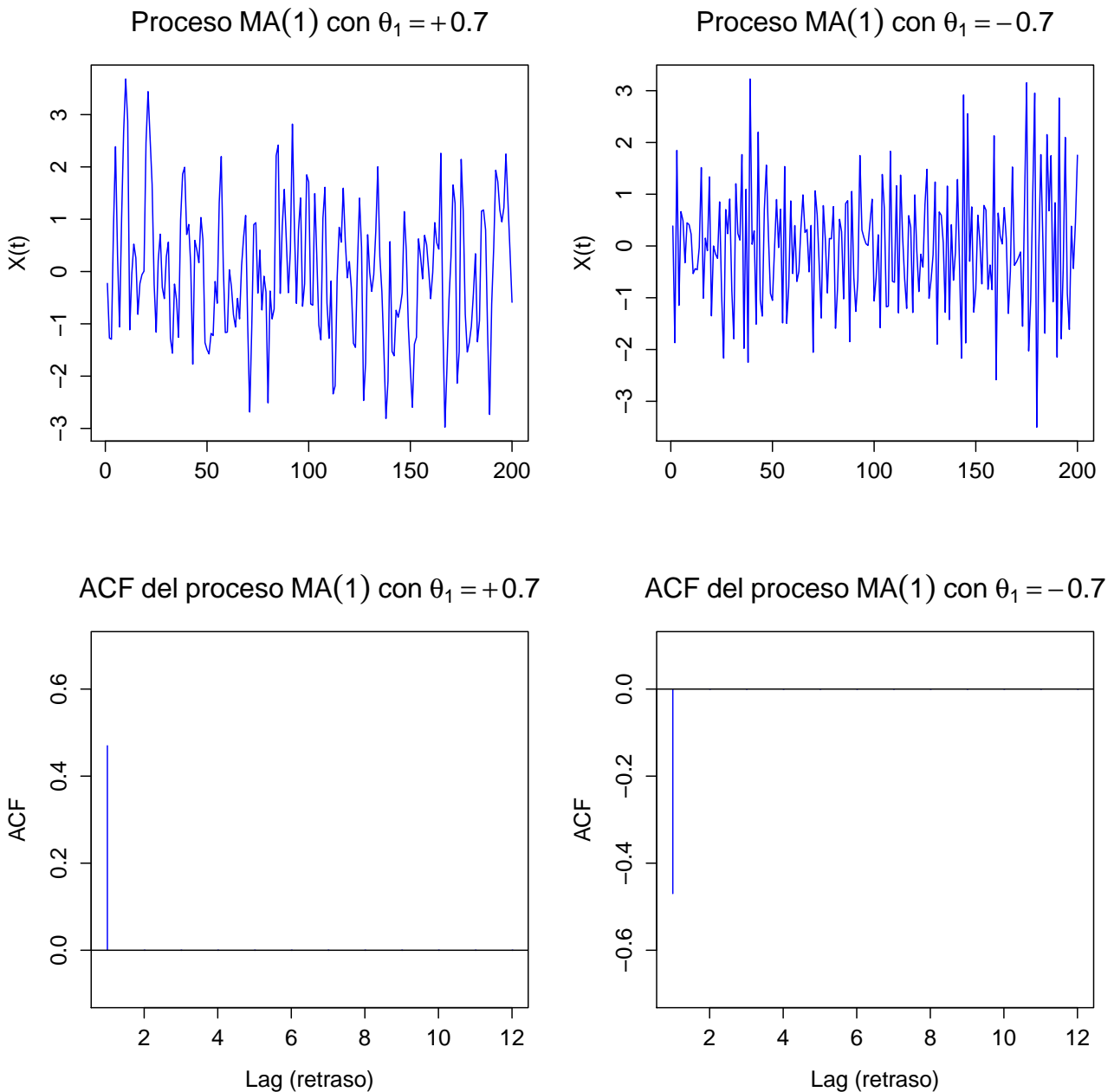


Figura 3.3.1: Figura que muestra dos procesos $MA(1)$ simulados con $\theta_1 = 0.7$ y $\theta_1 = -0.7$

La correspondiente sucesión de autocorrelaciones teórica (ACF) del proceso MA(1) simulado con parámetro $\theta_1 = 0.7$ y el proceso MA(1) simulado con parámetro $\theta_1 = -0.7$, también se muestra en la **Figura 3.3.1**. La simulación de ambos procesos la podemos obtener en R con el siguiente código:

```
> par(mfrow=c(2,2))
> plot(arima.sim(list(order=c(0,0,1),ma=.7),n=200),xlab="",ylab="X(t)",col="blue",
      main=(expression(Proceso~MA(1)~con~theta[1]==+.7)))
> plot(arima.sim(list(order=c(0,0,1),ma=-.7),n=200),xlab="",ylab="X(t)",col="blue",
      main=(expression(Proceso~MA(1)~con~theta[1]==-.7)))
> ACF1=ARMAacf(ar=0,ma=.7,12)[-1]
> ACF2=ARMAacf(ar=0,ma=-.7,12)[-1]
> plot(ACF1,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.1,0.7),col="blue",
      main=(expression(ACF~del~proceso~MA(1)~con~theta[1]==+.7)))
> abline(h=0)
> plot(ACF2,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.7,0.1),col="blue",
      main=(expression(ACF~del~proceso~MA(1)~con~theta[1]==-.7)))
> abline(h=0)
```

3.3.2. Modelo MA(2)

Un proceso MA(2) viene dado por la relación:

$$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} \quad (3.3.5)$$

donde μ, θ_1 y $\theta_2 \in \mathbb{R}$ y el proceso $\{\epsilon_t\}$ es un ruido blanco con $\mathbb{E}(\epsilon_t) = 0$ y $Var(\epsilon_t) = \sigma_\epsilon^2$. Tomemos nuevamente el proceso centrado $\{Y_t\} = \{X_t - \mu\}$, es decir, obtenemos la nueva relación dada por $Y_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$. A continuación calcularemos el primer momento, la varianza y la sucesión de autocovarianzas teórica correspondientes al proceso MA(2). Comencemos con el primer momento. La esperanza del proceso $\{Y_t\} = \{X_t - \mu\}$ está dada por:

$$\begin{aligned} \mathbb{E}(Y_t) &= \mathbb{E}(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}) \\ &= \mathbb{E}(\epsilon_t) + \theta_1\mathbb{E}(\epsilon_{t-1}) + \theta_2\mathbb{E}(\epsilon_{t-2}) \\ &= 0 \end{aligned} \quad (3.3.6)$$

ya que $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. La varianza se obtiene a partir de:

$$\begin{aligned} Var(Y_t) &= Var(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}) \\ &= Var(\epsilon_t) + \theta_1^2 Var(\epsilon_{t-1}) + \theta_2^2 Var(\epsilon_{t-2}) \\ &= \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2) \end{aligned} \quad (3.3.7)$$

recordando que el proceso $\{\epsilon_t\}$ representa una sucesión de variables aleatorias no correlacionadas con media 0 y varianza constante e igual a σ_ϵ^2 y se podría asumir que hay un término adicional $\theta_0 = 1$. A continuación, construiremos la sucesión de autocovarianzas teórica y veremos los valores de τ para los cuales dicha sucesión es distinta de cero.

Para comenzar dicha construcción tenemos la relación siguiente:

$$\begin{aligned}
\gamma_Y(\tau) &= \text{Cov}(Y_t, Y_{t+\tau}) \\
&= \text{Cov}(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}, \epsilon_{t+\tau} + \theta_1\epsilon_{t+\tau-1} + \theta_2\epsilon_{t+\tau-2}) \\
&= \text{Cov}(\epsilon_t, \epsilon_{t+\tau}) + \theta_1\text{Cov}(\epsilon_t, \epsilon_{t+\tau-1}) + \theta_2\text{Cov}(\epsilon_t, \epsilon_{t+\tau-2}) \\
&\quad + \theta_1\text{Cov}(\epsilon_{t-1}, \epsilon_{t+\tau}) + \theta_1^2\text{Cov}(\epsilon_{t-1}, \epsilon_{t+\tau-1}) + \theta_1\theta_2\text{Cov}(\epsilon_{t-1}, \epsilon_{t+\tau-2}) \\
&\quad + \theta_2\text{Cov}(\epsilon_{t-2}, \epsilon_{t+\tau}) + \theta_2\theta_1\text{Cov}(\epsilon_{t-2}, \epsilon_{t+\tau-1}) + \theta_2^2\text{Cov}(\epsilon_{t-2}, \epsilon_{t+\tau-2})
\end{aligned} \tag{3.3.8}$$

o bien:

$$\gamma_Y(\tau) = \begin{cases} \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2) & \text{si } \tau = 0 \\ \sigma_\epsilon^2(\theta_1 + \theta_1\theta_2) & \text{si } \tau = 1 \\ \sigma_\epsilon^2\theta_2 & \text{si } \tau = 2 \\ 0 & \text{si } \tau \geq 3 \end{cases}$$

La sucesión de autocorrelaciones teórica (ACF) vendrá dada entonces por:

$$\rho_Y(\tau) = \begin{cases} 1 & \text{si } \tau = 0 \\ \frac{\theta_1(1+\theta_2)}{1+\theta_1^2+\theta_2^2} & \text{si } \tau = 1 \\ \frac{\theta_2}{1+\theta_1^2+\theta_2^2} & \text{si } \tau = 2 \\ 0 & \text{si } \tau \geq 3 \end{cases}$$

Ejemplo 3.3.3 A partir de la sucesión de autocorrelaciones teórica (ACF) de un proceso MA(2), demuestre que se satisface $\rho_1^2 \leq 0.5$ y $|\rho_2| \leq 0.5$.

Tomemos de la sucesión de autocorrelaciones teórica (ACF) para un proceso MA(2) a ρ_2 y expresémosla como $\theta_2^2\rho_2 - \theta_2 + \rho_2(1 + \theta_1^2) = 0$. Utilizando la fórmula general para resolver una ecuación cuadrática en función de θ_2 considerando a θ_1 como valor constante tenemos:

$$\theta_2 = \frac{1 \pm \sqrt{1 - 4\rho_2^2(1 + \theta_1^2)}}{2\rho_2}$$

Para que las raíces de θ_2 sean reales se requiere que el discriminante de la fórmula general, $4\rho_2^2(1 + \theta_1^2) \leq 1$, lo cual implica que $1 \leq 4\rho_2^2(1 + \theta_1^2) \leq 1$. Despejando a ρ_2 de la relación anterior obtenemos parte de lo que queríamos demostrar, a saber, $|\rho_2| \leq 0.5$.

Ahora tomemos de la sucesión de autocorrelaciones teórica (ACF) a ρ_1 y expresémoslo como, $\theta_1^2\rho_1 - \theta_1(1 + \theta_2) + \rho_1(1 + \theta_2^2) = 0$. Utilizando nuevamente la fórmula general para resolver una ecuación cuadrática en función de θ_1 considerando ahora a θ_2 un valor constante tenemos:

$$\theta_1 = \frac{(1 + \theta_2) \pm \sqrt{(1 + \theta_2)^2 - 4\rho_1^2(1 + \theta_2^2)}}{2\rho_1}$$

Para que las raíces de θ_1 sean reales se requiere que $(1 + \theta_2)^2 \geq 4\rho_1^2(1 + \theta_2^2)$, o bien, manipulando un poco la ecuación anterior obtenemos:

$$4\rho_1^2 - 1 \leq \frac{2\theta_2}{1 + \theta_2^2}$$

Hemos visto anteriormente en el **Ejemplo 3.3.1.** que para un proceso $MA(1)$ se cumple $\rho_1 \leq 0.5$ (esta ρ_1 es aquella del proceso $MA(1)$). Utilizando este hecho y la ecuación anterior tenemos:

$$4\rho_1^2 - 1 \leq \frac{2\theta_2}{1 + \theta_2^2} \leq 2(0.5)$$

Despejando a ρ_1^2 de la relación anterior obtenemos lo que nos faltaba por demostrar, a saber, $\rho_1^2 \leq 0.5$.

Ejemplo 3.3.4 Pasemos a simular en R un proceso $MA(2)$ con parámetros $\theta_1 = 0.3$, $\theta_2 = -0.9$ y 250 observaciones. También obtendremos la correspondiente sucesión de autocorrelaciones teórica (ACF). La **Figura 3.3.2** muestra tanto el proceso simulado $MA(2)$ como la respectiva sucesión de autocorrelaciones teórica (ACF). El código necesario será entonces:

```
> par(mfrow=c(1,2))
> MA2=arima.sim(list(order=c(0,0,2),ma=c(0.3,-0.9)),n=250)
> plot(MA2,xlab="",ylab="X(t)",col="blue",
      main=(expression(MA(2)~con~theta[1]==+.3~y~theta[2]==-.9)))
> ACF=ARMAacf(ar=0,ma=c(0.3,0.9),12)[-1]
> plot(ACF,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.1,1),col="blue",
      main=(expression(ACF~teórica~del~proceso~MA(2))))
> abline(h=0)
```

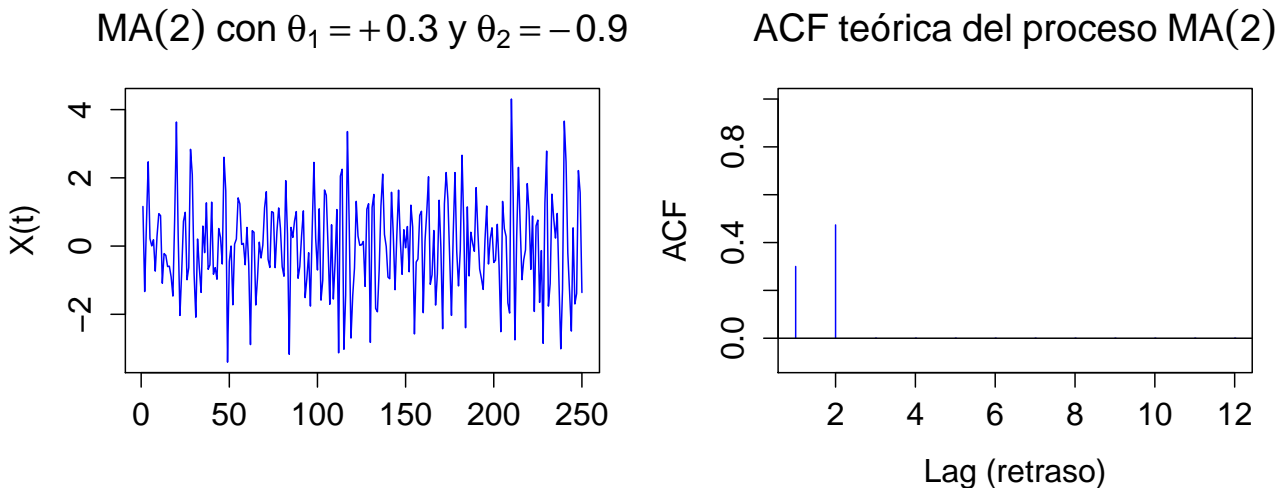


Figura 3.3.2: Figura que muestra un proceso $MA(2)$ simulado con $\theta_1 = 0.3$ y $\theta_2 = -0.9$

3.3.3. Modelo MA(q)

Decimos que el proceso $\{X_t\}$ es de tipo promedios móviles de orden q y media μ si $\{X_t\}$ se puede escribir como:

$$X_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q} \quad (3.3.9)$$

con $\mu, \theta_1, \theta_2, \dots, \theta_q \in \mathbb{R}$ y donde $\{\epsilon_t\}$ es un ruido blanco con $\mathbb{E}(\epsilon_t) = 0$ y $Var(\epsilon_t) = \sigma_\epsilon^2$. A este proceso se le denota como MA(q).

Tomemos el proceso centrado $\{Y_t\} = \{X_t - \mu\}$ (es centrado ya que su esperanza es cero). Para demostrar que el proceso MA(q) es estacionario, pasemos a calcular su primer y segundo momento. Obtendremos como siempre la sucesión de autocovarianzas y autocorrelaciones teórica asociadas al proceso.

La esperanza está dada por:

$$\begin{aligned} \mathbb{E}(Y_t) &= \mathbb{E}(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q}) \\ &= \mathbb{E}(\epsilon_t) + \theta_1\mathbb{E}(\epsilon_{t-1}) + \theta_2\mathbb{E}(\epsilon_{t-2}) + \dots + \theta_q\mathbb{E}(\epsilon_{t-q}) \\ &= 0 \end{aligned} \quad (3.3.10)$$

ya que la media del proceso $\{\epsilon_t\}$ es cero. Para la varianza, recordemos que el proceso $\{\epsilon_t\}$ es un conjunto de variables aleatorias no correlacionadas. Por tanto tenemos:

$$\begin{aligned} Var(Y_t) &= Var(\epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \dots + \theta_q\epsilon_{t-q}) \\ &= Var(\epsilon_t) + \theta_1^2 Var(\epsilon_{t-1}) + \theta_2^2 Var(\epsilon_{t-2}) + \dots + \theta_q^2 Var(\epsilon_{t-q}) \\ &= \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \end{aligned} \quad (3.3.11)$$

Asumiendo $\theta_0 = 1$ tenemos que el segundo momento viene expresado por:

$$\begin{aligned} \mathbb{E}(Y_t^2) &= Var(Y_t) + \mathbb{E}^2(Y_t) \\ &= Var(Y_t) \\ &= \sigma_\epsilon^2(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \end{aligned} \quad (3.3.12)$$

Obtengamos a continuación la sucesión de autocovarianzas teórica. Dicha sucesión puede verse como $\gamma_X(\tau) = Cov(X_t, X_{t+\tau}) = \mathbb{E}[(X_t - \mu)(X_{t+\tau} - \mu)] = \mathbb{E}(Y_t Y_{t+\tau})$. Suponiendo $\theta_0 = 1$ tenemos:

$$\begin{aligned} Y_t Y_{t+\tau} &= \left(\sum_{j=0}^q \theta_j \epsilon_{t-j} \right) \left(\sum_{l=0}^q \theta_l \epsilon_{t+\tau-l} \right) \\ &= \sum_{j=0}^q \sum_{l=0}^q \theta_j \theta_l \epsilon_{t-j} \epsilon_{t+\tau-l} \end{aligned}$$

Calculando su esperanza tenemos:

$$\begin{aligned}
\mathbb{E}(Y_t Y_{t+\tau}) &= \sum_{j=0}^q \sum_{l=0}^q \theta_j \theta_l \mathbb{E}(\epsilon_{t-j} \epsilon_{t+\tau-l}) \\
&= \sum_{j=0}^q \sum_{l=0}^q \theta_j \theta_l [\mathbb{E}(\epsilon_{t-j} \epsilon_{t+\tau-l}) - \mathbb{E}(\epsilon_{t-j}) \mathbb{E}(\epsilon_{t+\tau-l})] \\
&= \sum_{j=0}^q \sum_{l=0}^q \theta_j \theta_l \gamma_{\tau-l+j}
\end{aligned} \tag{3.3.13}$$

Hemos visto que para un ruido blanco, $\gamma_X(\tau)$ vale σ_X^2 cuando $\tau = 0$, en este caso va a suceder si $\tau - l + j = 0$, es decir, $l = \tau + j$. Como $l \leq q$ entonces $\tau + j \leq q$ y como $j \geq 0$ entonces $0 \leq j \leq q - \tau$.

Por tanto, la sucesión de autocovarianzas teórica es:

$$\gamma_Y(\tau) = \mathbb{E}(Y_t Y_{t+\tau}) = \begin{cases} \left(\sum_{j=0}^{q-\tau} \theta_j \theta_{j+\tau} \right) \sigma_\epsilon^2 & \text{si } 0 \leq \tau \leq q \\ 0 & \text{si } \tau > q \end{cases}$$

Hemos demostrado en la **Sección 3.2** que $\gamma_X(\tau) = \gamma_X(-\tau)$, por lo que la sucesión de autocovarianzas teórica está dada finalmente por:

$$\gamma_Y(|\tau|) = \mathbb{E}(Y_t Y_{t+|\tau|}) = \begin{cases} \left(\sum_{j=0}^{q-|\tau|} \theta_j \theta_{j+|\tau|} \right) \sigma_\epsilon^2 & \text{si } 0 \leq |\tau| \leq q \\ 0 & \text{si } |\tau| > q \end{cases}$$

La sucesión de autocorrelaciones teórica (ACF) está dada por:

$$\rho_Y(|\tau|) = \begin{cases} \frac{\sum_{j=0}^{q-|\tau|} \theta_j \theta_{|\tau|+j}}{\sum_{j=0}^q \theta_j^2} & \text{si } 0 \leq |\tau| \leq q \\ 0 & \text{si } |\tau| > q \end{cases}$$

Se concluye que un MA(q) es un proceso estacionario ya que al calcular $\mathbb{E}(Y_t)$, $\mathbb{E}(Y_t^2)$ y la sucesión de autocovarianzas teórica $Cov(Y_t, Y_{t+\tau}) = \gamma_Y(\tau)$, hemos demostrado que son expresiones que no dependen del tiempo.

3.4. Modelos autorregresivos (AR)

3.4.1. Modelo AR(1)

Un proceso AR(1) está dado por:

$$X_t = \mu + \phi_1 X_{t-1} + \epsilon_t \quad (3.4.1)$$

donde μ y $\phi_1 \in \mathbb{R}$ y el proceso $\{\epsilon_t\}$ es un ruido blanco con media 0 y varianza constante e igual a σ_ϵ^2 . Tomemos el proceso centrado $\{Y_t\} = \{X_t - \mu\}$, es decir, $Y_t = \phi_1 Y_{t-1} + \epsilon_t$ con $|\phi_1| < 1$. El valor esperado del proceso $\{Y_t\}$ será:

$$\begin{aligned} \mathbb{E}(Y_t) &= \mathbb{E}(\phi_1 Y_{t-1} + \epsilon_t) \\ &= \phi_1 \mathbb{E}(Y_t) \\ &= 0 \end{aligned} \quad (3.4.2)$$

Nuestra siguiente tarea será encontrar la sucesión de autocovarianzas teórica del proceso, en la cual se necesitará la expresión de la varianza. Para el cálculo de la sucesión de autocovarianzas teórica para $\tau > 1$ tenemos la siguiente relación:

$$\begin{aligned} \gamma_Y(\tau) &= \text{Cov}(Y_t, Y_{t+\tau}) \\ &= \text{Cov}(Y_t, Y_{t-\tau}) \\ &= \text{Cov}(\phi_1 Y_{t-1} + \epsilon_t, Y_{t-\tau}) \\ &= \phi_1 \text{Cov}(Y_{t-1}, Y_{t-\tau}) + \text{Cov}(\epsilon_t, Y_{t-\tau}) \\ &= \phi_1 \text{Cov}(Y_{t-1}, Y_{t-\tau}) \\ &= \phi_1 \gamma_Y(\tau - 1) \\ &= \phi_1^\tau \gamma_Y(0) \end{aligned} \quad (3.4.3)$$

donde claro, $\text{Cov}(\epsilon_t, Y_{t-\tau}) = 0$ ya que el proceso $\{\epsilon_t\}$ no está correlacionado con el proceso Y_s para cada $s < t$. Para $\tau = 0$, la varianza del proceso, tenemos:

$$\begin{aligned} \gamma_Y(0) &= \text{Cov}(Y_t, Y_t) \\ &= \text{Cov}(\phi_1 Y_{t-1} + \epsilon_t, \phi_1 Y_{t-1} + \epsilon_t) \\ &= \phi_1^2 \text{Cov}(Y_{t-1}, Y_{t-1}) + \text{Cov}(\epsilon_t, \epsilon_t) \\ &= \phi_1^2 \text{Cov}(Y_0, Y_0) + \sigma_\epsilon^2 \\ &= \phi_1^2 \gamma_Y(0) + \sigma_\epsilon^2 \end{aligned} \quad (3.4.4)$$

donde, $\text{Cov}(Y_{t-1}, Y_{t-1}) = \text{Cov}(Y_0, Y_0)$ ya que el proceso $\{Y_t\}$ es estacionario. Despejando $\gamma_Y(0)$ de la **ec. 3.4.4** podemos obtener la varianza del proceso $\{Y_t\} = \{X_t - \mu\}$, es decir, obtenemos:

$$\gamma_Y(0) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \quad (3.4.5)$$

La sucesión de autocovarianzas teórica es entonces:

$$\gamma_Y(|\tau|) = \frac{\phi_1^{|\tau|} \sigma_\epsilon^2}{1 - \phi_1^2} \quad (3.4.6)$$

mientras que la sucesión de autocorrelaciones teórica (ACF) está dada por:

$$\rho_Y(|\tau|) = \phi_1^{|\tau|} \quad (3.4.7)$$

lo cual indica que conforme $\tau > 0$ crece, la sucesión de autocorrelaciones teórica (ACF) tiende a cero, con decaimiento exponencial cuando $0 < \phi_1 < 1$ y con signos alternados cuando $-1 < \phi_1 < 0$. Observar que de $\rho_Y(|\tau|) = \phi_1^{|\tau|}$ se obtiene que $\rho_Y(1) = \phi_1$, es decir, el coeficiente ϕ_1 es la autocorrelación a un paso.

Ejemplo 3.4.1 Realizaremos a continuación una simulación de dos procesos $AR(1)$ en R , con parámetros, $\phi_1 = 0.7$ y $\phi_1 = -0.7$, cada uno con 150 observaciones.

La sucesión de autocorrelaciones teórica (ACF) para el primer proceso de parámetro $\phi_1 = 0.7$ está dada por $\rho_Y(\tau) = \phi_1^\tau = (0.7)^\tau$, por lo que observaciones contiguas en el tiempo estarán muy cerca una de otra.

Para el segundo proceso $AR(1)$ con parámetro $\phi_1 = -0.7$ tenemos que $\rho_Y(\tau) = \phi_1^\tau = (-0.7)^\tau$, por lo cual, observaciones contiguas estarán correlacionadas de manera negativa, es decir, si X_t es positiva entonces X_{t+1} es negativa, pero X_{t+2} será de nuevo positiva. Dichas aseveraciones se verán de manera clara en la **Figura 3.4.1**.

También, en la **Figura 3.4.1**, se muestra la correspondiente sucesión de autocorrelaciones teórica (ACF) para el proceso $AR(1)$ simulado en R con parámetro $\phi_1 = 0.7$ y para el proceso $AR(1)$ simulado con parámetro $\phi_1 = -0.7$.

Observamos que la primer sucesión de autocorrelaciones teórica (ACF), es decir aquella con parámetro $0 < \phi_1 = 0.7 < 1$, decrece de manera exponencial, mientras que la segunda sucesión de autocorrelaciones teórica (ACF), aquella con parámetro $-1 < \phi_1 = -0.7 < 0$, lo hace con signos alternados.

El código para este ejemplo será entonces:

```
> par(mfrow=c(2,2))
> plot(arima.sim(list(order=c(1,0,0),ar=.7),n=150),xlab="",ylab="X(t)",col="blue",
  main=(expression(Proceso~AR(1)~con~phi[1]==+.7)))
> plot(arima.sim(list(order=c(1,0,0),ar=-.7),n=150),xlab="",ylab="X(t)",col="blue",
  main=(expression(Proceso~AR(1)~con~phi[1]==-.7)))
> ACF1=ARMAacf(ar=.7,ma=0,15)[-1]
> ACF2=ARMAacf(ar=-.7,ma=0,15)[-1]
> plot(ACF1,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.1,1),col="blue",
  main=(expression(ACF~del~proceso~AR(1)~con~phi[1]==+.7)))
> abline(h=0)
> plot(ACF2,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.8,1),col="blue",
  main=(expression(ACF~del~proceso~AR(1)~con~phi[1]==-.7)))
> abline(h=0)
```

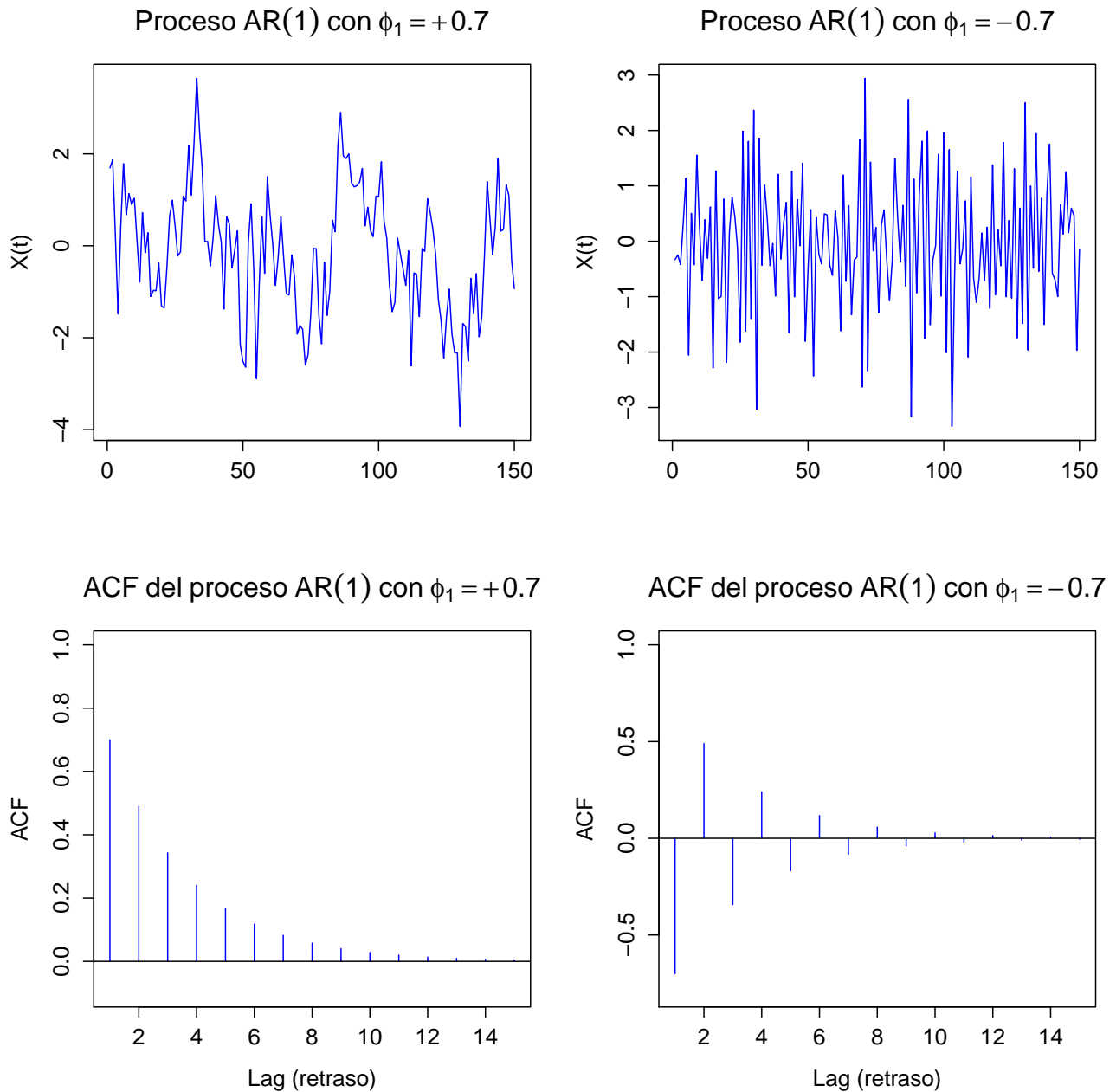


Figura 3.4.1: Figura que muestra dos procesos AR(1) simulados con $\phi_1 = 0.7$ y $\phi_1 = -0.7$

3.4.2. Modelo AR(2)

Un proceso AR(2) está dado por:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t \quad (3.4.8)$$

Tomemos nuevamente el proceso centrado $\{Y_t\} = \{X_t - \mu\}$, es decir, $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$ con $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. Para que el proceso anterior sea estacionario debe cumplirse que las raíces del polinomio $\phi(z) = 1 - \phi_1 z - \phi_2 z^2$ se encuentren fuera del círculo unitario. Esta proposición esta basada en el **Teorema 3.5.1** el cual estudiaremos a profundidad cuando estemos en la **Sección 3.5.1**.

Una forma alternativa para demostrar estacionariedad es a partir del **Teorema de Schur** el cual nos dice que el valor absoluto de las raíces del polinomio característico de orden p dado por la siguiente relación:

$$x^p - \alpha_1 x^{p-1} - \alpha_2 x^{p-2} - \dots - \alpha_{p-1} x - \alpha_p = 0 \quad (3.4.9)$$

serán menores a uno si y sólo si:

$$D_1 = \begin{vmatrix} -1 & \alpha_p \\ \alpha_p & -1 \end{vmatrix} > 0$$

$$D_2 = \begin{vmatrix} -1 & 0 & \alpha_p & \alpha_{p-1} \\ \alpha_{p-1} & -1 & 0 & \alpha_p \\ \alpha_p & 0 & -1 & \alpha_{p-1} \\ \alpha_{p-1} & \alpha_p & 0 & -1 \end{vmatrix} > 0$$

y así sucesivamente hasta el determinante:

$$D_p = \begin{vmatrix} -1 & 0 & \dots & 0 & \alpha_p & \alpha_{p-1} & \dots & \alpha_1 \\ \alpha_1 & -1 & \dots & 0 & 0 & \alpha_p & \dots & \alpha_2 \\ \vdots & & & \vdots & \vdots & & & \vdots \\ \alpha_{p-1} & \alpha_{p-2} & \dots & -1 & 0 & 0 & \dots & \alpha_p \\ \alpha_p & 0 & \dots & 0 & -1 & \alpha_1 & \dots & \alpha_{p-1} \\ \alpha_{p-1} & \alpha_p & \dots & 0 & 0 & -1 & \dots & \alpha_{p-2} \\ \vdots & & & \vdots & \vdots & & & \vdots \\ \alpha_1 & \alpha_2 & \dots & \alpha_p & 0 & 0 & \dots & -1 \end{vmatrix} > 0$$

Para el proceso AR(2) bastará con demostrar que:

$$D_1 = \begin{vmatrix} -1 & \phi_2 \\ \phi_2 & -1 \end{vmatrix} = 1 - \phi_2^2 > 0 \quad (3.4.10)$$

y

$$D_2 = \begin{vmatrix} -1 & 0 & \phi_2 & \phi_1 \\ \phi_1 & -1 & 0 & \phi_2 \\ \phi_2 & 0 & -1 & \phi_1 \\ \phi_1 & \phi_2 & 0 & -1 \end{vmatrix} = (1 + \phi_2)^2 ((1 - \phi_2)^2 - \phi_1^2) > 0 \quad (3.4.11)$$

La condición $D_1 > 0$ es equivalente a pedir que $|\phi_2| < 1$, mientras que la condición $D_2 > 0$ es equivalente a pedir que $(1 - \phi_2)^2 > \phi_1^2$. Esto nos facilitará el trabajo al momento de demostrar estacionariedad para un proceso AR(2).

Es decir, para que un proceso AR(2) sea estacionario basta con que se cumplan las siguientes condiciones:

$$\begin{aligned} |\phi_2| &< 1 \\ \phi_2 + \phi_1 &< 1 \\ \phi_2 - \phi_1 &< 1 \end{aligned} \tag{3.4.12}$$

Ahora bien, a partir de las ecuaciones de Yule-Walker, las cuales se verán en la **Sección 3.4.3**, obtenemos:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned}$$

las cuales permiten obtener los valores de ρ_1 y ρ_2 en función de los parámetros autorregresivos ϕ_1 y ϕ_2 respectivamente. Resolviendo las ecuaciones de Yule-Walker para ρ_1 y ρ_2 obtenemos:

$$\begin{aligned} \rho_1 &= \frac{\phi_1}{1 - \phi_2} \\ \rho_2 &= \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \end{aligned}$$

Como comentario final, al tomar el polinomio característico del proceso AR(2) haremos uso, en la mayoría de las ocasiones, de la fórmula general para la obtención de las raíces del polinomio $\phi(z)$.

Especial atención pondremos en el discriminante ya que si $\phi_1^2 + 4\phi_2 \geq 0$, entonces, las raíces de $\phi(z)$ son reales y la sucesión de autocorrelaciones teórica (ACF) decaerá rápidamente a cero (todas las autocorrelaciones serán positivas si la primera lo es y tendrán signos alternados si la primer autocorrelación es negativa). Ahora bien, si sucede lo contrario, es decir, $\phi_1^2 + 4\phi_2 < 0$, las raíces de $\phi(z)$ serán complejas y la sucesión de autocorrelaciones teórica (ACF) seguirá un comportamiento sinusoidal convergente a cero.

Ejemplo 3.4.2 Como en cada modelo, pasemos a simular ahora un proceso AR(2) en R con parámetros $\phi_1 = 0.7$, $\phi_2 = -0.3$ y 250 observaciones. La **Figura 3.4.2** nos muestra el proceso simulado AR(2) en R con dichos parámetros. De igual manera obtendremos la sucesión de autocorrelaciones teórica (ACF) del proceso AR(2) simulado, la cual puede observarse también en la **Figura 3.4.2**. Observamos de la sucesión de autocorrelaciones teórica (ACF) el comportamiento sinusoidal convergente a cero del que habíamos comentado ya que $\phi_1^2 + 4\phi_2 = -0.71 < 0$.

El código correspondiente para la simulación, gráfica y sucesión de autocorrelaciones teórica (ACF) del proceso estará dado por:

```
> par(mfrow=c(1,2))
> AR2=arima.sim(list(order=c(2,0,0),ar=c(0.7,-0.3)),n=250)
> plot(AR2,xlab="",ylab="X(t)",col="blue",
      main=(expression(AR(2)~con~phi[1]==+.7~y~phi[2]==-.3)))
> ACF=ARMAacf(ar=c(.7,-0.3),ma=0,15)[-1]
> plot(ACF,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.1,1),col="blue",
      main=(expression(ACF~teórica~del~proceso~AR(2))))
> abline(h=0)
```

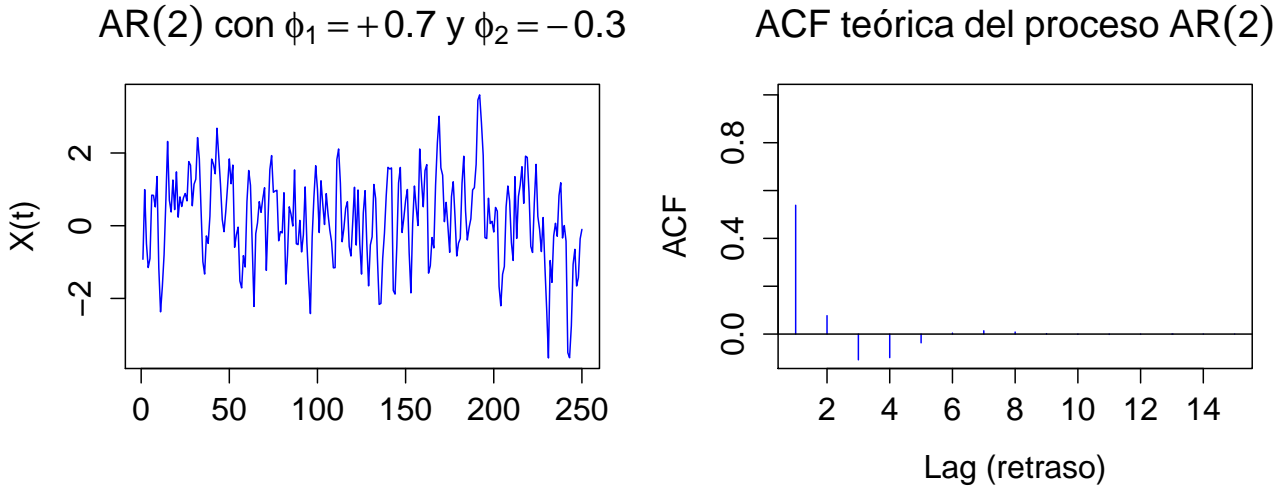


Figura 3.4.2: Figura que muestra un proceso AR(2) simulado con $\phi_1 = 0.7$ y $\phi_2 = -0.3$

3.4.3. Modelo AR(p)

Decimos que el proceso $\{X_t\}$ es autorregresivo de orden p y media μ si $\{X_t\}$ se puede escribir como:

$$X_t = \mu + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \epsilon_t \quad (3.4.13)$$

con $\mu, \phi_1, \phi_2, \dots, \phi_p \in \mathbb{R}$ y $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. A este proceso se le denota AR(p). De nuevo, tomaremos al proceso centrado $\{Y_t\} = \{X_t - \mu\}$, es decir:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (3.4.14)$$

La obtención de fórmulas más generales se pueden obtener de las ecuaciones de Yule-Walker para AR(p). Nuestra pregunta inmediata será entonces, ¿cómo llegar a dichas ecuaciones? Partimos de un proceso AR(p) dado por la **ec. 3.4.14**, suponiendo claro que el proceso $\{\epsilon_t\}$ es un ruido blanco con media 0 y varianza constante e igual a σ_ϵ^2 . Además supondremos causal al proceso $\{Y_t\}$. La definición de causalidad se dará en la **Sección 3.5.1** pero por el momento bastará explicar, en palabras, que un proceso $\{Y_t\}$ es causal si a tiempo t , no depende del futuro. Pasamos ahora a multiplicar la **ec. 3.4.14** por el término Y_{t-k} con $1 \leq k \leq p$, para después aplicar el operador esperanza, obteniendo:

$$\mathbb{E}(Y_t Y_{t-k}) = \phi_1 \mathbb{E}(Y_{t-1} Y_{t-k}) + \dots + \phi_p \mathbb{E}(Y_{t-p} Y_{t-k}) + \mathbb{E}(\epsilon_t Y_{t-k}) \quad (3.4.15)$$

Ahora bien, es posible demostrar, a partir de la definición de causalidad, que la esperanza del proceso $\{Y_t\}$ es 0 y así poder ver a $\mathbb{E}(Y_t Y_{t-k})$ como:

$$\begin{aligned} \mathbb{E}(Y_t Y_{t-k}) &= \mathbb{E}(Y_t Y_{t-k}) - \mathbb{E}(Y_t) \mathbb{E}(Y_{t-k}) \\ &= \text{Cov}(Y_t, Y_{t-k}) \\ &= \gamma_Y(k) \end{aligned}$$

Así como obtuvimos $\mathbb{E}(Y_t Y_{t-k}) = \gamma_Y(k)$, podemos obtener también:

$$\begin{aligned}\mathbb{E}(Y_{t-1} Y_{t-k}) &= \gamma_Y(k-1) \\ \mathbb{E}(Y_{t-2} Y_{t-k}) &= \gamma_Y(k-2) \\ \mathbb{E}(Y_{t-3} Y_{t-k}) &= \gamma_Y(k-3) \\ &\vdots \\ \mathbb{E}(Y_{t-p} Y_{t-k}) &= \gamma_Y(k-p)\end{aligned}$$

De tal forma que la **ec. 3.4.15** se convierte en:

$$\gamma_Y(k) - \phi_1 \gamma_Y(k-1) - \dots - \phi_p \gamma_Y(k-p) = \mathbb{E}(\epsilon_t Y_{t-k}) \quad (3.4.16)$$

Nos resta por saber lo que sucede con el término $\mathbb{E}(\epsilon_t Y_{t-k})$. Es posible demostrar, de nuevo a partir de la definición de causalidad (la cual se dará en la **Sección 3.5.1**), que $\mathbb{E}(\epsilon_t Y_{t-k}) = 0$. Al sustituir $\mathbb{E}(\epsilon_t Y_{t-k}) = 0$ en la **ec. 3.4.16** obtenemos, para $1 \leq k \leq p$, las siguientes expresiones:

$$\begin{aligned}k=1 \quad \gamma_Y(1) - \phi_1 \gamma_Y(0) - \dots - \phi_p \gamma_Y(p-1) &= 0 \\ k=2 \quad \gamma_Y(2) - \phi_1 \gamma_Y(1) - \dots - \phi_p \gamma_Y(p-2) &= 0 \\ k=3 \quad \gamma_Y(3) - \phi_1 \gamma_Y(2) - \dots - \phi_p \gamma_Y(p-3) &= 0 \\ &\vdots \\ k=p \quad \gamma_Y(p) - \phi_1 \gamma_Y(p-1) - \dots - \phi_p \gamma_Y(0) &= 0\end{aligned} \quad (3.4.17)$$

Dividiendo entre $\gamma_Y(0)$ obtenemos las ecuaciones de Yule-Walker:

$$\begin{aligned}k=1 \quad \rho_Y(1) - \phi_1 - \dots - \phi_p \rho_Y(p-1) &= 0 \\ k=2 \quad \rho_Y(2) - \phi_1 \rho_Y(1) - \dots - \phi_p \rho_Y(p-2) &= 0 \\ k=3 \quad \rho_Y(3) - \phi_1 \rho_Y(2) - \dots - \phi_p \rho_Y(p-3) &= 0 \\ &\vdots \\ k=p \quad \rho_Y(p) - \phi_1 \rho_Y(p-1) - \dots - \phi_p &= 0\end{aligned} \quad (3.4.18)$$

Lo siguiente será recordar que el proceso MA(q) visto en la **Sección 3.3.3** tiene como sucesión de autocorrelaciones teórica (ACF) a:

$$\rho_X(|\tau|) = \begin{cases} \frac{\sum_{j=0}^{q-|\tau|} \theta_j \theta_{|\tau|+j}}{\sum_{j=0}^q \theta_j^2} & \text{si } 0 \leq |\tau| \leq q \\ 0 & \text{si } |\tau| > q \end{cases}$$

la cual vale cero si $|\tau| > q$.

Así que si se pudieran obtener $\rho_X(1), \rho_X(2), \dots$, y en algún momento todas las $\rho_X(\tau)$ a partir de un cierto τ valen cero entonces se puede decir que estamos hablando de un MA(q). En un modelo autorregresivo no existe algo así, por ejemplo, para un proceso AR(1), $\rho_X(\tau) = \phi^\tau$, por lo cual no hay algún indicativo para saber a partir de la sucesión de autocorrelaciones teórica (ACF) qué modelo elegir. Más adelante veremos la sucesión de autocorrelaciones parciales teórica (PACF), la cual nos ayudará a identificar a los modelos AR(p).

3.5. Modelos autorregresivos de promedio móviles

Un proceso $\{X_t\}$ es tipo ARMA(p,q) si $\{X_t\}$ es estacionario y se puede escribir como:

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (3.5.1)$$

con $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q \in \mathbb{R}$ y $\{\epsilon_t\}$ es un ruido blanco con $\mathbb{E}(\epsilon_t) = 0$. Un proceso ARMA(p,q), dado por la **ec. 3.5.1**, puede describirse como:

$$\phi(B)X_t = \theta(B)\epsilon_t \quad (3.5.2)$$

$\phi(B)$ y $\theta(B)$ representan los polinomios asociados a los procesos $\{X_t\}$ y $\{\epsilon_t\}$ respectivamente, definidos de la siguiente forma:

$$\begin{aligned} \phi(z) &= 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p && \text{(polinomio autorregresivo de orden p)} \\ \theta(z) &= 1 + \theta_1 z + \theta_2 z^2 + \dots + \theta_q z^q && \text{(polinomio de promedios móviles de orden q)} \end{aligned}$$

Para justificar el porqué de esta expresión, recordemos que $BX_t = X_{t-1}$ y como consecuencia $B^j X_t = X_{t-j}$, por lo que tenemos:

$$\begin{aligned} \phi(B)X_t &= (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)X_t \\ &= X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \dots - \phi_p X_{t-p} \end{aligned}$$

Similarmente tenemos:

$$\begin{aligned} \theta(B)\epsilon_t &= (1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q)\epsilon_t \\ &= \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \end{aligned}$$

Ejemplo 3.5.1 Tomemos un proceso ARMA(1,1) causal dado por $X_t = \phi_1 X_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1}$ con $\{\epsilon_t\}$ ruido blanco con media igual a 0 y varianza constante e igual a σ_ϵ^2 . Obtendremos la sucesión de autocovarianzas teórica y la sucesión de autocorrelaciones teórica (ACF) asociadas al proceso $\{X_t\}$.

La varianza del proceso se obtiene de:

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \\ \mathbb{E}(X_t^2) &= \phi_1 \mathbb{E}(X_t X_{t-1}) + \mathbb{E}(X_t \epsilon_t) + \theta_1 \mathbb{E}(X_t \epsilon_{t-1}) \\ \gamma_X(0) &= \phi_1 \mathbb{E}(X_t X_{t-1}) + \mathbb{E}(X_t \epsilon_t) + \theta_1 \mathbb{E}(X_t \epsilon_{t-1}) \end{aligned}$$

en donde

$$\begin{aligned} \mathbb{E}(X_t \epsilon_t) &= \phi_1 \mathbb{E}(X_{t-1} \epsilon_t) + \mathbb{E}(\epsilon_t^2) + \theta_1 \mathbb{E}(\epsilon_{t-1} \epsilon_t) \\ &= \sigma_\epsilon^2 \end{aligned}$$

ya que X_{t-1} depende de la serie $\{\epsilon_t\}$ hasta el tiempo $t-1$, pero no depende de ϵ_t . Además, usando el resultado anterior pero con $t-1$ en lugar de t , tenemos:

$$\begin{aligned} \mathbb{E}(X_t \epsilon_{t-1}) &= \phi_1 \mathbb{E}(X_{t-1} \epsilon_{t-1}) + \mathbb{E}(\epsilon_t \epsilon_{t-1}) + \theta_1 \mathbb{E}(\epsilon_{t-1}^2) \\ &= \phi_1 \sigma_\epsilon^2 + \theta_1 \sigma_\epsilon^2 \end{aligned}$$

Por lo tanto, la varianza del proceso está dada por:

$$\gamma_X(0) = \phi_1 \gamma_X(1) + [1 + \theta_1(\phi_1 + \theta_1)] \sigma_\epsilon^2$$

De manera equivalente, la sucesión de autocovarianzas teórica del proceso está dada por:

$$\gamma_X(\tau) = \phi_1 \mathbb{E}(X_{t-1} X_{t-\tau}) + \mathbb{E}(X_{t-\tau} \epsilon_t) + \theta_1 \mathbb{E}(X_{t-\tau} \epsilon_{t-1})$$

es decir:

$$\gamma_X(\tau) = \begin{cases} \phi_1 \gamma_X(0) + \theta_1 \sigma_\epsilon^2 & \text{si } \tau = 1 \\ \phi_1 \gamma_X(\tau - 1) & \text{si } \tau \geq 2 \end{cases}$$

Un proceso $ARMA(p, q)$ es la generalización de los modelos $AR(p)$ y $MA(q)$. Esto se puede ver en la sucesión de autocovarianzas ya que sólo la primer autocovarianza refleja la inclusión del parámetro del modelo $MA(q)$ mientras que las demás siguen la relación de un proceso $AR(p)$.

Ahora bien, construyamos un sistema de ecuaciones para $\gamma_X(0)$ y $\gamma_X(1)$ para encontrar sus soluciones y expresar la sucesión de autocovarianzas teórica y la sucesión de autocorrelaciones teórica (ACF) de manera más sencilla:

$$\begin{aligned} \gamma_X(0) &= \phi_1 \gamma_X(1) + [1 + \theta_1(\phi_1 + \theta_1)] \sigma_\epsilon^2 \\ \gamma_X(1) &= \phi_1 \gamma_X(0) + \theta_1 \sigma_\epsilon^2 \end{aligned}$$

Dicho sistema de ecuaciones tiene por solución:

$$\begin{aligned}\gamma_X(0) &= \frac{[1 + 2\phi_1\theta_1 + \theta_1^2]\sigma_\epsilon^2}{1 - \phi_1^2} \\ \gamma_X(1) &= \frac{[(1 + \phi_1\theta_1)(\phi_1 + \theta_1)]\sigma_\epsilon^2}{1 - \phi_1^2}\end{aligned}$$

Puede concluirse que la sucesión de autocovarianzas teórica está dada por:

$$\gamma_X(\tau) = \frac{\phi_1^{\tau-1}[(1 + \phi_1\theta_1)(\phi_1 + \theta_1)]\sigma_\epsilon^2}{1 - \phi_1^2} \quad \tau = 1, 2, \dots$$

La sucesión de autocorrelaciones teórica (ACF) está dada por:

$$\rho_X(\tau) = \frac{\phi_1^{\tau-1}[(1 + \phi_1\theta_1)(\phi_1 + \theta_1)]}{1 + 2\phi_1\theta_1 + \theta_1^2} \quad \tau = 1, 2, \dots$$

Ejemplo 3.5.2 Nuevamente realizaremos una simulación de 100 observaciones pero ahora será la de dos modelos ARMA(1,1) con parámetros $\phi_1 = -0.8$, $\theta_1 = -0.3$ y $\phi_1 = 0.8$, $\theta_1 = 0.3$ respectivamente.

La **Figura 3.5.1** muestra dicha simulación así como la correspondiente sucesión de autocorrelaciones teórica (ACF) de ambos procesos ARMA(1,1) simulados. La simulación y gráfica de la correspondiente sucesión de autocorrelaciones teórica (ACF) la podemos obtener en R con el siguiente código:

```
> par(mfrow=c(2,2))
> ARMA=arima.sim(list(order=c(1,0,1),ar=-0.8,ma=-0.3),n=100)
> plot(ARMA,xlab="",ylab="X(t)",col="blue",
      main=(expression(ARMA(1,1)~con~phi[1]==-.8~y~theta[1]==.3)))
> ACF=ARMAacf(ar=-0.8,ma=-0.3,12)
> plot(ACF,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-1,1),col="blue",
      main=(expression(ACF~teórica~del~proceso~ARMA(1,1))))
> abline(h=0)
> ARMA11=arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.3),n=100)
> plot(ARMA11,xlab="",ylab="X(t)",col="blue",
      main=(expression(ARMA(1,1)~con~phi[1]==.8~y~theta[1]==.3)))
> ACF=ARMAacf(ar=0.8,ma=0.3,12)
> plot(ACF,type="h",xlab="Lag (retraso)",ylab="ACF",ylim=c(-0.1,1),col="blue",
      main=(expression(ACF~teórica~del~proceso~ARMA(1,1))))
> abline(h=0)
```

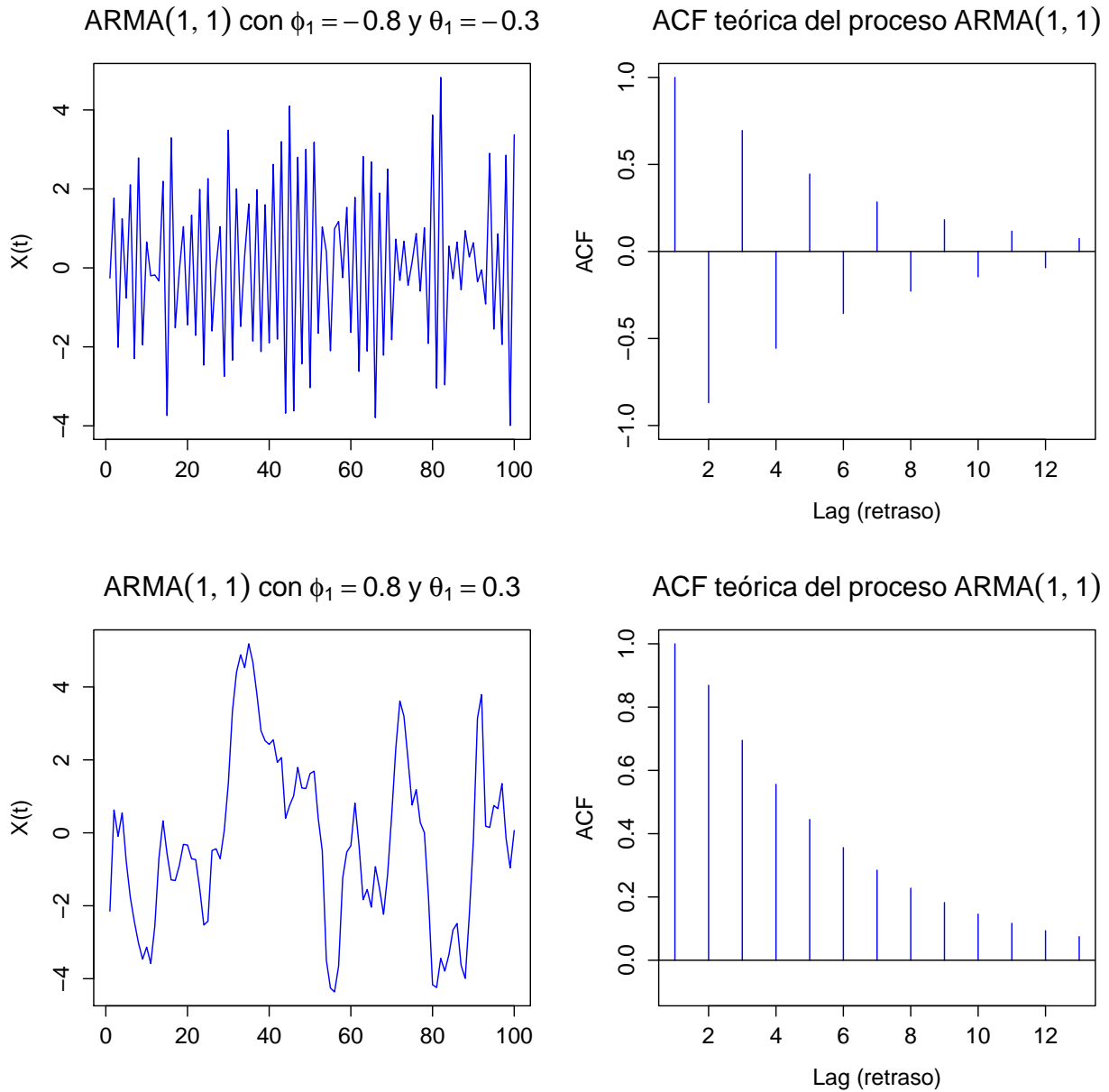


Figura 3.5.1: Figura que muestra dos procesos ARMA(1,1) simulados

Una propiedad deseable en un proceso ARMA(p,q) es la llamada causalidad, que quiere decir que el proceso $\{X_t\}$ puede escribirse como combinación lineal de los $\{\epsilon_t\}$ anteriores y el actual. Tomemos un proceso AR(1), al cual descomponemos iterando hacia atrás k-veces:

$$\begin{aligned}
 X_t &= \phi_1 X_{t-1} + \epsilon_t \\
 &= \phi_1(\phi_1 X_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
 &= \phi_1^2 X_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\
 &= \phi_1^2(\phi_1 X_{t-3} + \epsilon_{t-2}) + \phi_1 \epsilon_{t-1} + \epsilon_t \\
 &= \phi_1^3 X_{t-3} + \phi_1^2 \epsilon_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\
 &\vdots \\
 &= \phi_1^k X_{t-k} + \sum_{j=0}^{k-1} \phi_1^j \epsilon_{t-j}
 \end{aligned}
 \tag{3.5.3}$$

Suponiendo que $|\phi_1| < 1$ y por tanto, el proceso $\{X_t\}$ es causal, podemos representar, continuando con las iteraciones hacia atrás, al proceso AR(1) como un proceso lineal, definición que se dará en la **Sección 3.5.1**, dado por:

$$X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j} \quad (3.5.4)$$

En la **Sección 3.4.1** vimos que para un proceso AR(1) centrado, su primer momento era $\mathbb{E}(Y_t) = 0$ y su segundo momento estaba dado por $E(Y_t^2) = Var(Y_t) + \mathbb{E}(Y_t) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$, de tal manera que el proceso $\{Y_t\}$ resulta ser estacionario. Además, obtuvimos que la sucesión de autocorrelaciones teórica (ACF) estaba dada por $\phi_1^{|\tau|}$, lo cual indicaba que conforme $\tau > 0$ crece, la sucesión de autocorrelaciones teórica (ACF) tendía a cero.

Ahora bien, nos preguntamos si de casualidad existe algún proceso AR(1) estacionario con parámetro $|\phi_1| > 1$, ya que claramente, si existiese dicho proceso, entonces su sucesión de autocorrelaciones teórica (ACF), $\phi_1^{|\tau|}$, crecería de manera explosiva conforme $\tau \rightarrow \infty$ y además:

$$\sum_{j=0}^{k-1} \phi_1^j \epsilon_{t-j}$$

dejaría de converger cuando $k \rightarrow \infty$ y la descomposición dada por la **ec. 3.5.3** dejaría de ser útil. En efecto, existe tal proceso, conocido como **proceso AR(1) explosivo**. Para obtener estacionariedad tenemos que ver al proceso AR(1) explosivo de la siguiente forma: Tomemos $X_{t+1} = \phi_1 X_t + \epsilon_{t+1}$ y escribamos, iterando hacia adelante, a X_t como:

$$\begin{aligned} X_t &= \phi_1^{-1} X_{t+1} - \phi_1^{-1} \epsilon_{t+1} \\ &= \phi_1^{-1} (\phi_1^{-1} X_{t+2} - \phi_1^{-1} \epsilon_{t+2}) - \phi_1^{-1} \epsilon_{t+1} \\ &= \phi_1^{-2} X_{t+2} - \phi_1^{-2} \epsilon_{t+2} - \phi_1^{-1} \epsilon_{t+1} \\ &\vdots \\ &= - \sum_{j=1}^{\infty} \phi_1^{-j} \epsilon_{t+j} \end{aligned}$$

Hemos encontrado un ejemplo de un proceso estacionario que para predecir observaciones futuras necesita esas mismas observaciones futuras y por tanto no es causal.

3.5.1. Causalidad

Definición 3.5.1 *Un proceso $\{X_t\}$ es un proceso lineal si lo podemos escribir como:*

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} \quad (3.5.5)$$

$\forall t$, donde $\{\epsilon_t\}$ es un ruido blanco con media 0 y varianza constante e igual a σ_ϵ^2 .

Además, el proceso $\{\psi_j\}$ es una serie de constantes tales que:

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty \quad (3.5.6)$$

La condición dada por la **ec. 3.5.6** nos asegurará para el **Lema 3.5.1** y para la **Proposición 3.5.1** la convergencia en probabilidad 1 y en media cuadrática de series como la mostrada en la **ec. 3.5.5**, ya que como $\mathbb{E}(|\epsilon_t|) < \sigma_\epsilon$ tendremos:

$$\mathbb{E}(|X_t|) \leq \sum_{j=-\infty}^{\infty} |\psi_j| \mathbb{E}(|\epsilon_{t-j}|) \leq \left(\sum_{j=-\infty}^{\infty} |\psi_j| \right) \sigma_\epsilon < \infty$$

Definición 3.5.2 Un proceso $\{X_t\}$ ARMA(p, q) definido por la **ec. 3.5.1** es causal si existe una sucesión de pesos $\{\psi_j\}$ tales que se cumple la condición dada por la **ec. 3.5.6**, y el proceso $\{X_t\}$ se puede escribir, $\forall t \in T$, como:

$$X_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \quad (3.5.7)$$

En otras palabras, un proceso $\{X_t\}$ es causal si dicho proceso puede obtenerse a partir de otro proceso, $\{\epsilon_t\}$, viéndolo como un proceso lineal.

Lema 3.5.1 Si $\{X_t\}$ es una sucesión de variables aleatorias tales que $\sup_t \mathbb{E}(|X_t|) < \infty$ y se cumple la **ec. 3.5.6**, entonces la serie:

$$\psi(B)X_t = \sum_{j=-\infty}^{\infty} \psi_j B^j X_t = \sum_{j=-\infty}^{\infty} \psi_j X_{t-j} \quad (3.5.8)$$

converge absolutamente con probabilidad 1. Si adicionalmente $\sup_t \mathbb{E}(|X_t|^2) < \infty$, entonces la serie converge en media cuadrática al mismo límite.

Proposición 3.5.1 Sea $\{Y_t\}$ un proceso estacionario de 2º orden con sucesión de autocovarianzas $\{\gamma_Y(\tau)\}_\tau$, $\tau = 0, 1, 2, \dots$. Asumiendo cierta la condición dada por la **ec. 3.5.6**, la serie:

$$\psi(B)Y_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} \quad (3.5.9)$$

converge absolutamente con probabilidad 1 y en media cuadrática al mismo límite.

El proceso $X_t \equiv \psi(B)Y_t$ es estacionario de 2° orden con sucesión de autocovarianzas:

$$\gamma_X(\tau) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_Y(\tau - j + k) \quad (3.5.10)$$

Es decir, la **Proposición 3.5.1** nos permite asegurar que, dado un proceso $\{Y_t\}$ estacionario, el proceso $X_t \equiv \psi(B)Y_t$ es también estacionario de 2° orden e inclusive nos da la expresión para encontrar la sucesión de autocovarianzas correspondiente al proceso X_t . Entonces, si nosotros, después de realizar un análisis exhaustivo de unos datos, logramos identificar un proceso ARMA(p,q) causal, podremos también concluir que dicho proceso es estacionario de 2° orden. Si partimos del proceso $\{\epsilon_t\}$, el cual resultó ser estacionario en la **Sección 3.2**, entonces, el proceso dado por la **ec. 3.5.7** será estacionario de 2° orden, y gracias a la **Proposición 3.5.1** podremos obtener la sucesión de autocovarianzas correspondiente.

Teorema 3.5.1 Sea $\{X_t\}$ 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 si y sólo si el polinomio $\phi(z) \neq 0 \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\psi_j\}$ quedan determinados por la relación:

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)}{\phi(z)}; \quad \text{con } |z| \leq 1 \quad (3.5.11)$$

Ejemplo 3.5.3 Sea el proceso $X_t = \phi_1 X_{t-1} + \epsilon_t$, el cual corresponde a un proceso AR(1). El polinomio autorregresivo asociado a este proceso está dado por $\phi(z) = 1 - \phi_1 z$; su raíz será entonces $z = \frac{1}{\phi_1}$. El proceso es causal entonces, cuando $|\frac{1}{\phi_1}| > 1$, o bien, $|\phi_1| < 1$. El proceso puede escribirse entonces de acuerdo con la **ec. 3.5.7** y como el proceso de ruido blanco, $\{\epsilon_t\}$, es estacionario, la **Proposición 3.5.1** nos asegura la estacionariedad del proceso $\{X_t\}$. Usando la segunda parte de la **Proposición 3.5.1** podemos obtener $\gamma_X(\tau)$ de una manera más sencilla a la vista en la **Sección 3.4.1**. Tomemos la serie $1 + \phi_1 z + \phi_1^2 z^2 + \phi_1^3 z^3 + \dots$, es decir:

$$\begin{aligned} \sum_{j=0}^{\infty} \phi_1^j z^j &= \sum_{j=0}^{\infty} (\phi_1 z)^j \\ &= \frac{1}{1 - \phi_1 z} \end{aligned}$$

la cual converge a $\frac{1}{1-\phi_1 z}$. Por tanto, usando la **ec. 3.5.11**:

$$\begin{aligned} \frac{\theta(z)}{\phi(z)} &= \frac{1}{1 - \phi_1 z} \\ &= \sum_{j=0}^{\infty} \phi_1^j z^j \\ &= \sum_{j=0}^{\infty} \psi_j z^j \end{aligned}$$

con $\psi_j = \phi_1^j$ para $j = 1, 2, \dots$, y $\psi_j = 0$ para $j = -1, -2, \dots$

De acuerdo con la **Proposición 3.5.1** podemos obtener la sucesión de autocovarianzas asociada de la siguiente manera:

$$\begin{aligned}
 \gamma_X(\tau) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_\epsilon(\tau - j + k) \\
 &= \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \phi_1^j \phi_1^k \gamma_\epsilon(\tau - j + k) \\
 &= \sum_{k=0}^{\infty} \phi_1^{\tau+k} \phi_1^k \sigma_\epsilon^2 \\
 &= \sigma_\epsilon^2 \phi_1^\tau \sum_{k=0}^{\infty} (\phi_1^2)^k \\
 &= \sigma_\epsilon^2 \frac{\phi_1^\tau}{1 - \phi_1^2}
 \end{aligned}$$

usando para la convergencia de la última serie que como $|\phi_1| < 1$ entonces $|\phi_1^2| < 1$. Hemos obtenido, entonces, la misma expresión que en la **Sección 3.4.1**, es decir, la sucesión de autocovarianzas teórica de un proceso $AR(1)$.

Antes de continuar con otro ejemplo, expliquemos un poco el **Teorema 3.5.1**. En palabras, el **Teorema 3.5.1** dice que el proceso es causal si y sólo si las raíces del polinomio $\phi(z)$ no se encuentran dentro del círculo unitario, esto es, $\phi(z) = 0$ solamente cuando $|z| > 1$. Estas raíces pueden ser complejas y por eso nos referimos a su norma.

Ejemplo 3.5.4 Sea el proceso $X_t = 0.1X_{t-1} + 0.3X_{t-2} + \epsilon_t + \epsilon_{t-1} + 0.25\epsilon_{t-2}$. Uno podría rápidamente argumentar que estamos tratando con un proceso $ARMA(2,2)$, pero antes de aseverar dicha conclusión, obtengamos los polinomios $\phi(z)$ y $\theta(z)$ correspondientes. Dichos polinomios serán $\phi(z) = 1 - 0.1z - 0.3z^2 = (1 + 0.5z)(1 - 0.6z)$ y $\theta(z) = 1 + z + 0.25z^2 = (1 + 0.5z)^2$ respectivamente.

Recordemos que para poder aplicar el **Teorema 3.5.1**, los polinomios $\phi(z)$ y $\theta(z)$ no deben tener ceros en común, hecho que no sucede en este ejemplo. Eliminando los ceros en común de ambos polinomios concluimos que el proceso $\{X_t\}$ en realidad corresponde a un proceso $ARMA(1,1)$ con polinomios $\phi(z) = (1 - 0.6z)$ y $\theta(z) = (1 + 0.5z)$ ya que el término $(1 + 0.5z)$ se cancela en ambos polinomios.

En este ejemplo sólo nos interesa la parte de causalidad dada por el polinomio $\phi(z)$ el cual tiene como raíz a $z = \frac{10}{6}$, la cual se encuentra fuera del círculo unitario. El proceso $\{X_t\}$ será por tanto causal. Lo que se hará a continuación será tomar el producto de los polinomios $\phi(z)$ y $\psi(z)$, asociando los términos que tienen un mismo grado e igualándolos con los mismos términos para el polinomio $\theta(z)$. Los coeficientes $\{\psi_j\}$ quedan determinados por la relación $\phi(z)\psi(z) = \theta(z)$, dada por el **Teorema 3.5.1**, o bien por:

$$\begin{aligned}
 \phi(z)\psi(z) &= (1 - 0.6z)(\psi_0 + \psi_1z + \psi_2z^2 + \dots) \\
 &= \theta(z) \\
 &= (1 + 0.5z)
 \end{aligned}$$

Observemos que $\psi_0 = 1$, $\psi_1 = 0.5 + 0.6 = 1.1$ y $\psi_j = 0.6\psi_{j-1}$ para $j > 1$. Por tanto, $\psi_j = (0.5 + 0.6)0.6^{j-1}$ y el proceso $\{X_t\}$ puede expresarse como:

$$X_t = \epsilon_t + 1.1 \sum_{j=1}^{\infty} 0.6^{j-1} \epsilon_{t-j}$$

Ahora bien, sea $\theta_0 = 1 = -\phi_0$, $\phi_j = 0$ para $j > p$ y $\theta_j = 0$ para $j > q$. De acuerdo con la **ec. 3.5.11** $\psi(z)\phi(z)$ es un nuevo polinomio que se relaciona con $\theta(z)$ de la siguiente forma:

$$\psi(z)\phi(z) = \sum_{j=0}^{\infty} c_j z^j = \sum_{j=0}^q \theta_j z^j = \sum_{j=0}^{\infty} \theta_j z^j$$

Hemos extendido la suma hasta infinito ya que $\theta_j = 0$ para $j > q$. Nuevamente, de la **ec. 3.5.11** tenemos $\psi(z)\phi(z) = (\psi_0 + \psi_1 z + \psi_2 z^2 + \dots)(1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p)$. Al hacer el producto puede verse que el coeficiente c_j asociado a este polinomio es:

$$c_j = -\sum_{k=0}^j \phi_k \psi_{j-k} = -\phi_0 \psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j \quad (3.5.12)$$

Comenzamos con $c_0 = \psi_0 = \theta_0 = 1$. Los demás valores incluyen la siguiente lista:

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 - \phi_1 \psi_0 &= \theta_1 \\ \psi_2 - \phi_1 \psi_1 - \phi_2 \psi_0 &= \theta_2 \\ \psi_3 - \phi_1 \psi_2 - \phi_2 \psi_1 - \phi_3 \psi_0 &= \theta_3 \\ \psi_4 - \phi_1 \psi_3 - \phi_2 \psi_2 - \phi_3 \psi_1 - \phi_4 \psi_0 &= \theta_4 \\ &\vdots \end{aligned}$$

Ahora bien, en la **ec. 3.5.12**, si $j \geq p$ y además $j \geq q + 1$, se tiene que $\theta_j = 0$ (por ser un proceso de promedios móviles de orden q) y entonces:

$$c_j = -\psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = 0 \quad (3.5.13)$$

pero como para $k > p$, $\phi_k = 0$, en realidad tenemos:

$$c_j = -\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0 \quad (3.5.14)$$

Así que la ec. 3.5.14 queda escrita en 2 casos, a saber:

$$\psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} = \theta_j \quad 1 \leq j < \max\{p, q + 1\} \quad (3.5.15)$$

$$\psi_j - \sum_{k=1}^p \phi_k \psi_{j-k} = 0 \quad j \geq \max\{p, q + 1\} \quad (3.5.16)$$

Podemos usar las 2 ecuaciones anteriores y los valores iniciales $\theta_0 = 1 = -\phi_0$ para encontrar los pesos $\{\psi_j\}$ de forma sucesiva.

Ejemplo 3.5.5 Consideremos el proceso $X_t = 0.6X_{t-1} + \epsilon_t + 0.5\epsilon_{t-1}$ visto en el **Ejemplo 3.5.3**. Ya que el proceso $\{X_t\}$ es un proceso $ARMA(1,1)$ tenemos que $\max\{p, q + 1\} = 2$ por lo que, utilizando la ec. 3.5.15, obtenemos $\psi_0 = 1$ y $\psi_1 = 0.6 + 0.5 = 1.1$. Para $j > 1$ utilizamos la ec. 3.5.16 obteniendo así la serie de pesos $\{\psi_j\}$ que satisface $\psi_j - 0.6\psi_{j-1} = 0$.

El paquete estadístico *R* nos permite encontrar la serie de pesos $\{\psi_j\}$ e inclusive graficarla. No se necesita alguna librería en especial para realizar este ejemplo. La **Figura 3.5.2** muestra la serie de pesos $\{\psi_j\}$ correspondiente. El código para encontrar los primeros 25 términos de la serie de pesos $\{\psi_j\}$ y graficarlos, está dado por:

```
> ARMAtoMA(ar=.6,ma=.5,25)
> plot(ARMAtoMA(ar=.6,ma=.5,25),xlab="",ylab="Pesos",col="blue",
      main=(expression(Serie~de~pesos~psi[j])))
```

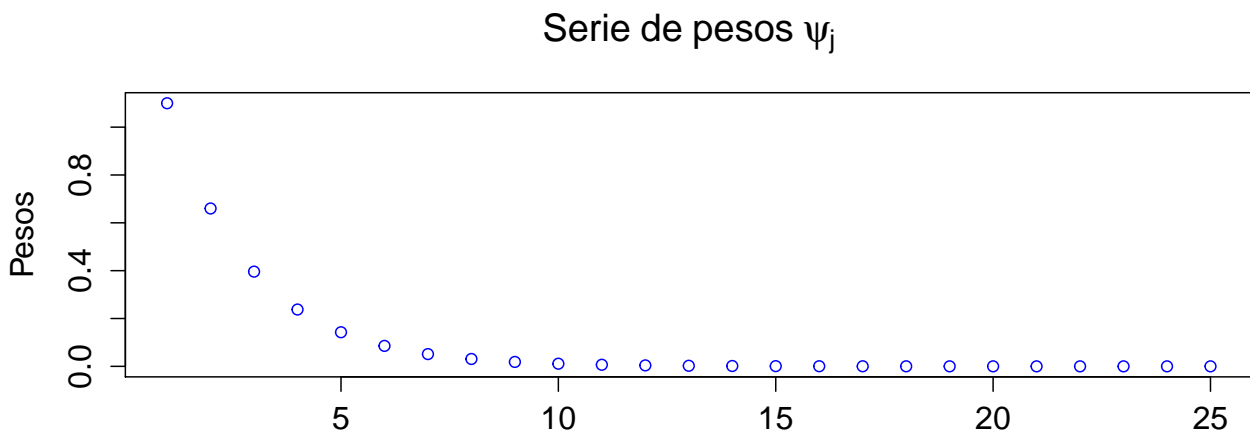


Figura 3.5.2: Figura que muestra la serie de pesos $\{\psi_j\}$ para un proceso $ARMA(1,1)$

Ejemplo 3.5.6 Para el siguiente modelo autorregresivo $Y_t - 1.097Y_{t-1} + 0.97Y_{t-2} = \epsilon_t$ con $\{\epsilon_t\}$ ruido blanco con $\mathbb{E}(\epsilon_t) = 0$ y $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$, identifique si el modelo es causal usando el **Teorema 3.5.1**.

El polinomio autorregresivo asociado al proceso $Y_t - 1.097Y_{t-1} + 0.97Y_{t-2} = \epsilon_t$ está dado por $\phi(z) = 1 - 1.097z + 0.97z^2$. Utilizando la fórmula general obtenemos que las raíces del polinomio $\phi(z)$ asociado son $z = 0.5654 \pm 0.8433i$.

$\therefore |z| = \sqrt{(0.5654)^2 + (0.8433)^2} = 1.015 > 1$ y por tanto el proceso $\{Y_t\}$ es causal, de acuerdo al **Teorema 3.5.1**. Como comentario y ejercicio, las raíces del polinomio asociado $\phi(z) = 1 - 1.097z + 0.97z^2$ se pueden calcular en R con el siguiente código:

```
> z=c(1,-1.097,.97)
> a=polyroot(z)[1]

[1] 0.5654639+0.8433139i
```

donde la primer línea del código representa un vector que incluye a los coeficientes del polinomio $\phi(z)$ y la segunda línea nos muestra una de las dos raíces (aquella que involucra el signo positivo) asociadas al polinomio $\phi(z)$. La gráfica de la serie de pesos $\{\psi_j\}$ asociada al proceso $\{Y_t\}$ se puede obtener con el siguiente código:

```
> ARMAtoMA(ar=c(1.097,0.97),ma=0,25)
> plot(ARMAtoMA(ar=c(1.097,0.97),ma=0,25),xlab="",ylab="Pesos",col="blue",
      main=(expression(Serie~de~pesos~psi[j])))
```

La **Figura 3.5.3** muestra la serie de pesos $\{\psi_j\}$ correspondiente.

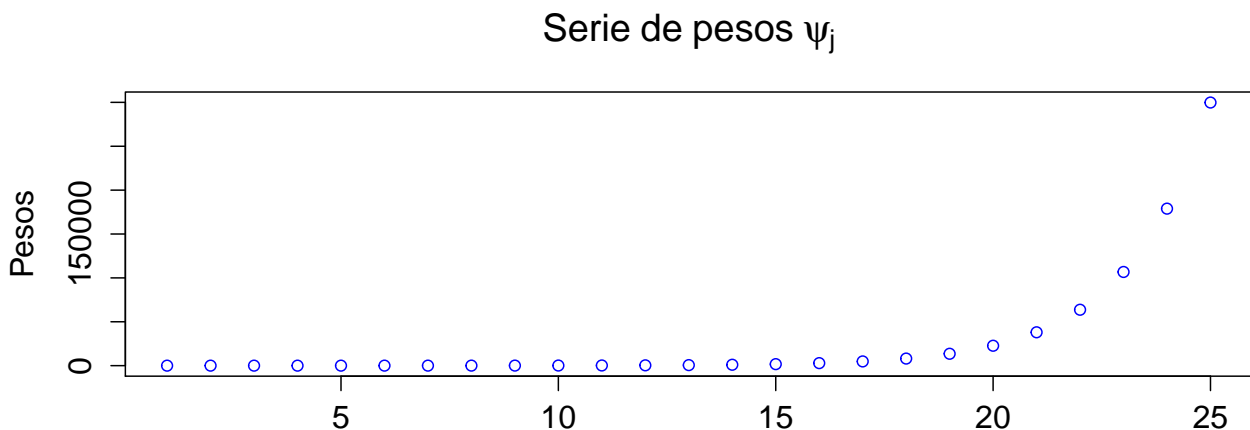


Figura 3.5.3: Figura que muestra la serie de pesos $\{\psi_j\}$ para un proceso ARMA(2,0)

Ejemplo 3.5.7 Regresando al caso $AR(1)$, $X_t = \phi X_{t-1} + \epsilon_t$ (es decir, $p = 1$ y $q = 0$), entonces $\max\{p, q + 1\} = \max\{1, 0 + 1\} = 1$. $1 \leq j < \max\{p, q + 1\} = 1$ indica $1 \leq j < 1$ lo cual no es posible así que sólo usamos la condición $\theta_0 = \psi_0 = 1$ y la ecuación para $j \geq \max\{p, q + 1\} = 1$:

$$\begin{aligned}\psi_1 &= \sum_{k=1}^1 \phi_k \psi_{1-k} = \phi_1 \psi_0 = \phi_1 \\ \psi_2 &= \sum_{k=1}^2 \phi_k \psi_{2-k} = \phi_1 \psi_1 + \phi_2 \psi_0 = \phi_1^2 \\ \psi_3 &= \sum_{k=1}^3 \phi_k \psi_{3-k} = \phi_1 \psi_2 + \phi_2 \psi_1 + \phi_3 \psi_0 = \phi_1^3\end{aligned}$$

y así sucesivamente. Por tanto, $X_t = \sum_{j=0}^{\infty} \phi_1^j z^j$, como ya sabíamos. Además $X_t = \sum_{j=0}^{\infty} \phi_1^j \epsilon_{t-j}$.

Existe un método para resolver la **ec. 3.5.16**, así que al final sólo se tendría que resolver la **ec. 3.5.15** y usar ésta y las condiciones iniciales $\theta_0 = \psi_0 = 1$ para resolver todas. A continuación veremos este procedimiento. Para un ψ_j fijo se define el siguiente polinomio:

$$P_p(B) = 1 + \alpha_1 B + \alpha_2 B^2 + \dots + \alpha_p B^p$$

de donde

$$P_p(B)\psi_j = \psi_j + \alpha_1 \psi_{j-1} + \alpha_2 \psi_{j-2} + \dots + \alpha_p \psi_{j-p}$$

Obtenemos ahora las raíces de $P_p(B)$, llamémoslas ξ_i , $i = 1, 2, \dots, l$ (con l el número de raíces distintas), r_i la multiplicidad de cada raíz $\forall i$, siendo evidente que al sumarlas se obtiene p , el número de raíces en total (considerando unas múltiples y otras no). La solución general de $P_p(B)\psi_j = 0$ es:

$$\psi_j = \sum_{i=1}^l \sum_{n=0}^{r_i-1} c_{in} j^n (\xi_i)^{-j} \quad (3.5.17)$$

En particular, a la **ec. 3.5.16** le corresponde el polinomio:

$$\begin{aligned}P_p(B)\psi_j &= \psi_j + \alpha_1 \psi_{j-1} + \alpha_2 \psi_{j-2} + \dots + \alpha_p \psi_{j-p} \\ &= \psi_j - \phi_1 \psi_{j-1} - \phi_2 \psi_{j-2} + \dots - \phi_p \psi_{j-p}\end{aligned}$$

Observe que este polinomio corresponde a la parte autorregresiva que se obtiene al escribir el modelo ARMA(p,q) como $X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$, así que al obtener las raíces del polinomio de esa parte obtenemos los valores r_i, ξ_i, l requeridos en la fórmula.

Entonces, el proceso a seguir para obtener el proceso causal explícito es el siguiente: obtenemos las raíces de $\phi(z)$, el polinomio de la parte autorregresiva, para después verificar que el proceso sea causal de acuerdo con el **Teorema 3.5.1** (estas raíces corresponden a los ξ_i vistos arriba). Proporcionamos las condiciones iniciales de acuerdo con la **ec. 3.5.15** y a $\theta_0 = \psi_0 = 1$ para finalmente usar estas condiciones y resolver la **ec. 3.5.16** de acuerdo con la **ec. 3.5.17**.

Ejemplo 3.5.8 Encontrar expresiones para $\{\psi_j\}$ en un proceso ARMA(p, q) causal dado por $X_t = X_{t-1} - \frac{1}{4}X_{t-2} + \epsilon_t + \epsilon_{t-1}$.

El proceso es equivalente a $X_t - X_{t-1} + \frac{1}{4}X_{t-2} = \epsilon_t + \epsilon_{t-1}$. Entonces $\phi_1 = 1$, $\phi_2 = -\frac{1}{4}$ y $\theta_1 = 1$. El polinomio de $\phi(z) = 1 - z + \frac{1}{4}z^2$ tiene una raíz $z = 2$ de multiplicidad 2, por lo que $\xi_1 = 2$ con $r_1 = 2$ (multiplicidad de la raíz) y $l = 1$ (número de raíces distintas). Además la raíz está fuera del círculo unitario por lo que el proceso es causal y podemos encontrar los pesos $\{\psi_j\}$ asociados. Las ecuaciones que queremos resolver son:

$$\begin{aligned} \psi_j - \sum_{k=1}^1 \phi_k \psi_{j-k} &= \theta_j & 1 \leq j < \max\{2, 1+1\} = 2 \\ \psi_j - \sum_{k=1}^2 \phi_k \psi_{j-k} &= 0 & j \geq \max\{2, 1+1\} = 2 \end{aligned}$$

Además $\theta_0 = \psi_0 = 1$. De la primera ecuación la única j posible es $j = 1$, así que $\psi_0 = 1$ y $\psi_1 = \theta_1 + \phi_1 \psi_{1-1} = \theta_1 + \phi_1 \psi_0 = 2$ y la solución de la **ec. 3.5.17** está dada por:

$$\begin{aligned} \psi_j &= \sum_{i=1}^l \sum_{n=0}^{r_i-1} c_{in} j^n (\xi_i)^{-j} \\ &= \sum_{n=0}^{r_1-1} c_{1n} j^n (\xi_1)^{-j} \\ &= c_{10} \left(\frac{1}{2}\right)^j + c_{11} j \left(\frac{1}{2}\right)^j \end{aligned}$$

Usando las condiciones iniciales para que la expresión anterior pueda usarse para cualquier j tenemos:

$$\begin{aligned} 1 &= c_{10} (2)^0 + c_{11} (0) (2)^0 & \therefore c_{10} &= 1 \\ 2 &= \frac{1}{2} c_{10} + \frac{1}{2} c_{11} (1) & \therefore c_{11} &= 3 \end{aligned}$$

$\therefore \psi_j = (1 + 3j)2^{-j} \forall j \geq 0$ y entonces $\{X_t\}$ se puede escribir como:

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \\ &= \sum_{j=0}^{\infty} (1 + 3j)2^{-j} \epsilon_{t-j} \\ &= \epsilon_t + \frac{4}{2} \epsilon_{t-1} + \frac{7}{4} \epsilon_{t-2} + \dots \end{aligned}$$

Ejemplo 3.5.9 Encontrar expresiones para $\{\psi_j\}$ en un proceso ARMA(p, q) causal dado por $X_t + 0.2X_{t-1} - 0.15X_{t-2} = \epsilon_t + 0.3\epsilon_{t-1} - 0.1\epsilon_{t-2}$.

