

CUADERNOS DE MATEMÁTICA  
DE LA ESCUELA POLITÉCNICA NACIONAL

HOLGER CAPA SANTOS

---

---

SERIES TEMPORALES

*La ciencia y el arte de los pronósticos*

---

---



## **Cuaderno de Matemática No. 10**

**SERIES TEMPORALES: LA CIENCIA Y EL ARTE DE LOS PRONÓSTICOS**

**HOLGER CAPA SANTOS**

**Responsable de la Edición: ?**

**Revisión técnica: ?**

**Asistentes: Andrés Merino**

**Portada: Andrés Merino**

Registro de derecho autoral No.

ISBN:

Publicado por la Unidad de Publicaciones de la Facultad de Ciencias de la Escuela Politécnica Nacional, Ladrón de Guevara E11-253, Quito, Ecuador.

Primera edición: 2015

Primera impresión: 2015

© Escuela Politécnica Nacional 2015

# Índice general

<b>1</b>	<b>Análisis a través de procesos estacionarios</b>	<b>3</b>
1.1	Introducción	3
1.2	Procesos Estacionarios	5
1.3	Representación Espectral de un Proceso Débilmente Estacionario	12
1.3.1	Representación espectral de $(\gamma_t)_{t \in \mathbb{Z}}$	12
1.3.2	Representación espectral de $(X_t)$	13
1.4	Teoremas Límites	18
1.4.1	Teorema Central del Límite	20
1.5	Predicción de un Proceso Estacionario y Descomposición de Wold	20
1.5.1	Predicción de un Proceso Estacionario	20
1.5.2	Descomposición de Wold	21
1.6	Ejercicios Propuestos	21
<b>2</b>	<b>Procesos Lineales, Modelos ARMA</b>	<b>25</b>
2.1	Procesos Lineales	25
2.2	Funciones de Autorrelación Simple y Parcial de un P.E.S.O.	28
2.3	Procesos Autoregresivos	29
2.3.1	El caso del modelo $AR(1)$	29
2.3.2	Unicidad de la descomposición para el caso del $AR(p)$ :	30
2.3.3	Autocorrelación asintótica de un proceso $AR(p)$	32
2.3.4	Autorrelación parcial asintótica de un proceso $AR(p)$	33
2.4	Procesos Medias Móviles	38
2.4.1	Autocorrelación de un $MA(q)$	39
2.4.2	Autocorrelación Parcial de un $MA(q)$	39
2.5	Procesos ARMA (Autoregresivos – Medias Móviles)	40
2.6	Procesos ARIMA (Autoregressive Integrates Moving Avarage)	43
2.6.1	Representación MA de un ARIMA	45
2.6.2	Representación AR de un ARIMA:	45
2.7	Ejercicios Propuestos	54
<b>3</b>	<b>Estimación, Verificación y Predicción en un Modelo ARIMA</b>	<b>57</b>
3.1	Metodología Box–Jenkins	57
3.2	Identificación <i>A Priori</i>	58
3.2.1	Elección de $d$	58
3.2.2	Elección de $p$ y $q$	59
3.3	Estimación de un Modelo ARIMA	60

3.3.1	Procedimiento condicional . . . . .	61
3.3.2	Procedimiento no condicional . . . . .	63
3.4	Identificación <i>A Posteriori</i> . . . . .	65
3.4.1	Fase de verificación . . . . .	65
3.4.2	Utilización de los residuos estimados para modificar el Modelo . . . . .	69
3.4.3	Pruebas de estabilidad . . . . .	69
3.4.4	Elección del Modelo . . . . .	71
3.5	Predicción de los Modelos ARIMA . . . . .	72
3.5.1	Cálculo de las predicciones óptimas en un modelo ARIMA . . . . .	72
3.5.2	Intervalos de predicción . . . . .	76
3.5.3	Funciones de predicción . . . . .	77
3.5.4	Contraste de la estabilidad estructural en la predicción . . . . .	78
3.6	Transformación de Datos . . . . .	79
3.7	Raíces Unitarias . . . . .	81
3.7.1	La prueba de Dickey-Fuller . . . . .	82
3.7.2	Extensiones de la prueba de Dickey-Fuller . . . . .	83
3.8	Modelos Estacionales . . . . .	85
3.9	Ejercicios Propuestos . . . . .	89
<b>4</b>	<b>Modelos de Heteroscedasticidad Condicional</b>	<b>93</b>
4.1	Modelos Arch–Garch Simétricos . . . . .	93
4.2	Modelos GARCH asimétricos . . . . .	95
4.3	Metodología de la Modelación . . . . .	97
4.4	Ejemplos con Heteroscedasticidad Condicional . . . . .	98
<b>5</b>	<b>Modelos Multivariantes de Series Temporales</b>	<b>103</b>
5.1	Procesos Estacionarios . . . . .	103
5.1.1	Proceso estrictamente estacionario . . . . .	103
5.1.2	Proceso débilmente estacionario . . . . .	103
5.2	Matrices de Correlación Cruzada (Cross-Correlation) . . . . .	104
5.2.1	Dependencia Lineal . . . . .	105
5.3	Modelos de Vectores Autoregresivos (VAR) . . . . .	109
5.3.1	El caso bivariante . . . . .	109
5.3.2	Representación general de un VAR . . . . .	111
5.4	Representación VARMA de una Serie Multivariante . . . . .	113
5.4.1	La representación VMA . . . . .	113
5.4.2	Representación lineal de un VAR(p) . . . . .	113
5.4.3	La representación VARMA . . . . .	115
5.5	Formulación de un modelo VAR . . . . .	116
5.5.1	Estimación . . . . .	116
5.5.2	Determinación del número de retardos . . . . .	117
5.5.3	Diagnóstico y validación del modelo . . . . .	118
5.6	Predicción . . . . .	121
5.7	La Causalidad . . . . .	132
5.7.1	La causalidad según Granger . . . . .	133
5.7.2	La causalidad según Sims . . . . .	134
5.8	Análisis de los “choques” . . . . .	136

---

5.9	Descomposición de la varianza . . . . .	144
5.9.1	Representación de errores ortogonales . . . . .	144
5.9.2	Descomposición de la Varianza . . . . .	144
5.9.3	Elección del orden de descomposición . . . . .	145
5.10	Ejemplo Práctico . . . . .	149
<b>6</b>	<b>La Cointegración</b>	<b>165</b>
6.1	Propiedades del orden de Integración de una serie . . . . .	165
6.2	Condiciones de cointegración . . . . .	167
<b>7</b>	<b>Modelo de Corrección del error (MCE)</b>	<b>171</b>
7.1	Planteamiento del MCE . . . . .	171
7.2	Prueba de Johansen . . . . .	173
7.3	Prueba de Exogeneidad Débil . . . . .	185
7.4	Cointegración Estacional . . . . .	186
	<b>Índice Alfabético</b>	<b>187</b>



# Índice de figuras

1.1	Ajuste adecuado de datos de tendencia cuadrática por una curva polinomial .	4
1.2	Ajuste no adecuado de una serie de ventas a través de una curva cuadrática .	4
1.3	Modelación de una serie de ventas a través de un modelo SARIMA (se estudiará más adelante) . . . . .	5
1.4	Ruido blanco fuerte simulado con distribución $N(0, 1)$ . . . . .	7
1.5	Proceso estacionario . . . . .	8
1.6	Proceso no estacionario: Tendencia creciente . . . . .	9
1.7	Proceso no estacionario: Varianza no homogénea . . . . .	10
1.8	Proceso no estacionario: Tendencia decreciente y varianza no homogénea . .	10
2.1	Modelo simulado $AR(eq4)$ . . . . .	34
2.2	Comportamiento de $r(\ell)$ . . . . .	34
2.3	Comportamiento de $r(\ell)$ . . . . .	34
2.4	Modelo $MA(1)$ . . . . .	40
2.5	Comportamiento de $\rho(l)$ . . . . .	40
2.6	Comportamiento asintótico de $r(l)$ . . . . .	40
2.7	Modelo $ARMA(1, 1)$ . . . . .	42
2.8	Comportamiento de $\rho(l)$ . . . . .	42
2.9	Comportamiento de $r(l)$ . . . . .	42
2.10	Efecto de la diferenciación en una tendencia lineal . . . . .	43
2.11	Efecto de la diferenciación de una tendencia cuadrática . . . . .	43
2.12	Efecto de la diferenciación de una tendencia exponencial . . . . .	44
2.13	Efecto de la transformación logarítmica en una tendencia exponencial . . . .	44
3.1	Serie de un Índice Bursátil (SIB) . . . . .	58
3.2	Función de autocorrelación de la SIB . . . . .	58
3.3	Serie estacionaria obtenida a partir de la SIB . . . . .	58
3.4	FAC estimada del modelo $ARIMA(0, 1, 0)$ , para la SIB . . . . .	60
3.5	FACP estimada del modelo $ARIMA(0, 1, 0)$ , para la SIB en primera diferencia	60
3.6	FAC residual estimada del Modelo $ARIMA(2, 1, 2)$ de la SIB . . . . .	68
3.7	FACP residual estimada del Modelo $ARIMA(2, 1, 2)$ de la SIB . . . . .	68
3.8	FAC residual estimada del Modelo $ARIMA(2, 1, 1)$ de la SIB . . . . .	68
3.9	FACP residual estimada del Modelo $ARIMA(2, 1, 1)$ de la SIB . . . . .	68
3.10	FAC residual estimada del Modelo $ARIMA(1, 1, 1)$ de la SIB . . . . .	68
3.11	FACP residual estimada del Modelo $ARIMA(1, 1, 1)$ de la SIB . . . . .	68

3.12	FAC y FACP residuales estimadas del Modelo $ARIMA(1,1,1)$ de la SIB en EViews . . . . .	69
3.13	Prueba de Chow para el punto de quiebre. . . . .	70
3.14	Prueba de predicciones de Chow . . . . .	79
3.15	Grafico de SVM . . . . .	86
3.16	FAC y FACP estimadas de SVM . . . . .	86
3.17	FAC Y FACP estimadas de la SVM con una diferencia estacional . . . . .	87
3.18	FAC y FACP residuales estimadas del Modelo 1 para SVM. . . . .	88
3.19	FAC y FACP estimadas del Modelo 3 para la SVM . . . . .	89
3.20	Prueba de Chow para las predicciones. . . . .	89
3.21	Comparación de predicciones con valores reales (2015) . . . . .	89
4.1	FAC y FACP estimadas residuales del Modelo 3 para la SVM . . . . .	99
4.2	FAC y FACP estimadas de los residuos cuadráticos del Modelo 3 para la SVM . . . . .	99
4.3	FAC y FACP estimadas residuales del Modelo 3-ARCH(1) para la SVM . . . . .	101
4.4	FAC y FACP estimadas de los residuos cuadráticos del Modelo 3-ARCH(1) para la SVM . . . . .	101
5.1	Correlaciones cruzadas entre $X_{1t}$ y $X_{2t}$ . . . . .	108
5.2	Correlaciones cruzadas entre $X_{1t}$ y $X_{3t}$ . . . . .	108
5.3	Correlaciones cruzadas entre $X_{2t}$ y $X_{3t}$ . . . . .	109
5.4	Gráfico de las variaciones trimestrales de las series PIB, CI y DFI . . . . .	123
5.5	Especificación de un modelo VAR en EViews . . . . .	125
5.6	Criterio de estabilidad para el VAR(1) estimado . . . . .	126
5.7	Comparación de las predicciones VAR y UNIVARIANTE para $Y_{1t}$ (PIB) . . . . .	131
5.8	Comparación de las predicciones VAR y UNIVARIANTE para $Y_{2t}$ (CI) . . . . .	131
5.9	Comparación de las predicciones VAR y UNIVARIANTE para $Y_{3t}$ (DFI) . . . . .	132
5.10	Respuesta de las variables $Y_{1t}$ , $Y_{2t}$ y $Y_{3t}$ ante un choque unitario de $Y_{1t}$ . . . . .	139
5.11	Respuesta de las variables $Y_{1t}$ , $Y_{2t}$ y $Y_{3t}$ ante un choque unitario de $Y_{2t}$ . . . . .	141
5.12	Correlaciones cruzadas entre $X_{1t}$ y $X_{2t}$ . . . . .	150
5.13	Gráfico de secuencia de las series . . . . .	151
5.14	Criterio de estabilidad para el VAR(1) estimado . . . . .	153
5.15	Comparación de las predicciones VAR y UNIVARIANTE para $X_{1t}$ (IPP) . . . . .	158
5.16	Comparación de las predicciones VAR y UNIVARIANTE para $X_{2t}$ (IAE) . . . . .	159
5.17	Respuesta de las variables $X_{1t}$ y $X_{2t}$ ante un choque unitario de $X_{1t}$ . . . . .	161
5.18	Respuesta de las variables $X_{1t}$ y $X_{2t}$ ante un choque unitario de $X_{2t}$ . . . . .	162
6.1	Las variables $X_{1t}$ y $X_{2t}$ no están cointegradas . . . . .	166
6.2	Las variables $X_{1t}$ y $X_{2t}$ están cointegradas . . . . .	166
6.3	Series consideradas en el ejemplo . . . . .	168
7.1	Series consideradas en el ejemplo . . . . .	175
7.2	Prueba de cointegración de Johansen . . . . .	176
7.3	Predicción para $X_{1t}$ . . . . .	182
7.4	Predicción para $X_{2t}$ . . . . .	182
7.5	Predicción para $X_{3t}$ . . . . .	183
7.6	Comparación de las predicciones para $X_{1t}$ entre el MCE y el modelo univariante. . . . .	184



---

7.7	Comparación de las predicciones para $X_{2t}$ entre el MCE y el modelo univariante. . . . .	184
7.8	Comparación de las predicciones para $X_{3t}$ entre el MCE y el modelo univariante. . . . .	185



# Índice de tablas

3.1	Cálculos para el caso condicionado . . . . .	62
3.2	Cálculos para el caso no condicionado . . . . .	64
3.3	Resumen estadístico para el modelo $ARIMA(2, 1, 2)$ de la SIB . . . . .	65
3.4	Resumen estadístico para el modelo $ARIMA(2, 1, 2)$ de la SIB . . . . .	66
3.5	Resumen estadístico para el modelo $ARIMA(2, 1, 1)$ de la SIB . . . . .	67
3.6	Resumen estadístico para el modelo $ARIMA(1, 1, 1)$ de la SIB . . . . .	67
3.7	Resumen estadístico para el modelo $ARIMA(1, 1, 1)$ de la SIB sin constante . . . . .	67
3.8	Información estadística para la prueba ADF de la SIB . . . . .	84
3.9	Información estadística para la prueba ADF de la SIB diferenciada . . . . .	85
3.10	Información estadística del Modelo 1 . . . . .	87
3.11	Información estadística del Modelo 2 . . . . .	88
3.12	Información estadística del Modelo 3 para la SVM . . . . .	89
4.1	Información sobre los coeficientes del Modelo 3 para la SVM . . . . .	98
4.2	Información estadística para el Modelo 3-ARCH(1) para la SVM . . . . .	100
5.1	Resumen de estadísticas y matrices de correlación cruzada para $X_{1t}$ , $X_{2t}$ y $X_{3t}$ . . . . .	107
5.2	Prueba DFA para $Y_{1t}$ . . . . .	123
5.3	Prueba DFA para $Y_{2t}$ . . . . .	123
5.4	Prueba DFA para $Y_{3t}$ . . . . .	124
5.5	Criterios para escoger el retardo del VAR . . . . .	124
5.6	Prueba de autocorrelación Pormanteau . . . . .	127
5.7	Prueba LM . . . . .	128
5.8	Prueba de Normalidad de los residuos . . . . .	128
5.9	Error Medio Cuadrático estimado para los modelos VAR y univariante . . . . .	132
5.10	Prueba de causalidad de Granger . . . . .	136
5.11	Respuesta de las variables $Y_{1t}$ , $Y_{2t}$ y $Y_{3t}$ ante un choque unitario de $Y_{1t}$ . . . . .	138
5.12	Respuesta de las variables $Y_{1t}$ , $Y_{2t}$ y $Y_{3t}$ ante un choque unitario de $Y_{2t}$ . . . . .	140
5.13	Respuesta de las variables $Y_{1t}$ , $Y_{2t}$ y $Y_{3t}$ ante un choque unitario de $Y_{3t}$ . . . . .	142
5.14	Descomposición hipotética de choques . . . . .	146
5.15	Otra descomposición hipotética de choques . . . . .	147
5.16	Descomposición de la Varianza (Orden de Cholesky $Y_3$ $Y_2$ $Y_1$ ) . . . . .	148
5.17	Resumen de estadísticas y matrices de correlación cruzada para $X_{1t}$ y $X_{2t}$ . . . . .	150
5.18	Prueba DFA para $X_{1t}$ . . . . .	151
5.19	Prueba DFA para $X_{2t}$ . . . . .	151
5.20	Criterios para escoger el retardo del VAR . . . . .	152

5.21	Prueba de autocorrelación Pormanteau . . . . .	154
5.22	Prueba LM . . . . .	155
5.23	Error Medio Cuadrático estimado para los modelos VAR y UNIVARIANTE .	159
5.24	Prueba de causalidad de Granger . . . . .	160
5.25	Respuesta de las variables $X_{1t}$ y $X_{2t}$ ante un choque unitario de $X_{1t}$ . . . . .	160
5.26	Respuesta de las variables $X_{1t}$ y $X_{2t}$ ante un choque unitario de $X_{2t}$ . . . . .	161
5.27	Descomposición de la Varianza (Orden de Cholesky $X_1 X_2$ ) . . . . .	163
6.1	Prueba de raíces unitarias para las $X_{1t}$ . . . . .	169
6.2	Prueba de raíces unitarias para las $X_{2t}$ . . . . .	169
6.3	Prueba de raíces unitarias para las $X_{3t}$ . . . . .	169
6.4	Prueba de raíces unitarias para $X_{1t}$ en primera diferencia . . . . .	169
6.5	Prueba de raíces unitarias para $X_{2t}$ en primera diferencia . . . . .	169
6.6	Prueba de raíces unitarias para $X_{3t}$ en primera diferencia . . . . .	170
7.1	Elección del retardo del VAR . . . . .	175
7.2	Resumen de la prueba de cointeración de Johansen para VAR(5), con solo el retardo 5 . . . . .	177
7.3	Resumen de la prueba de cointeración de Johansen para VAR(5), con los retardos desde el 1 al 5 . . . . .	177
7.4	Prueba Portmanteau . . . . .	180
7.5	Prueba de Jarque-Bera . . . . .	181
7.6	Prueba LM . . . . .	181
7.7	Predicciones para las tres series utilizando el MCE . . . . .	181

# Notaciones

- $E(.)$  : Esperanza matemática de una variable aleatoria.
- $V(.)$  : Varianza de una variable aleatoria.
- $\text{Cov}(.,.)$  : Covarianza de dos variables aleatorias.
- $N(\mu, \sigma^2)$  : Distribución normal con media  $\mu$  y varianza  $\sigma^2$ .
- $\hat{\cdot}$  : Denota un estimador del parámetro estadístico que se encuentra bajo el símbolo.
- $c$  : De manera general, representa a una constante numérica.
- i.i.d : Representa a la frase: independientes e idénticamente distribuidas.
- i.i.d (0,1) : Variables aleatorias independientes e idénticamente distribuidas con esperanza matemática 0 y varianza 1.
- $\gamma(l)$  : Función de autocovarianza de orden  $l$ .
- $\rho(l)$  : Función de autocorrelación de orden  $l$ .
- $r(l)$  : Función de autocorrelación parcial de orden  $l$ .
- $1_A$  : Función indicatriz sobre el conjunto  $A$ .
- $L_c^2(\Omega)$  : Espacio de variables aleatorias de  $\Omega$  en  $C$ , de cuadrado integrable.
- $L_c^2(E)$  : Espacio de funciones medibles de  $E$  en  $C$ , de cuadrado integrable.
- $\Sigma$  : Matriz de varianzas - covarianzas
- $\Sigma_u$  : Matriz de varianzas – covarianzas residuales
- $\therefore$  : Significa: por tanto



# Capítulo 1

## Análisis a través de procesos estacionarios

### 1.1 Introducción

Para el análisis de una variedad de fenómenos económicos o físicos se dispone, en general, de una cierta cantidad de observaciones, tomadas en momentos equidistantes. A una serie de observaciones de este tipo se le llama una serie cronológica o temporal. Como ejemplos de tales series se pueden mencionar las siguientes:

- a. La temperatura diaria promedio tomada en un lugar específico.
- b. El volumen de ventas diario de un cierto artículo.
- c. El índice mensual de precios al consumidor (IPC).
- d. La intensidad mensual del tráfico de viajeros por un cierto medio de transporte, en un determinado país.
- e. La posición en el instante  $t$ , de una partícula.
- f. El caudal promedio mensual de un río, en un sitio determinado.
- g. El producto interno bruto de un país (la periodicidad puede ser anual, trimestral o mensual).

Sea  $X_t$  la variable aleatoria (v.a.) que se observa en el instante  $t$ . Aunque son múltiples los problemas que se tratan dentro del contexto de las series temporales, uno de los principales consiste en predecir  $X_{T+h}$  cuando se han observado  $X_1, X_2, \dots, X_T$ . Se denotará a la predicción de  $X_{T+h}$  por  $\hat{X}_T(h)$ .

**Observación 1.1.** El signo “ $\hat{\phantom{x}}$ ” significa que la expresión teórica se reemplaza por su estimación; en todo el documento se utilizará esta notación.

Si el gráfico correspondiente a los puntos  $(t, X_t)$ ,  $t = 1, \dots, T$  es bastante regular se puede, en la mayoría de los casos, modelar el fenómeno por una curva simple (de tipo polinomial

o sinusoidal; ver figura 1.1). En caso contrario, es necesario utilizar modelos probabilísticos un poco más complejos (ver figuras 1.2 y 1.3).

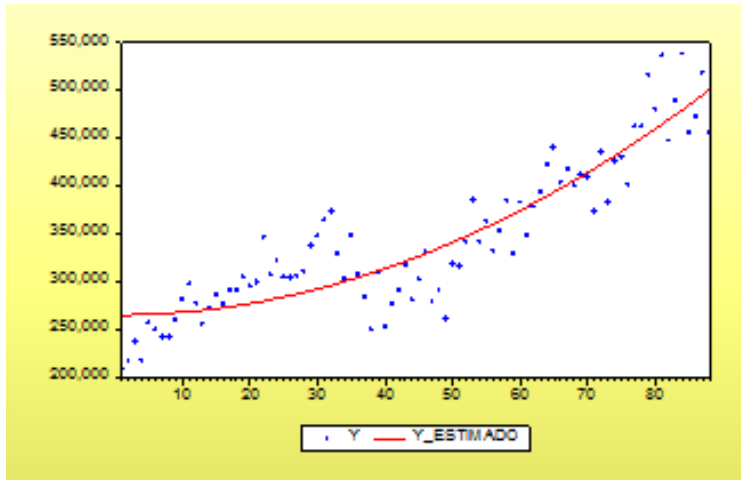


Figura 1.1: Ajuste adecuado de datos de tendencia cuadrática por una curva polinomial

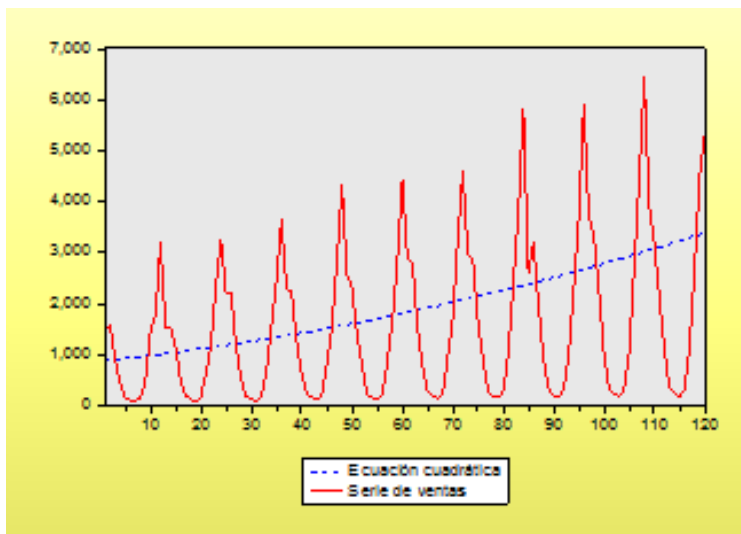
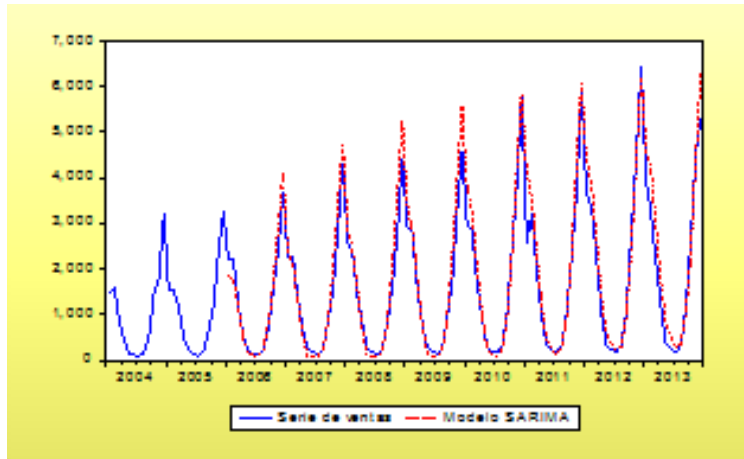


Figura 1.2: Ajuste no adecuado de una serie de ventas a través de una curva cuadrática





**Figura 1.3:** Modelación de una serie de ventas a través de un modelo SARIMA (se estudiará más adelante)

## 1.2 Procesos Estacionarios

Por simplicidad, se asumirá que un proceso estocástico es una familia de variables aleatorias, denotada por  $(X_t)_{t \in T}$  o  $(X_t, t \in T)$ , definidas sobre un espacio muestral  $\Omega$  y que toman valores en un conjunto  $E$  (regularmente  $\mathbb{R}$ , aunque también puede ser  $\mathbb{C}$ ,  $\mathbb{R}^k$  o  $\mathbb{C}^k$ ).  $T$  se dice el espacio de tiempos; por lo general es  $\mathbb{R}$  o un subconjunto de este, como  $\mathbb{N}$  o  $\mathbb{Z}$  (aquí se asumirá frecuentemente que  $T = \mathbb{Z}$ : el conjunto de números enteros).

**Definición 1.1.**  $(X_t, t \in T)$  es estrictamente (o fuertemente) estacionario si:

$$\text{Ley}(X_{t_1+l}, \dots, X_{t_k+l}) = \text{Ley}(X_{t_1}, \dots, X_{t_k}),$$

para todo  $k = 1, 2, \dots$  y  $t_1, \dots, t_k, t_1 + l, \dots, t_k + l \in T$ .

En este documento se utilizará la palabra Ley o Distribución (Dist.) indistintamente.

**Definición 1.2.**  $(X_t, t \in T)$  es débilmente estacionario si:

1.  $(X_t, t \in T)$  es real (toma valores en  $\mathbb{R}$ ) y de segundo orden ( $\forall t \in T, E(X_t^2) < \infty$ ).
2.  $E(X_t) = m$  ( $m$  es constante independiente de  $t$ ).
3.  $\text{Cov}(X_{s+l}, X_{t+l}) = \text{Cov}(X_s, X_t)$  para todo  $s, t, l, s+l, t+l \in T$ .

**Observación 1.2.** En adelante, un proceso débilmente estacionario se dirá estacionario.

**Observación 1.3.** 1. La parte 3 de la definición 1.2, tomando  $s = t$ , permite deducir que la varianza del proceso es constante.

2. Segundo orden y fuertemente estacionario implica débilmente estacionario. Lo recíproco no es cierto.

3. La parte 3 de la definición 1.2 es equivalente a

$$\text{Cov}(X_s, X_t) = \gamma(s - t) \text{ para todo } s, t, s - t \in T,$$

y equivalente a

$$\text{Cov}(X_t, X_{t+l}) = E(X_0, X_t) = \gamma(l) \text{ para todo } l, t, t + l \in T$$

4.  $\gamma(l)$  se llama la *función de autocovarianza* del proceso; ésta solo depende de la diferencia entre los índices  $t + l$  y  $t$ . En particular  $V(X_t) = \gamma(0)$ ; es decir, la varianza de un proceso estacionario es constante.

5. Si  $T = \mathbb{Z}$

- $\gamma(h)$  es par:  $\gamma(h) = \gamma(-h)$ ,  $\forall h \in \mathbb{Z}$
- $|\gamma(h)| \leq \gamma(0)$ ,  $\forall h \in \mathbb{Z}$
- $\gamma(h)$  es de tipo positivo:

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma(t_j - t_k) \geq 0, \forall n \in \mathbb{N}, \forall a_j \in \mathbb{R}, \forall t_j \in \mathbb{Z}$$

Se puede demostrar que esta condición caracteriza a una función de autocovarianza.

*Demostración.* Ejercicio (1.1). □

**Ejemplo 1.1.** Para  $T = \mathbb{Z}$ .

1. Sea  $X$  v.a. tal que  $E(X) = 0$ ,  $E(X^2) = 1$  y  $\text{Ley}(X)$  simétrica. El proceso definido por:

$$X_t = (-1)^t X, t \in \mathbb{Z},$$

es débilmente estacionario (d.e.). En efecto:

- (a)  $E(X_t^2) < \infty$ , se puede ver inmediatamente.
  - (b)  $E(X_t) = (-1)^t E(X) = (-1)^t \cdot 0 = 0$ .
  - (c)  $\text{Cov}(X_{s+l}, X_{t+l}) = (-1)^{s+t+2l} E(X^2) = (-1)^{s+t} E(X^2) = \text{Cov}(X_s, X_t)$ .
2. El proceso  $(X_t, t \in \mathbb{Z})$ , donde las variables aleatorias son i.i.d. (independientes e idénticamente distribuidas), es estrictamente estacionario.
3. El proceso  $(X_t)$  de v.a. reales de cuadrado integrable ( $E(X_t^2) < \infty$ ,  $\forall t$ ) tal que (t.q)

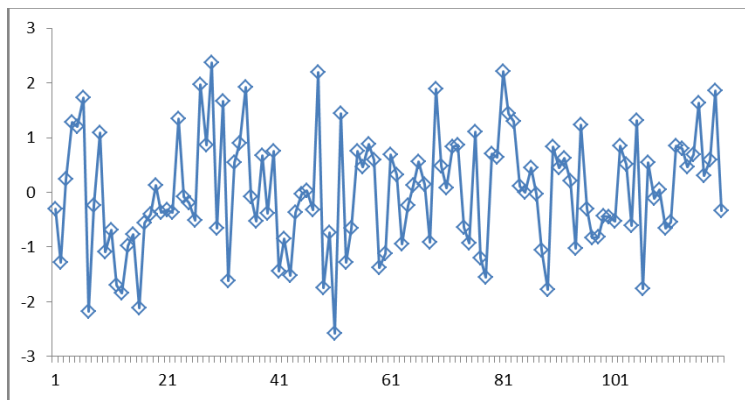
$$\begin{cases} E(X_t) \text{ es constante} \\ V(X_t) \text{ es constante} \\ s \neq t, X_t \perp X_s \text{ (Cov}(X_t, X_s) = 0) \end{cases}$$

es débilmente estacionario.

**4. Ruido Blanco**

- Sea  $(X_t, t \in \mathbb{Z})$  con v.a. independientes e idénticamente distribuidas, reales, de segundo orden, t.q.  $E(X_t) = 0$  y  $V(X_t) = \sigma^2 > 0, \forall t \in \mathbb{Z}$ .

De acuerdo al ejemplo (2),  $(X_t, t \in \mathbb{Z})$  es un proceso estrictamente estacionario, llamado ruido blanco fuerte. En la figura 1.4 se muestra una observación de un ruido blanco, simulado con observaciones de una distribución normal estándar.



**Figura 1.4:** Ruido blanco fuerte simulado con distribución  $N(0, 1)$

**Observación 1.4.** Nótese que en estricto rigor matemático el gráfico solo corresponderá a puntos, pues se definen los valores de las variables aleatorias en los puntos de  $\mathbb{Z}$  (conjunto discreto); la aparente continuidad de la curva constituye solamente un artificio para una mejor visualización, que se logra uniendo los puntos por segmentos de recta; algo similar a lo que sucede en los gráficos de los histogramas. Esto se considerará sobre entendido en gráficos posteriores.

- Sea  $(X_t, t \in \mathbb{Z})$  un proceso aleatorio real de segundo orden, t.q.  $E(X_t) = 0$  y  $V(X_t) = \sigma^2 > 0, X_t \perp X_s$  con  $s \neq t, \forall t, s \in \mathbb{Z}$ . Por el literal (3) del ejemplo este proceso es débilmente estacionario y se llama ruido blanco débil.

5. Sea  $(X_t, t \in \mathbb{Z})$ , tal que:

$$X_t = \sum_{j=1}^q (A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t))$$

donde,

- $\lambda_1, \dots, \lambda_q \in \mathbb{R}$ .
- $A_1, \dots, A_q, B_1, \dots, B_q$  son v.a. ortogonales dos a dos, centradas, tal que:

$$E(A_j^2) = E(B_j^2) = \sigma_j^2$$

Entonces,  $(X_t, t \in \mathbb{Z})$  es débilmente estacionario. Se puede verificar que:

$$\begin{cases} E(X_t) = 0 \\ \text{Cov}(X_s, X_t) = \sum_{j=1}^q \sigma_j^2 \cos(\lambda_j(t-s)) \end{cases}$$

En este y otros ejemplos, no se hace mención al hecho de que el proceso debe ser de segundo orden, porque es una consecuencia de la expresión de la covarianza, tomando  $s = t$ .

6. Sea  $(u_t, t \in \mathbb{Z})$  un ruido blanco débil de varianza  $\sigma^2$ .

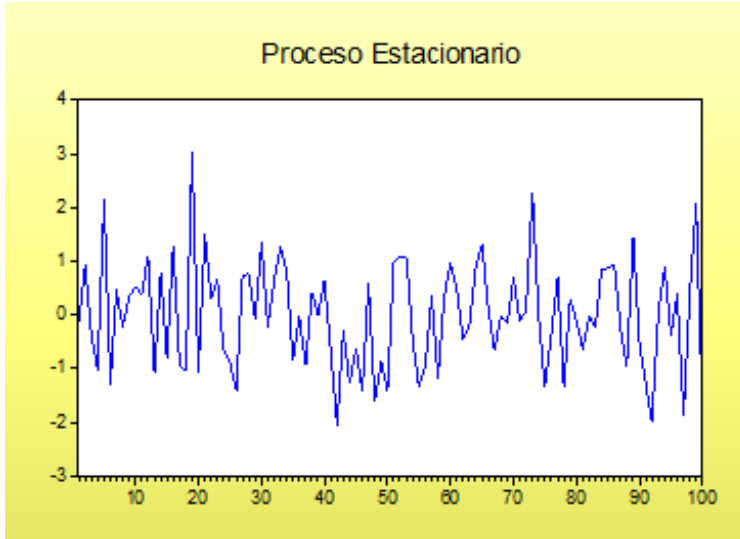
$$v_t = \sum_{j=0}^q \psi_j u_{t-j} \quad t \in \mathbb{Z} \quad \psi_0 = 1, \quad \psi_q \neq 0, \quad \psi_j \in \mathbb{R}$$

$(v_t)_{t \in \mathbb{Z}}$  es un proceso d.e. llamado *media móvil* de orden  $q$  ( $MA(q)$ , por su significado en inglés: Moving Average).

$$\begin{aligned} E(v_t) &= 0 \\ \text{Cov}(v_t, v_{t+k}) &= \text{Cov}\left(\sum_{i=0}^q \psi_i u_{t-i}, \sum_{j=0}^q \psi_j u_{t+k-j}\right) = E(u_{t-i}^2) \\ \sum_{i=1}^{q-k} \psi_i \psi_{i+k} &= \sigma^2 \sum_{i=1}^{q-k} \psi_i \psi_{i+k} = \gamma(k). \end{aligned}$$

**Observación 1.5.** Se puede demostrar que con  $(X_t)$  definida en el ejemplo 1.1 literal (5), se puede aproximar cualquier proceso débilmente estacionario.

Aunque teóricamente es algo complicado demostrar que un proceso estocástico es débilmente estacionario, gráficamente se puede intuir; su gráfico con respecto al tiempo aparecería con una forma similar a la siguiente:



**Figura 1.5:** Proceso estacionario

1. Los datos oscilan alrededor de una constante (cero en este caso), lo que permite suponer que la media es constante.
2. Los datos tienen una variabilidad algo homogénea, por lo que se puede considerar

que la varianza es constante (ésta solo es una condición necesaria, pero no suficiente para la propiedad (3) de la definición 1.2; sin embargo, para propósitos de aplicación es bastante útil).

3. También se puede decir que  $E(X_t^2) = V(X_t) + m^2 < \infty$ , ya que la varianza no crece indiscriminadamente.

A continuación, se presentan gráficamente algunos casos en los que no se cumplen los supuestos de estacionariedad:

- Proceso no estacionario: Media no constante.

En el caso que el proceso no tenga la media constante, se dice que presenta una *tendencia*.

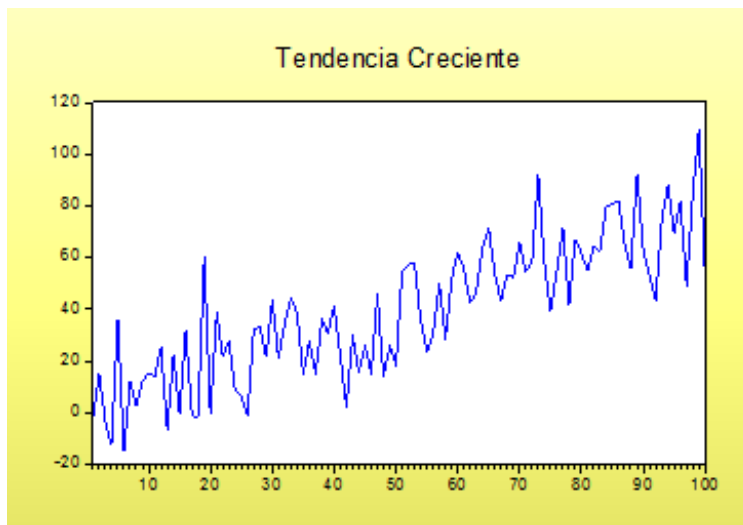
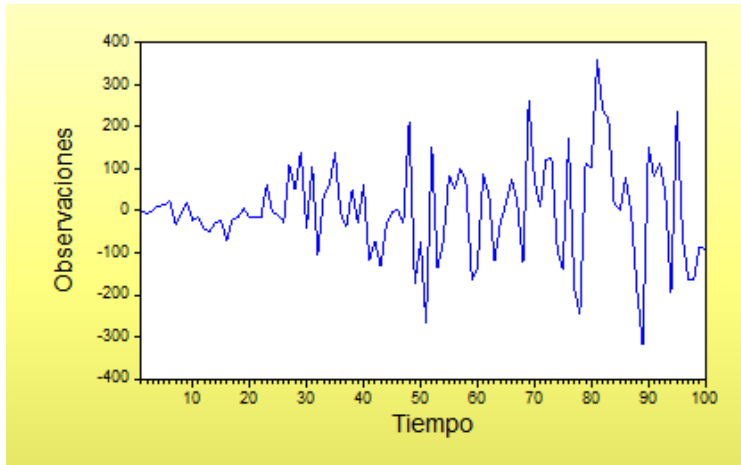


Figura 1.6: Proceso no estacionario: Tendencia creciente

- Proceso no estacionario: Varianza no homogénea.

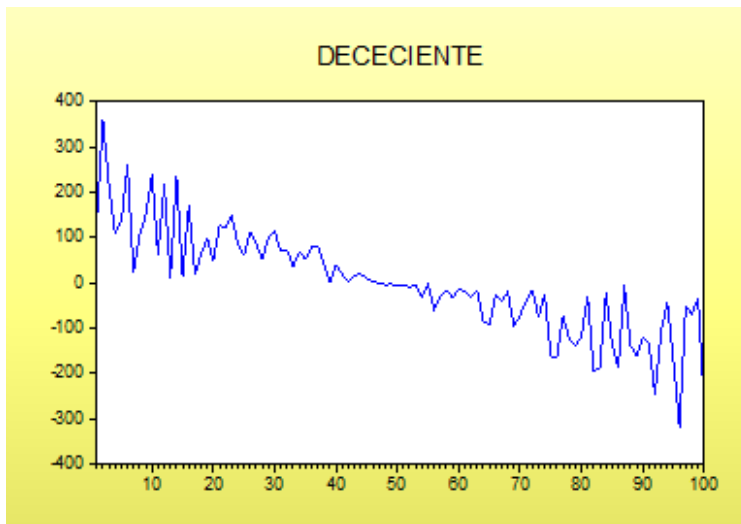
En este caso, gráficamente, se ve como si el proceso “explotara”; es decir, la oscilación de los datos se va haciendo más grande con el paso del tiempo.



**Figura 1.7:** Proceso no estacionario: Varianza no homogénea

- Proceso no estacionario: Tendencia no constante y varianza no homogénea.

Este caso es una combinación de los dos anteriores; es decir, se presenta una tendencia y la varianza del proceso no es homogénea.



**Figura 1.8:** Proceso no estacionario: Tendencia decreciente y varianza no homogénea

**Ejemplo 1.2.** Mostrar que las siguientes funciones son de autocovarianza (determinar los procesos asociados).

1.  $\gamma(h) = (-1)^{|h|}$ .

Sea el proceso  $X_t = (-1)^t X$  con  $E(X) = 0$  y  $V(X) = 1$

(a) Por demostrar que  $E(X_t) = \text{constante } \forall t$

$$E(X_t) = E[(-1)^t X] = (-1)^t E(X) = 0, \quad \forall t$$

(b) Por demostrar que  $E(X_t^2) < \infty$

$$E(X_t^2) = E\left[(-1)^{2t} X^2\right] = 1 < \infty, \quad \forall t$$

(c) Por demostrar que  $\text{Cov}(X_t, X_{t+h}) = \gamma(h) = (-1)^{|h|}$

• Sea  $h \geq 0$ :

$$\text{Cov}(X_t, X_{t+h}) = \text{Cov}\left((-1)^t X, (-1)^{t+h} X\right) = (-1)^h \text{Cov}(X, X)$$

$$\text{Cov}(X_t, X_{t+h}) = (-1)^h V(X) = (-1)^h = (-1)^{|h|}.$$

• Si  $h < 0$ :

$$\text{Cov}(X_t, X_{t+h}) = (-1)^h = (-1)^{h-2h} = (-1)^{-h} = (-1)^{|h|}$$

$$2. \gamma(h) = 1 + \cos\left(\frac{\pi}{2}\right) + \cos\left(\frac{\pi h}{4}\right).$$

$$X_t = A \cos \frac{\pi t}{2} + B \sin \frac{\pi t}{2} + D \cos \frac{\pi t}{4} + E \sin \frac{\pi t}{4} + F$$

donde  $A, B, D, E, F$  v.a. independientes y tales que  $E(A) = E(B) = E(D) = E(E) = E(F) = 0$ ,  $E(A^2) = E(B^2) = E(D^2) = E(E^2) = E(F^2) = 1$  y no correlacionadas dos a dos:

$$\text{Cov}(X_t, X_{t+h}) =$$

$$\begin{aligned} & \text{Cov}\left(A \cos \frac{\pi t}{2} + B \sin \frac{\pi t}{2} + D \cos \frac{\pi t}{4} + E \sin \frac{\pi t}{4} + F, \right. \\ & \left. A \cos \frac{\pi(t+h)}{2} + B \sin \frac{\pi(t+h)}{2} + D \cos \frac{\pi(t+h)}{4} + E \sin \frac{\pi(t+h)}{4} + F\right). \end{aligned}$$

(a) Por demostrar que  $E(X_t) = \text{constante}$

$$\begin{aligned} E(X_t) &= E\left(\cos \frac{\pi t}{2} + B \sin \frac{\pi t}{2} + D \cos \frac{\pi t}{4} + E \sin \frac{\pi t}{4}\right) \\ &= E(A) \cos \frac{\pi t}{2} + E(B) \sin \frac{\pi t}{2} + E(D) \cos \frac{\pi t}{4} + E(E) \sin \frac{\pi t}{4} + E(F) \\ &= 0, \quad \forall t. \end{aligned}$$

(b) Por demostrar que  $V(X_t) = E(X_t^2) < \infty$

$$\begin{aligned} E(X_t^2) &= E(A^2) \left(\cos \frac{\pi t}{2}\right)^2 + E(B^2) \left(\sin \frac{\pi t}{2}\right)^2 + E(D^2) \left(\cos \frac{\pi t}{4}\right)^2 \\ & \quad + E(E^2) \left(\sin \frac{\pi t}{4}\right)^2 + E(F^2) \\ &= \cos^2 \frac{\pi t}{2} + \sin^2 \frac{\pi t}{2} + \cos^2 \frac{\pi t}{4} + \sin^2 \frac{\pi t}{4} + 1 \\ &= 5 < \infty, \quad \forall t \end{aligned}$$

(c) Por demostrar que  $\text{Cov}(X_t, X_{t+h}) = \gamma(h) = 1 + \cos \frac{\pi h}{2} + \cos \frac{\pi h}{4}$

$$\begin{aligned}
 \text{Cov}(X_t, X_{t+h}) &= \\
 &= E(X_t, X_{t+h}) \\
 &= E\left(A^2 \cos \frac{\pi t}{2} \cos \frac{\pi(t+h)}{2} + A \cos \frac{\pi t}{2} B \sin \frac{\pi(t+h)}{2} + A \cos \frac{\pi t}{2} D \cos \frac{\pi(t+h)}{4} + \right. \\
 &\quad \left. A \cos \frac{\pi t}{2} E \sin \frac{\pi(t+h)}{4} + AF \cos \frac{\pi t}{2} + B \sin \frac{\pi t}{2} A \cos \frac{\pi(t+h)}{2} + \right. \\
 &\quad \left. B^2 \sin \frac{\pi t}{2} \sin \frac{\pi(t+h)}{2} + B \sin \frac{\pi t}{2} D \cos \frac{\pi(t+h)}{4} + \dots + F^2\right) \\
 &= E(A^2) \cos \frac{\pi t}{2} \cos \frac{\pi(t+h)}{2} + E(B^2) \sin \frac{\pi t}{2} \sin \frac{\pi(t+h)}{2} + \\
 &\quad E(D^2) \cos \frac{\pi t}{4} \cos \frac{\pi(t+h)}{4} + E(E^2) \sin \frac{\pi t}{4} \sin \frac{\pi(t+h)}{4} + E(F^2) \\
 &= \cos \frac{\pi h}{2} + \cos \frac{\pi h}{4} + 1.
 \end{aligned}$$

## 1.3 Representación Espectral de un Proceso Débilmente Estacionario

### 1.3.1 Representación espectral de $(\gamma_t)_{t \in \mathbb{Z}}$

Sea  $(X_t, t \in \mathbb{Z})$  d.e. y centrado. Recuerdese que se definió la función de autocovarianza  $(\gamma_t, t \in \mathbb{Z})$  por  $\gamma_t = E(X_0, X_t) = E(X_l, X_{l+t}), \forall t, l \in \mathbb{Z}$ .