Los polinomios asociados a la parte autorregresiva y de promedios móviles serán entonces $\phi(z) = 1 + 0.2z - 0.15z^2$ y $\theta(z) = 1 + 0.3z - 0.1z^2$ respectivamente. En esta sección nos interesará solamente el polinomio $\phi(z)$.

Al utilizar la fórmula general para obtener las raíces del polinomio $\phi(z)$ llegamos a $z_1 = -2$ y $z_2 = \frac{10}{3}$, ambas de multiplicidad 1, por lo que $\xi_1 = -2$, $\xi_2 = \frac{10}{3}$, $r_1 = r_2 = 1$ y $l = 2$ (número de raíces distintas). Además, las raíces $z_1 = -2$ y $z_2 = \frac{10}{3}$ se encuentran ambas fuera del círculo unitario por lo que el proceso $\{X_t\}$ es causal y podemos encontrar los pesos $\{\psi_j\}$ asociados. Las ecuaciones que queremos resolver ahora son:

$$\begin{aligned} \psi_j - \sum_{k=1}^j \phi_k \psi_{j-k} &= \theta_j & 1 \leq j < \max\{2, 2+1\} = 3 \\ \psi_j - \sum_{k=1}^2 \phi_k \psi_{j-k} &= 0 & j \geq \max\{2, 3\} = 3 \end{aligned}$$

Usaremos la primera ecuación para $j = 1, 2$, y la segunda para $j > 2$. Tenemos por un lado que $\psi_0 = 1$, y por otro, de la primera ecuación obtenemos $\psi_1 = 0.3 - 0.2 = 0.1$ y $\psi_2 = -0.1 - 0.2 + (0.15)(0.1) = -0.29$. La solución de la **ec. 3.5.17** está dada por:

$$\begin{aligned} \psi_j &= \sum_{i=1}^l \sum_{n=0}^{r_i-1} c_{in} j^n (\xi_i)^{-j} \\ &= \sum_{n=0}^{r_1-1} c_{1n} j^n (\xi_1)^{-j} + \sum_{n=0}^{r_2-1} c_{2n} j^n (\xi_2)^{-j} \\ &= -c_{10} \left(\frac{1}{2}\right)^j + c_{20} \left(\frac{3}{10}\right)^j \end{aligned}$$

Usando las condiciones iniciales tenemos:

$$\begin{aligned} 1 &= c_{10} + c_{20} \\ 0.1 &= -\frac{1}{2}c_{10} + \frac{3}{10}c_{20} \\ -0.29 &= \frac{1}{4}c_{10} + \frac{9}{100}c_{20} \end{aligned}$$

Tomando las dos primeras ecuaciones obtenemos $c_{10} = \frac{1}{4}$ y $c_{20} = \frac{3}{4}$. Por lo tanto la sucesión de pesos puede escribirse como:

$$\psi_j = \frac{1}{4} \left(\frac{1}{2}\right)^j + \frac{3}{4} \left(\frac{3}{10}\right)^j$$

Entonces, el proceso $\{X_t\}$ se puede expresar como:

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \\ &= \sum_{j=0}^{\infty} \left\{ \frac{1}{4} \left(-\frac{1}{2}\right)^j + \frac{3}{4} \left(\frac{3}{10}\right)^j \right\} \epsilon_{t-j} \\ &= \epsilon_t + \frac{1}{10} \epsilon_{t-1} + \frac{13}{100} \epsilon_{t-2} + \dots \end{aligned}$$

Ejemplo 3.5.10 Demostrar que en general, para un proceso $AR(2)$, las condiciones necesarias para que sea causal y estacionario son $-2 < \phi_1 < 2$ y $-1 < \phi_2 < 1$.

El polinomio $\phi(z)$ asociado al proceso $Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} = \epsilon_t$ es $\phi(z) = -\phi_2 z^2 - \phi_1 z + 1$. Resolvemos la ecuación $-\phi_2 z^2 - \phi_1 z + 1 = 0 \Leftrightarrow \phi_2 z^2 + \phi_1 z - 1 = 0$. A partir de aquí, y usando la fórmula general, podemos obtener las raíces z_1 y z_2 , a saber:

$$z_1 = \frac{-\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

$$z_2 = \frac{-\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$$

Ahora bien, a partir de $-\phi_2 - \frac{\phi_1}{z} + \left(\frac{1}{z}\right)^2 = 0$ obtenemos:

$$\frac{1}{z_1} = \frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

$$\frac{1}{z_2} = \frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

Para que el proceso $\{Y_t\}$ sea causal $|z_1| > 1$ y $|z_2| > 1$ o bien, $\left|\frac{1}{z_1}\right| < 1$ y $\left|\frac{1}{z_2}\right| < 1$, así que $\left|\frac{1}{z_1} \frac{1}{z_2}\right| < 1$, pero:

$$\left|\frac{1}{z_1} \frac{1}{z_2}\right| = \left| \left(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) \left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) \right| = |-\phi_2| = |\phi_2|$$

$\therefore |\phi_2| < 1$. Por otra parte:

$$\left|\frac{1}{z_1} + \frac{1}{z_2}\right| = \left| \left(\frac{\phi_1 + \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) + \left(\frac{\phi_1 - \sqrt{\phi_1^2 + 4\phi_2}}{2} \right) \right| = |\phi_1|$$

Además:

$$\left| \frac{1}{z_1} + \frac{1}{z_2} \right| \leq \left| \frac{1}{z_1} \right| + \left| \frac{1}{z_2} \right| < 1 + 1 = 2$$

$\therefore |\phi_1| < 2$. Además si el proceso $\{Y_t\}$ es causal entonces $\{Y_t\}$ es estacionario.

3.5.2. Invertibilidad

Ejemplo 3.5.11 En la **Sección 3.3** estudiamos a los procesos $MA(q)$. En particular, notemos que para dos modelos con parámetros $(\sigma_\epsilon^2 = 4, \theta = \frac{1}{2})$ y $(\sigma_\nu^2 = 1, \theta = 2)$ respectivamente, la sucesión de autocovarianzas teórica para ambos es la misma, a saber:

$$\gamma(\tau) = \begin{cases} 5 & \text{si } \tau = 0 \\ 2 & \text{si } |\tau| = 1 \\ 0 & \text{si } |\tau| > 1 \end{cases}$$

Entonces, los procesos $MA(1)$:

$$\begin{aligned} X_t &= \epsilon_t + \frac{1}{2}\epsilon_{t-1}, & \epsilon_t &\sim NIID(0, 4) \\ Y_t &= \nu_t + 2\nu_{t-1}, & \nu_t &\sim NIID(0, 1) \end{aligned}$$

son los mismos debido a la normalidad en las sucesiones $\{\epsilon_t\}$ y $\{\nu_t\}$. Para demostrar lo anterior, calculemos las varianzas de ambos procesos y veamos que son las mismas. Para el proceso $\{X_t\}$ tenemos:

$$\begin{aligned} \text{Var}(X_t) &= \text{Var}(\epsilon_t + 0.5\epsilon_{t-1}) \\ &= \text{Var}(\epsilon_t) + 0.25\text{Var}(\epsilon_{t-1}) \\ &= \sigma_\epsilon^2 + 0.25\sigma_\epsilon^2 \\ &= 1.25\sigma_\epsilon^2 \\ &= 5 && \text{(ya que } \sigma_\epsilon^2 = 4) \end{aligned}$$

mientras que para el proceso $\{Y_t\}$ tenemos:

$$\begin{aligned} \text{Var}(Y_t) &= \text{Var}(\nu_t + 2\nu_{t-1}) \\ &= \text{Var}(\nu_t) + 4\text{Var}(\nu_{t-1}) \\ &= \sigma_\nu^2 + 4\sigma_\nu^2 \\ &= 5\sigma_\nu^2 \\ &= 5 && \text{(ya que } \sigma_\nu^2 = 1) \end{aligned}$$

Nosotros podemos observar las series de tiempo $\{X_t\}$ y $\{Y_t\}$ pero no las secuencias $\{\epsilon_t\}$ y $\{\nu_t\}$, de tal forma que no podemos distinguir entre ambos modelos.

De la misma forma que expresamos un proceso $\{X_t\}$ como combinación lineal del proceso $\{\epsilon_t\}$, es decir, como un proceso de promedios móviles con infinitos términos, expresaremos al proceso $\{\epsilon_t\}$ como un proceso autorregresivo con infinitos términos. Dicho proceso se conoce como proceso invertible. De los dos modelos MA(1), escogeremos aquel que sea invertible. Sea un modelo MA(1) dado por $X_t = \epsilon_t + \theta_1 \epsilon_{t-1}$. Podemos escribir dicho proceso como $\epsilon_t = -\theta_1 \epsilon_{t-1} + X_t$ y realizar un procedimiento similar al hecho en la **ec. 3.5.4** para obtener:

$$\begin{aligned}
\epsilon_t &= -\theta_1 \epsilon_{t-1} + X_t \\
&= -\theta_1(-\theta_1 \epsilon_{t-2} + X_{t-1}) + X_t \\
&= \theta_1^2 \epsilon_{t-2} - \theta_1 X_{t-1} + X_t \\
&= \theta_1^2(-\theta_1 \epsilon_{t-3} + X_{t-2}) - \theta_1 X_{t-1} + X_t \\
&= -\theta_1^3 \epsilon_{t-3} + \theta_1^2 X_{t-2} - \theta_1 X_{t-1} + X_t \\
&\vdots \\
&= \sum_{j=0}^{\infty} (-\theta_1^j) X_{t-j}
\end{aligned} \tag{3.5.18}$$

la cual es la representación del proceso $\{\epsilon_t\}$ como un proceso autorregresivo con infinitos términos. Para invertibilidad utilizaremos la condición:

$$\sum_{j=-\infty}^{\infty} |\pi_j| < \infty \tag{3.5.19}$$

Definición 3.5.3 *Un proceso $\{X_t\}$ ARMA(p, q) definido por la **ec. 3.5.1** es invertible si existe una sucesión de pesos $\{\pi_j\}$ que cumplen la **ec. 3.5.19** y el proceso $\{\epsilon_t\}$ puede escribirse como:*

$$\epsilon_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}. \tag{3.5.20}$$

Esto significa despejar el ruido blanco en términos de un proceso autorregresivo infinito.

Teorema 3.5.2 *Sea $\{X_t\}$ 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 si y sólo si $\theta(z) \neq 0 \forall z \in \mathbb{C}$ tal que $|z| \leq 1$. Los coeficientes $\{\pi_j\}$ en la **ec. 3.5.20** quedan determinados por la relación:*

$$\pi(z) = \sum_{j=0}^{\infty} \pi_j Z^j = \frac{\phi(z)}{\theta(z)}; \quad \text{con } |z| \leq 1 \tag{3.5.21}$$

Similar al caso de causalidad se pueden encontrar expresiones para $\{\pi_j\}$ en el proceso ARMA(p, q) invertible.

De la **ec. 3.5.21** tenemos $\pi(z)\theta(z) = (\pi_0 + \pi_1z + \pi_2z^2 + \dots)(1 + \theta_1z + \theta_2z^2 + \dots + \theta_qz^q)$ con $c_0 = \pi_0$. Al hacerse el producto anterior puede verse que el coeficiente c_j asociado al nuevo polinomio $\pi(z)\theta(z)$ es:

$$c_j = \sum_{k=0}^j \theta_k \pi_{j-k} = \theta_0 \pi_j + \sum_{k=1}^j \theta_k \pi_{j-k} \quad (3.5.22)$$

Por conveniencia $\phi_0 = -1$, así que $c_0 = -\phi_0 = 1$, pero $c_0 = \pi_0$, de aquí que $\pi_0 = -\phi_0 = 1$. Nuevamente tenemos 2 casos, a saber:

$$\pi_j + \sum_{k=1}^j \theta_k \pi_{j-k} = -\phi_j \quad 1 \leq j < \max\{q, p+1\} \quad (3.5.23)$$

$$\pi_j + \sum_{k=1}^q \theta_k \pi_{j-k} = 0 \quad j \geq \max\{q, p+1\} \quad (3.5.24)$$

Así que se obtienen condiciones iniciales $\pi_0 = -\phi_0 = 1$ y con las **ecs. en 3.5.23** y las condiciones iniciales, se resuelven las **ecs. en 3.5.24** usando la misma técnica que para causalidad solo que todo en términos de $\theta(z)$. Veremos a continuación un ejemplo.

Ejemplo 3.5.12 Encontrar expresiones para $\{\pi_j\}$ en un proceso ARMA(p, q) dado por la expresión $X_t = \frac{1}{2}X_{t-1} + \epsilon_t + \frac{2}{5}\epsilon_{t-1}$.

Antes que nada, el proceso es equivalente a $X_t - \frac{1}{2}X_{t-1} = \epsilon_t + \frac{2}{5}\epsilon_{t-1}$. Observemos que $\phi_0 = 1$, $\phi_1 = \frac{1}{2}$, $\theta_0 = 1$, $\theta_1 = \frac{2}{5}$, $p = 1$, $q = 1$, de donde $\max\{q, p+1\} = \max\{1, 1+1\} = 2$. Además $\pi_0 = 1$, por tanto la **ec. 3.5.23** es para $1 \leq j < 2$, es decir:

$$\pi_1 + \sum_{k=1}^1 \theta_k \pi_{1-k} = -\phi_1$$

$$\therefore \pi_1 = -\frac{1}{2} - \frac{2}{5} = -\frac{9}{10}.$$

La **ec. 3.5.24**, para $j \geq 2$, está dada por:

$$\pi_j = -\sum_{k=1}^1 \theta_k \pi_{j-k} = -\theta_1 \pi_{j-1}$$

El polinomio asociado a $\theta(z)$ es $1 + \frac{2}{5}z$, su raíz es $z = -\frac{5}{2}$, que satisface $|- \frac{5}{2}| > 1$, de tal forma que la raíz se encuentra fuera del círculo unitario y por tanto, el proceso es invertible. Además, la raíz de $\phi(z) = 1 - \frac{1}{2}z$ es 2, por lo cual, los polinomios no tienen ceros en común.

En este caso $l = 1$, $r_1 = 1$ (multiplicidad de la raíz) y $\xi_1 = -\frac{5}{2}$, por lo que la **ec. 3.5.17** queda:

$$\begin{aligned}\pi_j &= \sum_{i=1}^l \sum_{n=0}^{r_i-1} c_{in} j^n (\xi_i)^{-j} \\ &= \sum_{i=1}^1 \sum_{n=0}^{1-1} c_{in} j^n (\xi_i)^{-j} \\ &= c_{10} \left(-\frac{5}{2}\right)^{-j}\end{aligned}$$

con las condiciones iniciales:

$$\begin{aligned}1 &= c_{10} \left(-\frac{5}{2}\right)^0 && \text{de donde } c_{10} = 1 \\ -\frac{9}{10} &= c_{10} \left(-\frac{5}{2}\right)^{-1} && \text{de donde } c_{10} = \frac{45}{20}\end{aligned}$$

Tenemos por tanto, 2 valores de c_{10} , ¿cuál usamos? Se usan las condiciones para π_j con la j más grande.

$\therefore \pi_j = \left(\frac{45}{20}\right) \left(-\frac{5}{2}\right)^{-j}$ y el proceso $\{\epsilon_t\}$ se puede escribir como:

$$\epsilon_t = \sum_{j=0}^{\infty} \left(\frac{45}{20}\right) \left(-\frac{5}{2}\right)^{-j} X_{t-j}$$

El obtener la expresión dada por la **ec. 3.5.7** obtenida para un proceso causal puede ser útil para obtener la media $\mathbb{E}(X_t)$ y $Cov(X_t, X_{t+\tau})$ asociados.

Proposición 3.5.2 *El proceso de promedios móviles con infinitos términos, $MA(\infty)$, dado por la **ec. 3.5.7**, es estacionario con media 0 y sucesión de autocovarianzas teórica:*

$$\gamma_X(\tau) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|\tau|} \quad (3.5.25)$$

La demostración de la **ec. 3.5.25** es inmediata usando la **Proposición 3.5.1**, mientras que la media del proceso $\{X_t\}$ es cero ya que:

$$\mathbb{E}(X_t) = \mathbb{E} \left(\sum_{j=0}^{\infty} \psi_j \epsilon_{t-j} \right) = \sum_{j=0}^{\infty} \psi_j \mathbb{E}(\epsilon_{t-j}) = 0$$

ya que el proceso $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$ y además se cumple la condición dada por la **ec. 3.5.6** para los pesos $\{\psi_j\}$.

Ejemplo 3.5.13 En el proceso causal $X_t = X_{t-1} - \frac{1}{4}X_{t-2} + \epsilon_t + \epsilon_{t-1}$ visto en el **Ejemplo 3.5.8**, obtuvimos los coeficientes $\psi_j = (1 + 3j)2^{-j}$. Según la **proposición 3.5.2**, $\mathbb{E}(X_t) = 0$ pero también:

$$\begin{aligned}\gamma_X(\tau) &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} (1 + 3j)2^{-j} (1 + 3(j + |\tau|))2^{-(j+|\tau|)} \\ &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} (1 + 3j)(1 + 3j + 3|\tau|)2^{-2j-|\tau|} \\ &= \sigma_\epsilon^2 2^{-|\tau|} \sum_{j=0}^{\infty} [(3|\tau| + 1) + (3|\tau| + 1)3j + 3j + 9j^2] 4^{-j} \\ &= \sigma_\epsilon^2 2^{-|\tau|} \sum_{j=0}^{\infty} [(3|\tau| + 1)4^{-j} + 3j(3|\tau| + 2)4^{-j} + 9j^2 4^{-j}] \\ &= \sigma_\epsilon^2 2^{-|\tau|} \left\{ (3|\tau| + 1) \sum_{j=0}^{\infty} 4^{-j} + 3(3|\tau| + 2) \sum_{j=0}^{\infty} j 4^{-j} + 9 \sum_{j=0}^{\infty} j^2 4^{-j} \right\}\end{aligned}$$

Sea $X \sim \text{Geom}\left(\frac{3}{4}\right)$ con $\frac{1}{3} = \frac{\frac{1}{4}}{\frac{3}{4}} = \frac{1 - \frac{3}{4}}{\frac{3}{4}} = \mathbb{E}(x) = \sum_{x=0}^{\infty} x \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right) = \frac{3}{4} \sum_{x=0}^{\infty} x \left(\frac{1}{4}\right)^x$

$$\therefore \sum_{x=0}^{\infty} x \left(\frac{1}{4}\right)^x = \frac{4}{9}$$

Realizamos un procedimiento similar para $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$ con $X \sim \text{Geom}\left(\frac{3}{4}\right)$

$$\therefore \sum_{x=0}^{\infty} x^2 \left(\frac{1}{4}\right)^x = \frac{20}{27}$$

Regresando a la expresión vista, obtenemos la sucesión de autocovarianzas:

$$\begin{aligned}\gamma_X(\tau) &= \sigma_\epsilon^2 2^{-|\tau|} \left\{ \frac{4}{3}(3|\tau| + 1) + \frac{12}{9}(3|\tau| + 2) + 9 \frac{20}{27} \right\} \\ &= \sigma_\epsilon^2 2^{-|\tau|} \left\{ 4|\tau| + \frac{4}{3} + 4|\tau| + \frac{8}{3} + \frac{20}{3} \right\} \\ &= \frac{\sigma_\epsilon^2 2^{-|\tau|}}{3} \{24|\tau| + 32\}\end{aligned}$$

Proposición 3.5.3 Sea $\{X_t\}$ un proceso ARMA(p, q) causal. Demostrar que se cumple la siguiente relación:

$$\sum_{\tau=-\infty}^{\infty} |\gamma_X(\tau)| < \infty \quad (3.5.26)$$

De la **ec. 3.5.26** se deriva que $|\gamma_X(\tau)| \xrightarrow{\tau \rightarrow \infty} 0$, es decir, a medida que aumenta la diferencia entre los tiempos en un proceso causal la sucesión de autocovarianzas teórica se acerca a cero.

Demostración

Al ser $\{X_t\}$ un proceso ARMA(p, q) causal, podemos hacer uso de la **Proposición 3.5.1** para de este modo obtener la sucesión de autocovarianzas teórica:

$$\begin{aligned}\gamma_X(\tau) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \psi_j \psi_k \gamma_\epsilon(\tau - j + k) \\ &= \sigma_\epsilon^2 \sum_{k=0}^{\infty} \psi_k \psi_{k+|\tau|}\end{aligned}$$

ya que

$$\gamma_\epsilon(\tau - j + k) = \begin{cases} \sigma_\epsilon^2 & \text{si } \tau - j + k = 0 \\ 0 & \text{si } \tau - j + k \neq 0 \end{cases}$$

Sin pérdida de generalidad, supóngase $\sigma_\epsilon^2 = 1$, entonces:

$$|\gamma_X(\tau)| \leq \sum_{k=0}^{\infty} |\psi_k| |\psi_{k+|\tau|}|$$

y al sumar sobre todo τ obtenemos:

$$\sum_{\tau} |\gamma_X(\tau)| \leq \sum_{\tau} \sum_{k=0}^{\infty} |\psi_k| |\psi_{k+|\tau|}| = \sum_{k=0}^{\infty} \sum_{\tau} |\psi_k| |\psi_{k+|\tau|}| = \sum_{k=0}^{\infty} |\psi_k| \left\{ \sum_{\tau} |\psi_{k+|\tau|}| \right\}$$

En la relación anterior se aplicó el teorema de Fubini al tener términos positivos. Entonces:

$$\begin{aligned}\sum_{\tau=-\infty}^{\infty} |\psi_{k+|\tau|}| &\leq \sum_{\tau=-\infty}^0 |\psi_{k-\tau}| + \sum_{\tau=0}^{\infty} |\psi_{k+\tau}| \\ &= \sum_{l=k}^{\infty} |\psi_l| + \sum_{l=k}^{\infty} |\psi_l| \\ &\leq \sum_{l=0}^{\infty} |\psi_l| + \sum_{l=0}^{\infty} |\psi_l| \\ &= 2 \sum_{l=0}^{\infty} |\psi_l| \\ &= 2c < \infty\end{aligned}\tag{3.5.27}$$

ya que $c = \sum_{l=0}^{\infty} |\psi_l| < \infty$ al ser $\{X_t\}$ un proceso causal.

$\therefore \sum_{\tau} |\gamma_X(\tau)| \leq \sum_{k=0}^{\infty} |\psi_k| \left\{ \sum_{\tau} |\psi_{k+|\tau|}| \right\} \leq 2c \sum_{k=0}^{\infty} |\psi_k| < \infty$ ya que $\sum_{k=0}^{\infty} |\psi_k| < \infty$ de acuerdo con la **ec. 3.5.6**.

$\therefore \sum_{\tau=-\infty}^{\infty} |\gamma_X(\tau)| < \infty$ y entonces $|\gamma_X(\tau)| \xrightarrow{\tau \rightarrow \infty} 0$, que es lo que queríamos demostrar.

Hasta aquí hemos trabajado con el marco teórico de series de tiempo, calculando, por ejemplo, las sucesiones de autocovarianzas y autocorrelaciones teóricas de modelos autorregresivos (AR(1), AR(2) y AR(p)), de promedios móviles (MA(1), MA(2) y MA(q)) y combinaciones de los anteriores (ARMA(p,q)). En la práctica, para identificar un proceso $\{X_t\}$, debemos estimar de unos datos todas las propiedades estadísticas estudiadas en la secciones anteriores.

3.5.3. Identificación del modelo a partir de la muestra

Sea $\{X_t\}$ un proceso estacionario con media μ . Un estimador para $\mu = \mathbb{E}(X_t)$ es:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (3.5.28)$$

con n el número de observaciones en la muestra. Es obvio que $\mathbb{E}(\bar{X}_n) = \mu$, es decir, el estimador es insesgado. El error cuadrático medio del estimador \bar{X}_n está dado por $Var(\bar{X}_n)$, ya que es insesgado, el cual puede obtenerse de la siguiente forma:

$$\begin{aligned} \mathbb{E}(\bar{X}_n - \mu)^2 &= Var(\bar{X}_n) \\ &= Var\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{t=1}^n \sum_{s=1}^n Cov(X_t, X_s) \\ &= \frac{1}{n^2} \sum_{t-s=-n}^n (n - |t-s|) \gamma_X(t-s) \\ &= \frac{1}{n} \sum_{\tau=-n}^n \left(1 - \frac{|\tau|}{n}\right) \gamma_X(\tau) \end{aligned} \quad (3.5.29)$$

Si $\gamma_X(\tau) \rightarrow 0$ conforme $\tau \rightarrow \infty$, entonces la **ec. 3.5.29** converge a 0 y por tanto \bar{X}_n converge en media cuadrática a μ . Ahora bien, ¿nos surge algún otro comentario a partir de la **ec. 3.5.29**? Pensemos por un momento que si el proceso $\{X_t\}$ es un ruido blanco entonces $\gamma_X(\tau) \neq 0$ solo para $\tau = 0$, por lo que, en este caso:

$$Var(\bar{X}_n) = \frac{1}{n} \sum_{\tau=-n}^n \left(1 - \frac{|\tau|}{n}\right) \gamma_X(\tau) = \frac{\sigma_X^2}{n}$$

Otros resultados interesantes del error cuadrático medio se mencionan a continuación.

Teorema 3.5.3 Si $\{X_t\}$ es un proceso estacionario de 2° orden con media μ y sucesión de autocovarianzas $\gamma_X(\tau)$ entonces:

$$\text{Var}(\bar{X}_n) = \mathbb{E}[(\bar{X}_n - \mu)^2] \xrightarrow{n \rightarrow \infty} 0 \quad (3.5.30)$$

siempre que $\gamma_X(\tau) \xrightarrow{\tau \rightarrow \infty} 0$. Además:

$$n\mathbb{E}[(\bar{X}_n - \mu)^2] \longrightarrow \sum_{\tau=-\infty}^{\infty} \gamma_X(\tau) \quad (3.5.31)$$

siempre que se cumpla la **ec. 3.5.26**.

Demostración

$$\begin{aligned} n\text{Var}(\bar{X}_n) &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \text{Cov}(X_t, X_s) \\ &= \sum_{|\tau| < n} \left(1 - \frac{|\tau|}{n}\right) \gamma_X(\tau) \\ &\leq \sum_{|\tau| < n} |\gamma_X(\tau)| \end{aligned}$$

Si $\gamma_X(n) \rightarrow 0$ conforme $n \rightarrow \infty$ entonces:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{|\tau| < n} |\gamma_X(\tau)| = 2 \lim_{n \rightarrow \infty} |\gamma_X(n)| = 0$$

En palabras, lo que hemos hecho primero es ver que $\text{Var}(\bar{X}_n)$ está acotada por el “promedio” de $|\gamma_X(\tau)|$. Luego observamos que cuando n tiende a infinito este “promedio” converge a cero. Así que la varianza, la cual es no negativa cuando n tiende infinito, está acotada por algo que tiende a cero así que necesariamente debe tender a cero dicha varianza. Si además se cumple la **ec. 3.5.26** entonces, utilizando el teorema de convergencia dominada obtenemos:

$$\lim_{n \rightarrow \infty} n\text{Var}(\bar{X}_n) = \lim_{n \rightarrow \infty} \sum_{|\tau| < n} \left(1 - \frac{|\tau|}{n}\right) \gamma_X(\tau) = \sum_{\tau=-\infty}^{\infty} \gamma_X(\tau)$$

El **Teorema 3.5.3** nos asegura que el $ECM(\bar{X}_n)$ se hace cada vez más chico, así que \bar{X}_n es un estimador insesgado cuyo error cuadrático medio disminuye al aumentar n , es decir, si $\gamma_X(n) \rightarrow 0$ conforme $n \rightarrow \infty$, entonces \bar{X}_n converge en media cuadrática a μ . Además, si la **ec. 3.5.26** se cumple entonces:

$$\text{Var}(\bar{X}_n) \sim \frac{1}{n} \sum_{\tau=-\infty}^{\infty} \gamma_X(\tau)$$

A partir de aquí, se puede demostrar que asintóticamente:

$$\bar{X}_n \sim N \left(\mu, \frac{1}{n} \sum_{\tau=-\infty}^{\infty} \gamma_X(\tau) \right)$$

Es claro que para obtener un modelo ARMA(p,q) apropiado a partir de unos datos, no podemos utilizar la sucesión de autocorrelaciones teórica (ACF) ni la sucesión de autocovarianzas teórica, por lo que debemos utilizar estimaciones. La estimación de $\gamma_X(\tau)$ y $\rho_X(\tau)$ se logra usando la sucesión de autocovarianzas muestrales y la sucesión de autocorrelaciones muestrales (ACF). Como comentario, debemos decir que para que los estimadores correspondientes a ambas sucesiones funcionen, el tamaño de la muestra, n , debe ser mayor o igual a 50. Para $n < 50$ los estimadores de $\gamma_X(\tau)$ y $\rho_X(\tau)$ se vuelven poco confiables al explicar las propiedades de un proceso $\{X_t\}$.

La sucesión de autocovarianzas muestrales está dada por:

$$\hat{\gamma}_X(\tau) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-\tau} (X_t - \bar{X}_n) (X_{t+\tau} - \bar{X}_n) & \text{si } 0 \leq \tau \leq n-1 \\ \hat{\gamma}_X(-\tau) & \text{si } -n < \tau \leq 0 \end{cases}$$

mientras que la sucesión de autocorrelaciones muestrales (ACF) está dada por:

$$\hat{\rho}_X(\tau) = \frac{\hat{\gamma}_X(\tau)}{\hat{\gamma}_X(0)} \quad (3.5.32)$$

El estimador $\hat{\gamma}_X(\tau)$ es sesgado, pero su distribución asintótica cuando $n \rightarrow \infty$ tiene media $\gamma_X(\tau)$. La sucesión de autocovarianzas muestrales se pueden expresar con una matriz, la matriz de autocovarianzas muestrales, $\hat{\Gamma}_n$, dada por:

$$\hat{\Gamma}_n = \begin{pmatrix} Cov(X_1, X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Cov(X_2, X_2) & \cdots & Cov(X_2, X_n) \\ Cov(X_3, X_1) & Cov(X_3, X_2) & \cdots & Cov(X_3, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & Cov(X_n, X_n) \end{pmatrix} = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{n-1} \\ \hat{\gamma}_{-1} & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{n-2} \\ \hat{\gamma}_{-2} & \hat{\gamma}_{-1} & \cdots & \hat{\gamma}_{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{-(n-1)} & \hat{\gamma}_{-(n-2)} & \cdots & \hat{\gamma}_0 \end{pmatrix}$$

pero como $\gamma_X(\tau) = \gamma_X(-\tau)$, al ser estacionario el proceso $\{X_t\}$, tenemos:

$$\hat{\Gamma}_n = \begin{pmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{n-1} \\ \hat{\gamma}_1 & \hat{\gamma}_0 & \cdots & \hat{\gamma}_{n-2} \\ \hat{\gamma}_2 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{n-1} & \hat{\gamma}_{n-2} & \cdots & \hat{\gamma}_0 \end{pmatrix}$$

En consecuencia, la matriz de autocorrelaciones muestrales es:

$$\hat{\rho}_n = \begin{pmatrix} 1 & \hat{\rho}_1 & \cdots & \hat{\rho}_{n-1} \\ \hat{\rho}_1 & 1 & \cdots & \hat{\rho}_{n-2} \\ \hat{\rho}_2 & \hat{\rho}_1 & \cdots & \hat{\rho}_{n-3} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\rho}_{n-1} & \hat{\rho}_{n-2} & \cdots & 1 \end{pmatrix}$$

Recordemos que para modelos MA(q) vistos en la **Sección 3.3.3** la sucesión de autocorrelaciones teórica (ACF) es cero para $|\tau| > q$. La sucesión de autocorrelaciones teórica (ACF) nos brinda gran información acerca del grado q de los procesos de promedios móviles. Por lo anterior, para ver si unos datos tienen un proceso MA(q) debemos ver si las autocorrelaciones $\hat{\rho}_\tau$ satisfacen ser casi cero para $\tau > q$ y distintas de cero para $\tau \leq q$.

Debemos recalcar que la sucesión de autocorrelaciones teórica (ACF) solamente funciona para identificar modelos MA(q). Si el proceso es AR(p) o ARMA(p, q) debemos encontrar una función que se comporte de manera similar a la sucesión de autocorrelaciones teórica (ACF), a saber, la sucesión de autocorrelaciones parciales (PACF).

Para comprender y motivar aún más la idea de la sucesión de autocorrelaciones parciales (PACF) consideremos un modelo AR(1) causal dado por $Y_t = \phi_1 Y_{t-1} + \epsilon_t$; al calcular la sucesión de autocorrelaciones teórica (ACF) para el retraso 2 obtenemos:

$$\begin{aligned} \gamma_Y(2) &= Cov(Y_t, Y_{t-2}) \\ &= Cov(\phi_1 Y_{t-1} + \epsilon_t, Y_{t-2}) \\ &= Cov(\phi_1[\phi_1 Y_{t-2} + \epsilon_{t-1}] + \epsilon_t, Y_{t-2}) \\ &= Cov(\phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t, Y_{t-2}) \\ &= Cov(\phi_1^2 Y_{t-2}, Y_{t-2}) + Cov(\phi_1 \epsilon_{t-1}, Y_{t-2}) + Cov(\epsilon_t, Y_{t-2}) \\ &= \phi_1^2 Cov(Y_{t-2}, Y_{t-2}) + \phi_1 Cov(\epsilon_{t-1}, Y_{t-2}) + Cov(\epsilon_t, Y_{t-2}) \\ &= \phi_1^2 \gamma_Y(0) \end{aligned} \tag{3.5.33}$$

ya que la observación Y_{t-2} está relacionada con los términos $\{\epsilon_{t-2}, \epsilon_{t-3}, \epsilon_{t-4}, \dots\}$ pero no con los términos $\{\epsilon_t, \epsilon_{t-1}\}$. Observamos que la correlación entre Y_t y Y_{t-2} , $\gamma_Y(2) \neq 0$, a diferencia de lo que ocurre para un modelo MA(1), ya que Y_t depende de Y_{t-2} a través de Y_{t-1} . Imaginemos que removemos de manera “parcial” el efecto de la observación Y_{t-1} . Por tanto:

$$Cov(Y_t, Y_{t-2}) = Cov(\epsilon_t, Y_{t-2} - \phi_1 Y_{t-1}) = 0 \tag{3.5.34}$$

Ejemplo 3.5.14 A continuación mostramos la sucesión de autocorrelaciones muestrales (ACF) de los distintos procesos simulados a lo largo de este capítulo. La **Figura 3.5.4** muestra dichas sucesiones.

La representación de $\hat{\rho}_\tau$ es mediante autocorrelogramas; en el eje x se tienen los retrasos o “lags” y en el eje y las respectivas autocorrelaciones.

Las líneas azules representan las bandas de confianza. Dichas bandas de confianza se usarán posteriormente para decidir el valor de q en un proceso MA(q). Por el momento bastará decir que escogeremos aquel retraso q para el cual, $\forall \tau > q$, las autocorrelaciones no se salen de dichas bandas de confianza.

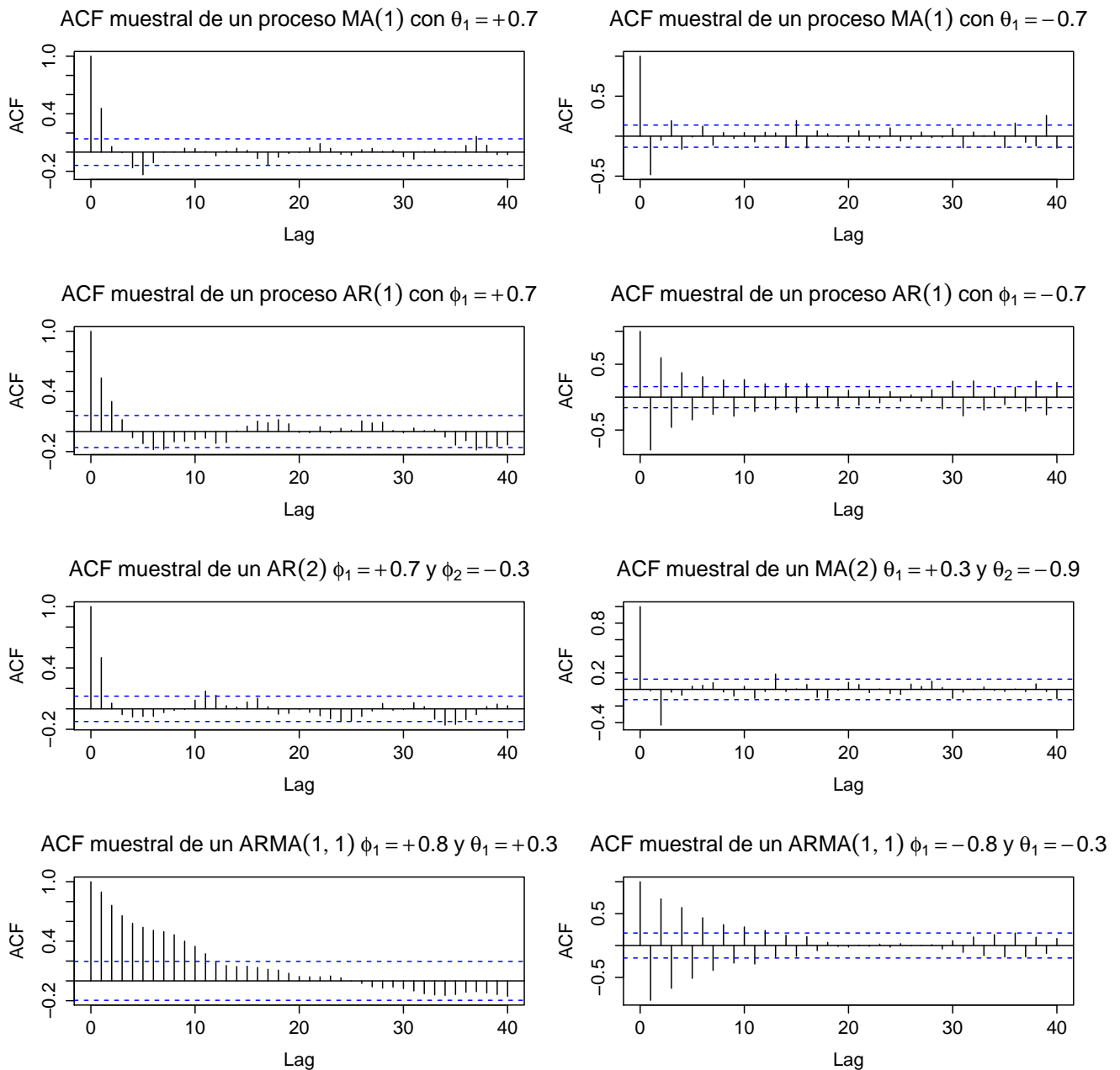


Figura 3.5.4: Figura que representa las ACF muestrales de varios procesos simulados

El código correspondiente a este ejemplo viene dado en R por:

```
> acf(arima.sim(list(order=c(0,0,1),ma=.7),n=200),lag.max=40,
main=(expression(ACF~muestral~de~un~proceso~MA(1)~con~theta[1]==+.7)))
> acf(arima.sim(list(order=c(0,0,1),ma=-.7),n=200),lag.max=40,
main=(expression(ACF~muestral~de~un~proceso~MA(1)~con~theta[1]==-.7)))
> acf(arima.sim(list(order=c(1,0,0),ar=.7),n=150),lag.max=40,
main=(expression(ACF~muestral~de~un~proceso~AR(1)~con~phi[1]==+.7)))
> acf(arima.sim(list(order=c(1,0,0),ar=-.7),n=150),lag.max=40,
main=(expression(ACF~muestral~de~un~proceso~AR(1)~con~phi[1]==-.7)))
> acf(arima.sim(list(order=c(0,0,2),ma=c(0.3,-0.9)),n=250),lag.max=40,
main=(expression(ACF~muestral~de~un~MA(2)~theta[1]==+.3~y~theta[2]==-.9)))
> acf(arima.sim(list(order=c(2,0,0),ar=c(0.7,-0.3)),n=250),lag.max=40,
main=(expression(ACF~muestral~de~un~AR(2)~phi[1]==+.7~y~phi[2]==-.3)))
```



```
> acf(arima.sim(list(order=c(1,0,1),ar=-0.8,ma=-0.3),n=100),lag.max=40,
      main=(expression(ACF~muestral~de~un~ARMA(1,1)~phi[1]==-.8~y~theta[1]==-.3)))
> acf(arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.3),n=100),lag.max=40,
      main=(expression(ACF~muestral~de~un~ARMA(1,1)~phi[1]==+.8~y~theta[1]==+.3)))
```

En general, la idea básica para determinar la parte autorregresiva de un modelo será la de calcular la correlación existente entre X_t y X_s sin tomar en cuenta el efecto de observaciones intermedias. Más adelante explicaremos a detalle toda la teoría relacionada con la sucesión de autocorrelaciones parciales (PACF).

Ahora bien, para seleccionar un modelo ARMA(p,q) apropiado a partir de unos datos, debemos tener la capacidad de reconocer cuando ciertas autocorrelaciones son distintas de 0. Para lograr esta meta debemos conocer la distribución asintótica conjunta de $\{\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_\tau\}$ conforme $n \rightarrow \infty$.

Teorema 3.5.4 Si $\{X_t\}$ es un proceso estacionario y

$$X_t - \mu = \sum_{j=-\infty}^{\infty} \psi_j \epsilon_{t-j} \quad (3.5.35)$$

donde $\{\epsilon_t\} \sim IID(0, \sigma_\epsilon^2)$, se cumple la **ec. 3.5.6** y $\mathbb{E}(\epsilon_t^4) < \infty$, entonces para cada τ :

$$\hat{\rho}_n(\tau) \sim N\left(\rho_{\mathbf{X}}(\tau), \frac{1}{n}W\right) \quad (3.5.36)$$

conforme $n \rightarrow \infty$ y donde $\hat{\rho}_n(\tau)$ es el vector de autocorrelaciones estimadas con un retraso τ que depende del tamaño de la muestra, $\rho_{\mathbf{X}}(\tau) = (\rho_1, \rho_2, \dots, \rho_\tau)'$ y la matriz W está formada por la **fórmula de Bartlett** dada por:

$$w_{ij} = \sum_{k=-\infty}^{\infty} \{\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} + 2\rho_i\rho_j\rho_k^2 - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i}\} \quad (3.5.37)$$

o bien, después de un poco de álgebra:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho_{k+i} + \rho_{k-i} - 2\rho_k\rho_i\} \times \{\rho_{k+j} + \rho_{k-j} - 2\rho_k\rho_j\} \quad (3.5.38)$$

Ejemplo 3.5.15 Si $\{X_t\} \sim IID(0, \sigma_X^2)$ entonces $\rho_l = 0 \forall |l| \geq 1$ y $\rho_0 = 1$, así que en la expresión w_{ij} solo permanecen los términos en los cuales $k = i = j$, de tal forma que:

$$w_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

conforme $n \rightarrow \infty$.

Esto quiere decir, que para n grande $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \dots, \hat{\rho}_\tau$ son aproximadamente independientes e idénticamente distribuidas con distribución normal, media 0 y varianza $\frac{1}{n}$. Si graficamos la sucesión de autocorrelaciones muestrales (ACF), $\hat{\rho}_X(\tau)$ como función de τ , entonces, aproximadamente el 95% de las autocorrelaciones deberán estar dentro de las bandas $\pm \frac{1.96}{\sqrt{n}}$. La aseveración anterior puede usarse para comprobar que las observaciones pertenecen a un conjunto de variables aleatorias independientes e idénticamente distribuidas.

Supongamos $\{X_t\}$ como en el **Ejemplo 3.5.15**. Ahora que conocemos la distribución de $\hat{\rho}_X(\tau)$ podemos obtener sus respectivos intervalos de confianza al $100(1 - \alpha)\%$ dados por:

$$\left(\hat{\rho}_j - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \hat{\rho}_j + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) \quad (3.5.39)$$

con $Z_{1-\frac{\alpha}{2}}$ el cuantil de una distribución Normal estándar que acumula una probabilidad de $1 - \frac{\alpha}{2}$. Para la prueba de hipótesis $H_0 : \rho_j = 0$ vs $H_1 : \rho_j \neq 0$ para $j = 1, 2, \dots, \tau$, se rechaza H_0 si el cero no está dentro del intervalo de confianza, es decir, si:

$$\hat{\rho}_j < -\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{o bien} \quad \hat{\rho}_j > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

Ahora bien, si $\{X_t\}$ no fuese ruido blanco, $\hat{\rho}_i \sim N\left(\rho_i, \frac{w_{ii}}{n}\right)$ y el intervalo de confianza para ρ_i al $100(1 - \alpha)\%$ para n suficientemente grande será:

$$\left(\hat{\rho}_i - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{w_{ii}}{n}}, \hat{\rho}_i + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{w_{ii}}{n}} \right) \quad (3.5.40)$$

Para este proceso la prueba de hipótesis es $H_0 : \rho_i = 0$ vs $H_1 : \rho_i \neq 0$ y rechazamos H_0 si el cero no está dentro del intervalo de confianza, es decir, si:

$$\hat{\rho}_i < -Z_{1-\frac{\alpha}{2}} \sqrt{\frac{w_{ii}}{n}} \quad \text{o bien} \quad \hat{\rho}_i > Z_{1-\frac{\alpha}{2}} \sqrt{\frac{w_{ii}}{n}}$$

Observemos que en este caso, para cada i tendríamos una banda de confianza distinta. Se puede ver que, en general, la banda de confianza para el ruido blanco es más chica que las bandas para modelos MA(q), por eso en la práctica se usa la banda del ruido blanco aunque el proceso $\{X_t\}$ no lo sea, para así decidir cuándo se rechaza la hipótesis $H_0 : \rho_i = 0$ vs $H_1 : \rho_i \neq 0$. Si se rechaza $H_0 : \rho_\tau = 0$ para $\tau = 1, 2, \dots, q$ y no se rechaza $H_0 : \rho_\tau = 0$ para $\tau > q$ entonces hablamos de un modelo MA(q), ya que en éste, las autocorrelaciones a partir de “ q ” deben ser cero; desde el punto de vista de las bandas de confianza, esto significa que las ρ_τ se salen de las bandas de confianza hasta “ q ” y a partir de allí quedan dentro.

Como ya fue mencionado anteriormente, en el caso de la parte autorregresiva se usan las autocorrelaciones parciales para decidir qué “ p ” usar para el modelo AR(p). La autocorrelación parcial mide el grado de asociación entre 2 variables aleatorias sin considerar el efecto de una variable o un conjunto de ellas.

Se define la autocorrelación parcial entre dos variables X e Y dada una variable Z como:

$$\rho_{xy,z} = \frac{\rho_{xy} - \rho_{xz}\rho_{yz}}{\sqrt{1 - \rho_{xz}^2}\sqrt{1 - \rho_{yz}^2}} \quad (3.5.41)$$

Ejemplo 3.5.16 Para un modelo $AR(1)$ estacionario $X_t = \phi_1 X_{t-1} + \epsilon_t$, ya habíamos obtenido $\rho_\tau = \phi_1^\tau$. La autocorrelación parcial entre X_t y X_{t-2} es:

$$\begin{aligned} \rho_{X_t X_{t-2}, X_{t-1}} &= \frac{\rho_{X_t X_{t-2}} - \rho_{X_t X_{t-1}} \rho_{X_{t-2} X_{t-1}}}{\sqrt{1 - \rho_{X_t X_{t-1}}^2} \sqrt{1 - \rho_{X_{t-2} X_{t-1}}^2}} = \frac{\rho_2 - \rho_1^2}{\sqrt{1 - \rho_1^2} \sqrt{1 - \rho_1^2}} \\ &= \frac{\phi_1^2 - \phi_1^2}{\sqrt{1 - \phi_1^2} \sqrt{1 - \phi_1^2}} = 0 \end{aligned}$$

es decir, en un modelo $AR(1)$ la autocorrelación parcial a 2 pasos es cero, o bien, la autocorrelación entre X_t y X_{t-2} controlando X_{t-1} es cero.

Ejemplo 3.5.17 Las ecuaciones de Yule-Walker dadas por la **ec. 3.4.18** para un modelo $AR(2)$ definido por $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$ son:

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 \end{aligned}$$

Despejando ρ_1 y ρ_2 y calculando la autocorrelación parcial entre X_t y X_{t-2} controlando X_{t-1} obtenemos:

$$\begin{aligned} \rho_{X_t X_{t-2}, X_{t-1}} &= \frac{\rho_{X_t X_{t-2}} - \rho_{X_t X_{t-1}} \rho_{X_{t-2} X_{t-1}}}{\sqrt{1 - \rho_{X_t X_{t-1}}^2} \sqrt{1 - \rho_{X_{t-2} X_{t-1}}^2}} = \frac{\rho_2 - \rho_1^2}{\sqrt{1 - \rho_1^2} \sqrt{1 - \rho_1^2}} \\ &= \frac{\frac{\phi_1^2}{1 - \phi_2} + \phi_2 - \frac{\phi_1^2}{(1 - \phi_2)^2}}{\sqrt{1 - \left(\frac{\phi_1}{1 - \phi_2}\right)^2} \sqrt{1 - \left(\frac{\phi_1}{1 - \phi_2}\right)^2}} = \phi_2 \end{aligned}$$

Es decir, la autocorrelación parcial entre X_t y X_{t-2} controlando X_{t-1} es ϕ_2 .

Para construir la sucesión de autocorrelaciones parciales (PACF) para series de tiempo estacionarias con media 0, sea $\hat{X}_{t+\tau}$ aquella combinación lineal de $\{X_{t+\tau-1}, X_{t+\tau-2}, X_{t+\tau-3}, \dots, X_{t+1}\}$ que minimiza el error cuadrático medio $\mathbb{E}[(X_{t+\tau} - \alpha_1 X_{t+1} - \alpha_2 X_{t+2} - \dots - \alpha_{\tau-1} X_{t+\tau-1})^2]$. Podemos escribir a $\hat{X}_{t+\tau}$ como:

$$\hat{X}_{t+\tau} = \beta_1 X_{t+\tau-1} + \beta_2 X_{t+\tau-2} + \dots + \beta_{\tau-1} X_{t+1}$$

Es decir, $\hat{X}_{t+\tau}$ representa la regresión de $X_{t+\tau}$ sobre $\{X_{t+\tau-1}, X_{t+\tau-2}, X_{t+\tau-3}, \dots, X_{t+1}\}$. Dicha regresión no posee intercepto ya que la esperanza del proceso $\{X_t\}$ es 0. Podemos simularmente denotar a \hat{X}_t como $\hat{X}_t = \beta_1 X_{t+1} + \beta_2 X_{t+2} + \dots + \beta_{\tau-1} X_{t+\tau-1}$. Ya que el proceso es estacionario entonces los coeficientes $\{\beta_1, \beta_2, \dots, \beta_{\tau-1}\}$ para $\hat{X}_{t+\tau}$ y para \hat{X}_t son los mismos.

Definición 3.5.4 La sucesión de autocorrelaciones parciales (PACF) para un proceso $\{X_t\}$ estacionario, denotada como $\phi_{\tau\tau}$, para $\tau = 1, 2, \dots$ está dada por:

$$\phi_{11} = \text{Corr}(X_{t+1}, X_t) \quad (3.5.42)$$

para $\tau = 1$ y

$$\phi_{\tau\tau} = \text{Corr}(X_{t+\tau} - \hat{X}_{t+\tau}, X_t - \hat{X}_t) \quad (3.5.43)$$

para $\tau \geq 2$. La sucesión de autocorrelaciones parciales (PACF) es la correlación entre $X_{t+\tau}$ y X_t sin tomar en cuenta las variables intermedias $\{X_{t+1}, X_{t+2}, \dots, X_{t+\tau-1}\}$.

Ejemplo 3.5.18 Tomemos un proceso AR(1) dado por $X_t = \phi_1 X_{t-1} + \epsilon_t$, recordando claro que $|\phi_1| < 1$. De acuerdo con la **Definición 3.5.4**, $\phi_{11} = \text{Corr}(X_{t+1}, X_t) = \rho_1 = \phi_1$. Para calcular ϕ_{22} consideremos $\hat{X}_{t+2} = \beta_1 X_{t+1}$ y $\hat{X}_t = \beta_1 X_{t+1}$. Vamos a escoger aquella β_1 que minimice tanto $\mathbb{E}[(X_{t+2} - \hat{X}_{t+2})^2]$ como $\mathbb{E}[(X_t - \hat{X}_t)^2]$, es decir:

$$\begin{aligned} \mathbb{E}[(X_{t+2} - \hat{X}_{t+2})^2] &= \mathbb{E}[(X_{t+2} - \beta_1 X_{t+1})^2] \\ &= \gamma(0) - 2\beta_1 \gamma(1) + \beta_1^2 \gamma(0) \\ \mathbb{E}[(X_t - \hat{X}_t)^2] &= \mathbb{E}[(X_t - \beta_1 X_{t+1})^2] \\ &= \gamma(0) - 2\beta_1 \gamma(1) + \beta_1^2 \gamma(0) \end{aligned}$$

Observamos que obtenemos la misma ecuación para minimizar β_1 . Por tanto, derivando con respecto a β_1 e igualando a cero obtenemos $\hat{\beta}_1 = \frac{\gamma(1)}{\gamma(0)} = \rho_1 = \phi_1$. Por tanto ϕ_{22} la podemos calcular como:

$$\begin{aligned} \phi_{22} &= \text{Corr}(X_{t+2} - \hat{X}_{t+2}, X_t - \hat{X}_t) \\ &= \text{Corr}(X_{t+2} - \hat{\beta}_1 X_{t+1}, X_t - \hat{\beta}_1 X_{t+1}) \\ &= \text{Corr}(X_{t+2} - \phi_1 X_{t+1}, X_t - \phi_1 X_{t+1}) \\ &= \text{Corr}(\epsilon_{t+2}, X_t - \phi_1 X_{t+1}) \\ &= 0 \end{aligned}$$

ya que ϵ_{t+2} no está correlacionada con $\{X_t, X_{t+1}\}$.

Asumamos que partimos de un conjunto de observaciones $\{X_1, X_2, X_3, \dots, X_n\}$ para un modelo ARMA(p,q) causal e invertible donde por el momento, el orden p y q del modelo autorregresivo y de promedios móviles, respectivamente, son conocidos.

Nuestro objetivo ahora será el de estimar los parámetros $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q$ y σ_ϵ^2 . Comencemos con el método de estimadores por momentos. La idea básica es igualar los momentos poblacionales a los respectivos muestrales.

En la **Sección 3.4.3** estudiamos a los procesos AR(p). De dicha sección nos interesa recordar las ecuaciones de Yule-Walker, dadas por la **ec. 3.4.18**.

Hemos visto cómo surgen dichas ecuaciones, pero ahora, con un mayor conocimiento de los temas que involucran al análisis de series de tiempo, nos gustaría profundizar un poco más para ver de dónde surge la idea de las ecuaciones de Yule-Walker y así, comprender la teoría necesaria para estimar al conjunto $\{\phi_1, \phi_2, \dots, \phi_p, \sigma_\epsilon^2\}$.

Sea $\{X_t\}$ un proceso ARMA(p,q) causal. El proceso puede escribirse de acuerdo con la **ec. 3.5.7**, es decir, en términos de una serie de pesos $\{\psi_j\}$ y la secuencia de ruido blanco $\{\epsilon_t\}$. Al aplicar el operador esperanza es claro que el proceso $\{X_t\}$ tiene media 0. Además, la sucesión de autocovarianzas teórica del proceso puede describirse como:

$$\gamma_X(\tau) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+\tau}$$

gracias a la **Proposición 3.5.1**, ya que estamos construyendo una nueva serie a partir de la secuencia $\{\epsilon_t\}$. Observamos que nuestra siguiente meta será la de obtener la serie de pesos $\{\psi_j\}$. Nuestra pregunta inmediata será ¿cómo obtener dicha serie de pesos? Nuestra primera opción sería utilizar la **ec. 3.5.15** y la **ec. 3.5.16** pero también podríamos obtener la serie de pesos $\{\psi_j\}$ resolviendo directamente la sucesión de autocovarianzas teórica del proceso asociado. Utilizando el hecho de que $\forall \tau \geq 0$:

$$\begin{aligned} Cov(\epsilon_{t+\tau-j}, X_t) &= Cov\left(\epsilon_{t+\tau-j}, \sum_{k=0}^{\infty} \psi_k \epsilon_{t-k}\right) \\ &= \sigma_\epsilon^2 \psi_{j-\tau} \end{aligned}$$

podemos escribir $\gamma_X(\tau)$ como:

$$\begin{aligned} Cov(X_{t+\tau}, X_t) &= Cov\left(\sum_{j=1}^p \phi_j X_{t+\tau-j} + \sum_{j=0}^q \theta_j \epsilon_{t+\tau-j}, X_t\right) \\ &= \sum_{j=1}^p \phi_j \gamma_X(\tau-j) + \sigma_\epsilon^2 \sum_{j=0}^q \theta_j \psi_{j-\tau} \end{aligned}$$

Podemos expresar la ecuación anterior para $\tau \geq \max(p, q+1)$ obteniendo:

$$\gamma_X(\tau) - \phi_1 \gamma_X(\tau-1) - \dots - \phi_p \gamma_X(\tau-p) = 0 \quad (3.5.44)$$

con condiciones iniciales:

$$\gamma_X(\tau) - \sum_{j=1}^p \phi_j \gamma_X(\tau-j) = \sigma_\epsilon^2 \sum_{j=0}^q \theta_j \psi_{j-\tau} \quad 0 \leq \tau < \max(p, q+1) \quad (3.5.45)$$

En particular, para un AR(p) causal, se tiene de la **ec. 3.5.45**, $0 \leq \tau < \max\{p, 1\}$ y $\theta_j = 0$ excepto cuando $j = \tau = 0$ en cuyo caso se tiene $\gamma_X(0) - \phi_1 \gamma_X(1) - \dots - \phi_p \gamma_X(p) = \sigma_\epsilon^2$, ya que $\gamma_X(-\tau) = \gamma_X(\tau)$.

Para $\tau = 1, 2, \dots, p-1$ se tendría $\gamma_X(\tau) = \phi_1\gamma_X(\tau-1) + \dots + \phi_p\gamma_X(\tau-p)$ ya que $\theta_j = 0$ para $j > 0$ al ser un modelo AR(p) y no contener elementos del modelo MA(q) salvo θ_0 . De la **ec. 3.5.44** se obtienen $\gamma_X(p) - \phi_1\gamma_X(p-1) + \dots + \phi_p\gamma_X(0) = 0$. Por lo tanto, para un AR(p), las primeras $p+1$ ecuaciones de la **ec. 3.5.44** y la **ec. 3.5.45** son las llamadas ecuaciones de Yule-Walker:

$$\begin{aligned} \sigma_\epsilon^2 &= \gamma_X(0) - \phi_1\gamma_X(1) - \dots - \phi_p\gamma_X(p) \\ \tau = 1 \quad \gamma_X(1) &= \phi_1\gamma_X(0) + \phi_2\gamma_X(1) + \dots + \phi_p\gamma_X(p-1) \\ \tau = 2 \quad \gamma_X(2) &= \phi_1\gamma_X(1) + \phi_2\gamma_X(0) + \dots + \phi_p\gamma_X(p-2) \\ &\vdots \\ \tau = p \quad \gamma_X(p) &= \phi_1\gamma_X(p-1) + \phi_2\gamma_X(p-2) + \dots + \phi_p\gamma_X(0) \end{aligned}$$

En términos matriciales, las ecuaciones de Yule-Walker están dadas por:

$$\sigma_\epsilon^2 = \gamma_X(0) - \Phi'\gamma_p \quad \text{y} \quad \gamma_p = \Gamma_p\Phi \quad (3.5.46)$$

donde

$$\begin{aligned} \Phi &= (\phi_1, \phi_2, \dots, \phi_p)' \\ \Gamma_p &= [\gamma_X(i-j)]_{i,j=1}^p \\ \gamma_p &= (\gamma_1, \gamma_2, \dots, \gamma_p)' \end{aligned}$$

Usando el método de momentos, podemos sustituir $\gamma(\tau)$ por $\hat{\gamma}(\tau)$ y resolviendo para $\hat{\Phi}$ y $\hat{\sigma}_\epsilon^2$ obtenemos:

$$\hat{\sigma}_\epsilon^2 = \hat{\gamma}_X(0) - \hat{\gamma}'_p \hat{\Gamma}_p^{-1} \hat{\gamma}_p \quad \text{y} \quad \hat{\Phi} = \hat{\Gamma}_p^{-1} \hat{\gamma}_p \quad (3.5.47)$$

o en términos de autocorrelaciones:

$$\hat{\sigma}_\epsilon^2 = \hat{\gamma}_X(0)[1 - \hat{\rho}'_p \hat{R}_p^{-1} \hat{\rho}_p] \quad \text{y} \quad \hat{\Phi} = \hat{R}_p^{-1} \hat{\rho}_p \quad (3.5.48)$$

donde

$$\begin{aligned} \hat{R}_p &= [\rho_X(i-j)]_{i,j=1}^p \\ \hat{\rho}_p &= (\hat{\rho}_1, \hat{\rho}_2, \dots, \hat{\rho}_p)' \end{aligned}$$

Ejemplo 3.5.19 El número de manchas solares durante 100 periodos fue medido, obteniendo un proceso $\{X_t\}$. Las autocovarianzas muestrales obtenidas a partir de los datos son $\hat{\gamma}_0 = 1382.2$, $\hat{\gamma}_1 = 1114.4$ y $\hat{\gamma}_2 = 591.72$.

Encuentre los estimadores vía Yule-Walker para $\hat{\phi}_1$, $\hat{\phi}_2$ y $\hat{\sigma}_\epsilon^2$ para el modelo dado por la expresión $Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \epsilon_t$, con $\{\epsilon_t\}$ ruido blanco con $\mathbb{E}(\epsilon_t) = 0$ y $\text{Var}(\epsilon_t) = \sigma_\epsilon^2$, donde $\{Y_t\}$ es la serie centrada (quitando la media) $Y_t = X_t - 46.93$, para $t = 1, 2, \dots, 100$. ¿El proceso $\{Y_t\}$ obtenido es causal? ¿Se satisfacen los requisitos para estimar vía Yule-Walker?

Para un modelo AR(2) las ecuaciones de Yule-Walker están dadas por:

$$\begin{aligned}\hat{\sigma}_\epsilon^2 &= \hat{\gamma}_0 - \phi_1 \hat{\gamma}_1 - \phi_2 \hat{\gamma}_2 \\ \hat{\gamma}_1 &= \phi_1 \hat{\gamma}_0 + \phi_2 \hat{\gamma}_1 \\ \hat{\gamma}_2 &= \phi_1 \hat{\gamma}_1 + \phi_2 \hat{\gamma}_0\end{aligned}$$

es decir:

$$\begin{aligned}\hat{\sigma}_\epsilon^2 &= 1382.2 - \phi_1(1114.4) - \phi_2(591.72) \\ 1114.4 &= \phi_1(1382.2) + \phi_2(1114.4) \\ 591.72 &= \phi_1(1114.4) + \phi_2(1382.2)\end{aligned}$$

resolviendo el sistema de ecuaciones para $\hat{\phi}_1$ y $\hat{\phi}_2$ obtenemos:

$$\begin{aligned}\hat{\phi}_2 &= -0.634451149 \\ \hat{\phi}_1 &= 0.80625 - 0.80625(-0.6344511) = 1.317976239 \\ \hat{\sigma}_\epsilon^2 &= 1382.2 - (1.317976239)(1114.4) + (0.6344511488)(591.72) = 288.8647126\end{aligned}$$

El polinomio asociado al proceso $\{Y_t\}$ es $\phi(z) = 0.634451148z^2 - 1.317976239z + 1 = 0$. Aplicando la fórmula general obtenemos las raíces del polinomio autorregresivo dadas por:

$$z = \frac{1.317976239 \pm \sqrt{(1.317976239)^2 - 4(0.6344511488)(1)}}{2(0.6344511488)}$$

cuya norma:

$$\begin{aligned}|z| &= \sqrt{(1.038674)^2 + (0.7053162213)^2} \\ &= \sqrt{1.57631465} \\ &= 1.2555 > 1\end{aligned}$$

∴ Sí se satisfacen los requisitos para la estimación vía Yule-Walker ya que el proceso es un proceso autorregresivo y causal.

Nos queda por demostrar la distribución de los estimadores de Yule-Walker cuando el tamaño de la muestra es grande. Se demostrará que dicha distribución es Normal y además el estimador $\hat{\sigma}_\epsilon^2$ se acerca a su valor real σ_ϵ^2 . Antes de enunciar el teorema donde obtenemos la distribución de los estimadores Yule-Walker debemos introducir el algoritmo de Durbin-Levinson y para ello necesitamos hablar un poco del tema de pronósticos.

Técnicas de pronóstico para modelos ARMA

Sea $\{X_{n+\tau}\}$, $\tau > 0$, un proceso estacionario con media μ y sucesión de autocovarianzas $\{\gamma_X(\tau)\}$ en términos de $\{X_1, X_2, \dots, X_n\}$. Nuestro objetivo será el de encontrar el mejor predictor para $X_{n+\tau}$ dado el conjunto de observaciones $\{X_0 = 1, X_1, X_2, \dots, X_n\}$.

Definición 3.5.5 *El mejor predictor lineal (BLP) de $X_{n+\tau}$ en términos del conjunto de observaciones $\{X_1, X_2, \dots, X_n\}$, se define como:*

$$P_n X_{n+\tau} := a_0 + a_1 X_n + \dots + a_n X_1 \quad (3.5.49)$$

que es la combinación lineal que mejor aproxima a $X_{n+\tau}$, minimizando el error cuadrático medio.

Dicho predictor lineal es la proyección de $X_{n+\tau}$ en el plano generado por la combinación lineal de los términos $\{X_0, X_1, X_2, \dots, X_n\}$. Pero, ¿cómo obtenemos los coeficientes $\{a_0, a_1, \dots, a_n\}$? Para encontrar $P_n X_{n+\tau}$ tenemos que minimizar la función:

$$S(a_0, a_1, \dots, a_n) = \mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})^2] \quad (3.5.50)$$

La función $S(a_0, a_1, \dots, a_n)$ es cuadrática en a_0, a_1, \dots, a_n ; además está acotada por debajo por cero. Por tanto, $\exists a_0, a_1, \dots, a_n$ tales que $S(a_0, a_1, \dots, a_n)$ es mínima. Esto implica resolver:

$$\frac{\partial S(a_0, a_1, \dots, a_n)}{\partial a_j} = 0 \quad \forall j = 0, 1, 2, \dots, n \quad (3.5.51)$$

Por ejemplo, si $j = 0$ obtenemos de la **ec. 3.5.51**:

$$\begin{aligned} \frac{\partial S(a_0, a_1, \dots, a_n)}{\partial a_0} &= \frac{\partial}{\partial a_0} \mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})^2] \\ &= \mathbb{E} \left[\frac{\partial}{\partial a_0} (X_{n+\tau} - P_n X_{n+\tau})^2 \right] \\ &= -2\mathbb{E}[(X_{n+\tau} - a_0 - a_1 X_n - \dots - a_n X_1)] \end{aligned} \quad (3.5.52)$$

es decir, tenemos que resolver:

$$\mathbb{E}[(X_{n+\tau} - a_0 - a_1 X_n - \dots - a_n X_1)] = 0 \quad (3.5.53)$$

Por otro lado:

$$\begin{aligned} \frac{\partial S(a_0, a_1, \dots, a_n)}{\partial a_j} &= \frac{\partial}{\partial a_j} \mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})^2] \\ &= \mathbb{E} \left[\frac{\partial}{\partial a_j} (X_{n+\tau} - P_n X_{n+\tau})^2 \right] \\ &= -2\mathbb{E}[(X_{n+\tau} - a_0 - a_1 X_n - \dots - a_n X_1) X_{n-j+1}] \end{aligned} \quad (3.5.54)$$

Es decir, tenemos que resolver:

$$\mathbb{E}[(X_{n+\tau} - a_0 - a_1X_n - \dots - a_nX_1)X_{n-j+1}] = 0 \quad (3.5.55)$$

Al tomar esperanzas en la **ec. 3.5.53** obtenemos:

$$\mathbb{E}(X_{n+\tau}) = a_0 + \sum_{i=1}^n a_i \mathbb{E}(X_{n-i+1}) \quad (3.5.56)$$

Además $\mathbb{E}(X_{n+\tau}) = \mu$ y al despejar a_0 de la **ec. 3.5.56** obtenemos:

$$a_0 = \mu \left(1 - \sum_{i=1}^n a_i \right) \quad (3.5.57)$$

Al tomar esperanzas en la **ec. 3.5.55**, para $j = 1, 2, \dots, n$, obtenemos:

$$\mathbb{E}(X_{n-j+1}X_{n+\tau}) = a_0\mathbb{E}(X_{n-j+1}) + \sum_{i=1}^n a_i\mathbb{E}(X_{n-j+1}X_{n-i+1}) \quad (3.5.58)$$

o bien:

$$\gamma_X(\tau + j - 1) = \mu a_0 + \sum_{i=1}^n a_i \gamma_X(i - j) \quad (3.5.59)$$

El desarrollo que hemos realizado con la función $S(a_0, a_1, \dots, a_n)$ implica la ortogonalidad de $P_n X_{n+\tau} - X_{n+\tau}$ con respecto al conjunto de variables $\{1, X_1, X_2, \dots, X_n\}$. De hecho, la ortogonalidad de $P_n X_{n+\tau} - X_{n+\tau}$ con respecto a la variable $X_0 = 1$ implica que podemos sustraer μ de todas las variables ($P_n X_{n+\tau}$ y X_i). Por lo tanto, para realizar pronósticos, podemos asumir $\mu = 0$.

La **ec. 3.5.59**, se puede escribir en términos matriciales como:

$$\Gamma_n \mathbf{a}_n = \gamma_n(\tau) \quad (3.5.60)$$

donde

$$\begin{aligned} \mathbf{a}_n &= (a_1, \dots, a_n)' \\ \Gamma_n &= [\gamma_X(i - j)]_{i,j=1}^n \\ \gamma_n(\tau) &= [\gamma_X(\tau), \gamma_X(\tau + 1), \dots, \gamma_X(\tau + n - 1)]' \end{aligned}$$

Por lo tanto, el mejor predictor lineal de $X_{n+\tau}$ es aquel que cumple:

$$P_n X_{n+\tau} = \mu + \sum_{i=1}^n a_i (X_{n-i+1} - \mu) \quad (3.5.61)$$

con a_n tal que resuelve la **ec. 3.5.60**. Notemos que el valor esperado del error de predicción, $\mathbb{E}(X_{n+\tau} - P_n X_{n+\tau})$, es cero ya que:

$$\mathbb{E}(X_{n+\tau} - P_n X_{n+\tau}) = \mathbb{E}(X_{n+\tau}) - \mu - \sum_{i=1}^n a_i \mathbb{E}(X_{n-i+1} - \mu) = 0 \quad (3.5.62)$$

mientras que el error cuadrático medio de la predicción está dado por:

$$\begin{aligned} \mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})^2] &= \mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})(X_{n+\tau} - P_n X_{n+\tau})] \\ &= \mathbb{E}[X_{n+\tau}(X_{n+\tau} - P_n X_{n+\tau})] \\ &= \gamma_X(0) - \gamma'_n(\tau) \Gamma_n^{-1} \gamma_n(\tau) \end{aligned} \quad (3.5.63)$$

Veamos el caso cuando $\tau = 1$, es decir, estudiaremos el mejor predictor lineal a un paso. Deseamos, a partir del conjunto de observaciones $\{X_1, X_2, \dots, X_n\}$, predecir el valor de X_{n+1} . Supongamos por el momento que $\mu = 0$, de tal forma que $a_0 = 0$ y por tanto no se necesita X_0 en el conjunto $\{X_1, X_2, \dots, X_n\}$ ya que la media es cero de entrada. El mejor predictor lineal está dado por:

$$P_n X_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \dots + \phi_{nn} X_1 \quad (3.5.64)$$

Notemos que hemos escrito $\phi_{n,n+1-k}$ en lugar de a_k para $k = 1, 2, \dots, n$. El conjunto $\{\phi_{n1}, \phi_{n2}, \phi_{n3}, \dots, \phi_{nn}\}$ satisface, en alusión a la **ec. 3.5.59**:

$$\gamma_X(k) = \sum_{i=1}^n \phi_{ni} \gamma_X(k-i) \quad (3.5.65)$$

para $k = 1, 2, \dots, n$. La **ec. 3.5.65**, conocida como ecuaciones de predicción, se puede escribir en términos matriciales como:

$$\Gamma_n \Phi_n = \gamma_n(1) \quad (3.5.66)$$

donde

$$\begin{aligned} \Phi_n &= (\phi_{n1}, \phi_{n2}, \dots, \phi_{nn})' \\ \Gamma_n &= [\gamma_X(k-i)]_{k,i=1}^n \\ \gamma_n(1) &= [\gamma_X(1), \gamma_X(2), \dots, \gamma_X(n)]' \end{aligned}$$

Gracias al **Teorema de proyección ortogonal**, el mejor predictor lineal de X_{n+1} , $P_n X_{n+1}$, es único. Si Γ_n es no singular, entonces los elementos Φ_n son únicos y están dados por

$$\Phi_n = \Gamma_n^{-1} \gamma_n(1) \quad (3.5.67)$$

El error cuadrático medio de la predicción, caso particular de la **ec. 3.5.64** para $\tau = 1$, está dado por:

$$\begin{aligned} \mathbb{E}[(X_{n+1} - P_n X_{n+1})^2] &= \mathbb{E}[(X_{n+1} - \Phi_n' \mathbf{X})^2] \\ &= \mathbb{E}[(X_{n+1} - \gamma_n' \Gamma_n^{-1} \mathbf{X})^2] \\ &= \mathbb{E}[X_{n+1}^2 - 2\gamma_n' \Gamma_n^{-1} \mathbf{X} X_{n+1} + \gamma_n' \Gamma_n^{-1} \mathbf{X} \mathbf{X}' \Gamma_n^{-1} \gamma_n] \quad (3.5.68) \\ &= \gamma_X(0) - 2\gamma_n' \Gamma_n^{-1} \gamma_n + \gamma_n' \Gamma_n^{-1} \Gamma_n \Gamma_n^{-1} \gamma_n \\ &= \gamma_X(0) - \gamma_n' \Gamma_n^{-1} \gamma_n \end{aligned}$$

con $\mathbf{X} = (X_1, X_2, \dots, X_n)'$.

Ejemplo 3.5.20 Consideremos la serie estacionaria $AR(1)$ dada por $X_t = \phi X_{t-1} + \epsilon_t$ (es decir, $\mu = 0$), donde $t \in \mathbb{Z}$, $|\phi| < 1$ y $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. De lo visto anteriormente, el mejor predictor lineal a un paso de X_{n+1} en términos de $\{1, X_1, X_2, \dots, X_n\}$ está dado por:

$$P_n X_{n+1} = \phi_{n1} X_n + \phi_{n2} X_{n-1} + \dots + \phi_{nn} X_1$$

El conjunto $\{\phi_{n1}, \phi_{n2}, \phi_{n3}, \dots, \phi_{nn}\}$ satisface la **ec. 3.5.66**, a saber:

$$\begin{bmatrix} \gamma_X(0) & \gamma_X(1) & \gamma_X(2) & \dots & \gamma_X(n-1) \\ \gamma_X(1) & \gamma_X(0) & \gamma_X(1) & \dots & \gamma_X(n-2) \\ \gamma_X(2) & \gamma_X(1) & \gamma_X(0) & \dots & \gamma_X(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_X(n-1) & \gamma_X(n-2) & \gamma_X(n-3) & \dots & \gamma_X(0) \end{bmatrix} \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \phi_{n3} \\ \vdots \\ \phi_{nn} \end{bmatrix} = \begin{bmatrix} \gamma_X(1) \\ \gamma_X(2) \\ \gamma_X(3) \\ \vdots \\ \gamma_X(n) \end{bmatrix}$$

o bien en términos de autocorrelaciones:

$$\begin{bmatrix} 1 & \phi & \phi^2 & \dots & \phi^{n-1} \\ \phi & 1 & \phi & \dots & \phi^{n-2} \\ \phi^2 & \phi & 1 & \dots & \phi^{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \phi^{n-1} & \phi^{n-2} & \phi^{n-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{n1} \\ \phi_{n2} \\ \phi_{n3} \\ \vdots \\ \phi_{nn} \end{bmatrix} = \begin{bmatrix} \phi \\ \phi^2 \\ \phi^3 \\ \vdots \\ \phi^n \end{bmatrix}$$

Supongamos que queremos predecir X_3 dadas dos observaciones, X_1 y X_2 . Entonces el mejor predictor lineal a un paso para X_3 está dado por:

$$P_2 X_{2+1} = \phi_{21} X_2 + \phi_{22} X_1$$

Para encontrar ϕ_{21} y ϕ_{22} tendríamos que resolver:

$$\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}^{-1} \begin{bmatrix} \phi \\ \phi^2 \end{bmatrix}$$

pero claramente:

$$\begin{aligned} \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}^{-1} \begin{bmatrix} \phi \\ \phi^2 \end{bmatrix} &= \frac{1}{1-\phi^2} \begin{bmatrix} 1 & -\phi \\ -\phi & 1 \end{bmatrix} \begin{bmatrix} \phi \\ \phi^2 \end{bmatrix} \\ &= \frac{1}{1-\phi^2} \begin{bmatrix} \phi - \phi^3 \\ \phi^2 - \phi^2 \end{bmatrix} \\ &= \begin{bmatrix} \phi \\ 0 \end{bmatrix} \end{aligned}$$

Así pues, para $n \geq 2$, se puede demostrar que una solución al sistema dado por la **ec. 3.5.67** es $\Phi_n = (\phi, \dots, 0)'$. Entonces el mejor predictor lineal a un paso para X_{n+1} en términos de $\{1, X_1, X_2, \dots, X_n\}$ está dado por:

$$P_n X_{n+1} = \phi X_n$$

Recordemos que la varianza para un modelo $AR(1)$ está dada por la **ec. 3.4.5** mientras que la sucesión de autocovarianzas teórica viene dada por la **ec. 3.4.6**, por lo que el error cuadrático medio está dado por:

$$\begin{aligned} \mathbb{E}[(X_{n+1} - P_n X_{n+1})^2] &= \gamma_X(0) - \Phi_n' \gamma_n(1) \\ &= \frac{\sigma_\epsilon^2}{1-\phi^2} - \frac{\sigma_\epsilon^2 \phi^2}{1-\phi^2} \\ &= \sigma_\epsilon^2 \end{aligned}$$

Ejemplo 3.5.21 Supongamos que tenemos un modelo $AR(2)$ dado por $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$ y dos observaciones, X_1 y X_2 . Entonces el mejor predictor lineal a un paso para X_3 está dado por:

$$P_2 X_{2+1} = \phi_{21} X_2 + \phi_{22} X_1$$

A partir de la **ec. 3.5.66** tenemos, para $k = 1, 2$, el sistema de ecuaciones:

$$\begin{aligned} \phi_{21} \gamma(0) + \phi_{22} \gamma(1) &= \gamma(1) \\ \phi_{21} \gamma(1) + \phi_{22} \gamma(0) &= \gamma(2) \end{aligned}$$

que podemos expresar también en términos de la **ec. 3.5.67** como:

$$\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix}$$

Además:

$$\begin{bmatrix} \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(0) \end{bmatrix}^{-1} \begin{bmatrix} \gamma(1) \\ \gamma(2) \end{bmatrix} = \frac{1}{\gamma^2(0) - \gamma^2(1)} \begin{bmatrix} \gamma(0)\gamma(1) - \gamma(1)\gamma(2) \\ \gamma(0)\gamma(2) - \gamma^2(1) \end{bmatrix}$$

Utilizando las ecuaciones de Yule-Walker para un modelo $AR(2)$, es decir aquellas dadas por la **ec. 3.4.17**, obtenemos que:

$$\begin{aligned} \gamma(1) &= \frac{\phi_1\gamma(0)}{1 - \phi_2} \\ \gamma(2) &= \phi_2\gamma(0) + \phi_1\gamma(1) \end{aligned}$$

Realizando un poco de álgebra obtenemos:

$$\begin{aligned} \gamma^2(0) - \gamma^2(1) &= \frac{\gamma^2(0)}{(1 - \phi_2)^2} [1 - 2\phi_2 + \phi_2^2 - \phi_1^2] \\ \gamma(0)\gamma(1) - \gamma(1)\gamma(2) &= \frac{\gamma^2(0)\phi_1}{(1 - \phi_2)^2} [1 - 2\phi_2 + \phi_2^2 - \phi_1^2] \\ \gamma(0)\gamma(2) - \gamma^2(1) &= \frac{\gamma^2(0)\phi_2}{(1 - \phi_2)^2} [1 - 2\phi_2 + \phi_2^2 - \phi_1^2] \end{aligned}$$

por lo que:

$$\begin{bmatrix} \phi_{21} \\ \phi_{22} \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$$

Así pues el mejor predictor lineal a un paso para X_3 dadas dos observaciones, X_1 y X_2 , está dado por $P_2X_3 = \phi_1X_2 + \phi_2X_1$. De hecho, se puede comprobar que para $n \geq 2$:

$$P_nX_{n+1} = \phi_1X_n + \phi_2X_{n-1}$$

donde $\phi_{n1} = \phi_1$, $\phi_{n2} = \phi_2$ y $\phi_{nj} = 0$ para $j = 3, 4, \dots, n$.

Como comentario final, si nosotros quisiéramos calcular el mejor predictor lineal a τ pasos, $\tau \geq 1$ y $\mu = 0$, utilizando al conjunto $\{\phi_{n1}, \phi_{n2}, \phi_{n3}, \dots, \phi_{nn}\}$, entonces, dado el conjunto de observaciones $\{X_1, X_2, \dots, X_n\}$, dicho predictor será:

$$P_nX_{n+1} = \phi_{n1}^m X_n + \phi_{n2}^m X_{n-1} + \dots + \phi_{nn}^m X_1 \quad (3.5.69)$$

donde el conjunto $\{\phi_{n1}^m, \phi_{n2}^m, \phi_{n3}^m, \dots, \phi_{nn}^m\}$ satisface las ecuaciones de predicción:

$$\gamma_X(\tau + k - 1) = \sum_{i=1}^n \phi_{ni}^m \gamma_X(k - i) \quad (3.5.70)$$

para $k = 1, 2, \dots, n$.

En términos matriciales tenemos:

$$\Gamma_n \Phi_n^m = \gamma_n^m(1) \quad (3.5.71)$$

donde

$$\begin{aligned} \Phi_n^m &= (\phi_{n1}^m, \phi_{n2}^m, \dots, \phi_{nn}^m)' \\ \Gamma_n &= [\gamma_X(k-i)]_{k,i=1}^n \\ \gamma_n^m(1) &= [\gamma_X(m), \gamma_X(m+1), \dots, \gamma_X(m+n-1)]' \end{aligned}$$

El error cuadrático medio de la predicción está dado por:

$$\mathbb{E}[(X_{n+\tau} - P_n X_{n+\tau})^2] = \gamma_X(0) - \gamma_n^m(1) \Gamma_n^{-1} \gamma_n^m(1) \quad (3.5.72)$$

El procedimiento anterior no resulta ser tan sencillo para los modelos ARMA(p,q) ya que involucra el uso de matrices de gran tamaño. Para evitar lo anterior existen soluciones iterativas que no necesitan el cálculo de matrices y sus respectivas inversas. Uno de ellos es el algoritmo de Durbin-Levinson, que enunciamos a continuación.

Proposición 3.5.4 *La ec. 3.5.67 y la ec. 3.5.72 pueden resolverse de forma iterativa mediante el algoritmo de Durbin-Levinson definiendo $\phi_{00} = 0$, $P_1^0 = \gamma(0)$ y para $n \geq 1$:*

$$\phi_{nn} = \frac{\rho(n) - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(n-k)}{1 - \sum_{k=1}^{n-1} \phi_{n-1,k} \rho(k)} \quad (3.5.73)$$

y

$$P_{n+1}^n = P_n^{n-1} (1 - \phi_{nn}^2) \quad (3.5.74)$$

donde para $n \geq 2$

$$\phi_{nk} = \phi_{n-1,k} - \phi_{nn} \phi_{n-1,n-k} \quad (3.5.75)$$

para $k = 1, 2, \dots, n-1$.