**Teorema 1.1.** *Existe una medida única  $\mu$  acotada y simétrica sobre  $[\pi, \pi]$  tal que:*

$$\gamma_t = \int_{-\pi}^{\pi} \cos(\lambda t) d\mu(\lambda), \quad t \in \mathbb{Z}$$

$\mu$  se dice la medida espectral de  $(X_t)$ . La densidad de la parte absolutamente continua de  $\mu$  se dice densidad espectral de  $(X_t)$ ; la función de distribución de  $\mu$  se llama función de distribución espectral de  $(X_t)$ .

*Demostración.* Aquí se abordará solamente un caso particular; la demostración completa se encuentra en el Anexo A.2.

$$\sum_t |\gamma_t| < \infty.$$

Se puede definir:

$$f(\lambda) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma_t \cos(\lambda t), \quad \lambda \in [-\pi, \pi]$$

( $f$  está bien definida). Entonces:

$$\gamma_t = \int_{-\pi}^{\pi} \cos(\lambda t) f(\lambda) d\lambda$$

si se llega a demostrar que  $f \geq 0$ , entonces  $f(\lambda)d\lambda$  puede considerarse como  $d\mu(\lambda)$ . □



**Observación 1.6.** También  $\gamma_t$  se puede expresar por:

$$\gamma_t = \int_{-\pi}^{\pi} e^{i\lambda t} d\mu(\lambda).$$

**Ejemplo 1.3.** 1. Sea  $(X_t)_{t \in \mathbb{Z}}$  ruido blanco (débil) tal que  $\gamma_t = 0, |t| \geq 1$  y  $\gamma_0 = \sigma^2$

$$\sum |\gamma_t| < \infty \implies f(\lambda) = \frac{1}{2\pi} \sum_{t \in \mathbb{Z}} \gamma_t \cos(\lambda t) = \frac{\sigma^2}{2\pi}, \quad \lambda \in [-\pi, \pi].$$

Esta es la densidad con respecto a la medida de Lebesgue sobre  $[\pi, \pi]$ .

2. Si  $X_t = X \forall t \in \mathbb{Z}$ , con  $E(X) = 0$  y  $E(X^2) = \sigma^2$ ,  $\gamma_t = E(X^2) = \sigma^2 \Rightarrow \sum |\gamma_t| = \infty$ .

En este caso,  $\mu = \sigma^2 \delta_{(0)}$ , pues:

$$\int \cos(\lambda) d\mu(\lambda) = \sigma^2 \cos(0) = \sigma^2 = \gamma_t$$

donde  $\delta_{(a)}$  representa a la medida de Dirac en el punto  $a$  (la distribución que corresponde a la variable aleatoria constante  $X = a$ ).

3. Para el proceso:

$$X_t = \sum_{j=1}^q (A_j \cos(\lambda_j t) + B_j \sin(\lambda_j t)), \quad \forall t \in \mathbb{Z}$$

se tiene:

$$\mu = \sum_{j=1}^q \frac{\sigma_j^2}{2} [\delta_{(-\lambda_j)} + \delta_{(\lambda_j)}].$$

### 1.3.2 Representación espectral de $(X_t)$

El objetivo es representar  $(X_t)$  de la manera siguiente:

$$X_t = \int_{-\pi}^{\pi} \cos(\lambda t) dW(\lambda)$$

donde  $W$  es una medida aleatoria o estocástica.

#### 1.3.2.1 Medida estocástica (medida de Wiener)

Sean  $(\Omega, A, P)$  un espacio de probabilidad;  $(E, \beta, \mu)$  espacio medido, donde  $\mu$  es  $\sigma$ -finita,  $\beta_\mu$  anillo de conjuntos de medida finita tal que la  $\sigma$ -álgebra generada por  $\beta_\mu$ ,  $\sigma(\beta_\mu) = \beta$ . Se considera la aplicación:

$$W : \beta_\mu \rightarrow L_c^2(\Omega, A, P)$$

$L_c^2(\Omega)$  es la notación simplificada de  $L_c^2(\Omega, A, P)$ , representa al espacio de las variables aleatorias de cuadrado integrable que toman valores sobre el cuerpo de los números complejos; este es un espacio de Hilbert sobre  $\mathbb{C}$ , con producto escalar  $\langle X, Y \rangle = E(X\bar{Y})$ . La raya sobre  $Y$  representa el conjugado de  $Y$ .

**Definición 1.3.**  $W$  es una medida estocástica basada sobre  $(\Omega, A, P)$  y de medida espectral  $\mu$  si:

- $\forall A, B \in \beta_\mu$ , con  $A \cap B = \emptyset$ , se tiene:
  1.  $W(A \cup B) = W(A) + W(B)$ ,
  2.  $E[W(A)\overline{W(B)}] = 0$ .
- $\forall A \in \beta_\mu$ ,  $E|W(A)|^2 = \mu(A)$ .

### Propiedades de $W$ :

1.  $\sigma$ -aditividad en media cuadrática (m.c.).

Sea  $(A_n)$  una sucesión creciente de elementos de  $\beta_\mu$  que tiene por límite  $A \in \beta_\mu$ , entonces

$$W(A_n) \xrightarrow{m.c.} W(A)$$

Esto se verifica puesto que:

$$A = A_n \cup (A - A_n) \implies W(A) = W(A_n) + W(A - A_n).$$

Por tanto:

$$E|W(A) - W(A_n)|^2 = E|W(A - A_n)|^2 = \mu(A - A_n) \rightarrow 0,$$

pues  $\mu$  es  $\sigma$ -aditiva.

2.  $\forall A, B \in \beta_\mu$ ,  $E[W(A)\overline{W(B)}] = \mu(A \cap B)$ .

La operación  $A \cap B$  también se denotará por  $AB$ . Puesto que:  $A = AB + AB^c$  y  $B = BA + BA^c$ , se tiene que:

$$\begin{aligned} E[W(A)\overline{W(B)}] &= E[(W(AB) + W(AB^c))\overline{W(BA) + W(BA^c)}] \\ &= E[W(AB)\overline{W(BA)} + W(AB)\overline{W(BA^c)} + W(AB^c)\overline{W(BA)} + \\ &\quad W(AB^c)\overline{W(BA^c)}] \\ &= E[W(AB)\overline{W(AB)}] \\ &= E|W(AB)|^2 = \mu(AB). \end{aligned}$$

**Ejemplo 1.4.** Sea  $(Z_t, t \in \mathbb{R}^+)$  de segundo orden, centrado, con incrementos estacionarios y ortogonales, continuo a la derecha en m.c.

Se inicia recordando las definiciones de incrementos ortogonales, incrementos estacionarios y continuidad a la derecha:

- Incrementos ortogonales:

$\forall t_1 < t_2 < \dots < t_k$  en  $\mathbb{R}^+$ , las v.a.  $Z_{t_1}Z_{t_2} - Z_{t_1}, \dots, Z_{t_k} - Z_{t_{k-1}}$  son ortogonales 2 a 2.

- Incrementos estacionarios:

$$\forall s, t, s + l, t + l \in \mathbb{R}^+, E(Z_{t+l} - Z_{s+l})^2 = E(Z_t - Z_s)^2.$$

- Continuo a la derecha:

$$\text{Si } t_n \rightarrow t^+, \text{ entonces } Z_{t_n} \rightarrow Z_t, \text{ con } t_n \in \mathbb{R}^+.$$

Se define:  $Z_0 = 0$ . Supóngase que  $E(Z_1^2) = 1$ , entonces se demostrará que:  $E(Z_t^2) = t$   $\forall t \in \mathbb{R}^+$ .

*Demostración.* Para todo  $s, t \in \mathbb{R}^+$ , se tiene que:  $E(Z_t^2) = E(Z_t - Z_s)^2 + E(Z_s^2)$ , para  $t > s$ . Sea  $f(t) = E(Z_t^2) \implies f(t) \geq f(s)$  para  $t > s$ .

- Ahora, se va a demostrar que:  $f(n) = n \forall n \in \mathbb{N}$

Considere el caso en que  $n = 2$ , se tiene:

$$E(Z_2^2) = E(Z_2 - Z_1)^2 + E(Z_1^2) = (Z_1 - Z_0)^2 + E(Z_1^2) = E(Z_1^2) + E(Z_1^2) = 2$$

Luego, realizando el mismo análisis  $n$  veces, se tiene:

$$E(Z_n^2) = E(Z_n - Z_{n-1})^2 + E(Z_{n-1}^2) = (Z_1 - Z_0)^2 + (n-1) = E(Z_1^2) + (n-1) = n$$

- Considere  $p$  y  $q$  enteros positivos; entonces:

$$p = f(p) = f\left(q \frac{p}{q}\right) = qf\left(\frac{p}{q}\right)$$

En efecto:

$$\begin{aligned} f(p) &= E\left(Z_{q \frac{p}{q}}^2\right) \\ &= E\left(Z_{q \frac{p}{q}}^2 - Z_{(q-1) \frac{p}{q}}^2\right)^2 + E\left(Z_{(q-1) \frac{p}{q}}^2\right) \\ &= E\left(Z_{\frac{p}{q}}^2\right) + E\left(Z_{(q-1) \frac{p}{q}}^2\right) \\ &= E\left(Z_{\frac{p}{q}}^2\right) + E\left(Z_{\frac{p}{q}}^2\right) + E\left(Z_{(q-2) \frac{p}{q}}^2\right) \\ &\vdots \\ &= qE\left(Z_{\frac{p}{q}}^2\right) = qf\left(\frac{p}{q}\right). \end{aligned}$$

Así se tiene:

$$p = qf\left(\frac{p}{q}\right)$$

y por tanto,

$$f(r) = r, \quad \forall r \in \mathbb{Q}^+.$$

- Si  $t$  es un número real positivo, existe una sucesión de racionales  $\{t_n\}$  tal que  $t_n \rightarrow t$ . Por tanto:

$$t_n = f(t_n) = E(Z_{t_n}^2) \rightarrow Z_t^2 = f(t)$$

Entonces,

$$f(t) = t, \quad \forall t \in \mathbb{R}^+.$$

- Si se define  $W([0, t]) = Z_t$ , entonces  $W([s, t]) = Z_t - Z_s$ .

□

### 1.3.2.2 Integral Estocástica

Sea

$$\varphi = \sum_{j \in J} \alpha_j 1_{A_j}$$

Donde,

- $J$  es finito.
- $(A_j)$  es una partición de  $E$  en  $\beta_\mu$ .
- $\alpha_j \in \mathbb{C}$ .

Sea  $\varepsilon_\mu$  el espacio de las funciones en escalera, cuyo elemento genérico es  $\varphi$ . Se define:

$$\int \varphi dW =: \sum_{j \in J} \alpha_j W(A_j) \in L_c^2(\Omega, A, P).$$

Entonces,

$$E \left( \left| \int \varphi dW \right|^2 \right) = \sum |\alpha_j|^2 E |W(A_j)|^2 = \sum |\alpha_j|^2 \mu(A_j) = \int |\varphi|^2 d\mu.$$

Por tanto, la aplicación:  $\varphi \rightarrow \int \varphi dW$  conserva la norma (además es lineal), entonces la aplicación correspondiente:  $\varepsilon_\mu \rightarrow L_c^2(\Omega, A, P)$  es una isometría. Pero  $\varepsilon_\mu$  es denso en  $L_c^2(E, \beta, \mu)$ ; por tanto, se puede prolongar la isometría de manera única sobre todo el espacio:

$$\begin{aligned} W: L_c^2(E, \beta, \mu) &\longrightarrow L_c^2(\Omega, A, P) \\ \varphi &\longmapsto \int \varphi dW =: \int_E \varphi(\lambda) dW(\lambda) \end{aligned}$$

#### Propiedades de la Integral estocástica:

1. Si  $\forall A \in \beta_\mu$ ,  $W(A)$  centrada; entonces  $\mathcal{I} = \{\varphi : E(\int \varphi dW) = 0\}$  es cerrado:

Sea  $\{\varphi_n\}$  una sucesión de  $\mathcal{I}$  tal que  $\varphi_n \xrightarrow{m.c.} \varphi \implies \int \varphi_n dW \xrightarrow{m.c.} \int \varphi dW$  (pues la integral estocástica es una isometría y entonces es continua).

Entonces existe convergencia en media cuadrática. Puesto que las  $\varphi_n \in \mathcal{I}$ , se puede concluir que  $\varphi \in \mathcal{I}$ .

2.  $E \left( \int \varphi dW \int \bar{\psi} dW \right) = \int \varphi \bar{\psi} dW$ , pues una isometría conserva el producto escalar.

1.3.2.3 Representación espectral de  $(X_t)$ 

**Teorema 1.2.** Sea  $(X_t, t \in T)$  de segundo orden y centrado (a valores complejos). Se tiene también:

$$\text{Cov}(X_s, X_t) = \int_E f(s, \lambda) \overline{f(t, \lambda)} d\mu(\lambda)$$

donde,  $\{f(s, \cdot), s \in T\}$  es una familia libre y total en  $L_c^2(E, \beta, \mu)$ ; entonces, existe  $W$  medida estocástica de medida espectral  $\mu$  t.q.:

$$X_t = \int_E f(t, \lambda) dW(\lambda), \quad t \in T$$

Se dice que una familia es total si:  $\forall f \in L_c^2(E, \beta, \mu), \exists \{f_n\}$  t.q.  $f_n \rightarrow f$ , donde  $f_n$  es combinación lineal de elementos de la familia.

*Demostración.* Se define:

$$W(f_t) = X_t, \quad y \quad W\left(\sum_{j \in J} \alpha_j f_{t_j}\right) =: \sum \alpha_j W(f_{t_j})$$

donde,  $t \in T$  y  $J$  finito.

La función está bien definida, pues la familia de las  $f_s = f(s, \cdot)$  Es libre.

Ahora, se define:

$$v = \sigma\left(\sum_{j \in J} \alpha_j f_{t_j}\right), \quad J \text{ finito}$$

como el subespacio generado por las combinaciones lineales de elementos  $f_{t_j}$ .

Se va a verificar que  $W$  es una isometría:

$$\begin{aligned} E \left| W\left(\sum_{j \in J} \alpha_j f_{t_j}\right) \right|^2 &= \left| \sum \alpha_j X_{t_j} \right|^2 \\ &= \sum_{j, j'} \alpha_j \bar{\alpha}_{j'} E(X_{t_j}, \bar{X}_{t_{j'}}) \\ &= \sum_{j, j'} \alpha_j \bar{\alpha}_{j'} \int f(t_j, \lambda) \overline{f(t_{j'}, \lambda)} d\mu(\lambda) \\ &= \int \left| \sum_j \alpha_j f(t_j, \lambda) \right|^2 d\mu(\lambda) \end{aligned}$$

entonces  $W$  es una isometría de  $v$  en  $L_c^2(\Omega, A, P)$ ; luego,  $W$  se prolonga de manera única (llamada también  $W$ ):

$$W: L_c^2(E, \beta, \mu) \rightarrow L_c^2(\Omega, A, P)$$

A continuación se va a verificar que la restricción de  $W$  a  $\beta_\mu$  es una medida estocástica.  $1_C$  representa la función indicatriz sobre el conjunto  $C$ .

Sean  $A, B \in \beta_\mu$  t.q.  $A \cap B = \emptyset$ .

- $W(1_{A \cup B}) = W(1_A + 1_B) = W(1_A) + W(1_B)$
- $E(W(1_A) \overline{W(1_B)}) = \int 1_A \overline{1_B} d\mu = \mu(A \cap B) = 0$

- $E |W(1_A)|^2 = \int |1_A|^2 d\mu = \mu(A).$

Se ha demostrado en realidad, que toda isometría de  $L_c^2(E, \beta, \mu) \rightarrow L_c^2(\Omega, A, P)$  es una integral estocástica. Entonces se pueden identificar las isometías con las integrales estocásticas.  $\square$

**Aplicación a los Procesos Estacionarios** Sea  $(X_t)_{t \in \mathbb{Z}}$  d.e. y centrado. Se tiene que:

- $\text{Cov}(X_s, X_t) = \text{Cov}(X_{s-t}, X_0) = \int e^{i(s-t)\lambda} d\mu(\lambda) = \int e^{is\lambda} \overline{e^{it\lambda}} d\mu(\lambda).$
- $\{f(t, \lambda) =: e^{it\lambda}, t \in \mathbb{Z}\}$ , es una familia libre y total en  $L_c^2(E, \beta, \mu)$  con  $E = [-\pi, \pi]$

Por tanto, existe  $W$  medida estocástica sobre  $[-\pi, \pi]$ , asociada a la medida espectral  $\mu$ , t.q.

$$X_t = \int_{[-\pi, \pi]} e^{it\lambda} dW(\lambda).$$

## 1.4 Teoremas Límites

**Teorema 1.3** (Teorema ergódico débil). Sea  $(X_t, t \in \mathbb{Z})$  d.e, centrado, de autovarianza  $(\gamma_t)$ , de medida espectral  $\mu$ , de medida estocástica  $W$ , entonces:

1.  $\bar{X} = \frac{1}{n} \sum_{t=1}^n X_t \xrightarrow{m.c.} W(\{0\})$
2.  $\bar{X} \xrightarrow{m.c.} 0 \iff \frac{1}{n} \sum_{t=0}^{n-1} \gamma_t \rightarrow 0$
- 3.

Un proceso que satisface 2 o 3 se llama un proceso ergódico (débil). Por ejemplo, el proceso definido por  $X_t = X$  no es ergódico, mientras que el proceso definido por  $X_t = (-1)^t X$  si lo es (en ambos casos se considera  $t \in \mathbb{Z}$ ).

Este teorema indica que solamente en los procesos ergódicos la media muestral converge a la media del proceso (extensión de la Ley de los Grandes Números).

*Demostración.* Se tiene que:

$$\bar{X} - W(\{0\}) = \int \left[ \frac{1}{n} \sum_{t=1}^n e^{it\lambda} - 1_{\{0\}}(\lambda) \right] dW(\lambda)$$

Se considera ahora:

$$E |\bar{X} - W(\{0\})|^2 = \int \left| \frac{1}{n} \sum_{t=1}^n e^{it\lambda} - 1_{\{0\}}(\lambda) \right|^2 d\mu(\lambda)$$

pero:

$$\left| \frac{1}{n} \sum_{t=1}^n e^{it\lambda} - 1_{\{0\}}(\lambda) \right|^2 = \begin{cases} \left( \frac{\text{sen}\left(\frac{n\lambda}{2}\right)}{n \text{sen}\left(\frac{\lambda}{2}\right)} \right)^2 & \lambda \neq 0 \\ 0 & \lambda = 0 \end{cases}$$

y por el teorema de la convergencia dominada,  $E|X - W(\{0\})|^2 \rightarrow 0$  por lo tanto,  $\bar{X} \xrightarrow{m.c.} W(\{0\})$

- $\gamma_t = \int_{-\pi}^{\pi} e^{i\lambda t} d\mu(\lambda)$ , entonces:

$$\frac{1}{n} \sum_{t=0}^{n-1} \gamma_t = \int_{-\pi}^{\pi} \left[ \frac{1}{n} \sum_{t=0}^{n-1} e^{i\lambda t} \right] d\mu(\lambda)$$

pero:

$$\frac{1}{n} \sum_{t=0}^{n-1} e^{i\lambda t} \xrightarrow{n \rightarrow \infty} 1_{\{0\}}(\lambda) \quad y \quad \frac{1}{n} \sum_{t=0}^{n-1} e^{i\lambda t} = \frac{1}{n} \frac{1 - e^{i\lambda n}}{1 - e^{i\lambda}}$$

entonces por el teorema de la convergencia dominada:

$$\frac{1}{n} \sum_{t=0}^{n-1} \gamma_t \rightarrow \mu(0)$$

Puesto que  $E|W(\{0\})|^2 = \mu(\{0\})$  se puede concluir. □

**Observación 1.7.**  $\hat{\gamma}_t = \frac{1}{n-t} \sum_{s=1}^{n-t} X_s X_{s+t}$  es un estimador natural de  $\gamma_t$ , ( $t < n, t \in \mathbb{N}$ ).

**Teorema 1.4.** Sea  $(X_t, t \in \mathbb{Z})$  d.e. y centrado; entonces se tiene:

1.  $E(X_s X_{s+t})^2 < \infty, \forall s, t \in \mathbb{Z}$ .
2.  $E(X_{s_1} X_{s_1+t} X_{s_1+s_2} X_{s_1+s_2+t})$  indep. De  $s_1 \forall s_1, s_2, t \in \mathbb{Z}$ .
3.  $\hat{\gamma}_t \xrightarrow{m.c.} \gamma_t \iff \frac{1}{n} \sum E(X_0 X_s X_t X_{s+t}) \rightarrow \gamma_t^2$ .

*Demostración.* Ejercicio (1.3). □

**Teorema 1.5** (Teorema ergódico fuerte). Sea  $(X_t, t \in \mathbb{Z})$  de segundo orden, centrado, estrictamente estacionario; entonces:

$$\bar{X} \xrightarrow{c.s.} 0 \iff \mu(\{0\})$$

(c.s: expresa casi seguramente).

*Demostración.* Admitida □

**Teorema 1.6** (Condición suficiente de convergencia casi segura). Sea  $(X_t, t \in \mathbb{Z})$  de segundo orden, centrado tal que:

$$V(X_t + \dots + X_{t+p-1}) \leq k p^\gamma \quad k = cte; \quad 0 \leq \gamma < 2; \quad p = 1, 2, \dots; \quad \forall t \in \mathbb{Z}$$

Entonces:  $\bar{X} \rightarrow 0$  en media cuadrática y casi seguramente.

*Demostración.* Anexo A.2 □

**Ejemplo 1.5.** 1. Sea  $(X_t)$  con v.a. 2 a 2 ortogonales tales que  $E(X_t) = 0, \sup_t E(X_t^2) \leq k$ . El teorema se aplica con  $\gamma = 1$

2. Si  $X_t = (-1)^t X$ ,  $t \in \mathbb{Z}$  y  $E(X^2) = 1$ ; el teorema se aplica con  $\gamma = 0$ .

### 1.4.1 Teorema Central del Límite

**Teorema 1.7.** Sea  $X_t = \sum_{j=-\infty}^{\infty} a_j u_{t-j}$ ,  $t \in \mathbb{Z}$ , con  $\sum_{j=-\infty}^{\infty} |a_j| < \infty$ ,  $(u_t)$  r.b. fuerte de varianza  $\sigma^2$ .

Entonces:

$$\sqrt{n}\bar{X} \xrightarrow{L} N\left(0, \sigma^2 \left(\sum_{s=-\infty}^{\infty} a_s\right)^2\right).$$

La letra  $L$  indica convergencia en ley o distribución.

*Demostración.* Admitida. □

**Observación 1.8.** Para determinar la velocidad de convergencia:

- Con v.a. independientes se aplica la desigualdad de Berestein:

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \leq \frac{K}{\sqrt{n}}$$

donde  $F_n$  y  $F$  representan a las distribuciones muestral y teórica, respectivamente, de una v. a.  $X$ .

- Con v.a. correlacionadas (entre otras hipótesis  $\text{Cov}(X_t, X_{t+h}) = O(\rho^h)$ ; es decir, la covarianza dividida por  $\rho^h$  converge a una velocidad constante):

$$\sup_{X \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow \frac{(\log n)^2}{\sqrt{n}}.$$

## 1.5 Predicción de un Proceso Estacionario y Descomposición de Wold

### 1.5.1 Predicción de un Proceso Estacionario

Se observa  $(X_s, s \leq t)$  y se busca predecir lo mejor posible a  $g(X_{t+h}) \in L^2$ . Hay dos procedimientos:

**Predicción no lineal:** La predicción de  $g(X_{t+h})$  se define por:  $E^{\beta_t}(g(X_{t+h}))$  = proyección ortogonal de  $g(X_{t+h})$  sobre  $L^2(\beta_t)$ , donde:  $\beta_t = \sigma(X_s, s \leq t)$ , la  $\sigma$ -álgebra generada por las variables aleatorias  $X_s, s \leq t$ . ( $L^2(\beta_t) \subseteq L^2$ , cerrado)

**Predicción lineal:** Sean  $(X_t)$  d.e., centrado;  $\mu_t = \overline{e.v(X_s; s \leq t)} \subseteq L^2$  (clausura del e.v. generado por las  $X_s, s \leq t$ ). Se toma como predicción de  $g(X_{t+h})$  a la proyección ortogonal de  $g(X_{t+h})$  sobre  $\mu_t$ :

$$\text{Pr}^{\mu_t} g(X_{t+h}).$$

**Observación 1.9.** Si  $(X_t)$  es un proceso gaussiano, las dos predicciones coinciden.



**Error de predicción:**

- no lineal:  $E(g(X_{t+h}) - E^{\beta_t} g(X_{t+h}))^2$
- lineal:  $E(g(X_{t+h}) - \text{Pr}^{u_t} g(X_{t+h}))^2$

**Observación 1.10.** Se puede demostrar que si  $\hat{X}_{t,p}$  = proyección ortogonal de  $X_t$  sobre el espacio vectorial generado por:  $X_{t-1}, \dots, X_{t-p}$ , entonces

$$\hat{X}_{t,p} \xrightarrow[p \rightarrow \infty]{m.c.} \hat{X}$$

Donde  $\hat{X}_t = \text{Pr}^{u_{t-1}}(X_t)$ . Este resultado permite utilizar en la práctica  $\hat{X}_{t,p}$  en lugar de  $\hat{X}_t$ , con un número suficientemente grande de observaciones.

**1.5.2 Descomposición de Wold**

**Teorema 1.8.** Sea  $(X_t, t \in \mathbb{Z})$  d.e, centrado y regular ( $\sigma^2 = E(X_t - \hat{X}_t)^2 > 0$ ). (Ejercicio: Verificar que  $E(X_t - \hat{X}_t)^2$  no depende de  $t$ ). Entonces:  $X_t = \sum_{j=0}^{\infty} \lambda_j u_{t-j} + v_t, t \in \mathbb{Z}$  donde  $(u_t)$  r.b. (débil) de varianza  $\sigma^2, \lambda_0 = 1, \sum \lambda_j^2 < \infty, u_t \in \mu_t, u_t \perp \mu_{t-1} (v_t)$  centrado, los  $u_t$  son ortogonales a los  $v_s, v_t \in \bigcap_{s=0}^{\infty} \mu_{t-s}$ . Además, la descomposición es única.

**Observación 1.11.**  $u_t = X_t - \hat{X}_t \Rightarrow X_t = \hat{X}_t + u_t$ .  $u_t$  se dice la innovación del proceso y  $v_t$  la parte determinista del proceso.

*Demostración.* Anexo A.2 □

**Definición 1.4.** Un proceso no regular se dice determinista.

**Ejercicio 1.** 1.  $(v_t)$  es un proceso determinista, es decir, si  $V_t = \text{ev}$  cerrado, generado por  $v_s, s \leq t$  y  $\tilde{v}_t = \text{Pr}^{V_{t-1}} v_t$  entonces  $\tilde{v}_t = v_t$ .

2.  $W_t = \sum_{j=0}^{\infty} \lambda_j u_{t-j}, t \in \mathbb{Z}$ , es un proceso puramente no determinista y su descomposición de Wold es:

$$W_t = \sum_{j=0}^{\infty} \lambda_j u_{t-j}$$

Así, el teorema de Wold permite concluir que todo proceso estacionario se puede expresar como la suma de un proceso lineal ( $W_t$ ) y un proceso determinista ( $v_t$ ).

**1.6 Ejercicios Propuestos**

1. Sea  $(X_t)_{t \in \mathbb{Z}}$  un proceso d.e. Se define  $Y_t = X_{t+1} - X_t, t \in \mathbb{Z}$ . Mostrar que  $(Y_t)_{t \in \mathbb{R}}$  es d.e., con media cero y función de autocovarianza:

$$\gamma_Y(t) = 2\gamma_X(t) - \gamma_X(t-1) - \gamma_X(t+1)$$

2. Sea  $(u_t)$  ruido blanco. Mostrar que los procesos definidos por:  $X_t = u_t$  y  $Y_t = (-1)^t u_t$  son estacionarios. ¿Qué puede decir del proceso  $Z_t = X_t + Y_t, t \in \mathbb{Z}$ ?
3. Sea  $(X_t)_{t \in \mathbb{Z}}$  el proceso estocástico definido por:

$$X_t = \sum_{j=1}^q (A_j \cos \lambda_j t + B_j \sin \lambda_j t), \quad t \in \mathbb{Z}; \lambda_1, \dots, \lambda_q \in \mathbb{R}; q \in \mathbb{N}^* \quad (\mathbb{N}^* = \mathbb{N} \setminus \{0\})$$

con  $A_1, \dots, A_q, B_1, \dots, B_q$  v.a. ortogonales 2-2, centradas, que verifican

$$E A_j^2 = E B_j^2 = \sigma_j^2 \quad (j = 1, \dots, q)$$

- (a) Probar que  $(X_t)_{t \in \mathbb{Z}}$  es d.e.
- (b) Mostrar que su medida espectral viene dada por:

$$\mu = \sum_{j=1}^q \frac{\sigma_j^2}{2} [\delta_{(-\lambda_j)} + \delta_{(\lambda_j)}]$$

4. Sean  $Z_1$  y  $Z_2$  variables aleatorias i.i.d. con distribución  $N(0, \sigma^2)$ . Se define al proceso estocástico

$$Z_t = Z_1 \cos \lambda t + Z_2 \sin \lambda t \quad t \in \mathbb{Z} \text{ y } \lambda \in [-\pi, \pi]. \quad (1.1)$$

- (a) ¿ $(Z_t)_{t \in \mathbb{Z}}$  es débilmente estacionario? En caso afirmativo, calcular su función de autocovarianza y la medida espectral correspondiente.

**Definición 1.5.** Se dice que un proceso  $(X_t)_{t \in T}$  es gaussiano si toda combinación lineal (finita) de v.a. de  $(X_t)$  es normal.

- (b) ¿Un proceso gaussiano es de segundo orden? ¿es estacionario?
- (c) Demuestre que el proceso definido por (1.1) es gaussiano.
5. Sea  $U$  una v.a. que sigue una ley uniforme en  $[0, 1]$ :  $U_{[0,1]}$ . Se define:

$$X_n = \sin(2\pi n U) \quad \forall n \in \mathbb{Z}$$

- (a) ¿El proceso  $(X_n)_{n \in \mathbb{Z}}$  es débilmente estacionario? ¿Qué se puede decir del proceso  $(X_n)_{n \in \mathbb{N}^*}$ ?
- (b) Mostrar que para todo boreliano real  $A$  se tiene:

$$P(X_n \in A) = \int_0^1 1_A \sin 2\pi y \frac{dy}{n}$$

y a partir de esto probar que la ley de  $X_n$  es independiente de  $n$ .

- (c) Probar que el proceso  $(X_n)_{n \in \mathbb{N}^*}$  no es estrictamente estacionario.
6. Mostrar que:
- (a)  $(X_n)_{n \in \mathbb{N}}$  es un proceso estrictamente estacionario si y sólo si  $\text{Dist}(X_0, \dots, X_n) = \text{Dist}(X_1, \dots, X_{n+1}) \forall n \in \mathbb{N}$ .

- (b)  $(X_n)_{n \in \mathbb{Z}}$  es un proceso estrictamente estacionario si y sólo si  $\text{Dist}(X_{-n}, \dots, X_0, \dots, X_n) = \text{Dist}(X_{-n+1}, \dots, X_1, \dots, X_{n+1}) \quad \forall \quad n \in \mathbb{N}$ .
- (c) Para todo proceso estacionario en sentido estricto  $(X_n)_{n \geq 0}$ , existe un proceso estrictamente estacionario único  $(Y_n)_{n \in \mathbb{Z}}$ , salvo por una equivalencia, t.q.  $(X_n)_{n \geq 0}$  y  $(Y_n)_{n \geq 0}$  sean equivalentes.

**Definición 1.6.** Se dice que dos procesos  $(X_t)_{t \in T}$  y  $(Y_t)_{t \in T}$  son equivalentes (en el sentido amplio) si coinciden sus respectivas distribuciones finitas.

7. ¿Cuáles de las siguientes funciones son funciones de autocovarianza? En caso afirmativo, dar el proceso asociado con su medida espectral sobre  $[-\pi, \pi]$ .

(a)  $\gamma(h) = 1 + |h|$

(b)  $\gamma(h) = \begin{cases} 1 & \text{si } h = 0 \\ -\frac{1}{2} & \text{si } |h| = 1 \\ 0 & \text{en caso contrario} \end{cases}$

(c)  $\gamma(h) = 1 + \frac{1}{4} \sin 4h$  (los ángulos se miden en radianes)

(d)  $\gamma(h) = 1 + \frac{1}{4} \cos 4h$  (los ángulos se miden en radianes)

(e)  $\gamma(h) = \begin{cases} 1 & \text{si } |h| \leq 1 \\ 0 & \text{en otro caso} \end{cases}$

Considere  $T = \mathbb{Z}$

8. Sean  $\varepsilon_1, \varepsilon_2, \dots$  variables aleatorias independientes, centradas t.q.  $V(\varepsilon_i) = \sigma^2$ . Considere que:

$$\begin{cases} X_1 = \varepsilon_1 \\ \vdots \\ X_n = \theta X_{n-1} + \varepsilon_n \\ \vdots \end{cases}$$

- (a) Encontrar el predictor lineal óptimo de  $X_{n+1}$ , dadas  $X_1, \dots, X_n$ .
- (b) Encontrar el predictor óptimo de  $X_{n+1}$ , dadas  $X_1, \dots, X_n$ .



## Capítulo 2

# Procesos Lineales, Modelos ARMA

### 2.1 Procesos Lineales

**Definición 2.1.**  $(X_t, t \in \mathbb{Z})$  centrado, de segundo orden, se dice un proceso lineal si se puede expresar como:

$$X_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}, \quad t \in \mathbb{Z} \quad (2.1)$$

donde,

- $\psi_0 = 1$
- $\sum_{j=0}^{\infty} |\psi_j|^2 < \infty$
- $(u_t)_{t \in \mathbb{Z}}$  es un ruido blanco de varianza  $\sigma^2$
- $\psi_j \in \mathbb{R}$  o  $\psi_j \in \mathbb{C}$  (Se tomará  $\psi_j \in \mathbb{R}$  en este documento).

Bajo estas condiciones la serie converge en media cuadrática (m.c.).

**Observación 2.1.** 1. (2.1) no es necesariamente la descomposición de Wold de  $(X_t)_{t \in \mathbb{Z}}$ .

2.  $u_t \perp u_{t-1}$  pues:

$$X_{t-s} = \sum_{j=0}^{\infty} \psi_j u_{t-s-j} \implies u_t \perp X_{t-s}, \quad s > 0$$

3.  $(X_t)$  d.e., pues es centrado y además:

$$\begin{aligned} E(X_t, X_{t+h}) &= \sum_{i,j} \psi_i \psi_j E(u_{t-i} u_{t+h-j}) \\ &= \sigma^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+h} = \gamma_h \quad (t-i = t+h-j \implies j = i+h) \end{aligned}$$

4. Se pasa de  $(u_t)$  a  $(X_t)$  por una transformación lineal, llamada *filtro lineal*.
5. Supóngase que  $\sum_j |\psi_j| < \infty$

$$\sum_h |\gamma_h| \leq \sigma^2 \sum_{h,j} |\psi_j \psi_{j+h}| \leq \sigma^2 \left( \sum_j |\psi_j| \right) \left( \sum_{j'} |\psi_{j'}| \right) < \infty$$

$$\therefore \sum |\gamma_h| < \infty$$

entonces existe una medida espectral  $\mu$  de  $(X_t)$ , absolutamente continua, con una densidad  $f$  continua (si  $f \geq 0$ ).

**Definición 2.2.** Se dirá que el proceso lineal es invertible si:

$$u_t = \sum_{j=0}^{\infty} \pi'_j X_{t-j}, \quad t \in \mathbb{Z}$$

donde

$$\pi'_0 = 1 \quad y \quad \sum_{j=0}^{\infty} (\pi'_j)^2 < \infty$$

**Observación 2.2.** Si el proceso  $(X_t)_{t \in \mathbb{Z}}$  es invertible, entonces:

1.  $u_t \in \mu_t$
- 2.

$$X_t = \sum_{j=1}^{\infty} \pi_j X_{t-j} + u_t = \hat{X}_t + u_t \quad t \in \mathbb{Z}, \text{ con } \pi_j = -\pi'_j \quad (2.2)$$

donde  $u_t$  es la innovación en el instante  $t$ . Entonces el problema estadístico será la estimación de los  $\pi_j$ . Se puede aproximar el modelo (2.2) por:

$$X_t = \sum_{j=1}^p \pi_j X_{t-j} + u_t, \quad t \in \mathbb{Z}$$

(proceso autoregresivo de orden  $p$  :  $AR(p)$ ) pero es necesario estimar  $p$ .

Si  $X_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$ , no es invertible y los  $\psi_j$  son pequeños para  $j$  suficientemente gran-

de, se puede considerar también la aproximación por  $X_t = \sum_{j=0}^q \psi_j u_{t-j}$ ,  $t \in \mathbb{Z}$  (proceso media móvil de orden  $q$  :  $MA(q)$ ). Es necesario también estimar  $q$ .

**Notaciones:** Considérese el operador de retardo  $B$  (Backward es su expresión en inglés), definido por  $BX_t = X_{t-1}$  y sean:

$$\psi(B) = \sum_{j=0}^{\infty} \psi_j B^j, \quad \Pi'(B) = \sum_{j=0}^{\infty} \pi'_j B^j$$

entonces se puede escribir:

$$X_t = \psi(B) u_t \quad y \quad u_t = \Pi'(B) X_t$$

El operador  $B$  es lineal e invertible. Se define su inverso  $B^{-1} = F$  por  $FX_t = X_{t+1}$ ;  $F$  se llama operador de avance.

Las principales propiedades del operador  $B$  se describen en el Anexo B.1

**Teorema 2.1.** Sea  $(Y_t, t \in \mathbb{Z})$  d.e. centrado, de medida espectral  $\mu$ . Se define:

$$Z_t = \sum_{j=0}^{\infty} c_j y_{t-j}, \quad t \in \mathbb{Z},$$

entonces  $(Z_t, t \in \mathbb{Z})$  es d.e., centrado y de medida espectral:

$$f_z(\lambda) = \left| \sum_j c_j e^{i\lambda j} \right|^2 d\mu(\lambda), \quad \lambda \in [-\pi, \pi]$$

*Demostración.* Luego del ejercicio (1.1) del capítulo 1, queda por determinar la densidad espectral de  $(Z_t)$ . Se supondrá, de manera general, en esta demostración que los  $c_j$  son complejos.

$$E(Z_s Z_t) = \sum_{j,j'} c_j \bar{c}_{j'} E(Y_{s-j} Y_{t-j'}) = \sum_{j,j'} c_j \bar{c}_{j'} \int_{[-\pi, \pi]} \cos \lambda ((t-s) + (j' - j)) d\mu(\lambda)$$

Por el teorema de convergencia dominada:

$$\begin{aligned} &= \int_{[-\pi, \pi]} \underbrace{\sum_{j,j'} c_j \bar{c}_{j'} \cos \lambda ((t-s) + (j' - j))}_{\leq (\sum |c_j|)^2} d\mu(\lambda) \\ E(Z_s Z_t) &= \int \Re \left[ \sum_{j,j'} c_j c_{j'} e^{i\lambda((t-s)+(j'-j))} \right] d\mu(\lambda) = \int \Re e^{i\lambda(t-s)} \left| \sum_j c_j e^{i\lambda j} \right|^2 d\mu(\lambda) \\ \therefore \gamma_{t-s} &= \int \cos \lambda (t-s) \left| \sum_j c_j e^{i\lambda j} \right|^2 d\mu(\lambda) = \int \cos \lambda (t-s) f_z(\lambda) d\mu(\lambda) \end{aligned}$$

El resultado se deduce por la unicidad de la representación espectral de los  $\gamma_t$ . □

**Aplicación:** Sea  $X_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}$  proceso lineal, con

$$\sum_{j=0}^{\infty} |\psi_j| < \infty$$

Puesto que  $(u_t)$  ruido blanco, la medida espectral es  $\frac{\sigma^2}{2\pi} d\lambda$ , entonces la medida espectral de  $X_t$  se escribe:

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| \sum_j \psi_j e^{i\lambda j} \right|^2 d\lambda$$

**Consecuencia:** Un proceso lineal es ergódico (la medida en 0 es nula, pues  $\mu(\{0\}) = 0$ ). Así, entonces, para un proceso lineal se satisface la Ley de los Grandes Números.

## 2.2 Funciones de Autorrelación Simple y Parcial de un P.E.S.O.

**Definición 2.3.** Se llama función de autocorrelación de orden  $l$  de un p.e.s.o (denotada por  $\rho(l)$  o  $\rho_l$ ) a:

$$\rho(l) = \frac{\gamma(l)}{\gamma(0)}$$

Puesto que la varianza de  $X_t$  es constante, este es el coeficiente de correlación lineal entre  $X_t$  y  $X_{t+l}$ .

**Observación 2.3.** Para realizar estimaciones, se supondrá que se dispone de una observación de cada una de las variables  $X_1, \dots, X_T$

Dado que las covarianzas se conservan por saltos,  $\rho(l)$  puede estimarse por:

$$\hat{\rho}(l) = \frac{\sum_{t=l+1}^T (X_t - \bar{X})(X_{t-l} - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$$

**Definición 2.4.** Se llama función de autocorrelación parcial de orden  $l$  (denotada por  $r(l)$  o  $r_l$ ) de un p.e.s.o. al coeficiente  $a_l$  de  $X_{t-l}$  en la regresión afín de  $X_t$  sobre  $1, X_{t-1}, \dots, X_{t-l}$ .

El sistema se escribe:

$$X_t = a_0 + a_1 X_{t-1} + \dots + a_l X_{t-l} + u_t, \quad t = l+1, l+2, \dots, T$$

Se define también:

$$r_0 = 1 \quad \text{y} \quad r_1 = \rho_1$$

Se puede demostrar que una de las formas de estimar  $r(l)$  es resolver el siguiente sistema de ecuaciones lineales (con matriz simétrica):

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(l) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(l-1) \\ \rho(1) & 1 & \cdots & \rho(l-2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(l-1) & \rho(l-2) & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_l \end{bmatrix}$$

Los valores desconocidos de  $\rho(l)$  se reemplazan por sus estimaciones. Existen técnicas más eficientes, como por ejemplo el método de autocorrelaciones inversas (Cleveland, 1972).



## 2.3 Procesos Autoregresivos

**Definición 2.5.** Sea  $(X_t, t \in \mathbb{Z})$  d.e. que satisfice:

$$X_t = \sum_{j=1}^p \pi_j X_{t-j} + u_t \quad t \in \mathbb{Z}$$

t.q.  $\pi_p \neq 0$ ,  $(u_t)$  ruido blanco con  $u_t \perp \mu_{t-1}$ ,  $t \in \mathbb{Z}$ ; entonces  $(X_t, t \in \mathbb{Z})$  se llama un proceso autoregresivo de orden  $p$  (notación:  $AR(p)$ ).

### 2.3.1 El caso del modelo $AR(1)$

Sea  $(X_t, t \in \mathbb{Z})$  tal que:

$$X_t = \rho X_{t-1} + u_t \quad (2.3)$$

con  $(u_t)$  ruido blanco débil y  $|\rho| < 1$

El proceso definido por (2.3) es débilmente estacionario; además,  $u_t$  y  $(X_s, s \leq t-1)$  son ortogonales;  $u_t$  se llama la innovación en el instante  $t$ .

- **Existencia y unicidad de un proceso  $AR(1)$**

1. Se va a demostrar que existe una solución estacionaria para  $X_t = \rho X_{t-1} + u_t$ .

Puesto que:

$$X_{t-1} = \rho X_{t-2} + u_{t-1}, \quad t \in \mathbb{Z}$$

Se tiene que:

$$\begin{aligned} X_t &= u_t + \rho u_{t-1} + \rho^2 X_{t-2} \\ &\vdots \\ X_t &= u_t + \rho u_{t-1} + \dots + \rho^s u_{t-s} + \rho^{s+1} X_{t-s-1} \end{aligned}$$

Se supone que:  $\sup_i E(X_i^2) < \infty$ . Se va demostrar que existe convergencia en media cuadrática (m.c.) de las sumas finitas hacia  $X_t$ .

$$E \left( X_t - \sum_{j=0}^s \rho^j u_{t-j} \right)^2 = \rho^{2s+2} E X_{t-s-1}^2 \rightarrow 0$$

cuando  $s \rightarrow \infty$ , pues  $|\rho| < 1$ , por lo tanto,  $X_t = \sum_{j=0}^{\infty} \rho^j u_{t-j}$  será la solución del problema.

2. Se va a probar que  $(X_t)$  es d.e.:

(a)  $E X_t = 0$  (Se verifica rápidamente ya que se puede introducir la esperanza dentro del sumatorio)

(b) Por la bicotinuidad de un producto escalar en el espacio de Hilbert  $L^2$  se tiene:

$$\begin{aligned} \text{Cov}(X_t, X_s) &= \sum_{j,j'} \rho^{j+j'} E(u_{t-j} u_{s-j'}) \\ &= \sum_j \rho^{2j+(s-t)} E(u_{t-j}^2) \end{aligned}$$

$$= \sigma^2 \sum_j \rho^{2j+(s-t)}$$

$X_s = \sum_{j=0}^{\infty} \rho^j u_{s-j}$  depende solamente de  $u_s, u_{s-1}, \dots$ ; pero  $\{u_s\}$  es un r.b.d. y entonces se verifica que  $u_t$  es la innovación en el instante  $t$ .

• ¿Qué sucede si  $|\rho| \geq 1$ ?

Supóngase que exista una solución de  $X_t = \rho X_{t-1} + u_t$ ,  $t \in \mathbb{Z}$ ;  $u_t$  independiente de  $(X_s, s < t)$ .

$$E X_t^2 = (1 + \rho^2 + \dots + \rho^{2s}) \sigma^2 + \rho^{2s+2} E X_{t-s-1}^2 \geq (1 + \rho^2 + \dots + \rho^{2s}) \sigma^2 \rightarrow \infty$$

cuando  $s \rightarrow \infty$ . Por tanto, no existe solución estacionaria. Sin embargo, existe solución no estacionaria; por ejemplo:

$$\begin{cases} X_t = u_1 + \dots + u_t & t \geq 1 \\ X_0 = 0 \\ X_t = -u_0 - \dots - u_{t+1} & t \leq -1 \end{cases}$$

Además, esta solución no satisface que sea no correlacionado  $u_t$  con  $(X_s, s < t)$ , pues  $E(u_0 X_{-1}) = -\sigma^2$ .

### 2.3.2 Unicidad de la descomposición para el caso del $AR(p)$ :

Supóngase que  $X_t$ , también se pueda expresar por:

$$X_t = \sum_{j=1}^p \pi'_j X_{t-j} + u'_t$$

con  $(u'_t)$  ruido blanco de igual varianza que  $(u_t)$ ,  $u'_t \perp \mu_{t-1}$ ,  $\pi'_p \neq 0$

Entonces  $\hat{X}_t = \sum_{j=1}^p \pi_j X_{t-j} = \sum_{j=1}^p \pi'_j X_{t-j}$  (pues  $u_t, u'_t \perp \mu_{t-1}$ )

Si  $\pi_1 \neq \pi'_1$ ,  $X_{t-1} = \sum_{j=1}^{p-1} \pi_j X_{t-1-j} \in \mu_{t-2}$ ,  $\forall t \in \mathbb{Z}$

Así,  $X_t \in \mu_{t-1} \Rightarrow u_t \in \mu_{t-1} \Rightarrow u_t = 0$ ; esto es una contradicción, pues  $E(u_t^2) = \sigma^2 > 0$

$$\therefore \pi_1 = \pi'_1$$

De la misma manera se demuestra que  $\pi_2 = \pi'_2$ , etc.

**Teorema 2.2.** Una condición necesaria y suficiente para que exista un  $AR(p)$ , que satisfaga  $\pi'(B)X_t = u_t$ , es que las raíces de la ecuación  $\pi'(z) = 0$  se encuentren fuera del círculo unidad. Bajo esta condición:

$$X_t = \sum_{j=0}^{\infty} \psi_j u_{t-j}, \quad t \in \mathbb{Z}$$

donde los  $\psi_j$ , son los coeficientes de Taylor de  $\frac{1}{\Pi'(z)}$

*Demostración.* Ver anexo B.3 □

**Observación 2.4.** Los coeficientes  $\psi_j$  pueden determinarse con uno de los siguientes métodos:

1. Puesto que  $\Pi'(z) \frac{1}{\Pi'(z)} = 1$ , se identifican los coeficientes de los términos del mismo grado. Se obtiene un sistema de ecuaciones en  $\psi_i$ , que puede resolverse sucesivamente:

$$\begin{aligned}\psi_0 &= 1 \\ \psi_1 - \pi_1 &= 0 \\ \psi_2 - \psi_1\pi_1 - \pi_2 &= 0 \\ \psi_p - \psi_{p-1}\pi_1 - \cdots - \psi_1\pi_{p-1} - \pi_p &= 0 \\ &\vdots \\ \psi_n - \psi_{n-1}\pi_1 - \cdots - \psi_{n-p+1}\pi_{p-1} - \psi_{n-p}\pi_p &= 0, \forall n > p \\ &\vdots\end{aligned}$$

2. Efectuar la división de 1 por  $\Pi(z)$  según las potencias crecientes; los coeficientes se caracterizan entonces por:

$$1 = \Pi(z)[\psi_0 + \psi_1 z + \cdots + \psi_p z^r] + z^{r+1} \lambda_r(z) \quad \forall r$$

donde  $\lambda_r(z)$  es un polinomio en  $z$ .

3. Descomponer en elementos simples la fracción  $\frac{1}{\Pi'(z)}$  y escribir el desarrollo en series de cada término. Supóngase que todas las raíces de  $\pi'(z)$  sean reales y distintas:

$$\begin{aligned}\frac{1}{\Pi'(z)} &= \frac{1}{\prod_{i=1}^p (1 - \lambda_i z)} \\ &= \sum_{i=1}^p \frac{a_i}{1 - \lambda_i z} \\ &= \sum_{i=1}^p a_i \sum_{j=0}^{\infty} \lambda_i^j z^j \\ &= \sum_{j=0}^{\infty} \left[ \sum_{i=1}^p a_i \lambda_i^j \right] z^j\end{aligned}$$

El caso de raíces complejas o múltiples se trata de la manera habitual.

**Ejemplo 2.1.** Se vio en el inicio de esta sección el proceso  $AR(1)$ :

$$X_t = \rho X_{t-1} + u_t, \quad |\rho| < 1$$

La condición  $|\rho| < 1$ , se requiere para que este proceso exista, pues: puede resolverse sucesivamente:

$$X_t - \rho X_{t-1} = u_t$$

$$(1 + \rho B) X_t = u_t$$

$$1 - \rho z = 0 \iff z = \frac{1}{\rho}$$

**Teorema 2.3.** La autocovarianza de un  $AR(p)$  satisface las ecuaciones de Yule-Walker.

$$\begin{cases} \sum_{j=1}^p \pi_j \gamma_{k-j} = \gamma_k & k = 1, 2, \dots \\ \sum_{j=1}^p \pi_j \gamma_j + \sigma^2 = \gamma_0 \end{cases}$$

y su densidad espectral se escribe:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^p \pi_j e^{i\lambda j} \right|^{-2}$$

*Demostración.* Sea  $(X_t)$  un  $AR(p)$

- $k \geq 1: 0 = E(u_t X_{t-k}) = E\left(\left(X_t - \sum_{j=1}^p \pi_j X_{t-j}\right) X_{t-k}\right) = \gamma_k - \sum_{j=1}^p \pi_j \gamma_{k-j}$
- $k = 0: \sigma^2 = E(u_t X_t) = E\left(\left(X_t - \sum_{j=1}^p \pi_j X_{t-j}\right) X_t\right) = \gamma_0 - \sum_{j=1}^p \pi_j \gamma_j$
- $u_t = X_t - \sum_{j=1}^p \pi_j X_{t-j} \Rightarrow \frac{\sigma^2}{2\pi} d\lambda = \left| 1 - \sum_{j=1}^p \pi_j e^{i\lambda j} \right|^2 f(\lambda) d\lambda = \left| 1 - \sum_{j=1}^p \pi_j e^{i\lambda j} \right|^2 d\mu(\lambda)$

Estas ecuaciones permiten estimar  $\sigma^2$  y los  $\pi_j$ , poniendo en lugar de  $\gamma_i$  su estimación.  $\square$

### 2.3.3 Autocorrelación asintótica de un proceso $AR(p)$

Considerando la definición 2.3, se tiene que el coeficiente de autorrelación es:

$$\rho(l) = \frac{\gamma(l)}{\gamma(0)} = \frac{\gamma_l}{\gamma_0} = \rho_l, \quad l \geq 0$$

Si se dividen por  $\gamma_0$  las ecuaciones de Yule – Walker, se tiene:

$$\sum_{j=1}^p \pi_j \rho_{\ell-j} = \rho_\ell \Leftrightarrow \pi'(B) \rho_\ell = 0 \quad (\text{raíces de } \Pi'(z) \text{ de } || > 1)$$

Así, la sucesión  $\rho_\ell$  satisface una ecuación en diferencias.

Se sabe que:

$$\rho_\ell = \sum_{i=1}^p a_i G_i^\ell$$

donde,

- $G_i$  : inversas de las raíces de  $\pi'(z)$  (se suponen distintas y de multiplicidad 1).
- $a_i$  : constantes arbitrarias.

De esta ecuación se puede deducir que  $\rho_\ell \rightarrow 0$  a velocidad exponencial. El resultado también se cumple en el caso general (Ver Anexo B.2).

### 2.3.4 Autorrelación parcial asintótica de un proceso $AR(p)$

**Teorema 2.4.** Si  $(X_t)_{t \in \mathbb{Z}}$  es  $AR(p)$  y  $r_\ell$  el coeficiente de autocorrelación parcial, entonces:

$$\begin{cases} r_p = \pi_p \neq 0 \\ r_\ell = 0, & \ell > p \end{cases}$$

*Demostración.* Se tiene que:

$$X_t = \sum_{j=1}^p \pi_j X_{t-j} + u_t, \quad t \in \mathbb{Z}$$

Se denota por  $X_t^*$  (respectivamente  $X_{t-l}^*$ ) a la proyección ortogonal de  $X_t$  sobre el espacio vectorial generado por  $X_{t-1}, X_{t-2}, \dots, X_{t-l+1}$  (respectivamente  $X_{t-l}$ ).

- Si  $l = p$ :

$$X_t^* = \sum_{j=1}^{p-1} \pi_j X_{t-j} + \pi_p X_{t-p}^* + 0$$

luego:

$$X_t - X_t^* = \pi_p (X_{t-p} - X_{t-p}^*) + u_t$$

entonces:

$$E((X_t - X_t^*)(X_{t-p} - X_{t-p}^*)) = \pi_p V(X_{t-p} - X_{t-p}^*) + 0$$

Por la estacionariedad del proceso se tiene que:

$$V(X_t - X_t^*) = V(X_{t-p} - X_{t-p}^*) > 0$$

$$\therefore r_p = \frac{E((X_t - X_t^*)(X_{t-p} - X_{t-p}^*))}{[V(X_t - X_t^*) V(X_{t-p} - X_{t-p}^*)]^{\frac{1}{2}}} = \pi_p$$

puesto que  $V(X_t - X_t^*) > 0$ .

- Para  $k > p$ ,  $X_t = \sum_{j=1}^p \pi_j X_{t-j} + u_t$

$$X_t^* = \sum_{j=1}^p \pi_j X_{t-j} + 0$$

$$\therefore X_t - X_t^* = u_t \Rightarrow E((X_t - X_t^*)(X_{t-\ell} - X_{t-\ell}^*)) = 0$$

$$\therefore r_k = 0$$

□

**Observación 2.5** (Cómo reconocer y determinar un proceso AR). A partir del comportamiento de las autocorrelaciones asintóticas (simple y parcial) de un  $AR(p)$  se puede concluir lo siguiente: Si  $(X_t)$  es un proceso  $AR(p)$  entonces,  $r(l) \rightarrow 0$  cuando  $l > p$  y  $\rho(l) \rightarrow 0$  a velocidad exponencial (descienden rápidamente hacia cero). Ver Figuras 2.1, 2.2 y 2.3.

**Ejemplo 2.2.** Considérese la serie  $X_t + 0,4X_{t-1} - 0,5X_{t-2} = u_t$ , de la cual se han simulado 200 observaciones. El ruido blanco corresponde a observaciones  $u_t$  independientes e idénticamente distribuidas de la distribución normal estándar ( $N(0,1)$ )

**Figura 2.1:** Modelo simulado  $AR(eq4)$

Se puede observar a continuación el comportamiento de las funciones de autocorrelación (FAC) y autocorrelación parcial (FACP) estimadas.

Las bandas de confianza de los gráficos sirven para determinar cuáles parámetros (autocorrelación o autocorrelación parcial), se pueden considerar significativos; la manera de construirlas se verá más adelante.

**Figura 2.2:** Comportamiento de  $r(\ell)$

**Figura 2.3:** Comportamiento de  $r(\ell)$

**Observación 2.6.** 1. El inverso del operador  $\Pi'(B)$  existe, si las raíces de  $\Pi'(z)$  son de  $|| \neq 1$ . Supóngase que las primeras  $r$  raíces sean de  $|| > 1$  y las  $(p-r)$  últimas de  $|| < 1$  entonces,

$$\begin{aligned}\Pi'(B) &= \prod_{i=1}^r (1 - \lambda_i B) \prod_{i=r+1}^p \left(1 - \frac{1}{\lambda_i} B\right) \prod_{i=r+1}^p (-\lambda_i B) \\ &= \Phi_1(B) \Phi_2(F) \lambda B^{p-r} \quad \text{con } \lambda = \prod_{i=r+1}^p (-\lambda_i)\end{aligned}$$

Las  $\lambda_i$  son las inversas de las raíces del polinomio  $\Pi'(z)$ . Puesto que  $\Phi_1(B)$  y  $\Phi_2(B)$  son invertibles, entonces  $\Pi'(z)$  es invertible. Además:

$$\frac{1}{\Pi'(B)} = \frac{1}{\Phi_1(B)} \frac{1}{\Phi_2(F)} \frac{1}{\lambda} F^{p-r}$$

Se puede observar que las potencias negativas de  $B$  no intervienen si y solamente si  $p = r$ .

Esto significa, utilizando el ejercicio 1.1, que para que un proceso autoregresivo sea estacionario únicamente requiere que el polinomio asociado  $\Pi'(z)$  tenga las raíces de  $|| \neq 1$ .

2. Cambiando el ruido blanco, se puede suponer que las raíces de  $\Pi'(z)$  son  $|| > 1$ .

Se sabe que la densidad espectral de  $(X_t)$  es:

$$\begin{aligned} f_X(\lambda) &= \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^p \pi_j e^{i\lambda j} \right|^{-2} \\ &= \frac{\sigma^2}{2\pi} \left| \Pi'(e^{i\lambda}) \right|^{-2} \\ &= \frac{\sigma^2}{2\pi} \left| \prod_{j=1}^p \left( 1 - \frac{e^{i\lambda}}{z_j} \right) \right|^{-2} \end{aligned}$$

donde  $z_j = \frac{1}{\lambda_j}$  son las raíces de  $\Pi'(z)$ .

Supóngase que las primeras  $r$  raíces de  $\Pi'(z)$  sean de  $|| > 1$  y las  $p - r$  últimas de  $|| < 1$ . Si remplazamos las raíces de  $|| < 1$  por sus inversas, se tiene:

$$\Phi(B) = \prod_{j=1}^r \left( 1 - \frac{B}{z_j} \right) \prod_{j=r+1}^p (1 - z_j B)$$

Considérese ahora el proceso estacionario  $(\eta_t; t \in \mathbb{Z})$  definido por:

$$\eta_t = \Phi(B) X_t, \quad t \in \mathbb{Z}$$

Se calcula su densidad espectral:

$$\begin{aligned} f_\eta(\lambda) &= \left| \Phi(e^{i\lambda}) \right|^2 f_X(\lambda) \\ &= \left| \Phi(e^{i\lambda}) \right|^2 \frac{\sigma^2}{2\pi} \left| \Pi'(e^{i\lambda}) \right|^{-2} \\ &= \frac{\sigma^2}{2\pi} \prod_{j=r+1}^p \frac{|1 - z_j e^{i\lambda}|^2}{\left| 1 - \frac{1}{z_j} e^{i\lambda} \right|^2} \\ &= \frac{\sigma^2}{2\pi} \prod_{j=r+1}^p |z_j|^2 \end{aligned}$$

(independiente de  $\lambda$ ), por lo tanto,  $(\eta_t, t \in \mathbb{Z})$  es un ruido blanco con varianza menor a  $\sigma^2$

La representación obtenida a partir de la elección del polinomio con raíces de  $|| > 1$ , se llama representación canónica.

**Ejemplo 2.3.** 1. En un modelo  $AR(2)$

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \varepsilon_t$$

( $\varepsilon_t$  es la innovación)

- (a) ¿Podría deducir el valor  $\varphi_2$  si se da la información que  $\varphi_1 = -0,4$  y  $\rho_1 = 0,7$ ?
- (b) La misma pregunta si se conoce que  $\varphi_1 = 0,8$  y  $\rho_2 = 0,6$ .

*Resolución.* (a) Se supone que el proceso es estacionario. Para esto, se requiere que las raíces de  $1 - \phi_1 z - \phi_2 z^2$  estén fuera del círculo unidad. De las ecuaciones de Yule-Walker se obtiene:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \iff \phi_2 = \frac{\rho_1 - \phi_1}{\rho_1} = \frac{0,7 + 0,4}{0,7} = 1,5714$$

Ahora, considérese la ecuación asociada al polinomio autoregresivos  $\varphi(z) : 1 + 0,4z - 1,5714z^2 = 0$ . Las raíces vienen dadas por la expresión:

$$z = \frac{0,4 \pm \sqrt{0,4^2 + 4(1,5717)}}{2 \cdot 1,5714} = \frac{0,4 \pm 2,5388}{3,1428}$$

de donde se obtiene:

$$z_1 = 0,93 \quad \text{o} \quad z_2 = -0,68$$

Se obtiene que  $z_1$  y  $z_2$  son de  $|| < 1$ ; luego, el proceso no es estacionario y por tanto no se pueden utilizar estas soluciones; es decir no se puede obtener el coeficiente  $\phi_2$ .

(b) Supóngase que el proceso es estacionario. De las ecuaciones de Yule-Walker se obtiene:

$$\rho_1 = \phi_1 + \phi_2 \rho_1 \quad \text{y} \quad \rho_2 = \phi_2 + \phi_1 \rho_1$$

Así:

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1 - \phi_2} \iff \rho_2 - \phi_2 \rho_2 = \phi_1^2 + \phi_2 - \phi_2^2 \iff \phi_2^2 - \phi_2(1 + \rho_2) + (\rho_2 - \phi_1^2) = 0$$

Reemplazando los valores de  $\phi_1$  y  $\rho_2$  se obtiene:

$$\phi_2^2 + 1,6\phi_2 - 0,04 = 0 \iff \phi_2 = 1,625 \quad \text{o} \quad \phi_2 = -0,025$$

Si  $\phi_2 = 1,625$ , se obtiene la ecuación variada al polinomio autoregresivo:

$$\varphi(z) = 1 - 0,8z - 1,625z^2 = 0 \iff z_1 = 0,576 \quad z_2 = -1,068$$

Puesto que una raíz se encuentra dentro del círculo unidad, no se pueden aplicar los resultados deducidos de las ecuaciones de Yule-Walker, desarrolladas para procesos estacionarios.

Si  $\phi_2 = -0,025$ , se obtiene la ecuación asociada al polinomio autoregresivo.

$$\varphi(z) = 1 - 0,8z + 0,02z^2 \iff z_1 = 38,70z_2 = 1,292$$

Puesto que las dos raíces son de  $|| < 1$ , el proceso es estacionario. Así, entonces,  $\phi_2 = -0,025$ . □

2. Sea  $(X_t)$  definido por:

$$X_t - \rho X_{t-1} = u_t \quad t \in \mathbb{Z}$$

donde  $(u_t, t \in \mathbb{Z})$  r.b. de varianza  $\sigma^2 > 0$ .

- Si  $|\rho| = 1$ , no existe un proceso estacionario que satisfaga esta relación. Supón-



se que  $\rho = 1$ . Entonces se tendría:

$$\begin{aligned} X_t - X_{t-n} &= u_t + u_{t-1} + \dots + u_{t-n+1} \quad \forall n \in \mathbb{N} \\ \therefore E(X_t - X_{t-n})^2 &= n\sigma^2 \quad \forall n \in \mathbb{N} \end{aligned}$$

Si  $(X_t)$  fuera estacionario, se tendría:

$$\begin{aligned} E(X_t - X_{t-n})^2 &\leq 4\sigma_X^2, \quad \text{donde } \sigma_X^2 = V(X_t) \\ \therefore n\sigma^2 &\leq 4\sigma^2 \quad \forall n \in \mathbb{N} \text{ (imposible)} \end{aligned}$$

Un razonamiento análogo se puede hacer para  $\rho = -1$

- Si  $|\rho| \neq 1$ , como se demostrará a continuación, existe un proceso estacionario único que satisface:

$$X_t - \rho X_{t-1} = u_t \iff (1 - \rho B) X_t = u_t, \quad t \in \mathbb{Z}$$

- Si  $|\rho| > 1$ , se tiene:

$$X_t - \rho X_{t-1} = u_t \iff -\rho B \left(1 - \frac{1}{\rho} F\right) X_t = u_t, \quad t \in \mathbb{Z}$$

luego:

$$\begin{aligned} X_t &= -\frac{1}{\rho} F \left(1 - \frac{1}{\rho} F\right)^{-1} u_t \\ &= \left(-\frac{1}{\rho} F \sum_{i=0}^{\infty} \left(\frac{1}{\rho}\right)^i F^i\right) u_t \\ &= \left(-\sum_{i=0}^{\infty} \frac{1}{\rho^{i+1}} F^{i+1}\right) u_t \\ &= -\sum_{i=1}^{\infty} \rho^{-i} u_{t+i} \end{aligned}$$

y la representación canónica es:

$$X_t - \frac{1}{\rho} X_{t-1} = \eta_t$$

donde  $\eta_t$  es la innovación en el instante  $t$ .

- Si  $|\rho| < 1$ ,  $X_t = (1 - \rho B)^{-1} u_t = \sum_{i=0}^{\infty} \rho^i u_{t-i}$  donde  $u_t$  es la innovación en el instante  $t$ .

La representación de avance del proceso es entonces:

$$X_t - \rho X_{t+1} = \varepsilon_t$$

Ahora bien,

$$\varepsilon_t = (1 - \rho F) X_t = \frac{1 - \rho F}{1 - \rho B} u_t$$

$$\begin{aligned}
&= \sum_{i=0}^{\infty} \rho^i u_{t-i} - \sum_{i=0}^{\infty} \rho^{i+1} u_{t+1-i} \\
&= \sum_{i=0}^{\infty} \rho^i u_{t-i} - \rho u_{t+1} - \sum_{i=1}^{\infty} \rho^{i+1} u_{t+1-i} \\
&= -\rho u_{t+1} + (1 - \rho^2) \sum_{i=0}^{\infty} \rho^i u_{t-i}
\end{aligned}$$

Nótese igualmente que:

$$X_t = \frac{1}{1 - \rho F} \varepsilon_t = \sum_{i=0}^{\infty} \rho^i \varepsilon_{t+i}$$

y así,  $\varepsilon_t \perp X_s, s \geq t+1$

También se tiene:

- $\rho(\ell) = \rho(\rho(\ell-1))$  para  $\ell > 0$ , entonces  $\rho(\ell) = \rho^\ell$  (puesto que  $\rho(0) = 1$ )
- $\rho(\ell) = \rho^{|\ell|} \quad \forall \ell$
- $\gamma(0) = \sigma^2 + \rho\gamma(1) = \sigma^2 + \rho(\rho\gamma(0)) \Rightarrow \gamma(0) = \frac{\sigma^2}{1-\rho^2}$
- $\rho(0) = 1, \quad r(1) = \rho(1) = \rho; \quad r(\ell) = 0, \quad \ell > 1$

## 2.4 Procesos Medias Móviles

**Definición 2.6.** Sea  $(X_t, t \in \mathbb{Z})$  t.q.

$$X_t = u_t - \sum_{j=1}^q \theta_j u_{t-j} \quad t \in \mathbb{Z}$$

con  $\theta_q \neq 0; (u_t)$  r.b. de varianza  $\sigma^2$  t.q.  $u_t \perp \mu_{t-1}$ .  $(X_t, t \in \mathbb{Z})$  se dice un proceso media móvil de orden  $q$  (notación:  $MA(q)$ ).

**Observación 2.7.** 1. Un proceso  $MA(q)$  es centrado, d.e. y tiene una representación única (a causa de la descomposición de Wold).

2. Si  $\Theta(z) = 1 - \theta_1 z - \dots - \theta_q z^q$  y  $\Theta(z)$  no tiene raíces de  $|| = 1$  (es posible suponerlas de  $|| > 1$ , eventualmente con un nuevo ruido blanco  $(\eta_t)$ ), se puede demostrar que el proceso es invertible; es decir,

$$X_t = \sum_{i=1}^{\infty} \pi_i X_{t-i} + \eta_t \quad \text{con} \quad \sum_{i=1}^{\infty} |\pi_i| < \infty$$

Los  $\pi_i$  se obtienen con los métodos descritos en los procesos AR (expresión obtenida a partir de la representación canónica).

3. Puesto que la densidad espectral de  $(X_t)$  es:

$$f_X(\lambda) = \frac{\sigma^2}{2\pi} \left| 1 - \sum_{j=1}^q \theta_j e^{i\lambda j} \right|^2$$

se puede demostrar que la densidad espectral de  $\eta_t$  es:

$$\frac{\sigma^2}{2\pi \left( \prod_{j=r'+1}^q |z_j|^2 \right)}$$

donde  $z_j$  raíz de  $|| < 1$ . Se supondrá que  $\Theta$  tiene raíces de  $|| > 1$ .

### 2.4.1 Autocorrelación de un $MA(q)$ .

$$\begin{aligned} \gamma(h) &= E(X_t X_{t+h}) \\ &= E[u_t - \theta_1 u_{t-1}, \dots, \theta_q u_{t-q}] [u_{t+h} - \theta_1 u_{t+h-1} - \dots - \theta_q u_{t+h-q}] \\ &= \begin{cases} (-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q \theta_{q-h})\sigma^2 & \text{si } 1 \leq h \leq q \\ 0 & \text{si } h > q \end{cases} \end{aligned}$$

Puesto que  $\gamma(0) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma^2$

$$\rho_h = \begin{cases} \frac{-\theta_h + \theta_{h+1}\theta_1 + \dots + \theta_q \theta_{q-h}}{1 + \theta_1^2 + \dots + \theta_q^2} & 1 \leq h \leq q \\ 0 & h > q \end{cases}$$

Se puede observar que  $\rho_q = -\frac{\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} \neq 0$

### 2.4.2 Autocorrelación Parcial de un $MA(q)$

Las fórmulas son complicadas, pero se puede demostrar que los coeficientes decrecen exponencialmente hacia cero.

**Ejemplo 2.4.** Sea el proceso definido por:

$$X_t = u_t - \theta u_{t-1} = (1 - \theta B)u_t \quad t \in \mathbb{Z}$$

donde  $(u_t)$  r.b. de varianza  $\sigma^2$ .

- Si  $|\theta| < 1$ ,  $u_t$  es la innovación al instante  $t$ .
- Si  $|\theta| > 1$ , la representación canónica es:

$$X_t = \varepsilon_t - \frac{1}{\theta} \varepsilon_{t-1}$$

y  $\varepsilon_t$  es la innovación.

Además:

$$\rho(1) = -\frac{\theta}{1 + \theta^2} \Rightarrow -\frac{1}{2} \leq \rho(1) \leq \frac{1}{2} \quad \text{y} \quad \rho(h) = 0, \quad h \geq 2$$

Se puede demostrar que  $r(h) = \frac{-\theta^h(1-\theta^2)}{1-\theta^{2(h+1)}}, h \geq 1$

**Ejemplo 2.5.** Sea la función de covarianza:  $\gamma(h) = \begin{cases} 1 & \text{si } h = 0 \\ 0,4 & \text{si } h = \pm 1 \\ 0 & \text{en otro caso} \end{cases}$

Considérese el proceso media móvil  $X_t = u_t - \theta u_{t-1}$  con  $(u_t)$  r.b. de varianza  $\sigma^2$ .

Se sabe que  $\gamma(0) = \sigma^2(1 + \theta^2)$  y  $\gamma(1) = -\theta\sigma^2$ .

Así:  $0,4 = -\theta\sigma^2$  y  $1 = \sigma^2(1 + \theta^2)$ , de donde se obtiene que  $\theta = -0,5$  y  $\sigma^2 = 0,8$ . Por tanto, el proceso asociado es:  $X_t = u_t + 0,5u_{t-1}$ ,  $(u_t)$  r.b. con varianza  $\sigma^2 = 0,8$ .

Ahora, analizando la  $\text{Cov}(X_t, X_{t+h}) = \gamma(h) = \begin{cases} 1 & \text{si } h = 0 \\ 0,4 & \text{si } h = \pm 1 \\ 0 & \text{en otro caso} \end{cases}$ , se tiene:

$$\begin{aligned} \text{Cov}(X_t, X_{t+h}) &= E(X_t X_{t+h}) \\ &= E[(u_t + 0,5u_{t-1})(u_{t+h} + 0,5u_{t+h-1})] \\ &= E(u_t u_{t+h}) + 0,5 E(u_t u_{t+h-1}) + 0,5 E(u_{t-1} u_{t+h}) + 0,25 E(u_{t-1} u_{t+h-1}) \end{aligned}$$

- si  $h = 0$ :  $\text{Cov}(X_t, X_t) = V(X_t) = E(X_t^2) = (1 + 0,25)0,8 = 1$
- si  $h = 1$ :  $\text{Cov}(X_t, X_{t-1}) = 0,5 E(u_t^2) = 0,5 \cdot 0,8 = 0,4$
- si  $h = -1$ :  $\text{Cov}(X_t, X_{t+1}) = 0,5 E(u_{t-1}^2) = 0,5 \cdot 0,8 = 0,4$
- Si  $|h| > 1$   $\text{Cov}(X_t, X_{t+h}) = 0$

**Observación 2.8** (Cómo reconocer un proceso MA). Si  $(X_t)$  es un proceso  $MA(q)$  entonces,  $\rho(l) \rightarrow 0$  cuando  $l \rightarrow \infty$  a velocidad exponencial (ver Figuras 2.4, 2.5 y 2.6).

**Ejemplo 2.6.** Considérese la serie  $X_t = u_t + 0,8u_{t-1}$ , de la cual se han simulado 200 observaciones (como en el ejemplo precedente). Se presentan las FAC y FACP estimadas.

$$X_t = u_t + 0,8u_{t-1}$$

**Figura 2.4:** Modelo  $MA(1)$

**Figura 2.5:** Comportamiento de  $\rho(l)$

**Figura 2.6:** Comportamiento asintótico de  $r(l)$

## 2.5 Procesos ARMA (Autoregresivos – Medias Móviles)

**Definición 2.7.** Se llama proceso autoregresivo-media móvil de orden  $(p, q)$  o  $ARMA(p, q)$ , a todo proceso estacionario  $(X_t, t \in \mathbb{Z})$  que satisface:

$$\Phi(B)X_t = \theta_0 + \Theta(B)u_t \iff X_t - \sum_{i=1}^p \phi_i X_{t-i} = \theta_0 + u_t - \sum_{i=1}^q \theta_i u_{t-i}$$

$\phi_i, \theta_i$  reales,  $\phi_p \neq 0, \theta_q \neq 0$ ;  $(u_t)$  r.b. de varianza  $\sigma^2$ ; además  $\Phi(z)$  y  $\Theta(z)$  no tienen raíces de  $|| = 1$ .

**Observación 2.9.** Se puede suponer que  $\theta_0 = 0$ , reemplazando  $X_t$  por  $X_t - E X_t$ ; es decir, por  $X_t - \frac{\theta_0}{1 - \sum_{i=1}^p \phi_i}$  (bien definida si  $\Phi(z)$  no tiene raíces de módulo 1).

En lo que sigue se supondrá que  $\theta_0 = 0$ .

### Propiedades:

1. Si  $\Phi(z)$  tiene raíces de  $|| > 1$ :

$$X_t = \sum_{i=0}^{\infty} \psi_i u_{t-i} \quad \text{con } \psi_0 = 1, \sum_{i=1}^{\infty} |\psi_i| < \infty$$

Si  $\Theta(z)$  tiene raíces de  $|| > 1$

$$X_t = \sum_{i=1}^{\infty} \Pi_i X_{t-i} + u_t, \quad \text{con } \sum_{i=1}^{\infty} |\Pi_i| < \infty$$

$u_t$  es la innovación en el instante  $t$ .

2. Si  $\Phi(z)$  y  $\Theta(z)$  tienen raíces de  $|| \neq 1$ , cambiando eventualmente el ruido blanco, se puede obtener una representación con polinomios de raíces con  $|| > 1$ . El ruido blanco que interviene es la innovación. Esta se llamará la **representación canónica** si, además,  $\Phi(z)$  y  $\Theta(z)$  son primos entre si. Se supondrá siempre que  $\Phi(z)$  y  $\Theta(z)$  tienen raíces de  $|| > 1$ .

La varianza del nuevo ruido blanco se obtiene integrando los resultados para los procesos AR y MA, que se presentaron anteriormente.

3. La densidad espectral del proceso, es entonces:

$$f(\lambda) = \frac{\sigma^2}{2\pi} \frac{|\Theta(e^{i\lambda})|^2}{|\Phi(e^{i\lambda})|^2}, \quad \lambda \in [-\pi, \pi]$$

4.  $\Phi(F)X_t = \Theta(F)\varepsilon_t$  es la representación en avance del proceso. Puesto que  $\Phi(z)$  y  $\Theta(z)$  tienen raíces en  $|| > 1$ ,  $\varepsilon_t \perp X_s, s \geq t + 1$ .
5. Los coeficientes de autocorrelación de un ARMA se comportan asintóticamente como los coeficientes de autorrelación de la parte AR (convergen a velocidad exponencial hacia cero).
6. Los coeficientes de autocorrelación parcial de un ARMA se comportan como los coeficientes de autocorrelación parcial de la parte MA (convergen a velocidad exponencial hacia cero).
7. Es posible demostrar que todo proceso estacionario se puede aproximar por un MA. La introducción de procesos AR y ARMA, permite disminuir los retardos y por ende los parámetros a estimar.

**Ejemplo 2.7.** Considérese el modelo  $ARMA(1,1)$ ,  $X_t + 0,3X_{t-1} = u_t + 0,7u_{t-1}$ , del cual también se han simulado 200 observaciones. Se puede ver que el comportamiento de las FAC y FACP estimadas, no es similar a las funciones correspondientes de un  $AR(p)$  o un  $MA(q)$ .

**Figura 2.7:** Modelo  $ARMA(1,1)$

**Figura 2.8:** Comportamiento de  $\rho(l)$

**Figura 2.9:** Comportamiento de  $r(l)$

Estas condiciones permiten determinar, a priori, ciertas estimaciones de los valores de  $p$  y  $q$ , observando a partir de qué índice se anulan las estimaciones de  $\rho(l)$  y  $r(l)$ , según ciertos criterios estadísticos; sin embargo, estas estimaciones se pueden considerar cotas superiores, según se puede constatar en las Figuras 2.8 y 2.9, las que podrían sugerir, en inicio, un modelo  $ARMA(2,2)$ , cuando en realidad las observaciones provienen de un  $ARMA(1,1)$ . También es importante para no sobreparametrizar (asumir valores de  $p$  o  $q$  extremadamente altos), lo que es muy común en quienes están iniciando en la modelación de series temporales, observar con cuidado los comportamientos asintóticos de las FAC y FACP estimadas. Así, en el proceso autoregresivo precedente (ejemplo 2.2), si se trata de asumir un modelo ARMA, al mirar la FAC estimada se podría pensar en un valor de  $q$  demasiado alto, cuando en verdad este comportamiento está confirmando lo que muestran las FACP estimadas: que se trataría de un  $AR(2)$ .

Las estimaciones de los coeficientes  $a_j$ ,  $b_j$  y  $\sigma^2$  se tratan en el capítulo siguiente. La técnica más utilizada es la popularizada por Box y Jenkins; este método se encuentra implantado en algunos paquetes estadísticos (por ejemplo SPSS, STATGRAPHICS, Eviews).

**Observación 2.10.** No siempre un p.e.s.o. se puede representar como un ARMA (los ejemplos están fuera del alcance de este documento).

**Observación 2.11.** En resumen, se tiene que un proceso  $X_t$ :

- Corresponde a un AR, cuando las correlaciones estimadas decrecen rápidamente hacia cero. El orden de los retardos en el modelo viene dado por las correlaciones parciales estimadas que se encuentran fuera de las bandas de confianza; para fines prácticos, se ignoran aquellas que se encuentran muy cerca de estas bandas.
- Corresponde a una MA, cuando las correlaciones parciales decrecen rápidamente hacia cero. El orden de los retardos en el modelo viene dado por las correlaciones estimadas que se encuentran fuera de las bandas de confianza; para fines prácticos, se ignoran aquellas que se encuentran muy cerca de estas bandas.
- Corresponde a un ARMA, si se consideran los órdenes dados por los dos casos precedentes. En general, el modelo final tendrá órdenes más pequeños que los inicialmente considerados.

## 2.6 Procesos ARIMA (Autoregressive Integrates Moving Avarage)

Una buena parte de las series económicas son generadas por procesos no estacionarios; por esta razón, es útil considerar tales procesos, pero que sean fácilmente transformables en procesos estacionarios. Por ejemplo, ciertas series se vuelven estacionarias mediante “diferenciación”:

$$Y_t = X_t - X_{t-1}$$

Realizando una segunda diferenciación se tiene:

$$Z_t = Y_t - Y_{t-1}$$

Es decir:

$$Z_t = X_t - 2X_{t-1} + X_{t-2}$$

Un proceso que después de diferenciarlo  $d$  veces se transforma en un proceso  $ARMA(p, q)$  se dice  $ARIMA(p, d, q)$ . Si se denota  $\Delta = 1 - B$ , entonces el proceso  $(X_t)$  es un  $ARIMA(p, d, q)$ , si  $\Delta^d X_t$  es un  $ARMA(p, q)$ . Se tiene así, la siguiente definición:

**Definición 2.8.** Sea  $(X_t, t \in \mathbb{Z})$  de segundo orden. El proceso se dice un  $ARIMA(p, d, q)$  si se puede representar por:

$$\Phi(B)(1 - B)^d X_t = \Theta(B)u_t, \quad t \in \mathbb{Z} \quad (2.4)$$

donde,  $\Phi(z)$  y  $\Theta(z)$  tienen sus raíces fuera del círculo unitario complejo y  $(u_t)$  es un ruido blanco de varianza  $\sigma^2$ .

**Ejemplo 2.8.**  $(1 - 0,3B)(1 - B)X_t = u_t$  es un proceso  $ARIMA(1, 1, 0)$ .

A continuación, se presentan los “efectos” gráficos de la diferenciación para algunas tendencias polinomiales:

- **Tendencia Lineal:** Considérese la tendencia dada por la ecuación  $X_t = 20 + 4t$ ; una vez que se realiza la diferenciación se obtiene  $\Delta X_t = 4(\text{constante})$ . Así, gráficamente se tiene:

**Figura 2.10:** Efecto de la diferenciación en una tendencia lineal

- **Tendencia Cuadrática:** Considérese la tendencia cuadrática dada por la ecuación  $X_t = 20 + 4t + 2t^2$ ; una vez que se realiza la diferenciación se obtiene  $\Delta X_t = 2 + 4t$ . Entonces, todavía se tiene una tendencia lineal; sin embargo, al realizar la segunda diferenciación se llega a:  $\Delta^2 X_t = 4(\text{constante})$ . Gráficamente se observa:

**Figura 2.11:** Efecto de la diferenciación de una tendencia cuadrática

**Observación 2.12.** Se puede demostrar que con una diferenciación de orden  $d$  se anula una tendencia polinomial de grado  $d$ .

**Observación 2.13.** Existen series temporales con tendencias no polinomiales, en las cuales la diferenciación ya no produce el efecto visto en los casos anteriores; por ejemplo, en el caso de una tendencia exponencial, se tiene:

**Figura 2.12:** Efecto de la diferenciación de una tendencia exponencial

Como se puede ver, la diferenciación no elimina la tendencia de los datos. En este caso es necesario aplicar previamente una transformación de los datos; por ejemplo, una transformación logarítmica. En el ejemplo, se tiene:

**Figura 2.13:** Efecto de la transformación logarítmica en una tendencia exponencial

Entonces, luego de aplicar la transformación logarítmica se puede efectuar la diferenciación. Como se vio en los ejemplos precedentes, se requiere efectuar adicionalmente una diferenciación de orden 1.

**Observación 2.14.** 1. Si  $\phi(B) = \Phi(B)(1-B)^d = 1 - \phi_1 B - \dots - \phi_{p'} B^{p'}$ ,  $p' = p + d$ , la relación (2.4) se escribe como aquella de un  $ARMA(p', q)$ .

$$\phi(B)X_t = \Theta(B)u_t$$

pero  $\phi(z)$  admite como raíz múltiple de orden  $d$  a 1.

2. Se puede observar que (2.3) no define  $(X_t)$  completamente, por lo cual se hace necesario introducir un mecanismo para inicializar el proceso. Considérese, por ejemplo, el caso particular de una “marcha aleatoria”:

$$X_t - X_{t-1} = u_t$$

Para completar la definición de  $X_t$  se introduce un valor inicial que se denotará por  $X_{-s+1}$  ( $s \in \mathbb{N}$ ). Se puede escribir entonces:

$$X_t = X_{-s+1} + \sum_{i=0}^{t+s-2} u_{t-i} \quad t \geq -s+2$$

3. En general un proceso  $ARIMA(p, d, q)$  estará definido por  $p' = p + d$  valores iniciales:

$$X_{-s+1}, \dots, X_{-s+p'}$$

que se supondrán no aleatorios.

La relación (2.3) es válida para  $t > -s + p'$ , con  $u_t = 0, \forall t \leq -s + p'$ . Todos los  $u_t$  que intervienen en (2.3) serán no nulos si  $t \geq -s + p + q + 1$ .

4. La introducción de un mecanismo de inicialización, hace que el propósito inicial de que el proceso  $(X_t)$  diferenciado  $d$  veces sea  $ARMA(p, q)$ , se cumpla solo asintóticamente.



mente. Es decir, el proceso ARIMA

$$X_t = m_t + \sum_{i=0}^{t+s-p'-1} \psi_i u_{t-i} \quad \psi_0 = 1, \quad t > -s + p' + 1$$

es t.q.  $E(\Delta^d X_t - Z_t)^2 \rightarrow 0$ , donde  $Z_t$  es un  $ARMA(p, q)$  definido por:  $\Phi(B)Z_t = \Theta(B)u_t$ .

### 2.6.1 Representación MA de un ARIMA

**Teorema 2.5.** Se tiene que  $X_t = m_t + Y_t \forall t \geq -s + 1$ , donde:

1.  $m_t$  es la solución de  $\phi(B)X_t = 0$  con  $X_{-s+1}, \dots, X_{-s+p'}$  como valores iniciales (v.i.).

$$2. Y_t = \begin{cases} \sum_{i=0}^{t+s-p'-1} \psi_i u_{t-i} & t \geq -s + p' + 1 \\ 0 & -s + 1 < t < -s + p' + 1 \end{cases}$$

Los  $\psi_i$  se obtienen por división creciente de  $\Theta$  por  $\phi$  (hasta el orden  $t + s - p' - 1$ ).