Ejemplo 3.5.22 *Sea $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$. A partir de las ecuaciones de Yule-Walker podemos obtener $\{\rho(1), \rho(2)\}$.*

Dichas autocorrelaciones serán:

$$\begin{aligned}\rho(1) &= \phi_1 + \rho(1)\phi_2 \\ \rho(2) &= \phi_1\rho(1) + \phi_2\end{aligned}$$

Utilizando el algoritmo de Durbin-Levinson podemos obtener $\{\phi_{11}, \phi_{22}, \phi_{33}, \phi_{21}, \phi_{32}, \phi_{31}\}$. Comencemos con la obtención de $\{\phi_{11}, \phi_{22}, \phi_{33}\}$ de la siguiente forma:

$$\begin{aligned}\phi_{11} &= \rho(1) = \frac{\phi_1}{1 - \phi_2} \\ \phi_{22} &= \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2} = \frac{\left[\frac{\phi_1^2}{1 - \phi_2} + \phi_2\right] - \left[\frac{\phi_1}{1 - \phi_2}\right]^2}{1 - \left[\frac{\phi_1}{1 - \phi_2}\right]^2} = \phi_2 \\ \phi_{33} &= \frac{\rho(3) - \phi_1\rho(2) - \phi_2\rho(1)}{1 - \phi_1\rho(1) - \phi_2\rho(2)} = 0\end{aligned}$$

Finalmente para $\{\phi_{21}, \phi_{32}, \phi_{31}\}$ tenemos:

$$\begin{aligned}\phi_{21} &= \phi_{11} - \phi_{22}\phi_{11} = \rho(1)[1 - \phi_2] = \phi_1 \\ \phi_{32} &= \phi_{22} - \phi_{33}\phi_{21} = \phi_2 \\ \phi_{31} &= \phi_{21} - \phi_{33}\phi_{22} = \rho(1)[1 - \phi_2] = \phi_1\end{aligned}$$

Existe otro enfoque para encontrar al conjunto $\{\phi_{n1}, \phi_{n2}, \phi_{n3}, \dots, \phi_{nn}\}$. El fondo de dicho enfoque es similar al que hemos usado pero se realiza a partir del **Teorema de proyección ortogonal** y el uso del producto interior. Queremos encontrar una combinación lineal \hat{X}_{k+1} :

$$\hat{X}_{k+1} = \sum_{j=1}^k \phi_{kj} X_{k+1-j} \quad (3.5.76)$$

que aproxime “lo mejor posible” a X_{k+1} en el sentido de que la distancia $\|X_{k+1} - \hat{X}_{k+1}\|^2$ sea mínima. Dicha distancia está definida como:

$$\|X_{k+1} - \hat{X}_{k+1}\|^2 = \mathbb{E} \left[\left| X_{k+1} - \sum_{j=1}^k \phi_{kj} X_{k+1-j} \right|^2 \right] \quad (3.5.77)$$

Buscamos una combinación lineal cuya distancia hacia X_{k+1} sea mínima. Sea

$$\mu = \sum_{j=1}^k \alpha_j X_{k+1-j} = \langle \overline{X_1, X_2, \dots, X_k} \rangle$$

Es menester decir que $\alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R}$ y donde el operador $\langle \rangle$ denota el producto interior. En este contexto la notación incluyendo la raya superior indica el espacio generado por el conjunto X_1, \dots, X_k . Se puede aplicar el **Teorema de proyección ortogonal**, el cual nos garantiza la existencia de dicha proyección, $\hat{X}_{k+1} = P_\mu(X_{k+1})$, proyección sobre μ de X_{k+1} , la cual es un elemento en μ y satisface $\hat{X}_{k+1} - X_{k+1} \in \mu^\perp$, es decir, $\hat{X}_{k+1} \in \mu$ y además $\langle X_{k+1} - \hat{X}_{k+1}, Y \rangle = 0 \forall Y \in \mu$ y donde $\langle \rangle$ denota un producto interior, el cual en este caso queda en función de las covarianzas.

En particular para $X_i \in \mu$ con $i = 1, 2, \dots, k$ tenemos:

$$\begin{aligned}
 i = 1, 2, \dots, k & \quad 0 = \langle X_{k+1} - \sum_{j=1}^k \phi_{kj} X_{k+1-j}, X_i \rangle \\
 i = 1, 2, \dots, k & \quad \langle X_{k+1}, X_i \rangle = \langle \sum_{j=1}^k \phi_{kj} X_{k+1-j}, X_i \rangle \\
 i = 1, 2, \dots, k & \quad \langle X_{k+1}, X_i \rangle = \sum_{j=1}^k \phi_{kj} \langle X_{k+1-j}, X_i \rangle \\
 i = 1, 2, \dots, k & \quad Cov(X_{k+1}, X_i) = \sum_{j=1}^k \phi_{kj} Cov(X_{k+1-j}, X_i)
 \end{aligned}$$

A partir de aquí obtenemos, para $i = 1, 2, \dots, k$:

$$Cov(X_{k+1}, X_i) = \phi_{k1} Cov(X_k, X_i) + \phi_{k2} Cov(X_{k-1}, X_i) + \dots + \phi_{kk} Cov(X_1, X_i)$$

Éstas corresponden nuevamente a las llamadas ecuaciones de predicción. Las ecuaciones de predicción en su 1^{er} forma están dadas por:

$$\begin{aligned}
 i = k & \quad \gamma_X(1) = \phi_{k1} \gamma_X(0) + \phi_{k2} \gamma_X(1) + \dots + \phi_{kk} \gamma_X(k-1) \\
 i = k-1 & \quad \gamma_X(2) = \phi_{k1} \gamma_X(1) + \phi_{k2} \gamma_X(0) + \dots + \phi_{kk} \gamma_X(k-2) \\
 i = k-2 & \quad \gamma_X(3) = \phi_{k1} \gamma_X(2) + \phi_{k2} \gamma_X(1) + \dots + \phi_{kk} \gamma_X(k-3) \\
 & \quad \vdots \\
 i = 1 & \quad \gamma_X(k) = \phi_{k1} \gamma_X(k-1) + \phi_{k2} \gamma_X(k-2) + \dots + \phi_{kk} \gamma_X(0)
 \end{aligned}$$

o en términos de autocorrelaciones:

$$\begin{aligned}
 i = k & \quad \rho_X(1) = \phi_{k1} + \phi_{k2} \rho_X(1) + \dots + \phi_{kk} \rho_X(k-1) \\
 i = k-1 & \quad \rho_X(2) = \phi_{k1} \rho_X(1) + \phi_{k2} + \dots + \phi_{kk} \rho_X(k-2) \\
 i = k-2 & \quad \rho_X(3) = \phi_{k1} \rho_X(2) + \phi_{k2} \rho_X(1) + \dots + \phi_{kk} \rho_X(k-3) \\
 & \quad \vdots \\
 i = 1 & \quad \rho_X(k) = \phi_{k1} \rho_X(k-1) + \phi_{k2} \rho_X(k-2) + \dots + \phi_{kk}
 \end{aligned}$$

Observe que:

$$X_{k+1} - \hat{X}_{k+1} = X_{k+1} - P_{\langle X_1, X_2, \dots, X_k \rangle}(X_{k+1}) \quad (3.5.78)$$

representa al residual de la regresión de X_{k+1} con respecto al conjunto de observaciones $\{X_1, X_2, \dots, X_k\}$. De forma similar podemos calcular la regresión de X_{k+1} con respecto a $\{X_2, X_3, \dots, X_k\}$ y la regresión de X_1 con respecto a $\{X_2, X_3, \dots, X_k\}$, dadas por:

$$X_{k+1} - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_{k+1}) \quad \text{y} \quad X_1 - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_1)$$

Ahora bien, considerando de la **ec. 3.5.43** a $\tau + t = k$ (esto puede hacerse debido a que el proceso es estacionario), tenemos:

$$PACF(k) = Corr(X_{k+1} - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_{k+1}), X_1 - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_1))$$

Esta correlación es justamente el término ϕ_{kk} , de aquí que podemos estimar a $PACF(k)$ como $\widehat{PACF}(k) = \hat{\phi}_{kk}$. La proposición que asegura lo anterior la enunciamos a continuación. La demostración la podemos encontrar en Brockwell y Davis (1991).

Proposición 3.5.5 *Si $\{X_t\}$ es un proceso estacionario con media cero y sucesión de autocovarianzas $\gamma_X(\tau)$ tal que $\gamma_X(0) > 0$ y $\gamma_X(\tau) \rightarrow 0$ si $\tau \rightarrow \infty$, entonces:*

$$\phi_{kk} = Corr(X_{k+1} - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_{k+1}), X_1 - P_{\langle X_2, X_3, \dots, X_k \rangle}(X_1)) \quad (3.5.79)$$

En particular aplica a los procesos ARMA(p,q) causales ya que son estacionarios, tienen media cero y $\gamma_X(\tau) \rightarrow 0$.

Hasta aquí termina el subtema de pronósticos. Regresemos a la obtención de la distribución de los estimadores Yule-Walker. Para modelos AR(p), si el tamaño de la muestra es significativamente grande entonces los estimadores Yule-Walker tiene una distribución aproximadamente Normal y el estimador $\hat{\sigma}_\epsilon^2$ es muy cercano al valor real σ_ϵ^2 . Lo anterior se enuncia en el siguiente teorema.

Teorema 3.5.5 *Si $\{X_t\}$ es un proceso AR(p) causal, entonces el comportamiento asintótico, conforme $n \rightarrow \infty$, de los estimadores Yule-Walker es:*

$$\sqrt{n}(\hat{\Phi} - \Phi) \sim N(0, \sigma_\epsilon^2 \Gamma_p^{-1}) \quad (3.5.80)$$

En particular para $k > p$:

$$\sqrt{n}\hat{\phi}_{kk} \sim N(0, 1) \quad (3.5.81)$$

Podemos utilizar el algoritmo de Durbin-Levinson para calcular $\hat{\Phi}$ reemplazando $\rho(\tau)$ por $\hat{\rho}(\tau)$. De hecho, al realizar este procedimiento lo que hacemos es calcular $\hat{\Phi}_k = (\hat{\phi}_{k1}, \hat{\phi}_{k2}, \dots, \hat{\phi}_{kk})'$. Esto quiere decir que el algoritmo de Durbin-Levinson nos sirve para calcular la sucesión de autocorrelaciones parciales (PACF). A partir de aquí, se pueden obtener intervalos de confianza para ϕ_{kk} al $100(1 - \alpha)\%$ dados por:

$$\left(\hat{\phi}_{kk} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \hat{\phi}_{kk} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \right) \quad (3.5.82)$$

con $Z_{1-\frac{\alpha}{2}}$ el cuantil de una distribución Normal estándar que acumula una probabilidad de $1 - \frac{\alpha}{2}$. Para la prueba de hipótesis $H_0 : \phi_{kk} = 0$ vs $H_1 : \phi_{kk} \neq 0$ para $j = 1, 2, \dots, \tau$, se rechaza H_0 si el cero no está dentro del intervalo de confianza, es decir, si:

$$\hat{\phi}_{kk} < -\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} \quad \text{o bien} \quad \hat{\phi}_{kk} > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

Escogeremos por tanto, aquella “p” a partir de la cual no se rechaza la hipótesis nula de que los términos ϕ_{kk} relacionados con las autocorrelaciones parciales son cero. Así encontraremos el valor de p para los modelos AR(p); debemos observar hasta qué valor de p se rechaza H_0 y a partir del valor $p + 1$ no se rechaza H_0 . Englobando todo lo anterior, con la sucesión de autocorrelaciones muestrales (ACF) decidimos el grado “q” de la parte MA(q) y con la sucesión de autocorrelaciones parciales (PACF) decidimos el grado “p” de la parte AR(p) en un modelo ARMA(p,q).

Ejemplo 3.5.23 De manera similar al Ejemplo 3.5.14, mostramos la sucesión de autocorrelaciones muestrales parciales (PACF) de los distintos procesos simulados a lo largo de este capítulo. La Figura 3.5.5 muestra dichas sucesiones. Las líneas azules representan las bandas de confianza. El código correspondiente a este ejemplo viene dado en R por:

```
> pacf(arima.sim(list(order=c(0,0,1),ma=.7),n=200),lag.max=40,
  main=(expression(PACF~muestral~de~un~proceso~MA(1)~con~theta[1]==+.7)))
> pacf(arima.sim(list(order=c(0,0,1),ma=-.7),n=200),lag.max=40,
  main=(expression(PACF~muestral~de~un~proceso~MA(1)~con~theta[1]==-.7)))
> pacf(arima.sim(list(order=c(1,0,0),ar=.7),n=150),lag.max=40,
  main=(expression(PACF~muestral~de~un~proceso~AR(1)~con~phi[1]==+.7)))
> pacf(arima.sim(list(order=c(1,0,0),ar=-.7),n=150),lag.max=40,
  main=(expression(PACF~muestral~de~un~proceso~AR(1)~con~phi[1]==-.7)))
> pacf(arima.sim(list(order=c(0,0,2),ma=c(0.3,-0.9),n=250),lag.max=40,
  main=(expression(PACF~muestral~de~un~MA(2)~theta[1]==+.3~y~theta[2]==-.9)))
> pacf(arima.sim(list(order=c(2,0,0),ar=c(0.7,-0.3),n=250),lag.max=40,
  main=(expression(PACF~muestral~de~un~AR(2)~phi[1]==+.7~y~phi[2]==-.3)))
> pacf(arima.sim(list(order=c(1,0,1),ar=-0.8,ma=-0.3),n=100),lag.max=40,
  main=(expression(PACF~muestral~de~un~ARMA(1,1)~phi[1]==-.8~y~theta[1]==-.3)))
> pacf(arima.sim(list(order=c(1,0,1),ar=0.8,ma=0.3),n=100),lag.max=40,
  main=(expression(PACF~muestral~de~un~ARMA(1,1)~phi[1]==+.8~y~theta[1]==+.3)))
```

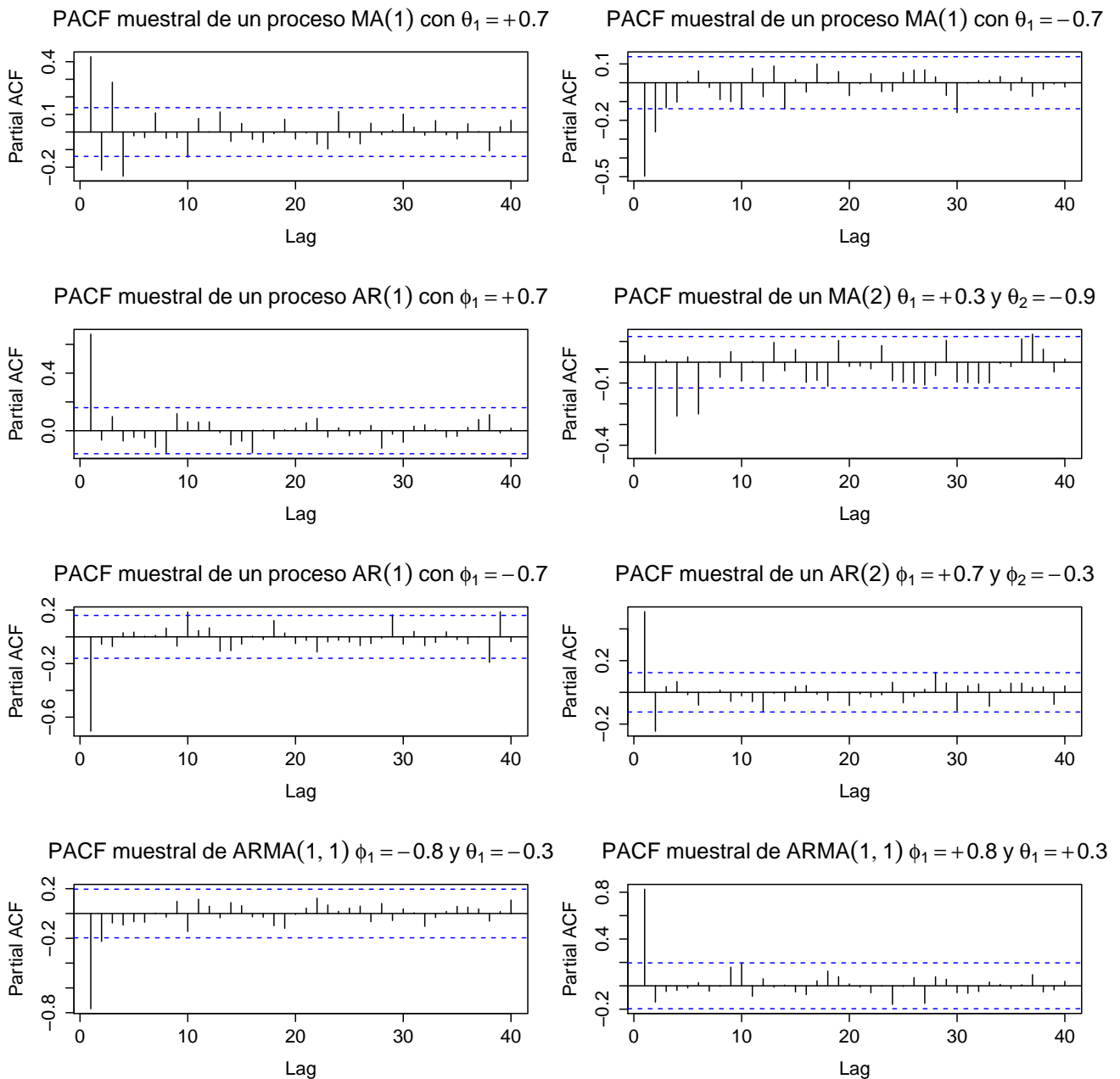


Figura 3.5.5: Figura que representa las PACF muestrales de varios procesos simulados

Ejemplo 3.5.24 Los siguientes valores corresponden a la sucesión de autocorrelaciones muestrales (ACF) y a la sucesión de autocorrelaciones muestrales parciales (PACF), para una serie de tiempo con 100 observaciones una vez que a la serie original se le aplicó el operador ∇^2 (es decir, se le aplicaron 2 diferencias). Identifique un modelo para la serie.

lag k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$.302	.198	.005	-0.179	-0.171	-0.149	.056	.011	-0.055	.101	.161	.022

Tabla 3.5.1: Tabla que muestra la sucesión de autocorrelaciones muestrales para $\nabla^2 X_t$

lag k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\phi}_{kk}$.388	.200	.097	-0.164	-0.052	-0.039	.033	.015	-0.125	.153	.091	.166

Tabla 3.5.2: Tabla que muestra la sucesión de autocorrelaciones parciales para $\nabla^2 X_t$

Para la parte de promedios móviles usamos la pruebas de hipótesis $H_0 : \rho_i = 0$ vs $H_1 : \rho_i \neq 0$ y consideramos las bandas de confianza para un ruido blanco. Las primeras “q” autocorrelaciones no deben ser cero para después serlo.

Se rechaza H_0 si $|\hat{\rho}_i| > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$, donde $Z_{1-\frac{\alpha}{2}}$ es el cuantil de una distribución Normal con media 0 y varianza 1, de aquí que:

$$\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} = \frac{1.96}{\sqrt{100}} = 0.196$$

Solo para los retrasos 1 y 2, ocurre que tanto $0.302 > 0.196$ y $0.198 > 0.196$, el resto caen dentro de la banda de confianza, así que se puede suponer $q = 2$. Para la parte autorregresiva se usa la sucesión de autocorrelaciones parciales (PACF).

La prueba de hipótesis es $H_0 : \phi_{kk} = 0$ vs $H_1 : \phi_{kk} \neq 0$ y se rechaza H_0 si $|\hat{\phi}_{kk}| > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$, de aquí que de nuevo:

$$\frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} = \frac{1.96}{\sqrt{100}} = 0.196$$

Nuevamente para los retrasos 1 y 2, ocurre que tanto $0.388 > 0.196$ como $0.200 > 0.196$, el resto caen dentro de la banda de confianza, así que se puede suponer $p = 2$.

∴ Un modelo plausible para $\nabla^2 X_t$ es $ARMA(2,2)$.

Ejemplo 3.5.25 Los siguientes valores corresponden a la sucesión de autocorrelaciones muestrales (ACF) y a la sucesión de autocorrelaciones muestrales parciales (PACF), para un proceso $Y_t = \nabla X_t$ donde $\{X_t\}$ es la serie original; sin embargo, observe que las primeras autocorrelaciones parciales no fueron calculadas. Sabiendo que la serie ∇X_t tiene 65 observaciones, calcule los PACF faltantes, es decir, $\hat{\phi}_{11}$ y $\hat{\phi}_{22}$. Una vez encontradas dichas autocorrelaciones proporcione un modelo $ARMA(p,q)$ plausible. Justificar la respuesta usando las bandas de confianza o pruebas de hipótesis asociados adecuados, considerando un 95% de confianza.

lag k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k$.546	.187	-0.050	-0.314	-0.239	-0.169	-0.060	.002	.001	.128	.148	.029

Tabla 3.5.3: Tabla que muestra la sucesión de autocorrelaciones muestrales para ∇X_t

lag k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\phi}_{kk}$	—	—	-0.069	-0.253	-0.078	-0.090	-0.111	-0.146	.042	.183	.102	-0.098

Tabla 3.5.4: Tabla que muestra la sucesión de autocorrelaciones parciales para ∇X_t

Usando las ecuaciones de predicción, para $k = 1$ tenemos $\rho_1 = \phi_{11}$, por lo que $\hat{\phi}_{11} = 0.546$. Ahora bien, para $k = 2$ tenemos el sistema de ecuaciones:

$$\begin{aligned}\hat{\rho}_1 &= \hat{\phi}_{21} + \hat{\phi}_{22}\hat{\rho}_1 \\ \hat{\rho}_2 &= \hat{\phi}_{21}\hat{\rho}_1 + \hat{\phi}_{22}\end{aligned}$$

Resolviendo el sistema de ecuaciones obtenemos los estimadores que buscábamos, a saber:

$$\begin{aligned}\hat{\phi}_{22} &= -0.158 \\ \hat{\phi}_{11} &= 0.546\end{aligned}$$

Para ver la parte de promedios móviles usamos la sucesión de autocorrelaciones muestrales (ACF) y suponemos las bandas de confianza de un ruido blanco. Para $\alpha = 0.05$, $Z_{1-\frac{\alpha}{2}} = 1.96$.

Rechazamos $H_0 : \rho_i = 0$ vs $H_1 : \rho_i \neq 0$ si $|\hat{\rho}_k| > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} = 0.2431$. Los valores que se salen de la banda de confianza son aquellos correspondientes a los retrasos 1 y 4. Por tanto, tomamos $q = 4$ en el modelo $ARMA(p, q)$.

Analizando ahora, la sucesión de autocorrelaciones muestrales parciales (PACF), incluyendo a los valores $\hat{\phi}_{11} = 0.546$ y $\hat{\phi}_{22} = -0.158$ calculados, rechazamos $H_0 : \phi_{kk} = 0$ vs $H_1 : \phi_{kk} \neq 0$ si $|\hat{\phi}_{kk}| > \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{n}} = 0.2431$.

Rechazo H_0 para $k = 1$ y $k = 4$ ya que:

$$\begin{aligned}|\hat{\phi}_{11}| &= 0.546 > 0.2431 \\ |\hat{\phi}_{44}| &= 0.253 > 0.2431\end{aligned}$$

Podría elegir una parte autorregresiva $p = 4$ ya que a partir de aquí toda la sucesión de autocorrelaciones parciales (PACF) es cero.

\therefore El proceso puede ser modelado con un modelo $ARMA(4, 4)$.

En el caso de los modelos $MA(q)$ no hay un sistema de ecuaciones lineales, similar al obtenido mediante Yule-Walker, que establezca relaciones entre los parámetros, en este caso, $\theta_1, \theta_2, \dots, \theta_q$ y las covarianzas (o correlaciones). A partir de la sucesión de autocorrelaciones teórica (ACF) de un modelo $MA(q)$, tenemos el sistema no lineal:

$$\begin{aligned}\rho_1 &= \frac{\theta_1 + \theta_1\theta_2 + \dots + \theta_{q-1}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \\ \rho_2 &= \frac{\theta_2 + \theta_1\theta_3 + \dots + \theta_{q-2}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} \\ &\vdots \\ \rho_q &= \frac{\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}\end{aligned}$$

el cual puede tener múltiples soluciones, aunque solamente una satisfará invertibilidad. Por lo anterior, la estimación de los parámetros del modelo $MA(q)$ se hace mediante el llamado **algoritmo de innovación**, el cual en su caso más general también permite obtener estimadores para un modelo $ARMA(p, q)$.

Dicho algoritmo se aplica a todas las series con segundos momentos finitos ya sea que sean estacionarios o no. Suponga un proceso $\{X_t\}$ centrado con $\mathbb{E}|X_t|^2 < \infty \forall t$ y:

$$\mathbb{E}(X_t, X_s) = \kappa(t, s) \tag{3.5.83}$$

Denotemos al mejor predictor lineal a un paso y su error cuadrático medio como:

$$\hat{X}_n = \begin{cases} 0 & \text{si } n = 1 \\ P_n X_{n+1} & \text{si } n = 2, 3, \dots \end{cases}$$

y $\nu_n = \mathbb{E} = (X_{n+1} - P_n X_{n+1})^2$. Introducimos las innovaciones o errores de predicción a un paso como $U_n = X_n - \hat{X}_n$. Ahora bien, en términos de los vectores $\mathbf{U}_n = (U_1, U_2, \dots, U_n)'$ y $\mathbf{X}_n = (X_1, X_2, \dots, X_n)'$, las ecuaciones anteriores se pueden escribir como:

$$\mathbf{U}_n = \mathbb{A}_n \mathbf{X}_n \quad (3.5.84)$$

donde \mathbb{A}_n está dada por:

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

Si $\{X_t\}$ es un proceso estacionario entonces $a_{ij} = -a_j$ tomados de la **ec. 3.5.60** con $\tau = 1$. La inversa de \mathbb{A}_n , denotada como \mathbb{C}_n está dada por:

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

El vector de predictores a un paso $\hat{\mathbf{X}}_n := (\mathbf{X}_1, \mathbf{P}_1 \mathbf{X}_2, \dots, \mathbf{P}_{n-1} \mathbf{X}_n)'$ se puede expresar como:

$$\begin{aligned} \hat{\mathbf{X}}_n &= \mathbf{X}_n - \mathbf{U}_n \\ &= \mathbb{C}_n \mathbf{U}_n - \mathbf{U}_n \\ &= \Theta_n (\mathbf{X}_n - \hat{\mathbf{X}}_n) \end{aligned} \quad (3.5.85)$$

donde

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

Además, \mathbf{X}_n satisface por sí misma:

$$\mathbf{X}_n = \mathbf{C}_n(\mathbf{X}_n - \hat{\mathbf{X}}_n) \quad (3.5.86)$$

La ec. 3.5.86 puede escribirse como:

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

y a partir de ésta, los predictores a un paso $\hat{X}_1, \hat{X}_2, \dots$ pueden ser calculados de manera recursiva una vez que los coeficientes θ_{ij} han sido determinados. La siguiente proposición nos ayuda a generar dichos coeficientes y sus correspondientes errores cuadráticos $\mathbb{E}(X_{i+1} - P_n X_{i+1})^2$ comenzando con las covarianzas $\kappa(i, j)$.

Proposición 3.5.6 *Los coeficientes $\theta_{n1}, \dots, \theta_{nn}$ se pueden obtener de manera recursiva utilizando las siguientes ecuaciones:*

$$\nu_0 = \kappa(1, 1) \quad (3.5.87)$$

$$\theta_{n,n-k} = \nu_k^{-1} \left(\kappa(n+1, k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j} \theta_{n,n-j} \nu_j \right) \quad (3.5.88)$$

para $0 \leq k < n$ y

$$\nu_n = \kappa(n+1, n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 \nu_j \quad (3.5.89)$$

Ejemplo 3.5.26 *Si $\{X_t\}$ es un proceso definido por $X_t = \epsilon_t + \theta_1 \epsilon_{t+1}$ con $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$ entonces $\kappa(i, j) = 0$ para $|i - j| > 1$, $\kappa(i, i) = \sigma_\epsilon^2(1 + \theta_1^2)$ y $\kappa(i, i+1) = \theta_1 \sigma_\epsilon^2$. La aplicación del algoritmo de innovación nos lleva a las siguientes ecuaciones:*

$$\begin{aligned} \theta_{nj} &= 0; & (\text{para } 0 \leq 2 < n) \\ \theta_{n1} &= \nu_{n-1}^{-1} \theta_1 \sigma_\epsilon^2 \\ \nu_0 &= \sigma_\epsilon^2(1 + \theta_1^2) \\ \nu_n &= \sigma_\epsilon^2(1 + \theta_1^2 - \nu_{n-1}^{-1} \theta_1^2 \sigma_\epsilon^2) \end{aligned}$$

Tomemos el caso $X_t = \epsilon_t + 0.8\epsilon_{t+1}$ con $\{\epsilon_t\} \sim WN(0, 1)$. Los coeficientes θ_{nj} para $1 \leq j \leq n$ utilizando el algoritmo de innovación serán:

$$\begin{aligned}
 \nu_0 &= 1(1 + 0.8^2) = 1.6400 \\
 \theta_{11} &= (\nu_0^{-1})(0.8)(1) = 0.4878 \\
 \nu_1 &= 1[1 + 0.8^2 - (\nu_0^{-1})(0.8^2)(1)] = 1.2498 \\
 \theta_{21} &= (\nu_1^{-1})(0.8)(1) = 0.6401 \\
 \theta_{22} &= 0 \\
 \nu_2 &= 1[1 + 0.8^2 - (\nu_1^{-1})(0.8^2)(1)] = 1.1279 \\
 \theta_{31} &= (\nu_2^{-1})(0.8)(1) = 0.7093 \\
 \theta_{32} &= 0 \\
 \theta_{33} &= 0 \\
 \nu_3 &= 1[1 + 0.8^2 - (\nu_2^{-1})(0.8^2)(1)] = 1.0726 \\
 \theta_{41} &= (\nu_3^{-1})(0.8)(1) = 0.7459 \\
 \theta_{42} &= 0 \\
 \theta_{43} &= 0 \\
 \theta_{44} &= 0 \\
 \nu_4 &= 1[1 + 0.8^2 - (\nu_3^{-1})(0.8^2)(1)] = 1.0433
 \end{aligned}$$

Note que conforme n se incrementa, ν_n se aproxima a la varianza del ruido blanco y θ_{n1} se aproxima a θ_1 . Estos resultados serán ciertos para cualquier proceso $MA(1)$ con $|\theta_1| < 1$. El algoritmo de innovación se ajusta de manera eficiente para pronosticar procesos $MA(q)$ ya que para dichos procesos, $\theta_{nj} = 0$ para $n - j > q$. Para procesos $AR(p)$ el algoritmo de Durbin-Levinson es preferible ya que $\phi_{nj} = 0$ para $n - j > p$.

Mediante las ecuaciones de Yule-Walker y el algoritmo de innovación podemos estimar los parámetros $\theta_1, \dots, \theta_q$ y ϕ_1, \dots, ϕ_p ; sin embargo, estas estimaciones no tienen las mejores propiedades estadísticas, así que se prefiere hacer estimaciones usando máxima verosimilitud y los valores estimados mediante Yule-Walker y el algoritmo de innovación se usan como valores iniciales en los métodos que usan máxima verosimilitud.

3.5.4. Estimaciones usando máxima verosimilitud

La mayor parte de este capítulo ha sido dedicado al cálculo de propiedades estadísticas y pronósticos de una serie de tiempo dado el conjunto $\{\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2\}$, el cual es conocido. En esta sección utilizaremos las estimaciones por máxima verosimilitud para encontrar el conjunto anterior a partir de unos datos. Supongamos que tenemos un conjunto de observaciones $\{x_1, x_2, \dots, x_N\}$. Denotemos además a $\Theta = (\mu, \phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \sigma^2)'$. La idea básica será la de calcular la función de densidad asociada dada por:

$$f_{X_N, \dots, X_1}(x_N, \dots, x_1; \Theta) \quad (3.5.90)$$

Es decir, la función de densidad representa la probabilidad de haber observado al conjunto $\{x_1, x_2, \dots, x_N\}$.

El estimador máximo verosímil de Θ es el valor para el cual el conjunto de observaciones $\{x_1, x_2, \dots, x_N\}$ es más probable de ser observado, es decir, es el valor de Θ para el cual se maximiza la **ec. 3.5.90**. Para las estimaciones por máxima verosimilitud vamos a suponer que el ruido blanco $\{\epsilon_t\}$ es Gaussiano, es decir, es un conjunto de variables aleatorias independientes e idénticamente distribuidas con distribución Normal, $\{\epsilon_t\} \sim WN_G(0, \sigma_\epsilon^2)$, y no como antes que sólo se pedía que $\mathbb{E}(\epsilon_t) = 0$ y $Var(\epsilon_t) = \sigma_\epsilon^2$ con las variables aleatorias no correlacionadas.

Comencemos con un proceso AR(1), dado por la **ec. 3.4.1**. Para este caso $\Theta = (\mu, \phi_1, \sigma^2)'$. Consideremos la distribución de X_1 . Es posible demostrar que la esperanza del proceso $\{X_1\}$ está dada por:

$$\mathbb{E}(X_1) = \frac{\mu}{1 - \phi_1} \quad (3.5.91)$$

mientras que la varianza viene dada por:

$$Var(X_1) = \frac{\sigma_\epsilon^2}{1 - \phi_1^2} \quad (3.5.92)$$

Debido a que $\{\epsilon_t\} \sim WN_G(0, \sigma_\epsilon^2)$, entonces $\{X_1\}$ también tiene distribución Normal con esperanza dada por la **ec. 3.5.91** y varianza dada por la **ec. 3.5.92**. Entonces, la función de densidad para la primera observación X_1 está dada por:

$$f_{X_1}(x_1; \Theta) = \sqrt{\frac{1 - \phi_1^2}{2\pi\sigma_\epsilon^2}} e \left[-\frac{\left\{x_1 - \left(\frac{\mu}{1 - \phi_1}\right)\right\}^2 \{1 - \phi_1^2\}}{2\sigma_\epsilon^2} \right] \quad (3.5.93)$$

Ahora bien, consideremos la distribución de la observación X_2 dada la observación $X_1 = x_1$, es decir, al condicionar a X_2 dado $X_1 = x_1$ significará tratar a la observación X_1 como una constante. De aquí que de la **ec. 3.4.1** tenemos a $X_2 = \mu + \phi_1 X_1 + \epsilon_2$ con función de distribución $X_2|X_1 = x_1 \sim N(\mu + \phi_1 x_1, \sigma_\epsilon^2)$. La función de densidad condicional para la observación X_2 dado $X_1 = x_1$ viene dada por:

$$f_{X_2|X_1}(x_2|x_1; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e \left[-\frac{(x_2 - \mu - \phi_1 x_1)^2}{2\sigma_\epsilon^2} \right] \quad (3.5.94)$$

La función de densidad conjunta de las observaciones $\{X_1, X_2\}$ viene dada por:

$$f_{X_2, X_1}(x_2, x_1; \Theta) = f_{X_2|X_1}(x_2|x_1; \Theta) f_{X_1}(x_1; \Theta) \quad (3.5.95)$$

Similarmente podemos encontrar la distribución de la observación X_3 dadas fijas $X_2 = x_2$ y $X_1 = x_1$, a saber:

$$f_{X_3|X_2, X_1}(x_3|x_2, x_1; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e \left[-\frac{(x_3 - \mu - \phi_1 x_2)^2}{2\sigma_\epsilon^2} \right] \quad (3.5.96)$$

La función de densidad conjunta de las observaciones $\{X_1, X_2, X_3\}$ viene dada por:

$$f_{X_3, X_2, X_1}(x_3, x_2, x_1; \Theta) = f_{X_3|X_2, X_1}(x_3|x_2, x_1; \Theta)f_{X_2, X_1}(x_2, x_1; \Theta) \quad (3.5.97)$$

Podemos repetir el mismo proceso hasta encontrar la función de densidad de la observación X_N dado fijo el conjunto $\{X_{N-1} = x_{N-1}, \dots, X_1 = x_1\}$, a saber:

$$f_{X_N|X_{N-1}, \dots, X_1}(x_N|x_{N-1}, \dots, x_1; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{\left[-\frac{(x_N - \mu - \phi_1 x_{N-1})^2}{2\sigma_\epsilon^2}\right]} \quad (3.5.98)$$

Observemos que para el cálculo de la función de densidad de X_N solamente importa el valor de la observación X_{N-1} . La función de densidad conjunta de las observaciones $\{X_1, \dots, X_N\}$ viene dada por:

$$f_{X_N, \dots, X_1}(x_N, \dots, x_1; \Theta) = f_{X_N|X_{N-1}}(x_N|x_{N-1}; \Theta)f_{X_{N-1}, \dots, X_1}(x_{N-1}, \dots, x_1; \Theta) \quad (3.5.99)$$

La función de verosimilitud del conjunto de observaciones $\{X_1, \dots, X_N\}$ puede ser calculada como:

$$f_{X_N, \dots, X_1}(x_N, \dots, x_1; \Theta) = f_{X_1}(x_1; \Theta) \prod_{n=2}^N f_{X_n|X_{n-1}}(x_n|x_{n-1}; \Theta) \quad (3.5.100)$$

mientras que la log-verosimilitud se obtiene al aplicar logaritmo a la función de verosimilitud:

$$\mathbb{L}(\Theta) = \ln[f_{X_1}(x_1; \Theta)] + \sum_{n=2}^N \ln[f_{X_n|X_{n-1}}(x_n|x_{n-1}; \Theta)] \quad (3.5.101)$$

Al sustituir $f_{X_1}(x_1; \Theta)$ y $f_{X_N|X_{N-1}, \dots, X_1}(x_N|x_{N-1}, \dots, x_1; \Theta)$ en la **ec. 3.5.101** obtenemos la versión final de la log-verosimilitud:

$$\begin{aligned} \mathbb{L}(\Theta) &= -\frac{\ln(2\pi)}{2} - \frac{1}{2} \ln\left(\frac{\sigma_\epsilon^2}{1 - \phi_1^2}\right) - \frac{\left[x_1 - \left(\frac{\mu}{1 - \phi_1}\right)\right]^2 [1 - \phi_1^2]}{2\sigma_\epsilon^2} \\ &- \frac{(N-1)\ln(2\pi)}{2} - \frac{(N-1)\ln(\sigma_\epsilon^2)}{2} \\ &- \sum_{n=2}^N \left[\frac{(x_n - \mu - \phi_1 x_{n-1})^2}{2\sigma_\epsilon^2} \right] \end{aligned} \quad (3.5.102)$$

Hemos visto como el condicionar para valores iniciales del conjunto $\{x_1, \dots, x_N\}$ nos facilita la obtención de la función de verosimilitud. El estimador máximo verosímil $\hat{\Phi}$ es aquel valor que se obtiene al maximizar la **ec. 3.5.102**.

Para maximizar debemos diferenciar e igualar el resultado a cero pero, en la práctica obtendríamos un sistema de ecuaciones no lineales sobre Θ y $\{x_1, x_2, \dots, x_N\}$. Debido a esto, maximizar la **ec. 3.5.102** requerirá de procedimientos numéricos. Una alternativa para evitar lo anterior es fijar la observación $X_1 = x_1$ y maximizar la verosimilitud condicionada a dicha primera observación de la siguiente forma:

$$f_{X_N, \dots, X_2 | X_1}(x_N, \dots, x_2 | x_1; \Theta) = \prod_{n=2}^N f_{X_n | X_{n-1}}(x_n | x_{n-1}; \Theta) \quad (3.5.103)$$

es decir, maximizar:

$$\mathbb{L}(\Theta) = -\frac{(N-1)\ln(2\pi)}{2} - \frac{(N-1)\ln(\sigma_\epsilon^2)}{2} - \sum_{n=2}^N \left[\frac{(x_n - \mu - \phi_1 x_{n-1})^2}{2\sigma_\epsilon^2} \right] \quad (3.5.104)$$

Maximizar la **ec. 3.5.104** es equivalente a minimizar:

$$\sum_{n=2}^N [(x_n - \mu - \phi_1 x_{n-1})^2] \quad (3.5.105)$$

por mínimos cuadrados ordinarios. Al diferenciar la **ec. 3.5.104** con respecto a σ_ϵ^2 e igualando a cero obtenemos:

$$\hat{\sigma}_\epsilon^2 = \sum_{n=2}^N \left[\frac{(x_n - \hat{\mu} - \hat{\phi}_1 x_{n-1})^2}{N-1} \right] \quad (3.5.106)$$

¿Nos sirvió de algo condicionar sobre la observación $X_1 = x_1$? Nuestra respuesta es afirmativa. El estimador máximo verosímil condicionado a $X_1 = x_1$ es ahora muy sencillo de calcular. Además, si el tamaño de la muestra es significativamente grande, entonces, la primera observación contribuye muy poco al total de la verosimilitud. También, la distribución de la función de verosimilitud exacta y la condicional es la misma conforme n crece, dado claro que $|\phi_1| < 1$. De hecho, cuando $|\phi_1| > 1$ la verosimilitud condicionada sigue funcionando adecuadamente mientras que la verosimilitud exacta deja de hacerlo.

Obtendremos a continuación la función de verosimilitud condicionada para un modelo MA(1). Consideremos la **ec. 3.3.1** que define a un proceso MA(1). Sea $\Theta = (\mu, \theta_1, \sigma^2)'$ el conjunto de parámetros a estimar. Es fácil demostrar que si ϵ_{n-1} es conocido, entonces $X_n | \epsilon_{n-1} \sim N(\mu + \theta_1 \epsilon_{n-1}, \sigma_\epsilon^2)$ o bien, en términos de su función de densidad:

$$f_{X_n | \epsilon_{n-1}}(x_n | \epsilon_{n-1}; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{-\frac{(x_n - \mu - \theta_1 \epsilon_{n-1})^2}{2\sigma_\epsilon^2}} \quad (3.5.107)$$

Supongamos $\epsilon_0 = 0$, entonces $X_1 | \epsilon_0 \sim N(\mu, \sigma_\epsilon^2)$ y podemos expresar a ϵ_1 en términos de la **ec. 3.3.1** como $\epsilon_1 = X_1 - \mu$.

La correspondiente función de densidad será:

$$f_{X_2|X_1, \epsilon_0}(x_2|x_1, \epsilon_0; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{\left[-\frac{(x_2 - \mu - \theta_1\epsilon_1)^2}{2\sigma_\epsilon^2}\right]} \quad (3.5.108)$$

Ahora bien, si ϵ_1 es conocido, podemos expresar a ϵ_2 en términos de la **ec. 3.3.1** como $\epsilon_2 = X_2 - \mu - \theta_1\epsilon_1$. Podemos realizar este procedimiento una y otra vez y obtener que la secuencia $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$ puede ser calculada a partir del conjunto de observaciones $\{x_1, \dots, x_N\}$ iterando la **ec. 3.3.1** como:

$$\epsilon_n = X_n - \mu - \theta_1\epsilon_{n-1} \quad (3.5.109)$$

para $n = 1, 2, \dots, N$. La función de densidad condicional para la observación X_n viene dada por:

$$f_{X_n|X_{n-1}, \dots, X_1, \epsilon_0}(x_n|x_{n-1}, \dots, x_1, \epsilon_0; \Theta) = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} e^{\left[-\frac{\epsilon_n^2}{2\sigma_\epsilon^2}\right]} \quad (3.5.110)$$

El valor de ϵ_n según la **ec. 3.5.109** depende de ϵ_{n-1} la cual depende de X_{n-1} y este a su vez de ϵ_{n-2} y así sucesivamente. Por eso la función sí está condicionada a los valores de x_{n-1}, \dots, x_1 . La función de verosimilitud será entonces:

$$f_{X_N, \dots, X_1|\epsilon_0}(x_N, \dots, x_1|\epsilon_0; \Theta) = f_{X_1|\epsilon_0}(x_1|\epsilon_0; \Theta) \prod_{n=2}^N f_{X_n|X_{n-1}, \dots, X_1, \epsilon_0}(x_n|x_{n-1}, \dots, x_1, \epsilon_0; \Theta)$$

mientras que la log-verosimilitud viene dada por:

$$\mathbb{L}(\Theta) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma_\epsilon^2) - \sum_{n=1}^N \frac{\epsilon_n^2}{2\sigma_\epsilon^2} \quad (3.5.111)$$

donde el logaritmo de $f_{X_1|\epsilon_0}(x_1|\epsilon_0; \Theta)$ está dado por:

$$-\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln(\sigma_\epsilon^2) - \frac{\epsilon_1^2}{2\sigma_\epsilon^2}$$

y de este modo podemos hacer correr la suma de la **ec. 3.5.111** de $n = 1, \dots, N$. Al diferenciar la **ec. 3.5.111** con respecto a σ_ϵ^2 e igualando a cero obtenemos:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N} \sum_{n=1}^N \epsilon_n^2 \quad (3.5.112)$$

Debemos recalcar que en los últimos dos ejemplos hemos supuesto procesos con media μ y no centrados en 0. De ser requerido, se puede realizar las estimaciones por máxima verosimilitud, similares a las anteriores, a procesos del tipo $\{Y_t\} = \{X_t - \mu\}$.

Por último, veremos el caso general, es decir, aquel que involucra a los modelos ARMA(p,q). Supóngase entonces un proceso ARMA(p,q) dado por la **ec. 3.5.1** pero con media μ y con proceso $\{\epsilon_t\} \sim WN_G(0, \sigma_\epsilon^2)$. El proceso a trabajar será $\{Y_t\} = \{X_t - \mu\}$.

Hemos visto que la función de verosimilitud para procesos autorregresivos se obtiene al condicionar sobre el conjunto $\{y_1, y_2, \dots, y_n\}$ (si trabajamos con procesos centrados) mientras que para procesos de promedios móviles se obtiene al condicionar sobre el conjunto $\{\epsilon_1, \epsilon_2, \dots, \epsilon_n\}$. Lo que haremos para los procesos ARMA(p,q) será condicionar sobre ambos conjuntos.

Sean $\mathbf{y}_0 = (y_0, y_{-1}, \dots, y_{-p+1})'$ y $\epsilon_0 = (\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-q+1})'$ valores iniciales dados. El conjunto $\{\epsilon_1, \dots, \epsilon_n\}$ se puede obtener a partir de $\{y_1, y_2, \dots, y_n\}$ iterando sobre:

$$\epsilon_t = Y_t - \phi_1 Y_{t-1} - \dots - \phi_q Y_{t-p} - \theta_1 \epsilon_{t-1} - \dots - \theta_q \epsilon_{t-q} \quad (3.5.113)$$

para $t = 1, 2, \dots, N$. La log-verosimilitud condicionada está dada entonces por:

$$\mathbb{L}(\Theta) = -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma_\epsilon^2) - \sum_{n=1}^N \frac{\epsilon_n^2}{2\sigma_\epsilon^2} \quad (3.5.114)$$

Una de las opciones por manejar será la de igualar los valores de $\mathbf{y}_0 = (y_0, y_{-1}, \dots, y_{-p+1})'$ y $\epsilon_0 = (\epsilon_0, \epsilon_{-1}, \dots, \epsilon_{-q+1})'$ con sus respectivos valores esperados, es decir, sea $y_s = \frac{1}{1-\phi_1-\dots-\phi_p}$ para $s = 0, -1, \dots, -p+1$ y $m = 0, -1, \dots, -q+1$ y $\epsilon_s = 0$ para $s = 0, -1, \dots, -p+1$ y $m = 0, -1, \dots, -q+1$ respectivamente y a partir de aquí iterar sobre la **ec. 3.5.113** para $t = 1, 2, \dots, N$. Otra opción a manejar para estimar la varianza $\hat{\sigma}_\epsilon^2$ es utilizar los valores y_1, \dots, y_p igual a sus valores actuales y establecer $\epsilon_p = \epsilon_{p-1} = \dots = \epsilon_{p-q+1} = 0$, por lo que la iteración de la **ec. 3.5.113** comenzará a partir de $t = p+1$. La log-verosimilitud condicionada estará dada por:

$$\mathbb{L}(\Theta) = -\frac{N-p}{2} \ln(2\pi) - \frac{N-p}{2} \ln(\sigma_\epsilon^2) - \sum_{n=p+1}^N \frac{\epsilon_n^2}{2\sigma_\epsilon^2} \quad (3.5.115)$$

o bien:

$$\mathbb{L}(\Theta) = -\frac{N-p}{2} \ln(2\pi) - \frac{N-p}{2} \ln(\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2} S(\Phi_p, \Theta_q) \quad (3.5.116)$$

donde $S(\Phi_p, \Theta_q)$ es la suma de los residuos al cuadrado a partir de cierta $n = p+1$ y hasta el total de la muestra. De nuevo, maximizar la **ec. 3.5.115** es equivalente a minimizar $S(\Phi_p, \Theta_q)$. Al diferenciar la **ec. 3.5.115** con respecto a σ_ϵ^2 e igualando a cero obtenemos:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{N-p} \sum_{n=p+1}^N \epsilon_n^2 \quad (3.5.117)$$

El minimizar $S(\Phi_p, \Theta_q)$ resulta ser una tarea sumamente complicada ya que la solución a dicha ecuación resulta ser no lineal en los parámetros Φ_p y Θ_q . Como ejemplo de dicha tarea, Guerrero (2003) considera un modelo MA(1) centrado dado por la **ec. 3.3.1**. Considere por un momento que podemos escribir dicho proceso como:

$$\begin{aligned}\epsilon_t &= X_t - \theta_1 \epsilon_{t-1} \\ &= X_t - \theta_1(X_{t-1} - \theta_1 \epsilon_{t-2}) \\ &= X_t - \theta_1 X_{t-1} + \theta_1^2 \epsilon_{t-2} \\ &\vdots \\ &= X_t - \theta_1 X_{t-1} + \theta_1^2 \epsilon_{t-2} + \dots + \theta_1^t \epsilon_0\end{aligned}$$

Recordando que $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$, nos resulta razonable suponer $\epsilon_0 = 0$ para así obtener:

$$S(\Theta_q = \theta_1) = \sum_{t=1}^N \epsilon_t^2$$

A continuación debemos minimizar $S(\theta_1)$ tomando:

$$\left. \frac{dS(\theta_1)}{d\theta_1} \right|_{\hat{\theta}_1} = 0$$

pero dicha ecuación no tiene solución analítica, por lo que se debe utilizar un procedimiento numérico para aproximarse al mínimo de $S(\theta_1)$. Guerrero (2003) propone evaluar la función $S(\theta_1)$ en distintos puntos del espacio de estados de θ_1 , obtener una gráfica como la que se muestra en la **Figura 3.5.6** y a partir de aquí encontrar el valor que minimiza $S(\theta_1)$. Dicho valor será la estimación de $\hat{\theta}_1$. El procedimiento es similar al momento de la estimación puntual de los parámetros Φ_p y Θ_q .

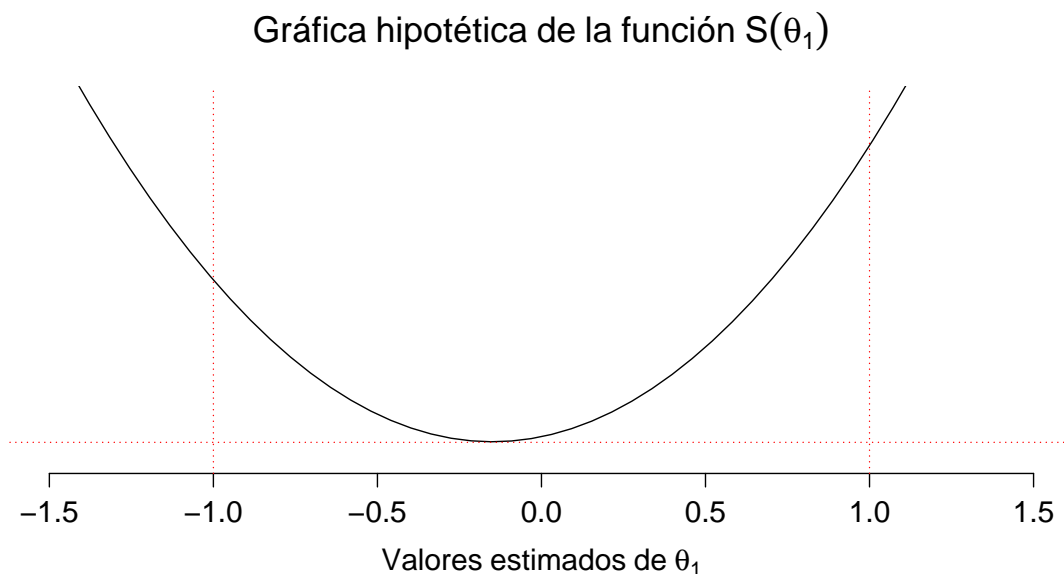


Figura 3.5.6: Figura que muestra la estimación de θ_1

3.5.5. Diagnósticos, verificación y corrección de supuestos

Hemos visto a lo largo del capítulo, los métodos de estimación que nos ayudan a obtener los valores de p y q de un modelo ARMA(p,q) para ajustar una serie de tiempo a partir de unos datos. Claro, para que dicho procedimiento tenga significado, el conjunto de datos observados debe ser una realización de un modelo ARMA(p,q) y, en particular, una realización de un proceso estacionario. De no ser un proceso estacionario, es decir, en caso de que el conjunto de datos presente tendencia, estacionalidad o bien variabilidad entorno a la media distinta de una constante, se prosigue a aplicar un conjunto de transformaciones que produzcan una nueva serie de tiempo que se acerque a las condiciones de estacionariedad.

Ante una fuerte dependencia en la variabilidad de los datos alrededor de la media podemos utilizar las transformaciones de Box-Cox vistas en la **Sección 3.2.2**. La tendencia y estacionalidad puede detectarse a través de la gráfica de los datos en cuestión. Además, el uso de la sucesión de autocorrelaciones muestrales (ACF), nos puede ayudar a determinar si la serie posee tendencia y estacionalidad ya que una lenta decaída y un comportamiento periódico en las autocorrelaciones, respectivamente, sería indicativo de la presencia de dichas componentes. Ante esta situación hemos visto, a lo largo de este escrito, que la descomposición clásica así como el diferenciar una serie, nos ayudan a eliminar la tendencia y estacionalidad.

Para poder aplicar un modelo ARMA(p,q) a unos datos, es menester que la serie en cuestión sea estacionaria en varianza y media. Cuando no lo es, se dice que presenta al menos una raíz unitaria. Cuando ocurre lo anterior, sabemos el conjunto de técnicas a utilizar para volver a una serie lo más estacionaria posible. Pues bien, una de las técnicas a utilizar será la de diferenciar la serie hasta volverla estacionaria. El número de diferencias por aplicar coincide con el número de raíces unitarias que presente la serie. El número de raíces unitarias se determina de manera formal con la prueba Dickey-Fuller aumentada.

Veamos un sencillo ejemplo. Tomemos un modelo AR(1) dado por la **ec. 3.4.1**. Recordemos que $|\phi_1| < 1$, así que cuando $\phi_1 = 1$ diremos que la serie en cuestión no es estacionaria. Si $|\phi_1| > 1$ tenemos el proceso AR(1) explosivo visto en la **Sección 3.5**. Pues bien, la hipótesis de estacionariedad de una serie puede ser evaluada analizando si el valor absoluto de ϕ_1 es menor estrictamente a la unidad.

La prueba Dickey-Fuller simple plantea contrastar $H_0 : \phi_1 = 1$ vs $H_1 : \phi_1 < 1$. Otra forma de realizar el planteamiento anterior es tomar la serie:

$$\begin{aligned}\nabla X_t &= \phi_1 X_{t-1} - X_{t-1} + \epsilon_t \\ &= (\phi_1 - 1)X_{t-1} + \epsilon_t \\ &= \pi X_{t-1} + \epsilon_t\end{aligned}$$

y establecer la prueba de hipótesis $H_0 : \pi = 0$ vs $H_1 : \pi < 0$. El estadístico bajo H_0 para la prueba Dickey-Fuller simple está dado por:

$$\hat{\tau} = \frac{\hat{\pi}}{\text{error estándar}(\hat{\pi})} \quad (3.5.118)$$

Aunque la prueba calcula un estadístico $\hat{\tau}$ sobre el valor estimado de $H_0 : \pi = 0$ y bajo la hipótesis nula de existencia de una raíz unitaria, este estadístico no tiene una distribución *t* de *Student* convencional.

Por ello, Dickey y Fuller simularon los valores críticos para una selección de distintos tamaños muestrales. Estas simulaciones serán las tablas que se utilizarán para determinar los valores críticos de la prueba Dickey-Fuller simple. El problema de la prueba Dickey-Fuller simple es que asume que el proceso estocástico subyacente a los datos sigue un modelo AR(1). En otras palabras, de manera estricta, la prueba de Dickey-Fuller sólo sirve para ver si el proceso es causal (su raíz es uno) en un proceso AR(1) o alternativamente para ver si conviene diferenciar una vez a dicho proceso. Por tanto, cuando el proceso siga otro esquema, la estimación de la regresión auxiliar de la prueba nos arrojará un incumplimiento de la condición de ruido blanco para el conjunto $\{\epsilon_t\}$, debido a una mala especificación. A partir de aquí es cuando se introduce la prueba Dickey-Fuller aumentada.

La prueba Dickey-Fuller aumentada es una generalización de la prueba Dickey-Fuller simple donde se supone al conjunto $\{\epsilon_t\}$ distinto a un ruido blanco. Para construir esta generalización se representó a una serie de tiempo como un proceso AR(p) dado por la **ec. 3.4.13**. Al tomar el operador diferencia ∇ obtenemos:

$$\nabla X_t = \Phi X_{t-1} + \sum_{i=1}^{p-1} c_i \nabla X_{t-i} + \epsilon_t \quad (3.5.119)$$

donde $\{c_i\}$ es un conjunto de constantes que dependen del proceso $\{X_t\}$ a través del conjunto $\{\phi_i\}$ para $i = 1, \dots, p$ y además:

$$\Phi = \left(\sum_{i=1}^p \phi_i - 1 \right) \quad (3.5.120)$$

A partir de la **ec. 3.5.119** se desprenden 2 modelos de series de tiempo, a saber aquel con deriva (en inglés *drift*) distinto de cero:

$$\nabla X_t = \alpha + \Phi X_{t-1} + \sum_{i=1}^{p-1} c_i \nabla X_{t-i} + \epsilon_t \quad (3.5.121)$$

y finalmente aquel con tendencia determinística dado por:

$$\nabla X_t = \alpha + \beta_t + \Phi X_{t-1} + \sum_{i=1}^{p-1} c_i \nabla X_{t-i} + \epsilon_t \quad (3.5.122)$$

La **ec. 3.5.119** es la que se recomienda para realizar la prueba Dickey-Fuller aumentada. El estadístico resulta ser el mismo que el de la **ec. 3.5.118** para probar $H_0 : \Phi = 0$ vs $H_1 : \Phi < 0$, es decir, para probar la existencia de una raíz unitaria en la serie. Si se rechaza H_0 se concluye que la serie no presenta raíz unitaria y por tanto es estacionaria.

Ejemplo 3.5.27 Consideremos a $\{Y_t\}$ un proceso AR(3) centrado, el cual podemos expresar como:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t$$

La prueba Dickey-Fuller aumentada se puede realizar de manera más sencilla escribiendo al modelo de la siguiente forma:

$$\begin{aligned} Y_t - Y_{t-1} &= (\phi_1 - 1)Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \epsilon_t \\ Y_t - Y_{t-1} &= (\phi_1 - 1)Y_{t-1} + (\phi_2 + \phi_3)Y_{t-2} + \phi_3(Y_{t-3} - Y_{t-2}) + \epsilon_t \\ Y_t - Y_{t-1} &= (\phi_1 + \phi_2 + \phi_3 - 1)Y_{t-1} + (\phi_2 + \phi_3)(Y_{t-2} - Y_{t-1}) + \phi_3(Y_{t-3} - Y_{t-2}) + \epsilon_t \\ \nabla Y_t &= \pi Y_{t-1} + c_1 \nabla Y_{t-1} + c_2 \nabla Y_{t-2} + \epsilon_t \end{aligned}$$

donde $\pi = \phi_1 + \phi_2 + \phi_3 - 1$, $c_1 = (\phi_2 + \phi_3)$, $c_2 = \phi_3$. A partir de aquí se construye la prueba de hipótesis $H_0 : \pi = 0$ vs $H_1 : \pi < 0$ y el estadístico dado por:

$$\hat{\tau} = \frac{\hat{\pi}}{\text{error estándar}(\hat{\pi})}$$

Una prueba alternativa fue introducida por Kwiatkowski, Phillips, Schmidt y Shin, conocida por tanto como la prueba KPSS, la cual asume como hipótesis nula la estacionariedad en una serie alrededor ya sea de la media o de una tendencia lineal y como hipótesis alternativa la no estacionariedad de una serie debido a la presencia de una raíz unitaria. En este sentido, es contraria a la prueba Dickey-Fuller aumentada en la cual la hipótesis nula asume la presencia de una raíz unitaria. Para la prueba KPSS una serie de tiempo se representa como la suma de 3 componentes, a saber, una tendencia determinística, una caminata aleatoria y un término estacionario de error. El modelo por tanto adquiere la forma:

$$\begin{aligned} Y_t &= \alpha t + \beta_t + \epsilon_t \\ \beta_t &= \beta_{t-1} + \nu_t \\ Y_t &= \alpha t + \beta_{t-1} + \nu_t + \epsilon_t \end{aligned} \tag{3.5.123}$$

donde claro Y_t denota a nuestra serie de tiempo, el parámetro t representa a la tendencia determinística, el conjunto $\{\beta_t\}$ representa a la caminata aleatoria mientras que el proceso $\{\epsilon_t\}$ representa al término aleatorio el cual hemos supuesto estacionario. Finalmente, el conjunto $\{\nu_t\}$ denota al término aleatorio de la segunda relación de la **ec. 3.5.122**, el cual es un conjunto de variables aleatorias independientes e idénticamente distribuidas con valor esperado igual a 0 y varianza σ_ν^2 . Como hipótesis supondremos β_0 una constante, la cual corresponde al intercepto.

La hipótesis nula que asume estacionariedad en la serie es equivalente al principio de que la varianza del proceso $\{\nu_t\}$, σ_ν^2 , es igual a 0, es decir, bajo la hipótesis nula se tiene que ν_t ya no es aleatorio sino un término constante (cero), en cuyo caso, el proceso no depende más que de la tendencia del proceso al tiempo anterior a través de β_{t-1} y de un proceso estacionario ϵ_t . En el caso de que $\alpha = 0$, la hipótesis nula significará que la serie Y_t es estacionaria alrededor de β_0 . Si $\alpha \neq 0$ entonces, la hipótesis nula significará que la serie Y_t es estacionaria alrededor de una tendencia lineal. Si la varianza del proceso $\{\nu_t\}$ es mayor a 0, entonces la serie Y_t resulta ser no estacionaria debido a la existencia de una raíz unitaria.

Al tomar el operador ∇ de la **ec. 3.5.123** obtenemos la siguiente relación:

$$\nabla Y_t = \alpha + \nu_t + \nabla \epsilon_t = \alpha + w_t \tag{3.5.124}$$

El conjunto w_t es generado por un proceso AR(1) $w_t = \gamma_t + \phi_1 \gamma_{t-1}$ si suponemos que los procesos $\{\epsilon_t\}$, $\{\nu_t\}$ son variables aleatorias independientes e idénticamente distribuidas. El modelo asociado a la prueba KPSS queda expresado entonces como:

$$\begin{aligned} Y_t &= \alpha + \lambda Y_{t-1} + w_t \\ w_t &= \gamma_t + \phi_1 \gamma_{t-1} \\ \lambda &= 1 \end{aligned} \quad (3.5.125)$$

La idea consiste en ver si el proceso $\{w_t\}$ es un proceso AR(1). El estadístico para probar la hipótesis nula de estacionariedad alrededor de una tendencia lineal contra la hipótesis alternativa de existencia de una raíz unitaria, se construye de la siguiente forma. Sea e_t para $t = 1, 2, \dots, N$, el residual correspondiente a la regresión de Y_t sobre una constante, es decir, $e_t = Y_t - \bar{Y}$. Sea $\hat{\sigma}_t^2$ el estimador para la varianza, el cual va de acuerdo con la **ec. 3.5.112**. Sea además

$$S_t = \sum_{i=1}^t e_i \quad (3.5.126)$$

con $t = 1, 2, \dots, N$. El estadístico para la prueba KPSS se define como:

$$LM = \frac{1}{\hat{\sigma}_t^2} \sum_{i=1}^N S_t^2 \quad (3.5.127)$$

Nos resta por estudiar la distribución asintótica del estadístico LM. La teoría concerniente a dicha distribución excede los objetivos de este escrito. Bastará decir que para estacionariedad alrededor de una tendencia lineal, el estadístico para la prueba KPSS converge de manera débil al límite de:

$$\hat{\nu}_\tau = \int_0^1 V_2(r)^2 dr \quad (3.5.128)$$

donde $V_2(r)$ es un puente Browniano de segundo orden, definido por:

$$V_2(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2) \int_0^1 W(s) ds \quad (3.5.129)$$

y $W(r)$ es un proceso de Wiener. Rechazamos H_0 si el estadístico LM es mayor que $\hat{\nu}_\tau$, al nivel de significancia dado. En el **capítulo 4** utilizaremos, tanto para la prueba Dickey-Fuller aumentada como para la prueba KPSS, el correspondiente p-value para así rechazar o no H_0 .

Después de obtener estacionariedad, nuestra tarea será la de ajustar un modelo ARMA(p,q) que represente de manera convincente al conjunto de datos observados. Estimaremos, a partir de la muestra, los valores p y q de dicho modelo. Se podría pensar por un momento que entre mayor sea la elección de los parámetros p y q del modelo mejor ajustados estarán los datos observados.

Pero la realidad es que debemos tener mucho cuidado de no ajustar de más un conjunto de datos ya que el cálculo de los coeficientes del modelo ARMA(p,q) puede ser extremadamente prohibitivo. Se han desarrollado criterios de información, en particular, el criterio de información de Akaike, que introducen un costo severo al momento de introducir parámetros de más. Se verá, más adelante, que debemos elegir aquel modelo ARMA(p,q) que minimice aquellos criterios de información.

Sabemos que nuestra primer herramienta para elegir los valores de p y q en un modelo ARMA(p,q) es el uso de la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF).

Asimismo, la bondad de ajuste de un modelo se juzga al comparar los valores observados y las predicciones obtenidas bajo el modelo. Así que puede ser conveniente analizar los residuos que miden la diferencia entre los valores reales y los valores predichos.

Cuando ajustamos un modelo ARMA(p,q) a unos datos, primero obtenemos los estimadores máximo verosímiles para $\hat{\Phi}$, $\hat{\Theta}$, $\hat{\sigma}^2$. Denótese como $\hat{X}_t(\hat{\Phi}_p, \hat{\Theta}_q)$ al valor estimado usando máxima verosimilitud y los valores X_1, X_2, \dots, X_{t-1} . Para ver que el estimado $\hat{X}_t(\hat{\Phi}_p, \hat{\Theta}_q)$ depende del conjunto $\{X_1, X_2, \dots, X_{t-1}\}$ consideraremos invertible al proceso $\{X_t\}$.

A partir de un proceso ARMA(p,q), dado por la **ec. 3.5.1** o bien, $\phi(B)X_t = \theta(B)\epsilon_t$, el valor de ϵ_t estimado es:

$$\begin{aligned} \hat{\epsilon}_t &= \hat{\phi}(B)\hat{\theta}^{-1}(B)X_t \\ &= \hat{\pi}(B)X_t \\ &= X_t + \hat{\pi}_1 X_{t-1} + \hat{\pi}_2 X_{t-2} + \dots \\ &= X_t - (-\hat{\pi}_1 X_{t-1} - \hat{\pi}_2 X_{t-2} - \dots) \\ &= X_t - \hat{X}_t \end{aligned} \tag{3.5.130}$$

La invertibilidad del proceso nos permite construir el nuevo polinomio $\hat{\pi}(B)$ ya que el polinomio de promedios móviles $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \neq 0$. Se concluye que \hat{X}_t , es una función de las observaciones hasta el tiempo $t-1$ que depende de los valores $\hat{\phi}_p$ y $\hat{\theta}_q$ calculados, y por eso lo denotamos como $\hat{X}_t(\hat{\Phi}_p, \hat{\Theta}_q)$. Los residuos a estudiar están definidos como:

$$\hat{W}_t = \frac{X_t - \hat{X}_t(\hat{\Phi}_p, \hat{\Theta}_q)}{\sqrt{r_{t-1}(\hat{\Phi}_p, \hat{\Theta}_q)}} \tag{3.5.131}$$

donde $r_i(\hat{\Phi}_p, \hat{\Theta}_q) = \mathbb{E}[(X_i - \hat{X}_i)^2]$ corresponde al error cuadrático medio para el tiempo i al aplicar el algoritmo de innovación. Si nuestra estimación máximo verosímil representa en realidad al conjunto de datos $\{X_1, X_2, \dots, X_n\}$ entonces, en teoría, $\{\hat{W}_t\} \sim WN(0, \sigma_W^2)$.

Es decir, $\{\hat{W}_t\}$ debería ser no correlacionada si $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$, independiente si la secuencia $\{\epsilon_t\} \sim IID(0, \sigma_\epsilon^2)$ y con distribución normal si $\{\epsilon_t\} \sim WN_G(0, \sigma_\epsilon^2)$. Los residuos re-escalados, $\{\hat{R}_t\}$, se obtienen al dividir $\{\hat{W}_t\}$ por el estimador $\hat{\sigma}_\epsilon$, la desviación estándar del ruido blanco, el cual está dado por $\hat{\sigma}_\epsilon = \sqrt{\hat{W}}$, donde W viene dado por:

$$W = \sum_{t=1}^n \frac{\hat{W}_t^2}{n} \tag{3.5.132}$$

Enunciamos a dichos residuos ya que el conjunto $\{\hat{R}_t\}$ debería comportarse como un ruido blanco con media 0 y varianza 1 si $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$ y como un ruido independiente con media 0 y varianza 1 si $\{\epsilon_t\} \sim IID(0, \sigma_\epsilon^2)$.

Supongamos esto último. La gráfica de $\{\hat{R}_t\}$ debería ser un indicativo de un ruido blanco con varianza 1. Desviaciones de la media de 0 son indicadores de fluctuaciones en la varianza de $\{\hat{R}_t\}$. Podemos usar también la sucesión de autocorrelaciones muestrales (ACF) de los residuos. Hemos visto en el **Ejemplo 3.5.15** que para n grande $\{\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, \dots, \hat{\rho}_\tau\} \sim WN_G(0, \frac{1}{n})$. Si graficamos la sucesión de autocorrelaciones muestrales (ACF), $\hat{\rho}_X(\tau)$, como función de τ , entonces, aproximadamente el 95% de las autocorrelaciones deberán estar dentro de las bandas $\pm \frac{1.96}{\sqrt{n}}$. Por ejemplo, si graficamos las primeras 40 autocorrelaciones entonces, para hablar de un ruido blanco Gaussiano, no más de 3 autocorrelaciones deben estar fuera de las bandas de confianza.

Para verificar que la media de los residuos $\{\hat{W}_t\}$ o $\{\hat{R}_t\}$ es cero podemos hacer uso de pruebas de hipótesis. Una prueba de hipótesis formal sería probar $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$, con estadístico:

$$T = \frac{\bar{X}\sqrt{n}}{S} \quad (3.5.133)$$

con \bar{X} la media de los residuos $\{\hat{W}_t\}$ o $\{\hat{R}_t\}$ y S su desviación estándar. Se rechaza H_0 si:

$$T < -t_{1-\frac{\alpha}{2}, n-1} \quad \text{o bien} \quad T > t_{1-\frac{\alpha}{2}, n-1}$$

con $t_{1-\frac{\alpha}{2}, n-1}$ el cuantil de una distribución t con $n - 1$ grados de libertad que acumula $\frac{\alpha}{2}$ de probabilidad.

Ahora bien, para demostrar normalidad en $\{\hat{W}_t\}$ o $\{\hat{R}_t\}$ podemos hacer uso de pruebas estadísticas como la prueba Kolmogorov-Smirnov, en particular las pruebas Shapiro-Wilk y Lilliefors para normalidad. También podemos hacer uso del **histograma** y de la **gráfica de probabilidad normal** de los residuos ya que en éstos se observa la forma de la distribución. En el caso de observaciones discrepantes, es decir, observaciones que se encuentran alejadas de manera significativa del resto, podemos usar las gráficas de residuos $\{\hat{W}_t\}$ o $\{\hat{R}_t\}$ contra el tiempo. Un residuo muy alejado del resto puede indicar la existencia de datos discrepantes o un evento extraño de poca probabilidad de ocurrencia. Como regla, consideraremos discrepantes a las observaciones cuyos residuos estén fuera del intervalo $(-3\sigma_\epsilon^2, 3\sigma_\epsilon^2)$.

Antes de tomar cualquier acción con dichas observaciones discrepantes debe investigarse la causa de que esas observaciones disten de las demás, ya que toda observación, incluyendo discrepantes, puede contener información muy valiosa para nuestros fines de estudio. La **Figura 3.5.7** muestra la existencia de valores extremos o datos discrepantes.

Las pruebas tipo $H_0 : \rho_\tau = 0$ vs $H_1 : \rho_\tau \neq 0$, para verificar si cierta autocorrelación independiente es cero, no son tan válidas para retrasos pequeños, en particular para $\tau \leq 3$, y por ello existen pruebas que conjuntamente sirven para ver si las autocorrelaciones hasta un retraso k son cero, y estas pueden servir para seguir verificando independencia.

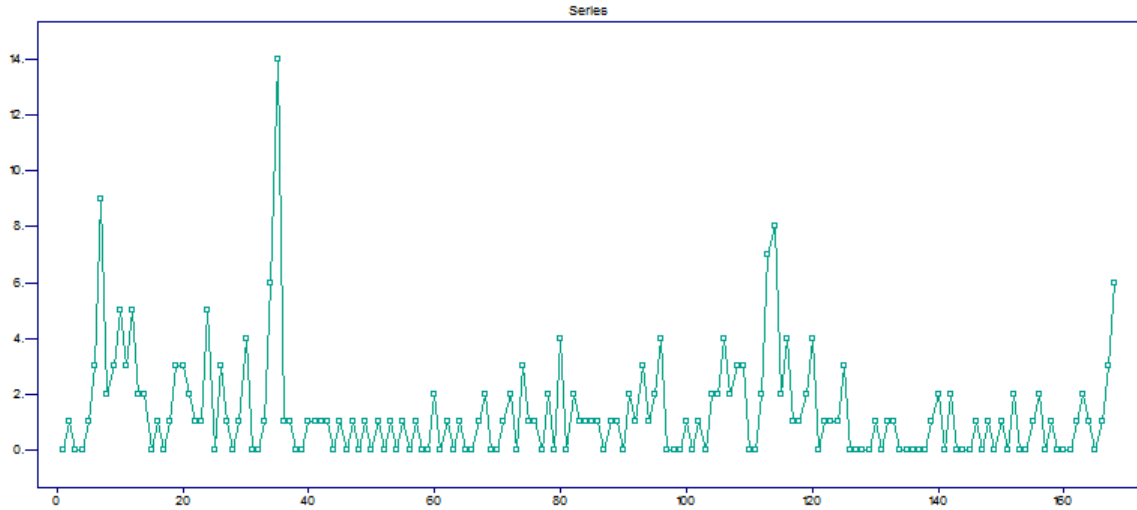


Figura 3.5.7: Figura que muestra la existencia de datos discrepantes

La primera de ellas es la llamada prueba de Ljung y Box (originalmente propuesta por Portmanteau) con prueba de hipótesis:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0 \quad \text{vs} \quad H_1 : \text{Al menos una } \rho_i \neq 0 \quad (3.5.134)$$

El estadístico asociado a la prueba de Ljung y Box es:

$$Q(k) = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_j^2}{n-j} \quad (3.5.135)$$

con n el tamaño de la muestra, $\hat{\rho}_j$ el estimador muestral de la autocorrelación con retraso j y k el número de retrasos bajo H_0 . El estadístico $Q(k)$ se distribuye asintóticamente como χ_{k-p-q}^2 así que se rechaza H_0 si:

$$Q(k) > \chi_{k-p-q, 1-\alpha}^2 \quad (3.5.136)$$

Otra modificación es la prueba McLeod-Li, quienes en lugar de usar $\hat{\rho}_j$ usan un estadístico que se basa en las estimaciones elevadas al cuadrado \hat{X}_t^2 , es decir, se usa:

$$\hat{\rho}_{xx}(j) = \frac{\sum_{t=1}^{n-j} (\hat{X}_t^2 - \bar{X}^2)(\hat{X}_{t+j}^2 - \bar{X}^2)}{\sum_{t=1}^n (\hat{X}_t^2 - \bar{X}^2)} \quad (3.5.137)$$

con \hat{X}_t los valores estimados y $\bar{X}^2 = \frac{1}{n} \sum_{t=1}^n \hat{X}_t^2$.

El estadístico obtenido será entonces:

$$Q_{xx}(k) = n(n+2) \sum_{j=1}^k \frac{\hat{\rho}_{xx}^2(j)}{n-j} \quad (3.5.138)$$

la distribución de $Q_{xx}(k)$ es χ_k^2 y se rechaza H_0 si:

$$Q_{xx}(k) > \chi_{k,1-\alpha}^2 \quad (3.5.139)$$

Existen pruebas para corroborar el supuesto de aleatoriedad en el conjunto $\{\hat{W}_t\}$, es decir, para demostrar que el proceso $\{\hat{W}_t\}$ es un conjunto de variables aleatorias independientes e idénticamente distribuidas. El demostrar aleatoriedad implicará que la serie de tiempo carece de tendencia. Nuestra primer prueba, utilizada para detectar la presencia de tendencia (creciente o decreciente) en unos datos, es la prueba de diferencia en signos. Sea $\{Y_t\}$ una serie de tiempo cualquiera. La idea de esta prueba es contar el número de valores de i para los cuales $Y_i > Y_{i-1}$, $i = 2, 3, \dots, n$, o bien, el número de veces en la cual la diferencia $Y_i - Y_{i-1} > 0$. Sea T dicho número. Es posible demostrar que:

$$T \sim N(\mu_T, \sigma_T^2) \quad (3.5.140)$$

con $\mu_T = \mathbb{E}(T) = \frac{1}{2}(n-1)$ y $\sigma_T^2 = \text{Var}(T) = \frac{1}{12}(n+1)$. Un valor positivo grande (o negativo) de $T - \mu_T$ es indicativo de una tendencia creciente (o decreciente) en nuestros datos. Rechazamos H_0 : No existe tendencia en los datos *vs* H_1 : Existe tendencia en los datos, si el estadístico:

$$S = \frac{1}{\sigma_T} |T - \mu_T| \quad (3.5.141)$$

es mayor a $Z_{1-\frac{\alpha}{2}}$, el cuantil de una distribución normal estándar que acumula una probabilidad de $1 - \frac{\alpha}{2}$.

Nuestra siguiente prueba es la llamada prueba de rangos. Dicha prueba es utilizada con gran frecuencia para detectar la presencia de tendencias lineales en unos datos. Sea $\{Y_t\}$ una serie de tiempo cualquiera. Sea P el número de parejas (i, j) tales que $Y_j > Y_i$, $j > i$ e $i = 1, 2, \dots, n-1$. Existe un total de $\frac{n}{2}(n-1)$ parejas tales que $j > i$ y para cada par de evento S , $\{Y_j > Y_i\}$, tenemos que su probabilidad de ocurrencia es $\frac{1}{2}$ si el conjunto $\{Y_t\}$ es una secuencia completamente aleatoria. La media y varianza de P serán $\mu_P = \mathbb{E}(P) = \frac{n}{4}(n-1)$ y $\sigma_P^2 = \text{Var}(P) = \frac{n}{8}(n-1)(2n+5)$ respectivamente, por lo que la distribución de P está dada por:

$$T \sim N(\mu_P, \sigma_P^2) \quad (3.5.142)$$

La hipótesis H_0 de existencia de tendencia lineal en el proceso $\{Y_t\}$ es rechazada, al nivel dado α , si el estadístico:

$$S = \frac{1}{\sigma_P} |P - \mu_P| \quad (3.5.143)$$

es mayor a $Z_{1-\frac{\alpha}{2}}$, el cuantil de una distribución normal estándar que acumula una probabilidad de $1 - \frac{\alpha}{2}$.

Ahora bien, consideremos al conjunto $\{\epsilon_t\} \sim WN_G(0, \sigma_\epsilon^2)$, proceso de ruido blanco generador de un proceso $\{Y_t\}$ ARMA(p,q). Sea $Y_{(1)} < Y_{(2)} < \dots < Y_{(n)}$ los estadísticos de orden de una muestra aleatoria $\{Y_i\}_{i=1}^n$ de una distribución normal con media μ y varianza σ_Y^2 . Si además, $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ son los estadísticos de orden de una muestra aleatoria $\{X_i\}_{i=1}^n$ de una distribución normal con media 0 y varianza 1, entonces la esperanza del j-ésimo estadístico de orden del proceso $\{Y_t\}$ está dado por:

$$\mathbb{E}[Y_{(j)}] = \mu + \sigma_Y \mathbb{E}[X_{(j)}] \quad (3.5.144)$$

con $j = 1, 2, \dots, n$. Si graficamos en un diagrama de dispersión al conjunto $\{(\mathbb{E}[X_{(j)}], Y_{(j)})\}_{j=1}^n$ esperaríamos que dicho conjunto sea aproximadamente lineal. Si la hipótesis de normalidad es adecuada entonces, el coeficiente de determinación, R^2 , debería ser cercano a 1. La hipótesis H_0 de normalidad en el proceso $\{\epsilon_t\}$ es rechazada entonces si el coeficiente R^2 es pequeño o cercano a 0.

Podemos aproximar $\mathbb{E}[X_{(j)}]$ con $Z^{-1}\left(\frac{i-0.5}{n}\right)$, con lo cual, el coeficiente de determinación, R^2 , se reduce a:

$$R^2 = \frac{\left[\sum_{i=1}^n (Y_{(i)} - \bar{Y}) Z^{-1}\left(\frac{i-0.5}{n}\right) \right]^2}{\sum_{i=1}^n (Y_{(i)} - \bar{Y})^2 \sum_{i=1}^n \left[Z^{-1}\left(\frac{i-0.5}{n}\right) \right]^2} \quad (3.5.145)$$

donde $\bar{Y} = \frac{1}{n}(Y_1 + Y_2 + \dots + Y_n)$.

3.5.6. Propiedades asintóticas de $\hat{\Phi}$ y $\hat{\Theta}$

Teorema 3.5.6 Sea $\{X_t\}$ un proceso ARMA(p,q) causal e invertible dado por la ec. 3.5.1. Sea $\Theta = (\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q)'$ el vector de parámetros del modelo ARMA(p,q) asociado.

Dejaremos atrás a los vectores $\Phi = (\phi_1, \phi_2, \dots, \phi_p)'$ y $\Theta = (\theta_1, \theta_2, \dots, \theta_q)'$. Entonces, los estimadores por máxima verosimilitud proveen estimadores óptimos de σ_ϵ^2 y Θ en el sentido de que σ_ϵ^2 es consistente. En particular para $n \rightarrow \infty$, la distribución asintótica de $\hat{\Theta}$ está dada por:

$$\sqrt{n}(\hat{\Theta} - \Theta) \sim N(0, \sigma_\epsilon^2 \Gamma_{p,q}^{-1}) \quad (3.5.146)$$

$\Gamma_{p,q}$ es una matriz de dimensión $(p+q) \times (p+q)$, dada por:

$$\Gamma_{p,q} = \begin{pmatrix} \Gamma_{\phi\phi} & \Gamma_{\phi\theta} \\ \Gamma_{\theta\phi} & \Gamma_{\theta\theta} \end{pmatrix} \quad (3.5.147)$$

La matriz de dimensión $p \times p$, $\Gamma_{\phi\phi}$, está dada por la **ec. 3.5.46**, es decir, el ij -ésimo elemento de la matriz $\Gamma_{\phi\phi}$ está dado por $\gamma_X(i-j)$ para un proceso autorregresivo de orden p , $\phi(B)X_t = \epsilon_t$, para $i, j = 1, 2, \dots, p$.

Similarmente, $\Gamma_{\theta\theta}$ representa la matriz de dimensión $q \times q$ con el ij -ésimo elemento, para $i, j = 1, 2, \dots, q$ igual a $\gamma_Y(i-j)$ de un proceso $AR(p)$, $\phi(B)Y_t = \epsilon_t$. La matriz $\Gamma_{\phi\theta}$, matriz de dimensión $p \times q$, representa la matriz de covarianzas de ambos procesos $\phi(B)X_t = \epsilon_t$ y $\phi(B)Y_t = \epsilon_t$. Finalmente $\Gamma_{\theta\phi} = \Gamma'_{\phi\theta}$ es una matriz de dimensión $q \times p$. La referencia al teorema anterior se puede encontrar en Shumway y Stoffer (2011).

Ejemplo 3.5.28 Veamos a continuación las distribuciones asintóticas para diversos procesos. Para un proceso $AR(1)$ tenemos $\gamma_X(0) = \frac{\sigma_\epsilon^2}{1-\phi_1^2}$. De acuerdo con el **Teorema 3.5.6** tenemos:

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \sim N(0, \sigma_\epsilon^2 \Gamma_{1,0}^{-1}) \quad \text{conforme } n \rightarrow \infty$$

con $\sigma_\epsilon^2 \Gamma_{1,0}^{-1} = 1 - \phi_1^2$ y entonces:

$$\sqrt{n}(\hat{\phi}_1 - \phi_1) \sim N(0, 1 - \phi_1^2) \quad \text{conforme } n \rightarrow \infty$$

o bien:

$$\hat{\phi}_1 \sim N\left(\phi_1, \frac{1 - \phi_1^2}{n}\right) \quad \text{conforme } n \rightarrow \infty$$

Así que $\text{Var}(\hat{\phi}_1) = \frac{1 - \phi_1^2}{n}$ y entonces podemos estimar la varianza como:

$$\widehat{\text{Var}}(\hat{\phi}_1) = \frac{1 - \hat{\phi}_1^2}{n}$$

y por tanto el error estándar y podemos dar intervalos de confianza para $\hat{\phi}_1$. Similarmente, para un $AR(2)$ se puede verificar:

$$\gamma_X(0) = \left(\frac{1 - \phi_2}{1 + \phi_2}\right) \frac{\sigma_\epsilon^2}{(1 - \phi_2)^2 - \phi_1^2}$$

y por tanto:

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 1 - \phi_2^2 & -\phi_1(1 + \phi_2) \\ -\phi_1(1 + \phi_2) & 1 - \phi_2^2 \end{pmatrix}\right]$$

Ahora bien, para calcular $\Gamma_{p,q}$ en un proceso $ARMA(1,1)$ debemos calcular primero $\gamma_{XY}(0)$, con procesos $\phi(B)Y_t = \epsilon_t$ y $\phi(B)X_t = \epsilon_t$, dado por:

$$\begin{aligned} \gamma_{XY}(0) &= Cov(X_t, Y_t) \\ &= Cov(\phi_1 X_{t-1} + \epsilon_t, -\theta_1 Y_{t-1} + \epsilon_t) \\ &= -\phi_1 \theta_1 \gamma_{XY}(0) + \sigma_\epsilon^2 \end{aligned}$$

Resolviendo para $\gamma_{XY}(0)$ tenemos $\gamma_{XY}(0) = \frac{\sigma_\epsilon^2}{1+\phi_1\theta_1}$ y por tanto:

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\theta}_1 \end{pmatrix} \sim N \left[\begin{pmatrix} \phi_1 \\ \theta_1 \end{pmatrix}, \frac{1 + \theta_1 \phi_1}{n(\phi_1 + \theta_1)^2} \begin{pmatrix} (1 - \phi_1^2)(1 + \theta_1 \phi_1) & -(1 - \theta_1^2)(1 - \phi_1^2) \\ -(1 - \theta_1^2)(1 - \phi_1^2) & (1 - \theta_1^2)(1 + \theta_1 \phi_1) \end{pmatrix} \right]$$

Pasemos a los procesos de promedios móviles. Para un $MA(1)$, $X_t = \epsilon_t + \theta_1 \epsilon_{t-1}$, se le asocia un proceso $AR(1)$, por lo que $\gamma_X(0) = \frac{\sigma_\epsilon^2}{1-\theta_1^2}$ y $\sigma_\epsilon^2 \Gamma_{1,0}^{-1} = 1 - \theta_1^2$. Se desprende que:

$$\hat{\theta}_1 \sim N \left(\theta_1, \frac{1 - \theta_1^2}{n} \right) \quad \text{conforme } n \rightarrow \infty$$

Similarmente, para un $MA(2)$:

$$\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \sim N_2 \left[\begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}, \frac{1}{n} \begin{pmatrix} 1 - \theta_2^2 & \theta_1(1 + \theta_2) \\ \theta_1(1 + \theta_2) & 1 - \theta_2^2 \end{pmatrix} \right]$$

Y así sucesivamente. El software estadístico ya nos da automáticamente los errores estándar de los estimadores e intervalos de confianza de cada uno.