### 2.6.2 Representación AR de un ARIMA:

**Teorema 2.6.** Se tiene:

$$X_t = - \sum_{i=1}^{t+s-p'-1} \pi_i X_{t-i} + \sum_{j=s-p'}^{s-1} \pi_{j,t}^* X_{-j} + u_t \quad \forall t > -s + p'$$

donde:

1. Los  $\pi_i$  se obtienen por división creciente de  $\phi$  por  $\Theta$  ( $\pi_0 = 1$ )
2. Los  $\pi_{j,t}^* \rightarrow 0$  cuando  $s \rightarrow \infty$

*Demostración.* Ver Anexo B.3

□

**Ejemplo 2.9.** Verifique si los siguientes modelos son estacionarios y/o invertibles. En caso de que no sean estacionarios, ¿puede transformarlos en estacionarios? ¿Cuál sería la varianza de la innovación?

1.  $(1 - 1,5B + 0,5B^2)(1 - B)X_t = (1 - 0,3B)u_t$ .

*Resolución.*

$$\begin{aligned} (1 - 1,5B + 0,5B^2)(1 - B)X_t &= (1 - 0,3B)u_t \\ \iff \left(1 - \frac{B}{2}\right)(1 - B)(1 - B)X_t &= (1 - 0,3B)u_t \\ \iff \left(1 - \frac{B}{2}\right)(1 - B)^2 &= (1 - 0,3B)\mu_t \\ \iff \varphi(B)X_t &= \Theta(B)u_t \end{aligned}$$

$\varphi(z)$  tiene una raíz unitaria de multiplicidad 2; por tanto,  $(X_t)$  no es estacionario.  $\Theta(z)$  tiene una raíz de  $|| > 1$ ; por tanto,  $(X_t)$  es invertible.

Para transformar el proceso en estacionario, se define un nuevo proceso  $Z_t = (1 - B)^2 X_t$ ; el nuevo proceso es estacionario, lineal (el polinomio AR asociado tiene como raíz 2) e invertible y su forma canónica se expresa por:

$$\left(1 - \frac{B}{2}\right) Z_t = (1 - 0,3B)u_t$$

La varianza de la innovación es  $V(u_t) = \sigma^2$  □

2.  $(1 - 2B + B^2)X_t = (1 + 1,1B)u_t$

*Resolución.*

$$(1 - 2B + B^2)X_t = (1 + 1,1B)u_t \Leftrightarrow (1 - B)^2 X_t = (1 + 1,1B)u_t \Leftrightarrow \varphi(B)X_t = \Theta(B)u_t$$

$\varphi(B) = (1 - B)^2$  tiene una raíz unitaria de multiplicidad 2; por tanto, el proceso no es estacionario

$\Theta(B) = (1 + 1,1B)$  tiene una raíz de  $|| < 1$ ; por tanto, el proceso no es invertible.

Para que el proceso sea estacionario se define  $Z_t = (1 - B)^2 X_t$ , es decir:

$$Z_t = (1 + 1,1B)u_t$$

Para que el proceso sea invertible, se invierte raíz  $z = \frac{1}{1,1} = 0,91$ .

La ecuación canónica es:

$$Z_t = (1 + 0,91B)\eta_t \quad (\eta_t \text{ es la innovación})$$

$$V(\eta_t) = \frac{\sigma^2}{|z|^2} = \frac{\sigma^2}{0,91^2} = \frac{\sigma^2}{0,828}$$

□

3.  $\left(1 + \frac{4}{3}B + \frac{4}{9}B^2\right) X_t = (1 - 0,4B)^2 u_t$

*Resolución.*

$$\left(1 + \frac{4}{3}B + \frac{4}{9}B^2\right) X_t = (1 - 0,4B)^2 u_t \Leftrightarrow \varphi(B)X_t = \Theta(B)u_t$$

$\varphi(B) = \left(1 + \frac{2}{3}B\right)\left(1 + \frac{2}{3}B\right)$  tiene raíces de  $|| < 1$ ; por tanto, el proceso es lineal y en consecuencia estacionario.

$\Theta(B) = (1 - 0,4B)^2$  tiene raíces de  $|| > 1$ , por tanto, el proceso es invertible y  $V(u_t) = \sigma^2$ . La ecuación canónica es la inicial. □

4.  $\left(1 - \frac{1}{4}B - \frac{3}{8}B^2\right) X_t = (1 + 1,1B)u_t$

*Solución.*

$$\begin{aligned} \left(1 - \frac{1}{4}B - \frac{3}{8}B^2\right) X_t = (1 + 1,1B)u_t &\Leftrightarrow \left(1 - \frac{3}{4}B\right) \left(1 + \frac{1}{2}B\right) X_t = (1 + 1,1B)u_t \\ &\Leftrightarrow \varphi(B)X_t = \Theta(B)u_t \end{aligned}$$

$\varphi(B)$  tiene raíces de  $|| > 1$ ; por tanto, el proceso es lineal y estacionario.

$\Theta(B)$  tiene una raíz de  $|| < 1$ ; por tanto, el proceso no es invertible.

Como  $(X_t)$  no es invertible, se invierte la raíz del polinomio media móvil. La ecuación canónica es:

$$\begin{aligned} \varphi(B)X_t &= (1 + 0,91B)\eta_t \\ V(\eta_t) &= \frac{\sigma^2}{0,91^2} = \frac{\sigma^2}{0,828} \end{aligned}$$

□

$$5. \left(1 - \frac{5}{3}B - \frac{2}{3}B^2\right) X_t = \left(1 - \frac{7}{10}B - \frac{3}{5}B^2\right) u_t$$

*Solución.*  $\varphi(B) = (1 - 2B)(1 + \frac{1}{3}B)$  tiene una raíz de  $|| < 1$  ( $z = \frac{1}{2}$ ); por tanto, el proceso es estacionario, pero no es lineal.

$\Theta(B) = (1 - \frac{6}{5}B) \left(1 + \frac{B}{2}\right) u_t$  tiene una raíz de  $|| < 1$  ( $z = \frac{5}{6}$ ); por tanto, no es invertible.

Invirtiéndolo las raíces correspondientes para que el proceso sea lineal e invertible, se tiene la siguiente ecuación ARMA canónica:

$$\begin{aligned} \left(1 - \frac{1}{2}B\right) \left(1 + \frac{1}{3}B\right) X_t &= \left(1 - \frac{5}{6}B\right) \left(1 + \frac{1}{2}B\right) u_t \\ V(\eta_t) &= \frac{\left|\frac{1}{2}\right|^2}{\left|\frac{5}{6}\right|^2} \sigma^2 = \frac{9}{25} \sigma^2 \end{aligned}$$

□

**Ejemplo 2.10.** Sea el proceso  $(X_t)_{t \in \mathbb{Z}}$ , definido por:

$$(1 - 4B)X_t = 3 + (1 - 0,5B)u_t$$

donde  $(u_t)$  r.b. de varianza 1.

1. Expresar  $X_t$  en función de los  $u_{t-i}$ ,  $i \in \mathbb{Z}$ .
2. Expresar el proceso en su forma canónica (y centrado). ¿Cuál es la expresión como proceso lineal?
3. Encuentre la covarianza entre  $X_t$  y  $u_{t-i}$ ,  $i \in \mathbb{Z}$ .
4. ¿Cuál es la varianza del error de predicción con horizonte 1?

*Solución.* 1. Se tiene que:

$$(1 - 4B)X_t = 3 + (1 - 0,5B)u_t \iff -4B(1 - \frac{1}{4}F)X_t = 3 + (1 - 0,5B)u_t$$

Por tanto:

$$\begin{aligned} X_t &= -\frac{1}{4}F \left( \frac{1}{1 - \frac{1}{4}F} \right) [3 + (1 - 0,5B)] u_t \\ &= -\frac{1}{4}F \sum_{i=0}^{\infty} \left( \frac{1}{4} \right)^i F^i [3 + (1 - 0,5B)] u_t \\ &= -\sum_{i=0}^{\infty} \left( \frac{1}{4} \right)^{i+1} F^{i+1} [3 + (1 - 0,5B)] u_t \\ &= -3 - \left[ \sum_{i=0}^{\infty} \left( \frac{1}{4} \right)^{i+1} F^{i+1} \right] u_t + \frac{1}{2} \left[ \sum_{i=0}^{\infty} \left( \frac{1}{4} \right)^{i+1} F^{i+1} \right] u_{t-1} \\ &= -3 - \sum_{i=1}^{\infty} \left( \frac{1}{4} \right)^i u_{t+i} + \left( \frac{1}{2} \right) \sum_{i=1}^{\infty} \left( \frac{1}{4} \right)^i u_{t+i-1} \\ &= -3 + \frac{1}{8}u_t + \sum_{j=1}^{\infty} \left( \left( \frac{1}{8} \right) \left( \frac{1}{4} \right)^j - \left( \frac{1}{4} \right)^j \right) u_{t+j} \\ &= -3 + \frac{1}{8}u_t - \frac{7}{8} \sum_{j=1}^{\infty} \left( \frac{1}{4} \right)^j u_{t+j} \end{aligned}$$

2. Se considera el proceso Sea  $(Y_t)$  en el que se invierte la raíz autoregresiva de  $|| < 1$  del proceso  $(X_t)$ , a la vez que se centra el nuevo proceso:

$$\begin{aligned} Y_t &= X_t - \frac{3}{1 - \frac{1}{4}} = X_t - 4 \\ v_t &= u_t = \left| \frac{1}{4} \right| u_t = \frac{1}{4}u_t \end{aligned}$$

Se define un nuevo proceso por:

$$\varphi(B)Y_t = \Theta(B)v_t \iff Y_t = \frac{\Theta(B)}{\varphi(B)}v_t = \frac{(1 - 0,5B)}{(1 - 0,25B)}v_t$$

Donde  $(v_t)$  es un ruido blanco. Esta es la forma canónica del proceso.

Considérese ahora la siguiente igualdad de operadores:

- $1 - 0,5B = (1 - 0,25B)(\psi_0 + \psi_1B + \psi_1B^2 + \dots)$
- $1 - 0,25 = \psi_0 + B(\psi_1 - 0,25\psi_0) + B^2(\psi_2 - 0,25\psi_1) + B^2(\psi_3 - 0,25\psi_2) + \dots$
- $\psi_0 = 1$
- $\psi_1 - 0,25\psi_0 = -0,5 \Rightarrow \psi_1 = -0,25$
- $\psi_t = 0,25\psi_{t-1} \implies \psi_t = (-0,25)^t$
- $Y_t = v_t + \sum_{j=2}^{\infty} \psi_j B^j v_t = v_t - \sum_{j=1}^{\infty} (0,25)^j v_{t-j}$

esta es la expresión como proceso lineal

3. Ahora se remplacea  $X_t$  por  $Y_t + 4$ , de donde se obtiene:

$$X_t = 4 + v_t - 0,5 \sum_{j=1}^{\infty} (0,25)^{j-1} v_{t-j} = 4 + \frac{1}{4} u_t - \frac{1}{8} \sum_{j=1}^{\infty} (0,25)^{j-1} u_{t-j}$$

Además:

$$\begin{aligned} \text{Cov}(X_t, u_{t-i}) &= \frac{1}{4} \text{Cov}(u_t, u_{t-i}) - \frac{1}{8} \sum_{j=1}^{\infty} 0,25^{j-1} \text{Cov}(u_{t-j}, u_{t-i}) \\ &= \begin{cases} \frac{1}{4} & \text{si } i = 0 \\ 0 & \text{si } i < 1 \\ \frac{1}{8} \cdot 0,25^{i-1} & \text{si } i \geq 1 \end{cases} \end{aligned}$$

4.

$$V(X_{t+1} - \hat{X}_t(1)) = V(v_t) = \frac{1}{16} V(u_t) = \frac{1}{16}$$

□

**Ejemplo 2.11.** Sea

$$f(\lambda) = \frac{\frac{13}{12} + \cos \lambda}{\frac{41}{40} + \cos \lambda} \cdot \frac{1}{2\pi} \quad \lambda \in [-\pi, \pi]$$

1. Probar que existe un proceso estacionario ARMA  $X = (X_t)_{t \in \mathbb{Z}}$  tal que la medida espectral  $\mu$  de  $X$  tenga por densidad  $f$  (con respecto a la medida de Lebesgue):

$$f(\lambda) = \frac{\frac{13}{12} + \cos \lambda}{\frac{41}{40} + \cos \lambda} \cdot \frac{1}{2\pi} = \frac{13 + 12 \cos \lambda}{41 + 40 \cos \lambda} \cdot \frac{40}{12} \cdot \frac{1}{2\pi}$$

2. Escribir la ecuación ARMA canónica (y centrado). ¿Cuál es la expresión como proceso lineal?

3. Representar  $u_t$  en términos de los  $X_j$ ,  $j \leq t$ .

*Resolución.* 1. Sea el modelo ARMA(1, 1). Su densidad espectral se expresa por:

$$f(\lambda) = \frac{|1 - \theta e^{i\lambda}|^2}{|1 - \varphi e^{i\lambda}|^2} \cdot \frac{\sigma^2}{2\pi}$$

$$\begin{aligned} |1 - \theta e^{i\lambda}|^2 &= (1 - \theta \cos \lambda - \theta i \sin \lambda)(1 - \theta \cos \lambda + \theta i \sin \lambda) \\ &= 1 - \theta \cos \lambda + \theta i \sin \lambda - \theta \cos \lambda - \theta^2 i \cos \lambda - \theta^2 i \cos \lambda \sin \lambda \\ &\quad - \theta i \sin \lambda + \theta^2 i \cos \lambda \sin \lambda + \theta^2 \sin^2 \lambda \\ &= 1 - 2\theta \cos \lambda + \theta^2 \end{aligned}$$

$$\therefore |1 - \theta e^{i\lambda}|^2 = 1 - 2\theta \cos \lambda + \theta^2$$

$$f(\lambda) = \frac{1 - 2\theta \cos \lambda + \theta^2}{1 - 2\varphi \cos \lambda + \varphi^2} \cdot \frac{\sigma^2}{2\pi} = \frac{-\frac{(1+\theta^2)}{2\theta} + \cos \lambda}{-\frac{(1+\varphi^2)}{2\varphi} + \cos \lambda} \cdot \frac{2\theta}{2\varphi} \cdot \frac{\sigma^2}{2\pi}$$

Del numerador de la expresión precedente se tiene que:

$$-\frac{(1+\theta^2)}{2\theta} = \frac{13}{6} \iff -1 - \theta^2 = \frac{13}{6}\theta \iff \theta^2 + \frac{13}{6}\theta + 1 = 0$$

$$\therefore \theta = -\frac{2}{3} \quad \text{o} \quad \theta = -\frac{3}{2}$$

Se elige  $\theta = -\frac{2}{3}$

De otro lado, analizando el denominador de  $f(\lambda)$  se tiene:

$$-\frac{(1+\varphi^2)}{2\varphi} = \frac{41}{40} \iff \varphi^2 + \frac{41}{20}\varphi + 1 = 0$$

$$\varphi = -\frac{5}{4} \quad \text{o} \quad \varphi = -\frac{4}{5}$$

Se elige  $\varphi = -4/5$  por ser de valor absoluto menor a 1

Además:

$$\sigma^2 = \frac{2\varphi}{2\varphi} = \frac{\theta}{\varphi}$$

2. Aunque en inicio se podrían considerar cuatro modelos posibles, se retiene aquel que toma  $\theta = -\frac{3}{2}$  y  $\varphi = -\frac{4}{5}$ . La ecuación ARMA canónica viene dada por:

$$X_t - \frac{4}{5}X_{t-1} = u_t - \frac{2}{3}u_{t-1} \iff \left(1 - \frac{4}{5}B\right) X_t = \left(1 - \frac{2}{3}B\right) u_t$$

La varianza de la innovación ( $u_t$ ) es  $\sigma^2 = \frac{5}{6}$ .

3. Representar  $u_t$  en términos de los  $X_j, j \leq t$

$$\left(1 - \frac{4}{5}B\right) X_t = \left(1 - \frac{2}{3}B\right) u_t \iff u_t = \frac{\left(1 - \frac{4}{5}B\right)}{\left(1 - \frac{2}{3}B\right)} X_t$$

$$\iff u_t = \left(1 - \frac{4}{5}B\right) \sum_{j=0}^{\infty} \left(\frac{2}{3}\right)^j X_{t-j} = \sum_{j=0}^{\infty} \left(\frac{2}{3}\right)^j X_{t-j} - \frac{4}{5} \sum_{j=0}^{\infty} \left(\frac{2}{3}\right)^j X_{t-1-j}$$

$$= X_t + \sum_{j=1}^{\infty} \left( \left(\frac{2}{3}\right)^j - \frac{3}{2} * \frac{4}{5} \left(\frac{2}{3}\right)^j \right) X_{t-j} = X_t - \frac{1}{5} \sum_{j=1}^{\infty} \left(\frac{2}{3}\right)^j X_{t-j}$$

□

**Ejemplo 2.12.** Sea  $(X_n)$  un proceso estacionario ARMA, que satisface la ecuación:

$$X_n - (4/9)X_{n-1} + (1/27)X_{n-2} = w_n - (19/2)w_{n-1} + (9/2)w_{n-2} \quad n \in \mathbb{Z}$$

donde  $(w_n)$  es un ruido blanco de varianza 1.

1. Encontrar la ecuación ARMA canónica satisfecha por  $X$ .
2. Si  $(v_n)$  es la innovación introducida en la ecuación ARMA canónica, expresar  $(v_n)$  con la ayuda de los  $(w_p, p \in \mathbb{Z})$ . Calcular la varianza de la innovación.
3. ¿Puede comparar los pasados de  $(v_n)$  y  $(w_n)$ ? ¿los futuros?.

Resolución. 1.

$$X_n - \frac{4}{9}X_{n-1} + \frac{1}{27}X_{n-2} = w_n - \frac{19}{2}w_{n-1} + \frac{9}{2}w_{n-2} \quad (2.5)$$

$$\text{Sea } \varphi(z) = 1 - \frac{4}{9}z + \frac{1}{27}z^2$$

$$\begin{aligned} \varphi(z) = 0 &\iff 27 - 12z + z^2 = 0 \\ &\iff z^2 - 12z + 27 = 0 \\ &\iff (z - 9)(z - 3) = 0 \\ &\iff z_1 = 9 \quad \text{o} \quad z_2 = 3 \end{aligned}$$

$$\text{luego, } \varphi(z) = \left(1 - \frac{1}{9}z\right) \left(1 - \frac{1}{3}z\right) \text{ con } |z_1| > 1 \text{ y } |z_2| > 1$$

$$\text{De otro lado: } \Theta(z) = 1 - \frac{19}{2}z + \frac{9}{2}z^2$$

$$\begin{aligned} \Theta(z) = 0 &\iff 2 - 19z + 9z^2 = 0 \\ &\iff 9z^2 - 19z + 2 = 0 \\ &\iff (z - 2)(9z - 1) = 0 \\ &\iff z_1 = 2 \quad \text{o} \quad z_2 = \frac{1}{9} \end{aligned}$$

$$\text{luego, } \Theta(z) = \left(1 - \frac{1}{2}z\right) (1 - 9z) \text{ con } |z_1| > 1 \text{ y } |z_2| < 1$$

La ecuación (2.5) se puede expresar:

$$\varphi(B)X_n = \Theta(B)w_n$$

$$\left(1 - \frac{1}{9}B\right) \left(1 - \frac{1}{3}B\right) X_n = \left(1 - \frac{1}{2}B\right) (1 - 9B) w_n \quad (2.6)$$

Como se tiene una raíz de valor absoluto menor que uno, tenemos que invertirla, de tal manera que cambiará el ruido blanco; ahora tendremos la innovación  $v_n$  y las raíces fuera del círculo unidad, así:

$$\left(1 - \frac{1}{9}B\right) \left(1 - \frac{1}{3}B\right) X_n = \left(1 - \frac{1}{2}B\right) \left(1 - \frac{1}{9}B\right) v_n$$

y la ecuación ARMA canónica se expresa por:

$$\left(1 - \frac{1}{3}B\right) X_n = \left(1 - \frac{1}{2}B\right) v_n$$

2. Puesto que:

$$v_n = \frac{\left(1 - \frac{1}{3}B\right)}{\left(1 - \frac{1}{2}B\right)} X_n$$

Se reemplaza  $X_n$  en función del ruido blanco  $w_n$  (ecuación (2.6)):

$$v_n = \frac{\left(1 - \frac{1}{3}B\right)}{\left(1 - \frac{1}{2}B\right)} \cdot \frac{\left(1 - \frac{1}{2}B\right)(1 - 9B)}{\left(1 - \frac{1}{9}B\right)\left(1 - \frac{1}{3}B\right)} w_n = \frac{(1 - 9B)}{\left(1 - \frac{1}{9}B\right)} w_n \quad n \in \mathbb{Z}$$

Para calcular la varianza de la innovación  $v_n$  se utiliza la función de densidad espectral de  $(v_n)$ :

$$f_v(\lambda) = \frac{|1 - 9e^{i\lambda}|^2}{\left|1 - \frac{1}{9}e^{i\lambda}\right|^2} \cdot \frac{\sigma^2}{2\pi} = \frac{81 \left|\frac{1}{9} - e^{i\lambda}\right|^2}{\left|1 - \frac{1}{9}e^{i\lambda}\right|^2} \cdot \frac{\sigma^2}{2\pi}$$

Multiplicando el numerador de esta expresión por  $|e^{-i\lambda}|^2 = 1$  se tiene:

$$f(v_n) = \frac{81 \left|\frac{1}{9}e^{-i\lambda} - 1\right|^2}{\left|1 - \frac{1}{9}e^{i\lambda}\right|^2} \frac{\sigma^2}{2\pi} = \frac{81\sigma^2}{2\pi}$$

Puesto que por hipótesis  $\sigma^2 = 1$ , tenemos que  $V(v_n) = 81$

3. Dado que:

$$v_n = \frac{(1 - 9B)}{\left(1 - \frac{1}{9}B\right)} w_n = \frac{\Theta_1(B)}{\varphi_1(B)} w_n$$

- $\varphi_1(z)$  tiene su raíz de valor absoluto mayor que 1; por tanto, la inversión de  $\varphi_1(B)$  nos da una serie en términos de potencias  $B \left( \sum_{j=0}^{\infty} \left(\frac{1}{9}\right)^j B^j \right)$ ; es decir,  $v_n$  se puede expresar en términos del presente y del pasado de los  $w_n$ . Por lo tanto  $v_n \in ev \{w_j | j \leq n\}$ .

Sin embargo, puesto que  $\Theta_1(z)$  tiene una raíz de valor absoluto menor que 1, al invertir  $\Theta_1(B)$  se obtiene una serie con potencias de  $F$ ; es decir, en términos del futuro de los  $v_j$  (e.d.  $v_j, j \geq n$ ); por tanto  $w_n$  no se puede expresar en términos del pasado de los  $(v_j)$ .

- De manera análoga, se puede observar que  $w_n$  está contenido en el futuro de los  $(v_j)$ , pero  $v_n$  no lo está en el futuro de los  $(w_j)$ .

□

**Ejemplo 2.13.** Sea  $X$  p.e. centrado, que satisface la ecuación ARMA:

$$9X_n - X_{n-2} = W_n + 3W_{n-1}$$

donde  $(X_n)_{es}$  un r.b. de varianza  $\sigma$

1. Encontrar la ecuación ARMA canónica satisfecha por  $X$  y la varianza de la innovación.



2. Calcular explícitamente las covarianzas  $\gamma_h = E(X_{n+h}X_n)$  del proceso  $X$  (se puede comenzar buscando la recurrencia satisfecha por  $\gamma_h$ )

*Resolución.* 1.  $9X_n - X_{n-2} = W_n + 3W_{n-1}$

Sea  $\Theta(z) = 1 + 3z \Rightarrow 1 + 3z = 0 \Leftrightarrow z = -1/3$  ( $|z| < 1$ )

Sea  $\phi(z) = 9 - z^2$  y  $\phi(z) = 0 \Rightarrow (3+z)(3-z) = 0 \Leftrightarrow z = 3$  o  $z = -3$

Puesto que:

$$9(X_n - X_{n-2}/9) = W_n + 3W_{n-1}$$

Se puede escribir a partir de un nuevo ruido blanco  $u_n$

$$9(X_n - X_{n-2}/9) = u_n + u_{n-1}/3 \Leftrightarrow 9(1 - B^2/9)X_n = (1 + B/3)u_n$$

$$\Leftrightarrow 9(1 - B/3)(1 + B/3)X_n = (1 + B/3)u_n$$

De donde se obtiene

$$9(1 - B/3)X_n = u_n \Leftrightarrow (X_n - X_{n-1}/3) = u_n/9$$

$$\Leftrightarrow X_n - (X_{n-1}/3) = v_n$$

(Ecuación canónica) donde  $(v_n)$  es un nuevo ruido blanco

La relación entre la innovación y el ruido blanco original es:

$$v_n = u_n/9 = \frac{W_n}{9|-1/3|} \Rightarrow v_n = W_n/3$$

Por lo tanto la varianza de  $v_n$  es:

$$V(v_n) = \frac{V(W_n)}{9} = \frac{\sigma^2}{9}$$

2. Puesto que  $X_n = v_n + X_{n-1}/3$ , se puede considerar:

$$\gamma_h = E(X_n X_{n-h}) = E(X_{n-1} X_{n-h})/3 + E(v_n X_{n-h})$$

$$\gamma_h = \frac{\gamma_{h-1}}{3} + E(v_n X_{n-h}) = \frac{1}{3}\gamma_{h-1} + 0 \text{ si } h \geq 1$$

$$\gamma_h = \frac{\gamma_{h-1}}{3} = \frac{1}{3}\left(\frac{\gamma_{h-2}}{3}\right) = \dots = (1/3)^h \gamma_0 \quad \forall h \geq 1$$

Utilizando las ecuaciones de Yule – Walker se obtiene:

$$\sum_{j=0}^k \pi_j \gamma_j = \sigma_v^2 \Leftrightarrow \gamma_0 - \frac{1}{3}\gamma_1 = \frac{\sigma^2}{9} \Leftrightarrow \gamma_0 - \left(\frac{1}{3}\right)\gamma_0 = \frac{\sigma^2}{9} \Leftrightarrow \frac{8}{9}\gamma_0 = \frac{\sigma^2}{9} \Leftrightarrow \gamma_0 = \frac{\sigma^2}{8}$$

En consecuencia:

$$\gamma_h = \frac{1}{8} \left(\frac{1}{3}\right)^h \sigma^2 \quad h \geq 1$$

□

## 2.7 Ejercicios Propuestos

### Modelos AR

1. Utilice la densidad espectral para probar que el proceso  $(Z_t, t \in \mathbb{Z})$  definido por:

$$\Phi(B) Z_t = \xi_t \quad \text{con } \xi_t = \prod_{j=r+1}^p |z_j| u_t$$

tiene los mismos elementos de segundo orden  $(\gamma_h, \rho(h), r(h), \text{densidad espectral})$  que  $(X_t, t \in \mathbb{Z})$ .

2. Sea el proceso autoregresivo de orden  $p$ , definido por  $\Phi(B) X_t = u_t$ . Probar que el proceso  $(\varepsilon_t, t \in \mathbb{Z})$  definido por:

$$\varepsilon_t = \Phi(F) X_t$$

es un ruido blanco de igual varianza que  $u_t$ . Además  $\varepsilon_t \perp X_s, s \geq t+1$ . Esta representación se dice la representación en avance de  $(X_t)$ .

### Modelos MA

3. El proceso  $(Z_t)_{t \in \mathbb{Z}}$  definido por:  $Z_t = \Theta(B) \xi_t$  con  $\xi_t = \frac{u_t}{\prod_{j=r'+1}^q |z_j|}$  donde  $z_j$  raíz de  $|\cdot| < 1$  tiene los mismos elementos de segundo orden que  $(X_t)$ .
4. El proceso  $(\varepsilon_t)_{t \in \mathbb{Z}}$  definido por  $X_t = \Theta(F) \varepsilon_t$  es r.b. de varianza  $\sigma^2$ . Se dice la representación en avance de  $(X_t)$ . Además  $\varepsilon_t$  no está correlacionada con  $(X_s, s \geq t+1)$ .

### Ejercicios Adicionales

5. ¿Cuáles de las siguientes funciones son funciones de autocovarianza? En caso afirmativo, dar el proceso asociado.

(a)  $\gamma(h) = 1 + |h|$

(b)  $\gamma(h) = \begin{cases} \leq 1 & \text{si } h = 0 \\ -\frac{1}{2} & \text{si } |h| = 1 \\ 0 & \text{caso contrario} \end{cases}$

(c)  $\gamma(h) = 1 + \frac{1}{4} \sin 4h$

6. Sea el proceso  $ARMA(1,1)$ :  $X_t = 0,5X_{t-1} + u_t - 1,5u_{t-1}$

(a) ¿El proceso es lineal?

(b) ¿El proceso es invertible?

(c) ¿El proceso es estacionario?

(d) Dar la ecuación canónica para este proceso. ¿Qué relación existe entre la innovación y los  $u_t$ ?

7. Considérense las siguientes estimaciones de un ruido blanco:

0,01	0,22	0,04	-0,01	-0,01	-0,06	0,05	0,25	0,20	0,06
-0,06	-0,14	0,00	0,19	0,22	0,14	0,20	0,36	0,37	0,14
-0,14	-0,26	-0,12	0,01	0,26	0,21	-0,04	-0,16	-0,35	-0,45

?Existe autocorrelación entre estas observaciones?



## Capítulo 3

# Estimación, Verificación y Predicción en un Modelo ARIMA por el Método de Box–Jenkins

### 3.1 Metodología Box–Jenkins

El procedimiento para identificar, estimar, verificar y predecir a partir de un modelo ARIMA compatible con los datos, se hace de una manera secuencial, que se puede describir (en forma resumida) de la siguiente forma:

1. Se comienza por buscar, a partir de las observaciones, los valores plausibles para  $(p, d, q)$ . Esta etapa constituye la fase de *identificación a priori* del modelo. Se requieren muchos cálculos y, en general, no se obtiene una sola tripleta de valores  $(p, d, q)$ .

Esto se explica porque un proceso estacionario puede aproximarse, tanto como se quiera, por un proceso  $ARMA(p, q)$ , donde el valor de  $p$  puede ser cualquiera.

2. Para cada tripleta de valores de  $(p, d, q)$  obtenida, se procede a la fase de *estimación de los coeficientes* de los polinomios autoregresivo y media móvil. Para fines prácticos, se los puede obtener con alguno de los paquetes estadísticos (EViews, Statgraphics, R, etc.).

3. Al final de la etapa 2, se dispone de varios modelos estimados. Estos pasan entonces por una fase de *pruebas estadísticas* (pruebas de significación para los coeficientes, pruebas concernientes a las no correlaciones de los ruidos blancos, etc.), para verificar los resultados obtenidos con las hipótesis hechas.

Esta es la fase de *verificación* y puede resultar que todos los modelos estimados sean rechazados. Bajo condiciones de normalidad de los residuos (o independencia), las pruebas para los coeficientes son del tipo ‘*t*-student’ usuales.

4. Puede darse el caso de que varios modelos pasen la fase de verificación. En este caso se escoge al modelo que tenga el mayor poder predictivo o la mayor cantidad de información (se explicarán posteriormente estos conceptos).

5. Finalmente, una vez escogido el mejor modelo se procede a realizar la predicción para un período posterior a la de los datos con los que se realizó la modelación.

Las etapas 3 y 4 constituyen la fase de *identificación a posteriori*; es decir, después de la estimación. Todas las etapas se explican en este capítulo tanto la parte teórica como la práctica.

## 3.2 Identificación *A Priori*

### 3.2.1 Elección de $d$

Si el proceso  $(X_t)_{t \in \mathbb{Z}}$ , generado por las observaciones, es un ARIMA con  $d > 0$ , no es estacionario; luego, su función de autocorrelación depende de  $l$  y  $t$ , así:

$$\rho(l, t) = \frac{\text{Cov}(X_t, X_{t+l})}{\sqrt{V(X_t)} \cdot \sqrt{V(X_{t+l})}}$$

Se puede demostrar, además, que la función  $\rho(l, t)$ , tiende a 1 cuando  $t$  tiende a  $+\infty$ . Si la función  $\hat{\rho}(l)$  queda relativamente próxima a uno para un número importante de valores de  $l$ , se puede entonces pensar que es necesario diferenciar la serie para volverla estacionaria. En la práctica, sin embargo, el criterio de proximidad a 1 de los primeros valores de la función  $\hat{\rho}(l)$  se reemplaza por el de la proximidad entre sí de los primeros valores de esta función, aun cuando el valor  $\hat{\rho}(l)$  sea muy diferente de uno. A la serie diferenciada una vez se le aplica nuevamente este criterio para ver si es conveniente diferenciarla una segunda vez, y así sucesivamente. Esto se muestra a continuación:

**Ejemplo 3.1.** Para este ejemplo se ha utilizado el programa Statgraphics Centurion XV. Considérense los datos de variación de un índice bursátil (ver Anexo C.2), cuyo gráfico se muestra a continuación:

**Figura 3.1:** Serie de un Índice Bursátil (SIB)

Al observar las autocorrelaciones estimadas del modelo original (Figura 3.2) se constata que se trata de un modelo donde es necesario realizar una diferenciación para volver a la serie estacionaria (Figura 3.3); esto se podría deducir también al mirar la forma lineal creciente de la tendencia de la serie (Figura 3.1).

**Figura 3.2:** Función de autocorrelación de la SIB

**Figura 3.3:** Serie estacionaria obtenida a partir de la SIB

Para las series económicas los valores de  $d$  que se obtienen, en general, son 0, 1 o 2; los valores superiores son muy raros. En el caso de la serie en consideración, no se requiere una segunda diferenciación no estacional.

**Observación 3.1.** El término utilizado para referirse a este tipo de diferenciación es el de: *diferenciación no estacional*; la *diferencia estacional* se explicará más adelante.

La consecuencia de aplicar en un modelo un número de diferencias mayor que el necesario es que el proceso final sea no invertible y con varianza mayor que si se hubiera aplicado un número de diferencias correcto.

### 3.2.2 Elección de $p$ y $q$

Para la elección de  $p$  y  $q$  se analiza esencialmente la forma de la función de autocorrelación estimada  $\hat{\rho}(l)$  y de la función de autocorrelación parcial estimada  $\hat{r}(l)$  de la serie diferenciada  $d$  veces. Si la función  $\hat{\rho}(l)$  se anula estadísticamente para  $l > q$ , se puede pensar que se trata de un  $MA(q)$ ; y si la función  $\hat{r}(l)$  se anula estadísticamente para valores de  $l > p$  se puede admitir que se trata de un  $AR(p)$ . Así, se obtienen valores de  $p$  y  $q$  (que pueden considerarse máximos) para el modelo; utilizando un paquete estadístico (en este capítulo se ha elegido Statgraphics) se puede graficar la función  $\hat{\rho}(l)$  y ver a partir de qué orden  $l$  los valores de la función quedan dentro de la banda. Similarmente, se procede con la función de autocorrelación parcial.

En cuanto a la obtención de la banda de confianza para las dos funciones estimadas, para juzgar si los valores de  $\hat{\rho}(l)$  o  $\hat{r}(l)$  son significativamente diferentes de cero es bueno tener una idea de su desviación estándar. Para un  $MA(q)$ , se tiene:

$$\forall l > q, \quad V[\hat{\rho}(l)] \approx \frac{1}{T} \left[ 1 + 2 \sum_{i=1}^{l-1} \rho^2(i) \right]$$

Se puede probar que para identificar el orden  $q$  de un proceso media móvil, se debe representar la sucesión de autocorrelaciones estimadas  $\hat{\rho}(l)$  y mirar a partir de qué valor  $q$  cada uno de los  $\hat{\rho}(l)$  se queda en el intervalo de extremos:

$$\pm \left( \frac{1,96}{\sqrt{T}} \right) \left[ 1 + 2 \left[ \rho^2(1) + \dots + \rho^2(l-1) \right]^{1/2} \right]$$

Los valores teóricos  $\rho(j)$  deben reemplazarse por sus estimaciones. Se asume también un nivel de significación del 5 % en la prueba de hipótesis respectiva.

Igualmente, se tiene que  $\sqrt{\frac{1}{T}}$  es la desviación estándar aproximada de  $\hat{r}(l)$  para  $l > p$ , si el proceso es un  $AR(p)$ . Para identificar el orden  $p$  de un proceso autoregresivo, asumiendo un nivel de significación del 5 %, se puede representar la sucesión de autocorrelaciones parciales estimadas  $\hat{r}(l)$  y mirar a partir de qué valor  $p$  cada una se queda en la banda dada por:

$$\left[ -\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}} \right]$$

**Ejemplo 3.2.** Volviendo al ejemplo 3.1, obsérvense los gráficos correspondientes a las autocorrelaciones estimadas (ver Figura 3.4), donde se aprecia que a partir de  $q = 2$  la función queda dentro de la banda de confianza, lo que en primera instancia hace suponer que se trata de un modelo  $MA(2)$ .

**Figura 3.4:** FAC estimada del modelo  $ARIMA(0, 1, 0)$ , para la SIB

Ahora bien, si además se revisa la función de autocorrelación parcial estimada (Figura 3.5), donde se observa un decrecimiento exponencial de la serie hacia cero, parece confirmarse el hecho de que la serie original se debe modelar por un  $MA(1)$  o un  $MA(2)$ .

En este punto es necesario realizar un análisis más profundo. En principio, dado que  $\hat{r}(6)$  es estadísticamente significativo, se podría pensar en que el Modelo más adecuado sería un  $AR(6)$ ; es decir, globalmente se consideraría un  $ARMA(6, 2)$ , lo que realmente llevaría a una sobreparametrización. En general, se aconseja que  $p + d + q \leq 5$ .

**Figura 3.5:** FACP estimada del modelo  $ARIMA(0, 1, 0)$ , para la SIB en primera diferencia

### 3.3 Estimación de un Modelo ARIMA

Considérese el siguiente modelo:

$$\Phi(B)\Delta^d X_t = \theta_0 + \Theta(B)u_t$$

donde  $\Phi(z)$  y  $\Theta(z)$  son primos entre si y tienen raíces de módulo mayor a 1 y  $(u_t)$  ruido blanco de varianza  $\sigma^2$ ; se tiene entonces que  $W_t = \Delta^d X_t$  es asintóticamente un  $ARMA(p, q)$ :

$$\Phi(B)W_t = \theta_0 + \Theta(B)u_t$$

Supóngase que se conocen las observaciones  $x_1, x_2, \dots, x_T$  del proceso  $(X_t)$ . Se pueden calcular las diferencias de orden  $d$ :

$$\begin{aligned} w_{d+1} &= \Delta^d x_{d+1} \\ w_{d+2} &= \Delta^d x_{d+2} \\ &\vdots \\ w_T &= \Delta^d x_T \end{aligned}$$

Así, se tienen  $N = T - d$  observaciones, que se denotarán por  $z_1, z_2, \dots, z_N$ , para estimar los parámetros del modelo:

$$\Phi(B)W_t = \theta_0 + \Theta(B)u_t$$

Se puede suprimir  $\theta_0$ , reemplazando  $W_t$  por  $\hat{W}_t = W_t - \mu$ , con  $\mu = \theta_0 / (1 - \varphi_1 - \dots - \varphi_p)$ . En lo que sigue se supondrá que  $\theta_0 = 0$ .

Se supondrá en esta sección que los ruidos blancos  $u_t$  son normales (en realidad solo se requiere independencia). Para calcular la función de verosimilitud  $L(\varphi_1, \dots, \varphi_p, \theta_1, \dots, \theta_q, \sigma^2) = L(\varphi, \theta, \sigma^2)$  se requieren conocer  $p$  valores iniciales  $z_0, \dots, z_{1-p}$  y  $q$  valores iniciales  $u_0, \dots, u_{1-q}$ , que se expresarán a través de los vectores  $z'_* = (z_0, \dots, z_{1-p})$  y  $u'_* = (u_0, \dots, u_{1-q})$ ; existen dos formas alternativas de obtenerlos, las que se describen a continuación.



### 3.3.1 Procedimiento condicional

En este caso se va a suponer que los elementos de  $z_*$  y  $u_*$  son iguales a sus esperanzas; es decir  $E u_t = 0$  y  $E z_t = 0$ . Para el caso en que  $\varphi(z)$  tenga raíces cercanas a la unidad o la serie sea estacional, una mejor solución consiste en utilizar la fórmula:

$$u_t = z_t - \varphi_1 z_{t-1} - \cdots - \varphi_p z_{t-p} + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q} \quad t = p+1, \dots, N$$

tomando  $u_p, u_{p-1}, \dots$  nulos, pero considerando valores de  $z_t$  realmente observados.

La función de verosimilitud en el caso condicional se expresa por:

$$\begin{aligned} L &= L(\varphi, \theta, \sigma^2 | z, z_*, u_*) \\ &= (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^N u_t^2\right) \\ &= (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{t=1}^N (z_t - \varphi_1 z_{t-1} - \cdots - \varphi_p z_{t-p} + \theta_1 u_{t-1} + \cdots + \theta_q u_{t-q})^2\right] \end{aligned}$$

tomando logaritmos se tiene que:

$$\ell = \ln L = cte - \frac{N}{2} \ln \sigma^2 - \frac{S_\bullet(\varphi, \theta | z, z_*, u_*)}{2\sigma^2}$$

se denotará:

$$S_\bullet(\varphi, \theta) = S(\varphi, \theta | z, z_*, u_*) = \sum_{t=1}^N u_t^2(\varphi, \theta, \sigma^2 | z, z_*, u_*)$$

llamada la suma cuadrática condicional. La expresión “cte” representa de manera general a un valor constante.

Derivando  $\ell$  con respecto a  $\sigma^2$  e igualando a cero se tiene:

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{N}{2\sigma^2} + \frac{S_*}{2\sigma^4} = 0 \iff \sigma^2 = \frac{S_*(\varphi, \theta | z, z_*, u_*)}{N} = \frac{S_*}{N}$$

Si se reemplaza esta expresión de  $\sigma^2$  en  $\ell$ , se obtiene:

$$\bar{\ell} = cte - \frac{N}{2} \ln \frac{S_*}{N} - \frac{N}{2} = cte + \frac{N}{2} \ln N - \frac{N}{2} \ln S_* - \frac{N}{2} = cte - \frac{N}{2} \ln S_*$$

Así, maximizar  $\bar{\ell}$  equivale a minimizar  $S_*$ ; esto es, a obtener una estimación mínimo cuadrática.

**Ejemplo 3.3.** Considérese una serie temporal que responde a un modelo  $ARIMA(0, d, 1)$ :

$$z_t = u_t - \theta u_{t-1} = (1 - \theta B) u_t$$

con

$$w_t = \Delta^d X_t, \quad t = d+1, \dots, T \quad y \quad z_t = w_{t+d}, \quad t = 1, \dots, N$$

En este caso:

$$S_*(\theta | z, u_0) = \sum_{t=1}^N u_t^2 = \sum_{t=1}^N (z_t + \theta u_{t-1}(\theta))^2$$

Si se considera entonces que  $u_0 = 0$ , se obtiene:

$$\begin{aligned} u_1(\theta) &= z_1 + \theta u_0(\theta) = z_1 + \theta u_0 \\ u_2(\theta) &= z_2 + \theta u_1(\theta) = z_2 + \theta(z_1 + \theta u_0) = z_2 + \theta z_1 + \theta^2 u_0 \\ u_3(\theta) &= z_3 + \theta u_2(\theta) = z_3 + \theta(z_2 + \theta z_1) = z_3 + \theta z_2 + \theta^2 z_1 + \theta^3 u_0 \\ &\vdots \\ u_t(\theta) &= z_t + \theta z_{t-1} + \cdots + \theta^{t-1} z_1 + \theta^t u_0 \\ &= z_t + \theta z_{t-1} + \cdots + \theta^{t-1} z_1 \end{aligned}$$

(considerando  $u_0 = 0$ )

Si  $N$  es suficientemente grande y puesto que  $|\theta| < 1$ , el valor que asuma  $u_0$  no es muy importante; sin embargo, si  $|\theta|$  está cerca de 1 y  $N$  es pequeño, la elección de  $u_0$  puede resultar significativa.

En la tabla siguiente se presentan los cálculos recursivos de  $u_t$ , para 11 valores dados de  $X_t$  en un modelo  $ARIMA(0, 1, 1)$ , en el caso en que  $\theta = 0,6$ . En este ejemplo,

$$S_*(\theta | z, u_0) = \sum_{t=1}^{10} u_t^2(\theta = 0,6 | u_0 = 0) = 511,98$$

**Tabla 3.1:** Cálculos para el caso condicionado

$t$	$X(t)$	$z(t)$	$u(t)$
0	150		0,00
1	149	-1	-1,00
2	146	-3	-3,60
3	144	-2	-4,16
4	148	4	1,50
5	155	7	7,90
6	149	-6	-1,26
7	156	7	6,24
8	161	5	8,75
9	171	10	15,25
10	170	-1	8,15

Observe que con esta disposición de datos  $z_j = w_j$  ( $j = 1, \dots, 10$ ).

Realizando el cálculo para diversos valores de  $\theta \in ]-1, 1]$ , se puede tener una idea de dónde se alcanza el mínimo para  $S_*(\theta | z, u_0)$ . En general, la solución se obtiene utilizando técnicas del análisis numérico, situación que no será abordada en este texto.