Ejemplo 3.5.29 Suponga que se identificó un modelo $Z_t - \phi_1 Z_{t-1} = \epsilon_t + \theta_1 \epsilon_{t-1}$ para la serie Z_t , que es una serie transformada obtenida a partir de una serie X_t . La serie Z_t consta de 150 observaciones. La etapa de estimación dio los siguientes resultados: $\hat{\phi}_1 = -0.613$, $\hat{\theta}_1 = -0.298$, residuos promedio igual a 0.0035 y desviación estándar igual a 0.1651.

La **Tabla 3.5.5** muestra la sucesión de autocorrelaciones (ACF) de los residuos para este ejemplo. Verifique que la secuencia $\{\epsilon_t\}$ tiene media cero y es no correlacionada. Obtenga mediante la aproximación asintótica vista en este capítulo estimaciones para $Var(\hat{\theta})$ y $Var(\hat{\phi})$.

Con estos valores construya intervalos de confianza al 95% para $\hat{\theta}$ y $\hat{\phi}$. ¿Está el cero contenido en estos intervalos? De ser así, esto sería indicativo de que los parámetros no son significativos y de que el modelo no está bien especificado.

Para demostrar que la serie $\{\epsilon_t\}$ no está correlacionada utilice la prueba de Ljung y Box usando $\alpha = 0.05$ y $\alpha = 0.1$.

lag k	1	2	3	4	5	6	7	8	9	10	11	12
$\hat{\rho}_k^W$	-0.09	.25	.01	.07	-0.05	.11	.07	-0.02	.04	.05	-0.08	-0.12

Tabla 3.5.5: Tabla que muestra la sucesión de autocorrelaciones de los residuos de $\{Z_t\}$

El modelo anterior corresponde a un proceso ARMA(1,1). Observemos que la media de los residuos es 0.0035 y su desviación estándar es $S=0.1651$. $H_0 : \mu = 0$ vs $H_1 : \mu \neq 0$, es la prueba de 2 colas a utilizar, la cual supone que los residuos se distribuyen $N(\mu, \sigma^2)$ con σ desconocida. El estadístico de prueba está dado por:

$$T = \frac{\sqrt{n}\bar{X}}{S} = \frac{\sqrt{150}[0.0035]}{0.1651} = 0.2596370109$$

Se rechaza H_0 si:

$$T < -t_{(1-\frac{\alpha}{2}, 149)} = -1.97601 \quad \text{o bien} \quad T > t_{(1-\frac{\alpha}{2}, 149)} = 1.97601$$

Ya que no se cumple la regla de decisión, no se rechaza H_0 y la media de los residuos es aproximadamente cero, así que $\{\epsilon_t\}$ tiene media cero.

Para demostrar que la secuencia $\{\epsilon_t\}$ son no correlacionados debemos probar la hipótesis $H_0 : \rho_1 = \rho_2 = \dots = \rho_{12} = 0$ vs $H_1 : \text{Al menos una } \rho_i \neq 0 \text{ para los residuos}$. Usando las autocorrelaciones muestrales para los residuos $\hat{\rho}_k^W$ de la **Tabla 3.5.5** obtenemos:

$$Q_W(12) = n(n+2) \sum_{j=1}^{12} \frac{\hat{\rho}_j^2(w)}{n-j} = (150)(152) \left[\frac{(-0.09)^2}{149} + \frac{(0.25)^2}{148} + \dots + \frac{(-0.12)^2}{138} \right] = 18.1708$$

Se rechaza H_0 si $Q_W(12) > \chi_{10, 1-\alpha}^2$. Para $\alpha = 0.05$ y $\alpha = 0.1$ respectivamente tenemos:

$$\begin{aligned} Q_{W,0.95}(12) &= 18.1708 \not> 18.307 && \therefore \text{No rechazo } H_0 \\ Q_{W,0.90}(12) &= 18.1708 > 15.987 && \therefore \text{Rechazo } H_0 \end{aligned}$$

Para $\alpha = 0.05$ no se rechaza H_0 , conjuntamente todas las autocorrelaciones son cero y entonces se puede pensar que estamos trabajando con un ruido blanco; sin embargo, para $\alpha = 0.1$ sí que se rechaza H_0 y la secuencia $\{\epsilon_t\}$ no se comporta como ruido blanco. Ahora bien, para un modelo ARMA(1,1) tenemos:

$$\begin{aligned} \text{Var}(\hat{\phi}) &= \frac{1 + \phi\theta}{n(\phi + \theta)^2} (1 - \phi^2)(1 + \phi\theta) \\ \text{Var}(\hat{\theta}) &= \frac{1 + \phi\theta}{n(\phi + \theta)^2} (1 - \theta^2)(1 + \phi\theta) \end{aligned}$$

y por tanto:

$$\begin{aligned} \widehat{\text{Var}}(\hat{\phi}) &= \frac{(1 + \hat{\phi}\hat{\theta})^2}{n(\hat{\phi} + \hat{\theta})^2} (1 - \hat{\phi}^2) \\ \widehat{\text{Var}}(\hat{\theta}) &= \frac{(1 + \hat{\phi}\hat{\theta})^2}{n(\hat{\phi} + \hat{\theta})^2} (1 - \hat{\theta}^2) \end{aligned}$$

Pero $\hat{\phi} = \hat{\phi}_1 = -0.613$ y $\hat{\theta} = \hat{\theta}_1 = -0.298$ de donde:

$$\begin{aligned}\widehat{Var}(\hat{\phi}) &= 0.007013703754 \\ \widehat{Var}(\hat{\theta}) &= 0.01023797089\end{aligned}$$

Como

$$\begin{pmatrix} \hat{\phi} \\ \hat{\theta} \end{pmatrix} \sim N \left[\begin{pmatrix} \phi \\ \theta \end{pmatrix}, \begin{pmatrix} Var(\phi) & Cov(\phi, \theta) \\ Cov(\phi, \theta) & Var(\theta) \end{pmatrix} \right]$$

y $\hat{\phi} \sim N(\phi, Var(\hat{\phi}))$, $\hat{\theta} \sim N(\theta, Var(\hat{\theta}))$ entonces:

$$\frac{\hat{\phi} - \phi}{\sqrt{Var(\hat{\phi})}} \sim N(0, 1) \quad y \quad \frac{\hat{\theta} - \theta}{\sqrt{Var(\hat{\theta})}} \sim N(0, 1)$$

Si pensamos los intervalos de confianza como los de una distribución normal con varianza conocida, es decir, si se conoce $Var(\theta)$ o $Var(\phi)$, tenemos:

$$\mathbb{P} \left(\hat{\phi} - Z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})} \leq \phi \leq \hat{\phi} + Z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})} \right) = 1 - \alpha$$

\therefore El intervalo de confianza está dado por $\left(\hat{\phi} - Z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})}, \hat{\phi} + Z_{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})} \right)$.

Si $Var(\hat{\phi})$ es desconocida, que es lo que sucede en realidad, entonces tenemos un intervalo de confianza para la media de una distribución normal con varianza desconocida, es decir:

$$\left(\hat{\phi} - t_{n-1}^{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})}, \hat{\phi} + t_{n-1}^{1-\frac{\alpha}{2}} \sqrt{Var(\hat{\phi})} \right)$$

Si n crece $t_{n-1}^{1-\frac{\alpha}{2}} = Z_{1-\frac{\alpha}{2}}$. Usemos $Z_{1-\frac{\alpha}{2}} = 1.96$, considerando $\alpha = 0.05$.

\therefore El intervalo de confianza para θ es $(-0.496318, -0.0996815)$.

\therefore El intervalo de confianza para ϕ es $(-0.7771058021, -0.4488541979)$.

\therefore El cero no está contenido en ningún intervalo de confianza, así que los parámetros son significativos y por tanto, el modelo está bien especificado.

3.5.7. Criterios de información

Hemos comentado anteriormente los riesgos que surgen al elegir los parámetros p y q en un modelo ARMA(p, q) de manera arbitraria, a saber uno de ellos, la sobre-parametrización. Ante tales circunstancias se han creado criterios de información que introducen una “penalización”, la cual ayudar a eliminar todo incentivo a usar parámetros de más.

En este escrito se estudiarán 4 criterios de información, a saber, el error de predicción final (FPE), el criterio de información de Akaike sin corregir, el criterio de información de Akaike corregido y el criterio de información Bayesiana (BIC).

Para modelos autorregresivos AR(p) se puede definir el error de predicción final (FPE) como un estimador del error cuadrático medio que minimiza la predicción a un paso. Sea $\{X_1, \dots, X_n\}$ una realización de un proceso AR(p) con coeficientes ϕ_1, \dots, ϕ_p y sea $\{Y_1, \dots, Y_n\}$ una realización independiente del mismo proceso. Sean $\hat{\phi}_1, \dots, \hat{\phi}_p$ los estimadores máximo verosímiles de los coeficientes basados en $\{X_1, \dots, X_n\}$, el predictor a un paso para Y_{n+1} es:

$$\hat{Y}_{n+1} = \hat{\phi}_1 Y_n + \dots + \hat{\phi}_p Y_{n+1-p} \quad (3.5.148)$$

Brockwell y Davis (1991) deriva el error cuadrático medio para la predicción a un paso dada por:

$$\mathbb{E}[(Y_{n+1} - \hat{Y}_{n+1})^2] \simeq \sigma_\epsilon^2 \left(\frac{n+p}{n} \right) \quad (3.5.149)$$

y se puede estimar σ_ϵ^2 como $\frac{n\hat{\sigma}_\epsilon^2}{n-p}$ así que se tiene a $\hat{\sigma}_\epsilon^2 \left(\frac{n+p}{n-p} \right)$ como un estimador para $\mathbb{E}[(Y_{n+1} - \hat{Y}_{n+1})^2]$, es decir:

$$\mathbb{E}[\widehat{(Y_{n+1} - \hat{Y}_{n+1})^2}] = \hat{\sigma}_\epsilon^2 \left(\frac{n+p}{n-p} \right) = FPE \quad (3.5.150)$$

Así que se selecciona el modelo AR(p) cuya “p” minimiza FPE porque de esta forma el error cuadrático medio de la predicción disminuye.

Un criterio mucho más general al error de predicción final (FPE) es el criterio de información de Akaike sin corregir. Dicho criterio fue construido para hacer un estimador insesgado del Índice de Kullback-Lieber. Definimos al Índice de Kullback-Lieber para $f(\cdot, \psi)$, distribución del n-vector aleatorio \mathbf{X} que usa como parámetros a ψ , respecto a $f(\cdot, \theta)$, distribución con los parámetros reales θ , como:

$$\begin{aligned} \Delta(\psi|\theta) &= \mathbb{E}_\theta(-2\ln f(\mathbf{X}; \psi)) \\ &= \int_{\mathbb{R}^n} -2\ln f(\mathbf{x}; \psi) f_{\mathbf{X}}(\mathbf{x}; \theta) d\mathbf{x} \end{aligned}$$

Así, $\Delta(\hat{\theta}|\theta)$ denota el índice entre el vector de parámetros estimados y los reales. El mejor modelo que podemos obtener es aquel que minimice la discrepancia Kullback-Lieber, dada por:

$$d(\psi|\theta) = \Delta(\psi|\theta) - \Delta(\theta|\theta) \quad (3.5.151)$$

La discrepancia Kullback-Lieber nos ayuda a medir la “distancia” entre dos distribuciones de probabilidad, una de las cuales actúa como referencia, definidas sobre la misma variable aleatoria.

En este caso estamos midiendo la distancia existente entre $f(\cdot, \psi)$ y $f(\cdot, \theta)$. De hecho, utilizando la desigualdad de Jensen podemos demostrar lo siguiente:

$$\begin{aligned} d(\psi|\theta) &= \int_{\mathbb{R}^n} -2\ln \frac{f(\underline{x}; \psi)}{f_X(\underline{x}; \theta)} f_X(\underline{x}; \theta) d\underline{x} \\ &\geq -2\ln \left[\int_{\mathbb{R}^n} \frac{f(\underline{x}; \psi)}{f_X(\underline{x}; \theta)} f_X(\underline{x}; \theta) d\underline{x} \right] \\ &= -2\ln \left[\int_{\mathbb{R}^n} f(\underline{x}; \psi) d\underline{x} \right] \\ &= 0 \end{aligned}$$

es decir, con la desigualdad de Jensen queremos demostrar que lo que buscamos es la menor distancia posible entre la distribución estimada con los parámetros ψ y la distribución real del modelo ARMA(p,q) asociado.

Ahora bien, Brockwell y Davis (1991), demuestra que si tenemos $\{X_1, \dots, X_n\}$ observaciones de un proceso ARMA Gaussiano (el ruido blanco es Gaussiano) con vector de parámetros $(\Phi_p, \Theta_q, \sigma_\epsilon^2)$ donde $(\hat{\Phi}_p, \hat{\Theta}_q, \hat{\sigma}_\epsilon^2)$ son sus correspondientes estimadores máximo verosímiles basados en $\{X_1, \dots, X_n\}$ y si tomamos $\{Y_1, \dots, Y_n\}$ una realización independiente del proceso verdadero con parámetros $(\Phi_p, \Theta_q, \sigma_\epsilon^2)$ entonces un estimador insesgado de:

$$\Delta((\hat{\Phi}_p, \hat{\Theta}_q, \hat{\sigma}_\epsilon^2) | (\Phi_p, \Theta_q, \sigma_\epsilon^2)) = \mathbb{E}_{(\Phi_p, \Theta_q, \sigma_\epsilon^2)}(-2\ln L_y(\Phi_p, \Theta_q, \sigma_\epsilon^2))$$

es:

$$AICC = -2\ln L_X \left(\Phi_p, \Theta_q, \frac{S(\Phi_p, \Theta_q)}{n} \right) + \frac{2(p+q+1)n}{n-p-q-2} \quad (3.5.152)$$

A este término se le conoce como el criterio de información de Akaike corregido, donde la verosimilitud y sus componentes $S(\Phi_p, \Theta_q)$ ya la habíamos definido cuando vimos máxima verosimilitud en la **Sección 3.5.4**.

El criterio de información de Akaike sin corregir está dado por:

$$AIC = -2\ln L_X \left(\Phi_p, \Theta_q, \frac{S(\Phi_p, \Theta_q)}{n} \right) + 2(p+q+1) \quad (3.5.153)$$

El criterio de información de Akaike sin corregir se basa en la entropía de la información, es decir, ofrece una estimación relativa de la información perdida cuando se utiliza un modelo ARMA(p,q) determinado para representar el proceso que realmente genera unos datos.

De hecho, el criterio de información de Akaike corregido es una corrección del criterio de información de Akaike sin corregir para muestras finita, y por tanto penaliza de manera más excesiva la inclusión de parámetros adicionales.

Seleccionaríamos el modelo ARMA(p,q) con AIC o AICC lo más bajo posible porque ese tendría menor índice de Kullback-Liebert, es decir, seleccionaríamos aquel modelo que posea la menor distancia, entre los modelos candidatos y el verdadero.

Intuitivamente, la primer parte del AIC y AICC corresponden a la verosimilitud así que al elegir el menor AIC o AICC elegimos el modelo con mayor verosimilitud; por otra parte, los siguientes términos $\frac{2(p+q+1)n}{n-p-q-2}$ o $2(p+q+1)$ son “penalizaciones” por parametrizar de más el modelo, así que al elegir los valores mínimos elegimos el modelo con menos parámetros.

En conclusión, si se tiene un grupo de modelos ARMA(p,q) que ajustan bien a una serie de tiempo, debemos entonces elegir aquel cuyo AIC o AICC sea mínimo. El AIC ó AICC pueden ser negativos, en este caso se elige el modelo con $|AIC|$ o $|AICC|$ mínimo.

Nuestro último criterio es el llamado criterio de información Bayesiana (BIC), el cual viene dado por:

$$BIC = (n - p - q) \ln \left(\frac{n \hat{\sigma}_\epsilon^2}{n - p - q} \right) + n(1 + \ln \sqrt{2\pi}) + (p + q) \ln \left[\frac{1}{p + q} \sum_{t=1}^n X_t^2 - n \hat{\sigma}_\epsilon^2 \right]$$

donde $\hat{\sigma}_\epsilon^2$ es el estimador máximo verosímil para la varianza del proceso $\{\epsilon_t\}$.

La mayoría de los trabajos que analizan y comparan distintos criterios de información se realizan con el objetivo de encontrar un criterio dominante en la búsqueda del modelo verdadero en el contexto de series de tiempo univariadas.

Una propiedad interesante del criterio de información Bayesiana (BIC), y que de hecho no la comparten los criterios de información de Akaike (sin corregir y corregido), es que, el criterio de información Bayesiana (BIC) es consistente, es decir, si tenemos una muestra $\{X_1, X_2, \dots, X_n\}$ y dicha muestra proviene de observaciones de un proceso ARMA(p,q), y los estimadores \hat{p} y \hat{q} son los estimadores de p y q encontrados al minimizar el criterio de información Bayesiana (BIC), entonces $\hat{p} \rightarrow p$ y $\hat{q} \rightarrow q$ con probabilidad 1. Debemos recalcar que para que suceda lo anterior, con probabilidad 1, el tamaño de nuestra muestra debe tender a infinito.

Por otro lado, la selección de parámetros, ya sea minimizando el criterio de información de Akaike sin corregir, el criterio de información de Akaike corregido o bien el error de predicción final (FPE), es asintóticamente eficiente para procesos autorregresivos, mientras que la minimización utilizando el criterio de información Bayesiana (BIC) no lo es. La eficiencia nos será de utilidad al querer minimizar el error correspondiente a nuestras predicciones.

3.6. Análisis de procesos ARMA(p,q) en ITSM

En este apartado realizaremos un análisis completo y encontraremos el modelo ARMA(p,q) óptimo para 4 series de tiempo. La primera de ellas es la serie consistente en el número de muertes accidentales en E.U.A. desde el año 1973 hasta el año 1978. La siguiente serie a analizar es sobre el número de acciones de atención otorgadas por Condusef desde el año 2007 y hasta el año 2012. Para el número de acciones de atención otorgadas por Condusef, realizaremos un pronóstico para el año 2013 y lo compararemos con los valores reales obtenidos ya transcurrido dicho año. Esta serie de datos no ha sido analizada antes. Las cifras correspondientes al número de acciones de atención han sido tomadas de los anuarios estadísticos proporcionados por Condusef. La tercera serie consiste en las ventas mensuales (kilolitros) de vino tinto por parte de viñedos australianos desde enero de 1980 hasta octubre de 1991. Por último, analizaremos la serie consistente en el nivel (pies) del lago Huron desde el año 1875 y hasta el año 1972. El análisis de dichas 4 series se realizará por completo en el paquete estadístico ITSM en su versión de estudiantes correspondiente a Brockwell y Davis (1991), dado una guía completa, de inicio a fin, de la serie de pasos a realizar.

Comencemos con el número de muertes accidentales en E.U.A. Abrimos el paquete estadístico ITSM y seleccionamos las opciones “File >Project >Open >Univariate” y elegimos el archivo “Deaths.tsm”. La **Figura 3.6.1** muestra dicha serie.

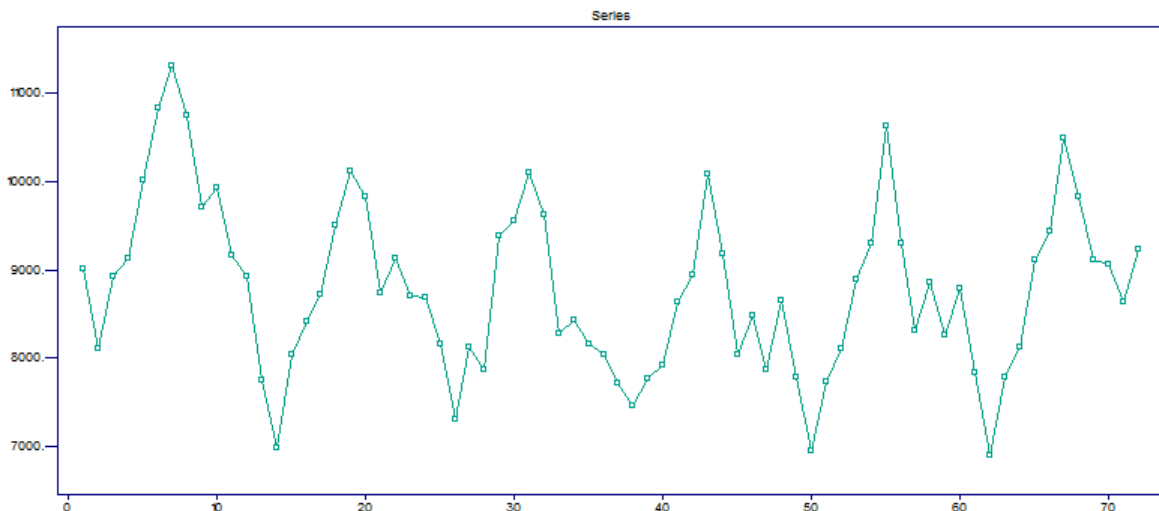


Figura 3.6.1: Figura que muestra el número de muertes accidentales en E.U.A.

De la **Figura 3.6.1** podemos observar una tendencia lineal descendente y un patrón estacional con un pico en julio y un declive en febrero de cada año. Ahora bien, para eliminar la tendencia aplicaremos el operador de diferencias, ∇ con un “lag” de 1 y para eliminar la estacionalidad, la cual es anual, se aplicará el operador ∇ con un “lag” de 12. Recordemos que después de eliminar la tendencia, y la estacionalidad debemos restar la media a la nueva serie para así centrar el proceso.

Comenzamos con la tendencia. Elegimos “Transform >Difference >Enter lag = 1”, obteniendo la **Figura 3.6.2**. Como siguiente paso, eliminamos la estacionalidad con el código “Transform >Difference >Enter lag = 12”, obteniendo la **Figura 3.6.3**. Nos resta centrar la nueva serie. El proceso centrado se obtiene eligiendo “Transform >Difference >Substract Mean”, obteniendo así, la **Figura 3.6.4**.

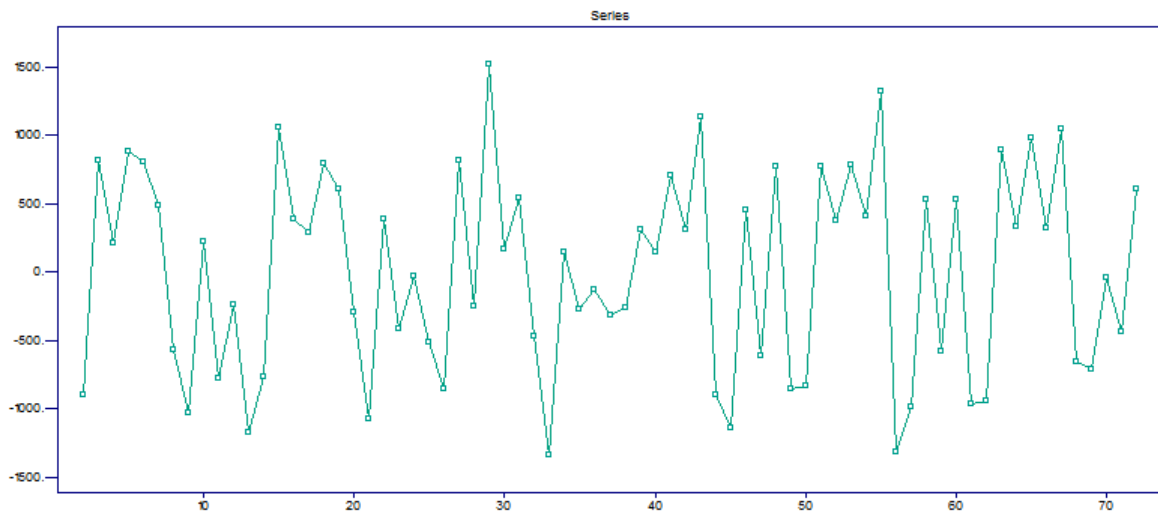


Figura 3.6.2: Figura que muestra el número de muertes accidentales sin tendencia

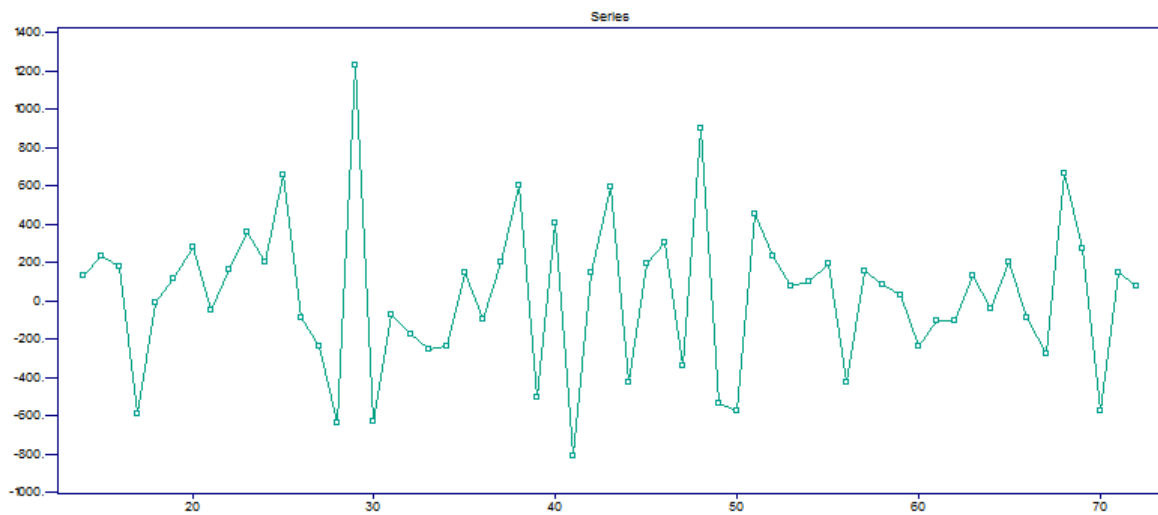


Figura 3.6.3: Figura que muestra el número de muertes accidentales sin estacionalidad

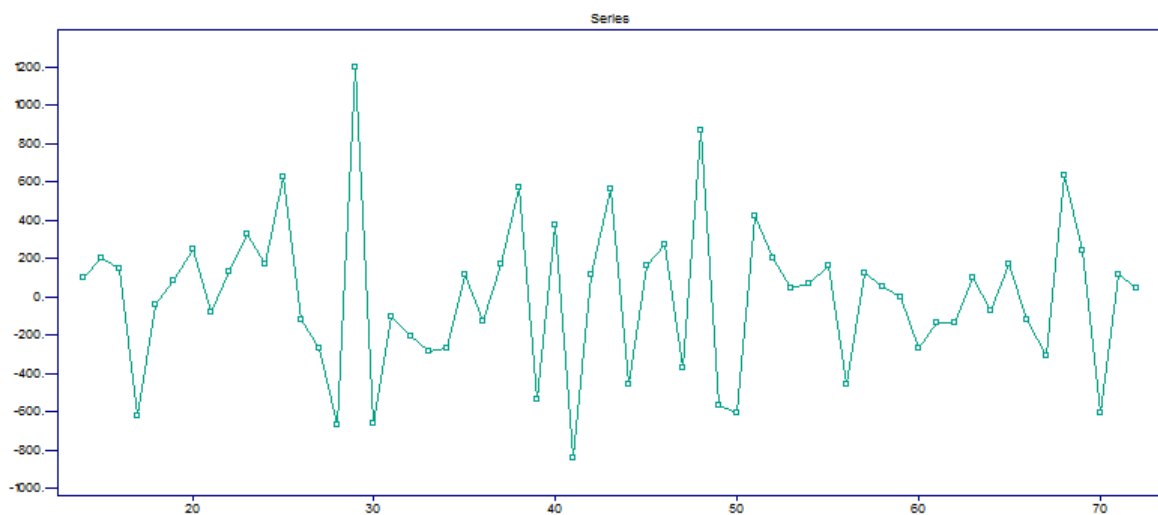


Figura 3.6.4: Figura que muestra el número de muertes accidentales restando la media

Nuestra siguiente tarea será la de obtener la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) para así poder elegir el modelo ARMA(p,q) óptimo. Elegimos “Statistics >ACF/PACF >Sample”, obteniendo la **Figura 3.6.5** y la **Figura 3.6.6** mostradas a continuación

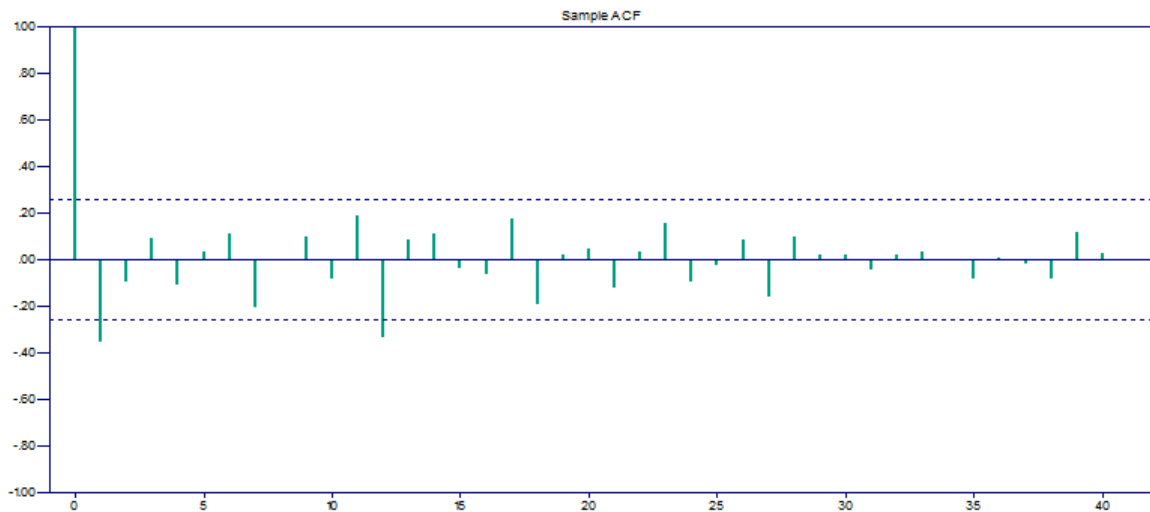


Figura 3.6.5: Figura ACF del número de muertes accidentales en E.U.A.

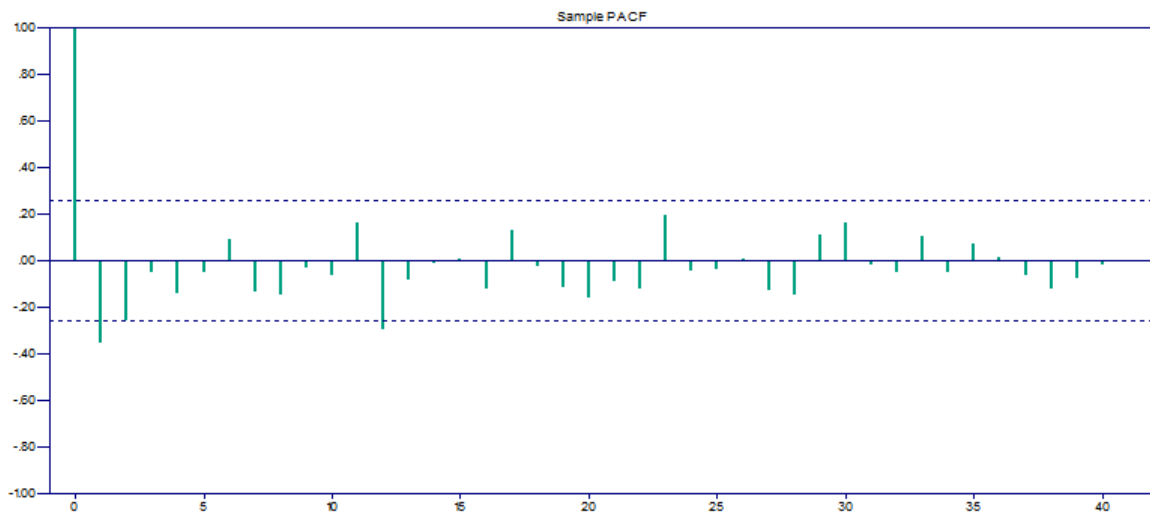


Figura 3.6.6: Figura PACF del número de muertes accidentales en E.U.A.

De la sucesión de autocorrelaciones muestrales (ACF) se salen la autocorrelaciones correspondientes al retraso o “lag” 1 y 12. De la sucesión de autocorrelaciones muestrales parciales (PACF) se salen las autocorrelaciones correspondientes al “lag” 1 y 12. La teoría nos dice que debemos escoger aquellos retrasos p y q, a partir de los cuales las autocorrelaciones de la sucesión de autocorrelaciones muestrales (ACF) y de la sucesión de autocorrelaciones muestrales parciales (PACF) no se salgan de las bandas de confianza. Por lo tanto, deberíamos escoger un modelo ARMA(12,12). Nosotros no haremos esto ya que al tomar dicho modelo el proceso carece de causalidad (de hecho, ITSM nos avisa que el proceso no es causal), además que para procesos con valores significativamente grandes para p y q, la estimación de los parámetros se vuelve un tanto problemática, se estaría violando el principio de parsimonia y los resultados obtenidos no distan mucho de aquellos con procesos con retrasos pequeños.

Algunas opciones para esta serie son las de un proceso ARMA(1,1) y un proceso ARMA(2,1). Veremos a continuación cuál de estas dos opciones es el proceso óptimo. Queremos hacer hincapié que nuestro objetivo es encontrar aquel modelo que represente de manera fehaciente la realidad sin importar qué tan complicado sea dicho modelo.

Supongamos que sólo deseamos el mejor modelo AR(p). Entonces, ITSM lo puede elegir por nosotros. Elegimos “`Model >Estimation >Preliminary >Find AR model with min AICC`”, e ITSM nos muestra que el mejor modelo es el AR(2) con coeficientes $\phi_1 = -0.447657$, $\phi_2 = -0.258017$ y criterio $AICC = 866.210$. Ahora bien, para estimar por máxima verosimilitud usando como valores iniciales los obtenidos vía Yule-Walker debemos elegir “`Model >Estimation >Max Likelihood`”, obteniendo los coeficientes $\phi_1 = -0.443450$, $\phi_2 = -0.253866$ y criterios $AICC = 866.208$, $BIC = 865.289$ y $FPE = 132,962$.

En el párrafo anterior sólo tomamos la parte autorregresiva del proceso. Si deseamos tomar en cuenta la parte de promedio móviles y construir un modelo, por ejemplo, ARMA(1,1) entonces, elegimos “`Model >Specify >AR order = 1 >MA order = 1`”, para después elegir las siguientes opciones “`Model >Estimation >Preliminary >Innovations`”, obteniendo los coeficientes $\phi_1 = 0.261474$, $\theta_1 = -0.693012$ y criterio $AICC = 865.440$. Para máxima verosimilitud debemos elegir “`Model >Estimation >Max Likelihood`”, obteniendo los coeficientes $\phi_1 = 0.485619$, $\theta_1 = -0.937805$ y criterios $AICC = 864.927$ y $BIC = 863.473$.

Ahora bien, si deseamos construir nuestro segundo modelo, es decir, un modelo ARMA(2,1), tenemos que elegir las opciones “`Model >Specify >AR order = 2 >MA order = 1`”, a continuación elegimos las opciones “`Model >Estimation >Preliminary >Innovations`”, obteniendo los coeficientes $\phi_1 = 0.286340$, $\phi_2 = -0.197289$, $\theta_1 = -0.714721$ y criterio $AICC = 870.474$. Para máxima verosimilitud elegimos “`Model >Estimation >Max Likelihood`”, obteniendo los coeficientes $\phi_1 = 0.454505$, $\phi_2 = 0.165270$, $\theta_1 = -0.986089$ y criterios $AICC = 865.859$ y $BIC = 863.643$.

Elegimos aquel modelo que tome en cuenta la parte autorregresiva y la parte de promedios móviles, pero también aquel que posea el criterio de información de Akaike corregido (AICC), el error de predicción final (FPE) (cuando corresponda) y el criterio de información Bayesiana (BIC) menor, es decir, elegimos el modelo ARMA(1,1).

Para realizar pronósticos elegimos las opciones siguientes “`Forecasting >ARMA >Number of forecasts = 12 >Plot 95 percent prediction bounds`”, obteniendo la **Figura 3.6.7**

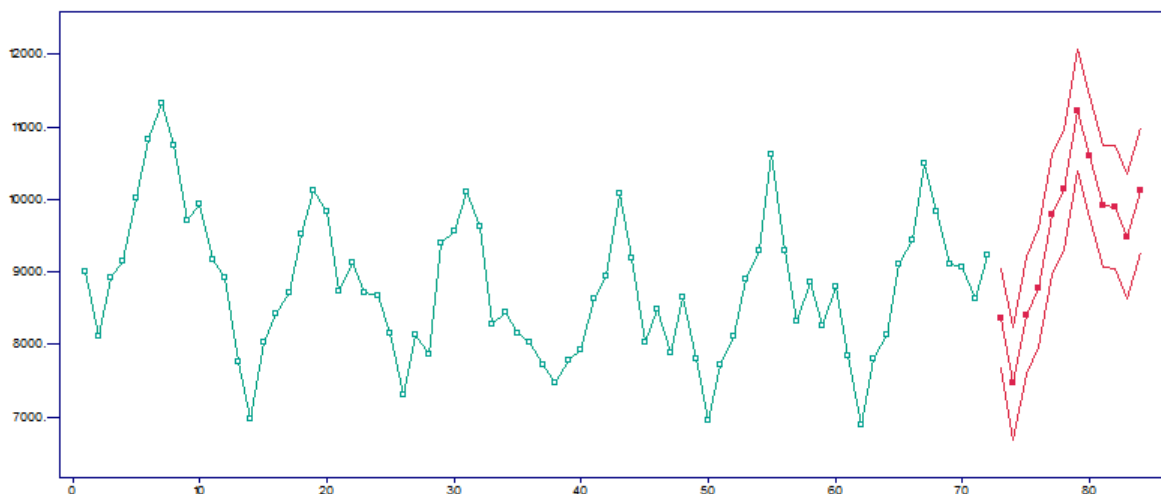


Figura 3.6.7: Pronóstico del número de muertes accidentales en E.U.A. para el año 1979

Sólo falta verificar los supuestos del modelo. Elegimos las opciones “`Statistics >Residual Analysis >Histogram >Default`”. La **Figura 3.6.8** muestra el histograma de los residuos, los cuales distan de poseer distribución Normal. Veamos ahora si la sucesión de autocorrelaciones muestrales (ACF) de los residuos se asemeja a un ruido blanco, es decir, nos interesa que aproximadamente el 95 % de las autocorrelaciones se encuentren dentro de las bandas de confianza, para entonces asegurar la semblanza con un ruido blanco. Observamos de la **Figura 3.6.9** que la mayoría de las autocorrelaciones caen, efectivamente, dentro de las bandas de confianza. Por tanto, estamos hablando de un ruido blanco.

La prueba de Ljung y Box y la prueba McLeod-Li pueden calcularse con “`Statistics >Residual Analysis >Tests of Randomness`”. Para la primer prueba, la prueba de Ljung y Box, dado $\alpha = 0.05$, tenemos un $p\text{-value} = 0.43909 > 0.05$ para la prueba de hipótesis $H_0 : \rho_1 = \rho_2 = \dots = \rho_k = 0$ vs $H_1 : \text{Al menos una } \rho_i \neq 0$, por lo cual no rechazamos H_0 y las autocorrelaciones hasta el retraso $k = 20$ son conjuntamente cero. La misma conclusión se obtiene para la prueba McLeod-Li ya que su correspondiente $p\text{-value} = 0.80059 > 0.05$. Se concluye que los residuos en verdad se comportan como un ruido blanco.

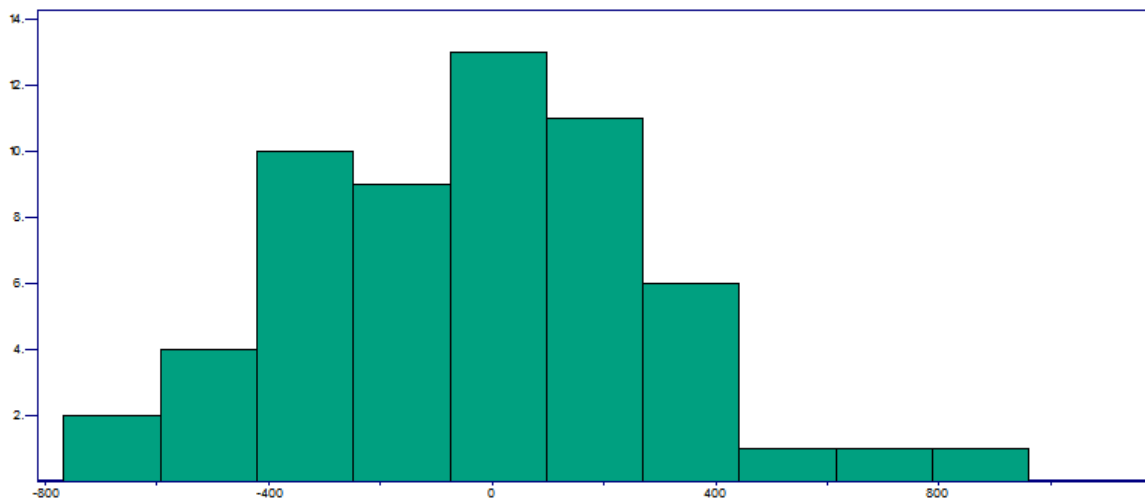


Figura 3.6.8: Figura que muestra el histograma de los residuos

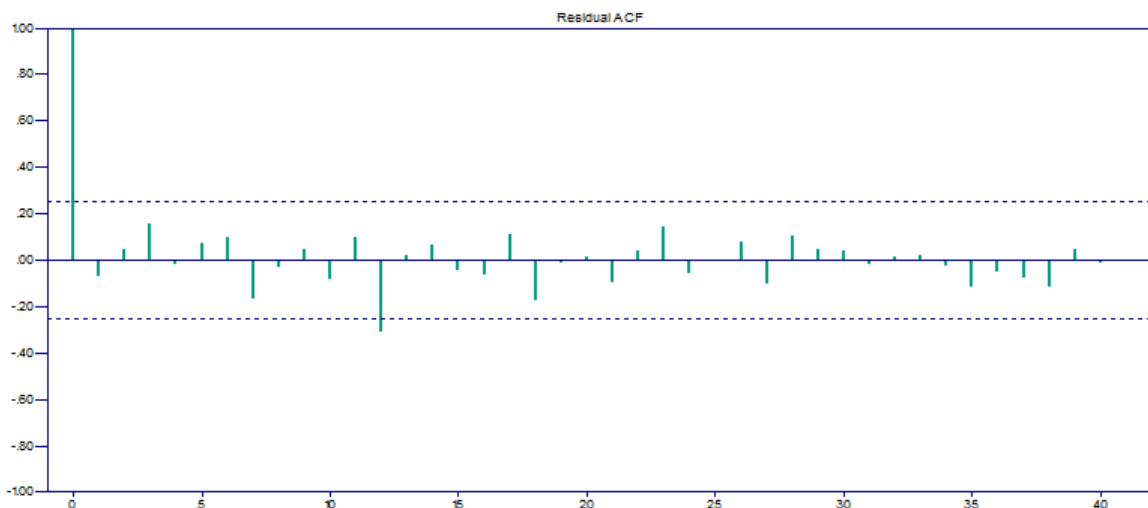


Figura 3.6.9: Figura ACF de los residuos

Pasemos ahora a analizar la serie correspondiente al número de acciones de atención otorgadas por Condusef desde el año 2007 al año 2012. Esta serie ya fue tratada en el **capítulo 2** de este trabajo y el análisis y los pronósticos se realizaron por medio del método de Holt-Winters.

Ahora realizaremos un análisis similar pero basándonos en los modelos ARMA(p,q). Veremos con qué método los pronósticos se acercan más a la realidad. Seleccionamos la siguiente opción “File >Project >Open >Univariate” y elegimos el archivo “Acciones.tsm”. La **Figura 3.6.10** muestra dicha serie.

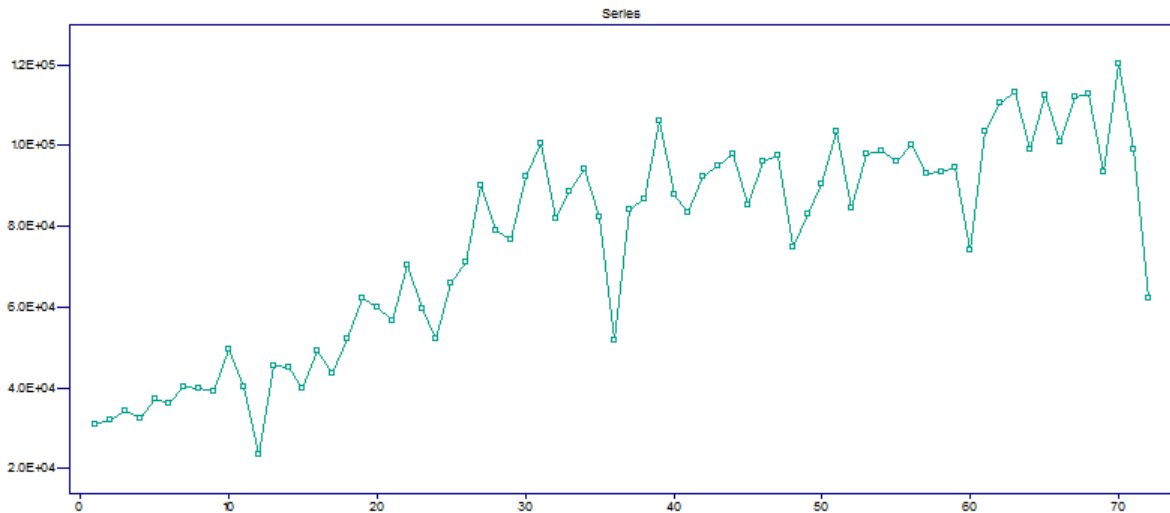


Figura 3.6.10: Figura que muestra el número de acciones de atención

De la **Figura 3.6.10** podemos observar una tendencia lineal ascendente y un patrón estacional con un pico en febrero y un declive en noviembre de cada año. El pico en el primer trimestre del año (especialmente en el mes de febrero) se debe al incremento en las acciones de atención después de la época navideña y el declive se debe al decremento en las acciones de atención por ser fin de año y comienzo de época navideña.

Ahora bien, para eliminar la tendencia aplicaremos el operador ∇^j con un “lag” de 1 y para eliminar la estacionalidad aplicaremos el operador ∇_s con un “lag” de 12. No olvidemos restar la media para así centrar al nuevo proceso.

Para eliminar la tendencia, elegimos “Transform >Difference >Enter lag = 1”, obteniendo la **Figura 3.6.11**. Para eliminar la componente estacional tenemos que seleccionar “Transform >Difference >Enter lag = 12”, obteniendo la **Figura 3.6.12** y para centrar la serie elegimos “Transform >Difference >Substract Mean”, obteniendo finalmente la **Figura 3.6.13**.

Ahora, mostramos la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) para elegir el modelo ARMA(p,q) óptimo. Elegimos “Statistics >ACF/PACF >Sample”, obteniendo la **Figura 3.6.14** y la **Figura 3.6.15** respectivamente.

Observamos de la sucesión de autocorrelaciones muestrales (ACF) que se salen las autocorrelaciones correspondientes al retraso 1, 12, 17 y 25. Por otro lado, de la sucesión de autocorrelaciones muestrales parciales (PACF) se salen las autocorrelaciones correspondientes al “lag” 1, 2 y 11. Nuestras opciones serán un modelo ARMA(1,1) y un modelo ARMA(2,1) ya que los demás procesos resultan ser no causales.

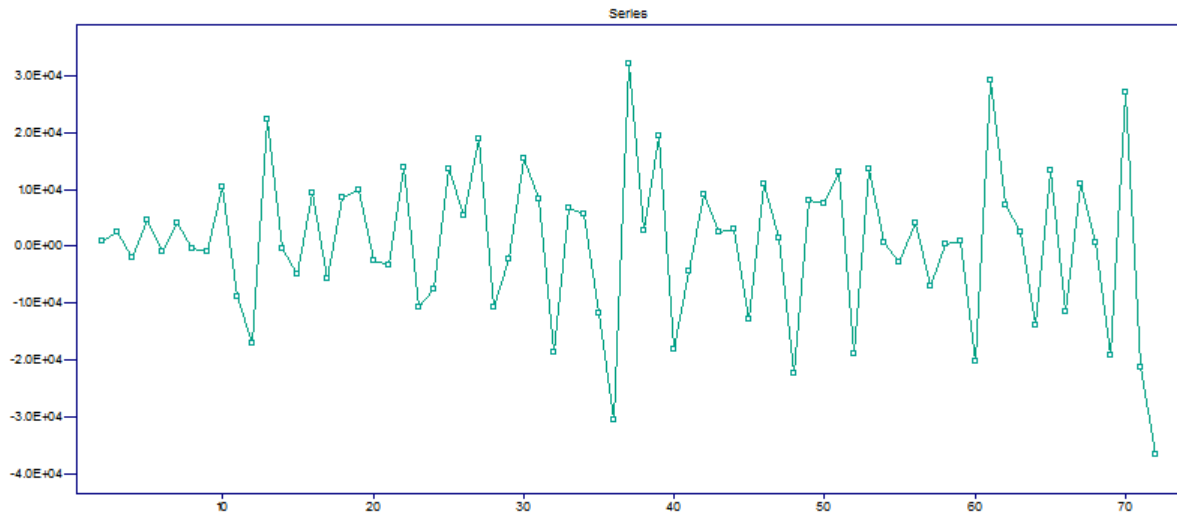


Figura 3.6.11: Figura que muestra el número de acciones de atención sin tendencia

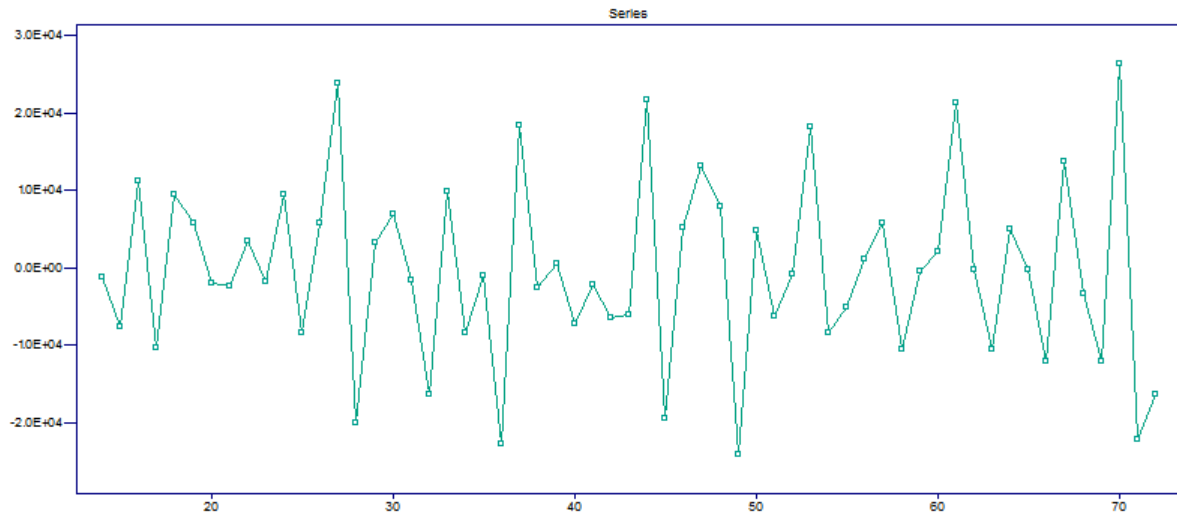


Figura 3.6.12: Figura que muestra el número de acciones de atención sin estacionalidad

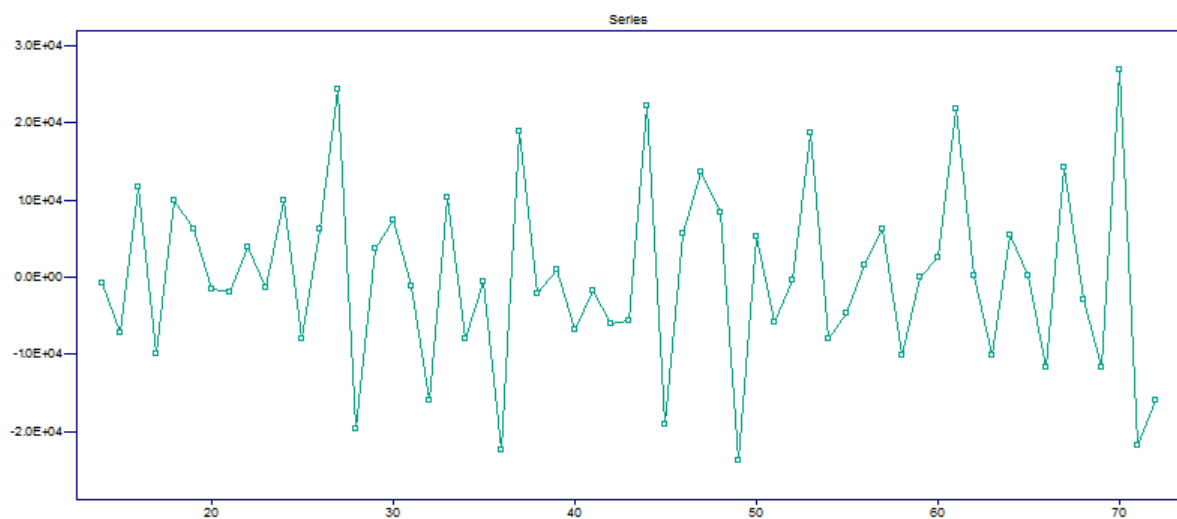


Figura 3.6.13: Figura que muestra el número de acciones de atención restando la media

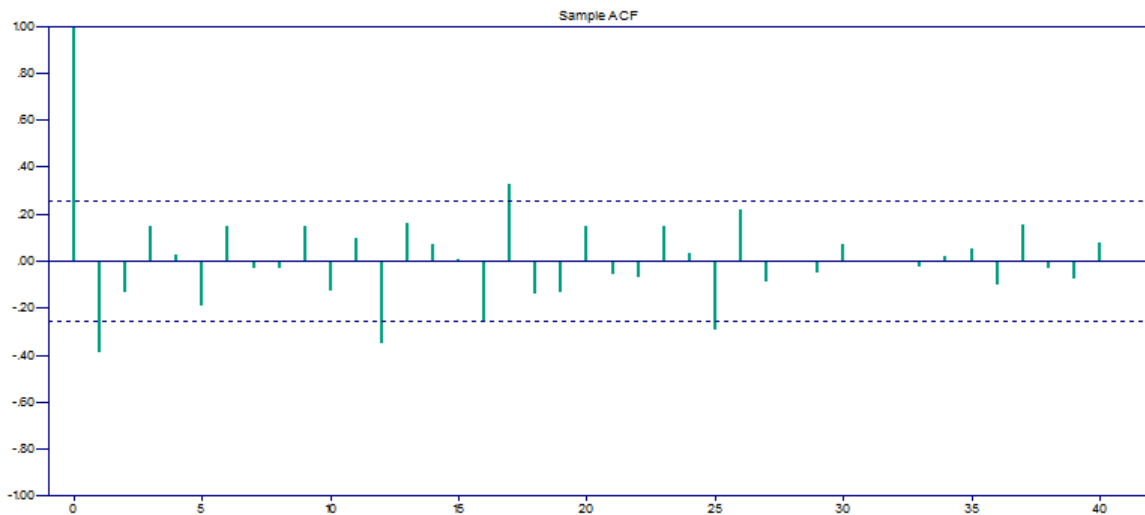


Figura 3.6.14: Figura ACF del número de acciones de atención

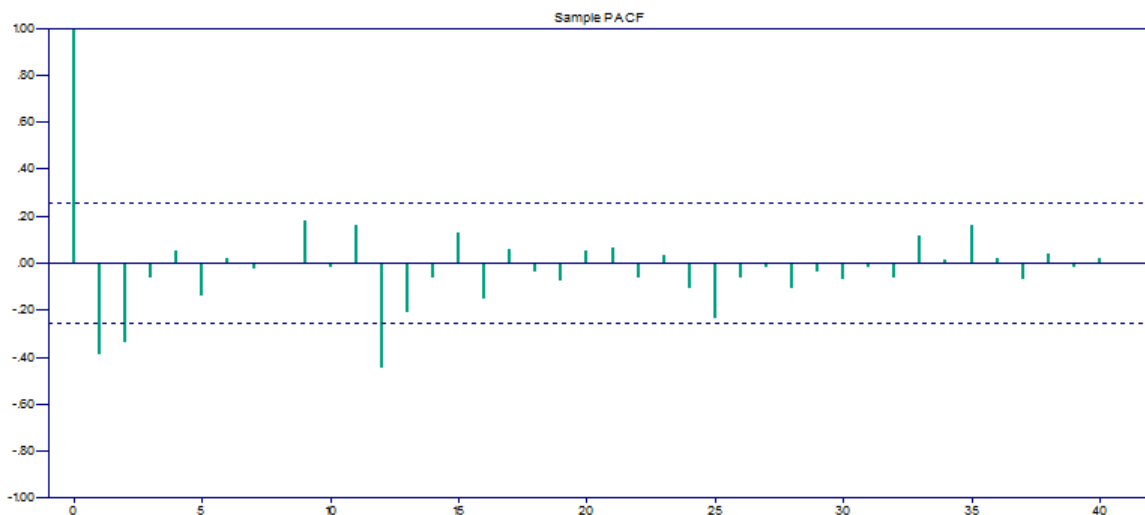


Figura 3.6.15: Figura PACF del número de acciones de atención

Para construir un modelo ARMA(1,1) tenemos que elegir las opciones “Model >Specify >AR order = 1 >MA order = 1”, a continuación elegimos las opciones “Model >Estimation >Preliminary >Innovations”, obteniendo los coeficientes $\phi_1 = 0.183155$, $\theta_1 = -0.702112$ y criterio $AICC = 1261.92$. Para máxima verosimilitud debemos elegir “Model >Estimation >Max Likelihood”, obteniendo los coeficientes $\phi_1 = -0.013740$, $\theta_1 = -0.598334$ y criterios $AICC = 1261.04$ y $BIC = 1261.82$.

Para la construcción del modelo ARMA(2,1) seleccionamos “Model >Specify >AR order = 2 >MA order = 1”. Después elegimos “Model >Estimation >Preliminary >Innovations”, obteniendo los coeficientes $\phi_1 = 0.203951$, $\phi_2 = -0.571660$, $\theta_1 = -0.720695$ respectivamente, y criterio $AICC = 1280.35$. Para máxima verosimilitud debemos elegir “Model >Estimation >Max Likelihood”, obteniendo los coeficientes autorregresivos $\phi_1 = -0.447829$, $\phi_2 = -0.364266$, de promedios móviles $\theta_1 = -0.164504$ y criterios $AICC = 1260.75$ y $BIC = 1260.69$. Elegimos aquel modelo con menor $AICC$ y BIC , a saber, el modelo ARMA(2,1).

Para realizar pronósticos elegimos las opciones siguientes “Forecasting >ARMA >Number of forecasts = 12 >Plot 95 percent prediction bounds”, obteniendo la **Figura 3.6.16**.

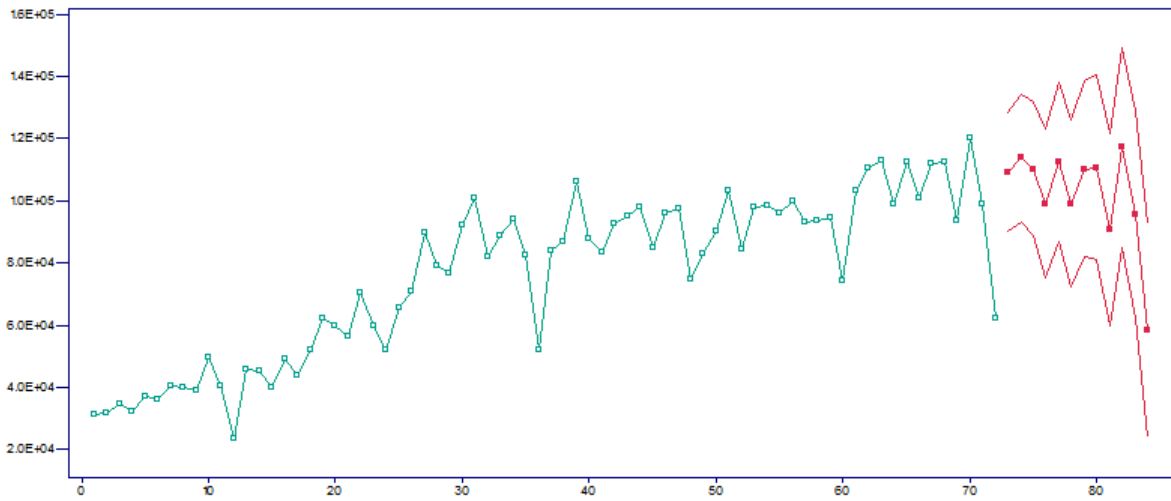


Figura 3.6.16: Pronóstico del número de acciones de atención para el año 2013

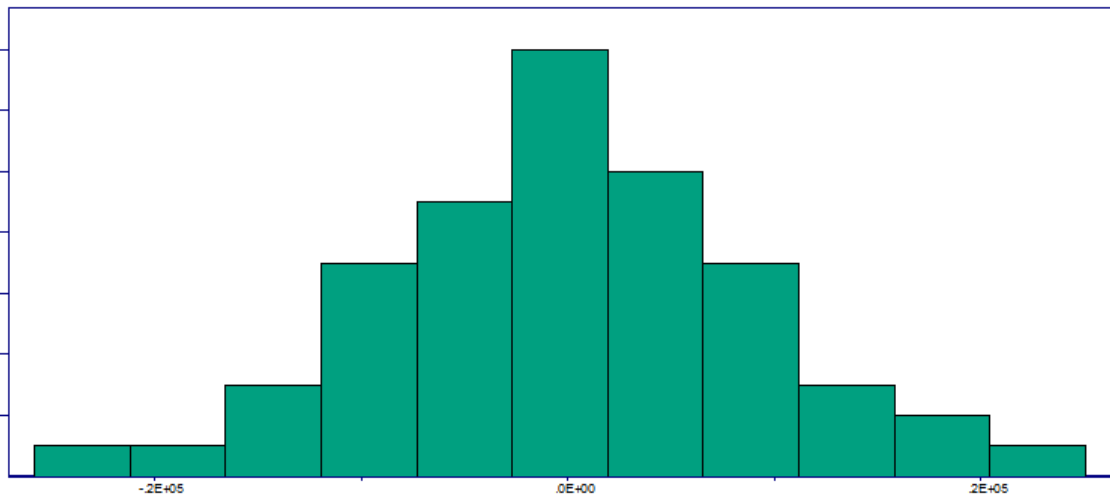


Figura 3.6.17: Figura que muestra el histograma de los residuos

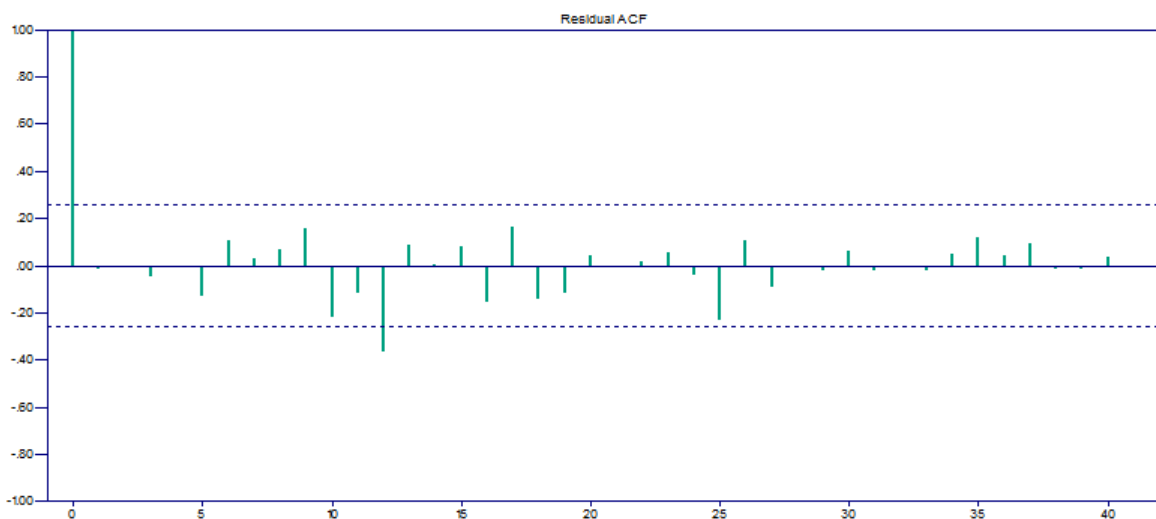


Figura 3.6.18: Figura ACF de los residuos

Resta verificar los supuestos del modelo. Elegimos las opciones “**Statistics >Residual Analysis >Histogram >Default**”. La **Figura 3.6.17** muestra el histograma de los residuos, los cuales claramente tienen distribución Normal. Además, observamos de la sucesión de autocorrelaciones muestrales (ACF) de los residuos, mostrada en la **Figura 3.6.17**, que más del 95% de las autocorrelaciones para 40 retrasos, caen dentro de las bandas de confianza. Los residuos en efecto se comportan como un ruido blanco.

Pasemos a nuestras dos pruebas de aleatoriedad. Seleccionamos “**Statistics >Residual Analysis >Test of Randomness >22**” y dado $\alpha = 0.05$, observamos de la prueba de Ljung y Box un $p\text{-value} = 0.09358 > 0.05$, por lo cual no rechazamos la hipótesis nula de que las autocorrelaciones hasta el retraso 22 son cero. A la misma conclusión se llega con la prueba McLeod-Li cuyo $p\text{-value} = 0.37013 > 0.05$.

Para exportar los valores pronosticados elegimos “**File >Export >ARMA Forecasts**” y los guardamos donde creamos conveniente. La **Tabla 3.6.1** muestra los correspondientes pronósticos con el método de Holt-Winters, aquellos con el modelo ARMA(2,1) y los valores reales, concluyendo que un modelo ARMA(p,q) es el indicado (en el sentido de que se aproxima de manera eficiente a la realidad y cumple a la perfección con los supuestos estadísticos) entre nuestras opciones para predecir el número de acciones de atención.

<i>Mes</i>	<i>Año</i>	<i>HW Aditivo</i>	<i>HW Multiplicativo</i>	<i>ARMA(2,1)</i>	<i>Valores reales</i>
Enero	2013	100,682	101,067	110,261	112,496
Febrero	2013	105,205	105,880	115,337	105,837
Marzo	2013	111,354	112,390	111,976	92,589
Abril	2013	97,316	96,889	101,318	118,764
Mayo	2013	103,867	104,761	115,333	113,360
Junio	2013	101,097	100,684	102,308	106,900
Julio	2013	106,919	107,286	113,673	127,885
Agosto	2013	106,097	107,282	114,752	116,374
Septiembre	2013	94,477	93,938	95,261	102,100
Octubre	2013	109,609	110,871	122,203	130,605
Noviembre	2013	98,579	98,542	100,984	106,262
Diciembre	2013	72,583	68,154	64,160	70,159

Tabla 3.6.1: Pronósticos para el número de acciones de atención para el año 2013

Si las propiedades estadísticas se mantienen constantes en el número de acciones otorgadas por Condusef, entonces, es nuestra creencia que un modelo ARMA(p,q), en particular aquél con parámetros $p = 2$ y $q = 1$, puede ser adoptado de ahora en adelante para así realizar una mejor predicción en el número de acciones y por tanto, con valores mucho más cercanos a la realidad, realizar propuestas que mejoren la atención brindada a los usuarios de servicios financieros.

Pasemos a analizar la tercera serie, es decir, aquella que muestra las ventas mensuales de vino tinto por parte de viñedos australianos. Seleccionamos las opciones “**File >Project >Open >Univariate**” y elegimos el archivo “**Wine.tsm**”. La **Figura 3.6.19** muestra dicha serie. Observamos inmediatamente una tendencia lineal ascendente y un patrón estacional con pico en el mes de julio y un declive en el mes de enero de cada año. Eliminaremos la tendencia y la estacionalidad por medio del método de descomposición clásica, para después ajustar un modelo ARMA(p,q). Seleccionamos las opciones “**Transform >Classical >Seasonal Fit = 12 >Polynomial Fit = Linear Trend**”. El ajuste por descomposición clásica se muestra en la **Figura 3.6.20**. La serie a analizar, sin tendencia ni estacionalidad se muestra en la **Figura 3.6.21**. No olvidemos restar la media a nuestra nueva serie.

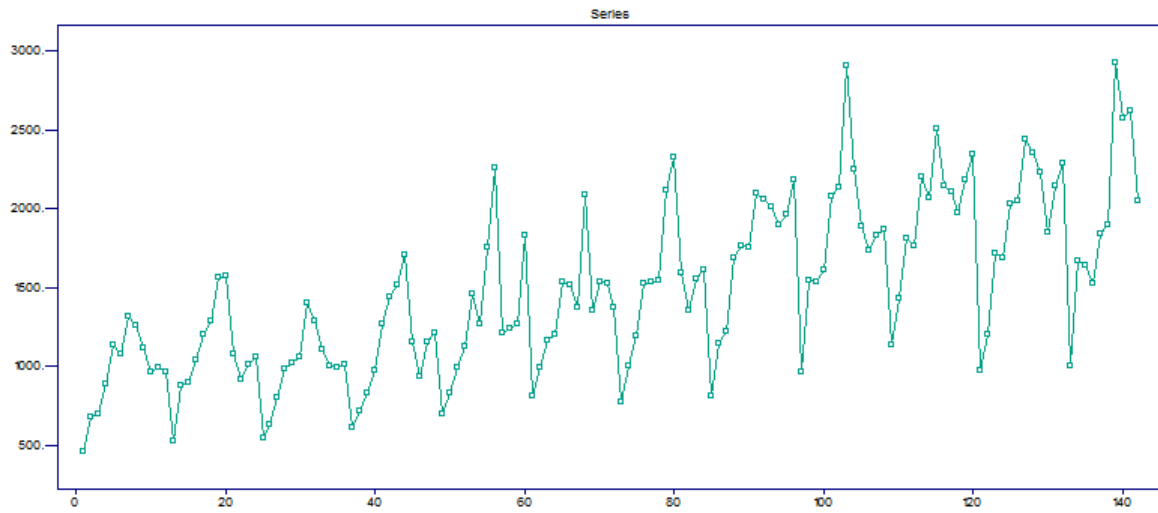


Figura 3.6.19: Figura que muestra el número de ventas mensuales de vino tinto

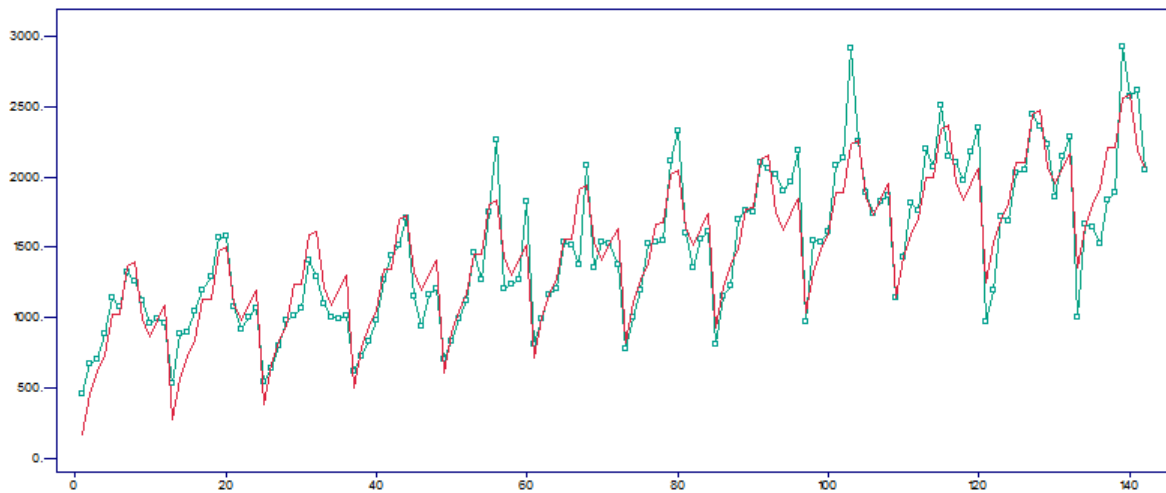


Figura 3.6.20: Figura que muestra la descomposición clásica de las ventas de vino tinto

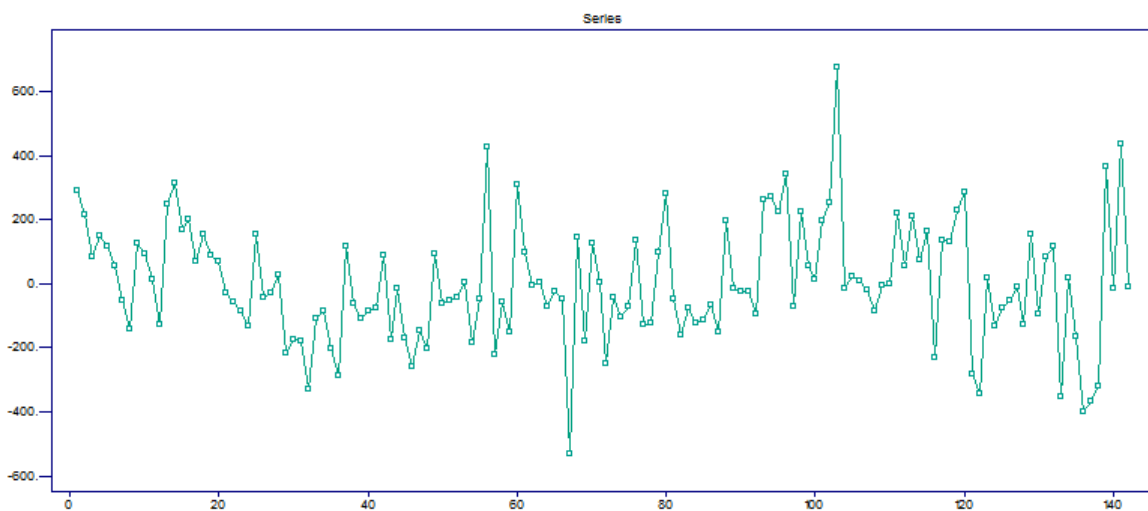


Figura 3.6.21: Figura que muestra ventas mensuales de vino sin tendencia ni estacionalidad

A continuación mostramos la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) para elegir el modelo ARMA(p,q) óptimo. Elegimos nuevamente “Statistics >ACF/PACF >Sample”, obteniendo la **Figura 3.6.22** y la **Figura 3.6.23** respectivamente.

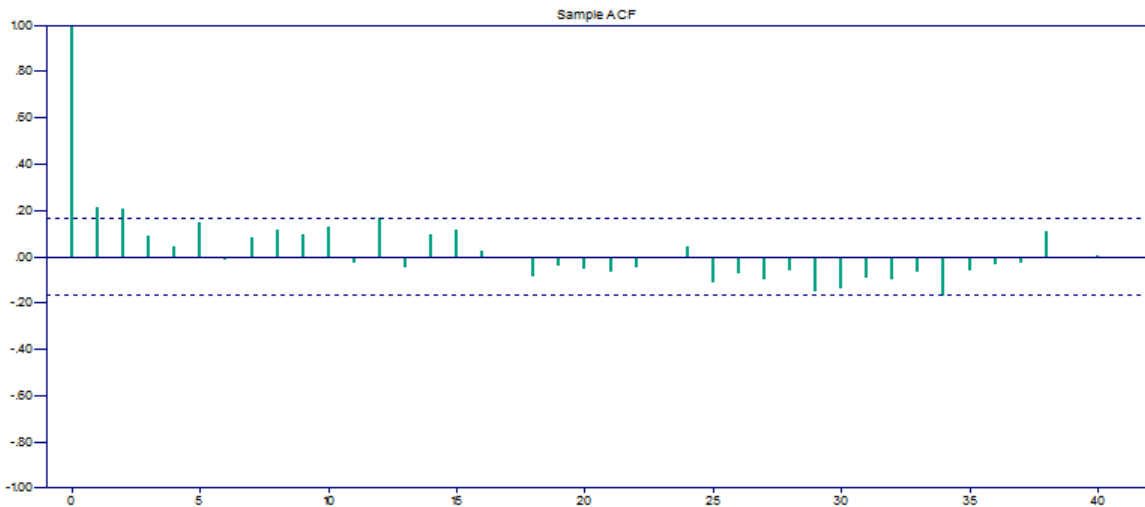


Figura 3.6.22: Figura ACF del número de ventas mensuales de vino tinto

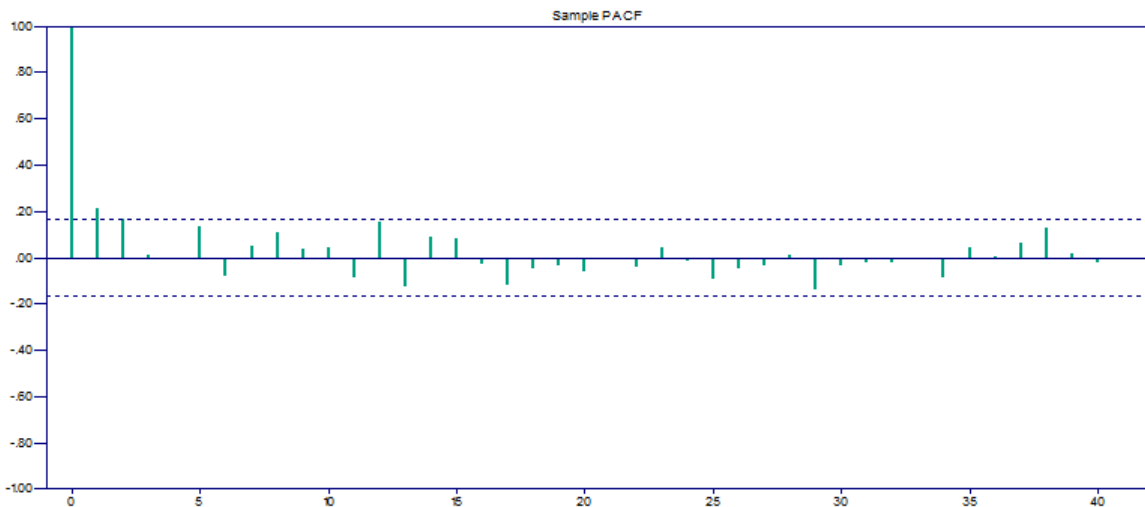


Figura 3.6.23: Figura PACF del número de ventas mensuales de vino tinto

De la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones parciales (PACF) observamos que el modelo ARMA(p,q) óptimo es aquel con parámetros $p = 1$ y $q = 2$. Para construir un modelo ARMA(1,2) tenemos que elegir las opciones “Model >Specify >AR order = 1 >MA order = 2”, a continuación elegimos las opciones “Model >Estimation >Preliminary >Innovations”, obteniendo los coeficientes $\phi_1 = 0.352789$, $\theta_1 = -0.138266$, $\theta_2 = 0.121670$ y criterio $AICC = 1879.25$. Para máxima verosimilitud debemos elegir “Model >Estimation >Max Likelihood”, obteniendo los coeficientes $\phi_1 = 0.602621$, $\theta_1 = -0.422800$, $\theta_2 = 0.167407$ y criterios $AICC = 1878.94$ y $BIC = 1877.68$.

Para realizar pronósticos elegimos las opciones siguientes “Forecasting >ARMA >Number of forecasts = 12 >Plot 95 percent prediction bounds”, obteniendo la **Figura 3.6.24**.

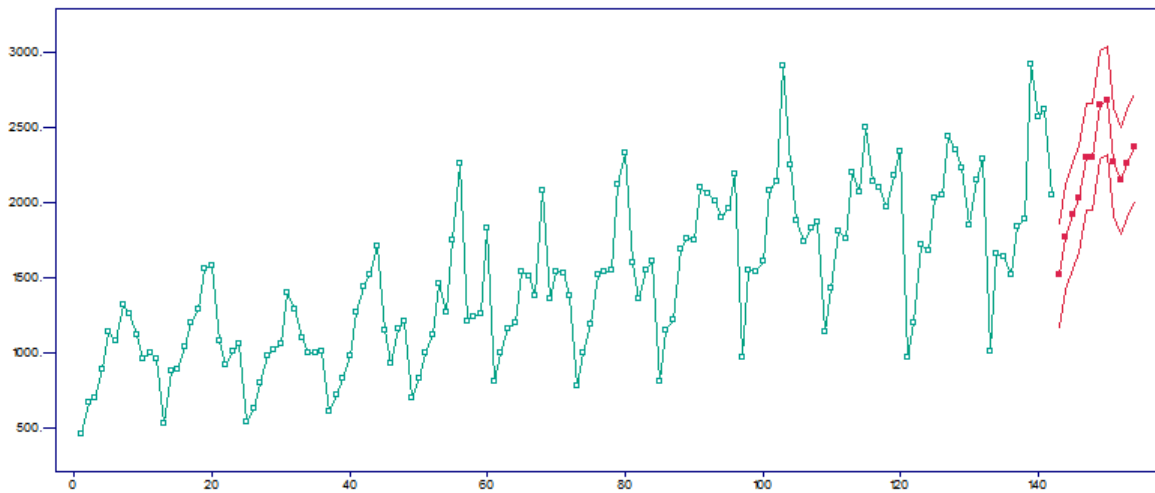


Figura 3.6.24: Pronóstico del número de ventas mensuales de vino tinto para el año 1992

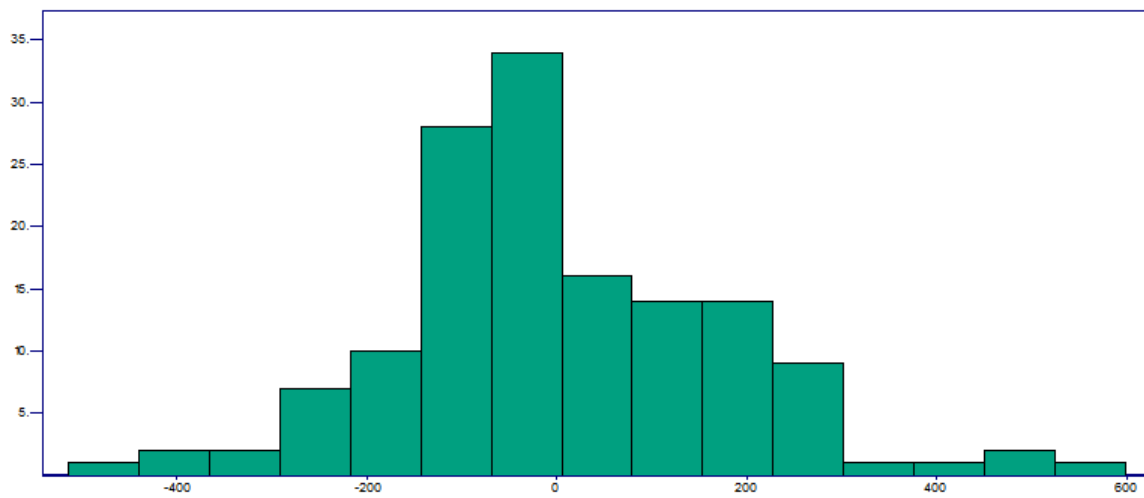


Figura 3.6.25: Figura que muestra el histograma de los residuos

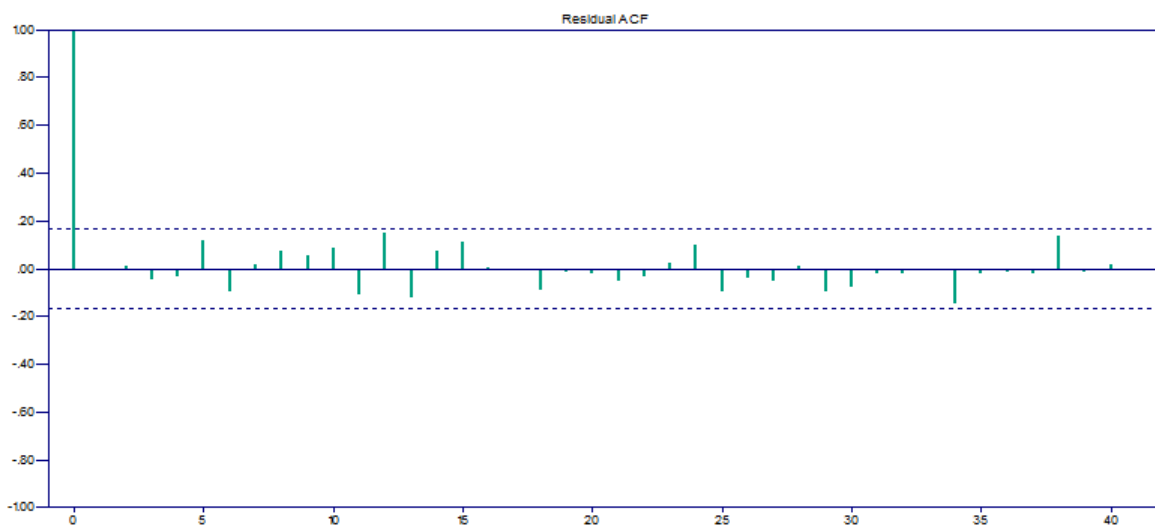


Figura 3.6.26: Figura ACF de los residuos

Pasemos a verificar los supuestos del modelo. Elegimos las opciones “**Statistics >Residual Analysis >Histogram >Default**”. La **Figura 3.6.25** muestra el histograma de los residuos, los cuales parecen tener distribución Normal. Ninguna de las autocorrelaciones en la sucesión de autocorrelaciones muestrales (ACF) de los residuos caen fuera de las bandas de confianza y por tanto estamos tratando con un ruido blanco. Para las pruebas de aleatoriedad seleccionamos “**Statistics >Residual Analysis >Test of Randomness >22**” y observamos la prueba de Ljung y Box cuyo $p\text{-value} = 0.42695 > 0.05$, por lo cual no rechazamos la hipótesis nula de que en conjunto, las autocorrelaciones hasta el retraso 23 son cero. Se concluye de manera análoga para la prueba McLeod-Li ya que su respectivo $p\text{-value} = 0.99821 > 0.05$.

Por último, tomemos la serie correspondiente al nivel (pies) del lago Huron. Realizaremos nuevamente el método de descomposición clásica, para después ajustar un modelo ARMA(p,q) con parámetros $p = 2$ y $q = 3$. La **Figura 3.6.27** muestra dicha serie.

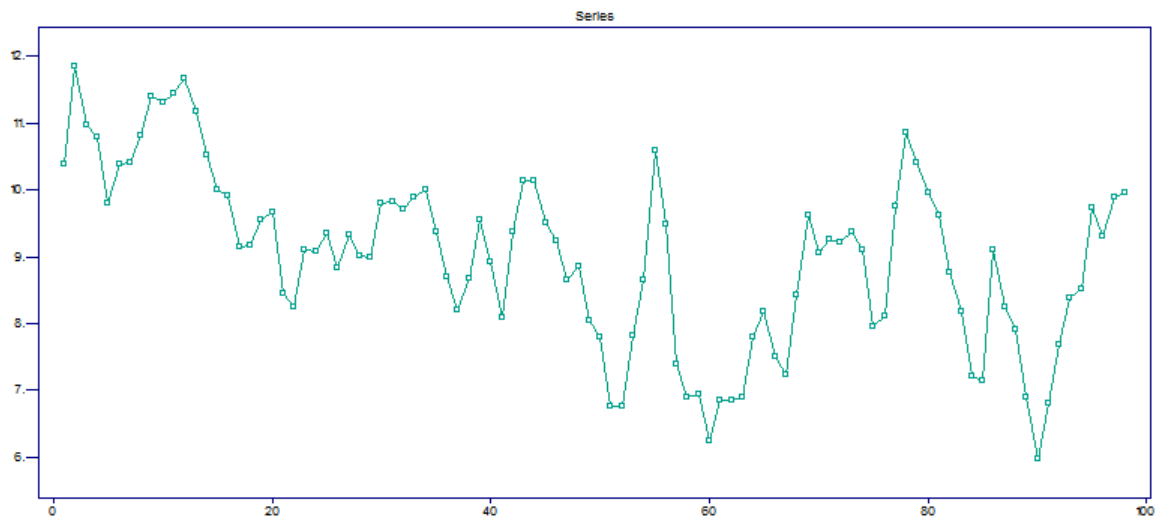


Figura 3.6.27: Figura que muestra el nivel del lago Hurón

Seleccionamos las opciones “**Transform >Classical >Seasonal Fit = 12 >Polynomial Fit = Linear Trend**”. El ajuste por descomposición clásica se muestra en la **Figura 3.6.28**.

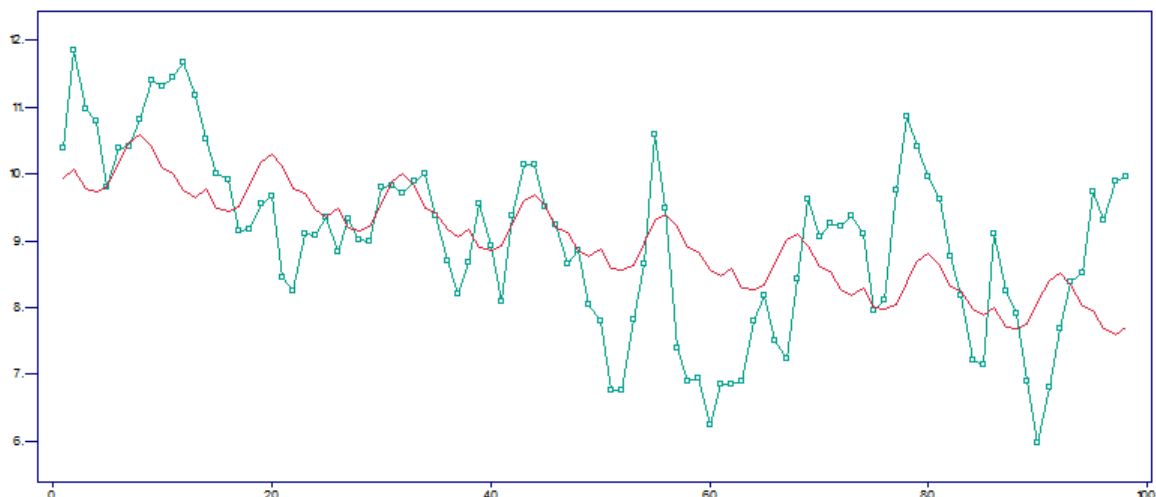


Figura 3.6.28: Figura que muestra la descomposición clásica del nivel del lago Hurón

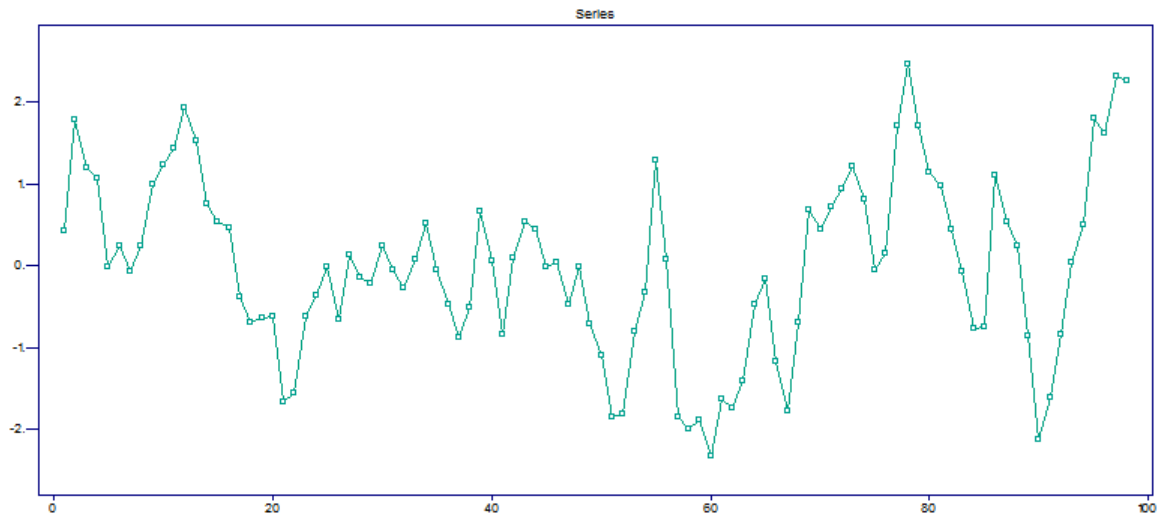


Figura 3.6.29: Figura que muestra el nivel del lago Hurón sin tendencia ni estacionalidad

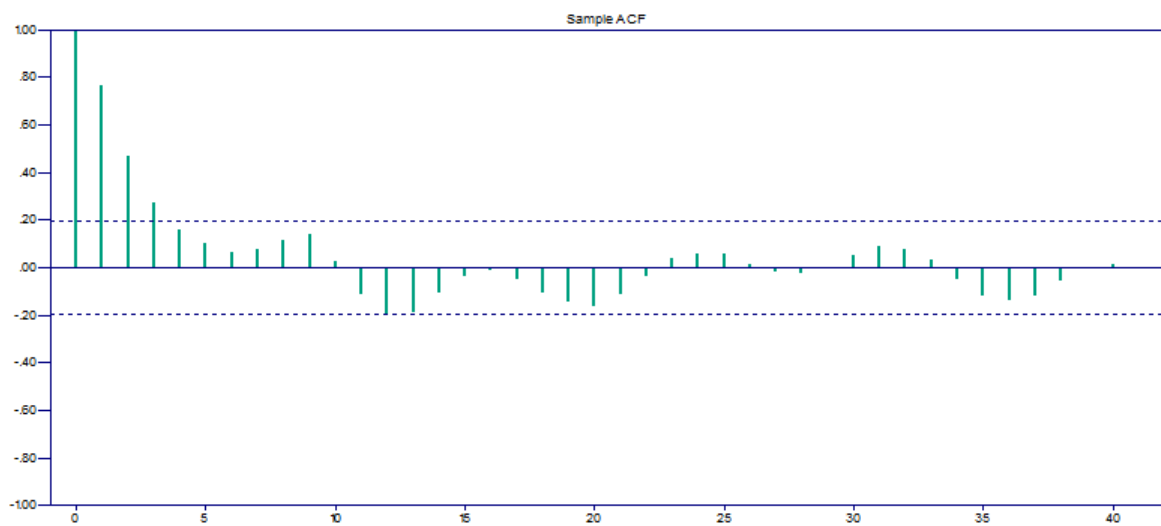


Figura 3.6.30: Figura ACF del nivel del lago Hurón

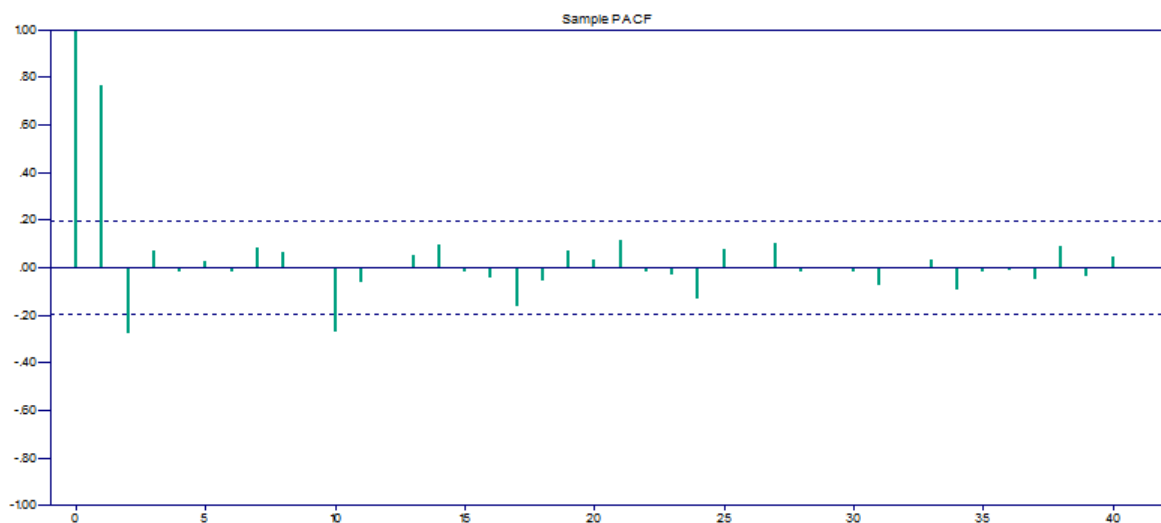


Figura 3.6.31: Figura PACF del nivel del lago Hurón

La serie a analizar, sin tendencia ni estacionalidad se muestra en la **Figura 3.6.29**. No olvidemos restar la media a nuestra nueva serie. La sucesión de autocorrelaciones muestrales (ACF) se observa en la **Figura 3.6.30** y la sucesión de autocorrelaciones muestrales parciales (PACF) se observa en la **Figura 3.6.31**.

Para construir un modelo ARMA(2,3) tenemos que elegir las opciones “Model >Specify >AR order = 2 >MA order = 3”, a continuación elegimos las opciones “Model >Estimation >Preliminary >Innovations”, obteniendo los coeficientes $\phi_1 = 0.202363$, $\phi_2 = 0.342407$, $\theta_1 = 0.758686$, $\theta_2 = 0.103300$, $\theta_3 = 0.013342$ y criterio $AICC = 206.829$. Para máxima verosimilitud debemos elegir “Model >Estimation >Max Likelihood”, obteniendo los coeficientes $\phi_1 = 0.244096$, $\phi_2 = 0.227036$, $\theta_1 = 0.792488$, $\theta_2 = 0.232734$, $\theta_3 = 0.023938$ y criterios $AICC = 205.675$ y $BIC = 214.404$.

Para realizar pronósticos elegimos las opciones siguientes “Forecasting >ARMA >Number of forecasts = 12 >Plot 95 percent prediction bounds”, obteniendo la **Figura 3.6.32**.

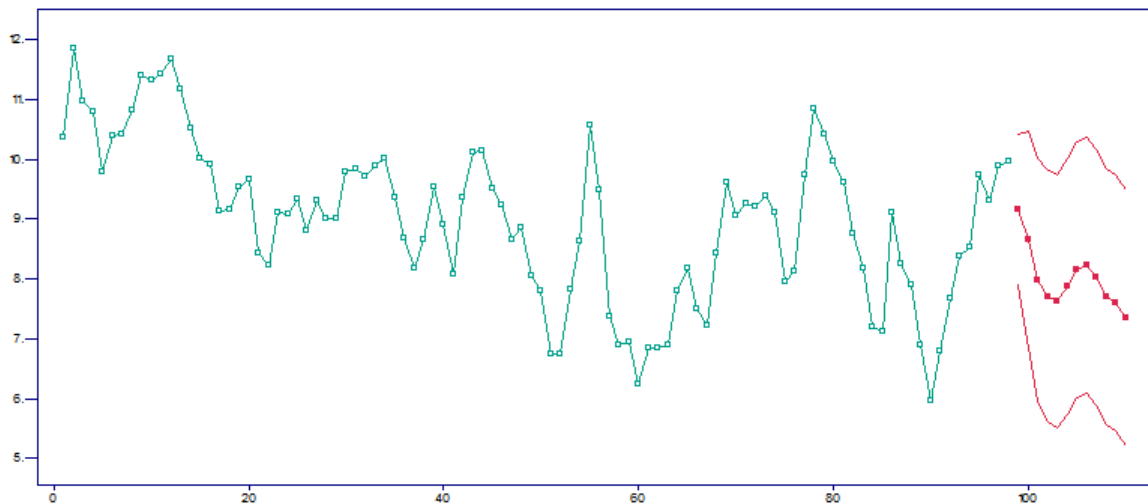


Figura 3.6.32: Pronóstico del nivel del lago Hurón para el año 1973

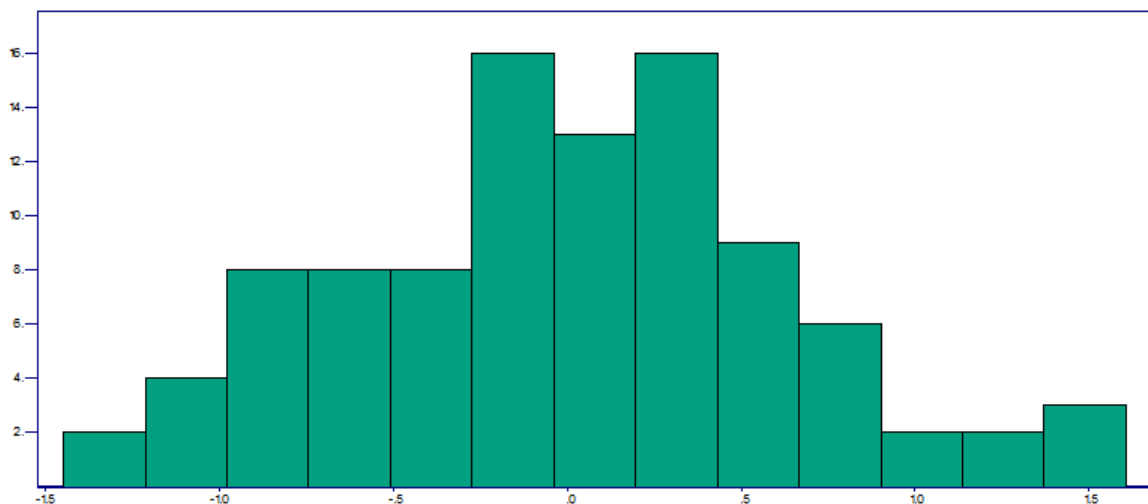


Figura 3.6.33: Figura que muestra el histograma de los residuos

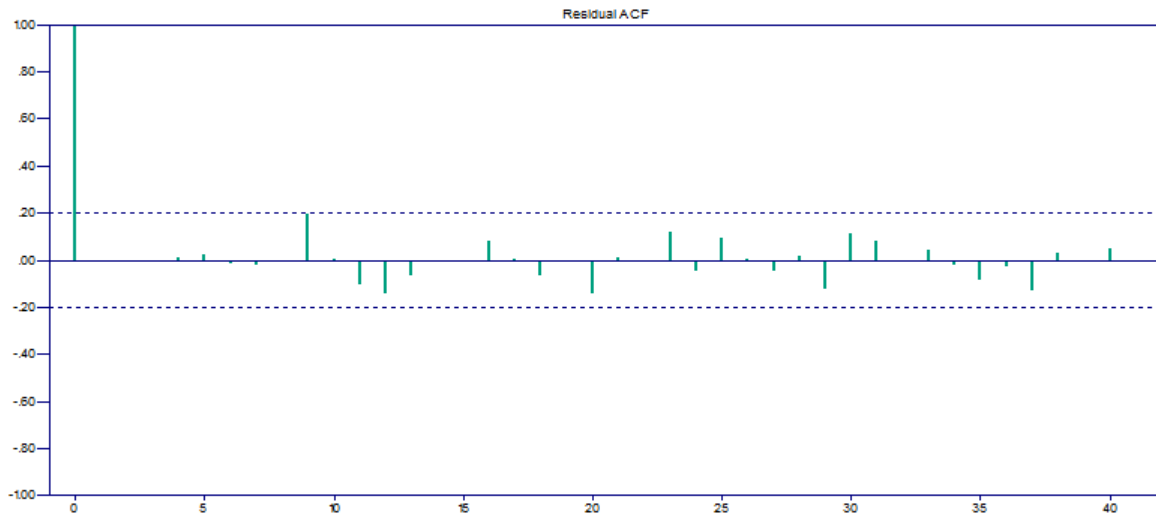


Figura 3.6.34: Figura ACF de los residuos

La **Figura 3.6.33** muestra el histograma de los residuos, los cuales no distan mucho de normalidad. Finalmente, volvemos a elegir “Statistics >Residual Analysis >Test of Randomness >22” y observamos la prueba de Ljung y Box cuyo p-value= 0.70133 > 0.05, por lo cual no rechazamos la hipótesis nula H_0 y concluimos que las autocorrelaciones hasta el “lag” $h = 25$ conjuntamente son cero. Asimismo obtenemos la prueba McLeod-Li cuyo p-value= 0.36907 > 0.05, con lo cual corroboramos lo obtenido con la prueba de Ljung y Box. Finalmente, ninguna de las autocorrelaciones mostradas en la **Figura 3.6.34** se encuentra fuera de las bandas de confianza, por lo que los residuos se comportan como un ruido blanco.

El paquete estadístico ITSM no nos brinda las pruebas no paramétricas para normalidad **Shapiro-Wilk** y **Lilliefors**, pero nos permite exportar los residuos mediante las opciones “File >Export >Residuals” para después llevarlos a R y aquí realizar las pruebas no paramétricas para normalidad. Para ilustrar lo anterior, veamos si la primera serie posee distribución Normal en los residuos. Hemos elegido esta serie ya que a nuestro parecer es la más complicada de asegurar normalidad. Habiendo exportado los residuos el código en R será:

```
> lillie.test(residuos)
```

```
Lilliefors (Kolmogorov-Smirnov) normality test
```

```
data:  residuos
```

```
D = 0.0653, p-value = 0.7684
```

```
> shapiro.test(residuos)
```

```
Shapiro-Wilk normality test
```

```
data:  residuos
```

```
W = 0.9658, p-value = 0.09585
```

El p-value para la prueba Lilliefors es 0.7684. La prueba de hipótesis para ambas pruebas será H_0 : Normalidad en los residuos vs H_1 : No normalidad en los residuos, y dado $\alpha = 0.05$, rechazamos H_0 ya que p-value= 0.7684 > 0.05 y concluimos que la distribución de los residuos es, en efecto, normal. A la misma conclusión se llega utilizando la prueba Shapiro-Wilk ya que p-value= 0.0958 > 0.05.

3.7. Modelos ARIMA

Los modelos ARIMA(p,d,q), modelos autorregresivos e integrados de promedios móviles, son una generalización de los modelos ARMA(p,q) previamente estudiados. Surgen ante la presencia de series de tiempo no estacionarias pero que pueden llegar a serlo mediante la toma sucesiva de diferencias de la serie original. Es decir, si el proceso original $\{X_t\}$ carece de estacionariedad debido a la existencia de una tendencia polinomial, entonces es posible construir, y haciendo uso del **Teorema 3.2.1**, un proceso estacionario $\{W_t\}$ dado por:

$$W_t = \nabla^d X_t \quad (3.7.1)$$

Para esta serie, $\{W_t\}$, podría ya ser posible obtener un proceso ARMA(p,q), dado por la expresión, $\phi(B)W_t = \theta(B)\epsilon_t$, lo cual sería equivalente a tomar un modelo ARIMA(p,d,q), en el cual se indica el orden del polinomio autorregresivo (p), el orden de diferencias a utilizar (d) y el orden del polinomio de promedios móviles(q). Así, el proceso quedará especificado como:

$$\phi(B)\nabla^d X_t = \theta(B)\epsilon_t \quad (3.7.2)$$

con $\{\epsilon_t\}$ ruido blanco con $\mathbb{E}(\epsilon_t) = 0$. El término “integrado” se refiere a que X_t se obtiene de la ec. **3.7.1** por inversión del operador ∇^d , dando como resultado una suma infinita (integración) de términos de W_t , es decir:

$$X_t = \nabla^{-1}W_t = W_t + W_{t-1} + W_{t-2} + W_{t-3} + \dots \quad (3.7.3)$$

Ahora bien, el orden de diferencias a utilizar puede obtenerse al observar la sucesión de autocorrelaciones muestrales (ACF), ya que un rápido decaimiento de las autocorrelaciones a cero es indicativo de que la serie es estacionaria. De no ser estacionaria la serie original, se procede a tomar diferencias, recordando claro no diferenciar de más, ya que esto complicaría el regreso a la serie original, a tal grado de encontrar una serie estacionaria. Además, el orden de la diferenciación depende del tipo de tendencia polinomial en los datos según el **Teorema 3.2.1**.

Un método alternativo, propuesto por Guerrero (2003), para encontrar el orden de diferencias es mediante el cálculo de la desviación estándar muestral de las series $\{\nabla X_t\}$, $\{\nabla^2 X_t\}$ y $\{\nabla^3 X_t\}$ denotada por:

$$S(j) = \sqrt{\frac{1}{N-j-1} \sum_{t=j+1}^N \left[\nabla^j X_t - \sum_{t=j+1}^N \frac{\nabla^j X_t}{N-j} \right]^2} \quad (3.7.4)$$

para $j = 0, 1, 2, 3$. El orden de diferenciación d será $S(d) = \min\{S(j), j = 0, 1, 2, 3\}$, es decir, seleccionamos aquella j que posea la desviación estándar muestral mínima. ¿Por qué solo tomamos $j = 0, 1, 2, 3$? Bueno pues, en la experiencia se ha notado que rara vez se necesita diferenciar para $j > 3$. Aún así, este método puede extenderse $\forall j$ pero debemos recordar los problemas que pueden surgir al diferenciar de más una serie de tiempo, a saber uno de ellos, la dificultad de identificar un modelo que reproduzca la serie original.

Como recomendación, este método debe ser siempre combinado con la obtención de la sucesión de autocorrelaciones muestrales (ACF) y con las reglas que daremos más adelante.

El modelo ARIMA(p,d,q) general conlleva una inmensa variedad de patrones en la sucesión de autocorrelaciones muestrales (ACF) y en la sucesión de autocorrelaciones muestrales parciales (PACF), por lo que es inapropiado el enunciar reglas para identificar dichos modelos. A pesar de esto, los modelos AR(p) y MA(q) proveen algunas características que nos ayudarán a identificar los modelos ARIMA(p,d,q).

La estacionalidad puede sugerirse así misma al observar los correspondientes valores en $1_s, 2_s, 3_s, \dots$ (el primer periodo, segundo periodo, etc.; por ejemplo, para $s = 12$ el primer mes del año 1, después el primer mes del año 2, etc.), de la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF). Estos valores serán grandes y significativamente distintos de cero. Los modelos AR(p) y MA(q) sabemos que pueden ser identificados por el patrón de la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF). Identificamos el orden p del proceso AR(p) y el orden q del proceso MA(q) al ligar los patrones de la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) con los patrones teóricos de modelos conocidos. La **Tabla 3.7.1** muestra las características principales de la sucesión de autocorrelaciones teóricas y de la sucesión de autocorrelaciones parciales teóricas.

<i>Modelo</i>	<i>Sucesión de autocorr. muestrales (ACF)</i>	<i>Sucesión de autocorr. muestrales parciales (PACF)</i>
AR(1)	Las autocorrelaciones caen de manera exponencial. Todas serán positivas si $\phi_1 > 0$ y alternarán signos en forma sinoidal si $\phi_1 < 0$.	Pico en el “lag” 1 seguido de autocorrelaciones cercanas a cero. El pico será positivo si $\phi_1 > 0$ y negativo si $\phi_1 < 0$.
AR(p)	Las autocorrelaciones caen de manera exponencial. El decaimiento será de forma sinoidal. El patrón exacto depende del signo y tamaño de $\phi_1, \phi_2, \dots, \phi_p$.	Picos a partir del retraso o “lag” 1 y hasta el retraso o “lag” p, seguidos de autocorrelaciones cercanas a cero.
MA(1)	Pico en el “lag” 1 seguido de autocorrelaciones cercanas a cero. El pico será positivo si $\theta_1 < 0$ y negativo si $\theta_1 > 0$.	Las autocorrelaciones caen de manera exponencial. Todas serán negativas si $\theta_1 > 0$ y alternarán signos comenzando positivas si $\theta_1 < 0$.
MA(q)	Picos a partir del retraso o “lag” 1 y hasta el retraso o “lag” q, seguidos de autocorrelaciones cercanas a cero.	Las autocorrelaciones caen de manera exponencial. El decaimiento será de forma sinoidal. El patrón exacto depende del signo y tamaño de $\theta_1, \theta_2, \dots, \theta_q$.

Tabla 3.7.1: Tabla que muestra las características teóricas de la ACF y PACF

La metodología Box & Jenkins para examinar la adecuación del modelo, se basa en las propiedades de la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) para modelos ARIMA(p,d,q) y modelos SARIMA, introducidos en la siguiente sección. Algunas propiedades generales que nos brindarán apoyo al identificar las estructuras del modelo ARIMA(p,d,q) son:

- No estacionariedad: La sucesión de autocorrelaciones muestrales (ACF) decae de manera lenta y la sucesión de autocorrelaciones muestrales parciales (PACF) tiene un valor muy grande (positivo o negativo) en el “lag” 1. Esto se puede eliminar al tomar el operador diferencia, es decir, $\nabla X_t, \nabla^2 X_t, \dots, \nabla^d X_t$.
- No estacionariedad estacional: La sucesión de autocorrelaciones muestrales (ACF) es cero a excepción del “lag” $1_s, 2_s, 3_s, 4_s, \dots$ y decae de manera muy lenta. Esto puede arreglarse tomando el operador diferencia para estacionalidad, es decir, $\nabla_s X_t$.

- Comportamiento autorregresivo: La sucesión de autocorrelaciones muestrales parciales (PACF) es distinta de cero para el “lag” k , con $k = 1, 2, \dots, p$, y cero en adelante. Ajustar un modelo AR(p).
- Comportamiento autorregresivo estacional: La sucesión de autocorrelaciones muestrales parciales (PACF) es cero excepto en el “lag” m , con $m = 1_s, 2_s, 3_s, \dots, P_s$, y cero para $m > P_s$. Ajustar un modelo AR(p) que incluya estacionalidad.
- Comportamiento de promedios móviles: La sucesión de autocorrelaciones muestrales (ACF) es distinta de cero para el “lag” k , con $k = 1, 2, \dots, q$, y cero en adelante. Ajustar un modelo MA(q).
- Comportamiento de promedios móviles estacional: La sucesión de autocorrelaciones muestrales (ACF) es cero excepto en el “lag” M , con $M = 1_s, 2_s, 3_s, \dots, Q_s$, y cero para $M > Q_s$. Ajustar un modelo MA(q) que incluya estacionalidad.

Ahora bien, para identificar el orden de diferenciación en modelos ARIMA(p, d, q) hemos visto el método propuesto por Guerrero (2003) y hemos dicho que debe ser acompañado por las reglas que enumeramos a continuación:

1. Si la serie posee autocorrelaciones positivas en un alto número de retrasos, entonces, probablemente necesitará un alto número de diferencias.

Diferenciar tiende a introducir correlación negativa. Si la serie muestra inicialmente una fuerte autocorrelación positiva, entonces, aplicar el operador ∇X_t , reduce la autocorrelación e incluso puede llevar al “lag” 1 a tomar un valor negativo. Aplicar el operador $\nabla^2 X_t$, enviaría al “lag” 1 a una dirección mucho más negativa.

Si la autocorrelación correspondiente al “lag” 1 es cero o inclusive negativa, no hay necesidad de aplicar el operador diferencia. Como advertencia, debemos resistirnos de diferenciar si no vemos un patrón en las autocorrelaciones. Es más, si la autocorrelación correspondiente al “lag” 1 es más negativa que el valor -0.5, esto querrá decir que la serie ha sido diferenciada de más.

2. Si la autocorrelación correspondiente al “lag” 1 es cero o inclusive negativa o las autocorrelaciones son pequeñas y sin patrón alguno entonces la serie no necesita más diferencias.
3. El orden óptimo de diferenciación es aquel en el cual, el valor de la desviación estándar sea el más pequeño.
4. Un modelo que no necesite la aplicación del operador diferencia asume que la serie es estacionaria. Un modelo cuyo orden de diferenciación sea 1, implica la existencia de tendencia lineal. Un modelo cuyo orden de diferenciación sea 2 asume que la serie original posee una tendencia que varía a lo largo del tiempo.
5. Un modelo que no necesite la aplicación del operador diferencia normalmente incluye un término constante (el cual representa la media de la serie). Un modelo cuyo orden de diferenciación sea 2 no incluye el término constante. Un modelo cuyo orden de diferenciación sea 1 debe incluir un término constante si la serie tiene una tendencia promedio distinta de cero.

Esta serie de reglas pueden ser utilizadas en conjunto con el operador ∇^j para eliminar el tipo de tendencia existente en un conjunto de datos (lineal, cuadrática, etc).

Un modelo particular de interés es el modelo ARIMA(0,1,1) o mejor conocido como IMA(1,1). Dicho modelo se utiliza en series de tiempo económicas. A partir del modelo anterior se construye el método de pronóstico llamada método de promedios móviles con pesos exponenciales, EWMA por sus siglas en inglés.

Dicho modelo se expresa de la siguiente forma:

$$X_t = X_{t-1} + \epsilon_t - \lambda\epsilon_{t-1} \quad (3.7.5)$$

con $|\lambda| < 1$, $t = 1, 2, \dots$ y $X_0 = 0$. Ahora bien, si denotamos:

$$Y_t = \epsilon_t - \lambda\epsilon_{t-1} \quad (3.7.6)$$

entonces podemos escribir la **ec. 3.7.5** como:

$$X_t = X_{t-1} + Y_t \quad (3.7.7)$$

Debido a que $|\lambda| < 1$, el proceso $\{Y_t\}$ es invertible, con representación:

$$Y_t = \sum_{j=1}^{\infty} \lambda^j Y_{t-j} + \epsilon_t \quad (3.7.8)$$

Sustituyendo $Y_t = X_t - X_{t-1}$ obtenemos:

$$X_t = \sum_{j=1}^{\infty} (1 - \lambda)\lambda^{j-1} X_{t-j} + \epsilon_t \quad (3.7.9)$$

como aproximación para t grande. Usando dicha aproximación obtenemos el mejor predictor lineal a un paso dado por:

$$\begin{aligned} P_n X_{n+1} &= \sum_{j=1}^{\infty} (1 - \lambda)\lambda^{j-1} X_{n+1-j} \\ &= (1 - \lambda)X_n + \lambda \sum_{j=1}^{\infty} (1 - \lambda)\lambda^{j-1} X_{n-j} \end{aligned}$$

Pero $P_{n-1}X_n = \sum_{j=1}^{\infty} (1 - \lambda)\lambda^{j-1} X_{n-j}$, por lo que el mejor predictor lineal a un paso será:

$$P_n X_{n+1} = (1 - \lambda)X_n + \lambda P_{n-1} X_n \quad (3.7.10)$$

es decir, el mejor predictor lineal a un paso es una combinación de anteriores predictores y nuevas observaciones. Este predictor ya lo hemos visto antes, a saber, es el suavizamiento exponencial simple visto en el **capítulo 2**.

Ejemplo 3.7.1 La **Tabla 3.7.2** muestra los datos correspondientes a la serie del índice de precios al mayoreo de 210 artículos en la Ciudad de México desde Enero de 1971 hasta Diciembre de 1980.

	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
Ene	30.60	31.19	33.48	42.29	47.06	54.38	78.87	92.36	109.22	133.37
Feb	30.69	31.36	33.92	43.97	47.62	55.08	80.23	94.35	111.52	136.51
Mar	30.86	31.53	34.27	44.91	47.86	55.89	81.76	96.26	112.92	138.61
Abr	31.17	31.67	34.87	45.16	48.39	56.15	84.12	97.52	113.45	139.60
May	31.14	31.88	36.12	45.16	49.34	57.01	85.96	100.00	116.34	142.02
Jun	31.31	31.93	36.16	45.28	50.49	57.51	86.78	101.34	117.59	145.68
Jul	31.27	32.00	37.07	45.81	51.02	58.95	87.39	102.49	118.92	151.69
Ago	31.29	32.08	37.90	45.76	50.99	58.76	89.83	102.27	120.83	153.35
Sep	31.26	32.20	39.10	45.84	51.07	62.43	90.29	101.42	122.50	153.97
Oct	30.98	32.29	39.72	46.03	51.52	67.66	90.10	102.59	124.78	155.45
Nov	31.08	32.32	39.89	46.19	52.24	73.21	90.12	103.90	125.04	156.78
Dic	31.17	32.84	41.13	46.62	52.86	77.12	91.07	105.50	126.54	159.95

Tabla 3.7.2: Tabla que muestra el índice de precios al mayoreo en la Ciudad de México

La **Figura 3.7.1** muestra la serie correspondiente al índice de precios al mayoreo. Observamos de dicha figura que la serie posee una tendencia cuadrática lo cual nos ayudará a confirmar, junto con el método visto en esta sección, el número de diferencias por aplicar. La no estacionariedad del índice de precios al mayoreo es clara al observar la sucesión de autocorrelaciones muestrales (ACF) de la serie original, mostrada en la **Figura 3.7.2**. Observamos una lenta caída en las autocorrelaciones, así como un valor cercano a 1 de la primer autocorrelación.

Realizando el cálculo de la desviación estándar muestral de la serie obtenemos $S(0) = 38.64$, $S(1) = 1.31$, $S(2) = 1.17$ y $S(3) = 1.79$. Nuestra primer impresión será que el orden de diferencias, para volver estacionaria la serie, es $d = 2$ ya que es el valor que minimiza $S(j)$, para $j=0, \dots, 3$. Concluyamos ahora el análisis de esta serie utilizando R, es decir, ajustaremos el modelo ARIMA(p, d, q) óptimo:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 3/Ejercicios")
> IPM<-read.table("IPM.csv",sep="," ,header=FALSE)
> Serie<-ts(IPM, frequency = 12, start = c(1971, 1))
```

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
1971	30.60	30.69	30.86	31.17	31.14	31.31	31.27	31.29	31.26	30.98	31.08	31.17
1972	31.19	31.36	31.53	31.67	31.88	31.93	32.00	32.08	32.20	32.29	32.32	32.84
1973	33.48	33.92	34.27	34.87	36.12	36.16	37.07	37.90	39.10	39.72	39.89	41.13
1974	42.29	43.97	44.91	45.16	45.16	45.28	45.81	45.76	45.84	46.03	46.19	46.62
1975	47.06	47.62	47.86	48.39	49.34	50.49	51.02	50.99	51.07	51.52	52.24	52.86
1976	54.38	55.08	55.89	56.15	57.01	57.51	58.95	58.76	62.43	67.66	73.21	77.12
1977	78.87	80.23	81.76	84.12	85.96	86.78	87.39	89.83	90.29	90.10	90.12	91.07
1978	92.36	94.35	96.26	97.52	100.00	101.34	102.49	102.27	101.42	102.59	103.90	105.50
1979	109.22	111.52	112.92	113.45	116.34	117.59	118.92	120.83	122.50	124.78	125.04	126.54
1980	133.37	136.51	138.61	139.60	142.02	145.68	151.69	153.35	153.97	155.45	156.78	159.95

```
> library(stats)
> plot(Serie,main="Indice de Precios al Mayoreo en la Ciudad de México",ylab="IPM")
```

Indice de precios al mayoreo en la Ciudad de México

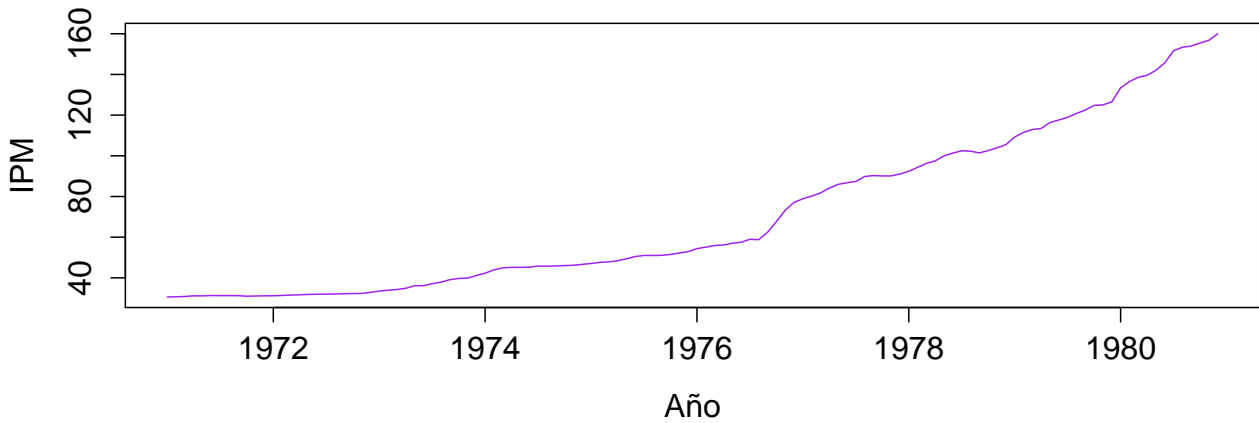


Figura 3.7.1: Figura que muestra el índice de precios al mayoreo en la Ciudad de México

ACF del proceso original

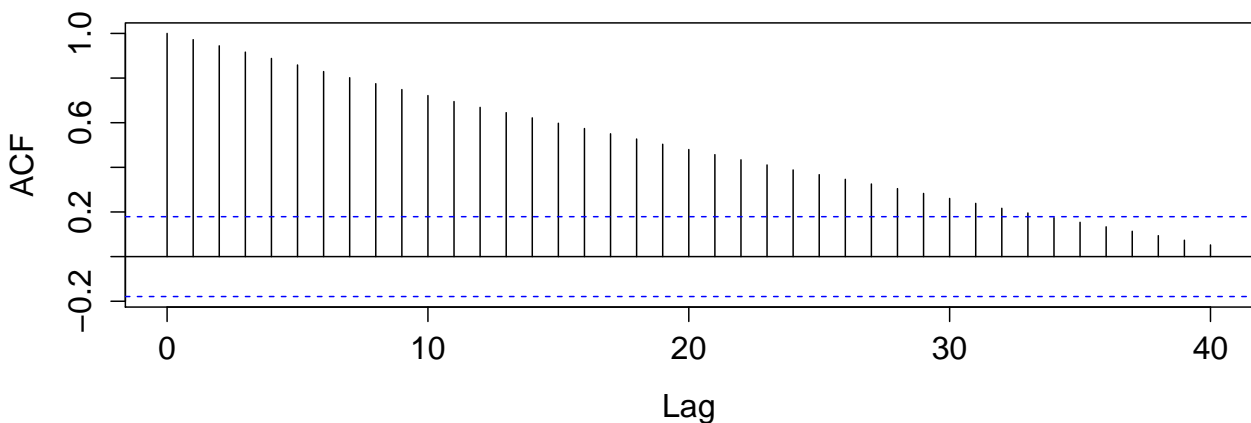


Figura 3.7.2: Figura ACF del índice de precios al mayoreo

Pasamos a diferenciar la serie original utilizando un “lag” o retraso de 2, con el fin de volverla estacionaria. Después observaremos el comportamiento de la sucesión de autocorrelaciones muestrales (ACF) y de la sucesión de autocorrelaciones muestrales parciales (PACF), las cuales se muestran en la **Figura 3.7.3** y en la **Figura 3.7.4** respectivamente, para, a partir de allí, ajustar un modelo $ARIMA(p,d,q)$ idóneo.

De la sucesión de autocorrelaciones muestrales (ACF) que aparece en la **Figura 3.7.3** se salen las autocorrelaciones correspondientes al “lag” 1, 2 y 3. Se sugiere ajustar $q = 3$ para la parte de promedios móviles. De la sucesión de autocorrelaciones muestrales parciales (PACF), mostrada en la **Figura 3.7.4**, se sugiere $p = 4$ para la parte autorregresiva pues se salen las autocorrelaciones correspondientes al retraso 2, 3 y 4 de las bandas de confianza. Nuestra opción óptima será la de ajustar un modelo $ARMA(4,3)$ a la serie diferenciada o bien, un modelo $ARIMA(4,2,3)$ a nuestros datos. Sin embargo, daremos otras opciones y veremos cómo, al comparar los criterios de información, nuestra mejor opción será en efecto aquella dada por un proceso $ARIMA(4,2,3)$. Nuestras dos alternativas a trabajar serán un modelo $ARIMA(4,2,1)$ y un modelo $ARIMA(4,2,2)$.

El código para obtener la serie diferenciada, así como la sucesión de autocorrelaciones muestrales (ACF) y la sucesión de autocorrelaciones muestrales parciales (PACF) será entonces:

```
> diferencialag2<-diff(Serie, lag = 1, differences = 2)
> Serieamodular2<-scale(diferencialag2, scale = FALSE)
> acf(ts(Serieamodular2,freq=1), lag.max=40, main="ACF del proceso")
> pacf(ts(Serieamodular2,freq=1), lag.max=40, main="PACF del proceso")
```

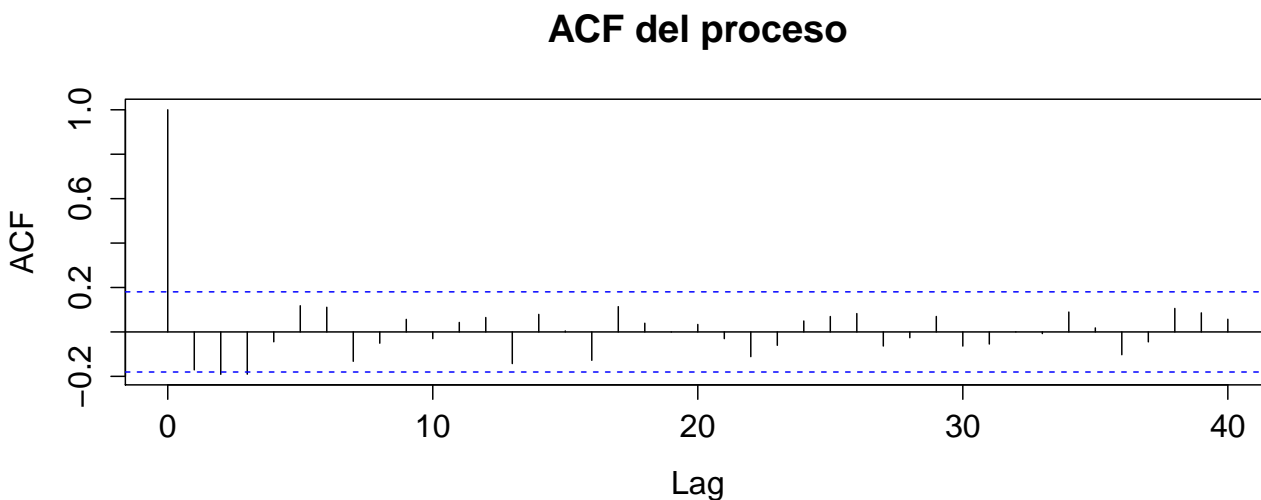


Figura 3.7.3: Figura ACF del índice de precios al mayoreo diferenciado

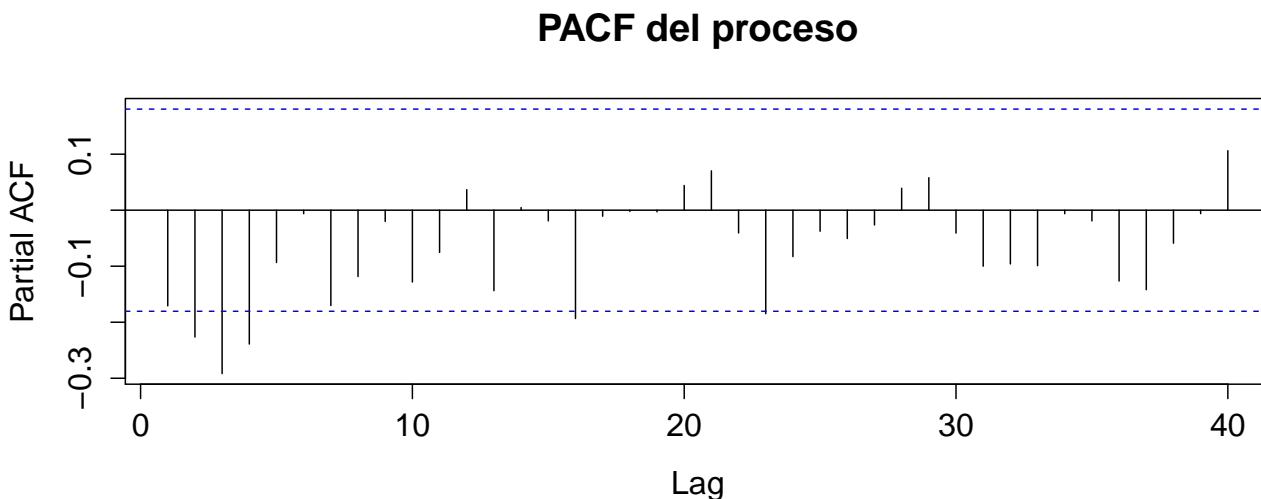


Figura 3.7.4: Figura PACF del índice de precios al mayoreo diferenciado

Si queremos ajustar el modelo $ARIMA(4,2,1)$ por máxima verosimilitud para la serie original se usa $order\ c(4,2,1)$ en la función $arima(\cdot)$, ya que el orden del modelo $AR(p)$ es 4, correspondiente al primer término en el paréntesis, la serie se diferencia 2 veces por lo cual el segundo lugar del paréntesis es 2 y el último corresponde a la parte $MA(q)$ del modelo. Además agregamos $include.mean=TRUE$ ya que el proceso no está centrado:

```
> ARIMA421<-arima(Serie,order=c(4,2,1), include.mean= TRUE)
```

```
Series: Serie
ARIMA(4,2,1)
```

```
Coefficients:
```

	ar1	ar2	ar3	ar4	ma1
	0.4899	-0.1057	-0.1008	0.0870	-0.9110
s.e.	0.1130	0.1055	0.1048	0.1062	0.0663

```
sigma^2 estimated as 1.024: log likelihood=-169.43
AIC=350.86 AICc=351.62 BIC=367.48
```

El código correspondiente para ajustar tanto el modelo ARIMA(4,2,2) como el modelo ARIMA(4,2,3) a nuestros datos originales viene dado en R por:

```
> ARIMA422<-arima(Serie,order=c(4,2,2), include.mean= TRUE)
```

```
Series: Serie
ARIMA(4,2,2)
```

```
Coefficients:
```

	ar1	ar2	ar3	ar4	ma1	ma2
	0.9303	-0.3071	-0.0544	0.1376	-1.3558	0.3949
s.e.	0.4429	0.2494	0.1395	0.0977	0.4392	0.4003

```
sigma^2 estimated as 1.018: log likelihood=-169.09
AIC=352.17 AICc=353.19 BIC=371.57
```

```
> ARIMA423<-arima(Serie,order=c(4,2,3), include.mean= TRUE)
```

```
Series: Serie
ARIMA(4,2,3)
```

```
Coefficients:
```

	ar1	ar2	ar3	ar4	ma1	ma2	ma3
	-0.1554	-0.7431	0.2677	-0.1674	-0.2431	0.4227	-0.8911
s.e.	0.1167	0.1158	0.1194	0.1112	0.0867	0.0719	0.0918

```
sigma^2 estimated as 0.9455: log likelihood=-166.56
AIC=349.12 AICc=350.44 BIC=371.28
```

Al observar los criterios de información seleccionamos aquel modelo con criterios más pequeños, a saber el que habíamos elegido en un principio, el modelo ARIMA(4,2,3).

*Para la verificación de los supuestos se usa la función `tsdiag(.)`, mostrada en la **Figura 3.7.5**, que proporciona gráficas para los residuos. En este caso lo estamos usando sobre el modelo que usa los datos originales. Observamos que las autocorrelaciones de los residuos no se salen de las bandas de confianza lo cual sugiere un ruido blanco. Además todos los *p*-values del estadístico de Ljung y Box están por arriba de un nivel de significancia de 0.05 hasta el lag 10 al menos, así que no rechazamos la hipótesis nula de que las autocorrelaciones conjuntamente son cero, con lo cual se confirma que el residuo es aproximadamente un ruido blanco.*

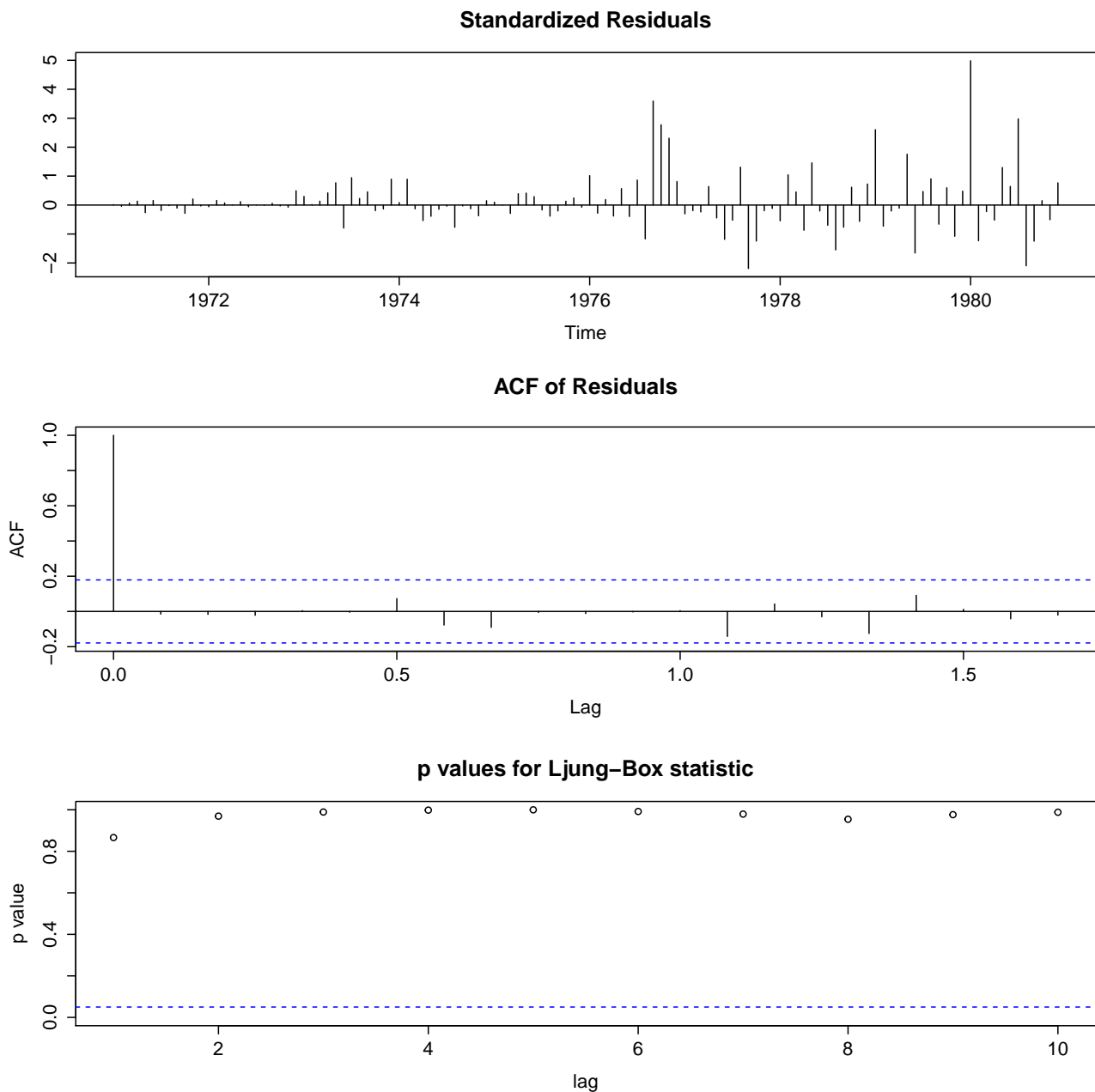


Figura 3.7.5: Figura que muestra la verificación de los supuestos del modelo

Finalmente en la **Figura 3.7.6** se presentan los residuos y su histograma, que sugieren una distribución más o menos normal y varianza distinta a una constante, respectivamente.

El código para la obtención de la función `tsdiag()`, histograma y gráfica de los residuos viene dado por:

```
> tsdiag(ARMA423)
> residuos<-ARMA423$residuals
> plot(residuos,xlab="Tiempo",ylab="Residuos",main="Gráfica de los residuos",pch=19)
> truehist(residuos,col=rainbow(4),border="white",
  main="Histograma de los residuos",xlab="Residuos",ylab="Frecuencia
  Relativa",col.main="black",col.axis="black",col.lab="black",prob=TRUE,ylim=c(0,0.4))
> curve(dnorm(x,mean(residuos),sd(residuos)),add=T,col="black")
```

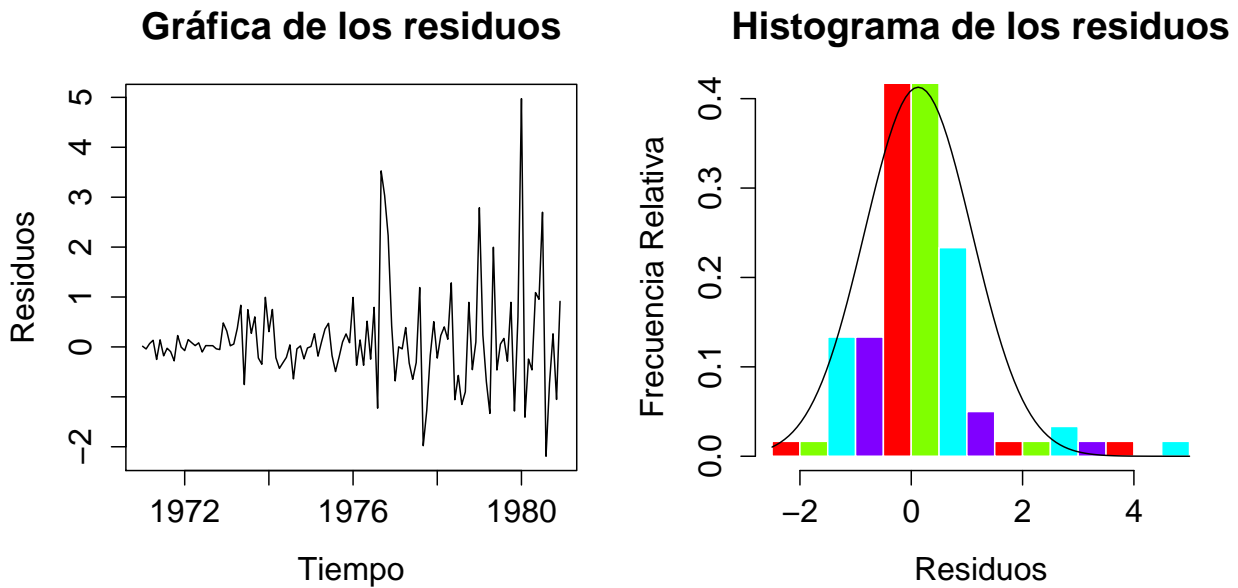



Figura 3.7.6: Figura que muestra a los residuos y su correspondiente histograma

Una prueba formal para la normalidad en los residuos es la de Lilliefors (Kolmogorov Smirnov). Observamos que el p -value de la prueba Lilliefors es $4.599e-07 \approx 0$, así que para un nivel de significancia del 5%, rechazamos la hipótesis de que los residuos se distribuyen normalmente:

```
> library(nortest)
> lillie.test(residuos)
```

Lilliefors (Kolmogorov-Smirnov) normality test

```
data:  residuos
D = 0.1508, p-value = 4.599e-07
```

Otra prueba formal para normalidad en los residuos es la prueba Shapiro-Wilk, con p -value $3.926e-09 \approx 0$, por lo que, para un nivel de significancia igual a 0.05, rechazamos la hipótesis nula y confirmamos que los residuos no se distribuyen normalmente. El código se muestra a continuación:

```
> shapiro.test(residuos)
```

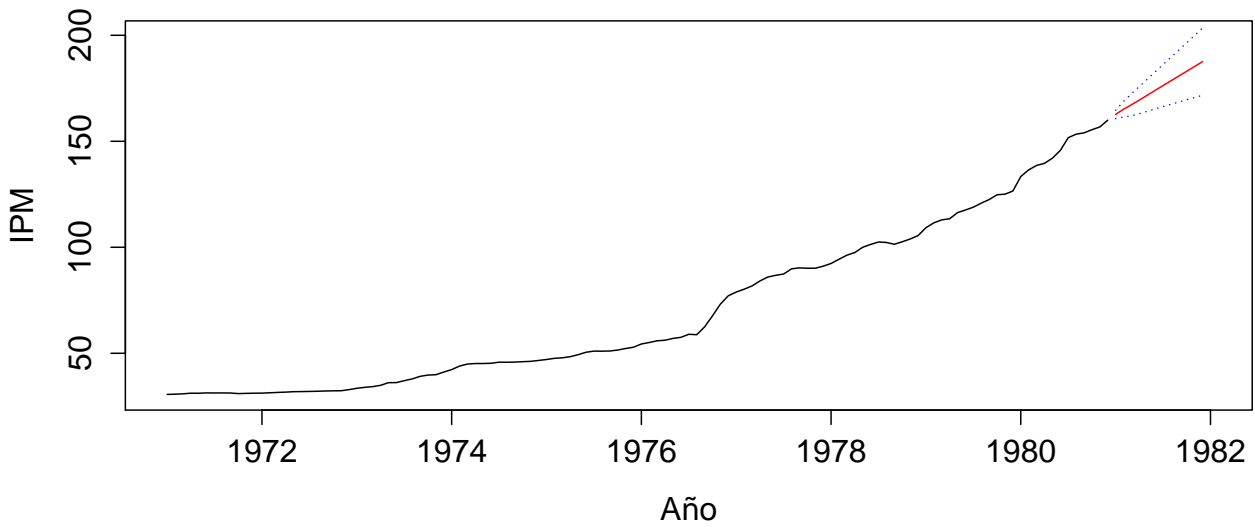
Shapiro-Wilk normality test

```
data:  residuos
W = 0.8634, p-value = 3.926e-09
```

\therefore Los datos cumplen los supuestos excepto el de normalidad y varianza constante en los residuos, aunque se pueden esperar pequeñas anomalías en el supuesto de normalidad. El modelo $ARIMA(4,2,3)$ es un buen modelo para explicar la serie del índice de precios al mayoreo si no tomamos en cuenta que la varianza no es constante y no hay normalidad en los residuos.

Podemos hacer predicciones sobre los siguientes 12 meses para nuestro modelo óptimo e incluso graficarlo. La **Figura 3.7.7** muestra un comparativo con las predicciones hechas con los modelos $ARIMA(4,2,1)$ y $ARIMA(4,2,3)$ con las bandas de confianza pintadas en color azul. Se escogieron estos dos modelos ya que poseen los criterios de información más pequeños y son claros candidatos a ajustar el índice de precios al mayoreo. Observamos que no existe una gran diferencia en las predicciones, así que cualquiera de los dos modelos servirá para nuestros propósitos aunque ya conocida la teoría, sabemos qué modelo elegir si tenemos que quedarnos con uno solo.

Predicción del IPM en la Ciudad de México



Predicción del IPM en la Ciudad de México

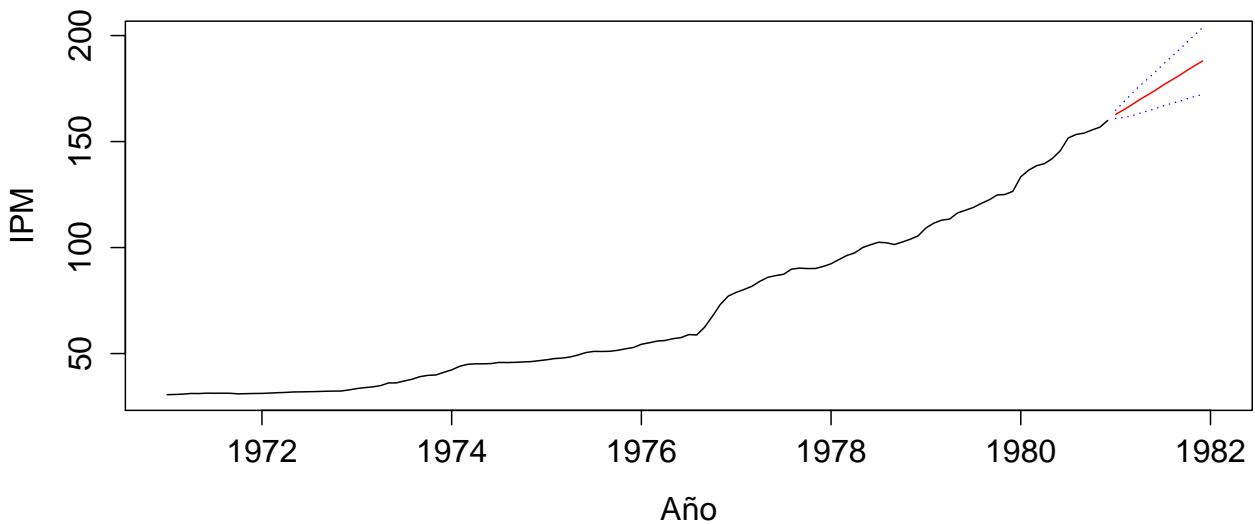


Figura 3.7.7: Pronóstico para el índice de precios al mayoreo en la Ciudad de México

El código para las predicciones está dado por:

```
> ARMA421.pred<-predict(ARMA421,n.ahead=12)
> plot(Serie, xlim=c(1971,1982),ylim=c(30,200),xlab="Año",ylab="IPM",
      main="Predicción del IPM en la Ciudad de México")
> lines(ARMA421.pred$pred,col="red")
> lines(ARMA421.pred$pred+1.96*ARMA421.pred$se,col="blue",lty=3)
> lines(ARMA421.pred$pred-1.96*ARMA421.pred$se,col="blue",lty=3)

> ARMA423.pred<-predict(ARMA423,n.ahead=12)
> plot(Serie, xlim=c(1971,1982),ylim=c(30,200),xlab="Año",ylab="IPM",
      main="Predicción del IPM en la Ciudad de México")
> lines(ARMA423.pred$pred,col="red")
> lines(ARMA423.pred$pred+1.96*ARMA423.pred$se,col="blue",lty=3)
> lines(ARMA423.pred$pred-1.96*ARMA423.pred$se,col="blue",lty=3)
```

Una de las opciones para estabilizar la varianza en los residuos (y que se aplica casi siempre a las series económicas) es tomar el logaritmo natural de la serie original y con esta nueva serie realizar un análisis similar al hecho en esta sección. El mejor modelo para explicar a la nueva serie será ahora un proceso $ARIMA(3,2,3)$. Habiendo realizado las predicciones requeridas, debemos regresar a la serie original aplicando la función exponencial a dichas predicciones. La gráfica de los residuos de la nueva serie e histograma de los mismos se muestran en la **Figura 3.7.8**. Observamos una mejor estabilidad en la varianza, el histograma muestra aparente normalidad en los residuos pero al aplicar la prueba Lilliefors concluimos que los residuos nuevamente poseen distribución distinta a la normal ya que el p-value es menor que $\alpha = 5\%$:

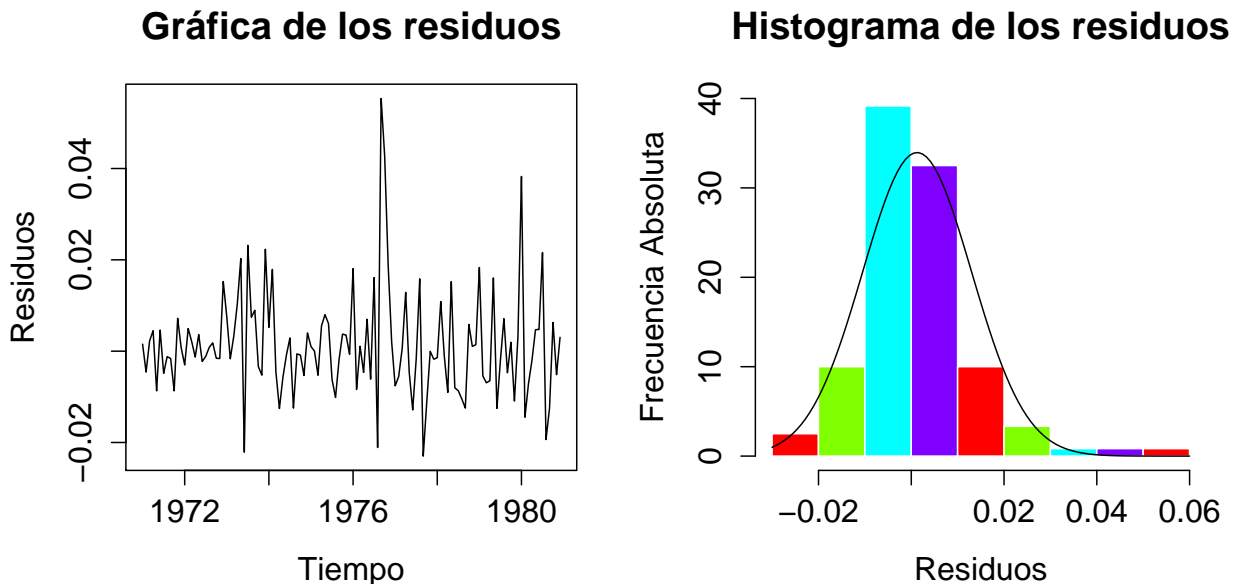


Figura 3.7.8: Figura que muestra los residuos y su correspondiente histograma (logaritmo)

```
> lillie.test(residuos)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: residuos

D = 0.1227, p-value = 0.0001346

La idea de este escrito es y será siempre, crear una conexión entre la parte teórica y la parte práctica proporcionada por los paquetes estadísticos. Hemos visto, en esta sección, el desarrollo, paso por paso, de la metodología ARIMA. Es decir, partimos desde la identificación de la estacionariedad de la serie hasta la comprobación de los supuestos del modelo. La estacionariedad de la serie se detectó a través del método propuesto por Guerrero (2003), es decir, dicho método nos brinda el número óptimo de diferencias por aplicar para volver a la serie estacionaria.

Ahora bien, en R, una forma alternativa y que a su vez debe ser combinada con los métodos propuestos anteriormente, es realizar la prueba Dickey-Fuller aumentada (la cual recordemos tiene como hipótesis nula la no estacionariedad de la serie) o bien, la prueba KPSS (la cual asume como hipótesis nula la estacionariedad de la serie) para de esta forma corroborar si la serie en cuestión es estacionaria o no. La librería necesaria para correr estas pruebas en R es “fpp: Data for forecasting: principles and practice”. Además utilizaremos el comando `ndiffs()`, el cual nos proporciona el número de diferencias por aplicar para volver a una serie estacionaria. El código en R viene dado por:

```
> kpss.test(Serie)

KPSS Test for Level Stationarity

data: Serie
KPSS Level = 3.8104, Truncation lag parameter = 2, p-value = 0.01

> adf.test(Serie,alternative="stationary")

Augmented Dickey-Fuller Test

data: Serie
Dickey-Fuller = -0.3732, Lag order = 4, p-value = 0.986
alternative hypothesis: stationary

> ndiffs(Serie,alpha=0.05,test=c("kpss"),max.d=2)

[1] 2
```

Observemos para la prueba KPSS, un p-value menor al 5% por lo que se rechaza la hipótesis nula y concluimos que la serie no es estacionaria. A la misma conclusión se llega al tomar la prueba Dickey-Fuller aumentada ya que para esta, no se rechaza la hipótesis nula de no estacionariedad en nuestra serie.

Además, observamos que el número de diferencias necesarias para volver a nuestra serie estacionaria es 2. Dicho número de diferencias coincide con el método propuesto por Guerrero (2003) y expuesto en esta sección. Entonces, tanto la parte teórica como los paquetes estadísticos nos han llevado a la misma conclusión respecto a la estacionariedad de la serie, lo cual siempre será deseable.

3.8. Pronósticos con modelos ARIMA

Así como en la **Sección 3.5.3** estudiamos las técnicas de pronóstico a τ pasos para modelos ARMA(p,q), en esta sección se realizará un estudio similar para modelos ARIMA(p,d,q). Supongamos a $\{W_t\}$ serie estacionaria con media cero, obtenida a partir de $W_t = \nabla^d X_t$. Pensemos por un momento que deseamos pronosticar a la observación $W_{t+\tau}$ a partir de toda la información existente. Vamos a distinguir a $\tilde{W}_t(\tau)$, un pronóstico cualquiera obtenido como combinación lineal de los valores de la serie $\{W_t\}$, esto es, W_t es combinación lineal de $\{X_1, \dots, X_t\}$ y de los residuos $\{\epsilon_t\} \forall t$, y al pronóstico óptimo denotado como $\hat{W}_t(\tau)$. De hecho, al hablar del pronóstico óptimo $\hat{W}_t(\tau)$, nos referimos a aquel pronóstico con menor error cuadrático medio, dado el conjunto $\{X_1, \dots, X_t\}$. $\hat{W}_t(\tau)$ deberá satisfacer:

$$\mathbb{E}_t[W_{t+\tau} - \hat{W}_t(\tau)]^2 = \min_{\tilde{W}_t(\tau)} \mathbb{E}_t[W_{t+\tau} - \tilde{W}_t(\tau)]^2 \quad (3.8.1)$$

donde el operador \mathbb{E} denota a la esperanza condicional dada toda la información a tiempo t, es decir, $\mathbb{E}_t[W_{t+\tau} - \tilde{W}_t(\tau)]^2 = \mathbb{E}_t\{[W_{t+\tau} - \tilde{W}_t(\tau)]^2 | X_t, X_{t-1}, \dots, X_1\}$. Es obvio que $\tilde{W}_t(\tau)$ puede escribirse en términos de los residuos $\{\epsilon_t\}$, es decir:

$$\tilde{W}_t(\tau) = \sum_{j=\tau}^{\infty} M_j \epsilon_{t+\tau-j} \quad (3.8.2)$$

La **ec. 3.8.2** implica escribir al proceso $\{\tilde{W}_t(\tau)\}$ en su forma causal. Nuestro único problema será el de obtener los valores $\{M_j\}$ de tal manera que se cumpla la **ec. 3.8.1**. Supondremos además que el conjunto $\{M_j\}$ converge absolutamente. Escribiremos a la observación $W_{t+\tau}$ como:

$$W_{t+\tau} = - \sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j} - \sum_{j=\tau}^{\infty} \psi_j \epsilon_{t+\tau-j} \quad (3.8.3)$$

es decir, dividimos a $W_{t+\tau}$ en dos sumas, una con información conocida, primera suma de la **ec. 3.8.3**, y la otra con información desconocida, las dos al tiempo t. A partir de la **ec. 3.8.2** y la **ec. 3.8.3** obtenemos:

$$W_{t+\tau} - \tilde{W}_t(\tau) = - \sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j} - \sum_{j=\tau}^{\infty} (\psi_j + M_j) \epsilon_{t+\tau-j} \quad (3.8.4)$$

y ya que $\mathbb{E}(\epsilon_{t+i} \epsilon_{t+j}) = 0$ para $i \neq j$ tenemos:

$$\mathbb{E}_t(W_{t+\tau} - \tilde{W}_t(\tau))^2 = \sum_{j=0}^{\tau-1} \psi_j^2 \sigma_\epsilon^2 + \sum_{j=\tau}^{\infty} (\psi_j + M_j)^2 \sigma_\epsilon^2 \quad (3.8.5)$$

Para pasar de la **ec. 3.8.4** a la **ec. 3.8.5** hemos elevado al cuadrado una suma finita y una con términos infinitos sin tomar en cuenta los productos cruzados ya que la secuencia $\{\epsilon_t\}$ no está correlacionada.

Además, hemos usado la linealidad del operador esperanza sólo en la primer suma de la **ec. 3.8.5**. Para la segunda parte hemos supuesto causalidad y convergencia absoluta en la serie $\{M_j\}$. Ahora bien, el mínimo de la **ec. 3.8.5** se obtiene haciendo $M_j = -\psi_j$ para $j = \tau, \tau+1, \dots$, por lo cual obtenemos:

$$\mathbb{E}_t(W_{t+\tau} - \hat{W}_t(\tau))^2 = \sum_{j=0}^{\tau-1} \psi_j^2 \sigma_\epsilon^2 \quad (3.8.6)$$

con $\hat{W}_t(\tau)$ dado, de manera similar a la **ec. 3.8.2**, por:

$$\hat{W}_t(\tau) = -\sum_{j=\tau}^{\infty} \psi_j \epsilon_{t+\tau-j} \quad (3.8.7)$$

Utilizando las propiedades de la esperanza condicional tenemos:

$$\mathbb{E}_t(W_{t+\tau} | X_1, X_2, \dots, X_t) = -\mathbb{E}_t\left(\sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j}\right) - \mathbb{E}_t\left(\sum_{j=\tau}^{\infty} \psi_j \epsilon_{t+\tau-j}\right) \quad (3.8.8)$$

La primer suma de la **ec. 3.8.8** es igual a 0 ya que estamos condicionando al conjunto $\{X_1, X_2, \dots, X_t\}$ y aplicando el operador esperanza al conjunto $\{\epsilon_{t+1}, \epsilon_{t+2}, \dots\}$ el cual es independiente del conjunto $\{\epsilon_t, \epsilon_{t-1}, \dots, \epsilon_1\}$; en notación, lo anterior puede expresarse como:

$$\mathbb{E}_t(\epsilon_{t+\tau-j}) = \begin{cases} \epsilon_{t+\tau-j} & \text{si } j \geq \tau \\ 0 & \text{si } j < \tau \end{cases}$$

Además, hemos hecho uso de algunas propiedades del operador esperanza al condicionar, a saber, $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X|Y)]$, primera suma de la **ec. 3.8.8**, y $\mathbb{E}(X|X)$ es una nueva variable aleatoria que corresponde a X , segunda suma de la **ec. 3.8.8**. Obtenemos por tanto:

$$\mathbb{E}_t(W_{t+\tau} | X_1, X_2, \dots, X_t) = -\sum_{j=\tau}^{\infty} \psi_j \epsilon_{t+\tau-j} \quad (3.8.9)$$

$\mathbb{E}_t(W_{t+\tau} | X_1, X_2, \dots, X_t)$ proporciona el pronóstico con error cuadrático medio mínimo. El error del pronóstico está dado por:

$$e_t(\tau) = W_{t+\tau} - \hat{W}_t(\tau) = -\sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j} \quad (3.8.10)$$

Por ejemplo, para $\tau = 1$ tenemos:

$$e_t(1) = W_{t+1} - \hat{W}_t(1) = \epsilon_{t+1} \quad (3.8.11)$$

La esperanza de $e_t(\tau)$ está dada por:

$$\mathbb{E}_t[e_t(\tau)] = -\mathbb{E}_t\left(\sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j}\right) = \sum_{j=0}^{\tau-1} \psi_j \mathbb{E}_t(\epsilon_{t+\tau-j}) = 0 \quad (3.8.12)$$

mientras que la varianza está dada por:

$$\text{Var}_t[e_t(\tau)] = \text{Var}_t\left(-\sum_{j=0}^{\tau-1} \psi_j \epsilon_{t+\tau-j}\right) = \sum_{j=0}^{\tau-1} \psi_j^2 \sigma_\epsilon^2 \quad (3.8.13)$$

La **ec. 3.8.13** representa el error cuadrático medio de la predicción. Es posible demostrar que, utilizando los pronósticos $\hat{W}_t(\tau)$, mientras mayor sea τ , mayor será la varianza del mismo. Esto se puede ver con mayor claridad en la **ec. 3.8.14**:

$$\text{Var}_t[e_t(\tau)] - \text{Var}_t[e_t(\tau-1)] = \psi_{\tau-1}^2 \sigma_\epsilon^2 \geq 0 \quad (3.8.14)$$

Ahora bien, tomemos el proceso ARIMA(p,d,q) especificado por $\phi(B)W_t = \theta(B)\epsilon_t$, con $W_t = \nabla^d X_t$; minimizamos la esperanza condicional de la **ec. 3.7.1** en la esperanza condicional de $W_{t+\tau}$ obteniendo:

$$\begin{aligned} \hat{W}_t(\tau) &= \mathbb{E}_t(W_{t+\tau}) \\ &= \mathbb{E}_t(\phi_1 W_{t+\tau-1} + \dots + \phi_p W_{t+\tau-p} + \epsilon_{t+\tau} - \theta_1 \epsilon_{t+\tau-1} - \dots - \theta_q \epsilon_{t+\tau-q}) \\ &= \phi_1 \mathbb{E}_t(W_{t+\tau-1}) + \dots + \phi_p \mathbb{E}_t(W_{t+\tau-p}) + \mathbb{E}_t(\epsilon_{t+\tau}) - \theta_1 \mathbb{E}_t(\epsilon_{t+\tau-1}) - \dots - \theta_q \mathbb{E}_t(\epsilon_{t+\tau-q}) \end{aligned} \quad (3.8.15)$$

donde

$$\mathbb{E}_t(W_{t+\tau-j}) = \begin{cases} W_{t+\tau-j} & \text{si } j \geq \tau \\ \hat{W}_t(\tau-j) & \text{si } j < \tau \end{cases}$$

y haciendo uso de la **ec. 3.8.11**, y de las propiedades del operador esperanza condicional aplicado al conjunto $\{\epsilon_t\}$ tenemos:

$$\mathbb{E}_t(\epsilon_{t+\tau-j}) = \begin{cases} W_{t+\tau-j} - \hat{W}_{t+\tau-j-1}(1) & \text{si } j \geq \tau \\ 0 & \text{si } j < \tau \end{cases}$$

Ejemplo 3.8.1 Recordando la condicionalidad al conjunto $\{X_1, X_2, \dots, X_t\}$ en los operadores esperanza $E_t(\cdot)$, sea $\{W_t\}$ el proceso $(1 + 0.8B)W_t = (1 - 0.4B)\epsilon_t$, donde $W_t = \nabla^d X_t$. Al despejar W_t obtenemos $W_t = -0.8W_{t-1} + \epsilon_t - 0.4\epsilon_{t-1}$. El pronóstico $\hat{W}_t(\tau)$ para $\tau = 1$ será:

$$\begin{aligned} \hat{W}_t(1) &= \mathbb{E}_t(W_{t+1}) \\ &= \mathbb{E}_t(-0.8W_t + \epsilon_{t+1} - 0.4\epsilon_t) \\ &= -0.8\mathbb{E}_t(W_t) + \mathbb{E}_t(\epsilon_{t+1}) - 0.4\mathbb{E}_t(\epsilon_t) \\ &= -0.8W_t - 0.4[W_t - \hat{W}_{t-1}(1)] \end{aligned}$$

Hemos aplicado $\mathbb{E}_t(W_{t+\tau-j})$ para $j \geq \tau$ y $j < \tau$. El pronóstico $\hat{W}_t(\tau)$ para $\tau = 2$ está dado por:

$$\begin{aligned}\hat{W}_t(2) &= \mathbb{E}_t(W_{t+2}) \\ &= \mathbb{E}_t(-0.8W_{t+1} + \epsilon_{t+2} - 0.4\epsilon_{t+1}) \\ &= -0.8\mathbb{E}_t(W_{t+1}) + \mathbb{E}_t(\epsilon_{t+2}) - 0.4\mathbb{E}_t(\epsilon_{t+1}) \\ &= -0.8[-0.8\mathbb{E}_t(W_t) + \mathbb{E}_t(\epsilon_{t+1}) - 0.4\mathbb{E}_t(\epsilon_t)] + \mathbb{E}_t(\epsilon_{t+2}) - 0.4\mathbb{E}_t(\epsilon_{t+1}) \\ &= -0.8\hat{W}_t(1)\end{aligned}$$

Utilizando la fórmula dada antes de empezar este ejemplo tenemos que los términos $\mathbb{E}_t(\epsilon_{t+2})$ y $\mathbb{E}_t(\epsilon_{t+1})$ son cero ya que $\tau - j = 2 > 0$ y $\tau - j = 1 > 0$ en el conjunto $\{\epsilon_t\}$. Es decir, ahora hemos aplicado $\mathbb{E}_t(\epsilon_{t+\tau-j})$ para $j \geq \tau$ y $j < \tau$. Los pronósticos se realizan de manera recursiva, y en general para $\tau \geq 2$, tenemos:

$$\hat{W}_t(\tau) = \mathbb{E}_t(W_{t+\tau}) = -0.8\hat{W}_t(\tau - 1)$$

En algún momento, al aplicar este procedimiento de manera recursiva, llegaremos a:

$$\hat{W}_1(1) = -0.8W_1 - 0.4[W_1 - \hat{W}_0(1)]$$

Claramente no tenemos información para obtener $\hat{W}_0(1)$. Para remediar dicho problema podemos usar $\hat{W}_0(1) = W_1$ y de ahí comenzar el procedimiento y obtener los pronósticos deseados.

Como bien sabemos, nosotros podemos aplicar una transformación $\{T(X_t)\}$ a nuestra serie original $\{X_t\}$, por ejemplo una transformación de Box-Cox para reducir la variabilidad en los datos. Claro, nuestra intención será volver a la nueva serie lo más estacionaria posible pero a veces, el realizar una transformación con ese fin, hace que se pierda el sentido de interpretación de los datos originales.

Ahora bien, en la realidad, las series con las que nos encontraremos distarán de ser estacionarias. Y claro, nosotros necesitamos una generalización para este tipo de situaciones. Dicha generalización está basada en el hecho de suponer que la no estacionariedad de la serie puede cancelarse al aplicarle una transformación (en este trabajo se maneja la transformación de Box-Cox), $T(X_t)$, a la serie original $\{X_t\}$.

El pronóstico óptimo de $T(X_{t+\tau})$ está dado por su esperanza condicional a tiempo t . Es decir, si el modelo para $T(X_t)$ puede escribirse como:

$$\omega(B)T(X_t) = \theta(B)\epsilon_t \quad \text{con } \omega(B) = \phi(B)\nabla^d \quad (3.8.16)$$

entonces el pronóstico, $\hat{T}(X_t)(\tau)$, estará dado por:

$$\begin{aligned}\hat{T}(X_t)(\tau) &= \mathbb{E}_t(T(X_{t+\tau})) \\ &= \mathbb{E}_t(\omega_1 T(X_{t+\tau-1}) + \dots + \omega_{p+d} T(X_{t+\tau-p-d}) + \epsilon_{t+\tau} - \theta_1 \epsilon_{t+\tau-1} - \dots - \theta_q \epsilon_{t+\tau-q}) \\ &= \omega_1 \mathbb{E}_t(T(X_{t+\tau-1})) + \dots + \omega_{p+d} \mathbb{E}_t(T(X_{t+\tau-p-d})) + \mathbb{E}_t(\epsilon_{t+\tau}) - \theta_1 \mathbb{E}_t(\epsilon_{t+\tau-1}) - \dots \\ &\quad - \theta_q \mathbb{E}_t(\epsilon_{t+\tau-q})\end{aligned}$$

donde

$$\mathbb{E}_t(T(X_{t+\tau-j})) = \begin{cases} T(X_{t+\tau-j}) & \text{si } j \geq \tau \\ \hat{T}(X_t)(\tau-j) & \text{si } j < \tau \end{cases}$$

y

$$\mathbb{E}_t(\epsilon_{t+\tau-j}) = \begin{cases} T(X_{t+\tau-j}) - \hat{T}(X_{t+\tau-j-1})(1) & \text{si } j \geq \tau \\ 0 & \text{si } j < \tau \end{cases}$$

Uno creará que después de realizar el pronóstico con la serie $\{T(X_t)\}$, la forma de regresar a la serie original sería simplemente tomar la transformación inversa, T^{-1} , y listo, obtenemos el pronóstico para la serie original. Pero muchas veces las propiedades del pronóstico $\hat{T}(X_t)(\tau)$ difieren de la optimalidad si la transformación no es lineal.