### 3.3.2 Procedimiento no condicional

Puede demostrarse que el logaritmo de la función de verosimilitud no condicional se expresa por:

$$\ell = \ell(\varphi, \theta, \sigma^2 | z) \approx -\frac{N}{2} \ln \sigma^2 - \frac{S(\varphi, \theta)}{2\sigma^2}$$

donde

$$S(\varphi, \theta) = \sum_{t=-\infty}^N [u_t(\varphi, \theta, z)]^2$$

es la suma de cuadrados no condicional y, además, los corchetes representan la esperanza condicional, definida por:

$$[u_t(\varphi, \theta, z)] = E(u_t | \varphi, \theta, z)$$

Se puede también demostrar que los estimadores mínimo cuadráticos, que se obtienen minimizando  $S(\varphi, \theta)$ , son buenas aproximaciones para los estimadores de máxima verosimilitud. Para ejecutar el proceso de minimización se requiere calcular las esperanzas condicionales de  $z_{-j}y_{-j}$ ,  $j = 0, 1, 2, \dots$ , denotadas por  $[z_{-j}]$  y  $[y_{-j}]$  .. Se utiliza para este efecto la representación en avance del proceso:

$$\Phi(F)z_t = \Theta(F)\varepsilon_t$$

donde  $F$  es el operador de avance ( $FX_t = X_{t+1}$ ) y  $(\varepsilon_t)$  es un ruido blanco con la misma varianza que  $(u_t)$ . En este caso  $z_t$  se expresa solamente en términos futuros de  $z_j$  y  $\varepsilon_j$ . Así entonces, los  $[z_{-j}]$  son las predicciones hacia atrás de  $z_{-j}$  dadas las  $z_1, \dots, z_N$ .

Para ilustrar el proceso, considérese el siguiente ejemplo:

**Ejemplo 3.4.** Sea nuevamente el modelo

$$z_t = (1 - \theta B)u_t = u_t - \theta u_{t-1}$$

cuya representación en avance se expresa por:

$$z_t = (1 - \theta F)\varepsilon_t = \varepsilon_t - \theta \varepsilon_{t+1}$$

se tiene que:

$$[u_t] = [z_t] + \theta [u_{t-1}] \quad \text{y} \quad [\varepsilon_t] = [z_t] + \theta [\varepsilon_{t+1}]$$

de donde se obtiene que:

- $[z_t] = z_t$ , si  $t = 1, \dots, N$  y son predicciones hacia atrás si  $t \leq 0$ .
- $[\varepsilon_0], [\varepsilon_{-1}], \dots$  son nulos, pues  $\varepsilon_0, \varepsilon_{-1}, \dots$  son ortogonales a  $z$ .
- $[u_{-1}], [u_{-2}]$  son nulos pues son independientes de  $z$  (en general son no nulos y se los obtiene por predicción hacia atrás).

En la tabla siguiente se ilustran los cálculos, considerando  $\theta = 0,6$

Tabla 3.2: Cálculos para el caso no condicionado

$t$	$X(t)$	$[z(t)]$	$[\varepsilon(t)]$	$[u(t)]$
-1			0,00	0,00
0	150	0,95	0,00	0,95
1	149	-1,00	-1,59	-0,43
2	146	-3,00	-0,98	-3,26
3	144	-2,00	3,36	-3,95
4	148	4,00	8,93	1,63
5	155	7,00	8,22	7,98
6	149	-6,00	2,03	-1,21
7	156	7,00	13,38	6,27
8	161	5,00	10,64	8,76
9	171	10,00	9,40	15,26
10	170	-1,00	-1,00	8,15
11			0	

Se parte calculando  $[\varepsilon_{10}]$ , bajo la suposición de que  $[\varepsilon_{11}] = 0$ :

$$[\varepsilon_{10}] = [z_{10}] + 0,6 [\varepsilon_{11}] = -1 + 0,6 \cdot 0 = -1$$

de la misma manera se procede para  $[\varepsilon_9], \dots, [\varepsilon_1]$ . Luego se tiene:

$$0 = [\varepsilon_0] = [z_0] + 0,6 \cdot [\varepsilon_1] = [z_0] + 0,6 \cdot (-1,59)$$

de donde:

$$[z_0] = 0,95$$

Luego, los  $[z_{-j}] = 0$ , para  $j = 1, 2, \dots$ . Ahora se procede a calcular los  $[u_t]$  no nulos, de la siguiente manera:

$$\begin{aligned} [u_0] &= [z_0] + \theta [u_{-1}] = [z_0] + 0 = 0,95 \\ [u_1] &= [z_1] + \theta [u_0] = -1 + 0,6 \cdot 0,95 = -0,43 \\ &\vdots \end{aligned}$$

Además:

$$[X_{-1}] = X_0 - [z_0] = 150 - 0,95 = 149,05$$

Así,

$$S(0,6) = \sum_{t=0}^{10} [u_t]^2 = 510,50$$

Como puede observarse en las dos tablas precedentes, aunque se pueden constatar ciertas diferencias entre los valores de  $u_t$  y  $[u_t]$  para valores pequeños de  $t$ , éstas van aminorándose conforme crece  $t$ , razón por la cual, aún con un pequeño número de observaciones,  $S_*(0,6)$  es bastante similar a  $S(0,6)$ .

Como se mencionó anteriormente, las minimizaciones se deben efectuar utilizando técnicas numéricas, para lo cual se requiere obtener valores iniciales para los parámetros.

El procedimiento general para obtener la función de verosimilitud exacta de un proceso

ARMA, así como para las estimaciones de las esperanzas condicionales para  $z_0, z_{-1}, \dots$  se presentan en el Anexo C.1. En todo caso, hay que recordar que los estimadores de máxima verosimilitud son normalmente asintóticos.

**Ejemplo 3.5.** En el ejemplo planteado (de la SIB) se ha considerado una serie que luego de realizar una diferenciación no estacional muestra su comportamiento estacionario; es decir, se trata de un modelo  $ARIMA(0, 1, 0)$

Con ayuda del utilitario Statgraphics y considerando el par  $p = 2$  y  $q = 2$  (es decir, se ha establecido la posibilidad de que se trate de un modelo  $ARIMA(2, 1, 2)$ ), se han obtenido los siguientes resultados:

**Tabla 3.3:** Resumen estadístico para el modelo  $ARIMA(2, 1, 2)$  de la SIB

Parámetro	Estimado	Error Estd.	Valor $t$	Valor $P$
AR(1)	1,44980	0,20709	7,00091	0,00000
AR(2)	-0,58809	0,12222	-4,81164	0,00000
MA(1)	-0,10109	0,23630	-0,42781	0,66948
MA(2)	0,60732	0,20394	2,97796	0,00344
Media	0,75431	0,30518	2,47172	0,01469
Constante	0,10431			

Esta estimación la realiza el paquete internamente a través de los métodos descritos anteriormente. En la sección siguiente se realizará la verificación, validación y elección del mejor modelo.

## 3.4 Identificación *A Posteriori*

Una vez realizadas las estimaciones de los modelos, se procede a efectuar las fases de verificación y de elección del mejor modelo.

### 3.4.1 Fase de verificación

La finalidad de esta fase consiste en analizar la adecuación entre el modelo y los datos. Las pruebas estadísticas utilizadas son de dos tipos: las concernientes a los parámetros  $a_j$  y  $b_j$  de los procesos autoregresivo y media móvil, respectivamente, del modelo propuesto  $ARMA(p, q)$  de la serie diferenciada  $d$  veces y las pruebas concernientes a las hipótesis del ruido blanco  $u_t$

#### 3.4.1.1 Pruebas concernientes a los parámetros:

Interesa saber si se puede disminuir el número de parámetros; es decir, si se dispone de un modelo  $ARMA(p, q)$ , si es posible formular un modelo  $ARMA(p - 1, q)$  o un modelo  $ARMA(p, q - 1)$ ; esto lleva a considerar la siguiente prueba de hipótesis:

$$\begin{cases} H_0 : \phi_p = 0 \\ H_1 : \phi_p \neq 0 \end{cases}$$

Si se acepta  $H_0$ , se asume que el modelo es  $ARMA(p-1, q)$ ; en caso contrario, se retiene la formulación  $ARMA(p, q)$ . Utilizando la normalidad asintótica de los estimadores, su región de rechazo está determinada por:

$$W = \left\{ \frac{|\hat{\varphi}_p|}{\sqrt{\hat{V}\hat{\varphi}_p}} > u_{1-\frac{\alpha}{2}} \right\}$$

Si ahora se considera  $p_1 = p + 1$ , la prueba a realizar es:

$$\begin{cases} H_0 : \phi_{p+1} = 0 \\ H_1 : \phi_{p+1} \neq 0 \end{cases}$$

Si se acepta  $H_0$  se retiene la forma  $ARMA(p, q)$ ; en caso contrario, se acepta el modelo  $ARMA(p_1, q)$ . En este caso la región de rechazo está determinada por:

$$W = \left\{ \frac{|\hat{\varphi}_{p+1}|}{\sqrt{\hat{V}\hat{\varphi}_{p+1}}} > u_{1-\frac{\alpha}{2}} \right\}$$

De manera similar se procede con el modelo  $ARMA(p, q_1)$  donde  $q_1 = q + 1$  o  $q_1 = q - 1$ .

**Observación 3.2.** No se deben disminuir o aumentar simultáneamente los dos parámetros.

Con el paquete estadístico Statgraphics, se obtiene una tabla (ver tabla 3.3) con los valores de los parámetros, el valor del estadístico  $T$  y el  $p$ -valor: la probabilidad obtenida con los datos observados de aceptar la hipótesis  $H_1$  de que el parámetro sea diferente de cero, cuando en verdad sea cero (el valor  $p$ , o nivel de significación empírico del contraste, es el mínimo nivel de significación tal que se rechaza la hipótesis nula).

**Ejemplo 3.6.** Retomando el ejemplo de la SIB, con propósitos ilustrativos, para mostrar los efectos de la sobreparametrización, se eligió el par  $p = 2$  y  $q = 2$ ; es decir, se ha establecido la posibilidad de que se trate de un modelo  $ARIMA(2, 1, 2)$ . Se obtienen los siguientes resultados:

Para el modelo  $ARIMA(2, 1, 2)$  se observa que los coeficientes pertenecientes a la parte MA son estadísticamente nulos (se tomará siempre como nivel de significación el 5 %. Ver Tabla 3.4), lo que indica la necesidad de reformularlo. Este hecho suele suceder como efecto de la sobreparametrización (en este caso  $q$  es muy grande).

**Tabla 3.4:** Resumen estadístico para el modelo  $ARIMA(2, 1, 2)$  de la SIB

Parámetro	Estimado	Error Estd.	Valor $t$	Valor $P$
AR(1)	1,44980	0,20709	7,00091	0,00000
AR(2)	-0,58809	0,12222	-4,81164	0,00000
MA(1)	-0,10109	0,23630	-0,42781	0,66948
MA(2)	0,60732	0,20394	2,97796	0,00344
Media	0,75431	0,30518	2,47172	0,01469
Constante	0,10431			

Considerando el modelo  $ARIMA(2, 1, 1)$ , se ve que el coeficiente que corresponde a  $AR(2)$  es estadísticamente nulo (ver tabla 3.5), según lo cual se debe volver a cambiar de modelo.

**Tabla 3.5:** Resumen estadístico para el modelo  $ARIMA(2,1,1)$  de la SIB

Parámetro	Estimado	Error Estd.	$t$	Valor- $P$
$AR(1)$	0,745854	0,100892	7,39258	0,000000
$AR(2)$	-0,116622	0,097009	-1,20218	0,231382
$MA(1)$	-0,825507	0,0626035	-13,1863	0,000000
Media	0,765563	0,416395	1,83855	0,068163
Constante	0,283847			

Considerando el modelo  $ARIMA(1, 1, 1)$  se obtiene que el coeficiente que corresponde a la constante en el modelo es estadísticamente igual a cero; por lo que, es necesario modificarlo nuevamente (ver Tabla 3.6).

**Tabla 3.6:** Resumen estadístico para el modelo  $ARIMA(1, 1, 1)$  de la SIB

Parámetro	Estimado	Error Estd.	$t$	Valor- $P$
$AR(1)$	0,64614	0,0645419	10,0112	0,000000
$MA(1)$	-0,855448	0,0491277	-17,4127	0,000000
Media	0,847445	0,45797	1,85044	0,066405
Constante	0,299877			

Finalmente, se adopta el mismo modelo  $ARIMA(1, 1, 1)$  pero sin la constante (ver Tabla 3.7). Luego, esta será la mejor estimación del modelo.

**Tabla 3.7:** Resumen estadístico para el modelo  $ARIMA(1, 1, 1)$  de la SIB sin constante

Parámetro	Estimado	Error Estd.	$t$	Valor- $P$
$AR(1)$	0,66579	0,0603313	11,0356	0,000000
$MA(1)$	-0,841889	0,0515922	-16,3181	0,000000

### 3.4.1.2 Pruebas concernientes al ruido blanco

Se pretende probar si la serie de residuos estimados:

$$\hat{u}_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} X_t$$

es coherente con la hipótesis de que los  $\hat{u}_t$  forman un ruido blanco; es decir, verificar si los residuos son no correlacionados y gaussianos (se puede probar que siempre son centrados). Estas pruebas pueden ser sustituidas por una prueba de independencia.

- **Utilización de las FAC y FACP residuales estimadas**

Las pruebas para los residuos consisten en verificar que son no correlacionados. Para ello, lo más fácil es verificar que no tienen una estructura ARMA; es decir, que sus

funciones de autocorrelación y autocorrelación parcial residuales estimadas son estadísticamente nulas (se encuentran dentro de las bandas de confianza).

- **Prueba de Box-Pierce (“Overcoat”)**

La prueba se basa en que el estadístico:

$$Q = T \sum_{l=1}^k \hat{\rho}_l^2(\hat{u}_t)$$

sigue asintóticamente una ley  $\chi^2$  con  $(k - p - q)$  grados de libertad, bajo la hipótesis de independencia de los  $u_t$  ( $\hat{\rho}_l(\hat{u}_t)$  representa la correlación estimada de orden  $l$ , de los residuos estimados). Luego, se rechaza la hipótesis de independencia de ruido blanco al nivel  $\alpha$ , si  $Q > \chi^2_{1-\alpha}[k - p - q]$  (cuantil de orden  $1 - \alpha$  de la ley  $\chi^2$  con  $(k - p - q)$  grados de libertad).

El número  $k$  debe ser suficientemente grande; habitualmente se lo toma entre 15 y 20. Cuando la hipótesis se declara inadecuada, se pueden utilizar las funciones de autocorrelación y autocorrelación parcial estimadas de los residuos, para ver en qué forma se modifica el modelo (ver la sección 3.8).

**Figura 3.6:** FAC residual estimada del Modelo  $ARIMA(2,1,2)$  de la SIB

**Figura 3.7:** FACP residual estimada del Modelo  $ARIMA(2,1,2)$  de la SIB

**Ejemplo 3.7.** En la serie planteada para el ejemplo y modelada por un  $ARIMA(2,1,2)$ , los gráficos de las funciones de autocorrelación residual estimada (Figura 3.6) y de autocorrelación parcial residual estimada (Figura 3.7), muestran los valores dentro de las bandas de confianza; lo mismo se verifica en la alternativa presentada con el modelo  $ARIMA(2,1,1)$  (Figuras 3.8 y 3.9) y, finalmente, para el modelo escogido  $ARIMA(1,1,1)$  sin constante, que se presentan en las Figuras 3.10 y 3.11.

**Figura 3.8:** FAC residual estimada del Modelo  $ARIMA(2,1,1)$  de la SIB

**Figura 3.9:** FACP residual estimada del Modelo  $ARIMA(2,1,1)$  de la SIB

**Figura 3.10:** FAC residual estimada del Modelo  $ARIMA(1,1,1)$  de la SIB

**Figura 3.11:** FACP residual estimada del Modelo  $ARIMA(1,1,1)$  de la SIB

Considerando el modelo seleccionado para la SIB ( $ARIMA(1,1,1)$ ), se presenta la salida en el paquete EViews de las autocorrelaciones simples y parciales estimadas del modelo, con



el respectivo estadístico  $Q$  y su valor  $p$  asociado (ver la columna Prob. de la figura 3.12). Dado que en todos los casos el valor  $p$  es mayor que 0,05, se consideran que los residuos no son significativos y por lo tanto se acepta la independencia de los mismos.

**Figura 3.12:** FAC y FACP residuales estimadas del Modelo  $ARIMA(1, 1, 1)$  de la SIB en EViews

### 3.4.2 Utilización de los residuos estimados para modificar el Modelo

Cuando la función de autocorrelación de los residuos estimados de un modelo ajustado indica que este es inadecuado, es necesario considerar de qué manera podría ser modificado. Supóngase que los residuos estimados  $\hat{u}_t$  del modelo:

$$\Phi_0(B) \Delta^{d_0} X_t = \Theta_0(B) u_t$$

parecen no verificar las hipótesis de independencia; utilizando las funciones de autocorrelación y autocorrelación parcial estimadas de los  $\hat{u}_t$  y los métodos explicados anteriormente, se puede ajustar para la serie  $u_t$  un modelo de la forma:

$$\Phi'(B) \Delta^{d'} u_t = \Theta'(B) \varepsilon_t$$

Eliminando  $u_t$ , de las ecuaciones anteriores, se puede asumir el siguiente modelo:

$$\Phi_0(B) \Phi'(B) \Delta^{d_0} \Delta^{d'} X_t = \Theta_0(B) \Theta'(B) \varepsilon_t$$

que puede ser ajustado nuevamente y verificado. Lo importante en el modelo anterior son los grados de los nuevos polinomios asociados a los operadores respectivos (se suman en cada producto):

$$\Phi(B) = \Phi_0(B) \Phi'(B) \quad \text{y} \quad \Theta(B) = \Theta_0(B) \Theta'(B)$$

Una aplicación de estos procedimientos se presenta en el capítulo 4.

### 3.4.3 Pruebas de estabilidad

En algunos casos, estimar un modelo de series temporales puede verse afectado por la presencia de cambios en la estructura de la serie (o series) en estudio, lo que puede llevar a conclusiones inexactas o incorrectas al momento de realizar predicciones o en la interpretación del comportamiento de la serie.

Una de las etapas de validación del modelo estimado consiste en estudiar si los parámetros estimados por el modelo son estables en toda la muestra. Si no se cumple este supuesto, la capacidad de predictibilidad del modelo será limitada.

Una de los métodos comunes para comprobar la estabilidad del modelo es dividir las  $T$  observaciones en dos grupos:

- $T_1$ : observaciones utilizadas en la estimación del modelo
- $T_2 = T - T_1$ : observaciones que se utilizarán para las pruebas y la evaluación del modelo.

Lo usual es que las primeras  $T_1$  observaciones se utilicen para modelar y las últimas observaciones para las pruebas.

Una de las pruebas tradicionales para evaluar esta estabilidad es la prueba de Chow, de punto de quiebre.

### Prueba de punto de quiebre de Chow

La idea de esta prueba consiste en estimar un modelo de cada submuestra por separado y ver si existen diferencias significativas en los modelos estimados, lo que se manifiesta en un cambio estructural en la relación.

La prueba de Chow compara la suma de residuos al cuadrado obtenidos a partir de un modelo estimado con toda la muestra con la suma de residuos al cuadrado que se obtiene cuando se estiman los modelos para cada subconjunto de las observaciones.

Para la prueba se utiliza el estadístico  $F$ , que se calcula de la siguiente manera:

$$F = \frac{(\tilde{u}'\tilde{u} - (u_1'u_1 + u_2'u_2)) / k}{(u_1'u_1 + u_2'u_2) / (T - 2k)}$$

donde,

- $\tilde{u}'\tilde{u}$ : Es la suma de cuadrados de los residuos considerando todas las observaciones.
- $u_i'u_i$ : Es la suma de cuadrados de los residuos de la submuestra  $i$  ( $i = 1, 2$ )
- $T$ : Número de observaciones.
- $k$ : Número de parámetros en la ecuación estimada.

Si no hubiera diferencias sustanciales en la estructura entre los dos modelos, el numerador sería muy pequeño; por tanto, se lo puede utilizar para contrastar la siguiente prueba de hipótesis:

$$\begin{cases} H_0: \text{No existen cambios estructurales} \\ H_1: \text{Existen cambios estructurales} \end{cases}$$

El estadístico  $F$  tiene una distribución de Fisher-Snedcor asintóticamente, bajo  $H_0$ ; es exacta si los errores son normales i.i.d.

Otra forma de evaluar este fenómeno es mediante el estadístico razón del logaritmo de verosimilitud o el estadístico de Wald. Las interpretaciones de los resultados de estos tres estadísticos son similares.

**Observación 3.3.** El tratamiento para dos submuestras se puede generalizar a más de un punto de quiebre.

**Ejemplo 3.8.** Considerando la SIB, se realiza la prueba de punto de quiebre para la observación 71 (aproximadamente, la mitad de los datos). Así, se obtienen los siguientes resultados:

**Figura 3.13:** Prueba de Chow para el punto de quiebre.

Dado que el  $p$ -valor correspondiente al estadístico  $F$  es mayor que 0,05, se acepta la hipótesis nula de no existencia de cambios estructurales de la serie.

**Observación 3.4.** En caso de que exista un cambio estructural de la serie, hay que tener mucho cuidado al realizar predicciones, ya que un cambio estructural al inicio del período de pronóstico afecta en gran medida el poder predictivo del modelo.

### 3.4.4 Elección del Modelo

Puede suceder que varios modelos pasen la fase de verificación. En este caso existen criterios que no poseen una base teórica sólida, pero son útiles a la hora de realizar una elección.

#### 3.4.4.1 Criterio del mayor poder predictivo

En un modelo ARMA, el error de predicción al horizonte 1 ( $e_t(1)$ ), es tal que su varianza  $V[e_t(1)] = \sigma^2$ . Se puede entonces, proponer escoger el Modelo que conduzca a un error de predicción más pequeño; es decir, aquel que tenga la varianza estimada del ruido blanco más pequeña. Diversos criterios son aceptables:

1. La estimación de la varianza ( $\hat{\sigma}^2$ ).
2. El coeficiente de determinación:

$$R^2 = 1 - \frac{(\hat{\sigma})^2}{V}$$

donde  $V$  es la varianza muestral de la serie diferenciada  $d$  veces. Este segundo criterio es simplemente una versión normada del anterior.

3. El coeficiente de determinación modificado:

$$\bar{R}^2 = 1 - \frac{(\hat{\sigma})^2 / (N - p - q)}{V / (N - 1)}$$

que toma en cuenta los grados de los polinomios autoregresivo y media móvil.

4. El estadístico de Fisher:

$$F = 1 - \frac{(V - (\hat{\sigma})^2) / (p + q)}{\hat{\sigma}^2 / (N - p - q)}$$

introducido por analogía con el caso del modelo lineal.

El criterio 1 es para minimizar mientras que los otros son para maximizar.

#### 3.4.4.2 Criterio de información

Otra aproximación, introducida por Akaike (1969), consiste en suponer que los modelos  $ARMA(p, q)$  dan aproximaciones de la realidad y que la verdadera distribución desconocida de las observaciones  $\Delta^d X_t$  (observaciones de  $(X_t)$  diferenciadas  $d$  veces) no satisface forzosamente tal modelo. Se puede entonces fundamentar la elección del modelo sobre una medida de la desviación entre la verdadera ley desconocida y el modelo propuesto.

La medida habitual es la cantidad de información de Kullback. Sean  $f_0(X)$  la densidad desconocida de las observaciones y  $\{f(x), f \in F_{p,q}\}$  la familia de densidades correspondientes

al modelo  $ARMA(p, q)$ ; la desviación entre la verdadera ley y el Modelo se mide por:

$$I(f_0, F_{p,q}) = \int \log \frac{f_0(x)}{f(x)} f_0(x) dx$$

Esta cantidad es siempre positiva y solo se anula si la verdadera densidad  $f_0$  pertenece a  $F_{p,q}$ .  $I(f_0, F_{p,q})$ , el valor de la cantidad de información es evidentemente desconocido; pero si se dispone de un buen estimador  $I(f_0, F_{p,q})$  de esta cantidad, será posible utilizarlo como criterio.

Se escoge entonces el modelo  $F_{p,q}$  que conduce al valor más pequeño de la estimación  $I(f_0, F_{p,q})$ . Los estimadores de la cantidad de información que han sido propuestos son:

1.  $AIC(p, q) = \log(\hat{\sigma})^2 + \frac{2(p+q)}{T}$  (Akaike, 1969)
2.  $BIC(p, q) = \log(\hat{\sigma})^2 + (p+q) \frac{\log T}{T}$  (Schwarz, 1978)
3.  $\varphi(p, q) = \log(\hat{\sigma})^2 + (p+q) c \frac{\log(\log T)}{T}$  con  $c > 2$  (Hannan-Quinn, 1979)

El primero de estos estimadores es el más utilizado; sin embargo, solo los dos restantes son convergentes y conducen a una selección asintótica correcta del modelo.

El objetivo es minimizar el error de predicción final, para lo cual los parámetros  $p$  y  $q$  pueden ser cambiados hasta obtener un valor mínimo del estimador. Para modelos autoregresivos, simulaciones de Monte Carlo sugieren que AIC tiene una tendencia a sobrestimar el parámetro  $p$ ; para corregir esta tendencia, se introduce el criterio modificado BIC (criterio de información bayesiano), donde  $\hat{\sigma}^2$  es el estimador de máxima verosimilitud para la varianza del ruido blanco.

En forma general, la introducción de estos criterios ayudan a la identificación del modelo. Sin embargo, la búsqueda de un modelo que los minimice puede ser muy larga si no se tiene alguna idea de la clase de modelo a explorar.

**Observación 3.5.** En ciertas ocasiones, no se obtienen los mismos resultados con todos los criterios; en este caso, se elige el modelo que cumpla con el mayor número de criterios (los paquetes estadísticos también suelen incorporar otros criterios como, por ejemplo, la razón de verosimilitud)

## 3.5 Predicción de los Modelos ARIMA

### 3.5.1 Cálculo de las predicciones óptimas en un modelo ARIMA

**Definición 3.1.** La predicción lineal óptima  $\hat{X}_T(h)$ , al horizonte  $h$  ( $h > 0$ , entero), en el instante  $T$ , se define como la regresión afín de  $X_{T+h}$  sobre el espacio vectorial generado por  $\{X_t, t \in \mathbb{Z}\}$ .

Supóngase que  $(X_t, t \in \mathbb{Z})$  es un proceso  $ARIMA(p, d, q)$ . La representación media móvil de tal proceso se escribe:

$$X_t = m_t + \sum_{i=0}^{t+s-p'-1} \psi_i u_{t-i} \quad t \in \mathbb{Z}$$

Por tanto, para el instante  $T + h$ , se tiene:

$$X_{T+h} = m_{T+h} + \sum_{i=0}^{T+h+s-p'-1} \psi_i u_{T+h-i}$$

donde  $(u_t)$  es un ruido blanco.

La regresión afín de  $X_{T+h}$  sobre el e.v. generado por las  $X_t$ ,  $t \leq T$ , es la misma que la regresión afín de  $X_{T+h}$  sobre el e.v. generado por los  $\{u_t, t \leq T\}$ ; además, ésta viene dada por:

$$\hat{X}_T(h) = m_{T+h} + \sum_{i=h}^{T+h+s-p'-1} \psi_i u_{T+h-i} = m_{T+h} + \sum_{j=0}^{T+s-p'-1} \psi_{j+h} u_{T-j} \quad (3.1)$$

Esta fórmula da la regresión afín de  $X_{T+h}$  sobre los  $X_t$ ,  $t \leq T$ , puesto que los  $u_t$  son no correlacionados. Si, además, los  $u_t$  son normales o independientes entre ellos,  $\hat{X}_T(h)$  es igualmente la esperanza condicional de  $X_{T+h}$  conocidas  $X_t$ ,  $t \leq T$ .

El error de predicción se deduce directamente de (3.1):

$$e_T(h) = X_{T+h} - \hat{X}_T(h) = \sum_{i=0}^{h-1} \psi_i u_{T+h-i}$$

En particular:  $e_T(1) = u_{T+1}$

Además, el error en media cuadrática de  $\hat{X}_T(h)$ , que en este caso coincide con la varianza de  $\hat{X}_T(h)$ , es:

$$EMC_T(h) = E(X_{T+h} - \hat{X}_T(h))^2 = \sigma^2 \sum_{i=0}^{h-1} \psi_i^2$$

En particular:  $EMC_T(1) = \sigma^2$

**Observación 3.6.** Por lo general, cuando  $X_t$  se puede expresar en la forma denominada  $MA(\infty)$ :

$$X_t = \sum_{i=0}^{\infty} \psi_i u_{t-i} \quad t \in \mathbb{Z}$$

se obtienen los mismos resultados para el error de predicción y el EMC de horizonte  $h$ . Los coeficientes  $\psi_i$  se pueden obtener de la expresión  $\psi(B)\phi(B) = \Theta(B)$ , donde:

$$\psi(B) = \sum_{i=0}^{\infty} \psi_i B^i$$

Obsérvese que  $MA(\infty)$  es solamente una notación, pues la serie  $\psi(B)$  no siempre es convergente; no confundir con la definición de un proceso lineal ni con la de una media móvil infinita.

La fórmula (3.1) no es utilizable directamente, pues los  $u_t$  no son observables; es necesario entonces buscar otras expresiones de  $\hat{X}_T(h)$ .

Considérese también la representación AR con  $s$  grande (se pueden despreciar los términos

con los  $\pi_{j,T+h}^*$ ):

$$X_{T+h} = - \sum_{i=1}^{T+h+s-p'-1} \pi_i X_{T+h-i} + u_{T+h}$$

entonces:

$$\hat{X}_T(h) = - \sum_{i=1}^{T+h+s-p'-1} \pi_i \hat{X}_{T+h-i} \quad (3.2)$$

donde:

$$\hat{X}_{T+h-i} = \begin{cases} \hat{X}_T(h-i) & \text{si } i < h \\ X_{T+h-i} & \text{si } i \geq h \end{cases}$$

Si se considera la representación ARIMA:

$$X_{T+h} = \sum_{i=1}^{p'} \varphi_i X_{T+h-i} + u_{T+h} - \sum_{i=1}^q \theta_i u_{T+h-i}$$

se obtiene:

$$\hat{X}_T(h) = \sum_{i=1}^{p'} \varphi_i \hat{X}_{T+h-i} - \sum_{i=1}^q \theta_i \hat{u}_{T+h-i} \quad (3.3)$$

donde:

$$\hat{u}_{T+h-i} = \begin{cases} 0 & \text{si } i < h \\ u_{T+h-i} & \text{si } i \geq h \end{cases}$$

Finalmente, considérese ahora la predicción de  $X_{T+h}$  en el instante  $T+1$ :

$$\hat{X}_{T+1}(h-1) = m_{T+h} + \sum_{i=h-1}^{T+h+s-p'-1} \psi_i u_{T+h-i}$$

por tanto:

$$\hat{X}_{T+1}(h-1) - \hat{X}_T(h) = \psi_{h-1} u_{T+1} = \psi_{h-1} (X_{T+1} - \hat{X}_T(1)) \quad (3.4)$$

Con esta fórmula se puede proceder a actualizar las predicciones, cuando se ha observado  $X_{T+1}$ .

En resumen:

1. Con (3.2) se calcula  $\hat{X}_T(h)$ ,  $h = 1, 2, 3, \dots, H$
2. Si se conoce  $X_{T+1}$  se calcula con (3.4),  $\hat{X}_{T+1}(h)$ ,  $h = 1, 2, \dots, H-1$  (actualización).
3.  $\hat{X}_T(H)$  se calcula con (3.3), bajo la condición de que  $H > q$  (caso frecuente).

**Observación 3.7.** Para el cálculo práctico de las predicciones se suponen conocidos los parámetros  $\psi_i$  y  $\theta_j$  (luego de que se han sido estimados); lo mismo se asume con  $u_1, \dots, u_T$ . En cuanto a  $u_0, u_{-1}, \dots$ , se los toma igual a sus esperanzas (es decir, como cero) o, alternatively, se los estima.

**Ejemplo 3.9.** A continuación se presentan algunos casos en los que se explicitan los predictores y los EMC:

1. Sea el modelo  $ARIMA(0, 0, 2)$ :

$$X_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} \quad t \in \mathbb{Z}$$

En este caso:

$$\psi_1 = -\theta_1; \quad \psi_2 = -\theta_2 \quad y \quad \psi_j = 0 \text{ para } j > 2$$

Se calculan a continuación para varios períodos de predicción el valor real u observado, el predictor y el correspondiente error medio cuadrático.

Período	Valor Real	Predictor	EMC
$T + 1$	$X_{T+1} = u_{T+1} - \theta_1 u_T - \theta_2 u_{T-1}$	$\hat{X}_T(1) = -\theta_1 u_T - \theta_2 u_{T-1}$	$\sigma_u^2$
$T + 2$	$X_{T+2} = u_{T+2} - \theta_1 u_{T+1} - \theta_2 u_T$	$\hat{X}_T(2) = -\theta_2 u_T$	$(1 + \theta_1^2) \sigma_u^2$
$T + 3$	$X_{T+3} = u_{T+3} - \theta_1 u_{T+2} - \theta_2 u_{T+1}$	$\hat{X}_T(3) = 0$	$(1 + \theta_1^2 + \theta_2^2) \sigma_u^2$

Puede verse que para  $h > 2$ , resultará que  $\hat{X}_T(h) = 0$  y que  $EMC(h) = (1 + \theta_1^2 + \theta_2^2) \sigma_u^2$ . En un modelo  $MA(q)$ , se verifica que  $\hat{X}_T(h) = 0$  para  $h > q$  y que el error medio cuadrático se estabiliza en el valor que toma para  $h = q + 1$ .

2. Sea un modelo  $ARIMA(0, 1, 0)$ ; es decir, un modelo “marcha aleatoria”

$$X_t = X_{t-1} + u_t \quad t \in \mathbb{Z}$$

Al pasar a la forma  $MA(\infty)$  se tiene que  $\psi_i = 1$  para  $i = 1, 2, 3, \dots$

Para los períodos  $T + 1$ ,  $T + 2$  y  $T + 3$  se obtienen los siguientes resultados:

Período	Valor Real	Predictor	ECM
$T + 1$	$X_{T+1} = X_T + u_{T+1}$	$\hat{X}_T(1) = X_T$	$\sigma_u^2$
$T + 2$	$X_{T+2} = X_{T+1} + u_{T+2}$	$\hat{X}_T(2) = \hat{X}_T(1) = X_T$	$2\sigma_u^2$
$T + 3$	$X_{T+3} = X_{T+2} + u_{T+3}$	$\hat{X}_T(3) = \dots = X_T$	$3\sigma_u^2$

En un modelo “marcha aleatoria” para todos los períodos futuros se predice con el valor correspondiente al último período muestral. El EMC se incrementa en  $\sigma_u^2$  para cada período adicional.

3. Sea un modelo  $ARIMA(1, 1, 0)$

$$(1 - \phi_1 B)(1 - B)X_t = u_t$$

Pasando a un ARMA mediante la utilización de parámetros autorregresivos generalizados se tiene que

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = u_t$$

donde  $\phi_1 = \phi_1 + 1$  y  $\phi_2 = -\phi_1$ .

A partir de la relación:

$$\psi(B)\phi(B) = 1$$

Se tiene que  $\psi_0 = 1$ , y  $\psi_1 = \varphi_1$ . Tomando  $\psi_0$  y  $\psi_1$  como valores iniciales, los siguientes valores de  $\psi_i$  se determinan recursivamente mediante la fórmula:

$$\psi_k = \varphi_1 \psi_{k-1} + \varphi_2 \psi_{k-2}, \quad k \geq 2$$

Para los períodos  $T+1$ ,  $T+2$  y  $T+3$  se obtienen los siguientes resultados:

Período	Valor Real	Predictor	ECM
$T+1$	$X_{T+1} = \varphi_1 X_T + \varphi_2 X_{T-1} + u_{T+1}$	$\hat{X}_T(1) = \varphi_1 X_T + \varphi_2 X_{T-1}$	$\sigma_u^2$
$T+2$	$X_{T+2} = \varphi_1 X_{T+1} + \varphi_2 X_T + u_{T+2}$	$\hat{X}_T(2) = \varphi_1 \hat{X}_T(1) + \varphi_2 X_T$	$(1 + \psi_1^2) \sigma_u^2$
$T+3$	$X_{T+3} = \varphi_1 X_{T+2} + \varphi_2 X_{T+1} + u_{T+3}$	$\hat{X}_T(3) = \varphi_1 \hat{X}_T(2) + \varphi_2 \hat{X}_T(1)$	$(1 + \psi_1^2 + \psi_2^2) \sigma_u^2$

4. Sea el modelo  $ARIMA(1,0,1)$

$$X_t = \varphi_1 X_{t-1} + u_t - \theta_1 u_{t-1}$$

El coeficiente  $\psi_1 = \varphi_1 - \theta_1$ . Los siguientes valores se calculan de forma recursiva a partir de

$$\psi_k = \varphi_1 \psi_{k-1}, \quad k \geq 1$$

Para este modelo, al predecir en los períodos  $T+1$ ,  $T+2$  y  $T+3$  se obtienen los siguientes resultados:

Período	Valor Real	Predictor	EMC
$T+1$	$X_{T+1} = \varphi_1 X_T + u_{T+1} - \theta_1 u_T$	$\hat{X}_T(1) = \varphi_1 X_T - \theta_1 u_T$	$\sigma_u^2$
$T+2$	$X_{T+2} = \varphi_1 X_{T+1} + u_{T+2} - \theta_1 u_{T+1}$	$\hat{X}_T(2) = \varphi_1 \hat{X}_T(1)$	$(1 + \psi_1^2) \sigma_u^2$
$T+3$	$X_{T+3} = \varphi_1 X_{T+2} + u_{T+3} - \theta_1 u_{T+2}$	$\hat{X}_T(3) = \varphi_1 \hat{X}_T(2)$	$(1 + \psi_1^2 + \psi_2^2) \sigma_u^2$

### 3.5.2 Intervalos de predicción

Si se asume que  $u_t \sim N(0, \sigma^2)$ , entonces:

$$e_T(h) = X_{T+h} - \hat{X}_T(h) = \sum_{i=0}^{h-1} \psi_i u_{T+h-i} \sim N\left(0, \sigma^2 \sum_{i=0}^{h-1} \psi_i^2\right)$$

Así, se puede obtener un intervalo de confianza, llamado intervalo de predicción, para  $X_{T+h}$ :

$$I = \hat{X}_T(h) \pm u_{1-\frac{\alpha}{2}} \sigma \sqrt{\sum_{i=0}^{h-1} \psi_i^2}$$

donde  $u_{1-\frac{\alpha}{2}}$  es el cuantil de orden  $(1 - \frac{\alpha}{2})$  de la distribución  $N(0, 1)$ .

Para calcular este intervalo, se requiere reemplazar  $\sigma$  por su estimación.



### 3.5.3 Funciones de predicción

Se conoce que:

$$\hat{X}_t(h) = \sum_{i=1}^{p'} \varphi_i \hat{X}_t(h-i) \iff \hat{X}_t(h) - \sum_{i=1}^{p'} \varphi_i \hat{X}_t(h-i) = 0$$

$h > q$ . Si  $(h \leq i, \hat{X}_t(h-i) = X_{t+h+i})$ ; es decir, se tiene una ecuación en recurrencia.

Se llamará función de predicción a  $\hat{X}_t(h)$ , (con  $t$  fijo). Entonces la solución general de  $\hat{X}_t(h)$  es:

$$\hat{X}_t(h) = \sum_{i=1}^{p'} b_i(t) f_i(h)$$

donde los  $f_i(h)$  son las soluciones elementales de la ecuación de recurrencia que tiene por polinomio característico a:

$$z^{p'} - \sum_{i=1}^{p'} \varphi_i z^{p'-i} = z^{p'} \left[ 1 - \sum_{i=1}^{p'} \varphi_i \left( \frac{1}{z} \right)^i \right] = z^{p'} \varphi \left( \frac{1}{z} \right)$$

es decir, las inversas de las raíces de  $\varphi(z) = 0$ . Los  $b_i(t)$  se determinan de los  $p'$  valores iniciales (ver Anexo B.3).

**Observación 3.8.** 1. Si no hay diferenciación:  $\varphi(z) = \Phi(z)$ ; además, si las raíces son distintas, se tiene:

$$\hat{X}_t(h) = \sum_{i=1}^p b_i(t) \lambda_i^h$$

(los  $\lambda_i$  inversas de las raíces de  $\Phi(z) = 0$ )

$$\therefore \hat{X}_t(h) \xrightarrow[m.c.]{h \rightarrow \infty} 0$$

El resultado es cierto también en el caso general, si las raíces de  $\Phi(z)$  son de  $|| > 1$ .

2. Si  $\phi(z) = (1-z)^d \Phi(z)$ ,  $\hat{X}_T(h)$  admite como límite un polinomio asintótico de grado  $(d-1)$ .
3. Si además en el modelo existe una constante  $\theta_0$ , hay que añadir una solución particular; por ejemplo:

$$\frac{\theta_0 h^d}{d! \Phi(1)}$$

**Ejemplo 3.10.** Considérese el modelo;

$$(1-B)^2(1-0,5B)X_t = 2 + (1-0,8B)u_t$$

con los siguientes valores iniciales;  $\hat{X}_t(1) = 1, X_t = 2, X_{t-1} = -1$ .

La forma general de  $\hat{X}_T(h)$  es:

$$\hat{X}_T(h) = a + bh + c(1/2)^h + \frac{2h^2}{2!(1/2)}$$

Para determinar los coeficientes  $a$ ,  $b$  y  $c$ , se deben utilizar los valores iniciales de  $\hat{X}_T(h)$  con  $h = 1, 0, -1$ :

$$\begin{aligned}\hat{X}_T(1) &= a + b + c(1/2) + 2 = 1 \\ \hat{X}_T(0) &= a + c = 2 \\ \hat{X}_T(-1) &= a - b + 2c + 2 = -1\end{aligned}$$

Se debe resolver entonces, el sistema de ecuaciones:

$$\begin{cases} 2a + 2b + c = -2 \\ a - b + 2c = -3 \\ a + c = 2 \end{cases}$$

que tiene como solución:  $a = 18$ ,  $b = -11$ ,  $c = -16$ ; entonces:

$$\hat{X}_T(h) = 18 - 11h - 16(1/2)^h + 2h^2$$

### 3.5.4 Contraste de la estabilidad estructural en la predicción

Cuando se ha utilizado un modelo ARIMA, ya validado, para predecir los valores de la variable en periodos extramuestrales, se puede plantear el siguiente interrogante: ¿sigue siendo válido el modelo para los periodos de predicción?; en otras palabras, ¿se ha producido un cambio estructural en los periodos de predicción?

Con objeto de obtener una respuesta a esta cuestión se puede utilizar el siguiente estadístico.

$$E_m^* = \frac{\sum_{i=0}^m e_{T+h}^2(i)}{\hat{\sigma}^2} \quad (3.5)$$

donde  $e_{T+h}(i)$  es el error de predicción de  $X_{T+h+i}$  utilizando la información disponible hasta el momento  $T + h$ .

Se puede demostrar que el estadístico tiene asintóticamente una distribución  $\chi^2$  con  $m$  grados de libertad, bajo la hipótesis nula de que no se ha dado una ruptura estructural al pasar a los periodos de predicción. Para comprender el significado de este contraste debe tenerse en cuenta que si el modelo sigue siendo correcto se verificará que:

$$e_{T+h}(i) \approx u_{T+h+i}$$

Por el contrario, en la medida que esto no ocurra, tenderá a aumentar el EMC de cada predicción, y por tanto el estadístico  $E_m^*$ , lo que llevaría a rechazar la hipótesis nula.

**Observación 3.9.** Esta prueba también se la encuentra con el nombre de *Prueba de predicciones de Chow*. Se puede utilizar el siguiente estadístico:

$$F = \frac{(\tilde{u}'\tilde{u} - u_1'u_1) / T_2}{u_1'u_1 / (T_1 - k)}$$

donde

- $\tilde{u}'\tilde{u}$ : Es la suma de cuadrados de los residuos considerando todas las observaciones.

- $u_1' u_1$ : Es la suma de cuadrados de los residuos de la submuestra  $T_1$
- $T_1$ : Es un subperíodo de  $T$ .
- $T$ : Número de observaciones.
- $k$ : Número de coeficientes estimados.
- $T_2$ : Número de puntos de predicción.

$F$  sigue una distribución exacta si los errores son i.i.d.

Por otro lado, también se puede considerar la estimación del logaritmo de la razón de verosimilitud que está basado en la comparación del máximo de las funciones de verosimilitud restringida y sin restringir. Ambos logaritmos de verosimilitud son obtenidos a partir de la regresión utilizando toda la muestra. La prueba tiene una distribución asintótica  $\chi^2$  con  $T_2$  grados de libertad, bajo  $H_0$ , para la siguiente prueba de hipótesis:

$$\begin{cases} H_0 : & \text{No existen cambios estructurales} \\ H_1 : & \text{Existen cambios estructurales} \end{cases}$$

Cabe recalcar que esta formulación es similar a la presentada anteriormente.

**Ejemplo 3.11.** Considerando la SIB se realiza la prueba anterior, teniendo en cuenta que se predicen desde la observación 135 hasta la 141. El resultado es el siguiente:

Chow Forecast Test			
Equation: UNTITLED			
Specification: $D(SER1, 1)$ $AR(1)$ $MA(1)$			
Test predictions for observations from 135 to 141			
	Value	df	Probability
F-statistic	0.363066	(7, 130)	0.9221
Likelihood ratio	2.691185	7	0.9120

**Figura 3.14:** Prueba de predicciones de Chow

Como se puede observar, el valor  $p$  asociado al estadístico  $F$  es mayor que 0,05; por lo tanto, se acepta la hipótesis nula de que no existen cambios estructurales en las predicciones.

## 3.6 Transformación de Datos

El operador  $\Delta^d = (1 - B)^d$  aplicado a  $X_t$  tiene por objeto volver a la serie estacionaria. Sin embargo, para ciertos tipos de series no se podrá obtener la estacionariedad de esta manera. Por ejemplo: si la serie  $X_t$  tiene una esperanza matemática que es función exponencial de  $t$  (tendencia exponencial), ningún operador  $\Delta^d$  podrá anularla, ya que  $\Delta^d (e^{at}) = e^a (e^t - e^{t-d})$ ; una idea natural en este caso consiste en transformar la serie  $X_t$  (supuesta positiva) por la función  $\ln$  antes de aplicar un operador  $\Delta^d$ .

De igual manera, si la serie es del tipo  $X_t = (at + b)Z_t$ , donde  $(Z_t)$  es un proceso estacionario con  $E Z_t = 1$ ,  $V Z_t = \sigma^2$  y  $a > 0$ , se tiene:

- $E X_t = at + b.$
- $V(X_t) = \sigma^2(at + b)^2.$
- $E \Delta X_t = a.$
- $V(\Delta X_t) = \sigma^2 \left\{ [at + b]^2 + [a(t-1) + b]^2 - 2[at + b][a(t-1) + b]\rho(1) \right\}$   
 $= \sigma^2 \left\{ a^2 + 2(1 - \rho(1)) [at + b][a(t-1) + b] \right\}$

Así,  $\Delta X_t$  tiene una esperanza constante y una varianza creciente con  $t$ ; si se toma previamente los logaritmos, se obtiene:

$$\Delta \ln X_t = \ln \frac{at + b}{a(t-1) + b} + \Delta \ln Z_t \approx \Delta \ln Z_t$$

Si  $(Z_t)$  es estacionario (en sentido estricto),  $\Delta \ln X_t$  lo será asintóticamente.

Se constata que la transformación  $\ln$  puede ser útil y las indicaciones generalmente admitidas de su utilidad son: un crecimiento exponencial de la tendencia o una crecimiento de la variabilidad de  $\Delta X_t$  junto con una estabilidad en media.

Si se supone que la serie  $Y_t = \ln X_t$  es un proceso ARIMA, se le puede aplicar las técnicas vistas y obtener así  $\hat{Y}_t(h) = E[Y_{t+h}|Y_t, Y_{t-1}, \dots]$ . El problema es entonces deducir de  $\hat{Y}_t(h)$  una predicción  $X_t^*(h)$  de  $X_{t+h}$ . La primera idea consiste en tomar:

$$X_t^*(h) = \exp [\hat{Y}_t(h)]$$

Es claro, sin embargo, que  $X_t^*(h)$  es diferente de la predicción óptima  $\hat{X}_t(h) = E[X_{t+h}|X_t, X_{t+1}, \dots]$ , pues por la desigualdad de Jensen (para convexidad) se tiene:

$$\begin{aligned} \hat{X}_t(h) &= E[X_{t+h}|X_t, X_{t-1}, \dots] \\ &= E[\exp(\ln X_{t+h})|X_t, X_{t-1}, \dots] \\ &> \exp E[\ln X_{t+h}|X_t, X_{t-1}, \dots] \\ &= \exp E[Y_{t+h}|Y_t, Y_{t-1}, \dots] \\ &= \exp [\hat{Y}_t(h)] = X_t^*(h) \end{aligned}$$

Supóngase ahora que  $Y_t$  es normal (o  $X_t$  log-normal). Se tiene entonces:

$$\hat{X}_t(h) = E[\exp Y_{t+h}|Y_t, Y_{t+1}, \dots] = \exp [\hat{Y}_t(h)] \exp \left[ \frac{1}{2} V(Y_{t+h})|Y_t, Y_{t-1}, \dots \right]$$

ya que la esperanza de la ley log-normal asociada a la ley normal  $N(m, \sigma^2)$  es:

$$\exp \left[ m + \frac{\sigma^2}{2} \right]$$

Así, se tiene:

$$\hat{X}_t(h) = X_t^*(h) \exp \left[ \frac{1}{2} V(Y_{t+h})|Y_t, Y_{t-1}, \dots \right]$$

Para calcular el factor de corrección, recordando que los  $u_t$  son el ruido blanco asociado a

$Y_t$ , obsérvese que:

$$\begin{aligned} V(Y_{t+h}|Y_t, Y_{t-1}, \dots) &= V(Y_{t+h}|u_t, u_{t-1}, \dots) \\ &= V\left(\sum_{i=0}^{h-1} \psi_i u_{t-1}\right) \\ &= \sigma^2 \sum_{i=0}^{h-1} \psi_i^2 \end{aligned}$$

Finalmente, se toma como predictor de  $X_{t+h}$ :

$$\hat{X}_t(h) = \exp \left[ \hat{Y}_t(h) + \frac{\sigma^2}{2} \sum_{i=0}^{h-1} \psi_i^2 \right]$$

En la práctica se reemplazan los parámetros por sus estimadores.

La transformación logarítmica es sin duda la más utilizada para las series económicas. Sin embargo, esta transformación es un elemento de una clase más general (llamada clase de transformaciones de Box-Cox) definida por:

$$T_\lambda(X_t) = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln X_t & \lambda = 0 \end{cases}$$

$X_t$  se supone positiva. Obsérvese que  $\ln X_t = \lim_{\lambda \rightarrow 0} T_\lambda(X_t)$ .

## 3.7 Raíces Unitarias

En el desarrollo de la metodología de Box-Jenkins para el tratamiento de las series temporales (ver capítulo 2), se exige que antes de la fase de identificación, si las series no son estacionarias se hagan las transformaciones oportunas para lograr la estacionariedad. Una de las transformaciones más comunes es la “diferenciación”, que consiste en considerar la transformación:

$$Y_t = (1 - B) X_t = X_t - X_{t-1}$$

Puesto que el polinomio autoregresivo asociado  $\Phi(z) = 1 - z$  tiene como raíz al valor 1, se dice que el proceso  $(Y_t)$  tiene una raíz unitaria; también se dice que la serie  $(Y_t)$  es integrada de orden 1 y se denota por  $I(1)$ . Obviamente, de manera similar, se pueden definir raíces unitarias de orden superior.

El gráfico de la serie temporal o el de sus correlogramas se utilizan frecuentemente para determinar la necesidad de “diferenciar” la serie; son instrumentos empíricos para detectar la presencia de raíces unitarias, pero pueden resultar algo imprecisos.

A partir de 1979 se han desarrollado pruebas de hipótesis apropiadas para saber si la serie tiene o no raíces unitarias; éstas se realizan utilizando la metodología propuesta por Dickey y Fuller (de ahí se desprende el nombre de la prueba). En este documento se utilizará el paquete EViews, que proporciona las pruebas Dickey-Fuller Aumentada (ADF) y la de Phillips-Perron.

### 3.7.1 La prueba de Dickey-Fuller

La forma más fácil de presentar esta prueba es considerar un modelo  $AR(1)$ :

$$X_t = \rho X_{t-1} + u_t, \quad t \in \mathbb{Z}$$

con  $|\rho| < 1$  y  $(u_t)$  un ruido blanco. La condición  $|\rho| < 1$  es necesaria y suficiente para que el proceso sea estacionario y lineal. En este caso  $X_t$  es la innovación en el instante  $t$ ; es decir,  $u_t$  influye sobre los valores de  $X_t$  correspondientes al mismo período, o a períodos posteriores, pero nunca ejerce influencia sobre los valores de  $X_t$  correspondientes a períodos anteriores. Considérese ahora el modelo  $AR(1)$  precedente. Si el coeficiente de  $X_{t-1}$  es en realidad igual a 1, surge el problema de raíz unitaria; es decir, una situación de no estacionariedad. Por consiguiente si se realiza la regresión:

$$X_t = \rho X_{t-1} + u_t$$

y si se encuentra que  $\rho = 1$ , entonces se dice que la variable  $X_t$  tiene una raíz unitaria (en este caso se dice que la serie es una *caminata aleatoria*). En este caso se realiza la prueba de hipótesis:

$$\begin{cases} H_0 : \rho = 1 \\ H_1 : \rho < 1 \end{cases}$$

Con frecuencia, a la ecuación anterior se la expresa de forma alternativa, utilizando el operador de primera diferencia  $\Delta = 1 - B$  y sustrayendo de cada lado  $X_{t-1}$ :

$$\Delta X_t = (\rho - 1) X_{t-1} + u_t$$

$$\Delta X_t = (\rho - 1) X_{t-1} + u_t = \delta X_{t-1} + u_t$$

donde  $\delta = (\rho - 1)$  y  $\Delta X_t = X_t - X_{t-1}$ . Ahora las hipótesis nula y alternativa son:

$$\begin{cases} H_0 : \delta = 0 \\ H_1 : \delta < 0 \end{cases}$$

Para averiguar si una serie temporal es o no estacionaria se efectúa una regresión, utilizando cualquiera de los modelos anteriores y se determina si  $\rho = 1$  estadísticamente (o  $\delta = 0$ ). Esta es la forma en que aparece la prueba en el paquete EViews.

Bajo la hipótesis nula que  $\rho = 1$ , el estadístico  $t$  no sigue una distribución  $t$ -student convencional. Dickey y Fuller mostraron que sigue una distribución especial y calcularon sus valores críticos con base en simulaciones de Monte Carlo. De esta manera, al estadístico  $t$  calculado convencionalmente se le denota como el estadístico  $\tau$ , y a su prueba como la de Dickey y Fuller.

Note que  $H_0$  se acepta si el valor observado de  $\tau$  es mayor al valor crítico correspondiente, de acuerdo al nivel de significación que se adopte para la prueba. Si la hipótesis nula se rechaza; es decir, si la serie de tiempo es estacionaria, se puede utilizar la prueba  $t$ -student usual.

Las tablas iniciales no eran del todo adecuadas, por lo cual fueron ampliadas por MacKinnon mediante simulaciones; éstas son las que utiliza el paquete Eviews.

Existen tres versiones del modelo de regresión:

1. Modelo sin componentes deterministas, que ya ha sido tratado:

$$X_t = \rho X_{t-1} + u_t \quad t = 2, 3, \dots, T$$

2. Modelo con un término constante:

$$X_t = \alpha + \rho X_{t-1} + u_t \quad t = 2, 3, \dots, T$$

en el cual la hipótesis nula conjunta es:  $H_0 : \alpha = 0, \rho = 1$ . Si se rechaza la hipótesis nula  $H_0 : \alpha = 0, \rho = 1$  y se concluye que  $\alpha \neq 0$  y  $\rho < 1$ , se tendrá que el proceso es estacionario con media no nula. Si  $\alpha = 0$ , se obtiene el caso precedente.

3. Modelo con un término constante y una tendencia lineal

$$X_t = \alpha + \beta t + \rho X_{t-1} + u_t \quad t = 2, 3, \dots, T$$

en el cual, si al rechazar la hipótesis nula  $H_0 : \alpha = \beta = 0, \rho = 1$ , se concluye que  $\beta \neq 0$  y  $\rho < 1$ , se dice que el proceso es estacionario alrededor de una tendencia lineal. Note que en este caso el proceso no es estacionario en el sentido de la definición; sin embargo, luego de una diferenciación se torna en estacionario. Así, el modelo solamente ahorra un paso en el análisis.

Si se ha aceptado  $H_0$  y se desea saber si se requiere una segunda diferenciación, se aplica el mismo procedimiento a la serie  $\Delta X_t$ . Los modelos de regresión precedentes también se pueden emplear en este caso.

### 3.7.2 Extensiones de la prueba de Dickey-Fuller

La prueba de raíz unitaria descrita hasta el momento es válida solamente si la serie obedece a un proceso  $AR(1)$ . Si la serie tiene una correlación de un orden mayor, se viola el supuesto de ruido blanco sobre los residuos, por lo cual se requiere aplicar la prueba de Dickey – Fuller aumentada (ADF), en la que se realiza una corrección paramétrica para correlaciones de orden superior asumiendo que la serie  $X_t$  sigue un proceso  $AR(p)$ ; esto es, añade retardos de la primera diferencia de  $X_{t-1}$  al lado derecho de la regresión. El modelo de manera general se expresa por:

$$\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \sum_{i=0}^{p-1} \gamma_i \Delta X_{t-i} + u_t$$

De esta manera, la hipótesis nula y alternativa a contrastarse son:  $H_0 : \delta = 0$  y  $H_1 : \delta < 0$ . Said y Dickey (1985) mostraron que esta prueba es asintóticamente válida incluso para series que presenten una componente media móvil; solamente, debe escogerse un valor de  $k = p - 1$  adecuado (en general se utiliza como  $k$  a la parte entera de la raíz cuarta del tamaño muestral); de manera que, para series grandes con componente media móvil pueden utilizarse los mismos valores críticos.

El contraste de Phillips-Perron (PP), es una generalización del contraste de DF, que permite flexibilidad en las hipótesis relativas a la distribución de errores.

**Observación 3.10.** Si se añade el término independiente y/o la tendencia determinista en

los modelos de regresión de DF o de DFA sin que lo exija el proceso, se reducen los grados de libertad, se aumenta el tamaño de la región crítica y se reduce además la potencia de la prueba, lo que puede conducir a una interpretación errónea del resultado de los contratos. Por tanto, es muy importante determinar si se incluyen o no el término independiente y/o tendencia determinista en la regresión seleccionada.

En la práctica, es mejor empezar tratando el modelo (c) y de acuerdo a los resultados de los coeficientes  $\alpha$  y  $\beta$  adoptar el modelo adecuado. También, se debe tener cuidado de elegir los coeficientes  $\gamma_i$  necesarios (estadísticamente significativos); en caso contrario, se puede incurrir en errores en la decisión sobre la prueba. El paquete Eviews, incluye una opción automática para esta selección.

**Ejemplo 3.12.** Se considera nuevamente la SIB (capítulo 2), que se consideró que se debía diferenciar, a partir del gráfico de la serie y de su FAC estimada. La información estadística sobre la prueba de raíces unitarias más adecuada aparece en la Tabla 3.8 (modelo (b)), en la que se observa que con los niveles de significación usuales (10 %, 5 % y 1 %) se debe aceptar la presencia de una raíz unitaria; sin embargo, el procedimiento aplicado a la serie diferenciada muestra que ésta ya no admite una raíz unitaria (la SIB es, entonces,  $I(1)$ ). Ver la Tabla 3.9.

**Tabla 3.8:** Información estadística para la prueba ADF de la SIB

	<i>t</i> -Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.112915	0.7096
Test critical values:		
1 % level	-3.479656	
5 % level	-2.883073	
10 % level	-2.578331	

\*MacKinnon (1996) one-sided *p*-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(SIB)

Method: Least Squares

Date: 05/20/13 Time: 13:14

Sample (adjusted): 8 141

Included observations: 134 after adjustments

Variable	Coefficient	Std. Error	<i>t</i> -Statistic	Prob.
SIB(-1)	-0.003310	0.002974	-1.112915	0.2679
D(SIB(-1))	1.466814	0.084727	17.31223	0.0000
D(SIB(-2))	-1.214989	0.146390	-8.299676	0.0000
D(SIB(-3))	0.930050	0.165351	5.624695	0.0000
D(SIB(-4))	-0.816091	0.165477	-4.931760	0.0000
D(SIB(-5))	0.514208	0.145415	3.536147	0.0006
D(SIB(-6))	-0.243578	0.082814	-2.941259	0.0039
C	0.375276	0.165658	2.265358	0.0252



Tabla 3.9: Información estadística para la prueba ADF de la SIB diferenciada

	<i>t</i> -Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.820465	0.0000
Test critical values: 1 % level	-3.479656	
5 % level	-2.883073	
10 % level	-2.578331	

\*MacKinnon (1996) one-sided *p*-values.

Augmented Dickey-Fuller Test Equation

Dependent Variable: D(SIB,2)

Method: Least Squares

Date: 05/20/13 Time: 13:15

Sample (adjusted): 8 141

Included observations: 134 after adjustments

Variable	Coefficient	Std. Error	<i>t</i> -Statistic	Prob.
D(SIB(-1))	-0.361056	0.062032	-5.820465	0.0000
D(SIB(-1),2)	0.834828	0.083637	9.981516	0.0000
D(SIB(-2),2)	-0.387908	0.095122	-4.078010	0.0001
D(SIB(-3),2)	0.544866	0.099687	5.465761	0.0000
D(SIB(-4),2)	-0.274558	0.090012	-3.050232	0.0028
D(SIB(-5),2)	0.240218	0.082837	2.899894	0.0044
C	0.224892	0.095922	2.344538	0.0206

### 3.8 Modelos Estacionales

Se conoce que, por ejemplo, ciertas series mensuales tienen un perfil estacional marcado; es decir, los datos relativos a un mismo mes de diferentes años tienen tendencia a situarse de manera similar respecto a la media anual. Esto hace pensar en incluir en el modelo ARIMA retardos múltiples de 12. En teoría, nada impide el incluir parámetros  $p, q$  lo suficientemente grandes para que esos retardos sean tomados en cuenta; sin embargo, su cálculo se convertiría en una tarea prácticamente imposible. Para evitar este aumento drástico de parámetros, Box y Jenkins propusieron un tipo particular de modelos ARIMA estacionales, que son modelos multiplicativos del tipo:

$$\Delta^d \Phi_p(B) \Delta_S^D \Phi_P(B^S) X_t = \Theta_q(B) \Theta_Q(B^S) u_t$$

donde  $S$  es el periodo de la estacionalidad ( $S = 12$  para series mensuales,  $S = 4$  para series trimestrales, ...).  $\Delta = 1 - B$ ,  $\Delta_S = 1 - B^S$ ;  $\Phi_p, \Phi_P, \Theta_q, \Theta_Q$  son polinomios de grado  $p, P, q$  y  $Q$ , respectivamente, con raíces de módulo superior al 1;  $(u_t)$  es un ruido blanco.

Un proceso  $X_t$  que satisface la ecuación precedente se llama un proceso  $SARIMA_S[(p, d, q)(P, D, Q)]$ . La justificación de este modelo consiste en aplicar a las  $S$  series obtenidas a partir de  $X_t$ , para los meses idénticos, la misma transformación:

$$\frac{\Delta_S^D \Phi_P(B^S)}{\Theta_Q(B^S)}$$

y suponer que la serie obtenida:

$$\alpha_t = \frac{\Delta_S^D \Phi_P(B^S)}{\Theta_Q(B^S)} X_t$$

ya no es estacional y se la puede modelar por un  $ARIMA(p, d, q)$ .

$$\Delta^d \Phi_p(B) \alpha_t = \Theta_q(B) u_t$$

Las series estacionales pueden detectarse examinando las funciones de autocorrelación y autocorrelación parcial estimadas pues éstas toman valores altos, en valor absoluto, en los índices múltiplos de  $S$ .

La identificación de los parámetros  $P$ ,  $D$ ,  $Q$  de la parte estacional se hace de manera análoga a lo descrito anteriormente, tomando en cuenta para las  $\hat{\rho}(h)$  y  $\hat{r}(h)$ , los valores de  $h$  múltiplos de  $S$ .

Las fases de estimación y verificación se realizan de la misma manera que para los modelos no estacionales.

**Ejemplo 3.13** (Un ejemplo con estacionalidad). Los datos que se utilizarán (ver Anexo C.3) corresponden a las ventas mensuales (SVM) en dólares, de una compañía comercial en el período comprendido desde enero de 2006 hasta diciembre de 2015 (Figura No. 3.15). Se modelará la serie con los datos hasta el año 2014 y se guardarán los del año 2015 para poder compararlos con las predicciones. Siguiendo la metodología explicada se procede a la identificación a priori del modelo.

**Figura 3.15:** Grafico de SVM

### Identificación a Priori

**Elección de  $d$ .** La modelación de esta serie se realizará utilizando el paquete Eviews.

Se observa que los primeros valores de la función de autocorrelación (Figura 4.16) no son significativamente grandes, por lo cual no es necesario realizar una diferenciación no estacional de la serie ( $d = 0$ ).

**Elección de  $D$ .** Los valores alrededor de 12 y 24 son grandes (Figura 4.16), por lo que se requiere una diferenciación estacional de orden 12; es decir,  $D = 1$ .

**Figura 3.16:** FAC y FACP estimadas de SVM

Luego de observar la serie diferenciada no estacionalmente de orden 12, de acuerdo a lo detallado, se observa que no hace falta una segunda diferenciación estacional.

**Observación 3.11.** Si existiera la posibilidad de diferenciar estacionalmente y no estacionalmente, siempre se debe realizar primero la diferencia estacional, pues hay que reconocer que el operador de diferenciación no estacional  $\Delta = 1 - B$  es un factor del operador de diferencia estacional  $\Delta^S = 1 - B^S$ ; es decir, al diferenciar estacionalmente a la serie, ya se la diferencia no estacionalmente.

**Observación 3.12.** Un error frecuente en que se suele incurrir es pensar que la prueba ADF o similares, que determinan la existencia de raíces unitarias (no estacionales), sirven para detectar estacionalidad; esto no es correcto, pues la estacionalidad es un fenómeno más complejo.

**Figura 3.17:** FAC Y FACP estimadas de la SVM con una diferencia estacional

### 3.8.0.1 Especificación del Modelo

En la Figura 3.17 de la serie diferenciada estacionalmente de orden 12, se observa que la autocorrelación estimada de orden 13 se sale de la banda de confianza (95 %), con lo que se escoge un coeficiente  $MA(13)$ ; de la misma manera, se observa que las autocorrelaciones parciales de orden 12 y 13 son significativamente distintas de cero, por lo que se eligen coeficientes  $AR(12)$  y  $AR(13)$

Se puede entonces considerar como modelo inicial, que se llamará Modelo 1, al siguiente:

$$SVMDE(t) = c_1 + c_2 \cdot SVMDE(t - 12) + c_3 \cdot SVMDE(t - 13) + c_4 \cdot u(t - 13)$$

donde SVMDE representa a la serie de ventas mensuales con una diferencia estacional. El modelo ingresa a un proceso iterativo de estimación, verificación y modificación (de ser el caso).

**Tabla 3.10:** Información estadística del Modelo 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	152.7693	28.75136	5.313462	0.0000
AR(12)	-0.284706	0.118481	-2.402962	0.0186
AR(13)	0.171932	0.235285	0.730737	0.4671
MA(13)	0.260089	0.253103	1.027601	0.3073
R-squared	0.172040	Mean dependent var	151.3253	
Adjusted R-squared	0.140599	S.D. dependent var	254.2020	
S.E. of regression	235.6552	Akaike info criterion	13.80961	
Sum squared resid	4387135.	Schwarz criterion	13.92618	
Log likelihood	-569.0988	Hannan-Quinn criter.	13.85644	
F-statistic	5.471747	Durbin-Watson stat	1.337219	
Prob(F-statistic)	0.001811			

### 3.8.0.2 Verificación

**Pruebas para los Parámetros.** Con el software Eviews, se realizó la estimación del modelo inicial, como se muestra en la Tabla 3.10; analizando la última columna de los  $p$ -valores se concluye que no todos los coeficientes tienen un  $p$ -valor menor a 0,05 (95 % de confianza), por lo cual no son estadísticamente significativos (diferentes de cero). Es el caso de los coeficientes asociados a  $AR(13)$  y  $MA(13)$ .

Por lo tanto, se debe modificar el modelo. Así, se deben eliminar los coeficientes que no son significativos; no es recomendable eliminar más de un coeficiente a la vez, sino de uno en

uno. En la tabla 3.11, se pueden ver los resultados obtenidos luego de quitar el coeficiente  $AR(13)$  (el que tiene el  $p$ -valor más grande); se denominará Modelo 2:

**Tabla 3.11:** Información estadística del Modelo 2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	154.4945	26.07034	5.926062	0.0000
$AR(12)$	-0.348977	0.119777	-2.913542	0.0046
$MA(13)$	0.408495	0.117771	3.468550	0.0008
R-squared	0.193596	Mean dependent var	157.1548	
Adjusted R-squared	0.173685	S.D. dependent var	258.2531	
S.E. of regression	234.7569	Akaike info criterion	13.79004	
Sum squared resid	4463975.	Schwarz criterion	13.87685	
Log likelihood	-576.1816	Hannan-Quinn criter.	13.82494	
F-statistic	9.722955	Durbin-Watson stat	1.285551	
Prob(F-statistic)	0.000164			

Como se puede ver en la Tabla 3.11, los coeficientes tienen un  $p$ -valor menor que 0,05, por lo cual son significativos. A continuación, se presentan las autocorrelaciones y autocorrelaciones parciales estimadas de los residuos del Modelo 2.

**Figura 3.18:** FAC y FACP residuales estimadas del Modelo 1 para SVM.

A partir de la Figura 3.18, se deduce que se debería modificar el Modelo 2, pues aunque pasa la prueba de los coeficientes, los residuos están correlacionados (se constata con la observación de los gráficos y del estadístico  $Q$ , con sus correspondientes  $p$ -valores, todos menores que 0,05).

### Modificación del Modelo

La Figura 3.18 muestra que las autocorrelaciones y autocorrelaciones parciales estimadas son significativas al orden 1, lo que sugiere modificar el modelo aumentando un parámetro no estacional auto-regresivo y/o media móvil. Los resultados fueron los siguientes:

Se aumentó un coeficiente para  $AR(1)$  que resultó significativo (se denominará Modelo 3); además los residuos se pueden considerar ruido blanco (todos los  $p$ -valores son mayores que 0,05).

**Tabla 3.12:** Información estadística del Modelo 3 para la SVM

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	156.1661	34.82541	4.484257	0.0000
AR(1)	0.325909	0.107494	3.031876	0.0033
AR(12)	-0.335945	0.114892	-2.924008	0.0045
MA(13)	0.480540	0.111556	4.307611	0.0000
R-squared	0.278803	Mean dependent var	157.1548	
Adjusted R-squared	0.251758	S.D. dependent var	258.2531	
S.E. of regression	223.3914	Akaike info criterion	13.70218	
Sum squared resid	3992299.	Schwarz criterion	13.81793	
Log likelihood	-571.4914	Hannan-Quinn criter.	13.74871	
F-statistic	10.30889	Durbin-Watson stat	1.912194	
Prob(F-statistic)	0.000008			

**Figura 3.19:** FAC y FACP estimadas del Modelo 3 para la SVM

### Prueba de predicciones de Chow

A continuación, se realiza la prueba de estabilidad de las predicciones (12 en este caso). Así, se obtienen los siguientes resultados:

**Figura 3.20:** Prueba de Chow para las predicciones.

Dado que el  $p$ -valor correspondiente al estadístico F es menor que 0,05, se rechaza la hipótesis nula de no existencia de cambios estructurales en las predicciones. Sin embargo, este fenómeno puede deberse a la estacionalidad de la serie y a un efecto de “volatilidad” que se estudiará en el siguiente capítulo.

### Predicciones

Como se puede ver, el Modelo 3 es el mejor modelo que describe los datos de ventas mensuales; se lo utilizará para comparar los valores de predicción del modelo retenido con los verdaderos valores para el año 2015. Para confirmar que el modelo retenido realiza buenas predicciones a corto plazo, se grafican éstas con los 12 datos correspondientes al año 2015 (Figura 3.17); también aparecen los intervalos de confianza correspondientes (nivel de confianza 95 %).

**Figura 3.21:** Comparación de predicciones con valores reales (2015)

## 3.9 Ejercicios Propuestos

1. Sea  $X = (X_n)_{n \in \mathbb{Z}}$  proceso estacionario, centrado que satisface la ecuación ARMA:

$$9X_n - X_{n-2} = W_n + 3W_{n-1}$$

donde  $(W_n)$  es un r.b. de varianza  $\sigma^2$ .

Calcular explícitamente el predictor lineal  $\hat{X}_{n+1}$  de  $X_{n+1}$  conociendo el pasado  $\{X_p, p \leq n\}$ , luego  $\hat{X}_{n+2}$  y mostrar como proseguir por recurrencia el cálculo de  $\hat{X}_{n+3}$ ,  $\hat{X}_{n+4}$ , etc.

2. Sea el siguiente modelo ajustado con una muestra de 64 observaciones

$$(1 - 0,8B)Y_t = (1 + 0,5)\hat{u}_t$$

La FAC y la FACP estimadas de los residuos vienen dadas por:

Coeficiente	Retardo				
	1	2	3	4	5
$\hat{r}_k$	0,38	0,16	0,05	0,02	0,01
$\hat{\rho}_k$	0,38	0,03	0,07	0,02	0,03

?Consideraría aceptable el modelo ajustado? En caso contrario, ?Cómo podría reformular el modelo? Justifique cada paso de su procedimiento.

3. Se ha estimado con una muestra de 110 observaciones del proceso  $(X_t)$  con el siguiente modelo:

$$X_t = u_t + 0,5u_{t-1} + 0,4u_{t-2} + 18; \quad \sigma_u^2 = 4$$

La información acerca de  $u_t$  y  $X_t$  se ha perdido, excepto para los siguientes valores de  $X_t$ :

$$X_{106} = 20 \quad X_{107} = 21 \quad X_{108} = 19 \quad X_{109} = 19 \quad X_{110} = 17$$

Bajo este supuesto:

- Efectuar la predicción de  $X_{t+h}$  para los períodos 111, 112 y 113 con origen en  $t = 110$ .
  - Calcular el EMC para las anteriores predicciones.
  - Suponiendo que se dispone de una observación adicional:  $X_{111} = 17$ , actualizar la predicción de los períodos 112 y 113.
4. Sea el modelo estimado:

$$(1 + 0,5B)(1 - B)X_t = (1 - 0,5B)u_t + 0,8$$

con  $\sigma^2 = 1$ . Se pide:

- Dado  $X_{79} = 40$ ,  $X_{80} = 41$ ,  $u_{80} = 0,2$ . Calcular:  $\hat{X}_{80}(1)$ ,  $\hat{X}_{80}(2)$  y  $\hat{X}_{80}(3)$ .
  - Calcular el error medio cuadrático de las predicciones.
  - Establecer una banda de confianza del 90 % para las predicciones.
5. Sea el proceso  $(X_t)_{t \in \mathbb{Z}}$ , definido por:

$$(1 - 4B)X_t = 3 + (1 - 0,5B)u_t$$

donde  $(u_t)$  r.b. de varianza 1.

- (a) Encuentre la covarianza entre  $X_t$  y  $u_{t-i}$ ,  $i \in \mathbb{Z}$
- (b) ¿Cuál es la función de autocorrelación del proceso?
- (c) Expresar  $X_t$  en función de los  $u_{t-i}$ ,  $i \in \mathbb{Z}$ .
- (d) Expresar el proceso en su forma canónica (y centrado). ¿Cuál es la expresión como proceso lineal?
- (e) ¿Cuál es la varianza del error de predicción con horizonte 1?

6. Sea  $f(\lambda) = \frac{\frac{13}{12} + \cos \lambda}{\frac{41}{40} + \cos \lambda} \cdot \frac{1}{2\pi}$ ,  $\lambda \in [-\pi, \pi]$ .

- (a) ¿Cuál es la ecuación ARMA canónica satisfecha por  $X$ ? ¿Cuál es la varianza de la innovación?
- (b) Dar una representación explícita de la innovación  $u_t$  en función de las  $X_j$ ,  $j \leq t$ .
- (c) Calcular  $\hat{X}_{T+1}$ , conociendo  $X_t$  ( $t \leq T$ ). Determinar el error de predicción.

¿Existe autocorrelación entre estas observaciones?

7. Suponga que se ha estimado un modelo  $ARIMA(0, 2, 1)$  con parámetro 0,30 para representar la serie trimestral de la población activa de un país, cuyos valores durante el año 2014 fueron:

Trimestre	(2014, I)	(2014, II)	(2014, III)	(2014, IV)
Población activa (Miles de personas)	14.000	14.150	14.200	14.350

Elabore predicciones para la población activa durante los cuatro trimestres de 2015, sabiendo que la predicción que se había hecho en (2014, III) para (2014, IV) era de 14.500.

8. Se considera el modelo  $(1 - B)(1 - 0,5B)X_t = (1 - 0,8B)u_t$ , donde  $(u_t)$  es un r.b. normal, con varianza 0,25. Se supone que  $X_T = 12$  y  $\hat{X}_T(1) = 10$ . Dar un intervalo de predicción, de nivel de confianza 95%, para  $X_{T+1}$ .

$t$	1	2	3	4	5	6	7	8	9	10
$X_t$	7	9	10	6	5	3	1	2	3	4

- (a) Calcular  $\hat{\rho}_1$ ,  $\hat{\rho}_2$  y  $\hat{\rho}_3$
- (b) Encontrar un intervalo de confianza ( $\alpha = 5\%$ ) para  $\hat{\rho}_1$ ,  $\hat{\rho}_2$  y  $\hat{\rho}_3$
- (c) Calcular  $\hat{r}_1$ ,  $\hat{r}_2$  y  $\hat{r}_3$ .
- (d) Encontrar un intervalo de confianza ( $\alpha = 5\%$ ) para  $\hat{r}_1$ ,  $\hat{r}_2$  y  $\hat{r}_3$

9. Sea el modelo  $ARIMA(1, 1, 1)$ :  $(1 - \phi B)\Delta X_t = \theta_0 + (1 - \theta B)u_t$ .

- (a) Calcular  $\hat{X}_t(h)$ ,  $h = 1, 2$
- (b) Dar una fórmula para  $\hat{X}_t(h)$ ,  $h \geq 2$

10. Sea el proceso  $(X_t)$  definido por  $X_t = u_t - 0,6u_{t-1}$ .

(a) Calcular  $S = \sum u_t^2$  para las observaciones:

$t$	1	2	3	4	5
$X_t$	3	2	0	2	1

condicionado al valor de  $u_0$ .

(b) Calcular el valor de  $u_0$  que minimiza  $S$ .



## Capítulo 4

# Modelos de Heteroscedasticidad Condicional

### 4.1 Modelos Arch–Garch Simétricos

En esta sección, se presentará un breve resumen de la teoría desarrollada alrededor de los modelos ARCH – GARCH, los cuales determinan un patrón de comportamiento estadístico para la varianza condicional, denominados modelos Auto-Regresivos con Heteroscedasticidad Condicional.

El análisis de series temporales económicas, tradicionalmente se ha centrado en el estudio de modelos para la media condicional en los que se asume que la varianza condicional es constante; en este caso se dice que existe homoscedasticidad. Si este no fuera el caso, se estaría enfrentado un problema de no estacionariedad de la serie.

En 1982, Robert Engle revolucionó los modelos de volatilidad introduciendo el estudio de las estructuras cuadráticas, ampliando así la visión de la metodología Box–Jenkins, en la cual los modelos lineales de tipo ARIMA admiten que las innovaciones son un ruido blanco, con media cero y varianza constante.

Los modelos de volatilidad condicional son importantes por el papel que juega el riesgo y el concepto de incertidumbre en el desarrollo de las teorías modernas de modelos financieros, que relacionan de forma directa el riesgo con la volatilidad. Estos modelos permiten relacionar el valor de la varianza condicional (no constante) en función del conjunto de información disponible en periodos anteriores; están específicamente diseñados para modelar y pronosticar varianzas condicionales.

**Definición 4.1.** *Un Modelo ARCH( $r$ ) se define por:*

$$\begin{aligned}Z_t &= \sqrt{h_t}u_t \\h_t &= \alpha_0 + \alpha_1 Z_{t-1}^2 + \dots + \alpha_r Z_{t-r}^2\end{aligned}$$

donde los  $(u_t)$  son independientes e idénticamente distribuidos con media cero y varianza

uno (i.i.d.  $(0, 1)$ ),  $\alpha_0 > 0$  y  $\alpha_i \geq 0$  para  $i > 0$ .

En la práctica se supone que  $u_t \sim N(0, 1)$  o que siguen una distribución t-student. Los coeficientes  $\alpha_i$  deben satisfacer ciertas condiciones (en general no negatividad:  $\alpha_0 > 0$  y  $\alpha_i \geq 0$  para  $i > 0$  y  $\sum_{i=1}^r \alpha_i < 1$ ), dependiendo del tipo de restricciones que se coloquen sobre el proceso  $Z_t$ .

Las restricciones de signo de los coeficientes de la ecuación de varianza garantizan que la varianza condicional sea positiva en todos los períodos. Lo que se necesita es que, una vez que se haya estimado el modelo, genere una serie de varianzas positiva, lo que puede suceder aún si algunos de los coeficientes  $\alpha_i$  fueran negativos. Esto se puede considerar como un contraste de validez del modelo.

Por la propia definición, a valores grandes de  $Z_t$  les siguen otros valores grandes de la serie.

$h_t$  es la varianza condicional de  $Z_t$  dado  $Z_s$ ,  $s < t$ .

**Observación 4.1.** Se puede demostrar que un proceso  $ARCH(r)$  implica que se puede representar  $X_t^2$  como un proceso  $AR(r)$ , con residuos que no son gaussianos. Además, el coeficiente de apuntamiento (curtosis) es mayor que 3, por lo cual las colas de la distribución serán más pesadas que en la distribución normal.

**Definición 4.2.** Un modelo  $GARCH(r, s)$ , está definido por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \sum_{j=1}^s \beta_j h_t$$

donde los  $u_t$  son i.i.d.  $(0, 1)$  con  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$  para  $i > 0$  y

$$\sum_{i=1}^r \alpha_i + \sum_{j=1}^s \beta_j < 1$$

Como en el caso de los modelos  $ARCH$ , usualmente se supone que los  $u_t$  son normales o siguen una distribución t-student.

Las restricciones de no negatividad impuestas sobre los coeficientes en la definición son para asegurarse que la varianza no llegue a tomar un valor negativo. Además, en algunos casos se pueden encontrar condiciones menos restrictivas para estos coeficientes, que aseguren que la varianza sea positiva.

**Observación 4.2.** Se puede demostrar que un proceso  $GARCH(r, s)$  implica que se puede representar  $Z_t^2$  como un proceso  $ARMA(r, s)$ , con residuos que no son gaussianos (más precisamente con residuos que representan una diferencia martingala). También, en este caso, el coeficiente de apuntamiento (curtosis) es mayor que 3, por lo cual las colas de la distribución serán más pesadas que en la distribución normal.

## Modelos IGARCH

Este modelo fue descrito originalmente por Engle y Bollerslev (1986). Si el modelo polinomial AR del modelo GARCH tiene una raíz unitaria, se tiene un modelo IGARCH. Los modelos IGARCH son modelos GARCH con una raíz unitaria. Estos modelos tienen la característica que los impactos de los choques al cuadrado sobre  $Z_t^2$  son persistentes; esto se puede expresar por:

$$\eta_{t-i} = Z_{t-i}^2 - h_{t-i}, \quad \text{para } i > 0$$

**Definición 4.3.** Un modelo IGARCH( $r, s$ ), se define por:

$$\begin{aligned} Z_t &= \sqrt{h_t} u_t \\ h_t &= \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \sum_{j=1}^s (1 - \beta_j) h_{t-j} \end{aligned}$$

tal que,

$$\sum_{j=1}^s \beta_j + \sum_{i=1}^r \alpha_i = 1$$

Como en el caso de los modelos ARCH, usualmente se supone que los  $u_t$  son normales o siguen una distribución *t*-student.

Los modelos IGARCH son un caso específico dentro de la familia de los “modelos con varianza persistente” en los que la información actual (en el instante “ $t$ ”) es importante para realizar predicciones óptimas con cualquier horizonte temporal.

## 4.2 Modelos GARCH asimétricos

Una característica de los modelos GARCH, dado que la varianza depende fundamentalmente de valores cuadráticos pasados, es que la volatilidad que generan frente a cambios positivos o negativos inesperados de la variable (piénsese en retornos de inversiones), dan una respuesta simétrica a estos. Sin embargo, se ha probado empíricamente que la reacción que tiene la volatilidad o varianza condicional de muchas variables financieras a este tipo de cambios es asimétrica; es decir, existe diferencia en la respuesta de la volatilidad de la variable, dependiendo de si el cambio es positivo o negativo.

Con el fin de modelar esta respuesta asimétrica se han desarrollado una variedad de modelos asimétricos; los más representativos son los modelos EGARCH, TARCH, PARCH.

## Modelos EGARCH

Las restricciones de no negatividad para asegurar la positividad de la varianza en los modelos GARCH en muchas ocasiones son difíciles de lograr. Nelson (1991) propuso los modelos EGARCH (*Exponential GARCH*), como solución a este problema; éstos, además, incorporan efectos asimétricos.

**Definición 4.4.** Un modelo  $EGARCH(r, s)$ , se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$\ln(h_t) = \alpha_0 + \sum_{j=1}^s \beta_j \ln(h_{t-j}) + \sum_{i=1}^r (\alpha_i |u_t| + \gamma_i u_t)$$

Nótese que el lado izquierdo de la ecuación es el logaritmo de la varianza condicional; esto implica que su efecto es exponencial y garantiza que las predicciones de ésta serán no negativas. Cuando  $\gamma_i \neq 0$ , el efecto asimétrico deber incorporarse al Modelo GARCH.

Obsérvese también que se puede escribir  $\frac{Z_t}{\sqrt{h_t}}$  en lugar de  $u_t$ , en las fórmulas anteriores.

Como en el caso de los modelos ARCH, usualmente se supone que los  $u_t$  son normales o siguen una distribución  $t$ -student.

## Modelos TARCH

Los modelos TARCH (*Threshold ARCH*) fueron introducidos independientemente por Zakoian (1990) y Glosten, Jaganathan y Runklen (1993), por lo que también se conocen como GJR-GARCH. Estos modelos incluyen una variable adicional  $d_t$ , que determina el carácter asimétrico del modelo.

**Definición 4.5.** Un modelo  $TARCH(r, s)$ , se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i Z_{t-i}^2 + \gamma Z_{t-1}^2 d_{t-1} + \sum_{j=1}^s \beta_j h_{t-j}$$

donde los  $u_t$  son i.i.d.  $(0, 1)$  con  $\alpha_0 > 0$ ,  $\alpha_i \geq 0$ ,  $\beta_j \geq 0$  para  $i, j > 0$  y  $\gamma \neq 0$

$$d_t = \begin{cases} 1, & \text{si } Z_t < 0 \\ 0, & \text{si } Z_t \geq 0 \end{cases}$$

Si  $\gamma = 0$  se pierde el efecto asimétrico del modelo.

En este modelo, las malas noticias ( $Z_t < 0$ ) y las buenas noticias ( $Z_t \geq 0$ ) (piénsese otra vez en retornos), tienen efectos diferentes sobre la varianza condicional.

## Modelos PARCH

Los modelos PARCH (*Power ARCH*) desarrollados independientemente por Taylor (1986) y Schwert (1989), introducen la desviación estándar a los modelos ARCH; donde se modela la desviación estándar en lugar de la varianza. Este modelo fue generalizado por Ding y otros (1993).

En el modelo PARCH, el parámetro de potencia  $\delta$  de la desviación estándar puede ser estimado antes que impuesto y los parámetros opcionales  $\gamma$  se agregan para capturar la asime-

tría dentro de los datos.

**Definición 4.6.** Un modelo  $PARCH(r, s)$ , se define por:

$$Z_t = \sqrt{h_t} u_t$$

$$h_t^\delta = \alpha_0 + \sum_{i=1}^r \alpha_i (|u_{t-i}| - \gamma_i u_{t-i})^\delta + \sum_{j=1}^s \beta_j h_{t-j}^\delta$$

donde  $\delta > 0$ , es el parámetro del término de la potencia.  $\gamma_i$  se dicen los parámetros de apalancamiento.

En series de valores simétricos  $\gamma_i = 0$  para todo  $i$ . Nótese que si  $\delta = 1$  y  $\gamma_i = 0$  para todo  $i$ , el modelo  $PARCH$  es simplemente una especificación  $GARCH$  estándar. Si los  $\gamma_i = 0$  se pierde el efecto asimétrico del modelo.

Los modelos  $GARCH$  asimétricos, se estiman por el método de máxima verosimilitud condicional, por lo cual se requiere de ciertos supuestos acerca del comportamiento de los errores. Por lo general, se suponen i.i.d con distribución normal o incluso con una distribución  $t$ -student.

### 4.3 Metodología de la Modelición

El objetivo es encontrar un modelo que represente adecuadamente a los datos históricos de una determinada variable, combinando especificaciones tanto para la media como para la varianza condicional. Los tipos de modelos que se considerarán serán los  $ARIMA - GARCH$ , de tal manera que la media condicional de la serie sea descrita por un modelo del tipo  $ARIMA$  y su varianza condicional por uno de la familia de modelos  $ARCH - GARCH$  o de sus extensiones asimétricas  $PARCH$ ,  $TARCH$  y  $EGARCH$ . La modelación se realizará utilizando el paquete  $EViews$ .