Guerrero (2003) propone un factor, al utilizar la transformación de Box-Cox, que permite corregir el sesgo que se produce al aplicar la transformación inversa. Dicho factor está dado por:

$$\hat{c}_{t,\lambda}(\tau) = \begin{cases} \left\{ \frac{1}{2} + \sqrt{1 - \frac{2(\lambda-1)}{\lambda} \left[\hat{T}(X_t)(\tau) \right]^{-2} \frac{\text{Var}_t[\epsilon_t(\tau)]}{2}} \right\}^{\frac{1}{\lambda}} & \text{si } \lambda < 1, \lambda \neq 0 \\ \exp \left\{ \frac{\text{Var}_t[\epsilon_t(\tau)]}{2} \right\} & \text{si } \lambda = 0 \end{cases}$$

A partir de este factor, el pronóstico aproximadamente insesgado de $X_{t+\tau}$ está dado por:

$$\begin{aligned} \hat{X}_t(\tau) &= \mathbb{E}_t\{T^{-1}[T(X_t)(\tau)]\} \\ &= T^{-1}[\hat{T}(X_t)(\tau)]\hat{c}_{t,\lambda}(\tau) \end{aligned} \quad (3.8.17)$$

3.9. Modelos SARIMA

La dependencia de información pasada tiende muchas veces a ocurrir en múltiplos de cierta estacionalidad de “lag” s . En esta sección, por tanto, realizaremos algunas modificaciones a los modelos ARIMA(p,d,q) para que tomen en cuenta comportamientos estacionales.

Como ejemplo tomemos la serie de ventas mensuales (kilolitros) de vino tinto por parte de viñedos australianos desde enero de 1980 a octubre de 1991. Dicha serie se muestra en la **Figura 3.9.1**. Observamos un patrón estacional con picos en los meses de julio de cada año, así como declives en los meses de enero de cada año.

Entonces, la característica fundamental de estos procesos estacionales es la similitud de comportamiento entre observaciones, por ejemplo, para el mismo mes pero en años consecutivos. Este hecho ocurre principalmente en series económicas, en las cuales, el retraso corresponde a múltiplos de $s = 12$.

O bien, si la serie está dada de manera semestral, las similitudes serán entre los mismos semestres pero en años consecutivos. La mayoría de la variabilidad de los procesos físicos, biológicos posee fluctuaciones estacionales de este tipo.

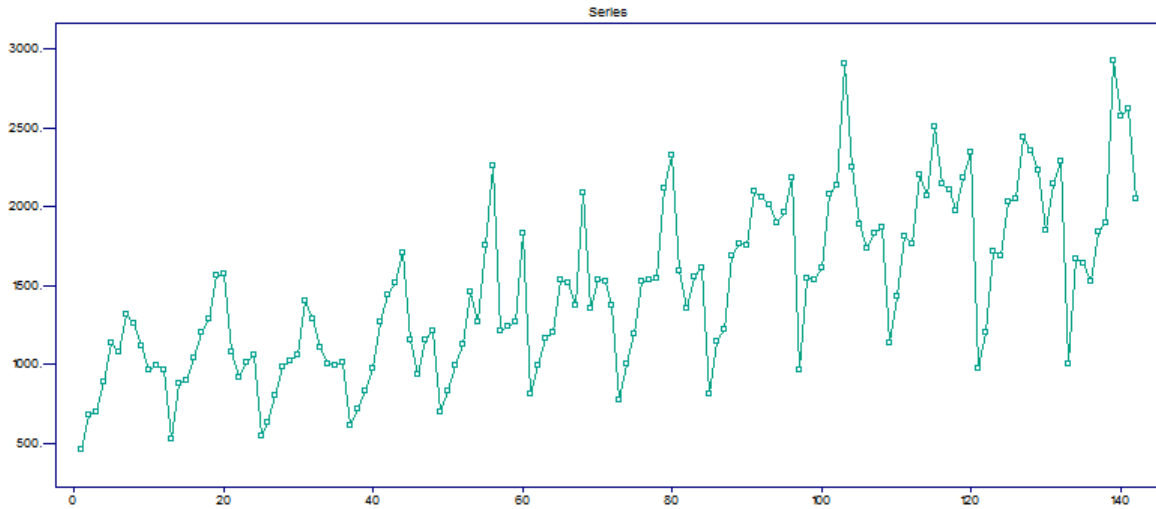


Figura 3.9.1: Figura de un proceso con estacionalidad

Para hacer frente a cambios estacionales debemos introducir polinomios tanto de promedios móviles como autorregresivos que tomen en cuenta dicho factor estacional. La parte meramente estacional de un proceso puede ser modelada por un proceso $ARMA(P, Q)_s$, el cual toma la forma de la **ec. 3.9.1**:

$$\Phi_P(B^s)X_t = \Theta_Q(B^s)\epsilon_t \quad (3.9.1)$$

y donde los operadores:

$$\begin{aligned} \Phi_P(B^s) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps} \\ \Theta_Q(B^s) &= 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs} \end{aligned}$$

representan el polinomio autorregresivo estacional y el polinomio de promedios móviles estacional de ordenes P y Q respectivamente con periodo estacional igual a s .

De la misma forma que definimos causalidad e invertibilidad para modelos $ARMA(p, q)$ podemos definirla para procesos $ARMA(P, Q)_s$, es decir, para que un modelo $ARMA(P, Q)_s$ sea causal e invertible, las raíces de los polinomios autorregresivos y de promedios móviles estacionales deben encontrarse fuera del círculo unitario.

Ahora bien, el proceso $ARMA(P, Q)_s$ solamente representa la parte estacional de un proceso mucho más amplio. Para modelar dicho proceso, es decir, aquel que conoceremos como $ARIMA(p, d, q) \times (P, D, Q)_s$ o bien $SARIMA(p, d, q) \times (P, D, Q)$, deberemos incluir la parte tendencial y estocástica de los modelos $ARMA(p, q)$ vistos a lo largo de este capítulo.

Pues bien, el modelo $ARIMA(p, d, q) \times (P, D, Q)_s$, modelo multiplicativo estacional autorregresivo e integrado de promedios móviles, está dado por:

$$\Phi_P(B^s)\phi(B)Y_t = \Theta_Q(B^s)\theta(B)\epsilon_t \quad (3.9.2)$$

con $Y_t = (1 - B)^d(1 - B^s)^D X_t$, $d, D \in \mathbb{Z}^+$, el número de diferencias a aplicar para la parte no-estacional y estacional del proceso $\{Y_t\}$, $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$ y polinomios:

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

La elección de P y Q en los procesos $ARIMA(p, d, q) \times (P, D, Q)_s$, se hace a partir de la sucesión de autocorrelaciones muestrales parciales (PACF) y de la sucesión de autocorrelaciones muestrales (ACF) pero ahora utilizando la estacionalidad.

Por ejemplo, si los datos son anuales, entonces $s = 12$ y elaboramos la sucesión de autocorrelaciones muestrales parciales (PACF) y la sucesión de autocorrelaciones muestrales (ACF) de manera anual y como en un modelo ARMA(p, q) usual utilizamos las bandas de confianza correspondientes para decidir el modelo que se va a elegir.

Esto es, vemos los valores de la sucesión de autocorrelaciones muestrales parciales (PACF) en $1_s, 2_s, 3_s, \dots$ y a partir de allí obtenemos P viendo a partir de qué valor todas las autocorrelaciones quedan dentro de las bandas. Por ejemplo si $P = 4$ y $s = 12$ debiera ocurrir que los valores de la sucesión de autocorrelaciones muestrales parciales (PACF) se salen de las bandas a los 12, 24, 36 y 48 meses y a partir de allí ya no se salen. Similarmente, para la sucesión de autocorrelaciones muestrales (ACF).

Ejemplo 3.9.1 Se realizará a continuación un estudio para modelar y pronosticar los datos del nivel de lluvia dados de forma trimestral en Port-Harcourt, Nigeria, a partir del primer trimestre del año 1971 hasta el último trimestre del año 2008. Los datos se muestran en la **Figura 3.9.2**. Observamos que la serie carece de tendencia por lo que sólo aplicaremos el operador $(\nabla_s)^D$ para filtrar la estacionalidad. El código en R está dado por:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 3/Ejercicios")
> Rainfall<-read.table("Rainfall.csv", sep=",", header=FALSE)
> Serie<-ts(Rainfall, frequency = 4, start = c(1971, 1))
> library(stats)
> plot(Serie, main="Nivel de lluvia en Port-Harcourt", xlab="Año", ylab="Nivel de lluvia",
col="blue")
```

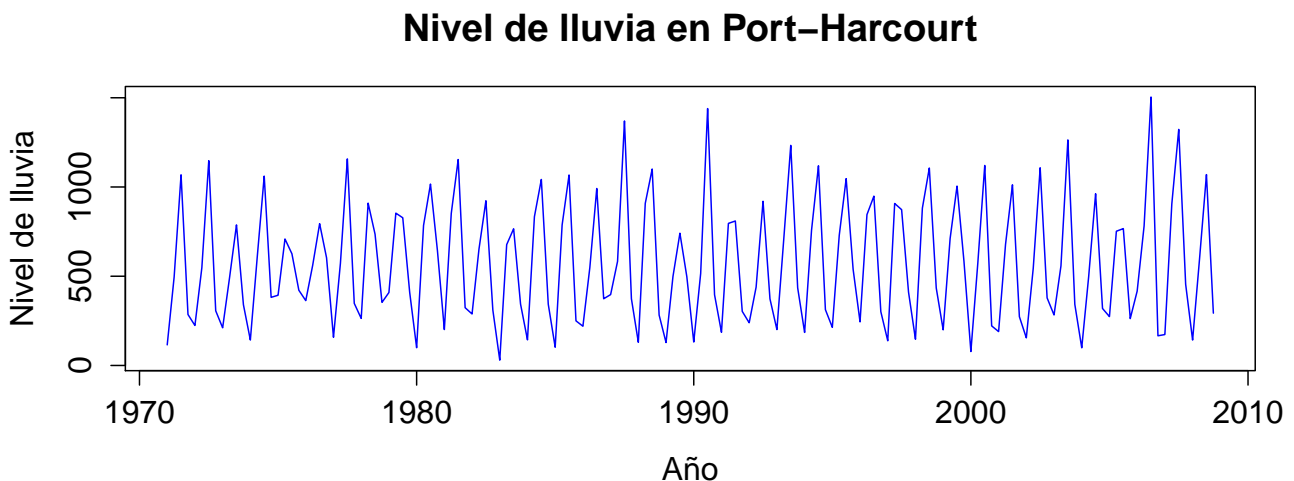
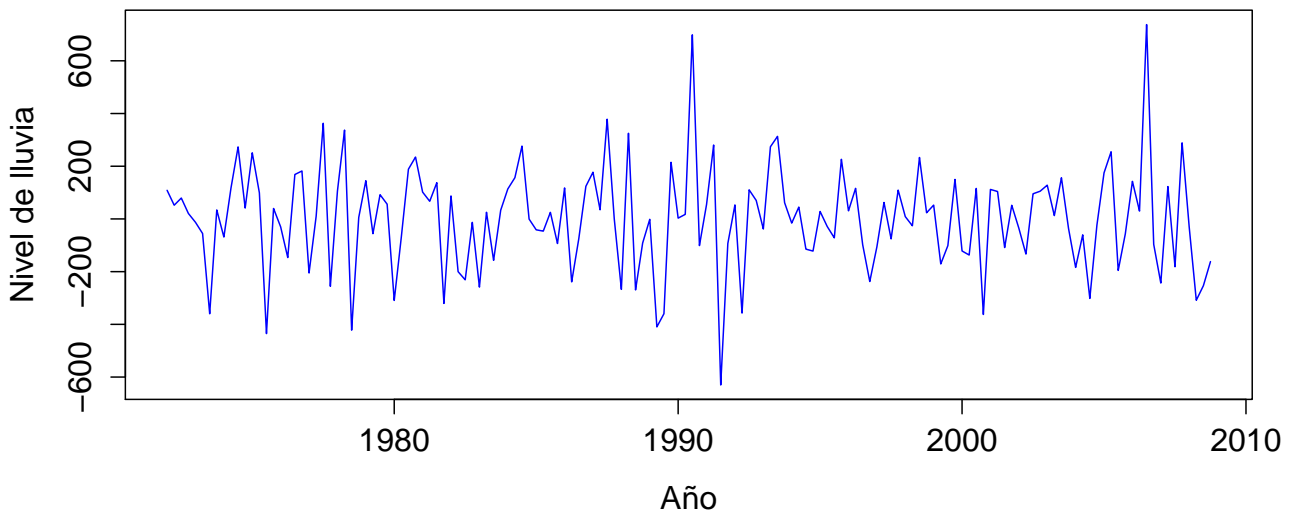


Figura 3.9.2: Figura que muestra el nivel de lluvia en Port-Harcourt

Ahora bien, como ya sabemos, la función *diff* sirve para obtener la serie diferenciada, “lag” indica el retraso, por ejemplo, para un modelo con estacionalidad *s* podríamos usar “lag” de *s* y “differences” indica el orden de las diferencias, es decir, el número de veces que se está diferenciando. Obtenemos la serie con “lag” de 4, ya que la serie es trimestral, para tratar de eliminar estacionalidad. No olvidemos restar la media de los datos. La **Figura 3.9.3** representa la serie sin estacionalidad y centrada:

```
> diferencialag4<-diff(Serie, lag = 4, differences = 1)
> plot(diferencialag4,main="Serie del nivel de lluvia sin estacionalidad",xlab="Año",
  ylab="Nivel de lluvia",col="blue")
> Serieamodelar<-scale(diferencialag4, scale =FALSE)
> plot(Serieamodelar,main="Serie del nivel de lluvia centrada",xlab="Año",
  ylab="Nivel de lluvia",col="blue")
```

Serie del nivel de lluvia sin estacionalidad



Serie del nivel de lluvia centrada

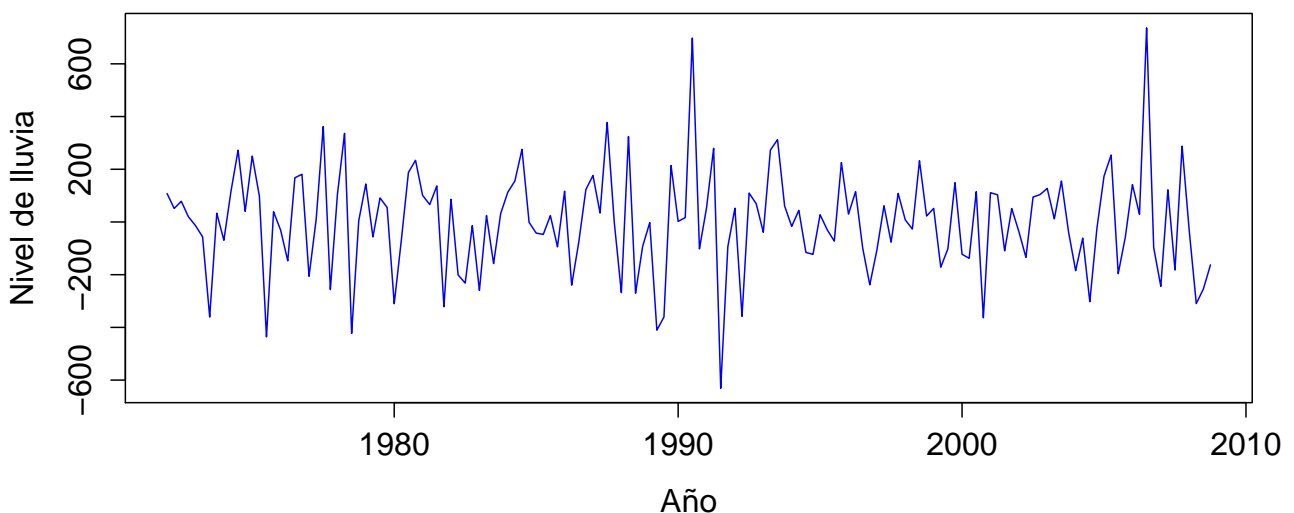


Figura 3.9.3: Figura que muestra el nivel de lluvia sin la presencia de estacionalidad

Ya que hemos logrado eliminar la estacionalidad y hemos centrado nuestra nueva serie, pasamos ahora a obtener la sucesión de autocorrelaciones muestrales (ACF) y a la sucesión de autocorrelaciones muestrales parciales (PACF), las cuales mostramos en la **Figura 3.9.4**. El código correspondiente viene dado por:

```
> acf(ts(Serieamodelar,freq=1), lag.max=20, main="ACF del proceso")
> pacf(ts(Serieamodelar,freq=1), lag.max=20, main="PACF del proceso")
```

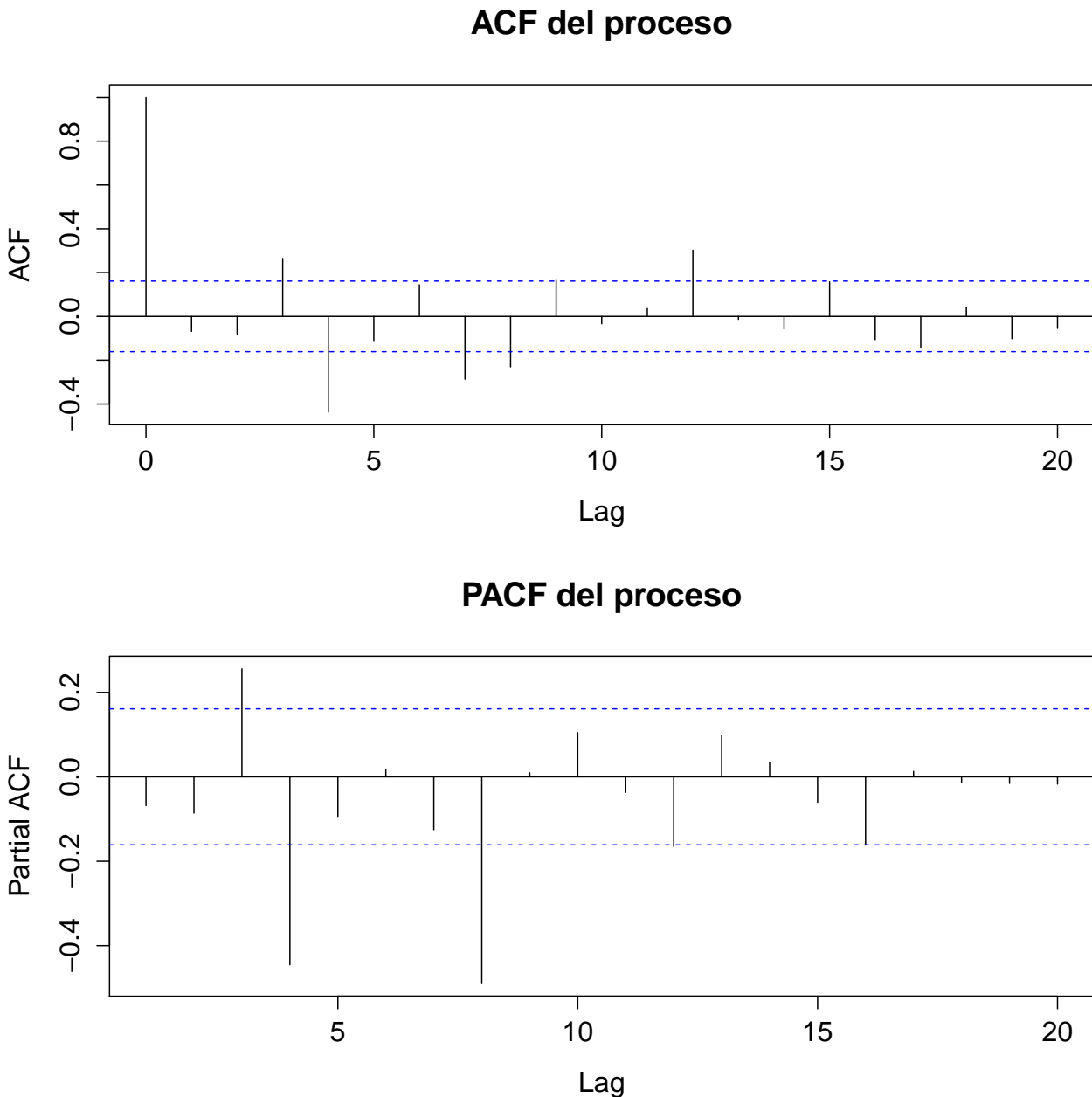


Figura 3.9.4: Figura ACF del nivel de lluvia en Port-Harcourt

De la sucesión de autocorrelaciones muestrales (ACF) se salen de las bandas de confianza las autocorrelaciones correspondientes al "lag" 4, 8 y 12 y de la sucesión de autocorrelaciones muestrales parciales (PACF) se salen las autocorrelaciones correspondientes al "lag" 4 y 8, confirmando con esto la estacionalidad con "lag" 4. Se modelará simplemente la parte estacional, de aquí que un modelo a escoger será el modelo $SARIMA(0,0,0)(2,1,0)$.

El código correspondiente está dado por:

```
> Modelolluvia<-arima(Serie,order=c(0,0,0), seasonal=c(2,1,0),include.mean= FALSE)
```

Call:

```
arima(x = Serie, order = c(0, 0, 0), seasonal = c(2, 1, 0), include.mean = FALSE)
```

Coefficients:

```
      sar1      sar2
-0.693  -0.5548
s.e.   0.069   0.0684
```

```
sigma^2 estimated as 21848:  log likelihood = -951.32,  aic = 1908.63
```

*Podemos hacer predicciones sobre los siguientes 4 años e incluso graficarlos. La **Figura 3.9.5** muestra dicha predicción. El código para la predicción será entonces:*

```
> Lluvia.pred<-predict(Modelolluvia,n.ahead=16)
```

	Qtr1	Qtr2	Qtr3	Qtr4
2009	298.6759	741.8320	1345.5735	245.0011
2010	207.4701	812.1175	1294.7300	368.0328
2011	183.9677	682.5567	1176.7361	309.4065
2012	250.8570	733.3424	1286.7112	281.7721

```
> plot(Serie,col="orange",xlim=c(1971,2014),xlab="Año",ylab="Nivel de lluvia",
      main="Predicción del nivel de lluvia en Port-Harcourt")
```

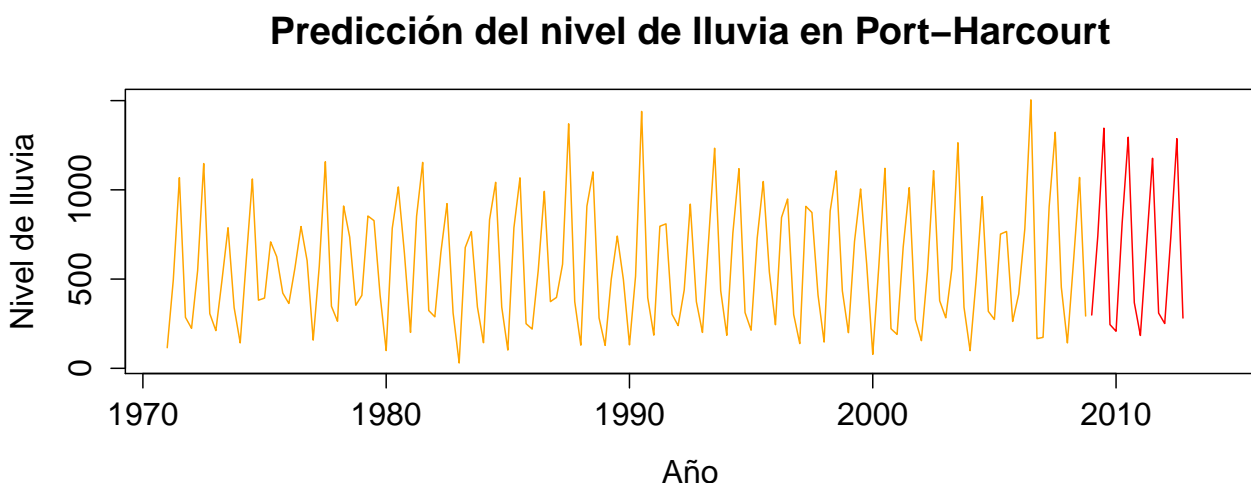


Figura 3.9.5: Predicción del nivel de lluvia en Port-Harcourt para los años 2009-2012

*Ahora pasemos a verificar los supuestos del modelo. Se usa la función `tsdiag`, mostrada en la **Figura 3.9.6**. En ella, observamos que las autocorrelaciones pertenecen a un ruido blanco ya que no se salen de las bandas de confianza. Además todos los *p*-values del estadístico de Ljung y Box están por arriba de un nivel de significancia de 0.05 hasta el “lag” 10 al menos, así que no rechazo la hipótesis nula de que las autocorrelaciones conjuntamente son cero.*

Para la normalidad en los residuos se presenta, en la **Figura 3.9.7**, el histograma, que sugiere una distribución aproximadamente normal. El código correspondiente de la verificación de los supuestos del modelo está dado por:

```
> tsdiag(Modelolluvia)
> par(mfrow=c(1,2))
> residuos<-Modelolluvia$residuals
> plot(residuos,xlab="Tiempo",ylab="Residuos",main="Gráfica de los residuos",pch=19)
> truehist(residuos,col=rainbow(4),border="white",
  main="Histograma de los residuos",xlab="Residuos",ylab="Frecuencia
  Relativa",col.main="black",col.axis="black",col.lab="black",prob=TRUE,ylim=c(0,0.003))
> curve(dnorm(x,mean(residuos),sd(residuos)),add=T,col="black")
```

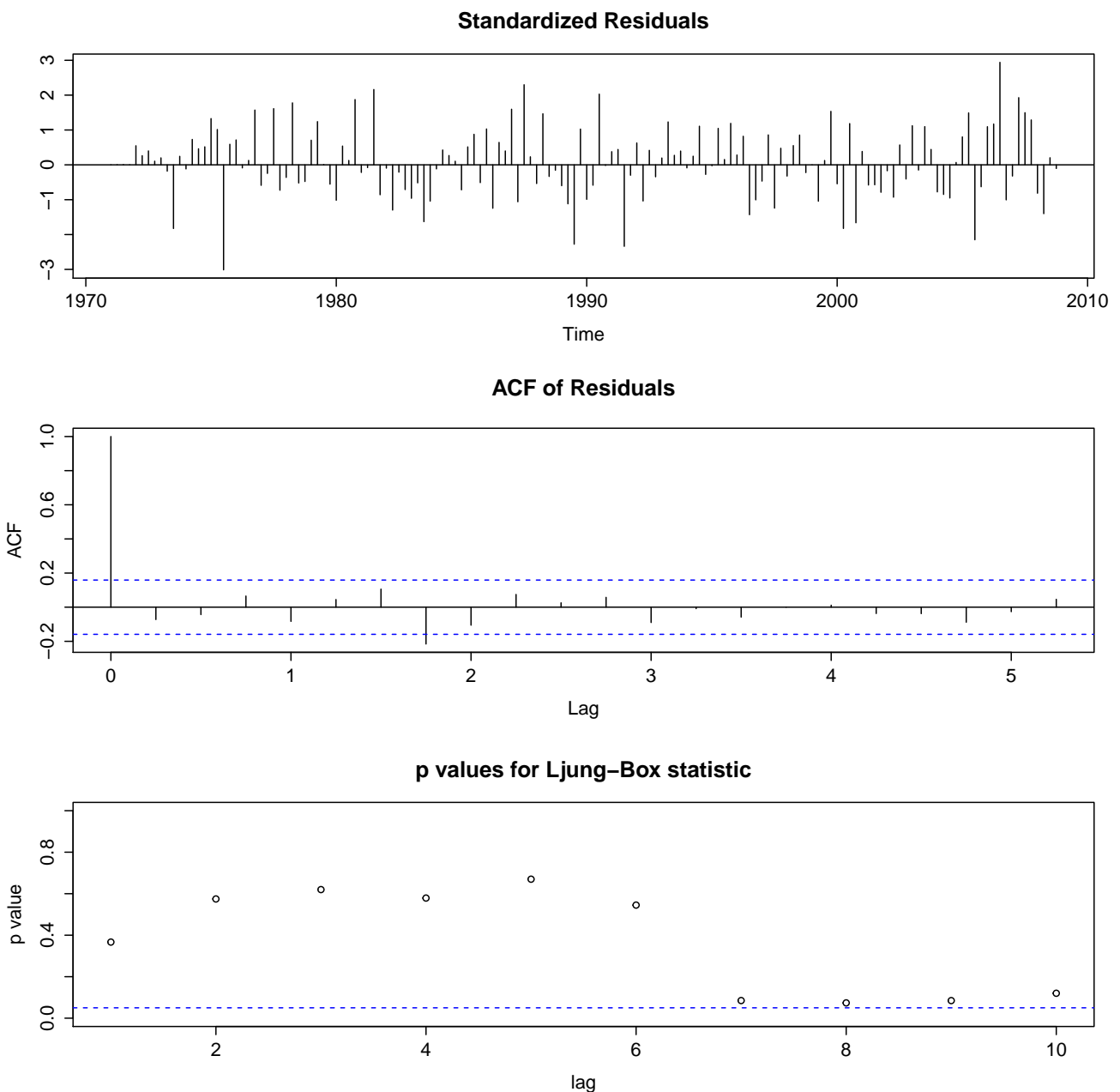


Figura 3.9.6: Figura que muestra la verificación de los supuestos del modelo

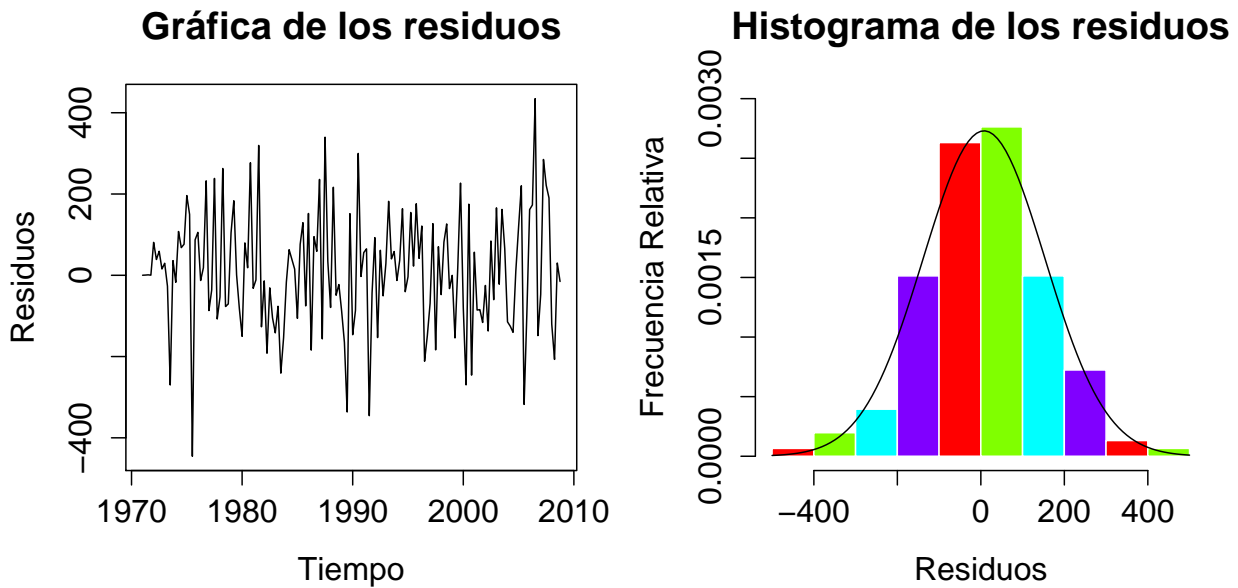


Figura 3.9.7: Figura que muestra los residuos e histograma

Como siempre, debemos corroborar el supuesto de normalidad con pruebas formales; en este caso utilizaremos la prueba de Lilliefors (Kolmogorov-Smirnov) y la prueba Shapiro-Wilk, las cuales están dadas en R por el siguiente código:

```
> library(nortest)
> lillie.test(residuos)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  residuos
D = 0.0385, p-value = 0.8412

> shapiro.test(residuos)

Shapiro-Wilk normality test

data:  residuos
W = 0.9959, p-value = 0.9489
```

El p-value de la prueba Lilliefors es 0.8412, mientras que el de la prueba Shapiro-Wilk es 0.9489, así que para un nivel de significancia de 0.05, no rechazamos la hipótesis de que los residuos se distribuyen normalmente.

Los datos cumplen los supuestos y el modelo $SARIMA(0,0,0)(2,1,0)$ es un buen modelo para explicar y pronosticar el nivel de lluvia en Port-Harcourt, Nigeria.

De nuevo, y solo como complemento al análisis exhaustivo y completo que se ha hecho de la serie en cuestión, se realizará en R, la prueba Dickey-Fuller aumentada y la prueba KPSS para corroborar nuestra conclusión de estacionariedad. Además, utilizaremos ahora el comando `nsdiffs(.)`, el cual nos proporciona el número de diferencias estacionales por aplicar para volver a una serie estacionaria alrededor de la media. La salida de este comando debería coincidir, la mayoría de las veces, con el número de diferencias propuesto por el investigador.

El código está dado por:

```
> kpss.test(Serie)

KPSS Test for Level Stationarity

data: Serie
KPSS Level = 0.095, Truncation lag parameter = 2, p-value = 0.1

> adf.test(Serie,alternative="stationary")

Augmented Dickey-Fuller Test

data: Serie
Dickey-Fuller = -5.1027, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary

> nsdiffs(Serie,m=frequency(Serie),test=c("ocsb"), max.D=2)

[1] 1
```

Observemos para la prueba KPSS, un p-value mayor al 5% por lo que no se rechaza la hipótesis nula y se concluye que la serie es estacionaria. A la misma conclusión se llega al tomar la prueba Dickey-Fuller aumentada ya que para esta, se rechaza la hipótesis nula de no estacionariedad en nuestra serie.

Además, observamos que el número de diferencias para eliminar estacionalidad es 1 (operador ∇_s), el cual coincide con lo propuesto anteriormente.

3.10. Modelos GARCH

A lo largo de este capítulo hemos recalcado la importancia de trabajar con series de tiempo estacionarias. En particular, los modelos ARMA(p,q) asumen varianza constante. Pero recientes problemas en finanzas, aquellos concernientes al estudio de la volatilidad de los rendimientos, han motivado a la creación de los llamados modelos **autorregresivos condicionales heterocedásticos** o **modelos ARCH**. De acuerdo con Engel (2001), estos modelos deben contar con la capacidad de pronosticar de manera eficiente la volatilidad además de tomar en cuenta las siguientes características:

- La volatilidad tiene tendencia a aparecer agrupada por periodos, es decir, la volatilidad puede ser alta durante un periodo y baja durante otro. A dicha característica la conoceremos como **aglomeración de la volatilidad**.
- La aglomeración de la volatilidad implica que a un periodo de alta volatilidad le seguirá otro de volatilidad normal y a un periodo de baja volatilidad le seguirá uno de volatilidad más alta. La volatilidad tiende a revertirse a la media, es decir, existe un nivel normal de volatilidad al cual ésta retorna eventualmente.
- La distribución de los rendimientos tiene colas pesadas y en general, exceso de curtosis.

- Evidentemente, nadie considera que los precios de los activos financieros evolucionan independientemente de los mercados alrededor de éstos y, por ello, cabría esperar que existan otras variables que contengan información relevante para la volatilidad de la serie.

Pensemos por un momento que x_t es el precio de una acción a tiempo t . Entonces, definimos al rendimiento y_t a tiempo t como:

$$y_t = \frac{x_t - x_{t-1}}{x_{t-1}} \quad (3.10.1)$$

Normalmente, el rendimiento y_t no tiene varianza condicional σ_t^2 constante (por varianza condicional nos referimos a que la varianza está condicionada a la información pasada) y además, como ya se mencionó anteriormente, periodos de alta volatilidad tienden a estar aglomerados.

El modelo ARCH más sencillo, es aquel que modela los rendimientos de la siguiente forma:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 \end{aligned} \quad (3.10.2)$$

Este modelo se conoce como **ARCH(1)**, donde $\{\epsilon_t\} \sim WN_G(0, 1)$ y con la condición de que α_1 no debe ser negativa ya que de serlo, σ_t^2 podría ser negativa. Observe como los modelos ARCH(1) solo dependen de la variabilidad de la serie y_t en el tiempo anterior, cuestión que no sucederá cuando veamos el caso general, los **modelos ARCH(m)**, los cuales dependerán de m tiempos de retraso.

Para la estimación de α_0 y α_1 recordemos que la distribución condicional de y_t dado y_{t-1} es Normal, es decir:

$$y_t | y_{t-1} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2) \quad (3.10.3)$$

Ahora bien, para estimar los parámetros del modelo ARCH(1) debemos obtener la verosimilitud condicional de nuestras observaciones y_2, y_3, \dots, y_n dado y_1 , es decir:

$$L(\alpha_0, \alpha_1 | y_1) = \prod_{t=2}^n f_{\alpha_0, \alpha_1}(y_t | y_{t-1}) \quad (3.10.4)$$

donde $f_{\alpha_0, \alpha_1}(y_t | y_{t-1})$ es la densidad especificada en la **ec. 3.10.3**. Entonces, después de tomar la log-verosimilitud, la función por ser minimizada será:

$$l(\alpha_0, \alpha_1) = \frac{1}{2} \sum_{t=2}^n \ln(\alpha_0 + \alpha_1 y_{t-1}^2) + \frac{1}{2} \sum_{t=2}^n \left(\frac{y_t^2}{\alpha_0 + \alpha_1 y_{t-1}^2} \right) \quad (3.10.5)$$

El cálculo de la estimación de los parámetros del modelo deberá realizarse por tanto, por métodos numéricos.

El modelo ARCH(1) puede extenderse a un modelo general, **ARCH(m)**, a saber, dado por:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \dots + \alpha_m y_{t-m}^2 \end{aligned} \quad (3.10.6)$$

Para la estimación de los parámetros del modelo ARCH(m), de nuevo se utiliza la verosimilitud condicional de nuestras observaciones $y_{m+1}, y_{m+2}, \dots, y_n$ dados y_1, y_2, \dots, y_m , es decir:

$$L(\alpha_0, \alpha_1, \dots, \alpha_m | y_1, y_2, \dots, y_m) = \prod_{t=m+1}^n f_{\alpha_0, \alpha_1, \dots, \alpha_m}(y_t | y_{t-1}, y_{t-2}, \dots, y_{t-m}) \quad (3.10.7)$$

donde

$$y_t | y_{t-1}, y_{t-2}, \dots, y_{t-m} \sim N(0, \alpha_0 + \alpha_1 y_{t-1}^2 + \alpha_2 y_{t-2}^2 + \dots + \alpha_m y_{t-m}^2) \quad (3.10.8)$$

Una generalización de los modelos ARCH son los **modelos GARCH**. Por ejemplo, un modelo GARCH(1,1) para los rendimientos está dado por:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \end{aligned} \quad (3.10.9)$$

es decir, depende de la variabilidad de la serie y_t y de la serie y_t en el tiempo anterior. Finalmente, el modelo GARCH(m,r) puede ser descrito como:

$$\begin{aligned} y_t &= \sigma_t \epsilon_t \\ \sigma_t^2 &= \alpha_0 + \sum_{j=1}^m \alpha_j y_{t-j}^2 + \sum_{j=1}^r \beta_j \sigma_{t-j}^2 \end{aligned} \quad (3.10.10)$$

La estimación por máxima verosimilitud condicional es similar a la expuesta para los modelos ARCH(m), donde la verosimilitud condicional dada por la **ec. 3.10.7** es el producto de densidades $N(0, \sigma_t^2)$ con σ_t^2 dada por la **ec. 3.10.10** y se condiciona sobre las primeras $\max(m, r)$ observaciones, con $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_r^2 = 0$. Una vez obtenidos los estimadores para los parámetros, el modelo se puede utilizar para realizar predicciones sobre la volatilidad a un paso, a saber:

$$\hat{\sigma}_{t+1}^2 = \hat{\alpha}_0 + \sum_{j=1}^m \hat{\alpha}_j y_{t+1-j}^2 + \sum_{j=1}^r \hat{\beta}_j \hat{\sigma}_{t+1-j}^2 \quad (3.10.11)$$

Como conclusión, una vez hecho el pronóstico para la volatilidad, $\hat{\sigma}_{t+1}^2$, podemos realizar el pronóstico de la observación \hat{y}_{t+1} con base en la **ec. 3.10.10** y una simulación a un paso del conjunto $\{\epsilon_t\}$.

Capítulo 4

Modelos de Regresión Dinámica

4.1. Introducción

En este capítulo introduciremos nociones básicas del modelo de espacio de estados y su uso en el análisis de series de tiempo. Veremos, en particular, los modelos de regresión dinámica, los cuales serán nuestro tema de interés en este capítulo. Comenzaremos con el modelo de espacio de estados para al final, abordar de manera extensiva el modelo de regresión dinámica. Para darnos una idea de las posibles aplicaciones de los modelos de espacio de estados consideremos la **Figura 4.1.1**, la cual representa el número de pasajeros mensuales de cierta aerolínea internacional desde el mes de enero de 1949 hasta el mes de diciembre de 1960.

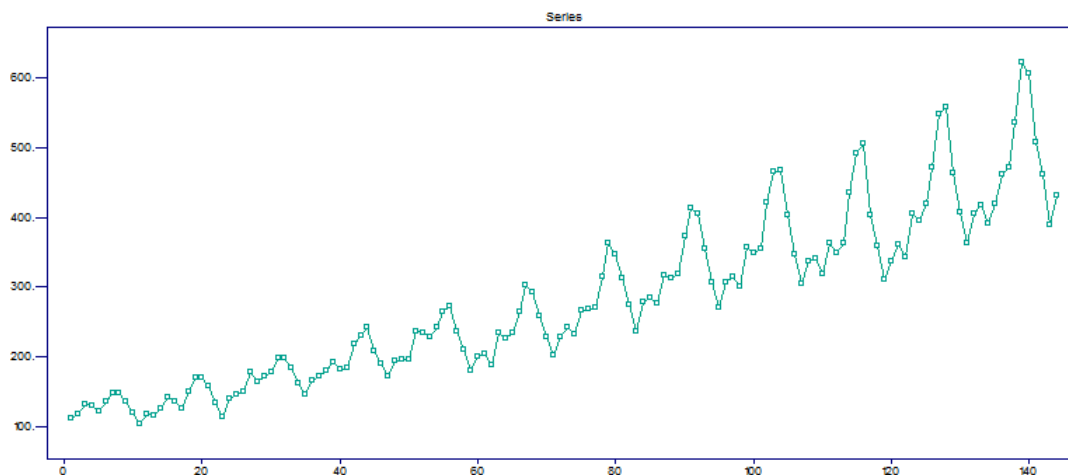


Figura 4.1.1: Figura que muestra el número de pasajeros de cierta aerolínea internacional

Esta serie posee tanto tendencia como estacionalidad y un pequeño incremento en la variabilidad de los datos alrededor de la media cada año. Podemos realizar, como en el capítulo anterior, un análisis exhaustivo de dicha serie obteniendo un modelo $SARIMA(1, 1, 1)(0, 1, 0)_{12}$. Como bien sabemos, el análisis de series de tiempo se basa en la posibilidad de encontrar regularidades en el comportamiento de una serie, para con ello, y manteniendo este supuesto a lo largo del tiempo, predecir el comportamiento futuro. Pero, ¿qué sucede con series como la mostrada en la **Figura 4.1.2**, la cual representa el consumo trimestral de gas en el Reino Unido, desde el año 1960 y hasta el año 1986? Observamos un cambio sustancial en la estacionalidad.

Consumo de gas británico

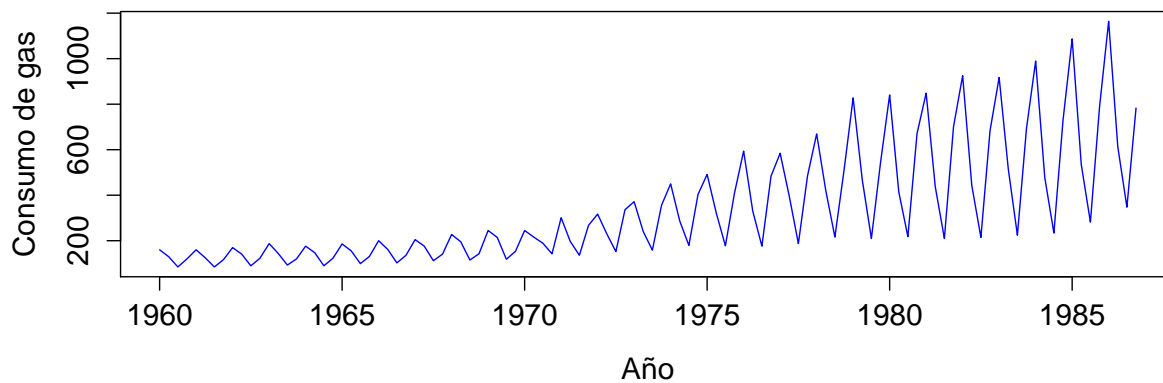


Figura 4.1.2: Figura que muestra el consumo de gas en el Reino Unido

Índice de utilidades Dow Jones

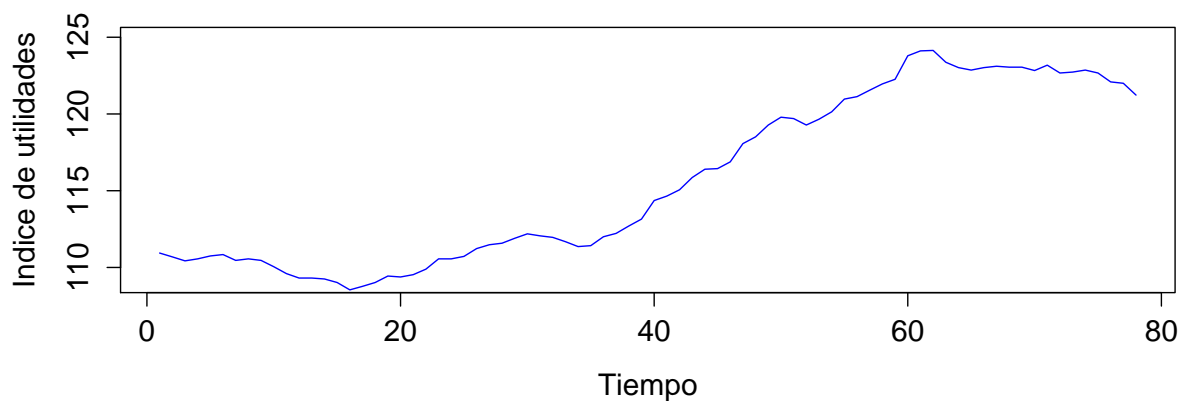


Figura 4.1.3: Figura que muestra el índice de utilidades del Dow Jones

Índice de precios al consumo

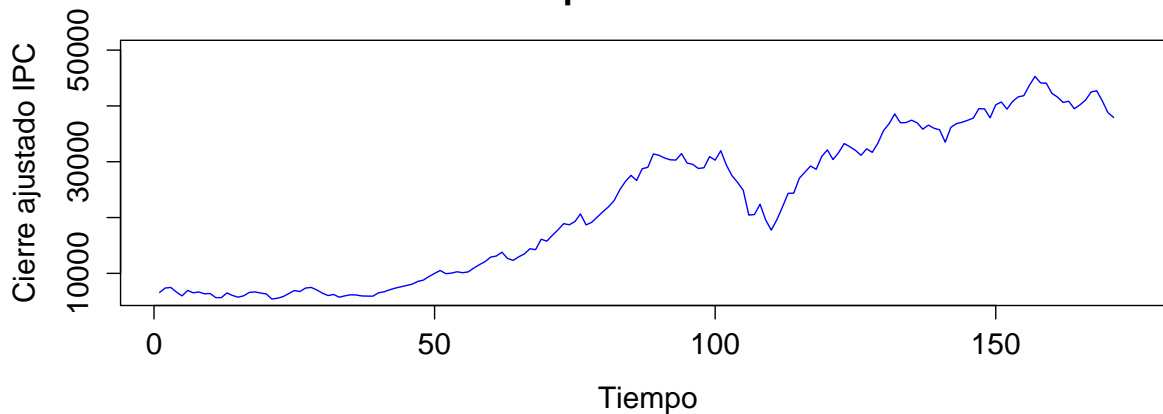


Figura 4.1.4: Figura que muestra el índice de precios al consumo (IPC)

O bien, con la **Figura 4.1.3**, que muestra el índice de utilidades del Dow Jones desde el 28 de agosto de 1972 hasta el 18 de diciembre del mismo año. Aún más irregular es la serie mensual del índice de precios al consumo (IPC) a partir de enero del año 2000 y hasta marzo del año 2014, graficada en la **Figura 4.1.4**. Esta serie claramente no es estacionaria y de hecho es bastante irregular. Comenzamos a sentir la necesidad de un modelo mucho más flexible, el cual no asuma un patrón regular pero que incluya cambios radicales en su estructura.

El análisis de series de tiempo con modelos ARMA(p,q) necesita, la mayoría de las ocasiones, de transformaciones en los datos para volverlos estacionarios. Pero, ¿qué pasaría si encontramos algún modelo que nos permita analizar de manera eficiente unos datos que posean inestabilidad en el nivel de tendencia, inestabilidad en su varianza y que tenga cambios estructurales radicales? Claramente estaríamos mucho más satisfechos.

Los modelos de espacio de estados incluyen a los modelos ARMA(p,q) como caso particular, pero pueden aplicarse a series de tiempo sin la necesidad de aplicar alguna transformación preliminar. Pero existe todavía una razón más poderosa para extender nuestro estudio a los modelos de espacio de estados. Cuando trabajamos con series económicas o financieras, los modelos de series de tiempo univariados se vuelven limitados para analizar dichas series. Expertos encargados de comprender series económicas querrán, por ejemplo, agregar variables macroeconómicas relevantes que puedan llegar a influir en la variable de interés. Un modelo de series de tiempo univariado puede satisfacer las necesidades del investigador al analizar series como la que se muestra en la **Figura 4.1.4**, pero dicho modelo puede ser ineficiente e incapaz al realizar predicciones si no tomamos en cuenta todos aquellos factores políticos, sociales y económicos que pueden llegar a interactuar con la variable de estudio. Antes de comenzar de lleno el tema de este capítulo, veamos algunos conceptos básicos relacionados con la Inferencia Bayesiana que serán útiles en la presentación de estos modelos.

4.2. Inferencia Bayesiana

Supongamos que tenemos un vector de n observaciones dado por $y' = (y_1, y_2, \dots, y_n)$, cuya función de densidad $f(y|\Theta)$ depende de los valores del vector de parámetros $\Theta' = (\theta_1, \theta_2, \dots, \theta_k)$. Suponga que el vector Θ posee su propia función de densidad, denotada por $f(\Theta)$. Entonces:

$$f(y|\Theta)f(\Theta) = f(y, \Theta) = f(\Theta|y)f(y) \quad (4.2.1)$$

Ahora bien, dado el vector de observaciones y , la probabilidad condicional de Θ será:

$$f(\Theta|y) = \frac{f(y|\Theta)f(\Theta)}{f(y)} \quad (4.2.2)$$

pero $f(y)$, la probabilidad incondicional de y está dada por:

$$f(y) = \mathbb{E}(f(y|\Theta)) = C^{-1} = \begin{cases} \int_{\Theta} f(y|\Theta)f(\Theta) d\Theta & \text{para } \Theta \text{ continuo} \\ \sum_{\Theta} f(y|\Theta)f(\Theta) & \text{para } \Theta \text{ discreto} \end{cases}$$

definiendo $f(y) = C^{-1}$ y donde la suma o la integral debe tomarse sobre el soporte de Θ .

De tal forma que la **ec. 4.2.2** puede escribirse como:

$$f(\Theta|y) = Cf(y|\Theta)f(\Theta) \quad (4.2.3)$$

La **ec. 4.2.2** o bien la **ec. 4.2.3** es conocida como el **Teorema de Bayes**. Veamos algunos términos involucrados en dicho teorema. $f(\Theta)$ nos habla de toda aquella información que poseamos de Θ sin tener información alguna proporcionada por unos datos; se le conoce como la distribución *a priori* de Θ . Similarmente, $f(\Theta|y)$ la podemos interpretar como todo aquello que se conoce de Θ dada la información obtenida de unos datos; se le conoce como la distribución *a posteriori* de Θ . La constante C es usada simplemente para asegurar que $f(\Theta|y)$ sume o integre uno.

Ahora bien, dado el vector de información y , $f(\Theta|y)$ puede ser visto como función de Θ . Cuando es así, se le conoce como la función de verosimilitud de Θ dado y y puede escribirse como $l(\Theta|y)$. Por tanto, la **ec. 4.2.3** puede escribirse como:

$$f(\Theta|y) \propto l(\Theta|y)f(\Theta) \quad (4.2.4)$$

¿Qué queremos decir con esto? Pues bien, el **Teorema de Bayes** nos dice que la distribución *a posteriori* de Θ está dada por el producto entre la función de verosimilitud y la distribución *a priori* de Θ . La función de verosimilitud modifica (al agregar la información de unos datos) el conocimiento que se tenía de Θ , es decir, es toda aquella información que obtenemos para Θ dados unos datos.

Debemos comentar también que la función de verosimilitud no se ve afectada al multiplicarla por una constante. El multiplicar la función de verosimilitud por una constante dejará intacta la distribución *a posteriori* de Θ . Ahora, llamaremos verosimilitud estandarizada al término:

$$\frac{l(\Theta|y)}{\int_{\Theta} l(\Theta|y) d\Theta} \quad (4.2.5)$$

y esto se realiza para que el área, volumen o hipervolumen debajo de la curva sea igual a uno. Por otro lado, es claro que la **ec. 4.2.4** nos ayuda a combinar información previa con nueva, es decir, el **Teorema de Bayes** nos permite actualizar continuamente la información de Θ , mientras más observaciones se vayan registrando. Para darnos una mejor idea de lo anterior, imaginemos una observación inicial y_1 , entonces, al aplicar la **ec. 4.2.4** obtenemos:

$$f(\Theta|y_1) \propto l(\Theta|y_1)f(\Theta) \quad (4.2.6)$$

Supongamos que se agrega una observación y_2 adicional e independiente de y_1 , entonces:

$$f(\Theta|y_2, y_1) \propto f(y_1|\Theta)f(y_2|\Theta) \propto l(\Theta|y_1)f(\Theta)l(\Theta|y_2) \propto f(\Theta|y_1)l(\Theta|y_2) \quad (4.2.7)$$

Esta última expresión es de la misma forma que la **ec. 4.2.6** pero $f(\Theta|y_1)$, la distribución *a posteriori* de Θ dado y_1 , jugará ahora el papel de la distribución *a priori* de Θ . Claramente, este proceso puede repetirse las veces que se desee.

En particular, si se tienen n observaciones independientes, entonces la distribución *a posteriori* de Θ puede recalcularse después de agregar una nueva observación, y por ejemplo, el m -ésimo paso de la función de verosimilitud asociada con la m -ésima observación y combinada con la distribución *a posteriori* de Θ después de $m - 1$ observaciones está dada por:

$$f(\Theta|y_1, y_2, \dots, y_m) \propto l(\Theta|y_m)f(\Theta|y_1, y_2, \dots, y_{m-1}) \quad m = 2, \dots, n \quad (4.2.8)$$

Consideremos el siguiente ejemplo. Dos actuarios, A y B, desean obtener estimadores más precisos acerca de la distribución de cierto parámetro α . Supongamos que el actuario A lleva mucho tiempo realizando análisis similares, teniendo claramente mucho más experiencia que el actuario B. Dicho actuario emite su opinión *a priori* acerca de α argumentando que la distribución del parámetro es Normal con media 800 y desviación estándar igual a 20. Es decir, α_A *a priori* $\sim N(800, 20^2)$. Por otro lado, el actuario B, con poca experiencia realizando este tipo de análisis emite su respectiva opinión acerca de la distribución de α , llegando a la conclusión de que α_B *a priori* $\sim N(700, 80^2)$. Las correspondientes funciones de densidad para el actuario A y B serán respectivamente:

$$f_A(\alpha) = \frac{1}{20\sqrt{2\pi}} e^{\left[-\frac{1}{2} \left(\frac{\alpha - 800}{20}\right)^2\right]} \quad \text{y} \quad f_B(\alpha) = \frac{1}{80\sqrt{2\pi}} e^{\left[-\frac{1}{2} \left(\frac{\alpha - 700}{80}\right)^2\right]} \quad (4.2.9)$$

El panel (a) de la **Figura 4.2.1** muestra las correspondientes funciones de densidad para el actuario A y para el actuario B. Supongamos que existe un método insesgado para estimar parámetros y se realiza una observación, y , con este método, arrojando la distribución, $y|\alpha \sim N(\alpha, 40^2)$. Si esta observación es tomada en cuenta, la verosimilitud estandarizada tiene distribución Normal centrada en α , y desviación estándar igual a 40. Veamos cómo se obtiene la distribución de $f(\alpha|y)$. Supongamos que α *a priori* $\sim N(\alpha_0, \sigma_0^2)$, es decir:

$$f(\alpha) = \frac{1}{\sigma_0\sqrt{2\pi}} e^{\left[-\frac{1}{2} \left(\frac{\alpha - \alpha_0}{\sigma_0}\right)^2\right]} \quad -\infty < \alpha < \infty \quad (4.2.10)$$

La función de verosimilitud de α está dada por:

$$l(\alpha|y) \propto e^{\left[-\frac{1}{2} \left(\frac{\alpha - x_y}{\sigma_1}\right)^2\right]} \quad (4.2.11)$$

donde x_y es una función del vector de observaciones, y . Un valor particular de x_y , por ejemplo, podría ser el promedio del conjunto de observaciones y_1, y_2, \dots, y_n , o bien el mismo vector de observaciones, y . La distribución *a posteriori* de α , de α dado el vector de observaciones, será:

$$f(\alpha|y) = \frac{l(\alpha|y)f(\alpha)}{\int_{-\infty}^{\infty} l(\alpha|y)f(\alpha) d\alpha} = \frac{s(\alpha|y)}{\int_{-\infty}^{\infty} s(\alpha|y) d\alpha} \quad (4.2.12)$$

donde

$$s(\alpha|y) \propto e^{\left\{-\frac{1}{2}\left[\left(\frac{\alpha - \alpha_0}{\sigma_0}\right)^2 + \left(\frac{x_y - \alpha}{\sigma_1}\right)^2\right]\right\}} \quad (4.2.13)$$

Haciendo uso de la identidad $A(z - a)^2 + B(z - b)^2 = (A + B)(z - c)^2 + \frac{AB}{A+B}(a - b)^2$ con $c = \frac{1}{A+B}(Aa + Bb)$ podemos escribir:

$$\left(\frac{\alpha - \alpha_0}{\sigma_0}\right)^2 + \left(\frac{x_y - \alpha}{\sigma_1}\right)^2 = (\sigma_0^{-2} + \sigma_1^{-2})(\alpha - \bar{\alpha})^2 + d \quad (4.2.14)$$

donde

$$\bar{\alpha} = \frac{1}{\sigma_0^{-2} + \sigma_1^{-2}}(\sigma_0^{-2}\alpha_0 + \sigma_1^{-2}x_y) \quad (4.2.15)$$

y d es una constante independiente de α . De aquí que:

$$s(\alpha|y) \propto e^{-\frac{(\sigma_0^{-2} + \sigma_1^{-2})(\alpha - \bar{\alpha})^2}{2} - \frac{d}{2}} \quad (4.2.16)$$

y además:

$$\begin{aligned} \int_{-\infty}^{\infty} l(\alpha|y)f(\alpha) d\alpha &= \int_{-\infty}^{\infty} e^{-\frac{(\sigma_0^{-2} + \sigma_1^{-2})(\alpha - \bar{\alpha})^2}{2} - \frac{d}{2}} d\alpha \quad (4.2.17) \\ &= e^{-\frac{d}{2}} \frac{\sqrt{2\pi}}{(\sigma_0^{-2} + \sigma_1^{-2})^{\frac{1}{2}}} \end{aligned}$$

de donde:

$$f(\alpha|y) = \frac{(\sigma_0^{-2} + \sigma_1^{-2})^{\frac{1}{2}}}{\sqrt{2\pi}} e^{-\frac{(\sigma_0^{-2} + \sigma_1^{-2})(\alpha - \bar{\alpha})^2}{2}} \quad (4.2.18)$$

es decir, $f(\alpha|y) \sim N\left(\bar{\alpha}, \sqrt{\frac{1}{\sigma_0^{-2} + \sigma_1^{-2}}}\right)$.

Regresando a nuestro ejemplo, supongamos que el resultado de una sola observación está dada por $y = x_y = 750$. La función de verosimilitud tendrá distribución Normal con media $y = 750$ y desviación estándar $\sigma_1 = 40$; es decir, $l(\alpha|y) \sim N(750, 40^2)$. Dicha función se muestra en el panel (b) de la **Figura 4.2.1**.

La opinión *a posteriori* del actuario A estará dada por $f_A(\alpha|y)$ con media $\bar{\alpha}$ y desviación estándar $\tilde{\sigma}^2$ dados por:

$$\bar{\alpha} = \frac{1}{\frac{1}{20^2} + \frac{1}{40^2}} \left(\frac{1}{20^2} 800 + \frac{1}{40^2} 750 \right) = 790 \quad \text{y} \quad \tilde{\sigma}^2 = \frac{1}{\frac{1}{20^2} + \frac{1}{40^2}} = 320$$

es decir, $f_A(\alpha|y = 750) \sim N(790, 17.89)$.

Asimismo, la opinión *a posteriori* del actuario B estará dada por $f_B(\alpha|y = 750)$ con media 740 y desviación estándar igual a 35.78. Los cálculos se muestran a continuación:

$$\bar{\alpha} = \frac{1}{\frac{1}{80^2} + \frac{1}{40^2}} \left(\frac{1}{80^2} 700 + \frac{1}{40^2} 750 \right) = 740 \quad \text{y} \quad \tilde{\sigma}^2 = \frac{1}{\frac{1}{80^2} + \frac{1}{40^2}} = 1280$$

El panel (c) de la **Figura 4.2.1** muestra tanto $f_A(\alpha|y)$ como $f_B(\alpha|y)$ cuando $y = 750$. Observamos que después de agregar la primer observación, las opiniones *a posteriori* de los actuarios acerca de α son mucho más parecidas aunque todavía difieren.

Ahora bien, ¿qué sucedería si agregamos, digamos, 99 observaciones más? Es decir, ¿qué pasaría si tuviésemos $n = 100$ observaciones y la media de dichas 100 observaciones fuese, por ejemplo, $\bar{y} = 770$? En general la función de verosimilitud de Θ dado n observaciones independientes provenientes de una población con distribución $N(\Theta, \sigma^2)$ es:

$$l(\Theta|y) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n e^{-\left[\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \Theta)^2 \right]} \quad (4.2.19)$$

Donde, sumando y restando \bar{y} en la expresión anterior:

$$\sum_{i=1}^n (y_i - \Theta)^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + n(\Theta - \bar{y})^2 \quad (4.2.20)$$

Como $\sum_{i=1}^n (y_i - \bar{y})^2$ es una constante que no depende de Θ , la función de verosimilitud será:

$$l(\Theta|y) \propto e^{-\left[-\frac{1}{2} \left(\frac{\Theta - \bar{y}}{\frac{\sigma}{\sqrt{n}}} \right)^2 \right]} \quad (4.2.21)$$

La función de verosimilitud tiene distribución Normal con media \bar{y} y desviación estándar $\frac{\sigma}{\sqrt{n}}$. En nuestro ejemplo de 100 observaciones, $\bar{y} = 770$, $n = 100$ y $\frac{\sigma}{\sqrt{n}} = 4$. Entonces, podemos tomar estos dos datos como si fueran una sola observación y realizar el procedimiento anterior. El panel (d) de la **Figura 4.2.1** muestra la función de verosimilitud dado $\bar{y} = 770$.

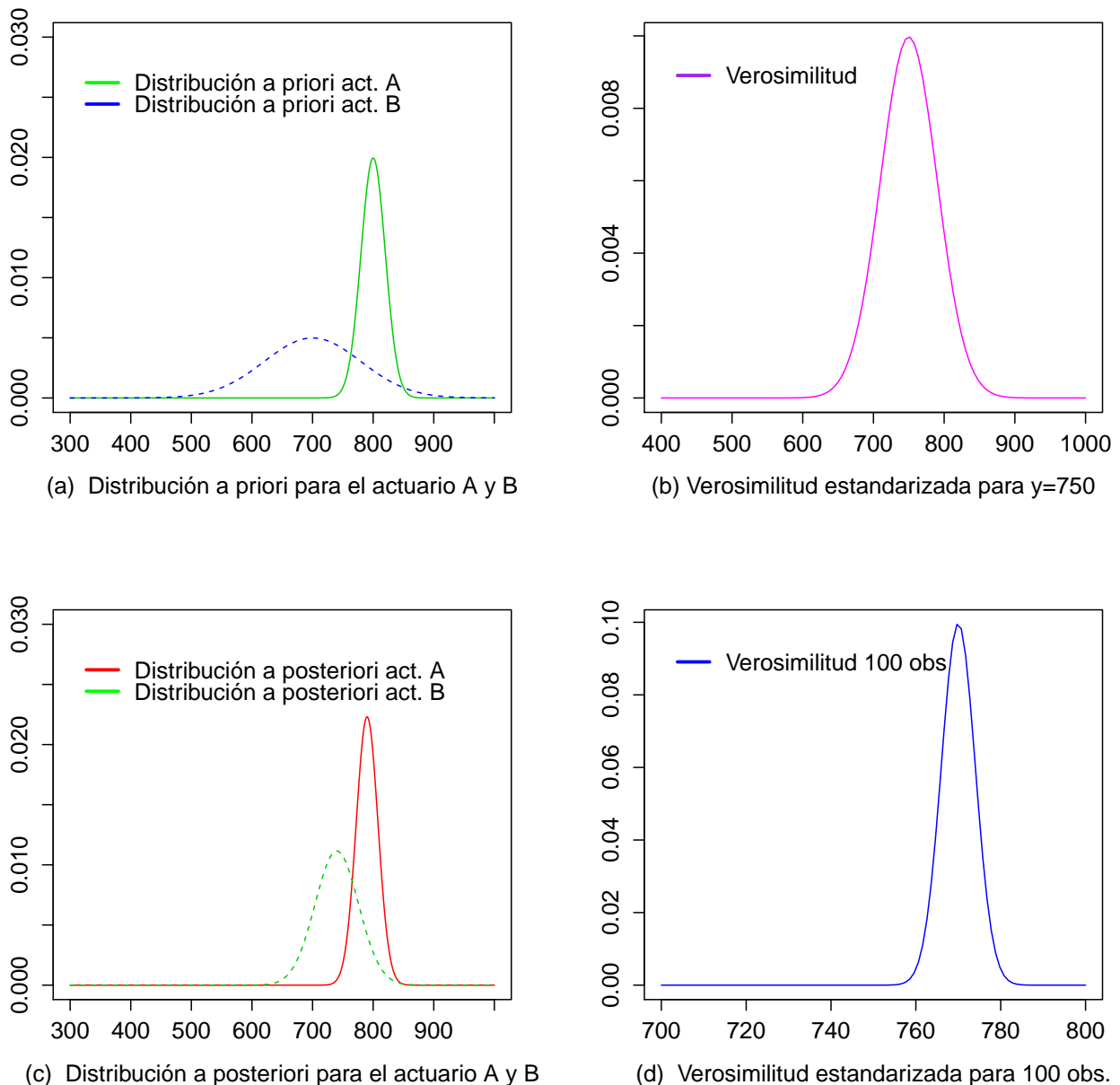


Figura 4.2.1: Figura que muestra la distribución *a priori* y *a posteriori* de los actuarios

Obtenemos, dado $\bar{y} = 770$, que la opinión *a posteriori* del actuario A estará dada por $f_A(\alpha|y)$ con media 771.2 y desviación estándar igual a 3.922, mientras que la opinión *a posteriori* del actuario B estará dada por $f_B(\alpha|y)$ con media 769.8 y desviación estándar igual a 3.995. Estas dos distribuciones se muestran en la **Figura 4.2.2**, observando que, después de 100 observaciones, el actuario A y el actuario B tendrán casi la misma opinión acerca de la distribución de α .

Con este ejemplo hemos querido explicar la teoría vista anteriormente, haciendo hincapié en el papel que desempeña la función de verosimilitud; gracias a ella, a la información que nos proporciona, la opinión o distribución *a posteriori* de ambos actuarios cambia, acercándose cada vez más, conforme aumenta la información disponible, a la realidad.

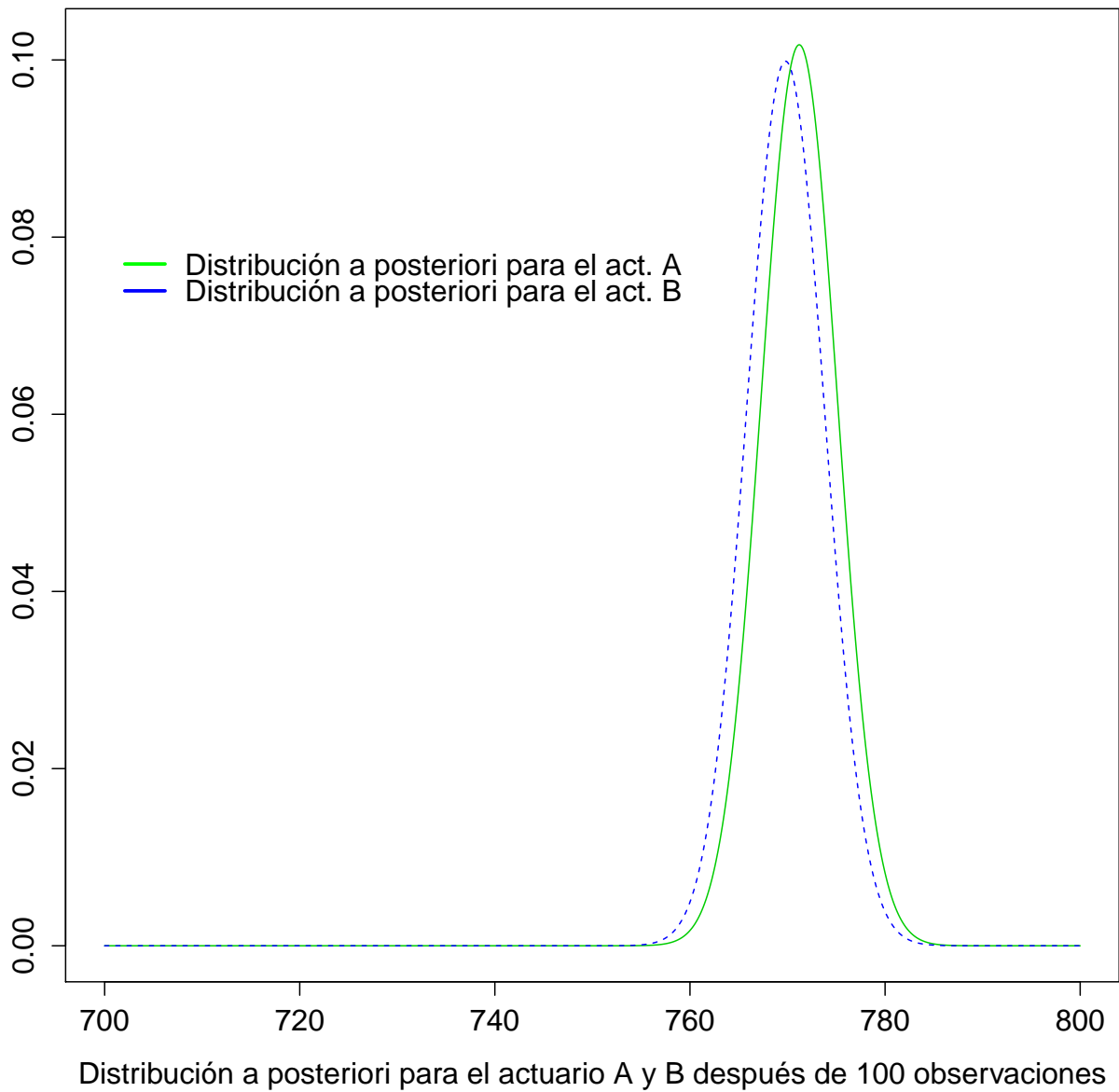


Figura 4.2.2: Figura que muestra la distribución *a posteriori* del actuario A y B | $\bar{y} = 770$

4.3. Modelos de espacio de estados

Consideremos $\{Y_t : t \in T\}$ un conjunto de variables aleatorias definidas en el mismo conjunto (Ω, F, \mathbb{P}) donde $T \neq \emptyset$ es un conjunto de índices, Ω es el espacio de eventos, F una σ -álgebra y \mathbb{P} una medida de probabilidad. Decimos que el proceso $\{Y_t\}$ es una cadena de Markov si $\forall t > 1$:

$$\mathbb{P}(Y_t = y_t | Y_1 = y_1, Y_2 = y_2, \dots, Y_{t-1} = y_{t-1}) = \mathbb{P}(Y_t = y_t | Y_{t-1} = y_{t-1}) \quad (4.3.1)$$

es decir, la **ec. 4.3.1** implica que la información acerca de Y_t dado toda la información que nos pueda dar Y_1, Y_2, \dots, Y_{t-1} es exactamente la misma que nos da simplemente la observación Y_{t-1} . Existen otras formas equivalentes de expresar esta propiedad; la que nos interesa será:

$$\mathbb{P}(y_1, y_2, \dots, y_t) = \mathbb{P}(y_1)\mathbb{P}(y_2|y_1)\dots\mathbb{P}(y_t|y_{t-1}) \quad (4.3.2)$$

En los modelos de espacio de estados ocuparemos a $\{\Theta_t\}$, serie de tiempo que nos ayudará en la tarea de especificar la distribución de otra serie de tiempo conocida como $\{Y_t\}$.

Definición 4.3.1 *Formalmente, un modelo de espacio de estados consiste en una serie de tiempo $\{\Theta_t : t = 0, 1, \dots\} \in \mathbb{R}^p$ y otra serie de tiempo $\{Y_t : t = 0, 1, \dots\} \in \mathbb{R}^m$, las cuales satisfacen las siguientes condiciones:*

1. $\{\Theta_t\}$ es una cadena de Markov.
2. Condicionando con respecto a $\{\Theta_t\}$, las Y_t 's son independientes y Y_t sólo depende de su correspondiente Θ_t .

4.3.1. Modelos dinámicos lineales

Un modelo dinámico lineal está especificado completamente por la distribución Normal multivariada *a priori*, Θ_0 *a priori* $\sim N_p(\mu_0, \sigma_0^2)$ y por las siguientes ecuaciones $\forall t \geq 1$:

$$Y_t = F_t \Theta_t + \alpha_t \quad \alpha_t \sim N_m(0, \psi_t) \quad (4.3.3)$$

$$\Theta_t = G_t \Theta_{t-1} + \beta_t \quad \beta_t \sim N_p(0, \Psi_t) \quad (4.3.4)$$

donde F_t es una matriz de tamaño $m \times p$, G_t es una matriz de tamaño $p \times p$ y $\{\alpha_t\}$, $\{\beta_t\}$ son dos secuencias independientes de variables aleatorias con distribución Normal, ambas con media 0 y varianza dada por las matrices ψ_t y Ψ_t , respectivamente. La **ec. 4.3.3** es conocida como la **ecuación de observación** mientras que la **ec. 4.3.4** es conocida como la **ecuación del sistema**.

Veamos algunos ejemplos. El primero de ellos está dado por:

$$Y_t = \mu_t + \alpha_t \quad \alpha_t \sim N(0, \psi) \quad (4.3.5)$$

$$\mu_t = \mu_{t-1} + \beta_t \quad \beta_t \sim N(0, \Psi) \quad (4.3.6)$$

donde $m = p = 1$, $\Theta_t = \mu_t$ y $F_t = G_t = 1$. Este modelo es conocido como **caminata aleatoria con ruido**.

Imaginemos la siguiente situación. Usted se encuentra afuera del estadio *Novo Castela*, en la ciudad de Fortaleza, esperando a su mejor amigo para ver el partido México-Brasil del mundial de fútbol 2014.

Usted y su amigo no quedaron en una puerta específica como lugar de encuentro. Usted tiene por ejemplo, la opción de quedarse en la puerta **X** y esperar a que su amigo recorra todas las puertas de entrada hasta encontrarlo. En este panorama usted, o más bien, su posición, $\{\Theta_t\}$, es la misma $\forall t$. El otro panorama es que usted se mueva también buscando a su amigo, por lo cual, su posición, $\{\Theta_t\}$, cambiará con el tiempo.

El primer escenario puede ser modelado modificando la **ec. 4.3.5**, de la siguiente forma:

$$Y_t = \mu + \alpha_t \quad \alpha_t \sim N(0, \sigma^2) \quad (4.3.7)$$

es decir, las observaciones dependerán de nuestra posición, $\Theta = \mu$, la cual es fija, y de la sucesión $\{\alpha_t\}$. El segundo escenario está modelado por la **ec. 4.3.5** y la **ec. 4.3.6** en la cual suponemos la posición $\{\Theta_t\} = \{\mu_t\}$ variante en el tiempo y descrita por una caminata aleatoria.

Este modelo es apropiado para analizar series de tiempo que no muestren una clara tendencia o componente estacional. La serie de tiempo, $\{Y_t\}$, es modelada a partir de ruidos sobre la serie $\{\mu_t\}$, la cual a su vez está descrita por una caminata aleatoria.

El siguiente ejemplo es conocido como **modelo de crecimiento lineal**. Es una pequeña variación del ejemplo anterior dado por las ecuaciones siguientes:

$$\begin{aligned} Y_t &= \mu_t + \alpha_t & \alpha_t &\sim N(0, \psi) \\ \mu_t &= \mu_{t-1} + \delta_{t-1} + \beta_{t,1} & \beta_{t,1} &\sim N(0, \sigma_\mu^2) \\ \delta_t &= \delta_{t-1} + \beta_{t,2} & \beta_{t,2} &\sim N(0, \sigma_\delta^2) \end{aligned} \quad (4.3.8)$$

$$\text{con } \Theta_t = \begin{pmatrix} \mu_t \\ \delta_t \end{pmatrix}, G_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \Psi_t = \begin{pmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\delta^2 \end{pmatrix} \text{ y } F_t = (1 \ 0).$$

Intuitivamente, el modelo de crecimiento lineal nos dice que la secuencia $\{\delta_t\}$, la cual podemos identificar como la pendiente (que varía en el tiempo) del proceso $\{\mu_t\}$, depende de observaciones anteriores así como del conjunto de errores no correlacionados $\{\beta_{t,2}\}$.

Los modelos lineales generalizados mixtos pueden ser vistos como una generalización del modelo de regresión lineal permitiendo que los coeficientes del modelo de regresión varíen con el paso del tiempo.

Lo anterior se traduce en la siguiente ecuación:

$$\begin{aligned} Y_t &= \alpha_{t,0} + \alpha_{t,1}x_t + \epsilon_t & \epsilon_t &\sim NID(0, \sigma_\epsilon^2) \\ &= (1 \quad x_t) \begin{pmatrix} \alpha_{t,0} \\ \alpha_{t,1} \end{pmatrix} + \epsilon_t & & (4.3.9) \\ &= F_t \Theta_t + \epsilon_t \end{aligned}$$

con $\alpha_{t,j}$, la evolución temporal, dada por $\alpha_{t,j} = \alpha_{t-1,j} + \beta_{t,j}$, donde $\beta_{0,j}$ y $\beta_{1,j}$ son variables aleatorias independientes Gaussianas. La **ec. 4.3.9**, corresponde a la ecuación de observación de un modelo dinámico lineal y cuya ecuación del sistema está dada por:

$$\Theta_t = G_t \Theta_{t-1} + \beta_t \quad \beta_t \sim N_2(0, \Psi_t) \quad (4.3.10)$$

con Ψ_t matriz diagonal de tamaño 2x2 tal que sus elementos son las varianzas de $\beta_{0,j}$ y $\beta_{1,j}$.

Este modelo es un modelo dinámico con $\Theta_t = \begin{pmatrix} \alpha_{t,0} \\ \alpha_{t,1} \end{pmatrix}$, $G_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $F_t = (1 \quad x_t)$ y $\psi = \sigma_\epsilon^2$.

El **modelo de regresión lineal simple**, $Y_t = \alpha + \alpha_1 x_t + \epsilon_t$, puede ser incorporado a los modelos dinámicos lineales, tomando simplemente a G_t como la matriz identidad ($G_t = I$), $\sigma_t^2 = \sigma^2$ y $\beta_t = 0$.

La **ec. 4.3.9** puede ser generalizada para el modelo de regresión lineal mixto múltiple como:

$$Y_t = x_t' \Theta_t + \alpha_t \quad \alpha_t \sim N(0, \sigma_t^2) \quad (4.3.11)$$

$$\Theta_t = G_t \Theta_{t-1} + \beta_t \quad \beta_t \sim N_p(0, \Psi_t) \quad (4.3.12)$$

donde claro, $\Theta_t = (\alpha_{t,0} \quad \alpha_{t,1} \quad \dots \quad \alpha_{t,m})'$, $F_t = x_t = (x_{t,1} \quad x_{t,2} \dots x_{t,m})'$, son todas las m variables explicativas incluidas en el modelo a tiempo t y en donde si se incluye un término constante, $x_{1,t}$ vale 1. Como primer opción tomaremos a la matriz G_t como la matriz identidad mientras que la matriz β_t será una matriz diagonal con los coeficientes de la regresión como elementos en la diagonal de la matriz. Finalmente, $\psi_t = \sigma_\epsilon^2$.

Nuestro siguiente ejemplo consistirá en representar, con un modelo dinámico lineal, a los modelos ARMA(p,q) y ARIMA(p,d,q). Comencemos con los modelos ARMA(p,q), los cuales supondremos estacionarios y centrados.

Petris *et al.* (2009) utiliza la siguiente representación para modelos ARMA(p,q), en los cuales se supone que el ruido blanco es Gaussiano y donde los parámetros $\{\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q\}$ son tales que el proceso $\{Y_t\}$ sea estacionario:

$$Y_t = \mu + \sum_{j=1}^p \phi_j (Y_{t-j} - \mu) + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t \quad (4.3.13)$$

Por simplicidad, supongamos $\mu = 0$; la **ec. 4.3.13** resulta en:

$$Y_t = \sum_{j=1}^r \phi_j Y_{t-j} + \sum_{j=1}^{r-1} \theta_j \epsilon_{t-j} + \epsilon_t \quad (4.3.14)$$

con $r = \max\{p, q + 1\}$, $\phi_j = 0$ para $j > p$ y $\theta_j = 0$ para $j > q$. Introducimos al r -vector de estados $\Theta_t = (\theta_{1,t} \ \theta_{2,t} \ \theta_{3,t} \ \dots \ \theta_{r,t})'$. Se definen también las matrices $F_t = (1 \ 0 \ 0 \ \dots \ 0)$, $R = (1 \ \theta_1 \ \theta_2 \ \dots \ \theta_{r-1})'$ y:

$$G_t = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \phi_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_{r-1} & 0 & 0 & \dots & 1 \\ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix}$$

La representación de los modelos ARMA(p,q) mediante modelos dinámicos lineales será entonces:

$$Y_t = F_t \Theta_t \quad (4.3.15)$$

$$\Theta_{t+1} = G_t \Theta_t + R \epsilon_t \quad (4.3.16)$$

con $\psi = 0$ y $\Psi = RR' \sigma_\epsilon^2 = \sigma_\epsilon^2 (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_{r-1}^2)$, la varianza del proceso de promedios móviles. $\Psi = RR' \sigma_\epsilon^2$ concuerda con lo obtenido en la **ec. 3.3.11**. De la ecuación de observación, es fácil ver que $Y_t = \theta_{1,t}$. Al desarrollar la ecuación del sistema obtenemos:

$$\begin{aligned} \theta_{1,t} &= \phi_1 \theta_{1,t-1} + \theta_{2,t-1} + \epsilon_t \\ \theta_{2,t} &= \phi_2 \theta_{1,t-1} + \theta_{3,t-1} + \theta_1 \epsilon_t \\ &\vdots \\ \theta_{r-1,t} &= \phi_{r-1} \theta_{1,t-1} + \theta_{r,t-1} + \theta_{r-2} \epsilon_t \\ \theta_{r,t} &= \phi_r \theta_{1,t-1} + \theta_{r-1} \epsilon_t \end{aligned} \quad (4.3.17)$$

De la ecuación $\theta_{2,t} = \phi_2 \theta_{1,t-1} + \theta_{3,t-1} + \theta_1 \epsilon_t$ podemos sustituir t por $t - 1$ obteniendo ahora $\theta_{2,t-1} = \phi_2 \theta_{1,t-2} + \theta_{3,t-2} + \theta_1 \epsilon_{t-1}$. Este valor, $\theta_{2,t-1}$, lo podemos sustituir en la ecuación $\theta_{1,t} = \phi_1 \theta_{1,t-1} + \theta_{2,t-1} + \epsilon_t$, obteniendo:

$$\theta_{1,t} = \phi_1 \theta_{1,t-1} + \phi_2 \theta_{1,t-2} + \theta_{3,t-2} + \theta_1 \epsilon_{t-1} + \epsilon_t \quad (4.3.18)$$

Realizamos el mismo procedimiento para $\theta_{3,t-2}$ y así sucesivamente obteniendo:

$$Y_t = \theta_{1,t} = \phi_1 \theta_{1,t-1} + \dots + \phi_r \theta_{1,t-r} + \theta_1 \epsilon_{t-1} + \dots + \theta_{r-1} \epsilon_{t-r+1} + \epsilon_t \quad (4.3.19)$$

Como $r = \max\{p, q + 1\}$, observamos que la **ec. 4.3.19** es efectivamente, la representación del proceso ARMA(p,q) mediante modelos dinámicos lineales. Observe que $\theta_{1,t}$ toma el lugar de Y_t , $\theta_{1,t-1}$ el de Y_{t-1} y así sucesivamente.

Veamos un caso particular de los procesos ARMA(p,q), a saber será el caso de los modelos AR(p). Suponiendo $\mu = 0$, un modelo AR(p) está dado por la **ec. 3.4.14** en la cual, el proceso $\{\epsilon_t\} \sim WN(0, \sigma_\epsilon^2)$. Sean $\Theta_t = (\theta_{1,t} \ \theta_{2,t} \ \theta_{3,t} \ \dots \ \theta_{p,t})'$, $\Psi = \text{diag}(\sigma_\epsilon^2 \ 0 \ \dots \ 0)$, $\psi = 0$, $F = (1 \ 0 \ 0 \ \dots \ 0)$ y:

$$G = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \phi_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_{r-1} & 0 & 0 & \dots & 1 \\ \phi_r & 0 & 0 & \dots & 0 \end{bmatrix}$$

El sistema de ecuaciones puede definirse como en la **ec. 4.3.17** o bien de la siguiente forma:

$$\Theta_{t+1} = \begin{bmatrix} \theta_{1,t} \\ \theta_{2,t} \\ \theta_{3,t} \\ \vdots \\ \theta_{p-1,t} \\ \theta_{p,t} \end{bmatrix} = \begin{bmatrix} \phi_1 & 1 & 0 & \dots & 0 \\ \phi_2 & 0 & 1 & \dots & 0 \\ \phi_3 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \phi_{p-1} & 0 & 0 & \dots & 1 \\ \phi_p & 0 & 0 & \dots & 0 \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

Al desarrollar como en la **ec. 4.3.17** obtenemos $Y_t = \theta_{1,t} = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t$, la representación de un proceso AR(p).

Para un modelo de espacio de estados tendremos como objetivos principales la inferencia de estados no observados y el pronóstico para futuras observaciones. La estimación y el pronóstico se realizan calculando las distribuciones condicionales de las cantidades de interés, dada la información.

Vamos a diferenciar tres distintos problemas, a saber, aquellos que involucran filtros, suavizamiento y pronóstico. Para comprender el término de filtro imaginemos la siguiente situación. Usted se encuentra trabajando en la Bolsa Mexicana de Valores, el día está por terminar y usted, con base en la información que obtuvo ese día, realiza un pronóstico para el precio de una acción de cierta compañía. Usted termina su trabajo, se retira y al día siguiente regresa y observa su pronóstico. ¿Debería usted quedarse con tal pronóstico, hecho con información del día anterior o debería usted tomar en cuenta lo que ha sucedido, es decir, los cambios que ocurrieron en el mercado, desde el día de ayer hasta el momento en que volvió a observar su pronóstico? Claramente, debemos tomar en cuenta todo aquel acontecimiento que haya ocurrido en este lapso de tiempo, por muy corto que este sea.

El filtro es un procedimiento que nos ayuda a estimar nuestro vector de estados, usando la información que tengamos disponible hasta el presente, y actualizar nuestros estimados y pronósticos conforme nueva información vaya surgiendo. El filtro de Kalman nos da la fórmula para pasar de $\mathbb{P}(\Theta_t | y_1, y_2, \dots, y_t)$ a $\mathbb{P}(\Theta_{t+1} | y_1, y_2, \dots, y_{t+1})$.

El problema de suavizamiento consiste simplemente en estudiar de manera retrospectiva el comportamiento de las observaciones. Para el problema de pronóstico, es claro que para poder pronosticar la observación Y_{t+1} dado y_1, y_2, \dots, y_t es menester primero obtener el pronóstico para Θ_{t+1} para después, y utilizando el pronóstico anterior, predecir el comportamiento de Y_{t+1} . Esto claramente puede generalizarse para Y_{t+k} con $k \geq 1$.

Hasta aquí abordaremos dichos problemas y no entraremos en más detalle ya que solamente deseamos dar una introducción a los modelos de espacio de estados y concentrarnos en los modelos de regresión dinámica. Para mayor detalle sobre estos procedimientos, recursiones y teoremas, consultar en Petris *et al.* (2009).

Pasemos ahora a la parte práctica. R nos da la opción de manejar modelos dinámicos lineales. Se encuentran dentro del paquete “dlm” y se manejan como objetos de la clase “dlm”. Dicho objeto se maneja como una lista con elementos $m_0, C_0, FF, GG, V = \psi, W = \Psi, JFF, JV, JGG$, y JW . Los primeros seis atributos corresponden a modelos dinámicos lineales donde no varía el tiempo, mientras que los cuatro restantes se utilizan para modelos de tiempo variante.

Los comandos `as.dlm` e `is.dlm` nos ayudarán a probar si cierto modelo representa un modelo dinámico lineal, es decir, estos comandos verificarán todos aquellos requisitos subyacentes de la **ec. 4.3.3** y la **ec. 4.3.4**.

Entonces, para un modelo que no varía en el tiempo debemos especificar:

1. La distribución *a priori*, Θ_0 *a priori*, con media inicial, m_0 , y su matriz de covarianzas, C_0 . Si no tenemos información *a priori*, señalaremos un vector de ceros y una matriz diagonal con varianzas muy grandes. De esta manera la distribución *a priori* corresponde a una distribución no informativa, es decir, corresponde a información general sobre una variable.
2. La matriz de transición $F_t = FF$.
3. La matriz de observación $G_t = GG$.
4. La matriz de covarianzas V de la ecuación del sistema o de estado β_t , $V = \psi$.
5. La matriz de covarianzas W de la ecuación de observación α_t , $W = \Psi$.

Veamos un ejemplo sencillo. Recordemos un poco el modelo **caminata aleatoria con ruido**. La ecuación de observación y la ecuación del sistema estaban dadas por:

$$\begin{aligned} Y_t &= \mu_t + \alpha_t & \alpha_t &\sim N(0, \psi) \\ \mu_t &= \mu_{t-1} + \beta_t & \beta_t &\sim N(0, \Psi) \end{aligned}$$

donde $m = p = 1$, $\Theta_t = \mu_t$ y $F_t = G_t = 1$. Apliquemos dicho ejemplo a la serie que representa el flujo anual del río Nilo en Ashwan desde el año 1871 y hasta el año 1970. La **Figura 4.3.1** muestra dicha serie.

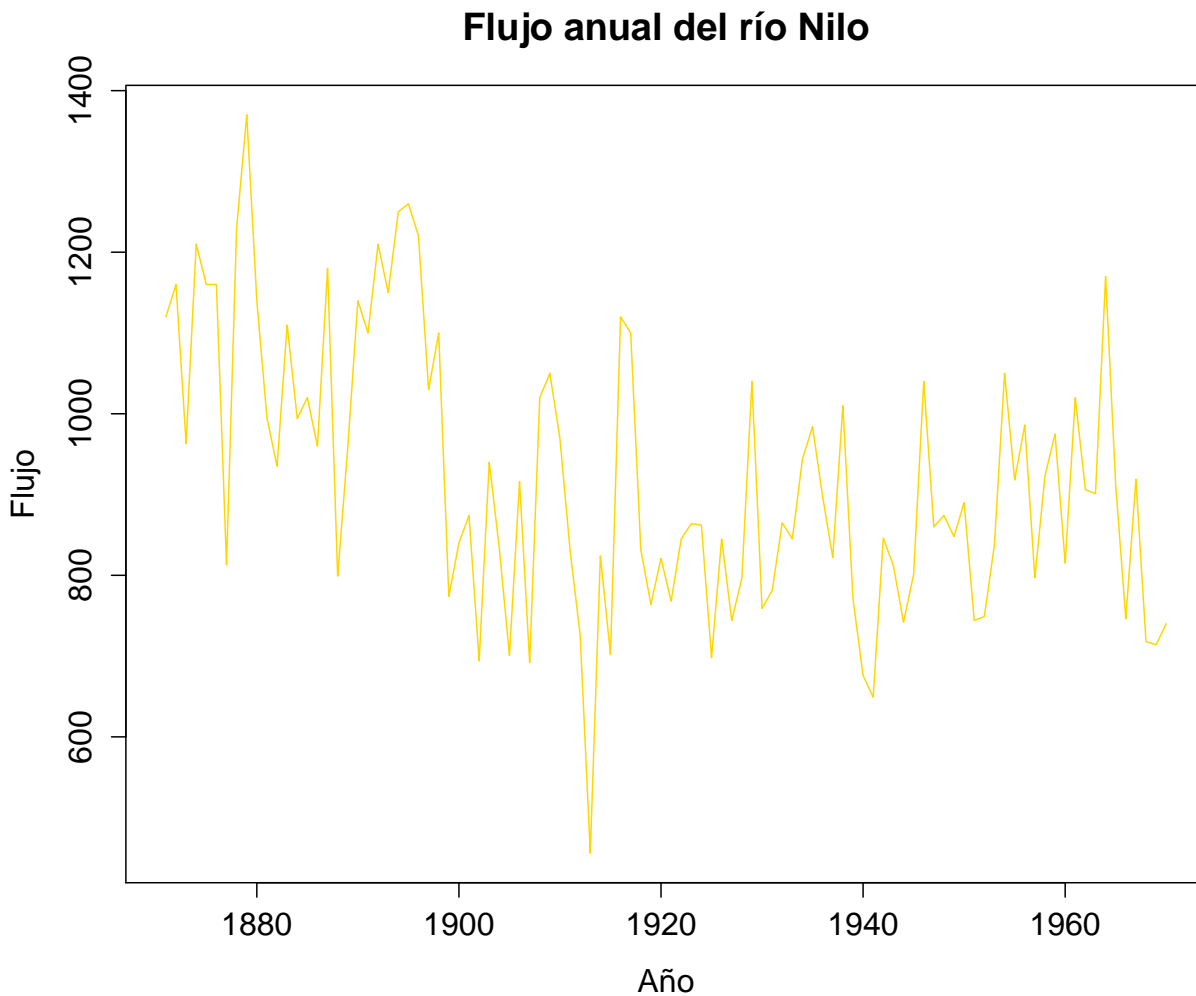


Figura 4.3.1: Figura que muestra el flujo anual del río Nilo en los años 1871-1970

Para especificar el modelo podríamos utilizar la función `d1m` con los siguientes parámetros:

```
> Nilo<-d1m(m0=1100,C0= 10^7,FF=1,GG=1,W=1468,V=15100)
```

Recordemos que $V = \psi$ es la varianza de la ecuación de observación, mientras que $W = \Psi$ representa a la varianza de la ecuación del sistema; también, hemos usado C_0 muy grande para que la distribución sea no informativa. Además, es claro de la **Figura 4.3.1** que la media de la serie es distinta de cero. Al realizar el cálculo de la media obtenemos un valor $m_0 = 1100$. Utilizaremos alternativamente para la especificación del modelo, el comando `d1mModPoly`, el cual crea un polinomio de grado n para el modelo dinámico lineal. El comando `d1mFilter` realiza el filtro correspondiente y además (dentro de este comando) provee, con base en nuestros datos, el pronóstico a un tiempo posterior. Finalmente el comando `d1mSmooth` realiza un suavizamiento del filtro.

El ajuste, filtro y suavizamiento del modelo están dados por el siguiente código:

```
> NiloPolinomio<-dlmModPoly(order=1,dV=15100,dW=1468,m0=1100)
> NiloAjuste<-dlmFilter(Nilo,NiloPolinomio)
> NiloSuaviza<-dlmSmooth(NiloAjuste)
```

Para la función *dlmModPoly* se requiere especificar el orden del polinomio, el valor esperado del vector de estados m_0 , la varianza del vector de estados C_0 , la varianza de la ecuación de observación dV (que en este escrito $dV = V = \psi$) y la varianza de la ecuación del sistema dW (de la misma forma $dW = W = \Psi$). Hemos seleccionado $V = 15100$ y $W = 1468$, ya que estos valores son los estimadores máximo verosímiles para cada una de las varianzas. La **Figura 4.3.2** muestra los datos originales así como el correspondiente filtro. La **Figura 4.3.3** muestra los datos originales con su respectivo suavizamiento. Finalmente, la **Figura 4.3.4** muestra los datos originales con el correspondiente pronóstico a un paso.

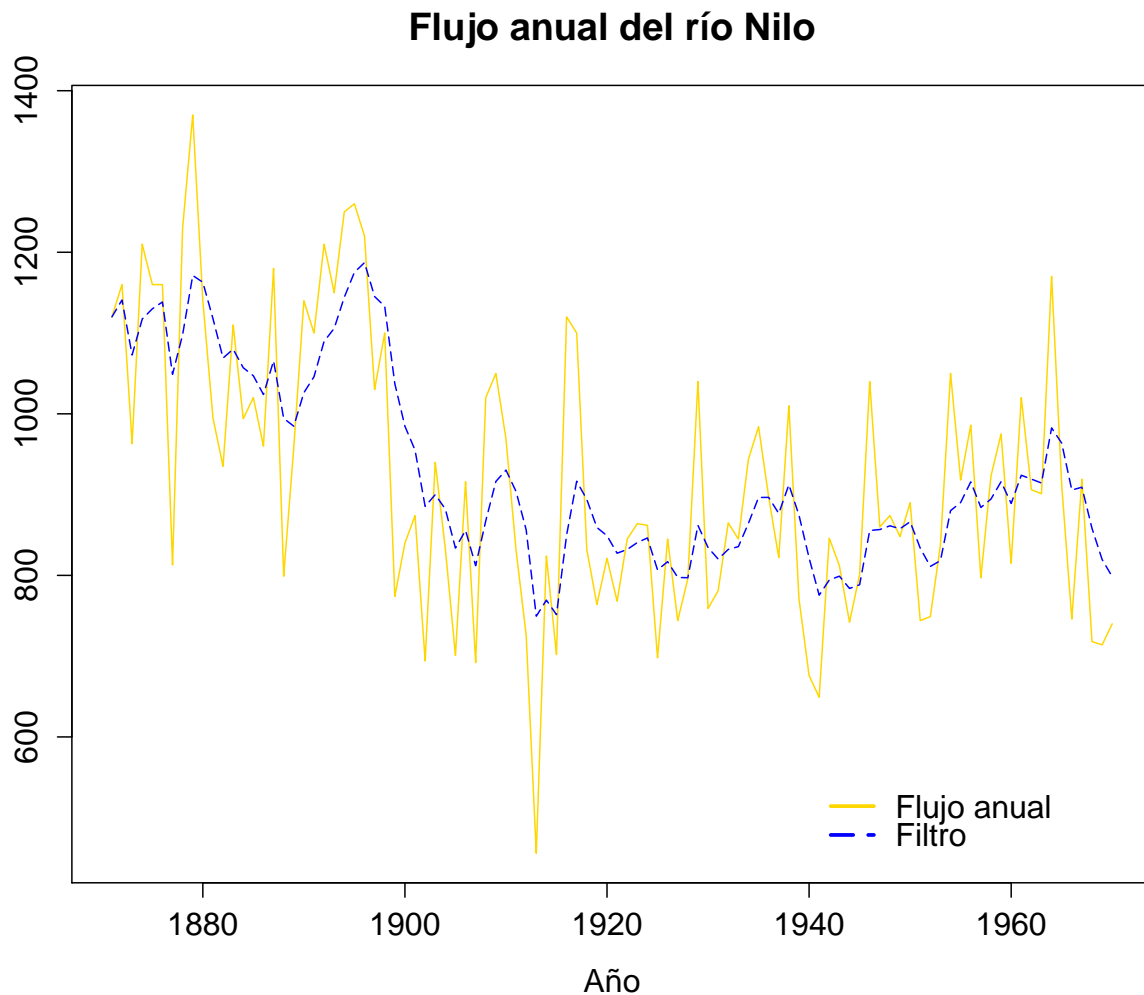


Figura 4.3.2: Figura que muestra el filtro para el flujo anual del río Nilo

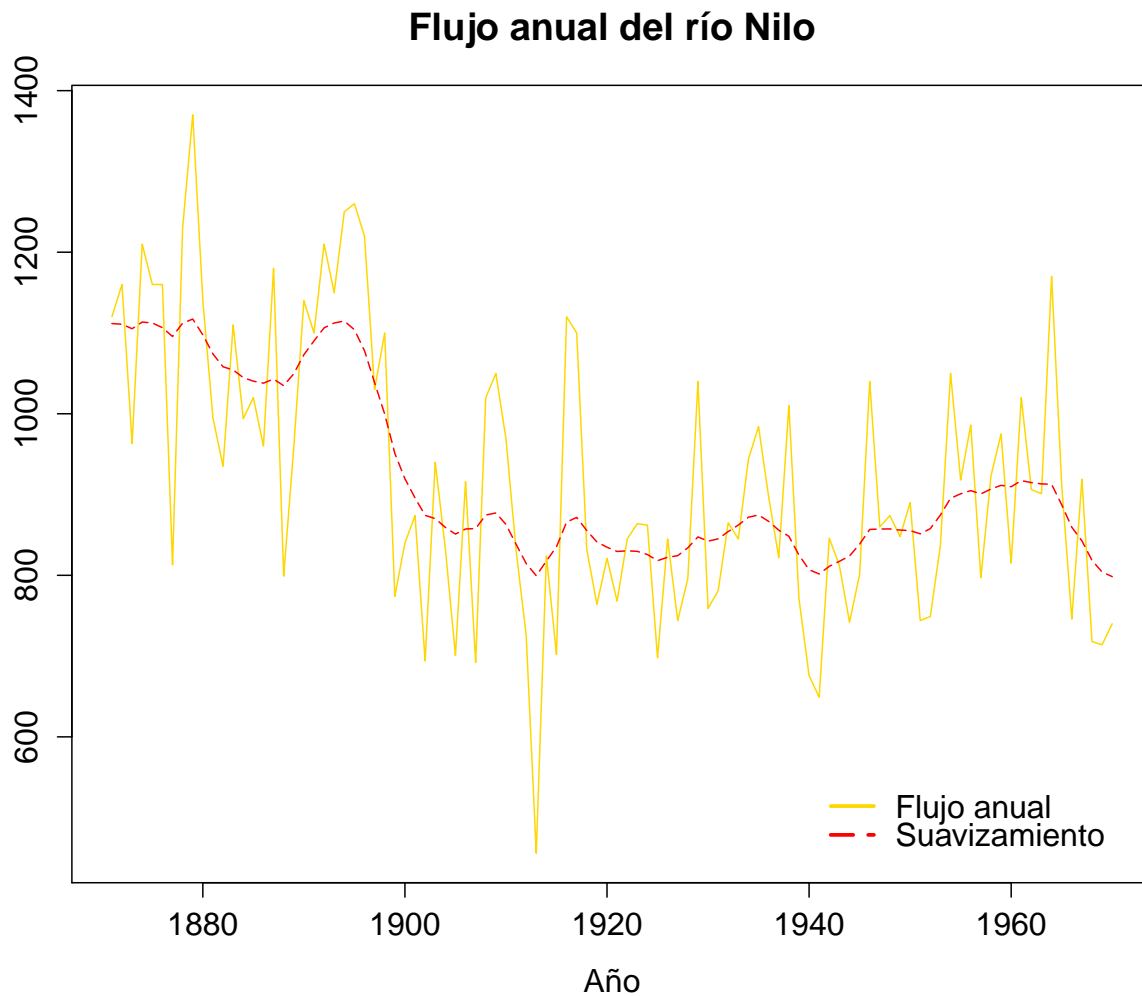


Figura 4.3.3: Figura que muestra el suavizamiento para el flujo anual del río Nilo

Recordemos que los valores dV y dW representan la varianza del proceso α_t y los elementos de la diagonal de la matriz del proceso β_t . Para terminar esta sección, y como ejercicio, modificaremos dV y dW para ver el cambio en el filtro y pronóstico. Las instrucciones que se presentan a continuación sirven para estimar el filtro ya visto arriba y otro filtro con valores $\psi = 15100$ y $\Psi = 150$. Mostraremos posteriormente comparativos entre ambos modelos y mediante el análisis de residuos concluiremos qué modelo nos conviene seleccionar:

```
> NiloPolinomio2<-dlmModPoly(order=1,dV=15100,dW=150)
> NiloAjuste2<-dlmFilter(Nile,NiloPolinomio2)
```

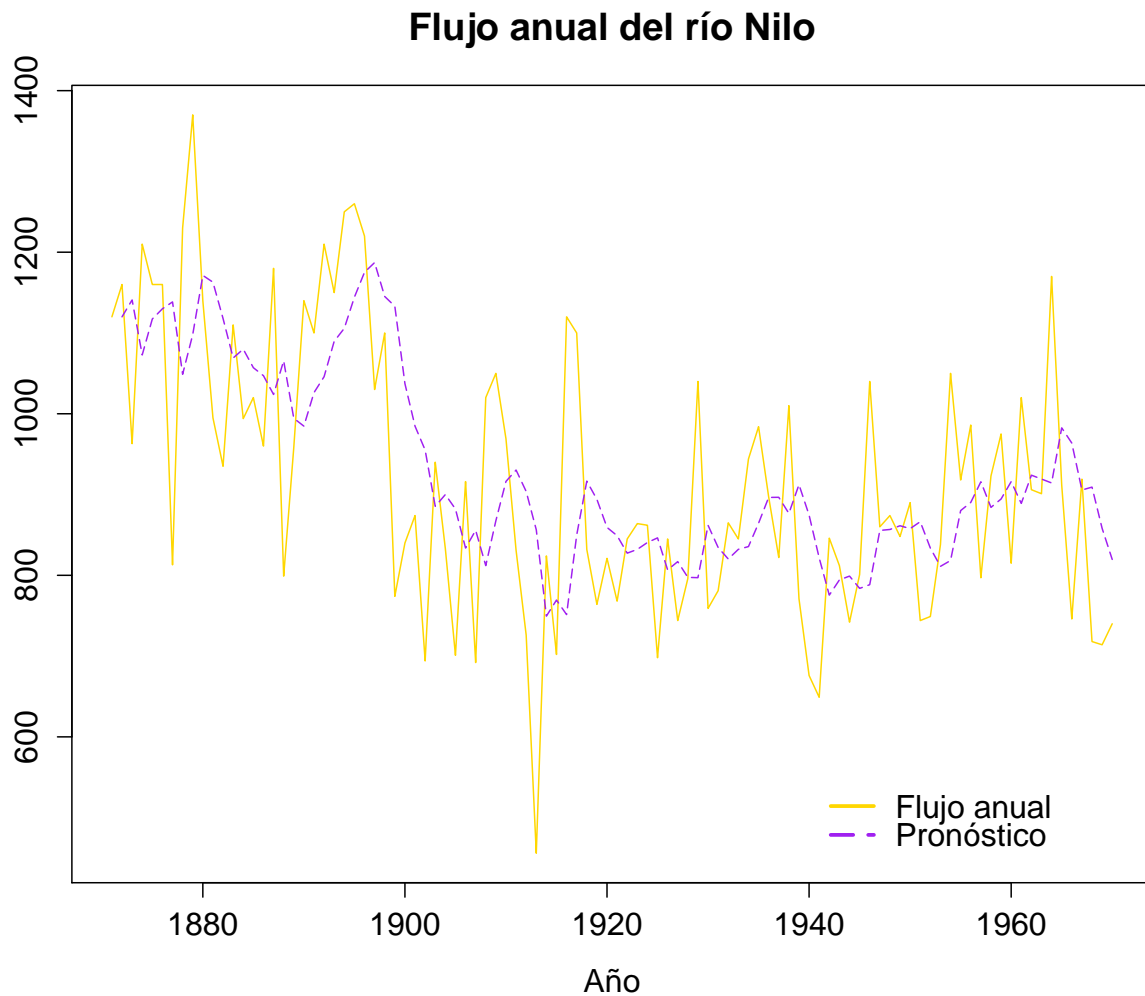


Figura 4.3.4: Figura que muestra el pronóstico a un paso para el flujo anual del río Nilo

El código para graficar ambos filtros está dado en R por:

```
> plot(Nile,pch=21,xlab="Año",ylab="",main="Flujo anual del río Nilo",col="gold")
> lines(dropFirst(NiloAjuste$m),lty="longdash",col="blue")
> lines(dropFirst(NiloAjuste2$m),lty="longdash",col="red")
> leg11<-c("Flujo anual","Filtro 1","Filtro 2")
> legend(1940,550,legend=leg11,col=c("gold","blue","red"),bty="n",
  lty=c("solid","longdash","longdash"),lwd=c(2.5,2.5,2.5))
```

El código para graficar ambos pronósticos está dado en R por:

```
> plot(Nile,pch=21,xlab="Año",ylab="",main="Flujo anual del río Nilo",col="gold")
> lines(dropFirst(NiloAjuste$f), lty="longdash",col="purple")
> lines(dropFirst(NiloAjuste2$f), lty="longdash",col="red")
> leg31<-c("Flujo anual","Pronóstico 1","Pronóstico 2")
> legend(1940,550,legend=leg31,col=c("gold","purple","red"),bty="n",
  lty=c("solid","longdash","longdash"),lwd=c(2.5,2.5,2.5))
```

La **Figura 4.3.5** muestra los datos originales junto con ambos filtros. La **Figura 4.3.6** muestra los datos originales con ambos pronósticos a un paso.

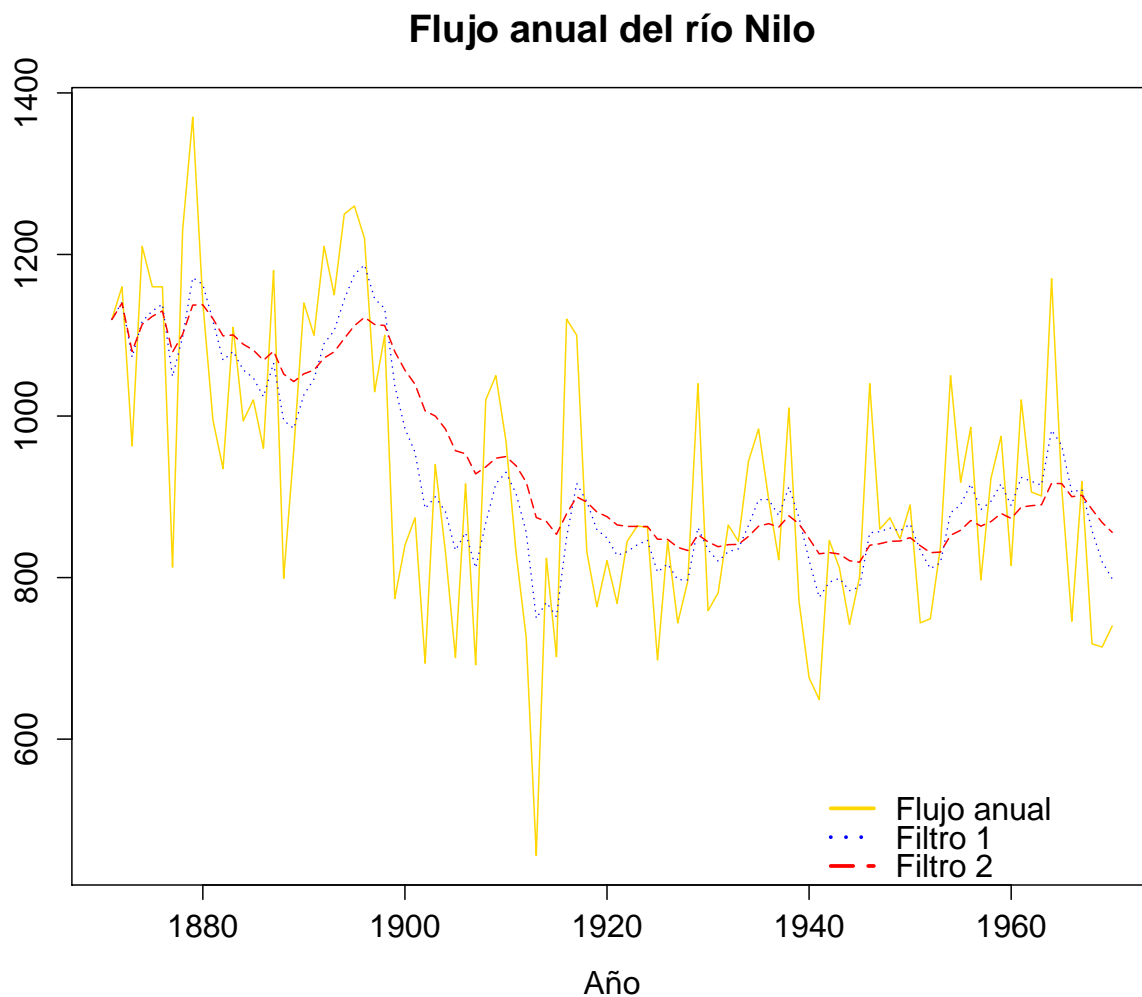


Figura 4.3.5: Figura que muestra ambos filtros para el flujo anual del río Nilo

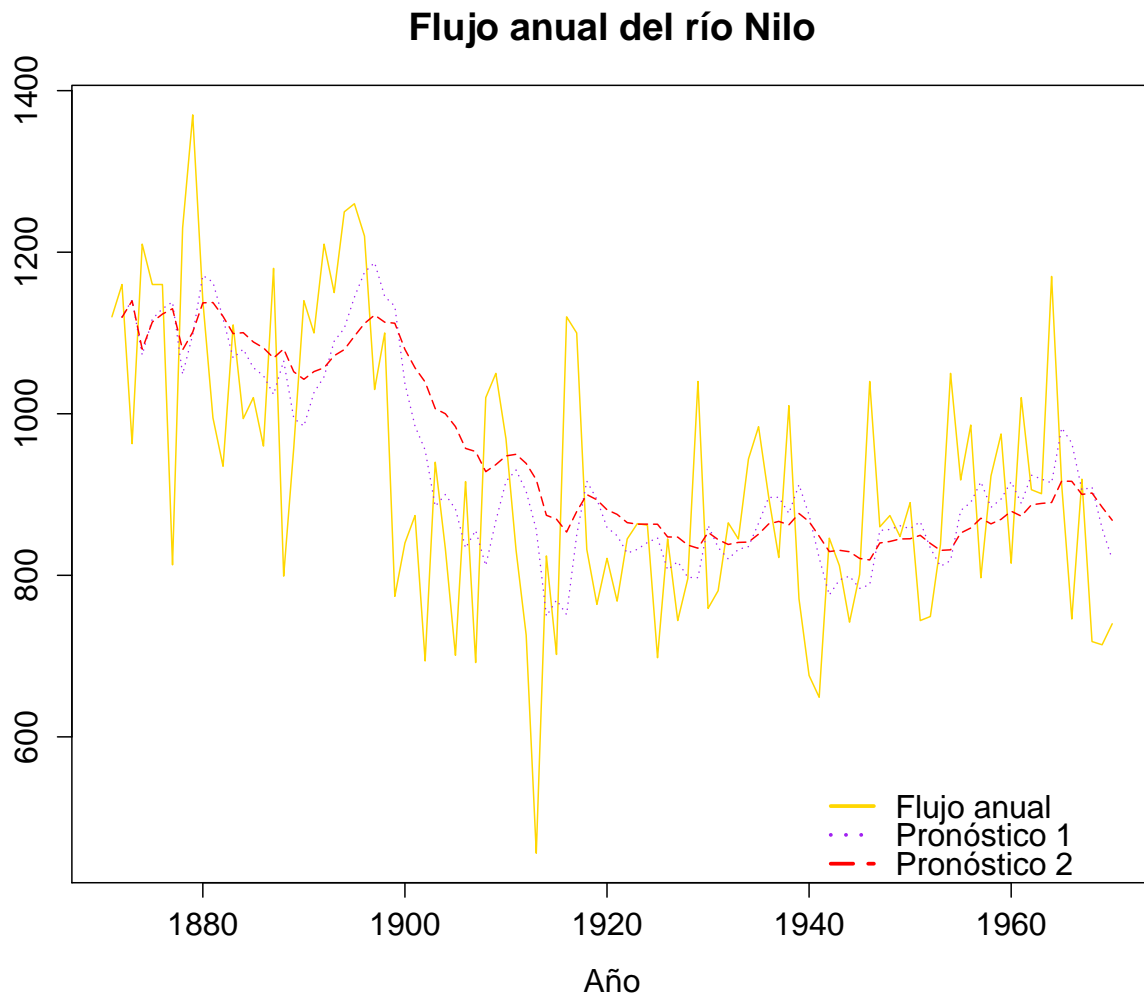


Figura 4.3.6: Figura que muestra ambos pronósticos a un paso para el flujo anual del Nilo

La definición y construcción de los residuos de los modelos dinámicos lineales excede el propósito de este escrito. Al final, después de la construcción de los mismos, se realiza el análisis de residuos de la misma forma que se ha hecho a lo largo de este trabajo. Será de nuestro interés, por ejemplo, obtener la **gráfica de probabilidad normal**, para así, encontrar desviaciones en la distribución de los residuos. El histograma también será de gran ayuda al identificar la distribución de dichos residuos. Pruebas formales para normalidad en los residuos como la prueba **Shapiro-Wilk** o **Lilliefors** son complementarias al análisis gráfico de los residuos y deberán tomarse en cuenta al concluir la distribución de los mismos.

Para el primer modelo tenemos el siguiente código en R:

```
> Residuos<-residuals(NiloAjuste)
> qqnorm(Residuos$res,xlab="Residuos",ylab="Valores Esperados",
  main="Gráfica de Probabilidad Normal",pch=19,col="blue")
> qqline(Residuos$res,xlab="Residuos",ylab="Valores Esperados",
  main="Gráfica de Probabilidad Normal",col="red")
> truehist(Residuos$res,col=rainbow(4),border="white",
  main="Histograma de los residuos",xlab="Residuos",ylab="Frecuencia
  Relativa",col.main="black",col.axis="black",col.lab="black",prob=TRUE,ylim=c(0,0.4))
> curve(dnorm(x,mean(Residuos$res),sd(Residuos$res)),add=T,col="black")
```

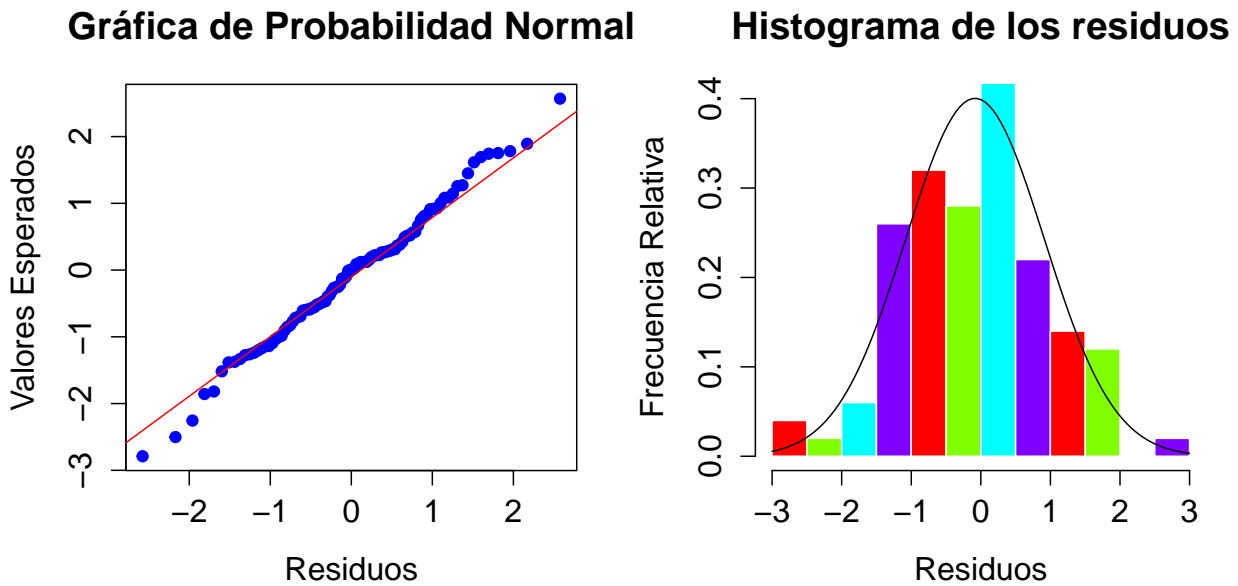



Figura 4.3.7: Figura que muestra gráficos de residuos para el primer modelo

Las pruebas formales están dadas por el siguiente código:

```
> shapiro.test(Residuos$res)
```

Shapiro-Wilk normality test

data: Residuos\$res

W = 0.9932, p-value = 0.8984

```
> lillie.test(Residuos$res)
```

Lilliefors (Kolmogorov-Smirnov) normality test

data: Residuos\$res

D = 0.0557, p-value = 0.6243

Para el segundo modelo tenemos el siguiente código en R:

```
> Residuos2<-residuals(NiloAjuste2)
> qqnorm(Residuos2$res,xlab="Residuos",ylab="Valores Esperados",
  main="Gráfica de Probabilidad Normal",pch=19,col="blue")
> qqline(Residuos2$res,xlab="Residuos",ylab="Valores Esperados",
  main="Gráfica de Probabilidad Normal",col="red")
> truehist(Residuos2$res,col=rainbow(4),border="white",
  main="Histograma de los residuos",xlab="Residuos",ylab="Frecuencia
  Relativa",col.main="black",col.axis="black",col.lab="black",prob=TRUE,ylim=c(0,0.2))
> curve(dnorm(x,mean(Residuos2$res),sd(Residuos2$res)),add=T,col="black")
```

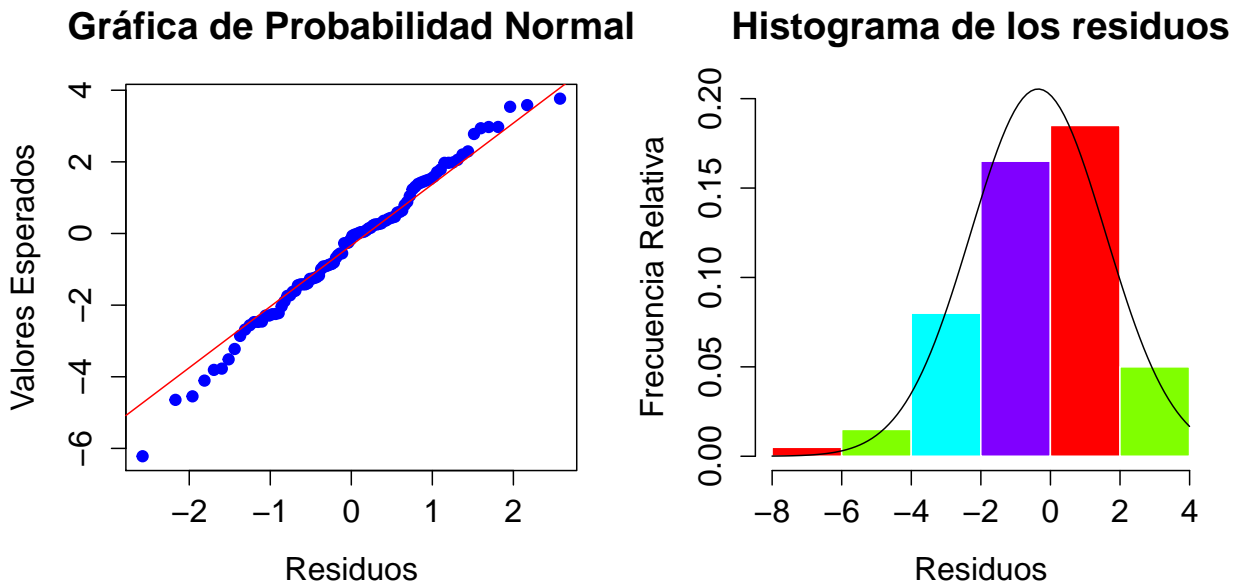


Figura 4.3.8: Figura que muestra gráficos de residuos para el segundo modelo

Las pruebas formales están dadas por el siguiente código:

```
> shapiro.test(Residuos2$res)

Shapiro-Wilk normality test

data:  Residuos2$res
W = 0.9904, p-value = 0.6934

> lillie.test(Residuos2$res)

Lilliefors (Kolmogorov-Smirnov) normality test

data:  Residuos2$res
D = 0.0594, p-value = 0.5201
```

Los residuos en ambos modelos poseen una **distribución Normal**. Esto se puede confirmar tanto en la **gráfica de probabilidad normal** como en el histograma (aunque si podemos comentar que para el segundo modelo los datos se encuentran más cargados a una de las colas de la distribución). Las pruebas de hipótesis no hacen más que confirmar la presencia de la distribución Normal en los residuos. A nuestro parecer, la diferencia entre ambos modelos estriba en que el primer modelo toma en cuenta de manera más aproximada, la variabilidad en nuestros datos. Lo anterior puede observarse claramente en la **Figura 4.3.5**. El modelo 2 lo podríamos utilizar como medio para identificar un poco la tendencia de nuestros datos ya que se enfoca más en mostrar dicha variable que en explicar la variabilidad en nuestros datos.

4.4. Modelos de regresión dinámica

Los modelos de series de tiempo nos permiten la inclusión de información de observaciones pasadas pero no nos permiten la inclusión de otra información que pueda llegar a ser relevante. Por ejemplo, los efectos de días festivos, las actividades de los competidores, cambios en las regulaciones, una expansión económica u otras variables externas pueden llegar a explicar algunas de las variaciones históricas en la serie y permitir una mejor predicción. Por otra parte, vimos en el **capítulo 1** que los modelos de regresión lineal (ya sea simple o múltiple) nos ayudan a explicar la relación que existe entre la variable respuesta y un conjunto de variables que conocemos como variables explicativas. Nuestro objetivo en esta sección será el de fusionar los modelos de regresión lineal (es decir, la información que podamos obtener de ellos) con los modelos ARIMA(p,d,q).

El **modelo de regresión lineal simple**, introducido en el **capítulo 1** es de la forma:

$$Y = \alpha + \alpha_1 X + \epsilon \quad (4.4.1)$$

Podemos, a partir de este modelo, combinar el **modelo de regresión lineal simple** con errores ARIMA(p,d,q), pero nos gustaría primero generalizarlo para los modelos de regresión lineal múltiple. El modelo regresión lineal múltiple está dado por:

$$Y_t = \alpha + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} + \epsilon_t \quad (4.4.2)$$

donde claro, Y_t es la llamada variable respuesta, $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$ (recuerde que los errores en los modelos de regresión lineal -simple o múltiple- en principio, son variables aleatorias no correlacionadas con media 0 y varianza σ_ϵ^2 aunque también podemos suponer cierta distribución en los mismos) y el conjunto $\{X_{k,i}\}$ son las k-variables explicativas del modelo en cada tiempo i . Para cada variable explicativa le corresponde un coeficiente α del conjunto $\{\alpha_k\}$, los cuales nos miden el efecto que tiene cada variable explicativa, $X_{k,i}$, después de haber considerado el efecto total de las demás variables explicativas. En palabras, los coeficientes $\{\alpha_k\}$ nos miden la aportación de cada variable explicativa al modelo en general. Para un mejor entendimiento del modelo de regresión lineal múltiple, consulte Montgomery y Peck (1992).

Ahora, con mayor razón podemos comprender por qué en el **capítulo 1** incluimos la prueba de **Durbin-Watson**, ya que ésta nos será de gran ayuda en esta sección para corroborar la correlación del conjunto $\{\epsilon_t\}$. Deseamos, a partir de ahora, que los residuos estén autocorrelacionados; para lograr este acometido, renombramos ϵ_t como ν_t de tal forma que ν_t seguirá un proceso ARIMA(p,d,q). Si, por ejemplo, suponemos que ν_t sigue un proceso ARIMA(1,1,1) podemos escribir la **ec. 4.4.2** como:

$$Y_t = \alpha + \alpha_1 X_{1,t} + \alpha_2 X_{2,t} + \dots + \alpha_k X_{k,t} + \nu_t \quad (4.4.3)$$

con

$$(1 - \phi_1 B)(1 - B)\nu_t = (1 + \theta_1 B)\epsilon_t \quad (4.4.4)$$

y $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$.

Notemos la existencia de dos errores, uno perteneciente al modelo de regresión lineal múltiple y otro perteneciente al modelo ARIMA(p,d,q). Solamente los errores ϵ_t los supondremos ruido blanco. Antes de proseguir con la estimación de los parámetros de los modelos de regresión dinámica, relacionemos estos modelos con los modelos de espacio de estados. Consideremos un ejemplo sencillo para comprender dicha relación. Tomemos un modelo de regresión lineal con una variable explicativa y errores que siguen un modelo AR(1). Este modelo particular puede ser visto como:

$$\begin{aligned} Y_t &= \alpha + \alpha_1 X_t + \nu_t \\ \nu_t &= \phi \nu_{t-1} + \epsilon_t \end{aligned}$$

donde X_t denota a la variable explicativa y el conjunto $\{\epsilon_t\}$ denota un ruido blanco. Nuestro modelo ejemplo se puede expresar en un modelo de espacio de estados de la siguiente forma:

$$\begin{aligned} Y_t &= (1 \quad X_t \quad 1) \Theta_t + \alpha_t \\ \Theta_t &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \phi \end{pmatrix} \Theta_{t-1} + \begin{pmatrix} 0 \\ 0 \\ \phi \alpha_{t-1} \end{pmatrix} \end{aligned}$$

donde $\Theta_t = (\alpha \quad \alpha_1 \quad \nu_t)'$ y $\alpha_t = \epsilon_t$. Además la varianza de $\alpha_t = \psi = \sigma^2$ y Ψ es una matriz con ceros en todas sus entradas excepto en la entrada (3,3) en donde tomaría el valor de $\psi\sigma^2$.

4.4.1. Estimación de parámetros

Recordemos que, para la estimación de los parámetros α y α_1 del modelo de regresión lineal simple, hacemos uso del método de **mínimos cuadrados ordinarios**, esto es, estimar $\hat{\alpha}$ y $\hat{\alpha}_1$ de tal manera que la suma de los cuadrados de las diferencias entre las observaciones Y_t y la línea recta estimada, $\hat{\alpha} + \hat{\alpha}_1 X_t$, sea mínima. De aquí que el criterio de mínimos cuadrados sea:

$$S(\hat{\alpha}, \hat{\alpha}_1) = \sum_{t=1}^n (Y_t - \hat{\alpha} - \hat{\alpha}_1 X_t)^2 \quad (4.4.5)$$

donde

$$Y_t = \alpha + \alpha_1 X_t + \epsilon_t, \quad t = 1, 2, \dots, n. \quad (4.4.6)$$

Recordamos también haber definido al t-ésimo residuo como la diferencia entre el valor observado Y_t y su correspondiente valor ajustado, \hat{Y}_t . Es decir, $\hat{\epsilon}_t = Y_t - \hat{Y}_t = Y_t - (\hat{\alpha} + \hat{\alpha}_1 X_t)$, $t = 1, 2, \dots, n$.

Con esto, el criterio de mínimos cuadrados se convierte simplemente en minimizar la suma de los residuos al cuadrado. Claramente, este procedimiento puede aplicarse de la misma forma para el modelo de regresión lineal múltiple dado por la **ec. 4.4.3**. Los errores a minimizar serán aquellos dados por el conjunto $\{\hat{\epsilon}_t\}$ ya que si minimizamos los errores ν_t aparecerían varios inconvenientes, ya que no estamos tomando en cuenta la autocorrelación en los errores.

Dichos inconvenientes serán los siguientes:

1. Los estimadores para el conjunto $\{\alpha_k\}$ dejan de ser los mejores estimadores lineales insesgados, ya que no toda la información ha sido tomada a consideración.
2. Pruebas para corroborar la significancia de los parámetros estimados dejan de ser correctas. Esto se traduce en que el p-value para los parámetros estimados será muy pequeño, implicando, de acuerdo con lo que sabemos, que dichos parámetros pueden ser de alta significancia al explicar al modelo, cuando en realidad no lo sean. Este evento es conocido como **regresión espuria**.
3. El criterio de información de Akaike corregido o bien el criterio de información de Akaike sin corregir dejan de ser una guía fiel para la elección del mejor modelo al momento de realizar predicciones.

Un enfoque más adecuado para modelos de regresión con errores autocorrelacionados es utilizar la estimación por mínimos cuadrados generalizada o bien la estimación por máxima verosimilitud vista en el **capítulo 1**. Los estimadores por mínimos cuadrados generalizados se obtienen al minimizar:

$$G = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \nu_i \nu_j \quad (4.4.7)$$

donde w_i, w_j son pesos relacionados con el patrón en las autocorrelaciones. Observe que en lugar de sumar sobre los errores ν_t al cuadrado, tomaremos en cuenta también los productos cruzados en los mismos.

El método de máxima verosimilitud resulta mucho más complejo pero los resultados son similares a los obtenidos con el método de mínimos cuadrados generalizado. Paquetes estadísticos calcularán dichos pesos de manera automática si nosotros especificamos un modelo ARIMA(p,d,q) inicial. Dicho modelo inicial se explica con todo detalle en la siguiente subsección.

Un punto importante a tomar en cuenta cuando estimamos un modelo de regresión lineal con errores ARIMA(p,d,q) es que todas las variables, tanto explicativas como respuesta, deben ser estacionarias; esto para que los parámetros estimados sean correctos.

Entonces, si alguna de las variables no es estacionaria, se procede a tomar diferencias (pero no solamente a dicha variable sino a todas las variables del modelo ya que deseamos mantener la relación original entre las variables explicativas y la variable respuesta) hasta volverla estacionaria.

Cuando volvemos estacionaria a una o varias variables, tomando diferencias, el modelo se conocerá como **modelo en diferencias** mientras que si trabajamos con los datos originales, es decir, todas las variables del modelo original son estacionarias, entonces el modelo se conocerá como **modelo en niveles**.

Si trabajamos con el modelo en diferencias ya no necesitamos modelar los errores con un modelo ARIMA(p,d,q) ya que, previamente, hemos tomado diferencias; entonces, sólo modelaremos a los errores con un modelo ARMA(p,q).

Tomemos nuevamente un proceso ARIMA(1,1,1) dado tanto por la **ec. 4.4.3** como por la **ec. 4.4.4**. Si tomamos diferencias, es decir, $y'_t = y_t - y_{t-1}$, $x'_{t,i} = x_{t,i} - x_{t-1,i}$ y $\nu'_t = \nu_t - \nu_{t-1}$, obtenemos el modelo dado por:

$$y'_t = \alpha_1 x'_{1,t} + \alpha_2 x'_{2,t} + \dots + \alpha_k x'_{k,t} + \nu'_t \quad (4.4.8)$$

con

$$(1 - \phi_1 B)\nu'_t = (1 + \theta_1 B)\epsilon_t \quad (4.4.9)$$

el cual es un modelo de regresión lineal en diferencias con errores ARMA(p,q).

4.4.2. Selección del modelo y pronósticos

En este apartado nos dedicaremos tanto a la selección del modelo óptimo para ajustar unos datos, como al pronóstico de los mismos. Observamos que para la selección del modelo se crea un ciclo infinito, ya que por ejemplo si deseamos calcular el conjunto $\{\alpha_k\}$ debemos primero haber identificado el modelo ARIMA(p,d,q) para los residuos $\{\nu_t\}$, pero para determinar la estructura de dichos residuos debemos primero conocer en su totalidad al conjunto $\{\alpha_k\}$. Se recomienda comenzar el proceso de ajuste con un modelo inicial AR(2) para los residuos (si no existe estacionalidad en los datos) o bien un modelo $ARIMA(2, 0, 0) \times (1, 0, 0)_s$ (si hay presencia de estacionalidad en los datos), para después estimar los coeficientes $\{\alpha_k\}$ del modelo de regresión lineal y calcular los valores de $\{\nu_t\}$, identificando un nuevo modelo ARIMA(p,d,q) apropiado para los mismos. Posteriormente, ajustamos el modelo completo usando el nuevo proceso ARIMA(p,d,q). Finalmente, lo único que nos resta será probar que el conjunto $\{\epsilon_t\}$ es un ruido blanco, es decir, $\epsilon_t \sim WN(0, \sigma_\epsilon^2)$.

En este capítulo trabajaremos exclusivamente con cuatro series. La primera de ellas nos habla del cambio porcentual en el ingreso y gasto en consumo personal en Estados Unidos desde el primer cuatrimestre del año 1970 hasta el último cuatrimestre del año 2010. La **Figura 4.4.1** muestra ambas series.

Cambios porcentuales en el ingreso y gasto en consumo en U.S.A.

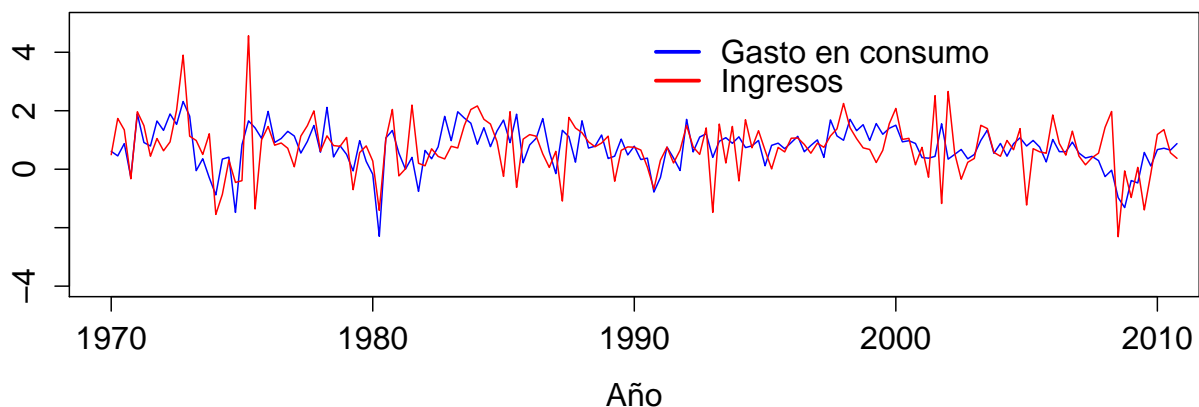


Figura 4.4.1: Figura que muestra los cambios porcentuales del ingreso y gasto en U.S.A.

La librería a utilizar es “fpp: Data for forecasting: principles and practice”. En ella podemos encontrar los datos correspondientes a nuestro primer ejemplo. La versión mínima de R a utilizar será la 3.0.3. El código correspondiente en R para la gráfica de ambas series está dado por:

```
> plot(c(1970,2010),main="Cambios porcentuales en el ingreso y gasto en consumo en
  U.S.A.",ylab="",xlab="Año",type="n",c(-4,5))
> lines(usconsumption[,1],col="blue")
> lines(usconsumption[,2],col="red")
> legend(1990,5,c("Gasto en consumo","Ingresos"),lty=c(1,1),bty="n",lwd=c(2.5,2.5),
  col=c("blue","red"))
```

El objeto correspondiente a este problema es “*us.consumption*”, donde la primer columna representa la variable gasto, mientras que la segunda hace referencia a la variable ingreso. La variable ingreso será nuestra variable explicativa mientras que la variable respuesta será el gasto.

Es claro que estas dos variables, ingreso y gasto en consumo, están altamente relacionadas. A mayor ingreso, mayor cantidad de dinero se destina al consumo de todo tipo de bienes. Pero atención, cuando aumenta o disminuye el ingreso, el cambio en el gasto en consumo no se da, a veces, de manera inmediata. Pensemos por un momento en una disminución del ingreso debido por ejemplo a la jubilación del jefe de familia después de varios años de trabajo. Claramente, la familia se tomará algunos meses en ajustar sus gastos dado el nuevo presupuesto familiar, es decir, el cambio en el gasto en consumo no se dará de manera inmediata, mientras que el cambio en el ingreso sí que lo hará. En este trabajo no entraremos en detalles como éste, pero simplemente hemos querido comentarlo. Observamos de la **Figura 4.4.1** que las series relacionadas al ingreso y gasto en consumo son ambas estacionarias y que carecen de estacionalidad. Ante esta situación podríamos bien ajustar un modelo AR(2). El código correspondiente en R para modelos de regresión dinámica con errores ARIMA(p,d,q) será:

```
> Ajuste<-Arima(usconsumption[,1], xreg=usconsumption[,2],order=c(2,0,0))
```

```
Series: usconsumption[, 1]
ARIMA(2,0,0) with non-zero mean
```

```
Coefficients:
```

	ar1	ar2	intercept	usconsumption[, 2]
	0.1325	0.2924	0.5641	0.2578
s.e.	0.0826	0.0747	0.0883	0.0530

```
sigma^2 estimated as 0.3452: log likelihood=-145.59
AIC=301.19 AICc=301.57 BIC=316.69
```

En el **capítulo 3** obtuvimos que el proceso $\{Y_t\}$ es tipo ARMA(p,q) si $\{Y_t\}$ es estacionario y se puede escribir como la **ec 3.5.1**. Este proceso está centrado en cero. Existe la posibilidad de que el proceso ARMA(p,q) no se encuentre centrado en cero por lo que sólo habría que agregar una constante μ a la **ec 3.5.1**. Ahora bien, el proceso ARIMA(p,d,q) está dado por $Y'_t - \mu - \phi_1 Y'_{t-1} - \dots - \phi_p Y'_{t-p} = \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q}$, donde $\{Y'_t\}$ es la serie diferenciada (una o más veces).

¿Por qué mencionamos lo anterior y hacemos énfasis en la constante μ ? En el **capítulo 3** utilizamos la función `arima(.)` para ajustar un modelo a unos datos. En este capítulo usaremos la función `Arima(.)` para lograr nuestros objetivos. ¿En qué radica la diferencia?

La función $\text{arima}(\cdot)$ establece $\mu = 0$ cuando $d > 0$ y nos brinda un estimado del parámetro μ cuando $d = 0$. El parámetro μ es muy cercano a la media muestral de la serie de tiempo pero no es idéntico ya que la media muestral no es el estimador máximo verosímil cuando $p + q > 0$. Dentro de las opciones de la función $\text{arima}(\cdot)$ tenemos `include.mean`, el cual sólo hace efecto cuando $d = 0$. Escribir `include.mean=FALSE` forzará a que $\mu = 0$. Cuando $d > 0$ en general da igual centrar o no la serie; por esto la función $\text{arima}(\cdot)$ por default considera la media de cero, ya que estamos eliminando la constante al aplicar el operador de diferencia. Ahora bien, la función $\text{Arima}(\cdot)$ ajustará un modelo de regresión con errores ARIMA(p,d,q) si el argumento `xreg` es utilizado. La función $\text{Arima}(\cdot)$ muestra mayor flexibilidad al incluir la constante μ . Sigue incluyendo la opción `include.mean` pero anexa la opción `include.drift` que nos permite que el intercepto de la regresión (deriva o *drift*) sea distinto de cero cuando $d = 1$. Esto no sucede con la función $\text{arima}(\cdot)$. Adicionalmente, existe la función `auto.arima` (utilizada más adelante) la cual automáticamente incluye la constante μ . Si $d = 0$ o $d = 1$, dicha constante se incluirá si ayuda a mejorar el criterio de Akaike; para $d > 1$ la constante siempre se omite. Del código tenemos los estimadores para el intercepto, $\alpha = 0.5641$, para la pendiente, $\alpha_1 = 0.2578$ (sólo tenemos 1 variable explicativa, el ingreso; de aquí que, estamos utilizando el modelo de regresión lineal simple), para $\phi_1 = 0.1325$ y para $\phi_2 = 0.2924$. El modelo, por tanto, podría ser descrito como:

$$\begin{aligned} y_t &= 0.5641 + 0.2578x_{1,t} + \nu_t \\ \nu_t &= 0.1325\nu_{t-1} + 0.2924\nu_{t-2} + \epsilon_t \\ \epsilon_t &\sim NID(0, 0.3452) \end{aligned} \quad (4.4.10)$$

donde $x_{1,t}$ corresponde al ingreso al tiempo t . La gráfica de residuos se muestra en la **Figura 4.4.2**. La sucesión de autocorrelaciones muestrales (ACF) y parciales (PACF) se muestran en la **Figura 4.4.3** y en la **Figura 4.4.4** respectivamente.

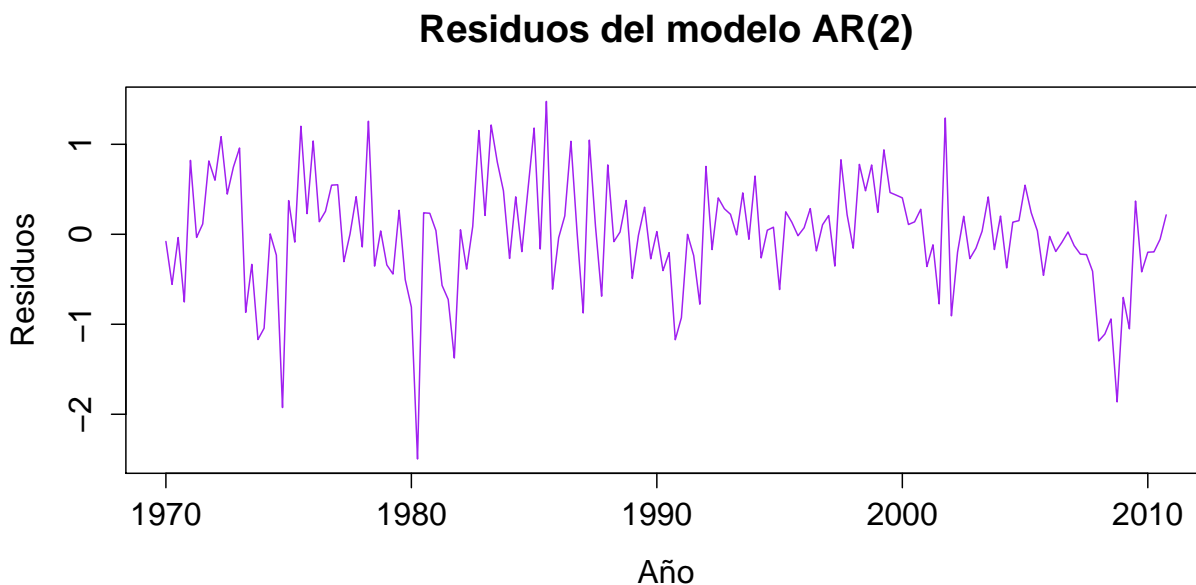


Figura 4.4.2: Figura que muestra los residuos del proceso AR(2)

Lo que sigue sería, a partir del modelo AR(2), buscar un nuevo modelo ARIMA(p,d,q) que se ajuste mejor a los datos, en el sentido de que posea el menor criterio de información de Akaike corregido o bien el menor criterio de información de Akaike sin corregir.

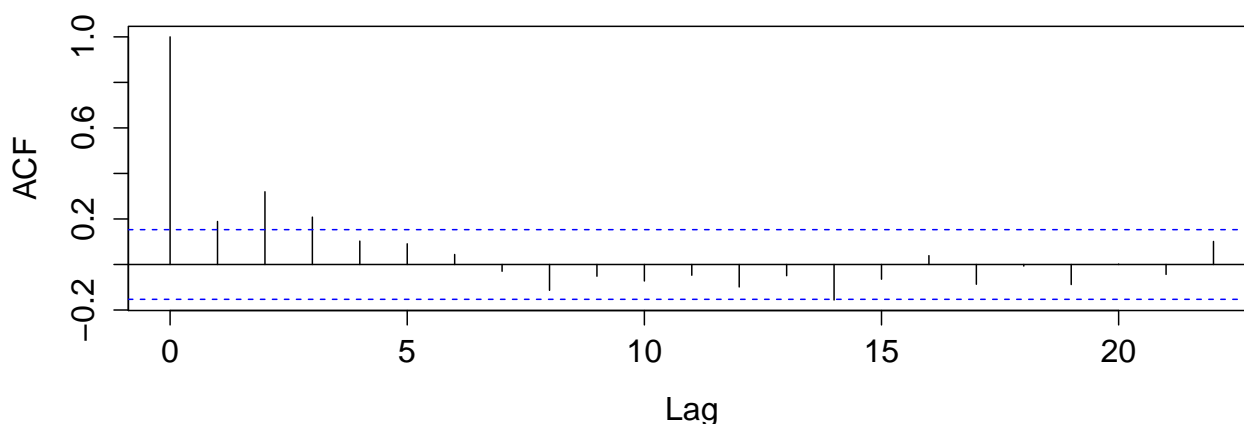
ACF del proceso AR(2)

Figura 4.4.3: Figura que muestra la sucesión de autocorrelaciones muestrales (ACF)

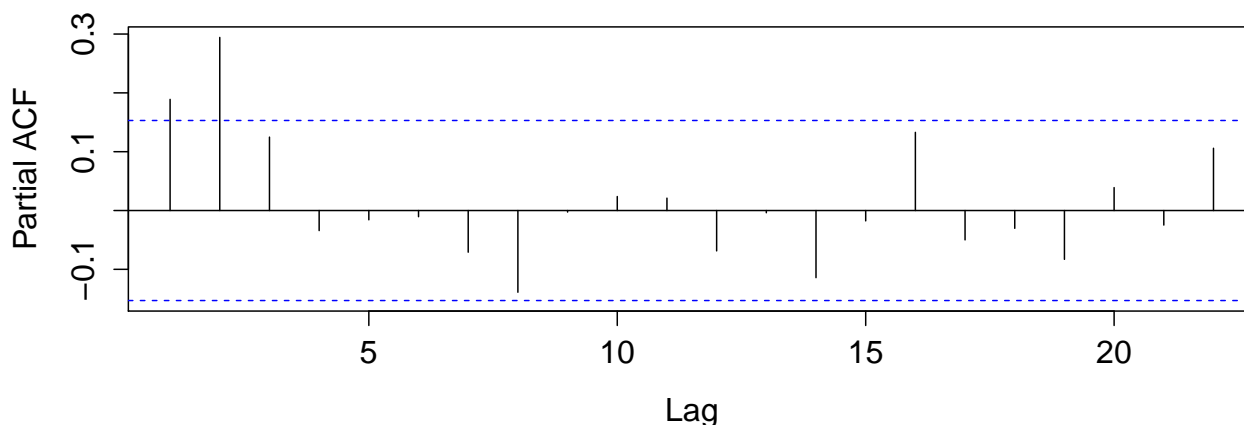
PACF del proceso AR(2)

Figura 4.4.4: Figura que muestra la sucesión de autocorrelaciones parciales (PACF)

Para lograr esto, podemos correr, una y otra vez, el código anterior, con las distintas combinaciones de modelos $ARIMA(p,d,q)$ y quedarnos con aquel que nos muestre el menor criterio de Akaike. De la **Figura 4.4.3** y la **Figura 4.4.4** podemos observar algunas opciones y combinaciones para el modelo $ARIMA(p,d,q)$ óptimo, a saber unos ejemplos, $AR(2)$, $MA(3)$, $ARMA(2,3)$, etc (por supuesto, pueden existir otras opciones y dentro de ellas podría estar el modelo óptimo). Recordemos que a partir de la sucesión de autocorrelaciones muestrales parciales (PACF) se selecciona el parámetro p del modelo $ARMA(p,q)$ mientras que de la sucesión de autocorrelaciones muestrales (ACF) el parámetro q . Recordemos asimismo que el valor de los parámetros dependerá de aquellos retrasos que se salgan de las bandas de confianza correspondientes.

Ahora bien, para cada uno de los modelos podemos utilizar el código anterior y ajustar dicho modelo a los datos y al final ver cuál de todos tiene menor criterio de Akaike. Dicha tarea, puede simplificarse utilizando el siguiente código en R, el cual, a través de la función `auto.arima`, nos dice automáticamente qué modelo posee menor criterio de Akaike (no necesariamente será uno de los propuestos anteriormente):

```
> Ajuste2<-auto.arima(usconsumption[,1], xreg=usconsumption[,2])
```

```
Series: usconsumption[, 1]
```

```
ARIMA(1,0,2) with non-zero mean
```

```
Coefficients:
```

	ar1	ma1	ma2	intercept	usconsumption[, 2]
	0.6516	-0.5440	0.2187	0.5750	0.2420
s.e.	0.1468	0.1576	0.0790	0.0951	0.0513

```
sigma^2 estimated as 0.3396: log likelihood=-144.27
```

```
AIC=300.54 AICc=301.08 BIC=319.14
```

Dentro de los parámetros a utilizar por la función `auto.arima()` se encuentra `xreg` el cual representa a las variables explicativas del modelo. Pero, ¿Cómo trabaja la función `auto.arima()`? Esta función se basa en el algoritmo creado por Hyndman-Khandakar donde, primero, el número de diferencias se determina usando la prueba KPSS (recordemos que esta prueba fue introducida por Kwiatkowski, Phillips, Schmidt y Shin; dicha prueba fue introducida en el **capítulo 3**). Acto seguido, se encuentran los valores de p y q que minimizan el criterio de Akaike una vez diferenciado la serie d veces. La función selecciona uno de cuatro modelos iniciales, a saber, $ARIMA(2,d,2)$, $ARIMA(0,d,0)$, $ARIMA(1,d,0)$, $ARIMA(0,d,1)$. A este modelo se le conocerá como “modelo actual”. Después se consideran variaciones de p y q en ± 1 o bien la inclusión o exclusión del parámetro μ , en el modelo actual para ver si existe algún modelo distinto a los iniciales que minimice el criterio de Akaike. Este último paso se repite hasta encontrar el modelo óptimo. Retomando nuestro ejemplo, el modelo óptimo para los errores es el modelo $ARIMA(1,0,2)$, es decir:

$$\begin{aligned}
 y_t &= 0.5750 + 0.2420x_{1,t} + \nu_t \\
 \nu_t &= 0.6516\nu_{t-1} + \epsilon_t - 0.5440\epsilon_{t-1} + 0.2187\epsilon_{t-2} \\
 \epsilon_t &\sim NID(0, 0.3396)
 \end{aligned}
 \tag{4.4.11}$$

Podemos, adicionalmente, agregar la prueba de Ljung y Box para autocorrelaciones en los residuos, $\{\epsilon_t\}$. Observamos que el p -value = 0.4673 > 0.05, por lo cual no rechazamos la hipótesis nula H_0 y concluimos que las autocorrelaciones hasta el “lag” $h = 10$ conjuntamente son cero. El código se muestra a continuación:

```
> Box.test(residuals(Ajuste2), fitdf=5, lag=10, type="Ljung")
```

```
Box-Ljung test
```

```
data: residuals(Ajuste2)
```

```
X-squared = 4.5948, df = 5, p-value = 0.4673
```

Pasemos ahora, a la parte de pronósticos. Es claro que para pronosticar un modelo de regresión dinámica debemos pronosticar tanto el modelo de regresión lineal (simple o múltiple) como el proceso $ARIMA(p,d,q)$ que mejor se ajuste a los errores del modelo de regresión, para después juntarlos. Para el modelo de regresión lineal, debemos predecir las variables explicativas. Cuando la variable explicativa es el tiempo o alguna otra variable que podamos conocer de antemano, el proceso se facilita.

Pero ¿qué sucede con variables explicativas, como el ingreso? Claramente es más complicado conocer el comportamiento del ingreso que el comportamiento del tiempo, el cual sí que es conocido. Entonces, nuestro objetivo ahora será el de predecir el ingreso. Una opción muy sencilla para predecir el comportamiento del ingreso es suponer que el cambio porcentual en el ingreso es igual al cambio porcentual medio de toda la serie. A partir de aquí, el pronóstico para los próximos 3 años (12 cuatrimestres) está dado por:

```
> Pred2<-forecast(Ajuste2,xreg=rep(mean(usconsumption[,2]),12),h=12)
> plot(Pred2,main="Pronóstico para el modelo ARIMA(1,0,2) en los residuos",xlab="Año")
```

La **Figura 4.4.5** muestra dicho pronóstico.

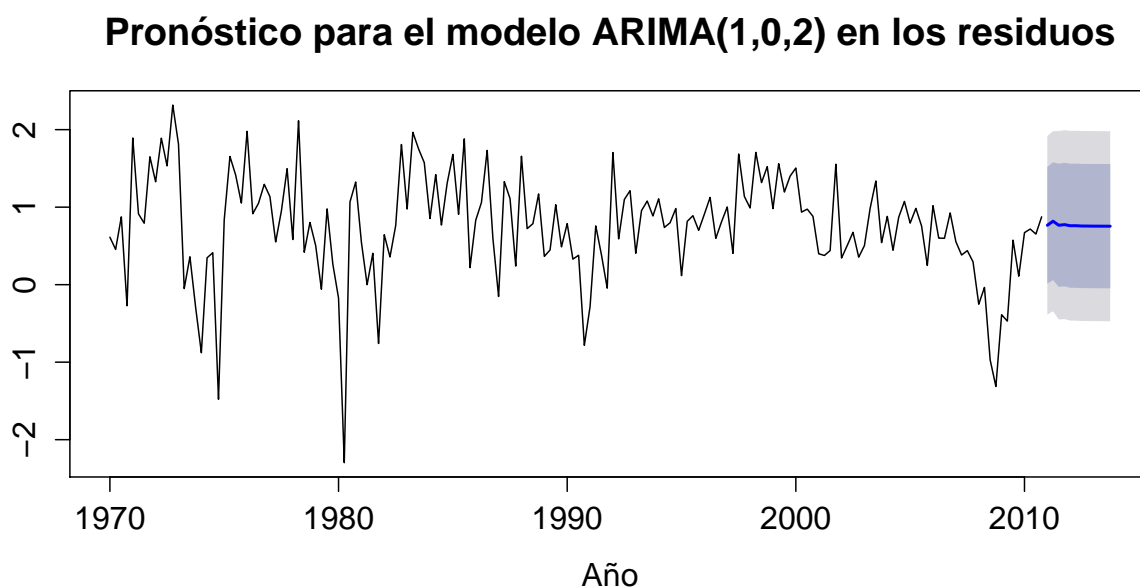


Figura 4.4.5: Figura que muestra el pronóstico del ingreso para los próximos 3 años

Si a nosotros se nos hubiera ocurrido pronosticar el ingreso con un modelo $ARIMA(p,d,q)$ y esos pronósticos utilizarlos para predecir la variable consumo, hubiésemos obtenido el mismo resultado, ya que el modelo óptimo para la variable ingreso es un modelo $ARIMA(0,0,0)$ o bien, un ruido blanco con media distinta a cero. En otras palabras, la variable ingreso tiene el comportamiento de un ruido blanco por lo cual no es posible modelarla a través de un proceso ARMA. De aquí que, el pronóstico para la variable ingreso será, en efecto, el mismo, es decir, el ingreso promedio de toda la serie. El código correspondiente será:

```
> Ajuste3<-auto.arima(usconsumption[,2])
```

```
Series: usconsumption[, 2]
ARIMA(0,0,0) with non-zero mean
```

```
Coefficients:
  intercept
      0.7366
s.e.      0.0732
```

```
sigma^2 estimated as 0.879: log likelihood=-222.13
AIC=448.26 AICc=448.33 BIC=454.46
```

La siguiente instrucción se usa para obtener los pronósticos para h periodos:

```
> Pred3<-forecast(Ajuste3,h=12)

      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
2011 Q1      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2011 Q2      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2011 Q3      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2011 Q4      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2012 Q1      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2012 Q2      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2012 Q3      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2012 Q4      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2013 Q1      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2013 Q2      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2013 Q3      0.7365571 -0.4649583 1.938073 -1.101002 2.574116
2013 Q4      0.7365571 -0.4649583 1.938073 -1.101002 2.574116

> mean(usconsumption[,2])
[1] 0.7365571
```

Recordemos que en el **capítulo 2** realizamos un análisis completo de la componente de tendencia ajustando finalmente una función del tipo $Y_t = f(t) + \epsilon_t$ a dicha componente, donde $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. Ahora, hablaremos un poco acerca de dos formas alternativas de modelar tendencia. Una de ellas es la tendencia determinista y la otra corresponde a la tendencia estocástica. Una tendencia determinista está dada por el modelo:

$$y_t = \alpha + \alpha_1 t + \nu_t \quad (4.4.12)$$

con ν_t proceso ARMA(p,q). Una tendencia estocástica está dada por el modelo:

$$y_t = \alpha + \alpha_1 t + \nu_t \quad (4.4.13)$$

con ν_t proceso ARIMA(p,1,q).

Recordemos del **capítulo 2** además, la serie correspondiente al comportamiento del PIB trimestral en México. Esta será nuestra segunda serie a trabajar en este capítulo. Los valores de la serie están dados a precios constantes con base en el año 2008. Las cifras datan de Enero de 1993 al tercer trimestre del año 2013. En el **capítulo 2** estimamos la tendencia de esta serie mediante el uso del modelo de regresión lineal simple. Hasta ese punto llegó nuestro estudio. Ahora realizaremos un ajuste diferente, suponiendo a los errores de dicho modelo, un proceso con tendencia tanto determinista como estocástica. Realizaremos un pronóstico para los próximos 3 años. Primero realizaremos la prueba Dickey-Fuller aumentada vista en el **capítulo 3** para corroborar la existencia de estacionariedad en la serie. Recordemos que la prueba Dickey-Fuller aumentada tiene como hipótesis:

$$H_0 : \text{La serie no es estacionaria} \quad \text{vs} \quad H_1 : \text{La serie presenta estacionariedad}$$

Un valor grande en el p-value es indicio de no estacionariedad mientras que un valor pequeño en el p-value indica la presencia de estacionariedad.

Usando $\alpha = 5\%$, se requerirá el uso del operador de diferencia (∇^j), visto en la **Sección 3.2.3**, para hacer estacionaria a la serie si p-value > 0.05 . Análogamente, realizaremos la prueba KPSS para estacionariedad. Recordamos, similarmente, que dicha prueba tiene como hipótesis:

$$H_0 : \text{La serie es estacionaria} \quad vs \quad H_1 : \text{La serie no presenta estacionariedad}$$

Un valor pequeño para el p-value implica que la serie no es estacionaria y que entonces es necesaria la aplicación del operador ∇^j . ¿Pero cuántas diferencias debemos aplicar? Es decir, ¿Cuál es el valor de j en el operador ∇^j ? Estas dos pruebas no nos dicen el número de veces a diferenciar. R provee dos funciones, a saber `ndiffs(.)` y `nsdiffs(.)`, las cuales nos determinan el número apropiado de diferencias para la parte no estacional y estacional de la serie, respectivamente. La función `ndiffs(.)` permite utilizar la prueba Dickey-Fuller aumentada o la prueba KPSS para determinar el número adecuado de diferencias a utilizar para la parte no estacional. La función `nsdiffs(.)` permite utilizar la prueba Osborn-Chui-Smith-Birchenhall para determinar el número óptimo de diferencias para la parte estacional de la serie. R ajustará de manera automática un modelo $ARIMA(p,d,q)$ o bien $ARIMA(p,d,q) \times (P,D,Q)_s$ óptimo, esto es, nos proporciona el valor de s óptimo en el operador ∇_s . Presentaremos finalmente el pronóstico para los próximos 3 años así como la prueba de Ljung y Box para autocorrelaciones.

Comenzamos con la tendencia determinista. Dentro de los parámetros a utilizar se encuentra `xreg`, que en este caso tendrá como variable explicativa al tiempo t de cada observación. El código en R para dicha tendencia está dado por las siguientes instrucciones:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 4/Ejercicios")
> PIBmex<-read.table("PIB a precios corrientes(MRD).txt",header=T)
> Serie<-ts(PIBmex, frequency = 4, start = c(1993, 1))
> kpss.test(Serie)
```

KPSS Test for Level Stationarity

```
data: Serie
KPSS Level = 2.8305, Truncation lag parameter = 2, p-value = 0.01
```

```
> adf.test(Serie, alternative = "stationary")
```

Augmented Dickey-Fuller Test

```
data: Serie
Dickey-Fuller = -3.7405, Lag order = 4, p-value = 0.02628
alternative hypothesis: stationary
```

```
> ndiffs(Serie,alpha=0.05,test=c("kpss"),max.d=2)
```

```
[1] 1
```

```
> nsdiffs(Serie,m=frequency(Serie), test=c("ocsb"), max.D=1)
```

```
[1] 1
```

Pasamos a continuación a ajustar el modelo óptimo, trabajo que dejaremos hacer a R. El análisis del código anterior y el correspondiente al ajuste se muestran al finalizar el siguiente código. Para ajustar el modelo óptimo tenemos:

```
> auto.arima(Serie,xreg=1:length(Serie),trace=FALSE,test=c("kpss"),stepwise=FALSE,
  approximation=FALSE,max.order=9,allowdrift=TRUE)
```

```
Series: Serie
ARIMA(2,0,0)(0,1,1)[4]
```

```
Coefficients:
      ar1      ar2      sma1  1:length(Serie)
      1.2195 -0.3652 -0.5431      181190.4
s.e.  0.1072  0.1109  0.1014      16059.2
```

```
sigma^2 estimated as 31486760360: log likelihood=-1012.35
AIC=2034.69  AICc=2035.51  BIC=2046.54
```

```
> Ajuste<-Arima(Serie,xreg=1:length(Serie),order=c(2,0,0), seasonal=c(0, 1, 1),
  include.constant=TRUE)
```

```
Series: Serie
ARIMA(2,0,0)(0,1,1)[4] with drift
```

```
Coefficients:
      ar1      ar2      sma1      drift
      1.2195 -0.3652 -0.5431  181190.4
s.e.  0.1072  0.1109  0.1014  16059.2
```

```
sigma^2 estimated as 3.149e+10: log likelihood=-1012.35
AIC=2034.69  AICc=2035.51  BIC=2046.54
```

```
> forecast(Ajuste,h=12)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2013 Q4	16757089	16529684	16984494	16409303	17104875
2014 Q1	16363276	16004642	16721911	15814792	16911760
2014 Q2	16579550	16139421	17019679	15906431	17252670
2014 Q3	16767495	16279895	17255095	16021775	17513215
2014 Q4	17455884	16900144	18011624	16605953	18305815
2015 Q1	17068149	16459375	17676923	16137109	17999189
2015 Q2	17289541	16646959	17932123	16306797	18272286
2015 Q3	17481508	16819289	18143727	16468731	18494284
2015 Q4	18172931	17473861	18872002	17103795	19242068
2016 Q1	17787429	17056137	18518720	16669015	18905842
2016 Q2	18010435	17257317	18763554	16858640	19162230
2016 Q3	18203555	17437331	18969778	17031717	19375392

```
> Box.test(residuals(Ajuste),fitdf=2,lag=22,type="Ljung")
```

```
Box-Ljung test
data: residuals(Ajuste)
X-squared = 24.2397, df = 20, p-value = 0.2321
```

Observamos de la prueba KPSS la necesidad de aplicar diferencias a nuestra serie (ya que el p -value = $0.01 < 0.05$ y por tanto rechazamos H_0 , es decir, la serie no es estacionaria), mientras que la prueba Dickey-Fuller aumentada nos dice que la serie ya es estacionaria (p -value < 0.05).

Tenemos dos pruebas que miden lo mismo y que envían resultados distintos. Ante esta situación debemos tomar en cuenta la gráfica de los datos, las salidas de los comandos `ndiffs(.)` y `nsdiffs(.)` (este último nos indica una diferencia estacional, la cual coincide con lo obtenido con la función `auto.arima`) y nuestra intuición para tomar la mejor decisión.

Si de nosotros dependiera, la serie carece de estacionariedad por el efecto tendencia, por lo que aplicaríamos el operador diferencia para eliminar dicha componente. Pero, como ya hemos dicho, hemos dejado que R ajuste el modelo óptimo, tomando a consideración las pruebas anteriores. Finalmente, el software recomienda la aplicación, de una diferencia estacional. El modelo óptimo es $ARIMA(2, 0, 0) \times (0, 1, 1)_4$. Podemos observar, para la prueba de Ljung y Box, que el p -value = $0.2321 > 0.05$, por lo cual concluimos que las autocorrelaciones hasta el “lag” $h = 22$ conjuntamente son cero. El pronóstico para los próximos 3 años se muestra dentro del código y su correspondiente gráfica se realizará a la par de la gráfica de pronósticos con tendencia estocástica.

Pasemos al cálculo de la tendencia estocástica. El código en R será:

```
> auto.arima(Serie,d=1,xreg=1:length(Serie),trace=FALSE,stepwise=FALSE,
  approximation=FALSE,max.order=9,allowdrift=TRUE)

Series: Serie
ARIMA(1,1,0)(0,1,2)[4]

Coefficients:
      ar1      sma1      sma2  1:length(Serie)
      0.2922 -0.6321  0.0816          192796.1
s.e.  0.1102  0.1179  0.1133              NaN

sigma^2 estimated as 32855834902:  log likelihood=-1055.98
AIC=2121.96  AICc=2122.79  BIC=2133.74

> Ajuste2<-Arima(Serie,xreg=1:length(Serie),order=c(1,1,0), seasonal=c(0,1,2),
  include.drift=TRUE)

> forecast(Ajuste2,h=12)

      Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
2013 Q4      16813075 16580778 17045371 16457808 17168341
2014 Q1      16464007 16084449 16843565 15883523 17044491
2014 Q2      16687300 16190846 17183755 15928039 17446562
2014 Q3      16896805 16302994 17490616 15988649 17804960
2014 Q4      17631058 16907773 18354344 16524888 18737228
2015 Q1      17257935 16412149 18103721 15964417 18551452
2015 Q2      17485384 16529275 18441493 16023142 18947627
2015 Q3      17691041 16635163 18746919 16076214 19305868
2015 Q4      18424170 17232200 19616140 16601209 20247131
2016 Q1      18050719 16723463 19377974 16020857 20080580
2016 Q2      18278072 16824465 19731679 16054972 20501172
2016 Q3      18483701 16912888 20054514 16081349 20886052
```

```
> Box.test(residuals(Ajuste2),fitdf=2,lag=22,type="Ljung")
```

Box-Ljung test