El primer paso es, por tanto, modelar la serie de datos por un modelo del tipo  $ARIMA$  o incluso  $SARIMA$ , con lo que se obtiene un modelo para la media condicional de la serie.

Luego de haberse eliminado toda correlación lineal en la serie, se debe indagar si existe heteroscedasticidad condicional residual, para lo cual deben analizarse los residuos estandarizados estimados al cuadrado; el correleograma correspondiente, permite llevar a cabo un análisis gráfico de identificación, para ver si algún valor es estadísticamente diferente de cero, y por tanto, existe autocorrelación en su forma residual cuadrática.

Si se verifica la existencia de heteroscedasticidad condicional en los residuos, se rechaza el supuesto de la varianza constante; se intentará entonces obtener una especificación para la varianza condicional, a través de la modelación de los residuos estimados obtenidos por el modelo  $ARIMA$ , mediante un modelo del tipo  $ARCH - GARCH$  o sus extensiones asimétricas.

Inicialmente se mantiene la estructura para la media condicional, obtenida por el modelo  $ARIMA$ , pero esta puede modificarse con la nueva especificación. Los residuos estimados

deben analizarse, tanto en su forma simple como en la cuadrática, para eliminar toda evidencia de autocorrelación lineal (deben aceptarse como un ruido blanco).

La estimación y verificación permiten encontrar uno o varios modelos que cumplan las condiciones que se impusieron en la modelación ARIMA; es decir, todos los coeficientes deben ser significativos; las raíces de los polinomios característicos, tanto de la parte auto-regresiva como de la media móvil, deben estar fuera del círculo unidad, para así asegurar la estacionariedad e invertibilidad del proceso. Además, los coeficientes de la ecuación de la varianza condicional deben satisfacer las restricciones de no negatividad para la varianza (modelos ARCH – GARCH).

Para la verificación de la presencia de una estructura ARIMA en los residuos (simples o cuadráticos) pueden utilizarse la FAC y la FACP; además, también se debe realizar la prueba global (estadístico  $Q$ ) de Box – Pierce – Ljung.

Una vez que un modelo ha sido estimado y ha superado las diversas verificaciones, se convierte en un instrumento útil para las predicciones de valores futuros. Como en la modelación ARIMA, si varios modelos son plausibles, se elige entre estos al mejor, mediante los criterios ya citados previamente.

## 4.4 Ejemplos con Heteroscedasticidad Condicional

Aunque los datos de las ventas que se vienen utilizando no corresponden al ámbito financiero, sirven muy bien para ilustrar la modelación para la varianza condicional. En esta ocasión se adoptará el Modelo 3 con el cual se modeló la media condicional (SARIMA). La Figura 4.2 no permite aceptar la hipótesis de que la serie tenga una varianza constante.

Una posibilidad para amortiguar los efectos de varianza no constante es utilizar la transformación logaritmo o, en general, la transformación de Box y Cox; sin embargo, en esta ocasión se tratará de modelar directamente la varianza a través de los Modelos ARCH-GARCH o sus extensiones asimétricas.

En la Tabla 4.1 y en las figuras 4.1 y 4.2 se presentan la información estadística y residual para el Modelo 3 de la SVM:

Variable	Coefficient	Std. Error	<i>t</i> -Statistic	Prob.
C	156.1661	34.82541	4.484257	0.0000
AR(1)	0.325909	0.107494	3.031876	0.0033
AR(12)	-0.335945	0.114892	-2.924008	0.0045
MA(13)	0.480540	0.111556	4.307611	0.0000

**Tabla 4.1:** Información sobre los coeficientes del Modelo 3 para la SVM

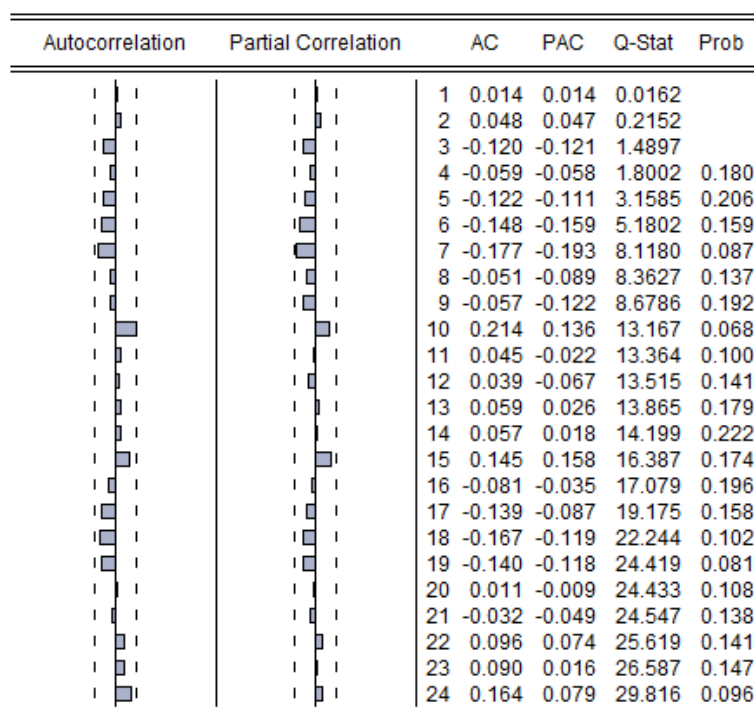


Figura 4.1: FAC y FACP estimadas residuales del Modelo 3 para la SVM

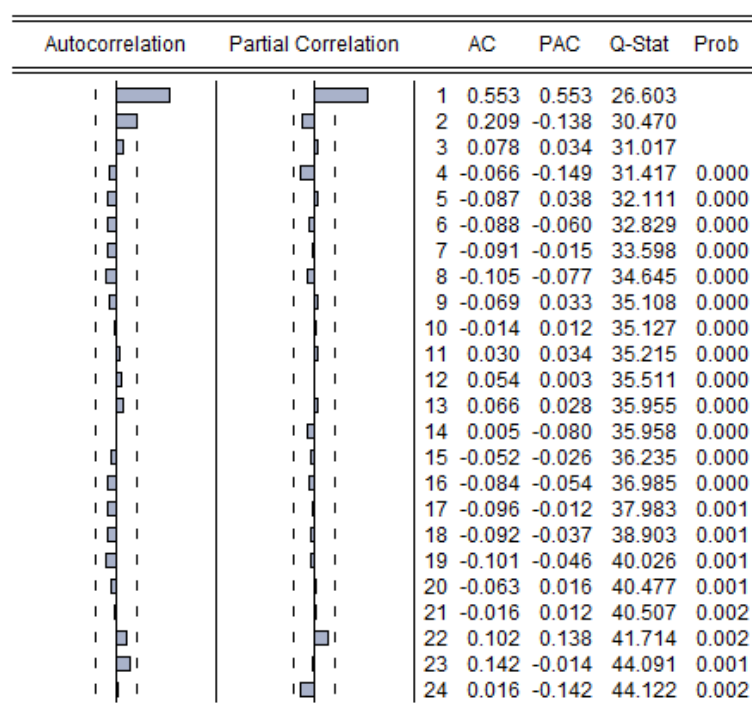


Figura 4.2: FAC y FACP estimadas de los residuos cuadráticos del Modelo 3 para la SVM

Se observan fuertes correlaciones entre los residuos cuadráticos estandarizados estimados, por lo cual se hace necesaria la modelación de la varianza condicional del Modelo 3.

En general, es difícil establecer el orden para los modelos ARCH-GARCH. Lo usual es probar los modelos con parámetros (1,0), (1,1), (1,2) o (2,2). Para este caso se empezó probando con el modelo  $ARCH(1)$ ; los resultados aparecen en la Tabla 4.2 y en las Figuras 4.3 y 4.4.

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	115.3958	36.57140	3.155355	0.0016
$AR(1)$	0.483776	0.103019	4.695993	0.0000
$AR(12)$	-0.449722	0.078830	-5.704935	0.0000
$MA(13)$	0.798959	0.041313	19.33915	0.0000
Variance Equation				
C	16129.50	3530.889	4.568113	0.0000
$RESID(-1)^2$	0.610436	0.249248	2.449115	0.0143

R-squared	0.335639	Mean dependent var	157.1548	
Adjusted R-squared	0.310726	S.D. dependent var	258.2531	
S.E. of regression	214.4082	Akaike info criterion	13.23093	
Sum squared resid	3677671.	Schwarz criterion	13.40456	
Log likelihood	-549.6992	Hannan-Quinn criter.	13.30073	
Durbin-Watson stat	2.201103			
Inverted AR Roots	.96 - .24i	.96 + .24i	.71 + .65i	.71 - .65i
	.28 - .89i	.28 + .89i	-.21 + .90i	-.21 - .90i
	-.63 - .66i	-.63 + .66i	-.87 + .24i	-.87 - .24i
Inverted MA Roots	.95-.24i	.95+.24i	.74-.65i	.74+.65i
	.35 + .92i	.35 - .92i	-.12 - .98i	-.12 + .98i
	-.56 - .81i	-.56 + .81i	-.87 + .46i	-.87 - .46i
	.98			

**Tabla 4.2:** Información estadística para el Modelo 3-ARCH(1) para la SVM



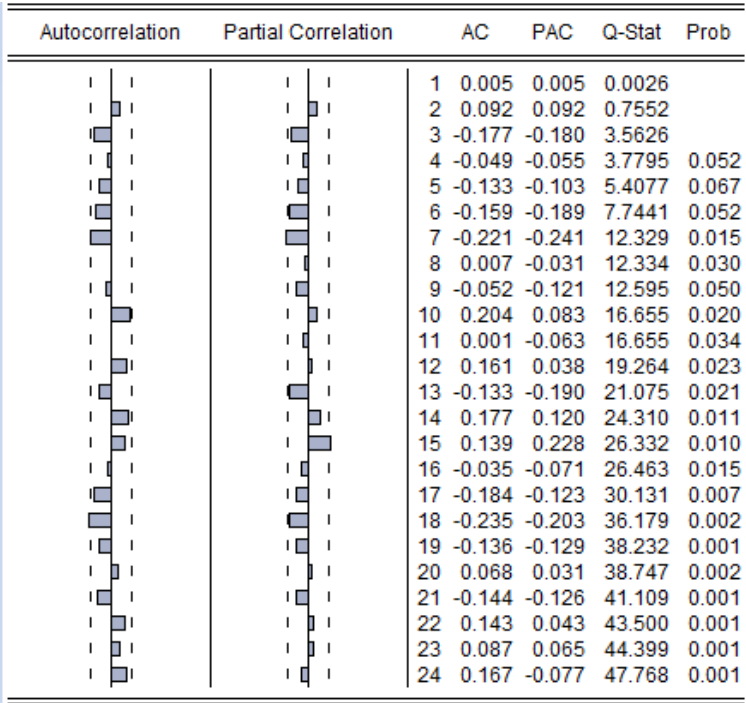


Figura 4.3: FAC y FACP estimadas residuales del Modelo 3-ARCH(1) para la SVM

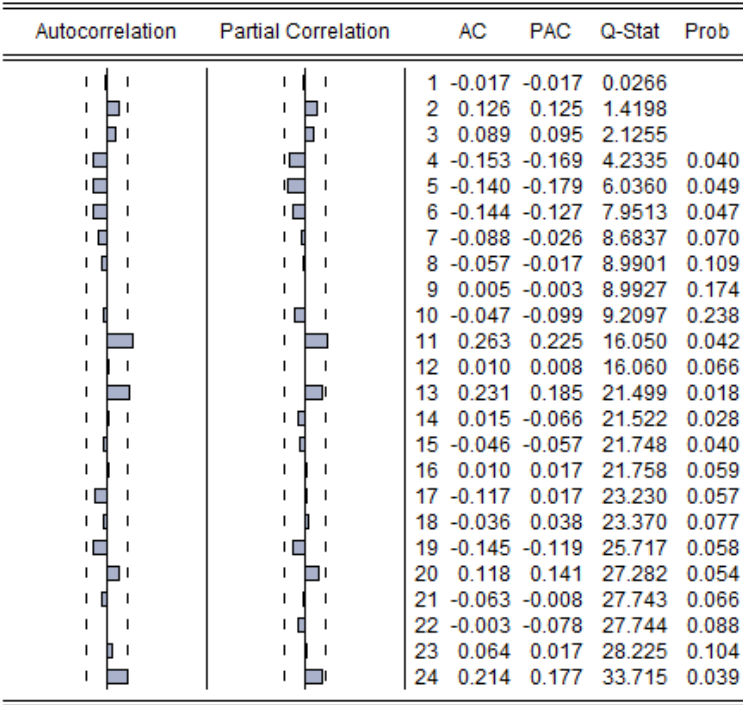


Figura 4.4: FAC y FACP estimadas de los residuos cuadráticos del Modelo 3-ARCH(1) para la SVM

Las Figuras 4.3 y 4.4 evidencian que existen problemas ya no solo en los residuos cuadráti-

cos, sino también en los residuos simples. En la figura 4.4, la FACP en el orden 13 es significativo (y cercano a la estacionalidad 12); por lo cual, se decidió incluir un término AR(13) en el Modelo 3; esto tampoco solucionó totalmente la falta de independencia de los residuos cuadráticos. Luego, de algunas pruebas se encontró como modelo final aquel que contiene términos  $c$ , SAR(12), MA(12) y AR(13) para la media (se lo llamará Modelo 4) y ARCH(1) para la varianza. Los resultados se muestran en la Tabla 4.3 y en las Figuras 4.5 y 4.6.

## Capítulo 5

# Modelos Multivariantes de Series Temporales

Una serie temporal multivariante es un proceso estocástico  $(X_t)_{t \in \mathbb{Z}}$ , con  $X_t$  un vector donde cada componente se define como una serie temporal univariante. En este documento se utilizará la notación de vectores como columnas. Así se denota:

$$X_t = (X_{1t}, \dots, X_{kt})' \text{ el vector de } k \text{ series univariantes en el instante } t$$

Lo importante de tratar series multivariantes es que, a más de considerar simultáneamente observaciones de dos o más series univariantes, también se puede analizar las correlaciones existentes entre ellas; esto evidentemente enriquece el análisis, aunque los procesos operativos serán más complejos que en el caso univariante.

### 5.1 Procesos Estacionarios

Para poder estimar las características de los procesos se necesita suponer que son estables a lo largo del tiempo; esto implica, que son estacionarios.

#### 5.1.1 Proceso estrictamente estacionario

Un proceso estocástico multivariado  $(X_t)_{t \in \mathbb{Z}}$ , con  $X_t = (X_{1t}, \dots, X_{kt})'$ , es estrictamente estacionario (o fuertemente estacionario) si las distribuciones conjuntas de cualquier conjunto finito de variables se mantienen por saltos.

Es decir, si:

$$F_{t_1+l, \dots, t_k+l}(x_{t_1+l}, \dots, x_{t_k+l}) = F_{t_1, \dots, t_k}(x_{t_1}, \dots, x_{t_k})$$

Para todo  $k \in \mathbb{N}$  y para todo  $t_1, \dots, t_k, l \in \mathbb{Z}$

Donde,  $F_{t_1, \dots, t_k}$  denota la distribución conjunta de  $X_{t_1}, \dots, X_{t_k}$ .

#### 5.1.2 Proceso débilmente estacionario

Un proceso estocástico multivariado  $(X_t)_{t \in \mathbb{Z}}$ , con  $X_t = (X_{1t}, \dots, X_{kt})'$ , se dice que es *débilmente estacionario* si sus momentos de primer y segundo orden son invariantes en el tiempo

(no dependen de  $t$ ); es decir:

1.  $E(X_t) = \mu \quad \forall t$  (el vector media es constante).
2.  $Cov(X_t, X_{t-l}) = E[(X_t - \mu)(X_{t-l} - \mu)'] = \Gamma_l \quad \forall t$ ; es decir, la matriz de *covarianzas cruzadas* entre  $X_t$  y  $X_{t-l}$  es independiente de  $t$  (solo depende del salto  $l$ ).

La media  $\mu$  es un vector  $k$ -dimensional compuesto por las esperanzas de las componentes de  $X_t$ . La matriz de covarianzas cruzadas es de orden  $k * k$ .

El  $i$ -ésimo elemento de la diagonal de  $\Gamma_0$  es la varianza de  $X_{it}$ ; mientras que, el elemento  $(i, j)$  de  $\Gamma_0$  es la covarianza entre  $X_{it}$  y  $X_{jt}$ . El elemento  $(i, j)$  de  $\Gamma_l$  es la covarianza entre  $X_{it}$  y  $X_{j,t-l}$ .

**Observación 5.1.** Se puede demostrar que si un proceso  $(X_t)_{t \in \mathbb{Z}}$  es débilmente estacionario entonces también lo será cada una de sus componentes.

## 5.2 Matrices de Correlación Cruzada (Cross-Correlation)

En lo que sigue se considera que  $(X_t)_{t \in \mathbb{Z}}$  es estacionaria.

Sea  $D$  una matriz diagonal de orden  $k * k$  compuesta por las desviaciones estándar de  $X_{it}$  para  $i = 1, \dots, k$ , que se denota por:  $D = \text{diag} \{ \sqrt{\Gamma_{11}(0)}, \dots, \sqrt{\Gamma_{kk}(0)} \}$ . La matriz de correlaciones cruzadas de  $X_t$  se define como:

$$\rho_0 \equiv [\rho_{ij}(0)] = D^{-1} \Gamma_0 D^{-1}$$

De manera particular, el elemento  $(i, j)$  de  $\rho_0$  es:

$$\rho_{ij}(0) = \frac{\Gamma_{ij}(0)}{\sqrt{\Gamma_{ii}(0)\Gamma_{jj}(0)}} = \frac{Cov(X_{it}, X_{jt})}{de(X_{it})de(X_{jt})},$$

donde,  $de(\cdot)$  es la desviación estándar.

$\rho_{ij}(0)$  es el coeficiente de correlación lineal entre  $X_{it}$  y  $X_{jt}$ . En el análisis de series de tiempo, dicho coeficiente se conoce como de concurrencia (en el mismo instante). Es fácil ver que:

- i)  $\rho_{ij}(0) = \rho_{ji}(0)$
- ii)  $-1 \leq \rho_{ij}(0) \leq 1$
- iii)  $\rho_{ii}(0) = 1$

Así,  $\rho(0)$  es una matriz simétrica con 1 en la diagonal.

Hay que mencionar que las matrices  $\Gamma_l$  contienen las **relaciones en retardo** entre las componentes de las series. Por lo tanto, las matrices de correlación cruzada se utilizan para medir la fuerza de la dependencia lineal entre las series de tiempo.

La matriz de correlación cruzada de  $X_t$  con  $X_{t-l}$  se define como:

$$\rho_l \equiv [\rho_{ij}(l)] = D^{-1} \Gamma_l D^{-1}$$

donde,  $D$  es la matriz diagonal de las desviaciones estándar de las series individuales. De la definición se tiene:

$$\rho_{ij}(l) = \frac{\Gamma_{ij}(l)}{\sqrt{\Gamma_{ii}(0) \Gamma_{jj}(0)}} = \frac{\text{Cov}(X_{it}, X_{jt-l})}{\text{de}(X_{it}) \text{de}(X_{jt})} = \frac{\text{Cov}(X_{it}, X_{jt-l})}{\text{de}(X_{it}) \text{de}(X_{jt-l})}$$

que es el coeficiente de correlación lineal entre  $X_{it}$  y  $X_{jt-l}$ . Cuando  $l > 0$ , este coeficiente de correlación mide la dependencia lineal de  $X_{it}$  con respecto a  $X_{jt-l}$  ( $X_{t-l}$  ocurre con anterioridad al instante  $t$ ). Consecuentemente, si  $\rho_{ij}(l) \neq 0$  y  $l > 0$ , se dice que la serie  $X_{jt}$  **conduce** a la serie  $X_{it}$  con retardo  $l$ .

Similarmente,  $\rho_{ji}(l)$  mide la dependencia lineal de  $X_{jt}$  con respecto a  $X_{it-l}$  y se puede decir que la serie  $X_{it}$  **conduce** a la serie  $X_{jt}$ , con retardo  $l$ , si  $\rho_{ji}(l) \neq 0$  y  $l > 0$ .

Se pueden mencionar las siguientes propiedades cuando  $l > 0$ :

1. En general,  $\rho_{ij}(l) \neq \rho_{ji}(l)$  para  $i \neq j$ , porque los dos coeficientes de correlación miden diferentes relaciones lineales entre las series. Por lo tanto,  $\Gamma_l$  y  $\rho_l$  son, generalmente, no simétricas.
2. Utilizando  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$  y suponiendo que las series son estacionarias, se tiene:

$$\text{Cov}(X_{it}, X_{jt-l}) = \text{Cov}(X_{jt-l}, X_{it}) = \text{Cov}(X_{jt}, X_{it+l}) = \text{Cov}(X_{jt}, X_{it-(-l)})$$

Así que  $\Gamma_{ij}(l) = \Gamma_{ji}(-l)$ , donde  $\Gamma_{ji}(-l)$  es el elemento  $(j, i)$  de  $\Gamma_{-l}$ ; la igualdad se cumple para  $1 \leq i, j \leq k$ . Es decir,  $\Gamma_l = \Gamma'_{-l}$ .

### 5.2.1 Dependencia Lineal

Considérense las matrices de correlación cruzada  $\{\rho(l) | l = 0, 1, 2, \dots\}$  de una serie temporal vectorial estacionaria; éstas contienen la siguiente información:

1. Los elementos de la diagonal de la matriz de correlación cruzada  $\rho_{ii}(l)$  son las funciones de autocorrelación de  $X_{it}$ .
2. El elemento fuera de la diagonal  $\rho_{ij}(0)$  mide la relación lineal de concurrencia entre  $X_{it}$  y  $X_{jt}$ .
3. Para  $l > 0$ , el elemento fuera de la diagonal  $\rho_{ij}(l)$  mide la dependencia lineal de  $X_{it}$  con respecto a  $X_{jt-l}$ .

Por lo tanto, si  $\rho_{ij}(l) = 0$  para todo  $l > 0$ ,  $X_{it}$  no depende linealmente de ningún valor del pasado  $X_{jt-l}$ .

#### *Resumen e interpretación.*

En general, la relación lineal entre dos series de tiempo  $X_{it}$  y  $X_{jt}$  puede resumirse en la siguiente forma:

1.  $X_{it}$  y  $X_{jt}$  no tienen relación lineal si  $\rho_{ij}(l) = \rho_{ji}(l) = 0, \quad \forall l \geq 0$ .
2.  $X_{it}$  y  $X_{jt}$  están al mismo tiempo correlacionadas si  $\rho_{ij}(0) \neq 0$ .
3.  $X_{it}$  y  $X_{jt}$  no tienen relación de avance-retardo si  $\rho_{ij}(l) = 0$  y  $\rho_{ji}(l) = 0, \quad \forall l > 0$ . En este caso, se dice que las series son desacopladas.
4. Existe una *relación unidireccional* desde  $X_{it}$  hacia  $X_{jt}$  si  $\rho_{ij}(l) = 0, \quad \forall l > 0$ , pero  $\rho_{ji}(v) \neq 0$  para algún  $v > 0$ . En este caso,  $X_{it}$  no depende de ningún valor del pasado de  $X_{jt}$ , pero  $X_{jt}$  depende de algún valor del pasado de  $X_{it}$ .
5. Existe una *relación de retroalimentación* entre  $X_{it}$  y  $X_{jt}$  si  $\rho_{ij}(l) \neq 0$  para algún  $l > 0$  y  $\rho_{ji}(v) \neq 0$  para algún  $v > 0$ .

Las formulaciones anteriores son suficientes para analizar la dependencia lineal entre series temporales. Un enfoque más informativo para estudiar las relaciones entre las series temporales es construir un modelo multivariante para las series, porque un modelo correctamente especificado considera simultáneamente el número de series y las correlaciones cruzadas de las mismas.

En la práctica se utilizan los estimadores de las matrices antes mencionadas; en particular para  $\Gamma(l)$ :

$$\hat{\Gamma}(l) = \frac{1}{T} \sum_{t=l+1}^T (X_t - \bar{X})(X_{t-l} - \bar{X})', \quad l \geq 0$$

donde,

T: número de observaciones

$$\bar{X} = \frac{\left( \sum_{t=1}^T X_t \right)}{T} : \text{vector de medias muestrales}$$

y para  $\rho(l)$  :

$$\hat{\rho}(l) = \hat{D}^{-1} \hat{\Gamma}_X(l) \hat{D}^{-1}, \quad l \geq 0$$

donde,

$\hat{D}$  : Es la matriz diagonal de orden  $(k * k)$  que contiene las desviaciones estándar muestrales del vector  $X_t$  en la diagonal.

**Ejemplo 5.1.** Se consideran tres series de datos económicos de un país sudamericano: el producto interno bruto (PIB), denotada por  $(X_{1t})$ ; el consumo interno (CI), denotada por  $(X_{2t})$  y la demanda final interna (DFI), denotada por  $(X_{3t})$ . Se dispone de 56 datos trimestrales, desde noviembre de 2010 hasta junio de 2015 (Ver Anexo D.1). Para efectos de comparaciones se trabajará únicamente con los primeros 50 datos y se guardarán los 6 restantes para comparar con predicciones posteriores (enero 2015 - junio de 2015). Se desea estimar las matrices de correlaciones cruzadas de las series.

#### Resolución.

Las matrices de correlación cruzada se las construyen de manera manual, considerando cada escenario de posibles combinaciones entre las variables; así, en este caso, se obtiene:

a) Estadísticos descriptivos de  $x_{1t}$ ,  $X_{2t}$  y  $X_{3t}$ .

	Media	Mediana	Máximo	Mínimo	Desv. Est.	Asimetría	Curtosis
$X_{1t}$	96,52	89,79	165,31	49,79	34,15	0,35	1,92
$X_{2t}$	78,10	73,72	126,01	43,05	24,89	0,31	1,84
$X_{3t}$	98,71	92,34	166,66	50,51	34,48	0,30	1,87

b) Matrices de correlación cruzada

	retardo 1			retardo 2			retardo 3		
$X_{1t}$	0,94	0,94	0,94	0,87	0,88	0,87	0,81	0,82	0,81
$X_{2t}$	0,93	0,94	0,94	0,87	0,88	0,87	0,80	0,81	0,81
$X_{3t}$	0,93	0,94	0,94	0,87	0,88	0,87	0,80	0,81	0,81

c) Representación simplificada

+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+

**Tabla 5.1:** Resumen de estadísticas y matrices de correlación cruzada para  $X_{1t}$ ,  $X_{2t}$  y  $X_{3t}$

Para representar a las matrices de correlación cruzada, se utiliza la forma gráfica simplificada, que utiliza el hecho que  $2/\sqrt{T}$  (0,28 en este caso) es el valor crítico de la correlación muestral con nivel de significación del 5 %, bajo la suposición que  $X_t$  es un ruido blanco:

- “+” representa a los coeficientes de correlación que son mayores o iguales a  $2/\sqrt{T}$ .
- “-” representa a los coeficientes de correlación que son menores o iguales que  $-2/\sqrt{T}$ .
- “.” Representa a los coeficientes que se encuentran entre a  $-2/\sqrt{T}$  y  $2/\sqrt{T}$ .

Es fácil ver que las correlaciones cruzadas son significativas en los primeros tres retardos. En algunos paquetes estadísticos se puede encontrar el cálculo de las matrices de correlación cruzada. En Eviews, por ejemplo, se presenta la siguiente salida:

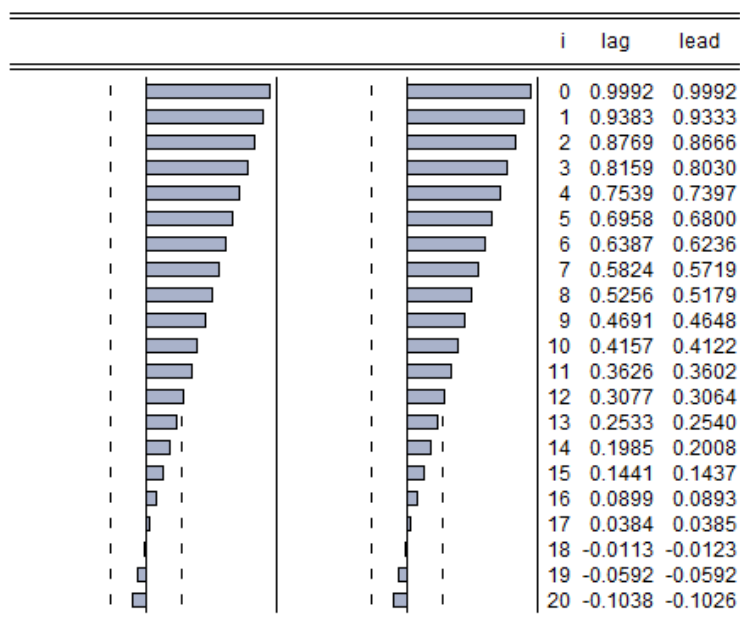


Figura 5.1: Correlaciones cruzadas entre  $X_{1t}$  y  $X_{2t}$

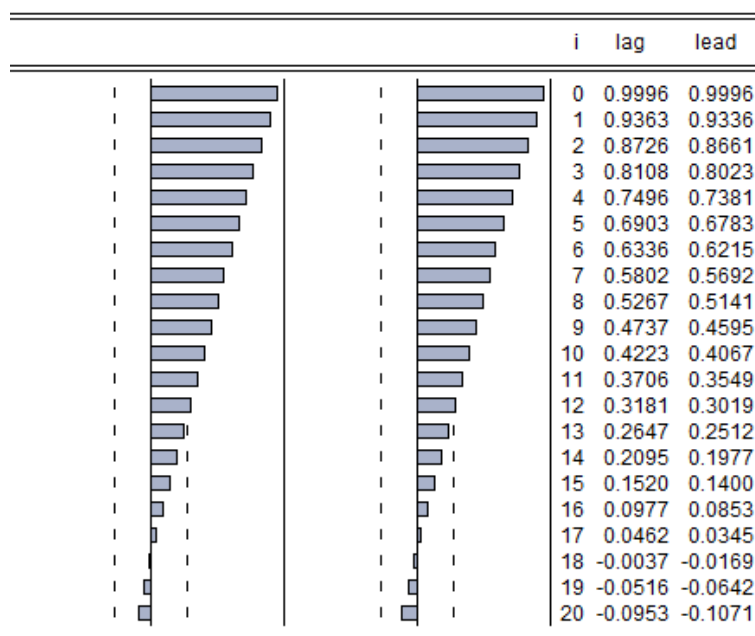


Figura 5.2: Correlaciones cruzadas entre  $X_{1t}$  y  $X_{3t}$



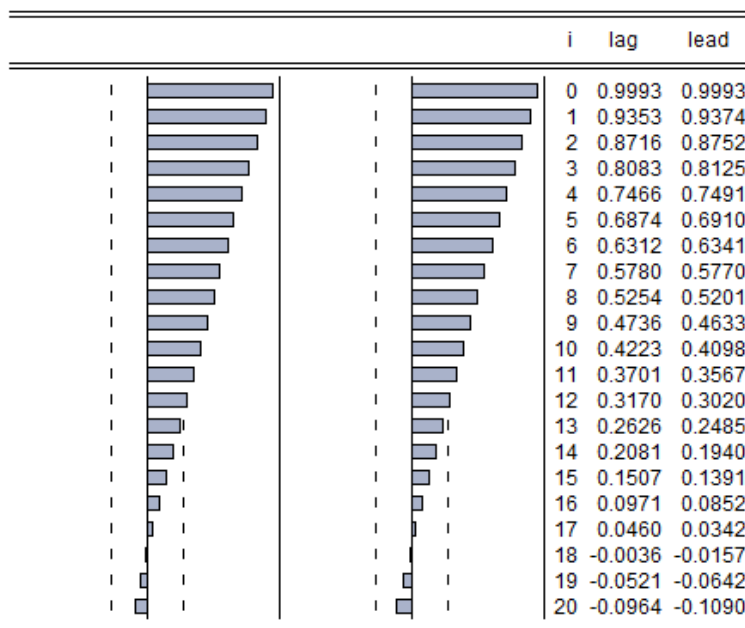


Figura 5.3: Correlaciones cruzadas entre  $X_{2t}$  y  $X_{3t}$

Como se puede observar, los valores calculados por el paquete son aquellos que están en la diagonal secundaria de las matrices calculadas manualmente. Para poder completar la matriz, se puede, ver a partir de la fórmula de cálculo que las  $\hat{\rho}_{ii}(l)$  corresponden a las autocorrelaciones simples de orden  $l$  de cada serie univariante dentro de  $X_t$ .

## 5.3 Modelos de Vectores Autoregresivos (VAR)

Este tipo de modelos no pertenecen a los modelos estocásticos desarrollados por Box y Jenkins; sin embargo, la representación VAR se puede considerar como la generalización de los modelos autoregresivos al caso multivariante.

### 5.3.1 El caso bivalente

Una representación VAR bivalente es aquella que consideran dos variables  $X_{1t}$  y  $X_{2t}$ . Cada una de ellas se expresa en función de sus propios valores del pasado y de los del presente y del pasado de la otra variable. Por ejemplo, se va a representar el modelo VAR bivalente de orden  $p = 3$  [VAR (3)]; se escribe:

$$X_{1t} = v_1 + \sum_{i=1}^3 b_{1i}X_{1t-i} + \sum_{i=1}^3 c_{1i}X_{2t-i} - d_1X_{2t} + u_{1t}$$

$$X_{2t} = v_2 + \sum_{i=1}^3 b_{2i}X_{1t-i} + \sum_{i=1}^3 c_{2i}X_{2t-i} - d_2X_{1t} + u_{2t}$$

Las variables  $X_{1t}$  y  $X_{2t}$  son estacionarias; las perturbaciones  $u_{1t}$  y  $u_{2t}$  son ruidos blancos de varianzas constantes y no correlacionados. Se puede ver inmediatamente la gran cantidad

de parámetros a estimar (aquí 16 coeficientes), con los problemas típicos de pérdida de grados de libertad. Hay que tomar en cuenta que  $X_{1t}$  tiene un efecto inmediato en  $X_{2t}$  y recíprocamente. Este sistema inicial se denomina *forma estructural* de la representación VAR. Su *forma matricial*, se expresa como:

$$BX_t = v + \sum_{i=1}^3 \tilde{A}_i X_{t-i} + u_t$$

con:

$$B = \begin{bmatrix} 1 & d_1 \\ d_2 & 1 \end{bmatrix} \quad X_t = \begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} \quad \tilde{A}_i = \begin{bmatrix} b_{1i} & c_{1i} \\ b_{2i} & c_{2i} \end{bmatrix} \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

Para obtener la *forma estándar* de un modelo VAR, se multiplica la ecuación anterior por  $B^{-1}$  (que se supone existe); es decir, se expresa por:

$$X_{1t} = v_1^0 + \sum_{i=1}^3 a_{1i}^1 X_{1t-i} + \sum_{i=1}^3 a_{1i}^2 X_{2t-i} + \vartheta_{1t}$$

$$X_{2t} = v_2^0 + \sum_{i=1}^3 a_{2i}^1 X_{1t-i} + \sum_{i=1}^3 a_{2i}^2 X_{2t-i} + \vartheta_{2t}$$

En esta especificación, los errores  $\vartheta_{1t}$  y  $\vartheta_{2t}$  son funciones de las innovaciones  $u_{1t}$  y  $u_{2t}$ ; en efecto, de  $\vartheta = B^{-1}u$ , se obtiene:

$$\vartheta_{1t} = \frac{(u_{1t} - d_1 u_{2t})}{(1 - d_1 d_2)} \quad \text{y} \quad \vartheta_{2t} = \frac{(u_{2t} - d_2 u_{1t})}{(1 - d_1 d_2)}$$

Se puede ver que:

$$E(\vartheta_{1t}) = 0; \quad E(\vartheta_{2t}) = 0; \quad E(\vartheta_{1t} \vartheta_{1t-i}) = 0; \quad E(\vartheta_{2t} \vartheta_{2t-i}) = 0$$

Por lo tanto, los elementos de cada familia de errores tienen esperanza nula y son no correlacionados. Además:

$$E(\vartheta_{1t}^2) = \frac{(\sigma_{u_1}^2 + d_1^2 \sigma_{u_2}^2)}{(1 - d_1 d_2)^2}; \quad E(\vartheta_{2t}^2) = \frac{(\sigma_{u_2}^2 + d_2^2 \sigma_{u_1}^2)}{(1 - d_1 d_2)^2}$$

Donde  $\sigma_{u_1}^2$  y  $\sigma_{u_2}^2$  son las varianzas de  $u_1$  y  $u_2$ , respectivamente. Así, la varianza de los errores es constante (independiente del tiempo). Además:

$$E(\vartheta_{1t} \vartheta_{2t}) = -\frac{(d_2 \sigma_{u_1}^2 + d_1 \sigma_{u_2}^2)}{(1 - d_1 d_2)^2}$$

Si  $d_1 = d_2 = 0$ , las variables  $X_{1t}$  y  $X_{2t}$  no tienen ninguna influencia sincrónica entre sí, pues los errores  $\vartheta_{1t}$  y  $\vartheta_{2t}$  serían no correlacionados. En caso contrario, los errores  $\vartheta_{1t}$  y  $\vartheta_{2t}$  estarían correlacionados y por tanto, una variación de uno de estos errores en un instante dado tiene impacto en el otro.

**Proposición 5.1.** *El modelo VAR no permite distinguir entre variables endógenas (variables propias del fenómeno estudiado) y exógenas (variables externas que ayudan a explicar las*

variables endógenas).

### 5.3.2 Representación general de un VAR

**Notación.** Un modelo VAR a  $k$  variables con  $p$  retardos se denota  $VAR(p)$ .

La generalización de la representación VAR a  $k$  variables con  $p$  retardos se escribe en su forma estándar como:

$$X_t = v_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t$$

donde,

$$X_t = \begin{bmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{k,t} \end{bmatrix}; \quad v_0 = \begin{bmatrix} v_1^0 \\ v_2^0 \\ \vdots \\ v_k^0 \end{bmatrix}; \quad A_i = \begin{bmatrix} a_{1i}^1 & \dots & a_{1i}^k \\ \vdots & \ddots & \vdots \\ a_{ki}^1 & \dots & a_{ki}^k \end{bmatrix}; \quad u_t = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{bmatrix}$$

$u_t$  es el vector compuesto por los ruidos blancos de cada una de las  $k$  ecuaciones del modelo.

Se denota por:  $\Sigma_u = E(u_t u_t')$ , la matriz desconocida, de dimensión  $k$ , de varianzas-covarianzas de los errores.

Esta representación puede escribirse mediante el operador de retardo  $B$ , como:

$$(I - A_1 B - A_2 B^2 - \dots - A_p B^p) X_t = v_0 + u_t, \quad \text{o también:} \quad A(B) X_t = v_0 + u_t$$

donde, el operador de retardo  $B$  se define de la siguiente manera:

$$B^i X_t = X_{t-i}, \quad i = 1, 2, \dots$$

$$B^0 X_t = X_t$$

#### 5.3.2.1 Estabilidad de un VAR

Considérese un modelo  $VAR(1)$ :

$$X_t = v_0 + A_1 X_{t-1} + u_t$$

Se dice que un  $VAR(1)$  es estable si todos los valores propios de  $A_1$  son de valor absoluto menor que 1; lo que se puede expresar también por:

$$\det(I_k - A_1 z) \neq 0, \quad \text{para} \quad |z| \leq 1$$

Esto implica que todas las raíces del polinomio característico están fuera del círculo unidad.

### 5.3.2.2 Representación de un proceso VAR(p) en la forma de VAR(1)

Un proceso  $VAR(p)$  se puede escribir como un proceso  $VAR(1)$  si se plantea en la siguiente forma:

$$X_t = A_0 + AX_{t-1} + U_t$$

donde,

$$X_t = \begin{pmatrix} X_t \\ X_{t-1} \\ \vdots \\ X_{t-p+1} \end{pmatrix}_{kp*1} \quad X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{kt} \end{pmatrix}_{k*1} \quad A_0 = \begin{pmatrix} v_0 \\ 0_{k*1} \\ \vdots \\ 0_{k*1} \end{pmatrix}_{kp*1} \quad U_t = \begin{pmatrix} (u_t)_{k*1} \\ 0_{k*1} \\ \vdots \\ 0_{k*1} \end{pmatrix}_{kp*1}$$

$$A = \begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_k & 0 & \dots & 0 & 0 \\ 0 & I_k & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_k & 0 \end{pmatrix}_{kp*kp}$$

Lo importante de esta representación es que para obtener las propiedades de los procesos VAR, es suficiente con probarlas para una VAR (1)

### 5.3.2.3 Procesos VAR(p) estables

Se dice que un  $VAR(p)$  es estable si:

$$\det(I_{kp} - Az) \neq 0, \text{ para } |z| \leq 1$$

Además, se puede demostrar que:

$$\det(I_{kp} - Az) = \det(I - A_1z - A_2z^2 - \dots - A_pz^p)$$

**Observación 5.2.** Se puede demostrar que si un proceso  $VAR(1)$  es estable, entonces es estacionario.

En general, se puede demostrar que un proceso  $VAR(p)$  es estacionario si el polinomio definido a partir de la expresión:  $\det(I - A_1z - A_2z^2 - \dots - A_pz^p)$  tiene sus raíces fuera del círculo unidad del plano complejo; es decir:

$$\det(I - A_1z - A_2z^2 - \dots - A_pz^p) \neq 0 \forall z \text{ tal que } |z| \leq 1$$

**Ejemplo 5.2.** Determine si el siguiente modelo es estacionario.

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \end{bmatrix} + \begin{bmatrix} 0,8 & 0,9 \\ 0,7 & 0,7 \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

**Resolución.**

Se tiene que:

$$\det \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,8 & 0,9 \\ 0,7 & 0,7 \end{bmatrix} z \right) = 1 - 1,5z - 0,07z^2, \quad \text{entonces} \quad \begin{matrix} z_1 = -16,17 \\ z_2 = -15,97 \end{matrix}$$

Las dos raíces son superiores a 1 en valor absoluto; por lo tanto, el proceso es estable; lo que implica que es estacionario.

## 5.4 Representación VARMA de una Serie Multivariante

### 5.4.1 La representación VMA

Un modelo media móvil vectorial de orden  $q$  ( $VMA(q)$  por sus siglas en inglés), tiene la siguiente forma:

$$X_t = m_0 + u_t - M_1 u_{t-1} - \dots - M_q u_{t-q} \quad \text{o} \quad X_t = m_0 + M(B)u_t$$

donde,

$m_0$  : Es un vector de dimensión  $k$  constante.

$M_i$  : Son matrices de dimensión  $k * k$ .

$M(B) = I - M_1 B - \dots - M_q B^q$  es el polinomio matriz MA en términos del operador de retardo  $B$ .

$\{u_t\}_{t \in \mathbb{Z}}$  : Es un ruido blanco multidimensional.

De manera similar al caso univariante, los procesos  $VMA(q)$  son débilmente estacionarios, siempre que la matriz de covarianzas ( $\Sigma_u$ ) de  $u_t$  exista. Si se toma la esperanza de  $X_t$ , se tiene:

$$\mu = E(X_t) = m_0$$

Así, el vector constante  $m_0$  es el vector media de  $X_t$  para un modelo VMA.

Se define  $\tilde{X}_t = X_t - m_0$  como el proceso corregido en media  $VAR(q)$ . Cuando se tiene un proceso  $VMA(q)$  y considerando el hecho de que los  $\{u_t\}$  no están correlacionados, se obtiene:

1.  $Cov(\tilde{X}_t, u_t) = \Sigma_u$
2.  $\Gamma_0 = \Sigma_u + M_1 \Sigma_u M_1' + \dots + M_q \Sigma_u M_q'$
3.  $\Gamma_l = 0$  si  $l > q$
4.  $\Gamma_l = \sum_{j=l}^q M_j \Sigma_u M_{j-l}'$  si  $1 \leq l \leq q$ , donde  $M_0 = -I$

Dado que  $\Gamma_l = 0$  para  $l > q$ , las matrices de correlación cruzada de un proceso  $VMA(q)$  satisfacen:

$$\rho_l = 0, \quad l > q$$

### 5.4.2 Representación lineal de un VAR(p)

Cuando se analizaron las series temporales univariantes, se mostró que bajo ciertas condiciones un proceso  $AR(1)$  se puede representar como un proceso lineal. De la misma manera,

para las series multivariantes se puede representar, en particular, un  $VAR(1)$  como un proceso lineal (se dice que es la representación lineal del proceso). Un modelo con esta forma permite medir el impacto en los valores presentes de una variación de innovaciones (o choques).