```
data: residuals(Ajuste2)
```

```
X-squared = 25.3743, df = 20, p-value = 0.1875
```

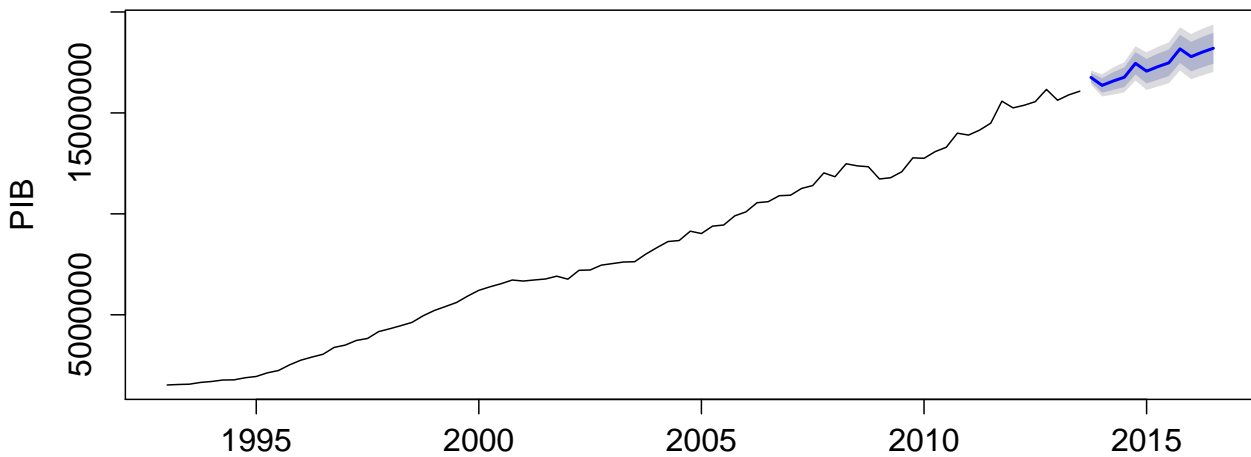
El modelo óptimo para la tendencia estocástica es $ARIMA(1, 1, 0) \times (0, 1, 2)_4$. La **Figura 4.4.6** muestra el pronóstico correspondiente. El código en R está dado por:

```
> par(mfrow=c(2,1)); options(scipen=3)
```

```
> plot(forecast(Ajuste,h=12),main="Pronóstico para el PIB mexicano",
      xlab="(a) Tendencia determinista",ylab="PIB")
```

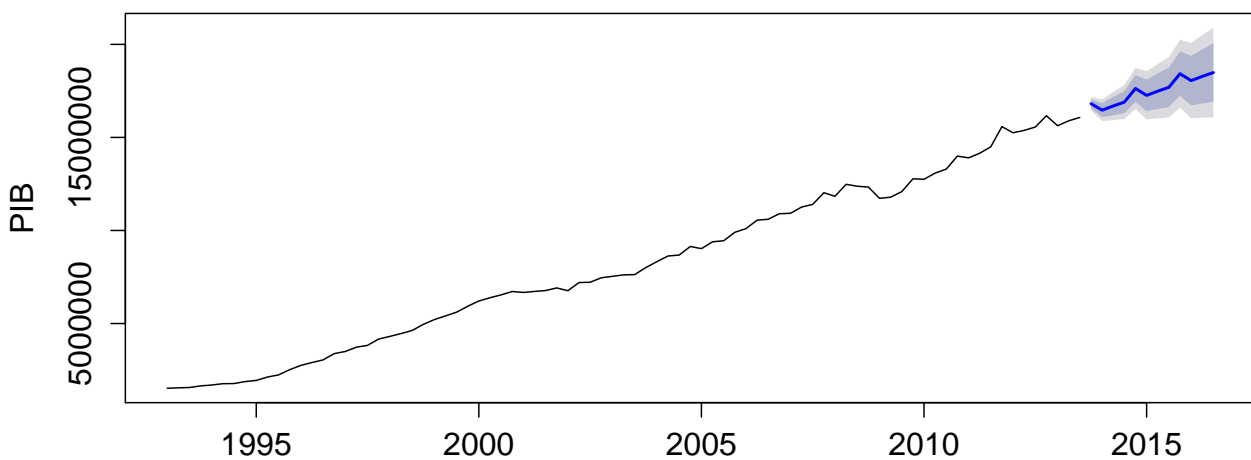
```
> plot(forecast(Ajuste2,h=12),main="Pronóstico para el PIB mexicano",
      xlab="(b) Tendencia estocástica",ylab="PIB")
```

Pronóstico para el PIB mexicano



(a) Tendencia determinista

Pronóstico para el PIB mexicano



(b) Tendencia estocástica

Figura 4.4.6: Figura que muestra el pronóstico para el PIB mexicano para los próx. 3 años

Algunas observaciones para estas dos tendencias de este ejemplo salen a la luz. Primero, cuando ajustamos un modelo para la tendencia determinista el número de diferencias, en conjunto -estacionales y no estacionales- es 1, por lo que las funciones `Arima(.)` y `auto.arima(.)` permiten la inclusión de una constante si esta ayuda a mejorar el criterio de Akaike. En nuestro ejemplo sí que se incluye un término constante (en el código podemos observar `drift = 181190.4`). Como segunda observación, notemos que el número de diferencias en conjunto para el ajuste del modelo de tendencia estocástica es 2, por lo que las funciones antes mencionadas omiten el valor de la constante.

La tercer serie a analizar será aquella correspondiente al PIB trimestral (millones de dólares) de E.U.A a partir del año 1947 hasta el año 1991. Esta serie también fue analizada (se calculó la componente de tendencia) en el **capítulo 2**. Realizaremos un ajuste de un modelo de tendencia estocástica en los datos, así como su correspondiente pronóstico para parte del año 1991 y los años 1992-1994. La prueba KPSS y la prueba Dickey-Fuller aumentada, ambas, concluyen en la necesidad de aplicar diferencias para volver estacionaria la serie. De la gráfica del PIB trimestral en Estados Unidos podemos observar, y esto se corrobora con el comando `ndiffs(.)`, la necesidad de aplicar solamente una diferencia para la parte no estacional. También observamos que no hay necesidad de aplicar diferencias para la parte estacional. Esto lo corroboramos con el comando `nsdiffs(.)`. El código en R para este proceso está dado por:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 4/Ejercicios")
> PIBusa<-read.table("Usgnp(MRD).txt",header=T)
> Serie2<-ts(PIBusa, frequency = 4, start = c(1947, 1))
> kpss.test(Serie2)

KPSS Test for Level Stationarity

data: Serie2
KPSS Level = 4.4492, Truncation lag parameter = 3, p-value = 0.01

> adf.test(Serie2,alternative="stationary")

Augmented Dickey-Fuller Test

data: Serie2
Dickey-Fuller = -2.3099, Lag order = 5, p-value = 0.4467
alternative hypothesis: stationary

> ndiffs(Serie2,alpha=0.05,test=c("kpss"),max.d=2)

[1] 1

> nsdiffs(Serie2,m=frequency(Serie2), test=c("ocsb"), max.D=2)

[1] 0

> Ajuste4<-auto.arima(Serie2,d=1,xreg=1:length(Serie2),trace=FALSE,test=c("kpss"),
  stepwise=FALSE,approximation=FALSE,max.order=9,allowdrift=TRUE)

Series: Serie2
ARIMA(2,1,0) with drift
```

Coefficients:

```

      ar1      ar2      drift
0.3005  0.1540  17.1097
s.e.  0.0748  0.0751  3.1964

```

```

sigma^2 estimated as 540.9:  log likelihood=-798.57
AIC=1605.14  AICc=1605.37  BIC=1617.82

```

```
> forecast(Ajuste4,h=12)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1991 Q2	4125.167	4095.361	4154.973	4079.583	4170.752
1991 Q3	4129.960	4081.062	4178.858	4055.177	4204.744
1991 Q4	4140.530	4073.364	4207.697	4037.808	4243.253
1992 Q1	4153.778	4070.273	4237.283	4026.068	4281.488
1992 Q2	4168.719	4070.447	4266.991	4018.425	4319.013
1992 Q3	4184.582	4072.920	4296.245	4013.810	4355.355
1992 Q4	4200.984	4077.071	4324.896	4011.475	4390.492
1993 Q1	4217.688	4082.473	4352.903	4010.895	4424.482
1993 Q2	4234.567	4088.840	4380.295	4011.696	4457.438
1993 Q3	4251.545	4095.969	4407.121	4013.612	4489.478
1993 Q4	4268.580	4103.718	4433.442	4016.445	4520.715
1994 Q1	4285.647	4111.981	4459.312	4020.048	4551.245

```
> Box.test(residuals(Ajuste4),fitdf=2,lag=22,type="Ljung")
```

Box-Ljung test

data: residuals(Ajuste4)

X-squared = 16.7376, df = 20, p-value = 0.6699

El modelo óptimo es ARIMA(2,1,0). De la prueba de Ljung y Box, podemos observar que el $p\text{-value} = 0.6699 > 0.05$, por lo cual concluimos que las autocorrelaciones hasta el “lag” $h = 22$ conjuntamente son cero. La **Figura 4.4.7** muestra nuestro pronóstico.

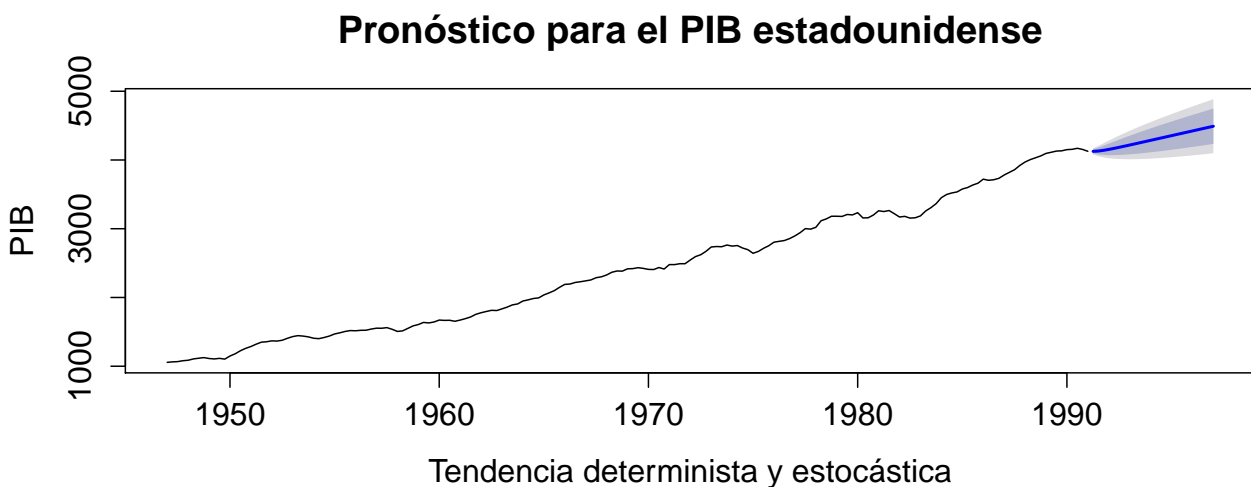


Figura 4.4.7: Figura que muestra el pronóstico para el PIB de U.S.A para años 1991-1994

Por último veamos el comportamiento de la cuarta serie, a saber, aquella correspondiente al índice de precios al consumo (IPC) a partir de enero del año 2000 y hasta marzo del año 2014. Dicha serie fue introducida al inicio de este capítulo y posee una periodicidad mensual. Realizaremos el mismo análisis que con las series anteriores y haremos un pronóstico para el próximo año. Observamos nuevamente la necesidad de aplicar diferencias, solamente en la parte no estacional, gracias a la prueba KPSS, a la prueba Dickey-Fuller aumentada y claro a los comandos `ndiffs(.)` y `nsdiffs(.)`. Para el cálculo de la tendencia determinista tenemos el siguiente código en R:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 4/Ejercicios")
> IPC<-read.table("IPC(MRD).csv",header=T)
> Serie3<-ts(IPC, frequency = 12, start = c(2000, 1))
> kpss.test(Serie3)
```

KPSS Test for Level Stationarity

```
data: Serie3
KPSS Level = 4.1148, Truncation lag parameter = 3, p-value = 0.01
```

```
> adf.test(Serie3,alternative="stationary")
```

Augmented Dickey-Fuller Test

```
data: Serie3
Dickey-Fuller = -2.8661, Lag order = 5, p-value = 0.2147
alternative hypothesis: stationary
```

```
> ndiffs(Serie3,alpha=0.05,test=c("kpss"),max.d=2)
```

```
[1] 1
```

```
> nsdiffs(Serie3,m=frequency(Serie3), test=c("ocsb"), max.D=2)
```

```
[1] 0
```

```
> Ajuste5<-auto.arima(Serie3,xreg=1:length(Serie3),trace=FALSE,test=c("kpss"),
  stepwise=FALSE,approximation=FALSE,max.order=9,allowdrift=TRUE)
```

```
Series: Serie3
ARIMA(5,0,3) with zero mean
```

Coefficients:

	ar1	ar2	ar3	ar4	ar5	ma1	ma2	ma3	1:length(Serie3)
	2.0487	-2.1409	2.1329	-1.1943	0.1384	-1.0490	1.0299	-0.9624	257.4221
s.e.	0.0714	0.1549	0.1488	0.1559	0.0753	0.0309	0.0517	0.0418	2.5857

```
sigma^2 estimated as 1067253: log likelihood=-1424.28
AIC=2868.56 AICc=2869.95 BIC=2899.92
```

```
> forecast(Ajuste5,xreg=c(171:182),h=12)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar 2014	37892.09	36567.54	39216.64	35866.37	39917.82
Apr 2014	38014.32	36140.93	39887.71	35149.22	40879.42
May 2014	38808.76	36561.09	41056.43	35371.24	42246.28
Jun 2014	39686.99	37110.40	42263.58	35746.43	43627.54
Jul 2014	40005.20	37177.24	42833.17	35680.20	44330.21
Aug 2014	40321.21	37340.79	43301.62	35763.05	44879.36
Sep 2014	41232.54	38141.59	44323.48	36505.34	45959.73
Oct 2014	42166.74	38969.57	45363.92	37277.09	47056.40
Nov 2014	42548.99	39273.98	45824.00	37540.29	47557.69
Dec 2014	42946.39	39635.10	46257.69	37882.20	48010.59
Jan 2015	43894.02	40561.37	47226.68	38797.16	48990.89
Feb 2015	44814.22	41460.33	48168.11	39684.88	49943.55

```
> Box.test(residuals(Ajuste5),fitdf=2,lag=22,type="Ljung")
```

Box-Ljung test

data: residuals(Ajuste5)

X-squared = 17.8676, df = 20, p-value = 0.5961

El modelo óptimo es ARIMA(5,0,3) con media igual a cero. Los pronósticos para el próximo año se muestran dentro del código. Para el cálculo de la tendencia estocástica tenemos el código siguiente en R, obteniendo ahora un modelo ARIMA(3,1,2) y cuyos pronósticos también se muestran dentro del código:

```
> setwd("C:/Users/Gerardo/Desktop/Tesis/Capítulo 4/Ejercicios")
> IPC<-read.table("IPC(MRD).csv",header=T)
> Serie3<-ts(IPC, frequency = 12, start = c(2000, 1))
> kpss.test(Serie3)
```

KPSS Test for Level Stationarity

data: Serie3

KPSS Level = 4.1148, Truncation lag parameter = 3, p-value = 0.01

```
> adf.test(Serie3,alternative="stationary")
```

Augmented Dickey-Fuller Test

data: Serie3

Dickey-Fuller = -2.8661, Lag order = 5, p-value = 0.2147

alternative hypothesis: stationary

```
> ndiffs(Serie3,alpha=0.05,test=c("kpss"),max.d=2)
```

[1] 1

```
> nsdiffs(Serie3,m=frequency(Serie3), test=c("ocsb"), max.D=2)
```

[1] 0

```
> Ajuste6<-auto.arima(Serie3,d=1,trace=FALSE,xreg=1:length(Serie3),stepwise=FALSE,
  approximation=FALSE,max.order=9,allowdrift=TRUE)
```

```
Series: Serie3
ARIMA(3,1,2)
```

```
Coefficients:
```

	ar1	ar2	ar3	ma1	ma2	1:length(Serie3)
	0.1855	-0.9431	0.2068	-0.0699	0.9746	176.4805
s.e.	0.0777	0.0339	0.0790	0.0224	0.0468	101.3177

```
sigma^2 estimated as 1153576: log likelihood=-1411.53
AIC=2837.06 AICc=2837.75 BIC=2858.97
```

```
> forecast(Ajuste6,xreg=c(171:182),h=12)
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
Mar 2014	37503.12	36126.37	38879.88	35397.56	39608.69
Apr 2014	37397.92	35335.28	39460.56	34243.38	40552.46
May 2014	37902.41	35286.94	40517.88	33902.39	41902.42
Jun 2014	38276.29	35125.92	41426.67	33458.21	43094.38
Jul 2014	38121.79	34518.75	41724.83	32611.42	43632.16
Aug 2014	38118.52	34167.09	42069.95	32075.32	44161.72
Sep 2014	38614.63	34337.59	42891.68	32073.45	45155.81
Oct 2014	38951.49	34327.67	43575.32	31879.96	46023.03
Nov 2014	38819.10	33878.95	43759.26	31263.78	46374.42
Dec 2014	38853.12	33652.27	44053.97	30899.11	46807.13
Jan 2015	39327.64	33872.48	44782.81	30984.69	47670.60
Feb 2015	39629.90	33899.89	45359.92	30866.60	48393.20

```
> Box.test(residuals(Ajuste6),fitdf=2,lag=22,type="Ljung")
```

```
Box-Ljung test
```

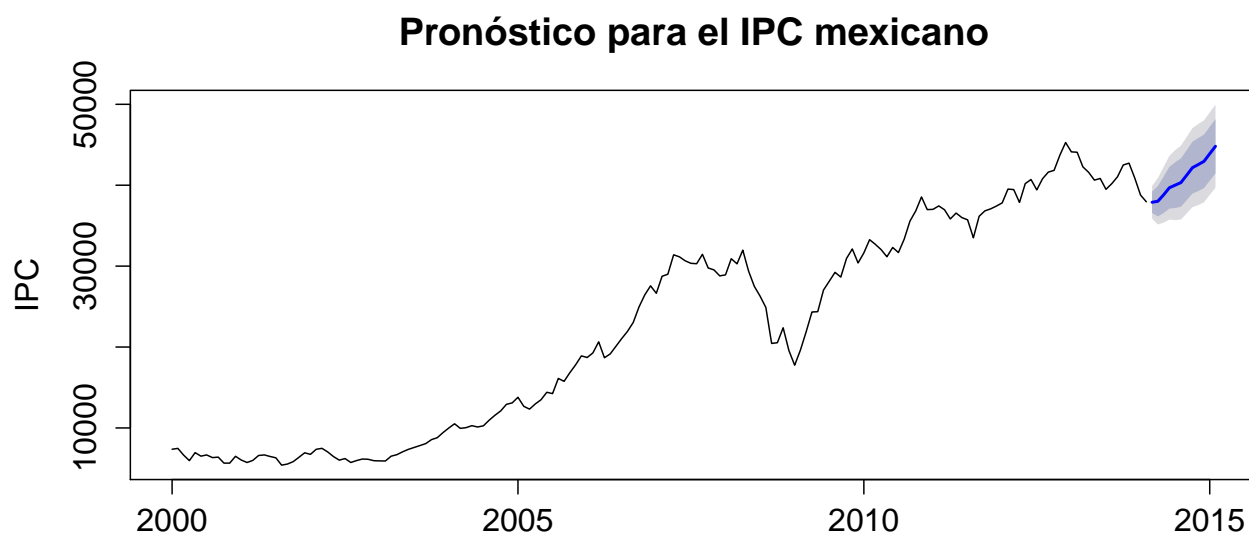
```
data: residuals(Ajuste6)
```

```
X-squared = 21.24, df = 20, p-value = 0.3831
```

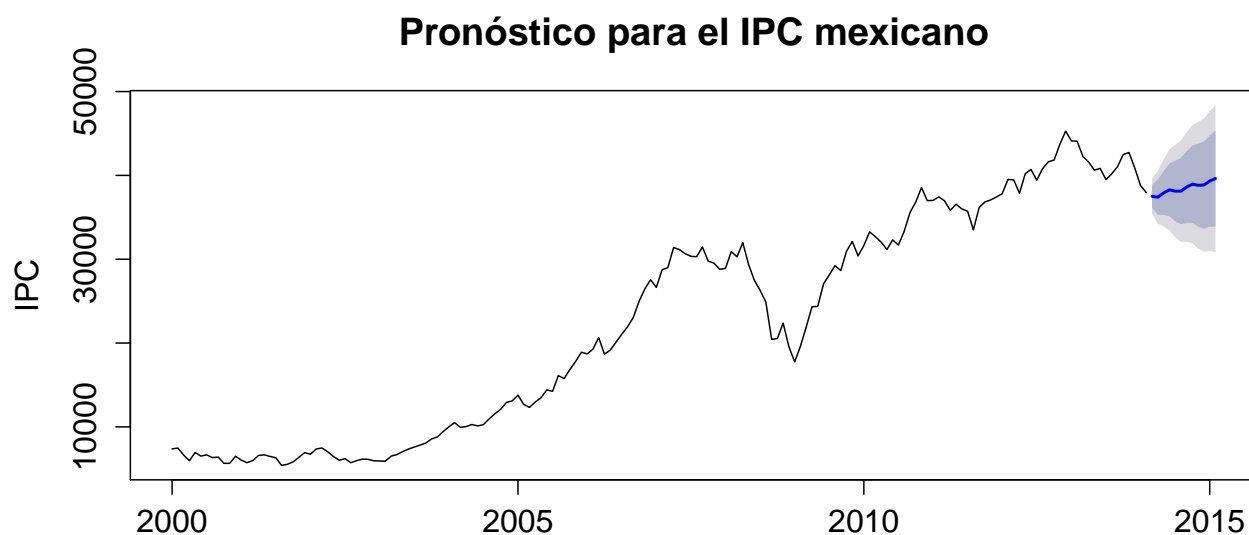
En el cálculo de ambas tendencias, vemos, gracias a la prueba de Ljung y Box, que las autocorrelaciones hasta el “lag” $h = 22$ conjuntamente son cero. El correspondiente p-value para la tendencia determinista es $0.5961 > 0.05$, mientras que el p-value para la tendencia estocástica es $0.3831 > 0.05$. La **Figura 4.4.8** muestra el pronóstico correspondiente con ambas tendencias. El código en R es el siguiente:

```
> par(mfrow=c(2,1)); options(scipen=3)
> plot(forecast(Ajuste5,xreg=c(171:182),h=24),main="Pronóstico para el IPC mexicano",
  xlab="(a) Tendencia determinista",ylab="IPC")
> plot(forecast(Ajuste6,xreg=c(171:182),h=24),main="Pronóstico para el IPC mexicano",
  xlab="(b) Tendencia estocástica",ylab="IPC")
```

Al ver estos ejemplos nos preguntamos, ¿por qué las bandas de predicción son un poco más amplias para la tendencia estocástica? En el modelo con tendencia determinista, la pendiente de dicha tendencia se supone invariante en el tiempo, mientras que en el modelo con tendencia estocástica puede cambiar en el tiempo, de aquí que las bandas de predicción para la tendencia estocástica permiten una mayor incertidumbre en la pendiente de la tendencia.



(a) Tendencia determinista



(b) Tendencia estocástica

Figura 4.4.8: Figura que muestra el pronóstico para el IPC mexicano para los próx. 2 años

Conclusiones

En este trabajo se revisaron un conjunto de modelos que ayudan a entender la estructura, de manera aproximada, de un sinnúmero de fenómenos. De acuerdo con unos datos, nuestro modelo por ajustar debe de cumplir ciertos supuestos que, de no cumplirse podrían, ocasionar una defectuosa interpretación del problema en cuestión.

Hemos visto en el capítulo 1 cómo un conjunto de datos, el financiamiento otorgado a través de la Banca Privada y Mixta, y una variable explicativa como el tiempo, logran un ajuste eficiente en el modelo de regresión lineal simple, cumpliendo inclusive el supuesto de normalidad en los residuos (el p-value de las pruebas Shapiro-Wilk y Lilliefors son 0.1538 y 0.1486, respectivamente, ambos mayores al nivel de significancia $\alpha = 0.05$, por lo que no se rechaza la hipótesis nula y se concluye normalidad en los residuos). Hasta este punto, un investigador podría concluir que dicho modelo es óptimo al describir la relación existente entre el financiamiento y el tiempo. Al incluir la prueba de Durbin-Watson se concluye que existe auto-correlación en los residuos. Lo anterior implica que el error que estamos cometiendo al ajustar el modelo de regresión a dichos datos va en aumento con el paso del tiempo y que por tanto no es recomendable utilizar el modelo de regresión lineal simple para ajustar este conjunto de datos.

En cada modelo existen supuestos que pueden ser obviados si difieren en una pequeña cuantía de lo que se busca demostrar ya que dichas desviaciones no son tan significativas y no influyen en nuestra decisión de rechazar o no rechazar el modelo. Pero en cambio, sí que existen otros supuestos que, de no cumplirse, impedirían de manera rotunda la aplicación de cierto modelo a unos datos.

Ahora bien, para poder ajustar un modelo ARMA(p,q) a una serie debemos ser capaces de primero estimar las componentes de tendencia y de estacionalidad para así, al sustraerlas, obtener una nueva serie lo más estacionaria posible. El capítulo 2 nos ha servido de guía para encontrar modelos y métodos de ajuste tendencial y estacional que nos servirán como paso previo al ajuste de los modelos estudiados en el capítulo 3, aunque claro que hay gente que puede optar por usar el método de Holt-Winters, por ejemplo, para modelar completamente a la serie.

El corazón de este escrito es el capítulo 3, en éste se discutió de forma exhaustiva la necesidad de volver una serie lo más estacionaria posible. La estructura de dependencia de las series que poseen estacionariedad, al mantenerse constante, nos permitirá modelar la serie y poder así prever su evolución a lo largo del tiempo. Al tener series estacionarias hemos podido desarrollar técnicas eficientes para predecir valores futuros de la serie y no sólo eso, sino de igual forma se logró obtener intervalos de confianza para las predicciones.

De especial interés fue lograr llevar la teoría de la mano de la parte gráfica, en particular el procedimiento que se llevó a cabo para identificar, a partir de la sucesión de autocorrelaciones muestrales (ACF) y de la sucesión de autocorrelaciones muestrales parciales (PACF), las características teóricas de los modelos $ARMA(p,q)$, $ARIMA(p,d,q)$ y $SARIMA(p,d,q) \times (P,D,Q)$.

Los ejercicios analizados en el capítulo 4 han querido ejemplificar la metodología del modelo de regresión dinámica. Nuestros pronósticos se basaron solamente en la teoría vista en este capítulo y a pesar de ser una buena base para realizar predicciones, muchas series manejadas a lo largo de este escrito deben tomar en cuenta un sinnúmero de variables explicativas extras para su mejor pronóstico. Nuestra intención no es modificar la metodología utilizada para realizar predicciones de series como las correspondientes al PIB o IPC mexicano, por ejemplo, sino más bien, con base en la información pasada de la serie, dar un estimado confiable para el futuro, para con ello tomar decisiones que involucren el menor riesgo posible. El entender el pasado de una serie de tiempo es de gran ayuda al realizar predicciones pero seguramente cada conjunto de datos podrá involucrar (o ha involucrado en sus observaciones anteriores) un sinnúmero de variables explicativas que ayudarán (o han ayudado) a predecir observaciones futuras.

Sabemos que el IPC es un indicador que se utiliza para expresar la variación promedio de los precios de las acciones que cotizan en la Bolsa Mexicana de Valores y que para su cálculo, se debe tomar en cuenta las acciones y precios de una muestra significativa de emisoras que representen a los sectores económicos en los cuales se clasifica a las emisoras inscritas en la Bolsa Mexicana de Valores. De cierta forma, al tomar la información pasada, estamos tomando a consideración muestras significativas de emisoras del pasado y junto al tiempo hemos realizado nuestro pronóstico. Pero estamos seguros que para la predicción real, expertos en la materia, tienen mucha más información, no sólo de la serie, sino de todos aquellos factores que han afectado, afectan y de cierta forma, afectarán, el comportamiento de la serie, en este caso, del IPC.

Cuando hablamos del PIB estadounidense o mexicano, no pretendemos ser expertos en el tema ya que no basta con información pasada de la serie, sino que claramente existirán, al momento de querer realizar predicciones, variables explicativas que mueven, en ese momento, el comportamiento de dichas series y de las cuales nosotros no estamos informados. Como hemos dicho, nosotros meramente hemos dado un pronóstico confiable en el cual expertos puedan basarse e incluir, si así lo creen conveniente, en sus modelos de predicción. El mismo argumento se utilizará para justificar el uso de todas las series manejadas en este escrito.

Bibliografía

- Bowerman, B. y O'Connell, R. (1987). *Forecasting and Time Series. An Applied Approach*. Duxbury Press, tercera edición.
- Box, G. y Tiao, G. (1973). *Bayesian Inference in Statistical Analysis*. Addison Wesley, primera edición.
- Brockwell, P. y Davis, R. (1991). *Time Series: Theory and Methods*. Springer, segunda edición.
- Chatfield, C. (2004). *The Analysis of Time Series. An Introduction*. Chapman & Hall, sexta edición.
- Engel, R. (2001). *What a Good is a Volatility Model?* University of California, primera edición.
- Fuller, W. (1996). *Introduction to Statistical Time Series*. Wiley, segunda edición.
- Greene, W. (1999). *Análisis Económico*. Prentice Hall, tercera edición.
- Guerrero, V. (2003). *Análisis Estadístico de Series de Tiempo Económicas*. Thomson, segunda edición.
- Hamilton, J. (1994). *Time Series Analysis*. Princeton University Press, primera edición.
- Harvey, A. (1990). *Forecasting, structural time series models and the kalman filter*. Cambridge University Press, primera edición.
- Kedem, B. y Konstantinos, F. (2002). *Regression models for time series analysis*. John Wiley & Sons, primera edición.
- Kutner, M., Neter, J., y Nachtsheim, C. (2005). *Applied Linear Statistical Models*. McGraw Hill Irwin, quinta edición.
- Madsen, H. (2008). *Time Series Analysis*. Chapman & Hall, segunda edición.
- Montgomery, D. y Peck, E. (1992). *Introduction to Linear Regression Analysis*. John Wiley & Sons, cuarta edición.
- Morilla, C. (2000). *Análisis de Series Temporales*. La Muralla, primera edición.
- Neter, J., Kutner, M., y Wasserman, W. (1990). *Applied Linear Statistical Models. Regression, Analysis of Variance and Experimental Designs*. Irwin, tercera edición.
- Novalés, A. (1993). *Econometría*. McGraw Hill, segunda edición.
- Petris, G., Petrone, S., y Campagnoli, P. (2009). *Dynamic Linear Models With R*. Springer, primera edición.
- Seber, G. y Lee, A. (2003). *Linear Regression Analysis*. John Wiley & Sons, segunda edición.

- Shumway, R. y Stoffer, D. (2011). *Time Series Analysis and its Applications*. Springer, tercera edición.
- Wei, W. (2006). *Time Series Analysis. Univariate and Multivariate Methods*. Pearson Addison Wesley, segunda edición.

Apéndice A

Tablas Estadísticas

Tabla A.1 Distribución Normal Estándar

Tabla A.2 Distribución t de Student

Tabla A.3 Distribución Ji-Cuadrada

Tabla A.4 Distribución $F_{.25,n,m}$ de Fisher

Tabla A.5 Distribución $F_{.10,n,m}$ de Fisher

Tabla A.6 Distribución $F_{.05,n,m}$ de Fisher

Tabla A.7 Distribución $F_{.025,n,m}$ de Fisher

Tabla A.8 Distribución $F_{.01,n,m}$ de Fisher

Tabla A.9 Valores críticos del estadístico Durbin-Watson

Tabla A1: Distribución Normal (0,1)

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

p	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.00		-3.0902	-2.8782	-2.7478	-2.6521	-2.5758	-2.5121	-2.4573	-2.4089	-2.3656
0.01	-2.3263	-2.2904	-2.2571	-2.2262	-2.1973	-2.1701	-2.1444	-2.1201	-2.0969	-2.0749
0.02	-2.0537	-2.0335	-2.0141	-1.9954	-1.9774	-1.9600	-1.9431	-1.9268	-1.9110	-1.8957
0.03	-1.8808	-1.8663	-1.8522	-1.8384	-1.8250	-1.8119	-1.7991	-1.7866	-1.7744	-1.7624
0.04	-1.7507	-1.7392	-1.7279	-1.7169	-1.7060	-1.6954	-1.6849	-1.6747	-1.6646	-1.6546
0.05	-1.6449	-1.6352	-1.6258	-1.6164	-1.6072	-1.5982	-1.5893	-1.5805	-1.5718	-1.5632
0.06	-1.5548	-1.5464	-1.5382	-1.5301	-1.5220	-1.5141	-1.5063	-1.4985	-1.4909	-1.4833
0.07	-1.4758	-1.4684	-1.4611	-1.4538	-1.4466	-1.4395	-1.4325	-1.4255	-1.4187	-1.4118
0.08	-1.4051	-1.3984	-1.3917	-1.3852	-1.3787	-1.3722	-1.3658	-1.3595	-1.3532	-1.3469
0.09	-1.3408	-1.3346	-1.3285	-1.3225	-1.3165	-1.3106	-1.3047	-1.2988	-1.2930	-1.2873
0.10	-1.2816	-1.2759	-1.2702	-1.2646	-1.2591	-1.2536	-1.2481	-1.2426	-1.2372	-1.2319
0.11	-1.2265	-1.2212	-1.2160	-1.2107	-1.2055	-1.2004	-1.1952	-1.1901	-1.1850	-1.1800
0.12	-1.1750	-1.1700	-1.1650	-1.1601	-1.1552	-1.1503	-1.1455	-1.1407	-1.1359	-1.1311
0.13	-1.1264	-1.1217	-1.1170	-1.1123	-1.1077	-1.1031	-1.0985	-1.0939	-1.0893	-1.0848
0.14	-1.0803	-1.0758	-1.0714	-1.0669	-1.0625	-1.0581	-1.0537	-1.0494	-1.0450	-1.0407
0.15	-1.0364	-1.0322	-1.0279	-1.0237	-1.0194	-1.0152	-1.0110	-1.0069	-1.0027	-0.9986
0.16	-0.9945	-0.9904	-0.9863	-0.9822	-0.9782	-0.9741	-0.9701	-0.9661	-0.9621	-0.9581
0.17	-0.9542	-0.9502	-0.9463	-0.9424	-0.9385	-0.9346	-0.9307	-0.9269	-0.9230	-0.9192
0.18	-0.9154	-0.9116	-0.9078	-0.9040	-0.9002	-0.8965	-0.8927	-0.8890	-0.8853	-0.8816
0.19	-0.8779	-0.8742	-0.8705	-0.8669	-0.8633	-0.8596	-0.8560	-0.8524	-0.8488	-0.8452
0.20	-0.8416	-0.8381	-0.8345	-0.8310	-0.8274	-0.8239	-0.8204	-0.8169	-0.8134	-0.8099
0.21	-0.8064	-0.8030	-0.7995	-0.7961	-0.7926	-0.7892	-0.7858	-0.7824	-0.7790	-0.7756
0.22	-0.7722	-0.7688	-0.7655	-0.7621	-0.7588	-0.7554	-0.7521	-0.7488	-0.7454	-0.7421
0.23	-0.7388	-0.7356	-0.7323	-0.7290	-0.7257	-0.7225	-0.7192	-0.7160	-0.7128	-0.7095
0.24	-0.7063	-0.7031	-0.6999	-0.6967	-0.6935	-0.6903	-0.6871	-0.6840	-0.6808	-0.6776
0.25	-0.6745	-0.6713	-0.6682	-0.6651	-0.6620	-0.6588	-0.6557	-0.6526	-0.6495	-0.6464
0.26	-0.6433	-0.6403	-0.6372	-0.6341	-0.6311	-0.6280	-0.6250	-0.6219	-0.6189	-0.6158
0.27	-0.6128	-0.6098	-0.6068	-0.6038	-0.6008	-0.5978	-0.5948	-0.5918	-0.5888	-0.5858
0.28	-0.5828	-0.5799	-0.5769	-0.5740	-0.5710	-0.5681	-0.5651	-0.5622	-0.5592	-0.5563
0.29	-0.5534	-0.5505	-0.5476	-0.5446	-0.5417	-0.5388	-0.5359	-0.5330	-0.5302	-0.5273
0.30	-0.5244	-0.5215	-0.5187	-0.5158	-0.5129	-0.5101	-0.5072	-0.5044	-0.5015	-0.4987
0.31	-0.4959	-0.4930	-0.4902	-0.4874	-0.4845	-0.4817	-0.4789	-0.4761	-0.4733	-0.4705
0.32	-0.4677	-0.4649	-0.4621	-0.4593	-0.4565	-0.4538	-0.4510	-0.4482	-0.4454	-0.4427
0.33	-0.4399	-0.4372	-0.4344	-0.4316	-0.4289	-0.4261	-0.4234	-0.4207	-0.4179	-0.4152
0.34	-0.4125	-0.4097	-0.4070	-0.4043	-0.4016	-0.3989	-0.3961	-0.3934	-0.3907	-0.3880
0.35	-0.3853	-0.3826	-0.3799	-0.3772	-0.3745	-0.3719	-0.3692	-0.3665	-0.3638	-0.3611
0.36	-0.3585	-0.3558	-0.3531	-0.3505	-0.3478	-0.3451	-0.3425	-0.3398	-0.3372	-0.3345
0.37	-0.3319	-0.3292	-0.3266	-0.3239	-0.3213	-0.3186	-0.3160	-0.3134	-0.3107	-0.3081
0.38	-0.3055	-0.3029	-0.3002	-0.2976	-0.2950	-0.2924	-0.2898	-0.2871	-0.2845	-0.2819
0.39	-0.2793	-0.2767	-0.2741	-0.2715	-0.2689	-0.2663	-0.2637	-0.2611	-0.2585	-0.2559
0.40	-0.2533	-0.2508	-0.2482	-0.2456	-0.2430	-0.2404	-0.2378	-0.2353	-0.2327	-0.2301
0.41	-0.2275	-0.2250	-0.2224	-0.2198	-0.2173	-0.2147	-0.2121	-0.2096	-0.2070	-0.2045
0.42	-0.2019	-0.1993	-0.1968	-0.1942	-0.1917	-0.1891	-0.1866	-0.1840	-0.1815	-0.1789
0.43	-0.1764	-0.1738	-0.1713	-0.1687	-0.1662	-0.1637	-0.1611	-0.1586	-0.1560	-0.1535
0.44	-0.1510	-0.1484	-0.1459	-0.1434	-0.1408	-0.1383	-0.1358	-0.1332	-0.1307	-0.1282
0.45	-0.1257	-0.1231	-0.1206	-0.1181	-0.1156	-0.1130	-0.1105	-0.1080	-0.1055	-0.1030
0.46	-0.1004	-0.0979	-0.0954	-0.0929	-0.0904	-0.0878	-0.0853	-0.0828	-0.0803	-0.0778
0.47	-0.0753	-0.0728	-0.0702	-0.0677	-0.0652	-0.0627	-0.0602	-0.0577	-0.0552	-0.0527
0.48	-0.0502	-0.0476	-0.0451	-0.0426	-0.0401	-0.0376	-0.0351	-0.0326	-0.0301	-0.0276
0.49	-0.0251	-0.0226	-0.0201	-0.0175	-0.0150	-0.0125	-0.0100	-0.0075	-0.0050	-0.0025

Tabla A1: Distribución Normal (0,1)

$$\phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

<i>p</i>	0.000	0.001	0.002	0.003	0.004	0.005	0.006	0.007	0.008	0.009
0.50	0.0000	0.0025	0.0050	0.0075	0.0100	0.0125	0.0150	0.0175	0.0201	0.0226
0.51	0.0251	0.0276	0.0301	0.0326	0.0351	0.0376	0.0401	0.0426	0.0451	0.0476
0.52	0.0502	0.0527	0.0552	0.0577	0.0602	0.0627	0.0652	0.0677	0.0702	0.0728
0.53	0.0753	0.0778	0.0803	0.0828	0.0853	0.0878	0.0904	0.0929	0.0954	0.0979
0.54	0.1004	0.1030	0.1055	0.1080	0.1105	0.1130	0.1156	0.1181	0.1206	0.1231
0.55	0.1257	0.1282	0.1307	0.1332	0.1358	0.1383	0.1408	0.1434	0.1459	0.1484
0.56	0.1510	0.1535	0.1560	0.1586	0.1611	0.1637	0.1662	0.1687	0.1713	0.1738
0.57	0.1764	0.1789	0.1815	0.1840	0.1866	0.1891	0.1917	0.1942	0.1968	0.1993
0.58	0.2019	0.2045	0.2070	0.2096	0.2121	0.2147	0.2173	0.2198	0.2224	0.2250
0.59	0.2275	0.2301	0.2327	0.2353	0.2378	0.2404	0.2430	0.2456	0.2482	0.2508
0.60	0.2533	0.2559	0.2585	0.2611	0.2637	0.2663	0.2689	0.2715	0.2741	0.2767
0.61	0.2793	0.2819	0.2845	0.2871	0.2898	0.2924	0.2950	0.2976	0.3002	0.3029
0.62	0.3055	0.3081	0.3107	0.3134	0.3160	0.3186	0.3213	0.3239	0.3266	0.3292
0.63	0.3319	0.3345	0.3372	0.3398	0.3425	0.3451	0.3478	0.3505	0.3531	0.3558
0.64	0.3585	0.3611	0.3638	0.3665	0.3692	0.3719	0.3745	0.3772	0.3799	0.3826
0.65	0.3853	0.3880	0.3907	0.3934	0.3961	0.3989	0.4016	0.4043	0.4070	0.4097
0.66	0.4125	0.4152	0.4179	0.4207	0.4234	0.4261	0.4289	0.4316	0.4344	0.4372
0.67	0.4399	0.4427	0.4454	0.4482	0.4510	0.4538	0.4565	0.4593	0.4621	0.4649
0.68	0.4677	0.4705	0.4733	0.4761	0.4789	0.4817	0.4845	0.4874	0.4902	0.4930
0.69	0.4959	0.4987	0.5015	0.5044	0.5072	0.5101	0.5129	0.5158	0.5187	0.5215
0.70	0.5244	0.5273	0.5302	0.5330	0.5359	0.5388	0.5417	0.5446	0.5476	0.5505
0.71	0.5534	0.5563	0.5592	0.5622	0.5651	0.5681	0.5710	0.5740	0.5769	0.5799
0.72	0.5828	0.5858	0.5888	0.5918	0.5948	0.5978	0.6008	0.6038	0.6068	0.6098
0.73	0.6128	0.6158	0.6189	0.6219	0.6250	0.6280	0.6311	0.6341	0.6372	0.6403
0.74	0.6433	0.6464	0.6495	0.6526	0.6557	0.6588	0.6620	0.6651	0.6682	0.6713
0.75	0.6745	0.6776	0.6808	0.6840	0.6871	0.6903	0.6935	0.6967	0.6999	0.7031
0.76	0.7063	0.7095	0.7128	0.7160	0.7192	0.7225	0.7257	0.7290	0.7323	0.7356
0.77	0.7388	0.7421	0.7454	0.7488	0.7521	0.7554	0.7588	0.7621	0.7655	0.7688
0.78	0.7722	0.7756	0.7790	0.7824	0.7858	0.7892	0.7926	0.7961	0.7995	0.8030
0.79	0.8064	0.8099	0.8134	0.8169	0.8204	0.8239	0.8274	0.8310	0.8345	0.8381
0.80	0.8416	0.8452	0.8488	0.8524	0.8560	0.8596	0.8633	0.8669	0.8705	0.8742
0.81	0.8779	0.8816	0.8853	0.8890	0.8927	0.8965	0.9002	0.9040	0.9078	0.9116
0.82	0.9154	0.9192	0.9230	0.9269	0.9307	0.9346	0.9385	0.9424	0.9463	0.9502
0.83	0.9542	0.9581	0.9621	0.9661	0.9701	0.9741	0.9782	0.9822	0.9863	0.9904
0.84	0.9945	0.9986	1.0027	1.0069	1.0110	1.0152	1.0194	1.0237	1.0279	1.0322
0.85	1.0364	1.0407	1.0450	1.0494	1.0537	1.0581	1.0625	1.0669	1.0714	1.0758
0.86	1.0803	1.0848	1.0893	1.0939	1.0985	1.1031	1.1077	1.1123	1.1170	1.1217
0.87	1.1264	1.1311	1.1359	1.1407	1.1455	1.1503	1.1552	1.1601	1.1650	1.1700
0.88	1.1750	1.1800	1.1850	1.1901	1.1952	1.2004	1.2055	1.2107	1.2160	1.2212
0.89	1.2265	1.2319	1.2372	1.2426	1.2481	1.2536	1.2591	1.2646	1.2702	1.2759
0.90	1.2816	1.2873	1.2930	1.2988	1.3047	1.3106	1.3165	1.3225	1.3285	1.3346
0.91	1.3408	1.3469	1.3532	1.3595	1.3658	1.3722	1.3787	1.3852	1.3917	1.3984
0.92	1.4051	1.4118	1.4187	1.4255	1.4325	1.4395	1.4466	1.4538	1.4611	1.4684
0.93	1.4758	1.4833	1.4909	1.4985	1.5063	1.5141	1.5220	1.5301	1.5382	1.5464
0.94	1.5548	1.5632	1.5718	1.5805	1.5893	1.5982	1.6072	1.6164	1.6258	1.6352
0.95	1.6449	1.6546	1.6646	1.6747	1.6849	1.6954	1.7060	1.7169	1.7279	1.7392
0.96	1.7507	1.7624	1.7744	1.7866	1.7991	1.8119	1.8250	1.8384	1.8522	1.8663
0.97	1.8808	1.8957	1.9110	1.9268	1.9431	1.9600	1.9774	1.9954	2.0141	2.0335
0.98	2.0537	2.0749	2.0969	2.1201	2.1444	2.1701	2.1973	2.2262	2.2571	2.2904
0.99	2.3263	2.3656	2.4089	2.4573	2.5121	2.5758	2.6521	2.7478	2.8782	3.0902

Tabla A2: Distribución t

$$F(t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi} \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} dx$$

n	α									
	0.75	0.80	0.85	0.875	0.90	0.95	0.975	0.99	0.995	0.9995
1	1.000	1.376	1.963	2.414	3.078	6.314	12.706	31.821	63.657	636.619
2	0.816	1.061	1.386	1.604	1.886	2.920	4.303	6.965	9.925	31.599
3	0.765	0.978	1.250	1.423	1.638	2.353	3.182	4.541	5.841	12.924
4	0.741	0.941	1.190	1.344	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.301	1.476	2.015	2.571	3.365	4.032	6.869
6	0.718	0.906	1.134	1.273	1.440	1.943	2.447	3.143	3.707	5.959
7	0.711	0.896	1.119	1.254	1.415	1.895	2.365	2.998	3.499	5.408
8	0.706	0.889	1.108	1.240	1.397	1.860	2.306	2.896	3.355	5.041
9	0.703	0.883	1.100	1.230	1.383	1.833	2.262	2.821	3.250	4.781
10	0.700	0.879	1.093	1.221	1.372	1.812	2.228	2.764	3.169	4.587
11	0.697	0.876	1.088	1.214	1.363	1.796	2.201	2.718	3.106	4.437
12	0.695	0.873	1.083	1.209	1.356	1.782	2.179	2.681	3.055	4.318
13	0.694	0.870	1.079	1.204	1.350	1.771	2.160	2.650	3.012	4.221
14	0.692	0.868	1.076	1.200	1.345	1.761	2.145	2.624	2.977	4.140
15	0.691	0.866	1.074	1.197	1.341	1.753	2.131	2.602	2.947	4.073
16	0.690	0.865	1.071	1.194	1.337	1.746	2.120	2.583	2.921	4.015
17	0.689	0.863	1.069	1.191	1.333	1.740	2.110	2.567	2.898	3.965
18	0.688	0.862	1.067	1.189	1.330	1.734	2.101	2.552	2.878	3.922
19	0.688	0.861	1.066	1.187	1.328	1.729	2.093	2.539	2.861	3.883
20	0.687	0.860	1.064	1.185	1.325	1.725	2.086	2.528	2.845	3.850
21	0.686	0.859	1.063	1.183	1.323	1.721	2.080	2.518	2.831	3.819
22	0.686	0.858	1.061	1.182	1.321	1.717	2.074	2.508	2.819	3.792
23	0.685	0.858	1.060	1.180	1.319	1.714	2.069	2.500	2.807	3.768
24	0.685	0.857	1.059	1.179	1.318	1.711	2.064	2.492	2.797	3.745
25	0.684	0.856	1.058	1.178	1.316	1.708	2.060	2.485	2.787	3.725
26	0.684	0.856	1.058	1.177	1.315	1.706	2.056	2.479	2.779	3.707
27	0.684	0.855	1.057	1.176	1.314	1.703	2.052	2.473	2.771	3.690
28	0.683	0.855	1.056	1.175	1.313	1.701	2.048	2.467	2.763	3.674
29	0.683	0.854	1.055	1.174	1.311	1.699	2.045	2.462	2.756	3.659
30	0.683	0.854	1.055	1.173	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.167	1.303	1.684	2.021	2.423	2.704	3.551
50	0.679	0.849	1.047	1.164	1.299	1.676	2.009	2.403	2.678	3.496
60	0.679	0.848	1.045	1.162	1.296	1.671	2.000	2.390	2.660	3.460
70	0.678	0.847	1.044	1.160	1.294	1.667	1.994	2.381	2.648	3.435
80	0.678	0.846	1.043	1.159	1.292	1.664	1.990	2.374	2.639	3.416
90	0.677	0.846	1.042	1.158	1.291	1.662	1.987	2.368	2.632	3.402
100	0.677	0.845	1.042	1.157	1.290	1.660	1.984	2.364	2.626	3.390

Tabla A2: Distribución t

$$F(t) = \int_{-\infty}^t \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right) \sqrt{n\pi} \left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}} dx$$

n	α									
	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	0.255	0.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
50	0.255	0.679	1.299	1.676	2.009	2.403	2.678	2.937	3.261	3.496
60	0.254	0.679	1.296	1.671	2.000	2.390	2.660	2.915	3.232	3.460
70	0.254	0.678	1.294	1.667	1.994	2.381	2.648	2.899	3.211	3.435
80	0.254	0.678	1.292	1.664	1.990	2.374	2.639	2.887	3.195	3.416
90	0.254	0.677	1.291	1.662	1.987	2.368	2.632	2.878	3.183	3.402
100	0.254	0.677	1.290	1.660	1.984	2.364	2.626	2.871	3.174	3.390

Tabla A3: Distribución χ^2

$$F(u) = \int_0^u \frac{x^{\frac{n-2}{2}} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} dx$$

n \ α	0.995	0.99	0.975	0.95	0.90	0.50	0.10	0.05	0.025	0.01	0.005
1	0.00	0.00	0.00	0.00	0.02	0.45	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	1.39	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	2.37	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	3.36	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	4.35	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	5.35	10.64	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	6.35	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	7.34	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	8.34	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	9.34	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	10.34	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	11.34	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	12.34	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	13.34	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	14.34	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	15.34	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	16.34	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	17.34	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	18.34	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	19.34	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	20.34	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	21.34	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	22.34	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	23.34	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	24.34	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	25.34	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	26.34	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	27.34	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	28.34	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	29.34	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	39.34	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	49.33	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	59.33	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	69.33	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	79.33	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	89.33	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	99.33	118.50	124.34	129.56	135.81	140.17

Tabla A4: Distribución F

$$F_{.25,n,m}$$

Grados de libertad para el numerador (n)

Grados de libertad para el denominador (m)

m \ n	1	2	3	4	5	6	7	8	9	10
1	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32
2	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38
3	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44
4	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08
5	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89
6	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77
7	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69
8	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.63	1.63
9	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59
10	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55
11	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52
12	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50
13	1.45	1.55	1.55	1.53	1.52	1.51	1.50	1.49	1.49	1.48
14	1.44	1.53	1.53	1.52	1.51	1.50	1.49	1.48	1.47	1.46
15	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45
16	1.42	1.51	1.51	1.50	1.48	1.47	1.46	1.45	1.44	1.44
17	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43
18	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42
19	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41
20	1.40	1.49	1.48	1.47	1.45	1.44	1.43	1.42	1.41	1.40
21	1.40	1.48	1.48	1.46	1.44	1.43	1.42	1.41	1.40	1.39
22	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39
23	1.39	1.47	1.47	1.45	1.43	1.42	1.41	1.40	1.39	1.38
24	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38
25	1.39	1.47	1.46	1.44	1.42	1.41	1.40	1.39	1.38	1.37
26	1.38	1.46	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.37
27	1.38	1.46	1.45	1.43	1.42	1.40	1.39	1.38	1.37	1.36
28	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36
29	1.38	1.45	1.45	1.43	1.41	1.40	1.38	1.37	1.36	1.35
30	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35
40	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33
50	1.35	1.43	1.41	1.39	1.37	1.36	1.34	1.33	1.32	1.31
60	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30
70	1.35	1.41	1.40	1.38	1.36	1.34	1.33	1.32	1.31	1.30
80	1.34	1.41	1.40	1.38	1.36	1.34	1.32	1.31	1.30	1.29
90	1.34	1.41	1.39	1.37	1.35	1.33	1.32	1.31	1.30	1.29
100	1.34	1.41	1.39	1.37	1.35	1.33	1.32	1.30	1.29	1.28

Tabla A4: Distribución F

$$F_{.25,n,m}$$

Grados de libertad para el numerador (n)

Grados de libertad para el denominador (m)

m \ n	11	12	13	14	15	20	30	40	50	100
1	9.37	9.41	9.44	9.47	9.49	9.58	9.67	9.71	9.74	9.80
2	3.39	3.39	3.40	3.41	3.41	3.43	3.44	3.45	3.46	3.47
3	2.45	2.45	2.45	2.45	2.46	2.46	2.47	2.47	2.47	2.47
4	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
5	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.88	1.87
6	1.77	1.77	1.77	1.76	1.76	1.76	1.75	1.75	1.75	1.74
7	1.69	1.68	1.68	1.68	1.68	1.67	1.66	1.66	1.66	1.65
8	1.63	1.62	1.62	1.62	1.62	1.61	1.60	1.59	1.59	1.58
9	1.58	1.58	1.58	1.57	1.57	1.56	1.55	1.54	1.54	1.53
10	1.55	1.54	1.54	1.54	1.53	1.52	1.51	1.51	1.50	1.49
11	1.52	1.51	1.51	1.51	1.50	1.49	1.48	1.47	1.47	1.46
12	1.49	1.49	1.49	1.48	1.48	1.47	1.45	1.45	1.44	1.43
13	1.47	1.47	1.47	1.46	1.46	1.45	1.43	1.42	1.42	1.41
14	1.46	1.45	1.45	1.44	1.44	1.43	1.41	1.41	1.40	1.39
15	1.44	1.44	1.43	1.43	1.43	1.41	1.40	1.39	1.38	1.37
16	1.43	1.43	1.42	1.42	1.41	1.40	1.38	1.37	1.37	1.36
17	1.42	1.41	1.41	1.41	1.40	1.39	1.37	1.36	1.36	1.34
18	1.41	1.40	1.40	1.40	1.39	1.38	1.36	1.35	1.34	1.33
19	1.40	1.40	1.39	1.39	1.38	1.37	1.35	1.34	1.33	1.32
20	1.39	1.39	1.38	1.38	1.37	1.36	1.34	1.33	1.32	1.31
21	1.39	1.38	1.37	1.37	1.37	1.35	1.33	1.32	1.32	1.30
22	1.38	1.37	1.37	1.36	1.36	1.34	1.32	1.31	1.31	1.29
23	1.37	1.37	1.36	1.36	1.35	1.34	1.32	1.31	1.30	1.29
24	1.37	1.36	1.36	1.35	1.35	1.33	1.31	1.30	1.29	1.28
25	1.36	1.36	1.35	1.35	1.34	1.33	1.31	1.29	1.29	1.27
26	1.36	1.35	1.35	1.34	1.34	1.32	1.30	1.29	1.28	1.27
27	1.35	1.35	1.34	1.34	1.33	1.32	1.30	1.28	1.28	1.26
28	1.35	1.34	1.34	1.33	1.33	1.31	1.29	1.28	1.27	1.25
29	1.35	1.34	1.33	1.33	1.32	1.31	1.29	1.27	1.27	1.25
30	1.34	1.34	1.33	1.33	1.32	1.30	1.28	1.27	1.26	1.25
40	1.32	1.31	1.31	1.30	1.30	1.28	1.25	1.24	1.23	1.21
50	1.30	1.30	1.29	1.28	1.28	1.26	1.23	1.22	1.21	1.19
60	1.29	1.29	1.28	1.27	1.27	1.25	1.22	1.21	1.20	1.18
70	1.29	1.28	1.27	1.27	1.26	1.24	1.21	1.20	1.19	1.16
80	1.28	1.27	1.27	1.26	1.26	1.23	1.21	1.19	1.18	1.16
90	1.28	1.27	1.26	1.26	1.25	1.23	1.20	1.19	1.18	1.15
100	1.27	1.27	1.26	1.25	1.25	1.23	1.20	1.18	1.17	1.14

Tabla A5: Distribución F
 $F_{.10,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	1	2	3	4	5	6	7	8	9	10
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94
21	2.96	2.57	2.36	2.23	2.14	2.08	2.02	1.98	1.95	1.92
22	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90
23	2.94	2.55	2.34	2.21	2.11	2.05	1.99	1.95	1.92	1.89
24	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87
26	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86
27	2.90	2.51	2.30	2.17	2.07	2.00	1.95	1.91	1.87	1.85
28	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84
29	2.89	2.50	2.28	2.15	2.06	1.99	1.93	1.89	1.86	1.83
30	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82
40	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73
60	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71
70	2.78	2.38	2.16	2.03	1.93	1.86	1.80	1.76	1.72	1.69
80	2.77	2.37	2.15	2.02	1.92	1.85	1.79	1.75	1.71	1.68
90	2.76	2.36	2.15	2.01	1.91	1.84	1.78	1.74	1.70	1.67
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66

Tabla A5: Distribución F
 $F_{.10,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	11	12	13	14	15	20	30	40	50	100
1	60.47	60.71	60.90	61.07	61.22	61.74	62.26	62.53	62.69	63.01
2	9.40	9.41	9.41	9.42	9.42	9.44	9.46	9.47	9.47	9.48
3	5.22	5.22	5.21	5.20	5.20	5.18	5.17	5.16	5.15	5.14
4	3.91	3.90	3.89	3.88	3.87	3.84	3.82	3.80	3.80	3.78
5	3.28	3.27	3.26	3.25	3.24	3.21	3.17	3.16	3.15	3.13
6	2.92	2.90	2.89	2.88	2.87	2.84	2.80	2.78	2.77	2.75
7	2.68	2.67	2.65	2.64	2.63	2.59	2.56	2.54	2.52	2.50
8	2.52	2.50	2.49	2.48	2.46	2.42	2.38	2.36	2.35	2.32
9	2.40	2.38	2.36	2.35	2.34	2.30	2.25	2.23	2.22	2.19
10	2.30	2.28	2.27	2.26	2.24	2.20	2.16	2.13	2.12	2.09
11	2.23	2.21	2.19	2.18	2.17	2.12	2.08	2.05	2.04	2.01
12	2.17	2.15	2.13	2.12	2.10	2.06	2.01	1.99	1.97	1.94
13	2.12	2.10	2.08	2.07	2.05	2.01	1.96	1.93	1.92	1.88
14	2.07	2.05	2.04	2.02	2.01	1.96	1.91	1.89	1.87	1.83
15	2.04	2.02	2.00	1.99	1.97	1.92	1.87	1.85	1.83	1.79
16	2.01	1.99	1.97	1.95	1.94	1.89	1.84	1.81	1.79	1.76
17	1.98	1.96	1.94	1.93	1.91	1.86	1.81	1.78	1.76	1.73
18	1.95	1.93	1.92	1.90	1.89	1.84	1.78	1.75	1.74	1.70
19	1.93	1.91	1.89	1.88	1.86	1.81	1.76	1.73	1.71	1.67
20	1.91	1.89	1.87	1.86	1.84	1.79	1.74	1.71	1.69	1.65
21	1.90	1.87	1.86	1.84	1.83	1.78	1.72	1.69	1.67	1.63
22	1.88	1.86	1.84	1.83	1.81	1.76	1.70	1.67	1.65	1.61
23	1.87	1.84	1.83	1.81	1.80	1.74	1.69	1.66	1.64	1.59
24	1.85	1.83	1.81	1.80	1.78	1.73	1.67	1.64	1.62	1.58
25	1.84	1.82	1.80	1.79	1.77	1.72	1.66	1.63	1.61	1.56
26	1.83	1.81	1.79	1.77	1.76	1.71	1.65	1.61	1.59	1.55
27	1.82	1.80	1.78	1.76	1.75	1.70	1.64	1.60	1.58	1.54
28	1.81	1.79	1.77	1.75	1.74	1.69	1.63	1.59	1.57	1.53
29	1.80	1.78	1.76	1.75	1.73	1.68	1.62	1.58	1.56	1.52
30	1.79	1.77	1.75	1.74	1.72	1.67	1.61	1.57	1.55	1.51
40	1.74	1.71	1.70	1.68	1.66	1.61	1.54	1.51	1.48	1.43
50	1.70	1.68	1.66	1.64	1.63	1.57	1.50	1.46	1.44	1.39
60	1.68	1.66	1.64	1.62	1.60	1.54	1.48	1.44	1.41	1.36
70	1.66	1.64	1.62	1.60	1.59	1.53	1.46	1.42	1.39	1.34
80	1.65	1.63	1.61	1.59	1.57	1.51	1.44	1.40	1.38	1.32
90	1.64	1.62	1.60	1.58	1.56	1.50	1.43	1.39	1.36	1.30
100	1.64	1.61	1.59	1.57	1.56	1.49	1.42	1.38	1.35	1.29

Tabla A6: Distribución F
 $F_{.05,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	1	2	3	4	5	6	7	8	9	10
1	161.4	199.5	215.7	224.5	230.1	233.9	236.7	238.8	240.5	241.8
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.22	2.18
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97
80	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95
90	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93

Tabla A6: Distribución F
 $F_{.05,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	11	12	13	14	15	20	30	40	50	100
1	242.9	243.9	244.6	245.3	245.9	248.0	250.1	251.1	251.7	253.0
2	19.40	19.41	19.42	19.42	19.43	19.45	19.46	19.47	19.48	19.49
3	8.76	8.74	8.73	8.71	8.70	8.66	8.62	8.59	8.58	8.55
4	5.94	5.91	5.89	5.87	5.86	5.80	5.75	5.72	5.70	5.66
5	4.70	4.68	4.66	4.64	4.62	4.56	4.50	4.46	4.44	4.41
6	4.03	4.00	3.98	3.96	3.94	3.87	3.81	3.77	3.75	3.71
7	3.60	3.57	3.55	3.53	3.51	3.44	3.38	3.34	3.32	3.27
8	3.31	3.28	3.26	3.24	3.22	3.15	3.08	3.04	3.02	2.97
9	3.10	3.07	3.05	3.03	3.01	2.94	2.86	2.83	2.80	2.76
10	2.94	2.91	2.89	2.86	2.85	2.77	2.70	2.66	2.64	2.59
11	2.82	2.79	2.76	2.74	2.72	2.65	2.57	2.53	2.51	2.46
12	2.72	2.69	2.66	2.64	2.62	2.54	2.47	2.43	2.40	2.35
13	2.63	2.60	2.58	2.55	2.53	2.46	2.38	2.34	2.31	2.26
14	2.57	2.53	2.51	2.48	2.46	2.39	2.31	2.27	2.24	2.19
15	2.51	2.48	2.45	2.42	2.40	2.33	2.25	2.20	2.18	2.12
16	2.46	2.42	2.40	2.37	2.35	2.28	2.19	2.15	2.12	2.07
17	2.41	2.38	2.35	2.33	2.31	2.23	2.15	2.10	2.08	2.02
18	2.37	2.34	2.31	2.29	2.27	2.19	2.11	2.06	2.04	1.98
19	2.34	2.31	2.28	2.26	2.23	2.16	2.07	2.03	2.00	1.94
20	2.31	2.28	2.25	2.22	2.20	2.12	2.04	1.99	1.97	1.91
21	2.28	2.25	2.22	2.20	2.18	2.10	2.01	1.96	1.94	1.88
22	2.26	2.23	2.20	2.17	2.15	2.07	1.98	1.94	1.91	1.85
23	2.24	2.20	2.18	2.15	2.13	2.05	1.96	1.91	1.88	1.82
24	2.22	2.18	2.15	2.13	2.11	2.03	1.94	1.89	1.86	1.80
25	2.20	2.16	2.14	2.11	2.09	2.01	1.92	1.87	1.84	1.78
26	2.18	2.15	2.12	2.09	2.07	1.99	1.90	1.85	1.82	1.76
27	2.17	2.13	2.10	2.08	2.06	1.97	1.88	1.84	1.81	1.74
28	2.15	2.12	2.09	2.06	2.04	1.96	1.87	1.82	1.79	1.73
29	2.14	2.10	2.08	2.05	2.03	1.94	1.85	1.81	1.77	1.71
30	2.13	2.09	2.06	2.04	2.01	1.93	1.84	1.79	1.76	1.70
40	2.04	2.00	1.97	1.95	1.92	1.84	1.74	1.69	1.66	1.59
50	1.99	1.95	1.92	1.89	1.87	1.78	1.69	1.63	1.60	1.52
60	1.95	1.92	1.89	1.86	1.84	1.75	1.65	1.59	1.56	1.48
70	1.93	1.89	1.86	1.84	1.81	1.72	1.62	1.57	1.53	1.45
80	1.91	1.88	1.84	1.82	1.79	1.70	1.60	1.54	1.51	1.43
90	1.90	1.86	1.83	1.80	1.78	1.69	1.59	1.53	1.49	1.41
100	1.89	1.85	1.82	1.79	1.77	1.68	1.57	1.52	1.48	1.39

Tabla A7: Distribución F
 $F_{.025,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	1	2	3	4	5	6	7	8	9	10
1	647.7	799.5	864.1	899.5	921.8	937.1	948.2	956.6	963.2	968.6
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77
21	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73
22	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70
23	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67
24	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61
26	5.66	4.27	3.67	3.33	3.10	2.94	2.82	2.73	2.65	2.59
27	5.63	4.24	3.65	3.31	3.08	2.92	2.80	2.71	2.63	2.57
28	5.61	4.22	3.63	3.29	3.06	2.90	2.78	2.69	2.61	2.55
29	5.59	4.20	3.61	3.27	3.04	2.88	2.76	2.67	2.59	2.53
30	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51
40	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32
60	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27
70	5.25	3.89	3.31	2.97	2.75	2.59	2.47	2.38	2.30	2.24
80	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28	2.21
90	5.20	3.84	3.26	2.93	2.71	2.55	2.43	2.34	2.26	2.19
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18

Tabla A7: Distribución F
 $F_{.025,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	11	12	13	14	15	20	30	40	50	100
1	973.0	976.7	979.8	982.5	984.8	993.1	1001.4	1005	1008	1013
2	39.41	39.41	39.42	39.43	39.43	39.45	39.46	39.47	39.48	39.49
3	14.37	14.34	14.30	14.28	14.25	14.17	14.08	14.04	14.01	13.96
4	8.79	8.75	8.71	8.68	8.66	8.56	8.46	8.41	8.38	8.32
5	6.57	6.52	6.49	6.46	6.43	6.33	6.23	6.18	6.14	6.08
6	5.41	5.37	5.33	5.30	5.27	5.17	5.07	5.01	4.98	4.92
7	4.71	4.67	4.63	4.60	4.57	4.47	4.36	4.31	4.28	4.21
8	4.24	4.20	4.16	4.13	4.10	4.00	3.89	3.84	3.81	3.74
9	3.91	3.87	3.83	3.80	3.77	3.67	3.56	3.51	3.47	3.40
10	3.66	3.62	3.58	3.55	3.52	3.42	3.31	3.26	3.22	3.15
11	3.47	3.43	3.39	3.36	3.33	3.23	3.12	3.06	3.03	2.96
12	3.32	3.28	3.24	3.21	3.18	3.07	2.96	2.91	2.87	2.80
13	3.20	3.15	3.12	3.08	3.05	2.95	2.84	2.78	2.74	2.67
14	3.09	3.05	3.01	2.98	2.95	2.84	2.73	2.67	2.64	2.56
15	3.01	2.96	2.92	2.89	2.86	2.76	2.64	2.59	2.55	2.47
16	2.93	2.89	2.85	2.82	2.79	2.68	2.57	2.51	2.47	2.40
17	2.87	2.82	2.79	2.75	2.72	2.62	2.50	2.44	2.41	2.33
18	2.81	2.77	2.73	2.70	2.67	2.56	2.44	2.38	2.35	2.27
19	2.76	2.72	2.68	2.65	2.62	2.51	2.39	2.33	2.30	2.22
20	2.72	2.68	2.64	2.60	2.57	2.46	2.35	2.29	2.25	2.17
21	2.68	2.64	2.60	2.56	2.53	2.42	2.31	2.25	2.21	2.13
22	2.65	2.60	2.56	2.53	2.50	2.39	2.27	2.21	2.17	2.09
23	2.62	2.57	2.53	2.50	2.47	2.36	2.24	2.18	2.14	2.06
24	2.59	2.54	2.50	2.47	2.44	2.33	2.21	2.15	2.11	2.02
25	2.56	2.51	2.48	2.44	2.41	2.30	2.18	2.12	2.08	2.00
26	2.54	2.49	2.45	2.42	2.39	2.28	2.16	2.09	2.05	1.97
27	2.51	2.47	2.43	2.39	2.36	2.25	2.13	2.07	2.03	1.94
28	2.49	2.45	2.41	2.37	2.34	2.23	2.11	2.05	2.01	1.92
29	2.48	2.43	2.39	2.36	2.32	2.21	2.09	2.03	1.99	1.90
30	2.46	2.41	2.37	2.34	2.31	2.20	2.07	2.01	1.97	1.88
40	2.33	2.29	2.25	2.21	2.18	2.07	1.94	1.88	1.83	1.74
50	2.26	2.22	2.18	2.14	2.11	1.99	1.87	1.80	1.75	1.66
60	2.22	2.17	2.13	2.09	2.06	1.94	1.82	1.74	1.70	1.60
70	2.18	2.14	2.10	2.06	2.03	1.91	1.78	1.71	1.66	1.56
80	2.16	2.11	2.07	2.03	2.00	1.88	1.75	1.68	1.63	1.53
90	2.14	2.09	2.05	2.02	1.98	1.86	1.73	1.66	1.61	1.50
100	2.12	2.08	2.04	2.00	1.97	1.85	1.71	1.64	1.59	1.48

Tabla A8: Distribución F
 $F_{.01,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	1	2	3	4	5	6	7	8	9	10
1	4052	4999	5403	5624	5763	5858	5928	5981	6022	6055
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13
26	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09
27	7.68	5.49	4.60	4.11	3.78	3.56	3.39	3.26	3.15	3.06
28	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03
29	7.60	5.42	4.54	4.04	3.73	3.50	3.33	3.20	3.09	3.00
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63
70	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59
80	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55
90	6.93	4.85	4.01	3.53	3.23	3.01	2.84	2.72	2.61	2.52
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50

Tabla A8: Distribución F
 $F_{.01,n,m}$
 Grados de libertad para el numerador (n)
 Grados de libertad para el denominador (m)

m \ n	11	12	13	14	15	20	30	40	50	100
1	6083	6106	6125	6142	6157	6208	6260	6286	6302	6334
2	99.41	99.42	99.42	99.43	99.43	99.45	99.47	99.47	99.48	99.49
3	27.13	27.05	26.98	26.92	26.87	26.69	26.50	26.41	26.35	26.24
4	14.45	14.37	14.31	14.25	14.20	14.02	13.84	13.75	13.69	13.58
5	9.96	9.89	9.82	9.77	9.72	9.55	9.38	9.29	9.24	9.13
6	7.79	7.72	7.66	7.60	7.56	7.40	7.23	7.14	7.09	6.99
7	6.54	6.47	6.41	6.36	6.31	6.16	5.99	5.91	5.86	5.75
8	5.73	5.67	5.61	5.56	5.52	5.36	5.20	5.12	5.07	4.96
9	5.18	5.11	5.05	5.01	4.96	4.81	4.65	4.57	4.52	4.41
10	4.77	4.71	4.65	4.60	4.56	4.41	4.25	4.17	4.12	4.01
11	4.46	4.40	4.34	4.29	4.25	4.10	3.94	3.86	3.81	3.71
12	4.22	4.16	4.10	4.05	4.01	3.86	3.70	3.62	3.57	3.47
13	4.02	3.96	3.91	3.86	3.82	3.66	3.51	3.43	3.38	3.27
14	3.86	3.80	3.75	3.70	3.66	3.51	3.35	3.27	3.22	3.11
15	3.73	3.67	3.61	3.56	3.52	3.37	3.21	3.13	3.08	2.98
16	3.62	3.55	3.50	3.45	3.41	3.26	3.10	3.02	2.97	2.86
17	3.52	3.46	3.40	3.35	3.31	3.16	3.00	2.92	2.87	2.76
18	3.43	3.37	3.32	3.27	3.23	3.08	2.92	2.84	2.78	2.68
19	3.36	3.30	3.24	3.19	3.15	3.00	2.84	2.76	2.71	2.60
20	3.29	3.23	3.18	3.13	3.09	2.94	2.78	2.69	2.64	2.54
21	3.24	3.17	3.12	3.07	3.03	2.88	2.72	2.64	2.58	2.48
22	3.18	3.12	3.07	3.02	2.98	2.83	2.67	2.58	2.53	2.42
23	3.14	3.07	3.02	2.97	2.93	2.78	2.62	2.54	2.48	2.37
24	3.09	3.03	2.98	2.93	2.89	2.74	2.58	2.49	2.44	2.33
25	3.06	2.99	2.94	2.89	2.85	2.70	2.54	2.45	2.40	2.29
26	3.02	2.96	2.90	2.86	2.81	2.66	2.50	2.42	2.36	2.25
27	2.99	2.93	2.87	2.82	2.78	2.63	2.47	2.38	2.33	2.22
28	2.96	2.90	2.84	2.79	2.75	2.60	2.44	2.35	2.30	2.19
29	2.93	2.87	2.81	2.77	2.73	2.57	2.41	2.33	2.27	2.16
30	2.91	2.84	2.79	2.74	2.70	2.55	2.39	2.30	2.25	2.13
40	2.73	2.66	2.61	2.56	2.52	2.37	2.20	2.11	2.06	1.94
50	2.63	2.56	2.51	2.46	2.42	2.27	2.10	2.01	1.95	1.82
60	2.56	2.50	2.44	2.39	2.35	2.20	2.03	1.94	1.88	1.75
70	2.51	2.45	2.40	2.35	2.31	2.15	1.98	1.89	1.83	1.70
80	2.48	2.42	2.36	2.31	2.27	2.12	1.94	1.85	1.79	1.65
90	2.45	2.39	2.33	2.29	2.24	2.09	1.92	1.82	1.76	1.62
100	2.43	2.37	2.31	2.27	2.22	2.07	1.89	1.80	1.74	1.60

Tabla A9: Estadístico Durbin-Watson
k = Número de predictores (sin incluir el intercepto)

Cola izquierda		1		2		3		4		5	
Muestra	α	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	0.01	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.96
	0.025	0.95	1.23	0.83	1.40	0.71	1.61	0.59	1.84	0.48	2.09
	0.05	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
20	0.01	0.95	1.15	0.86	1.27	0.77	1.41	0.63	1.57	0.60	1.74
	0.025	1.08	1.28	0.99	1.41	0.89	1.55	0.79	1.70	0.70	1.87
	0.05	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	0.01	1.05	1.21	0.98	1.30	0.90	1.41	0.83	1.52	0.75	1.65
	0.025	1.13	1.34	1.10	1.43	1.02	1.54	0.94	1.65	0.86	1.77
	0.05	1.20	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
30	0.01	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
	0.025	1.25	1.38	1.18	1.46	1.12	1.54	1.05	1.63	0.98	1.73
	0.05	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
40	0.01	1.25	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.05	1.58
	0.025	1.35	1.45	1.30	1.51	1.25	1.57	1.20	1.63	1.15	1.69
	0.05	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
50	0.01	1.32	1.40	1.28	1.45	1.24	1.49	1.20	1.54	1.16	1.59
	0.025	1.42	1.50	1.38	1.54	1.34	1.59	1.30	1.64	1.26	1.69
	0.05	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
60	0.01	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
	0.025	1.47	1.54	1.44	1.57	1.40	1.61	1.37	1.65	1.33	1.69
	0.05	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
80	0.01	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
	0.025	1.54	1.59	1.52	1.62	1.49	1.65	1.47	1.67	1.44	1.70
	0.05	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
100	0.01	1.52	1.56	1.50	1.58	1.48	1.60	1.45	1.63	1.44	1.65
	0.025	1.59	1.63	1.57	1.65	1.55	1.67	1.53	1.70	1.51	1.72
	0.05	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

Índice alfabético

- Índice de Kullback-Lieber, 175
- acciones de atención, 32, 33, 36, 52–54, 57, 64, 65, 178, 183
- ajuste del modelo a los datos, 3
- algoritmo de Durbin-Levinson, 138, 145, 146, 149
- algoritmo de innovación, 152, 155, 166
- análisis de regresión lineal, 37, 70
- Anderson-Darling, 20, 21, 28
- ANOVA, 12, 13
- AR(1), 22, 23, 93, 99, 100, 106, 109, 110, 112, 113, 117, 130, 134, 135, 142, 143, 156, 162, 163, 165, 171, 172, 247
- AR(1) explosivo, 110
- AR(2), 101–103, 120, 134, 138, 143, 144, 171, 181, 249–252
- AR(p), 93, 104, 106, 107, 130, 133, 135–137, 148, 149, 155, 163, 171, 175, 181, 196, 197, 201, 236
- ARCH(1), 221
- ARIMA(0,0,0), 254
- ARIMA(1,0,2), 253
- ARIMA(1,1,1), 246, 249
- ARIMA(p,d,q), 195–197, 199, 200, 210, 212, 234, 246–254, 256
- ARMA(1,1), 106, 108, 113, 115, 172, 173, 181, 183, 185
- ARMA(1,2), 189
- ARMA(2,1), 181, 183, 185, 187
- ARMA(2,2), 113, 151
- ARMA(2,3), 193, 252
- ARMA(p,q), 43, 81, 85, 88, 106, 107, 109, 111, 112, 117–119, 122, 123, 125, 126, 129, 130, 132, 135, 136, 145, 148, 149, 151, 152, 160, 162, 165, 166, 170, 174, 176–178, 180, 183, 187, 189, 191, 195, 213, 214, 220, 225, 234–236, 248–250, 252, 255, 267, 268
- autocorrelación parcial, 133, 134
- autorregresivo de orden p, 104
- Box & Jenkins, 30, 75, 196
- cadena de Markov, 232
- calidad del ajuste, 3
- causalidad, 3, 109
- cociente estacional, 34
- coeficiente de determinación, 15
- coeficiente de variación, 34
- coeficientes de la regresión, 1
- componente irregular, 31, 32, 48, 72
- Condusef, 32, 33, 52–54, 64, 65, 178, 183, 187
- consistentes, 10
- correlación de Spearman, 6, 7
- criterio de información Bayesiana (BIC), 175, 177, 181
- criterio de información de Akaike corregido, 175–177, 181, 248, 251
- criterio de información de Akaike sin corregir, 175–177, 248, 251
- criterios de información, 166
- curtosis, 18
- débilmente estacionario, 76
- datos de series de tiempo, 22
- datos de tipo longitudinales, 30
- datos de tipo transversales, 30
- deflactación, 31
- deflactor del PIB, 31
- descomposición clásica, 30
- desigualdad de Cauchy, 76, 79
- desigualdad de Jensen, 176
- diagrama de caja y bigotes, 18, 21, 28
- diferencia estacional, 34
- distribución a posteriori, 226, 227
- distribución a priori, 226
- distribución F, 13
- distribución ji-cuadrada, 5
- distribución Normal, 14, 156
- distribución t, 12, 14
- Durbin-Watson, 22–24, 28, 246
- ecuación de observación, 232, 235, 237
- ecuación del sistema, 232, 235, 237
- ecuaciones de predicción, 141, 147

ecuaciones de Yule-Walker, 103–105, 134–138, 145
 ecuaciones normales por mínimos cuadrados, 4
 enfoque determinista, 30
 enfoque estocástico, 30, 75
 error cuadrático medio, 5, 52, 56, 127, 128, 134, 139, 141–143, 145, 166, 175
 error de predicción final (FPE), 175, 177, 181
 esquema aditivo, 33, 35, 36, 48, 65, 68, 74
 esquema multiplicativo, 33–36
 estacionalidad, 31, 32, 36, 37, 48, 61–63, 69, 71, 72, 74, 83–85, 88, 178, 183, 187, 223
 estadísticas suficientes, 10
 estimador máximo verosímil, 157

 fórmula general, 103
 filtro de estacionalidad, 62, 88
 filtro de Kalman, 236
 filtro de promedios móviles, 85
 fuertemente estacionario, 76
 función de autocorrelación, 76
 función de autocovarianza, 76, 79
 función de verosimilitud, 157–159, 226–230

 GARCH(1,1), 222
 gráfica de probabilidad normal, 17, 18, 21, 28, 167, 243, 245

 heteroscedasticidad, 18, 28
 histograma, 18, 21, 28, 167
 homoscedasticidad, 17, 20

 Índice de Kullback-Lieber, 175
 intercepto, 1, 10, 11, 14

 Kolmogorov-Smirnov, 167

 Lilliefors, 194, 219

 mínimos cuadrados ordinarios, 3, 5, 7–10, 12, 22, 37, 42, 44, 247
 máxima verosimilitud, 3, 9, 10, 155, 156, 166, 170, 176, 181, 185, 189, 193, 201
 método de descomposición clásica, 187, 191
 método de Holt, 29, 53, 58, 59, 63, 85
 método de Holt-Winters, 29, 53, 63–66, 68, 88, 183, 187
 método del ajuste analítico, 37
 MA(1), 91–94, 96, 121, 122, 130, 155, 158, 161, 172
 MA(2), 94–96, 172
 MA(3), 252
 MA(q), 97, 98, 105–107, 121, 130, 133, 137, 149, 152, 155, 196, 197
 matriz de autocorrelaciones muestrales, 130
 matriz de autocovarianzas muestrales, 129
 mejor predictor lineal, 139, 141–144
 metodología ARIMA, 29, 30, 75
 modelo aditivo, 36
 modelo de espacio de estados, 223
 modelo de regresión dinámica, 253, 268
 modelo de regresión lineal, 1, 3, 9, 15, 16, 249, 253
 modelo de regresión lineal múltiple, 247
 modelo de regresión lineal simple, 2, 3, 5, 9, 16, 22, 24, 85, 234, 246, 247, 251
 modelo de regresión muestral, 4
 modelo dinámico lineal, 232, 234, 237
 modelo en diferencias, 248
 modelo en niveles, 248
 modelo multiplicativo, 36
 modelos ARCH, 220
 modelos de espacio de estados, 225
 modelos GARCH, 222
 modelos inestables, 16

 nivel, 57

 pendiente, 1, 10, 11, 14, 56, 57
 PIB nominal, 31
 PIB real, 31
 poder adquisitivo, 31
 precios constantes, 31
 precios corrientes, 31
 proceso autorregresivo, 22
 proceso estacionario, 80, 81, 85, 88–90, 92, 98, 112, 127, 128, 132, 148
 proceso estocástico, 30
 promedios móviles, 29, 37, 47, 48
 promedios móviles de orden q, 97, 114
 prueba de diferencia en signos, 169
 prueba de Ljung y Box, 168, 172, 182, 187, 191, 194, 253, 256, 258, 261, 264
 prueba Dickey-Fuller aumentada, 162–165, 207, 219, 220, 255, 256, 258, 260, 262
 prueba Dickey-Fuller simple, 162, 163
 prueba KPSS, 164, 165, 207, 219, 220, 253, 256, 258, 260, 262
 prueba Kruskal-Wallis, 73, 74
 prueba McLeod-Li, 168, 187, 191, 194

 regresión espuria, 248
 residuos estandarizados, 16, 18

residuos estudentizados, 16, 18
 ruido blanco, 80, 81, 94, 97–99, 104, 106, 116, 122, 133, 138, 151, 152, 156, 173, 195, 202
 serie de tiempo, 29, 30, 33, 34, 37, 68, 74, 75, 83–86, 150, 162, 163, 177, 195, 232, 233, 251
 sesgo, 18, 21
 Shapiro-Wilk, 194, 204, 219
 significancia de la regresión, 12
 Sistema Financiero Mexicano, 32
 suavizamiento exponencial simple, 52–54, 57
 sucesión de autocorrelaciones (ACF), 172
 sucesión de autocorrelaciones muestrales (ACF), 129, 130, 133, 149–152, 162, 166, 167, 180, 182, 183, 187, 189, 191, 193, 195–197, 199–201, 214, 216, 251, 252, 268
 sucesión de autocorrelaciones muestrales parciales (PACF), 166, 180, 183, 189, 193, 196, 197, 200, 201, 214, 216, 252, 268
 sucesión de autocorrelaciones parciales (PACF), 130, 132, 134, 135, 149, 151, 152, 189, 251
 sucesión de autocorrelaciones teórica (ACF), 80, 92, 94–96, 98, 100, 103, 105–108, 110, 129, 130, 152
 sucesión de autocovarianzas, 89–91, 107, 111–113, 125, 128, 139, 148
 sucesión de autocovarianzas muestrales, 129
 sucesión de autocovarianzas teórica, 79, 80, 92, 94, 97–100, 106–108, 113, 124–126, 129, 136, 143
 suma de cuadrados total, 5
 suma de los errores al cuadrado, 4, 5
 tendencia, 31–33, 36, 37, 42, 47, 48, 56, 62, 63, 69, 70, 72, 83–88, 178, 183, 187, 197, 233, 258, 260
 tendencia cuadrática, 42
 tendencia determinista, 255, 264
 tendencia estocástica, 255, 264
 tendencia exponencial, 44
 tendencia lineal, 42
 Teorema de Bayes, 226
 Teorema de Gauss-Markov, 8, 10, 37
 Teorema de proyección ortogonal, 147
 Teorema de Schur, 102
 transformación de Box-Cox, 211, 212
 validez del modelo, 3
 valores extremos, 16, 17
 variable aleatoria, 1, 2
 variable explicativa, 2, 15, 24, 28, 246
 variable respuesta, 2, 3, 6, 7, 12, 16, 24, 37, 246
 verosimilitud estandarizada, 226, 227