Sea  $X_t$  un  $VAR(1)$  estable:

$$X_t = v_0 + A_1 X_{t-1} + u_t$$

Si se realizan sustituciones repetidas en el proceso hasta el  $i$ -ésimo paso, se obtiene:

$$\begin{aligned} X_t &= v_0 + A_1 (v_0 + A_1 X_{t-2} + u_{t-1}) + u_t = (I + A_1) v_0 + A_1^2 X_{t-2} + (A_1 u_{t-1} + u_t) \\ X_t &= v_0 + A_1 \left( (I + A_1) v_0 + A_1^2 X_{t-3} + A_1 u_{t-2} + u_{t-1} \right) + u_t \\ &= (I + A_1 + A_1^2) v_0 + A_1^3 X_{t-3} + (A_1^2 u_{t-2} + A_1 u_{t-1} + u_t) \\ &\vdots \\ X_t &= (I + A_1 + \dots + A_1^i) v_0 + A_1^{i+1} X_{t-i} + \sum_{j=0}^i A_1^j u_{t-j}, \quad i = 0, 1, 2, \dots \end{aligned}$$

Por definición,  $A^0 = I$ .

Como el  $VAR$  es estable, se cumple que:

$$(I + A_1 + \dots + A_1^i) v_0 \rightarrow (I - A_1)^{-1} v_0 \quad \text{si } i \rightarrow \infty$$

Además,  $A_1^{i+1} \rightarrow 0$  rápidamente; así, se lo puede ignorar. Por lo tanto, se obtiene la siguiente representación:

$$X_t = (I - A_1)^{-1} v_0 + \sum_{i=0}^{\infty} A_1^i u_{t-i}$$

La generalización a un proceso  $VAR(p)$  se la realiza aplicando la representación de un  $VAR(p)$  como un  $VAR(1)$ . Así, se obtiene:

$$X_t = \mu + \sum_{i=0}^{\infty} M_i u_{t-i}$$

donde,

$$\begin{aligned} \mu &= (I - A_1 - A_2 - \dots - A_p)^{-1} v_0 \\ M_i &= \sum_{j=1}^{\min(p,i)} A_j M_{i-j} \quad i = 1, 2, \dots \quad \text{y} \quad M_0 = I \end{aligned}$$

Las matrices  $M_i$  aparecen como un “*factor de impacto*”, a través de las cuales se analiza el efecto de un choque a lo largo de todo el proceso.

### Observación 5.3.

1. Así, se obtiene que si un proceso  $VAR(p)$  es estable, tiene una representación lineal estacionaria.
2. No se profundiza sobre la modelación  $VMA(q)$  porque no está implementada en los

programas comerciales usuales.

**Ejemplo 5.3.** Considérese el proceso VMA (1):

$$X_t = \mu + u_t - M_1 u_{t-1} = \mu + u_t - M u_{t-1}$$

donde, por simplicidad, se ha quitado el subíndice de  $M_1$ . Este modelo puede escribirse explícitamente como:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} - \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}$$

Se dice que la serie de retardos ( $X_t$ ) solo depende del presente y del pasado de  $\{u_t\}$ . Por lo tanto, el modelo es de memoria finita.

El parámetro  $m_{12}$  denota la dependencia lineal de  $X_{1t}$  con  $u_{2,t-1}$  en la presencia de  $u_{1,t-1}$ . Si  $m_{12} = 0$ ,  $X_{1t}$  no depende de los retardos de  $u_{2t}$  y, entonces tampoco, de los retardos de  $X_{2t}$ . De manera similar, si  $m_{21} = 0$ ,  $X_{2t}$  no depende de los valores pasados de  $X_{1t}$ . Los elementos fuera de la diagonal de  $M$  muestran la dependencia entre las componentes de las series.

Para este ejemplo, se pueden clasificar las relaciones entre  $X_{1t}$  y  $X_{2t}$  así:

1. Son series desacopladas si  $m_{12} = m_{21} = 0$ .
2. Hay una relación dinámica unidireccional de  $X_{1t}$  sobre  $X_{2t}$  si  $m_{12} = 0$ , pero  $m_{21} \neq 0$  y viceversa.
3. Hay una relación de retroalimentación entre  $X_{1t}$  y  $X_{2t}$  si  $m_{12} \neq 0$  y  $m_{21} \neq 0$ .

Finalmente, la correlación actual entre los  $m_{ij}$  (coeficientes estimados para el modelo VMA) es la misma que entre los  $u_{it}$ . La descripción previa se puede generalizar para un modelo  $VMA(q)$ .

### 5.4.3 La representación VARMA

La representación VAR puede generalizarse (es una aplicación multivariante del teorema de descomposición de Wold (1954)), por analogía con los procesos  $ARMA(pq)$ .

$$X_t = A_0 + A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t + M_1 u_{t-1} + M_2 u_{t-2} + \dots + M_q u_{t-q}$$

Se trata de un proceso ARMA multivariante que se denota: VARMA.

Las condiciones de estacionariedad son análogas a las de un proceso ARMA univariante:

- Un proceso VAR es siempre invertible; es lineal (por ende estacionario) cuando es estable.
- Un proceso VMA es siempre estacionario. Es invertible si las raíces del polinomio característico asociado a  $M(z)$  están fuera del círculo unitario complejo.
- Las condiciones de estacionariedad e invertibilidad de un VARMA están dadas, respectivamente, por la parte VAR y la parte VMA del VARMA.

La generalización de los modelos ARMA encuentra nuevos temas que no ocurren en el desarrollo de los modelos VAR y VMA. Uno de ellos es el *problema de identificación*. A diferencia de los modelos ARMA, los modelos VARMA pueden no estar definidos de manera única.

**Ejemplo 5.4.** Considere un modelo bivalente  $VMA(1)$ :

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} - \begin{bmatrix} 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1,t-1} \\ u_{2,t-1} \end{bmatrix}$$

Es *idéntico* al modelo bivalente  $VAR(1)$ :

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} = \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

La equivalencia de los modelos se puede examinar fácilmente componente a componente. Es decir, para el modelo  $VMA(1)$  se tiene:

$$X_{1t} = u_{1t} - 2u_{2,t-1} \quad y \quad X_{2t} = u_{2t}$$

Por otro lado, para el modelo  $VAR(1)$  se tiene:

$$X_{1t} + 2X_{2,t-1} = u_{1t} \quad y \quad X_{2t} = u_{2t}$$

De los modelos se puede ver que:

$$X_{2,t-1} = u_{2,t-1}$$

Luego, los modelos para  $X_{1t}$  son idénticos. Este tipo de problema de identificación es inofensivo porque cualquiera de los modelos puede ser utilizado en una aplicación real. Sin embargo, existen casos en los que esta situación si se convierte en un problema y hay que tener en cuenta muchas restricciones para poder estimar un modelo VARMA.

## 5.5 Formulación de un modelo VAR

Los parámetros de un proceso VAR pueden estimarse solamente en las series temporales estacionarias. Se conoce que muchas series pueden volverse estacionarias a través de un proceso de diferenciación (en el caso de una tendencia determinista o una estacionalidad) o a través de una transformación de las variables (por ejemplo, una transformación logarítmica) en ciertos casos con heteroscedasticidad.

### 5.5.1 Estimación

En el caso de un proceso VAR, las ecuaciones pueden estimarse por MCO independientemente una de la otra (o por un método de máxima verosimilitud).

Sea el modelo  $VAR(p)$  estimado:

$$X_t = \hat{A}_0 + \hat{A}_1 X_{t-1} + \hat{A}_2 X_{t-2} + \dots + \hat{A}_p X_{t-p} + \hat{u}_t$$



siendo,  $\hat{u}_t$  el vector de dimensión  $(k, 1)$  de componentes  $\hat{u}_{1t}, \hat{u}_{2t}, \dots, \hat{u}_{kt}$ .

Se denotará por  $\hat{\Sigma}_{u,p}$  la matriz de varianzas covarianzas estimada de los residuos del modelo. Para cualquier orden  $p$ , se define por:

$$\hat{\Sigma}_{u,p} = \frac{1}{T - kp - 1} \sum_{t=p+1}^T \hat{u}_t (\hat{u}_t)', \quad p \geq 0$$

### 5.5.2 Determinación del número de retardos

El mayor problema que debe enfrentarse a la hora de estimar los modelos VAR es el de la determinación del número de retardos a incluir en la estimación; suele realizarse en forma cuantitativa, analizando los resultados de la estimación y comparando los resultados obtenidos entre distintos modelos alternativos, ya que no es frecuente encontrar evidencias teóricas al respecto.

Los criterios comúnmente utilizados para la selección entre modelos alternativos son el criterio informativo de Akaike (AIC), el criterio de información bayesiano (BIC), que también se conoce como el criterio de Schwarz (SC) o el criterio de Hanan-Quinn (HQ).

Para el caso de la representación VAR, estos criterios se pueden utilizar para determinar el orden  $p$  del modelo. EL proceso de selección del orden de la representación consiste en estimar todos los modelos VAR para retardos de 0 a  $p_0$  ( $p_0$  es el máximo retardo admisible por la teoría económica o por los datos disponibles y se fija de antemano). Los estadísticos AIC(p), SC(p) y HQ(p) para el caso multivariante tienen las siguientes expresiones:

$$\begin{aligned} AIC(p) &= \ln \left[ \left| \sum_u \right| \right] + \frac{2k^2 p}{T} \\ SC(p) &= \ln \left[ \left| \sum_u \right| \right] + \frac{k^2 p \ln(T)}{T} \\ HQ(p) &= \ln \left[ \left| \sum_u \right| \right] + \frac{2k^2 p \ln(\ln(T))}{T} \end{aligned}$$

donde,

$k$ = número de variables del sistema.

$T$ = número de observaciones.

$p$ = número de retardos.

$\sum_u$  = matriz de varianzas covarianzas de residuos del modelo con retardo  $p$  (fijo).

Otro criterio utilizado para determinar el retardo del modelo, es la razón de máxima verosimilitud. Para utilizar este criterio, es necesario que el vector de las innovaciones tenga una distribución normal; el logaritmo de la función de verosimilitud tiene la siguiente expresión:

$$l = -\frac{Tp}{2} (1 + \ln(2\pi)) - \frac{T}{2} \ln \left[ \left| \hat{\Sigma}_u \right| \right]$$

**Observación 5.4.** El retardo  $p$  que minimice la mayor cantidad de los criterios de AIC, HQ, BIC; o, maximice el logaritmo de la función de verosimilitud, se retiene. En la práctica se aconseja que  $p \leq 5$ , debido a que valores superiores implican incorporar una gran cantidad de parámetros.

También se puede utilizar el estadístico  $M(p)$  para probar la hipótesis nula  $H_0$  : El modelo es un  $VAR(p)$  contra la alternativa,  $H_1$  : El modelo es un  $VAR(p-1)$ .

Este estadístico se define por:

$$M(p) = - \left( T - k - p - \frac{3}{2} \right) \ln \left[ \frac{|\hat{\Sigma}_{u,p}|}{|\hat{\Sigma}_{u,p-1}|} \right]$$

donde,

$\hat{\Sigma}_{u,j}$  =matriz de varianzas covarianzas del modelo con retardo  $j$ .

$M(p)$  sigue asintóticamente una distribución Ji-cuadrado con  $k^2$  grados de libertad.

### 5.5.3 Diagnóstico y validación del modelo

Un buen punto de partida para la verificación de que el modelo estimado es el adecuado, es la significación de los parámetros estimados, para no tener parámetros no deseados o parámetros que no aportan al modelo. Por otro lado, esto puede ser engañoso, porque los parámetros estimados de un modelo pobre pueden ser también significativos. Por lo tanto, no se debe depender exclusivamente de la significación de los parámetros para evaluar el modelo.

Como en la mayoría de situaciones de modelación, la forma de evaluación se realiza a través del comportamiento de los residuos. Si el modelo es una representación adecuada de un proceso generado por las series de tiempo, los residuos no deben tener ninguna tendencia significativa ni patrón.

Una forma de observar esto es considerar los elementos individuales de la matriz de autocorrelación de los vectores de residuos. Otra forma es el uso del estadístico *Portmanteau*, que se analizará posteriormente.

#### 5.5.3.1 Matrices de autocorrelación multivariante

Sea  $\{u_t\}$  un ruido blanco  $k$ -dimensional con matriz de covarianza  $\Sigma_u$  y su correspondiente matriz de correlación  $R_u$ . La matriz de autocovarianza y la matriz de autocorrelación muestral de  $\{u_t\}$  con respecto al retardo  $i$  están dadas por:

$$\hat{C}_i = \frac{1}{T} \sum_{t=i+1}^T \hat{u}_t \hat{u}'_{t-i} \quad i = 0, 1, \dots; i < T$$

$$\hat{R}_i = V_u^{-\frac{1}{2}} \hat{C}_i V_u^{-\frac{1}{2}} \quad i = 0, 1, \dots; i < T$$

donde,  $T$  es el número de observaciones de las series de tiempo y  $V_u^{-\frac{1}{2}}$  es una matriz diagonal  $(k * k)$  con el inverso de la raíz cuadrada de los elementos de la diagonal de  $C_0$  en su

diagonal.

Sea  $R_l^* = (R_1, \dots, R_l)'$ .

### 5.5.3.2 La prueba "Portmanteau"

La prueba de bondad de ajuste para los residuos de Box-Pierce (1970), la prueba *Portmanteau*, fue extendida a modelos VAR multivariante por Hosking (1980) y Li-McLeod (1981). Esta prueba determina si las autocorrelaciones residuales, sobre un retardo específico, son estadísticamente nulos.

La hipótesis que se prueba es:

$$H_0 : R_l^* = (R_1, \dots, R_l)' = 0 \quad \text{contra} \quad H_a : R_l^* = (R_1, \dots, R_l)' \neq 0$$

Si no se rechaza la hipótesis nula, se puede asumir que los residuos se comportan como un ruido blanco y, por lo tanto, es adecuado el modelo ajustado.

La prueba multivariante *Portmanteau* propuesta por Hosking (1980) considera el estadístico:

$$Q(l) = T \sum_{i=1}^l \text{tr} \left( \hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right)$$

Este estadístico tiene aproximadamente una distribución Ji-Cuadrada con  $k^2(l-p)$  grados de libertad bajo la hipótesis nula, donde  $p$  es el orden estimado del modelo VAR ( $p$ ) y  $l$  es el número de retardos incluidos en la prueba para la significación total. Ljung-Box (1978) propusieron una modificación que conduce a propiedades mejores en el caso univariante; Hosking considera una modificación similar para el caso multivariante. El estadístico modificado de la prueba *Portmanteau* está dado por:

$$Q'(l) = T^2 \sum_{i=1}^l (T-i)^{-1} \text{tr} \left( \hat{C}_i' \hat{C}_0^{-1} \hat{C}_i \hat{C}_0^{-1} \right)$$

### 5.5.3.3 Prueba de Breusch - Godfrey o Prueba del Multiplicador de Lagrange (LM)

Se utiliza para detectar autocorrelación de cualquier orden, especialmente en aquellos modelos con o sin variables dependientes retardadas. Permite determinar si existe correlación en los residuos hasta un determinado orden.

Se realiza la siguiente prueba de hipótesis:

$$H_0 : \rho_l = 0, \quad \text{contra} \quad H_a : \rho_l \neq 0,$$

donde,  $l$  es el orden del modelo VAR ajustado.

El estadístico utilizado para la prueba es:

$$LM = TR^2$$

donde,  $T$  el número de observaciones y  $R^2$  corresponde a la bondad de ajuste de la regresión auxiliar entre las variables y los residuos.

Este estadístico, bajo  $H_0$ , sigue asintóticamente una distribución Ji-cuadrado con  $l$  grados de libertad,  $\chi_l^2$ .

#### 5.5.3.4 Prueba de Jarque-Bera

Es una prueba asintótica de normalidad para grandes muestras. La prueba de Jarque-Bera (JB) considera la relación entre los coeficientes de asimetría y apuntamiento de los residuos de la ecuación estimada y los correspondientes de una distribución normal, de forma tal que si estas relaciones son suficientemente diferentes se rechazará la hipótesis nula de normalidad.

Se realiza la siguiente prueba de hipótesis:

$H_0$  : los residuos siguen una distribución normal multivariante

$H_1$  : los residuos no siguen una distribución normal multivariante

Este estadístico se basa en las medidas de apuntamiento (*curtosis*) y la asimetría a través de la transformación de Mahalanobis.

La  $i$ -ésima componente del vector de asimetría estimado, se calcula de la siguiente manera:

$$as_i = \frac{\frac{1}{T} \sum_{j=1}^T \hat{\epsilon}_{ij}^3}{\frac{1}{T} \sum_{j=1}^T (\hat{\epsilon}_{ij}^2)^{3/2}} = \frac{\sum_{j=1}^T \hat{\epsilon}_{ij}^3}{\sum_{j=1}^T (\hat{\epsilon}_{ij}^2)^{3/2}}$$

La  $i$ -ésima componente del vector de apuntamiento estimado se calcula de la siguiente manera:

$$k_i = \frac{\frac{1}{T} \sum_{j=1}^T \hat{\epsilon}_{ij}^4}{\frac{1}{T} \sum_{j=1}^T (\hat{\epsilon}_{ij}^2)^2} = \frac{\sum_{j=1}^T \hat{\epsilon}_{ij}^4}{\sum_{j=1}^T (\hat{\epsilon}_{ij}^2)^2}$$

$\hat{\epsilon}_{ij}$  son los elementos de la matriz  $\hat{V}$ , que se define de la siguiente manera:

$$\hat{V} = \hat{U} S_{\hat{U}}^{-1}$$

donde,  $\hat{U}$  es la matriz de los residuos obtenidos a través de la estimación de las variables utilizando el método de mínimos cuadrados; mientras que  $S_{\hat{U}}$  es una matriz triangular superior tal que:

$$\hat{U}' \hat{U} = S_{\hat{U}}' S_{\hat{U}} \quad \text{y} \quad (\hat{U}' \hat{U})^{-1} = S_{\hat{U}}^{-1} (S_{\hat{U}}^{-1})'$$

En este caso,  $\hat{V}$  es la matriz ortogonalizada de los residuos estimados; es decir,  $as_i$  y  $k_i$  corresponden a la asimetría y el apuntamiento individual estimados, respectivamente.

Entonces, se define a la asimetría y al apuntamiento estimados de la distribución de la serie multivariante como:

$$AS = (as_1, \dots, as_T)' (as_1, \dots, as_T)$$

$$K = (k_1 - 3, \dots, k_T - 3)' (k_1 - 3, \dots, k_T - 3)$$

El estadístico utilizado para la prueba es:

$$JB = T \left[ \frac{AS}{6} + \frac{K}{24} \right]$$

Este estadístico se compara con una distribución Ji-Cuadrada con  $2T$  grados de libertad.

## 5.6 Predicción

Con los coeficientes estimados del modelo, se puede calcular la predicción para un horizonte  $h$ , dada la información hasta el período  $T$ ; por ejemplo, para un VAR (1) se tiene:

$$\hat{X}_T(1) = \hat{v}_0 + \hat{A}_1 X_T$$

Al horizonte de 2 períodos, la predicción es:

$$\hat{X}_T(2) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(1) = \hat{v}_0 + \hat{A}_1 \hat{v}_0 + \hat{A}_1^2 X_T$$

Al horizonte de 3 períodos, la predicción se escribe:

$$\hat{X}_T(3) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(2) = \left( I + \hat{A}_1 + \hat{A}_1^2 \right) \hat{v}_0 + \hat{A}_1^3 X_T$$

$$\hat{X}_T(h) = \hat{v}_0 + \hat{A}_1 \hat{X}_T(h-1) = \left( I + \hat{A}_1 + \dots + \hat{A}_1^{h-1} \right) \hat{v}_0 + \hat{A}_1^h X_T, \quad h \geq 0$$

Cuando  $h \rightarrow \infty$ , la previsión tiende a un valor constante (estado estacionario) puesto que  $\hat{A}_1^i \rightarrow 0$  si  $i \rightarrow \infty$  y existe el límite de  $\sum_{j=0}^{\infty} \hat{A}_1^j$ , que es igual a  $(I - \hat{A}_1)^{-1}$ . Por tanto:

$$\hat{X}_T(h) \rightarrow (I - \hat{A}_1)^{-1} \hat{v}_0 \quad \text{cuando } h \rightarrow \infty$$

El error de predicción al horizonte  $h$  viene dado por:

$$e_T(h) = X_{T+h} - \hat{X}_T(h)$$

En particular, para  $h=1$  y  $h=2$ , se tiene:

$$e_T(1) = u_{T+1}$$

$$e_T(2) = u_{T+2} + A_1 u_{T+1}$$

En general, para el horizonte  $h$ , se tiene:

$$e_T(h) = \sum_{i=0}^{h-1} A_1^i u_{T+h-i}$$

La esperanza del error de predicción es nula. La matriz de varianza-covarianza del error de predicción es:

$$\sum_u(h) = E \left[ \left( \sum_{i=0}^{h-1} A_1^i u_{T+h-i} \right) \left( \sum_{i=0}^{h-1} A_1^i u_{T+h-i} \right)' \right]$$

La varianza-covarianza estimada viene dada por:

$$\hat{\sum}_u(h) = E \left[ \left( \sum_{i=0}^{h-1} \hat{A}_1^i \hat{u}_{T+h-i} \right) \left( \sum_{i=0}^{h-1} \hat{A}_1^i \hat{u}_{T+h-i} \right)' \right]$$

Luego,

$$\hat{\sum}_T(h) = M_0 \hat{\sum}_u M'_0 + M_1 \hat{\sum}_u M'_1 + \dots + M_{h-1} \hat{\sum}_u M'_{h-1}$$

donde  $M_i$  son las matrices de la representación VMA.

Por lo tanto, se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 + \hat{A}_2 M_0 = \hat{A}_1^2 + \hat{A}_2;$$

$$M_3 = \hat{A}_1 M_2 + \hat{A}_2 M_1 + \hat{A}_3 M_0 = \hat{A}_1^3 + \hat{A}_1 \hat{A}_2 + \hat{A}_2 \hat{A}_1 + \hat{A}_3$$

La varianza del error de predicción para cada una de las predicciones de las  $k$  variables  $(\hat{\sigma}_i^2(h))$  se lee sobre la primera diagonal de la matriz  $\hat{\sum}_u(h)$ . El intervalo de predicción al nivel  $(1 - \alpha)$  está dado por:  $\hat{X}_{iT}(h) \pm z_{1-\frac{\alpha}{2}} \hat{\sigma}_i(h)$  donde  $z_{1-\frac{\alpha}{2}}$  es el cuantil de orden  $(1-\alpha/2)$  de la ley normal.

**Ejemplo 5.5.** Considerando las series del ejemplo 5.1, se busca modelar en su forma VAR. Sin embargo, las series en el ejemplo 5.1 están en niveles y como se pudo observar no son estacionarias; por esta razón, se trabajará con las variaciones trimestrales de las series. Así, se tendrá: la variación trimestral del PIB ( $Y_{1t}$ ), del CI ( $Y_{2t}$ ) y de la DFI ( $Y_{3t}$ ) de un país sudamericano. Nótese que se ha denotado a las variaciones de las series con  $Y$ ; para las variables originales se dejará la notación con  $X$ . Se trabajará con 49 datos, dado que al calcular las variaciones, se pierde el primer dato. La serie  $Y$  iniciará en la observación 2 hasta la observación 50. (Ver Anexo D.1).

Determinar:

- El orden del modelo VAR.
- Los parámetros del modelo.
- La predicción para las 6 siguientes observaciones y dar el intervalo de confianza al 95 %.

**Resolución.**

Se inicia presentando el gráfico de las series:

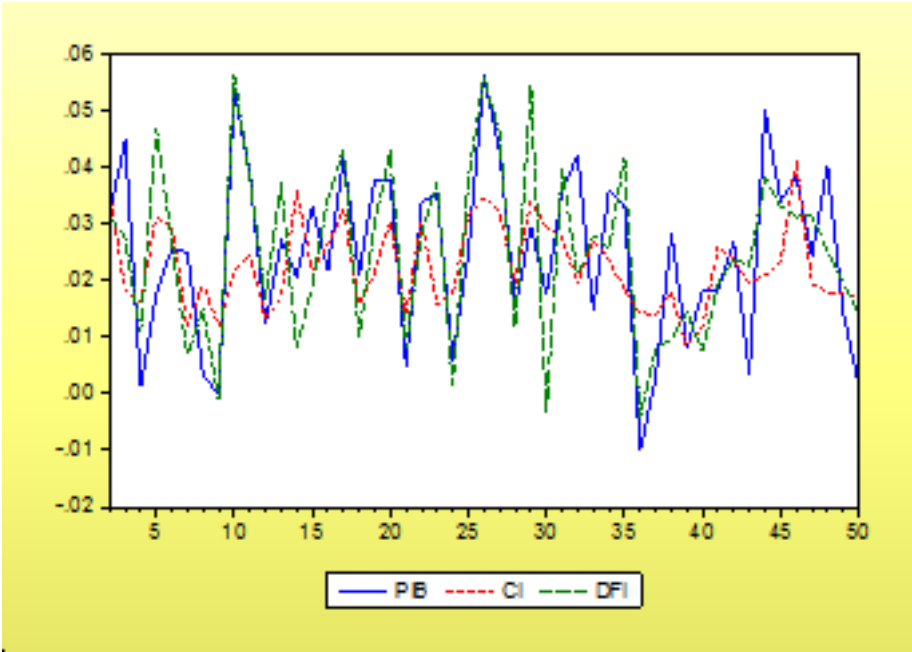


Figura 5.4: Gráfico de las variaciones trimestrales de las series PIB, CI y DFI

Antes de realizar los procedimientos para estimar los modelos, se debe verificar si las series a ser analizadas son estacionarias; para ello se realiza la prueba de raíces unitarias para cada serie utilizando el programa EViews:

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6,689257	0,0000
Test critical values:	1 % level	-4,161144	
	5 % level	-3,506374	
	10 % level	-3,183002	

Tabla 5.2: Prueba DFA para  $Y_{1t}$

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-6,231136	0,0000
Test critical values:	1 % level	-4,161144	
	5 % level	-3,506374	
	10 % level	-3,183002	

Tabla 5.3: Prueba DFA para  $Y_{2t}$

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7,729086	0,0000
Test critical values:		
1 % level	-4,161144	
5 % level	-3,506374	
10 % level	-3,183002	

**Tabla 5.4:** Prueba DFA para  $Y_{3t}$

Como se puede ver en las tablas 5.2, 5.3 y 5.4, las tres series son estacionarias.

- a) Se utilizarán los criterios de Akaike, Schwarz y el logaritmo de máxima verosimilitud para determinar el retardo  $p$  entre 1 y 4. Se deben estimar cuatro modelos diferentes y retener aquel que satisfaga la mayor cantidad de criterios óptimos.

Inicialmente, se tiene un modelo de la forma, para  $p=1$ :

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ Y_{3t} \end{bmatrix} = \begin{bmatrix} a_1^0 \\ a_2^0 \\ a_3^0 \end{bmatrix} + \begin{bmatrix} a_{11}^1 & a_{11}^2 & a_{11}^3 \\ a_{21}^1 & a_{21}^2 & a_{21}^3 \\ a_{31}^1 & a_{31}^2 & a_{31}^3 \end{bmatrix} \begin{bmatrix} Y_{1t-1} \\ Y_{2t-1} \\ Y_{3t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{bmatrix}$$

Algunos paquetes econométricos utilizan la estimación de MCO ecuación por ecuación ya que sólo los valores rezagados de las variables endógenas aparecen en el lado derecho de la ecuación, lo que hace que los estimadores sean eficientes.

Se obtiene lo siguiente:

$$\hat{Y}_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115$$

$$\hat{Y}_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197$$

$$\hat{Y}_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185$$

Con la ayuda del paquete EViews 7, se realiza la estimación de los 4 modelos. Así, se obtuvieron los siguientes resultados:

Criterio / Retardo	1	2	3	4
Log likelihood	484,96	479,76	476,41	477,71
Akaike	-18,92	-18,72	-18,60	-18,67
Schwarz	-18,46	-17,91	-17,43	-17,13

**Tabla 5.5:** Criterios para escoger el retardo del VAR

Como se puede observar en la tabla 5.5, es en el retardo 1 ( $p = 1$ ) donde los criterios de Akaike y Schwarz se minimizan y el valor del *log de verosimilitud* es el máximo. Por lo tanto se realiza la estimación del VAR(1).

En el caso del programa EViews, para calcular el VAR se debe ingresar el número de retardos como un rango; por ejemplo, para un VAR(1) se debe especificar como "1 1", para un



VAR(4) se lo especifica como "1 4", "2 4", "3 4" o "4 4", dependiendo del rango de retardos que se requiera en el modelo.

**Figura 5.5:** Especificación de un modelo VAR en EViews

b) El modelo VAR estimado se escribe:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

(-0,55)
(2,05)
(-0,09)
(-1,66)

$R^2 = 0,09$ ;  $n = 50$ ;  $(.)$  = estadístico correspondiente a la distribución t de Student.

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

(-0,56)
(0,40)
(0,83)
(5,85)

$R^2 = 0,03$ ;  $n = 50$ ;  $(.)$  = estadístico correspondiente a la distribución t de Student.

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

(0,92)
(1,18)
(-1,52)
(2,61)

$R^2 = 0,06$ ;  $n = 50$ ;  $(.)$  = estadístico correspondiente a la distribución t de Student.

Antes de realizar las predicciones, se debe verificar si el modelo cumple con el criterio de estabilidad (las raíces del polinomio característico están fuera del círculo unidad). Con ayuda

del programa EViews se pueden calcular las inversas de las raíces del polinomio característico autoregresivo, las que se espera que se encuentren dentro del círculo unidad. Así se obtiene:

Root	Modulus
-0,246672 - 0,062904i	0,254566
-0,246672 + 0,062904i	0,254566
0,097725	0,097725

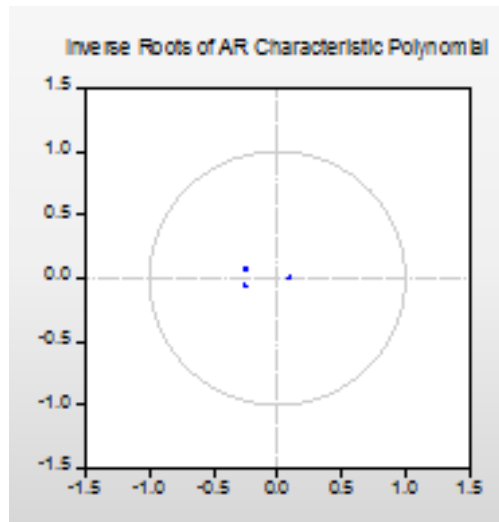


Figura 5.6: Criterio de estabilidad para el VAR(1) estimado

Analítica y gráficamente, se concluye que las inversas de las raíces del polinomio característico se encuentran dentro del círculo unidad; por lo tanto, se concluye que el modelo es estable y, por tanto, es estacionario.

- c) Ahora, se necesita verificar que los residuos del modelo sean ruidos blancos; en general, se prueba la independencia. Para ello, se utilizará el paquete EViews para obtener las pruebas sobre los residuos que se describieron anteriormente. Así se obtiene:

### Prueba Portmanteau

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	0,628382	NA*	0,641751	NA*	NA*
2	9,516110	0,3911	9,915902	0,3573	9
3	16,74629	0,5406	17,62809	0,4804	18
4	25,33301	0,5558	26,99543	0,4640	27
5	35,85563	0,4754	38,74160	0,3470	36
6	41,19646	0,6338	44,84542	0,4784	45
7	51,95004	0,5539	57,43497	0,3491	54
8	62,57741	0,4913	70,18781	0,2494	63
9	72,52842	0,4604	82,43521	0,1879	72
10	80,22987	0,5033	92,16336	0,1863	81
11	88,06759	0,5380	102,3312	0,1763	90
12	89,89910	0,7324	104,7732	0,3264	99
13	94,98418	0,8100	111,7471	0,3831	108
14	99,40624	0,8789	117,9900	0,4570	117
15	103,7800	0,9263	124,3517	0,5248	126
16	115,1927	0,8904	141,4708	0,3343	135
17	118,4536	0,9412	146,5199	0,4259	144
18	127,4406	0,9348	160,8991	0,3149	153
19	130,5767	0,9668	166,0899	0,3965	162
20	136,6630	0,9751	176,5237	0,3701	171

**Tabla 5.6:** Prueba de autocorrelación Pormanteau

En la tabla 5.6 se observa que los p-valores (Prob.) para los retardos 2 al 10 son no significativos; de esto se concluye que los residuos no están autocorrelacionados.

## Prueba LM

Lags	LM-Stat	Prob
1	6,309247	0,7086
2	8,996282	0,4376
3	7,384406	0,5972
4	8,563089	0,4785
5	11,86241	0,2212
6	5,615736	0,7777
7	13,14360	0,1562
8	12,15696	0,2046
9	13,95395	0,1240
10	9,385154	0,4025
11	9,988141	0,3514
12	2,713061	0,9746
13	5,772031	0,7625
14	6,030074	0,7369
15	6,067111	0,7332
16	16,42881	0,0584
17	5,150654	0,8210
18	13,63067	0,1361
19	4,863835	0,8460
20	7,838041	0,5505

Tabla 5.7: Prueba LM

Por los valores en la columna Prob. (ver tabla 5.7), se puede concluir que no existe correlación serial. Esto confirma que no hay que reformular el modelo planteado.

### Pruebas de Normalidad

Component	Jarque-Bera	df	Prob.
1	1,142571	2	0,5648
2	2,337811	2	0,3107
3	0,393216	2	0,8215
Joint	3,873599	6	0,6938

Tabla 5.8: Prueba de Normalidad de los residuos

De la prueba de Jarque-Bera se concluye que la distribución de los residuos es una distribución normal multivariante.

**Observación 5.5.** De los resultados obtenidos en los literales b y c se concluye que el modelo VAR(1) es adecuado para los datos.

d) La predicción calculada por el modelo, de manera recurrente es:

$$\hat{Y}_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115$$

$$\hat{Y}_{1:51} = -0,1168 * 0,0015 + 0,7168 * 0,0167 - 0,0198 * 0,0142 + 0,0115$$

$$\hat{Y}_{1:51} = 0,023$$

$$\hat{Y}_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197$$

$$\hat{Y}_{2:51} = -0,0575 * 0,0015 + 0,0674 * 0,0167 + 0,0900 * 0,0142 + 0,0197$$

$$\hat{Y}_{2:51} = 0,022$$

$$\hat{Y}_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185$$

$$\hat{Y}_{3:51} = 0,1972 * 0,0015 + 0,4192 * 0,0167 - 0,3462 * 0,0142 + 0,0185$$

$$\hat{Y}_{3:51} = 0,021$$

donde,  $\hat{Y}_{i:j}$  significa, la previsión de la variable  $Y_i$  para el período  $j$ .

**Observación 5.6.** A pesar que se trabaja con 49 datos, la última observación es la número 50, por los motivos explicados en el enunciado del ejemplo; es por esto, que la primera observación a predecir es la de orden 51 aunque en realidad correspondería al dato 50 de una serie temporal que inicie en el instante  $t=1$ .

De la misma manera se obtiene:

$$\hat{Y}_{1:52} = 0,024$$

$$\hat{Y}_{2:52} = 0,022$$

$$\hat{Y}_{3:52} = 0,025$$

$$\hat{Y}_{1:53} = 0,023$$

$$\hat{Y}_{2:53} = 0,022$$

$$\hat{Y}_{3:53} = 0,023$$

$\vdots$

Para calcular la varianza del error de predicción, si fuera un VAR (1) se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 = \hat{A}_1^2; \quad etc \dots$$

Para el caso del ejemplo, se obtiene:

$$\hat{A}_1 = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,04192 & -0,3462 \end{bmatrix}$$

La matriz de varianza covarianza estimada de la predicción, para el horizonte  $h=1$ , es:

$$\hat{\Sigma}_T(1) = \hat{\Sigma}_u = \begin{bmatrix} 0,0002 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0002 \end{bmatrix}$$

Así, la varianza del error de predicción para  $\hat{Y}_{1:51}$  es igual a 0,0002, la varianza del error de predicción para  $\hat{Y}_{2:51}$  es igual a 0,0001 y la varianza del error de predicción para  $\hat{Y}_{3:51}$  es igual a 0,0002.

Los intervalos de confianza para  $Y_{1:51}$ ,  $Y_{2:51}$  y  $Y_{3:51}$  vienen dados, respectivamente, por:

$$0,024 \pm 1,96 * \sqrt{0,0002} = [-0,007; 0,053]$$

$$0,022 \pm 1,96 * \sqrt{0,0001} = [0,007; 0,037]$$

$$0,025 \pm 1,96 * \sqrt{0,0002} = [-0,009; 0,051]$$

Para los horizontes  $h = 2$ ,  $h = 3$  se utilizan las siguientes fórmulas:

$$\hat{\Sigma}_T(2) = \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' = \begin{bmatrix} 0,0003 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0003 \end{bmatrix}$$

$$\hat{\Sigma}_T(3) = \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \hat{A}_1^2 \hat{\Sigma}_u \hat{A}_1^{2'} = \begin{bmatrix} 0,0005 & 0,0001 & 0,0004 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0004 & 0,0001 & 0,0005 \end{bmatrix}$$

Entonces los intervalos de confianza son:

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,016 = [-0,007; 0,056]$$

$$IC(Y_{2:52}) = 0,022 \pm 1,96 * 0,008 = [0,007; 0,037]$$

$$IC(Y_{3:52}) = 0,025 \pm 1,96 * 0,016 = [-0,006; 0,056]$$

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,023 = [-0,020; 0,068]$$

$$IC(Y_{1:52}) = 0,022 \pm 1,96 * 0,011 = [0,001; 0,043]$$

$$IC(Y_{1:52}) = 0,024 \pm 1,96 * 0,023 = [-0,020; 0,068]$$

$\vdots$

### Comparación con modelos univariantes

La teoría VAR sugiere que las predicciones logradas son de mejor calidad que si se realiza la modelación de las series de manera univariante. Para comprobar esto, se realizó un modelo univariante para cada una de las series analizadas en este documento. Así, para las variaciones PIB se encontró el modelo  $Y_{1t} = 0,94Y_{1t-5} + \hat{u}_t - 0,87\hat{u}_{1t-5}$ ; para la variación del CI se estimó el modelo  $Y_{2t} = 0,97Y_{2t-7} + \hat{u}_t - 0,88\hat{u}_{2t-7}$  y para la variación de la DFI se obtuvo el modelo  $Y_{3t} = 0,36Y_{3t-2} + 0,53Y_{3t-6} + \hat{u}_t$ .

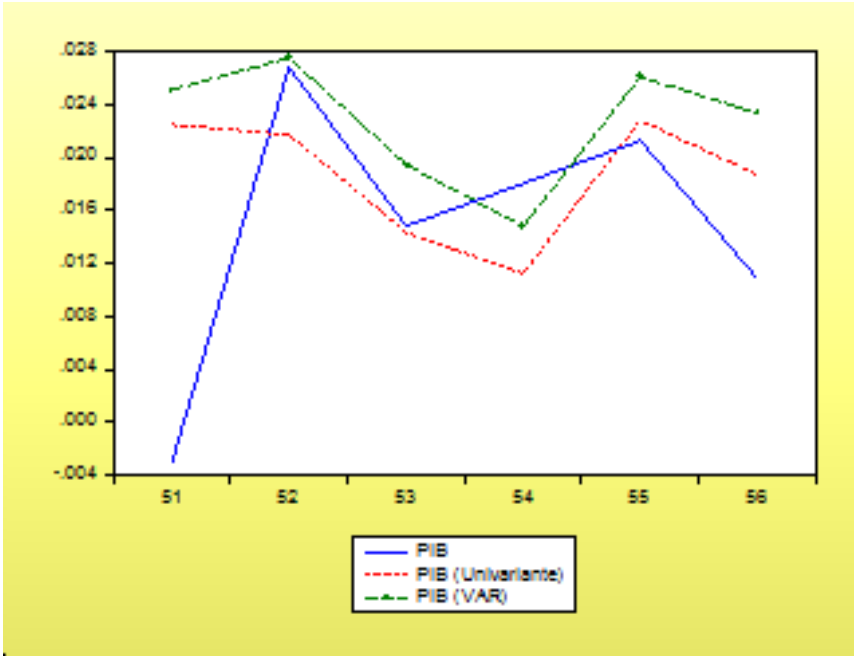


Figura 5.7: Comparación de las predicciones VAR y UNIVARIANTE para  $Y_{1t}$  (PIB)

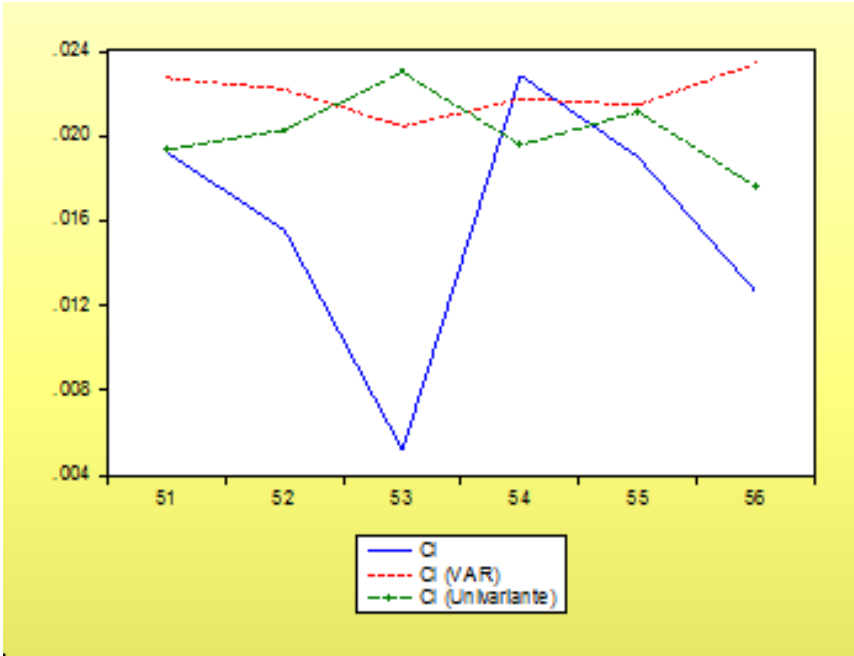


Figura 5.8: Comparación de las predicciones VAR y UNIVARIANTE para  $Y_{2t}$  (CI)

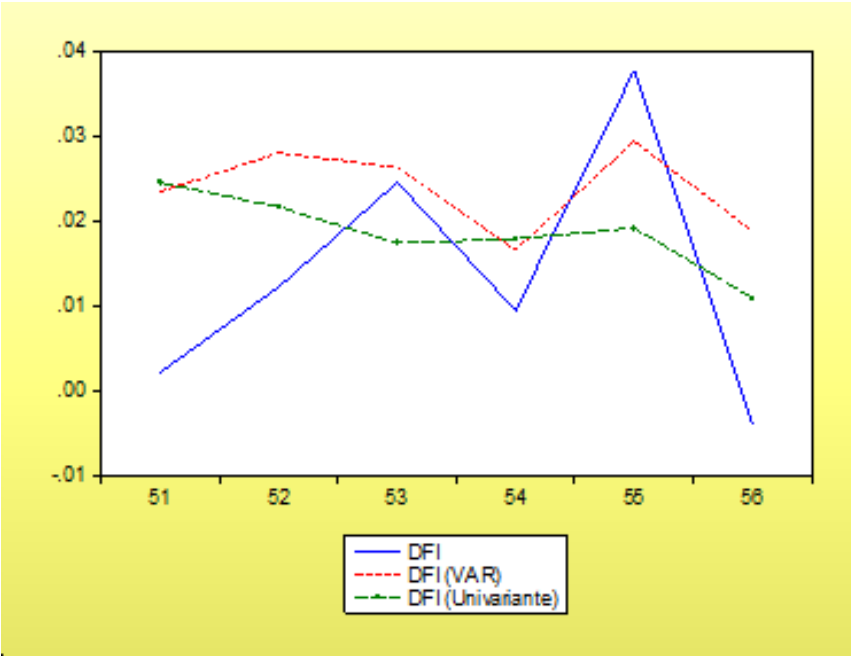


Figura 5.9: Comparación de las predicciones VAR y UNIVARIANTE para  $Y_{3t}$  (DFI)

Se puede observar que el ajuste que tienen las predicciones del modelo VAR; pero, para comparar analíticamente se calcula del error cuadrático medio estimado, para determinar el mejor ajuste. Así se tiene:

	PIB	CI	DFI
VAR	0,0005	0,0005	0,0006
UNIVARIANTE	0,0004	0,0004	0,0004

Tabla 5.9: Error Medio Cuadrático estimado para los modelos VAR y univariante

Como se puede observar, los errores son muy pequeños para ambos casos; la diferencia existente es estadísticamente no significativa entre los dos tipos de predicciones. Esto se explica porque las correlaciones cruzadas entre las series son muy poco significativas; sin embargo, el objetivo de esta presentación, es más bien, didáctico; posteriormente se mostrará un ejemplo más completo.

### 5.7 La Causalidad

En la teoría, la demostración de relaciones causales entre las variables de análisis proporciona los elementos de reflexión propicios para una mejor comprensión de los fenómenos, sobre todo los económicos. De manera práctica, *“the causal knowledge”* (“el conocimiento de la causalidad”) es necesario en la formulación correcta de la política económica. En efecto, saber la dirección de la causalidad es también importante en cuanto a poner un enlace entre las variables económicas.



### 5.7.1 La causalidad según Granger

Granger (1969) propuso los conceptos de causalidad y de exogeneidad: en el sentido de series de tiempo, se diría que la variable  $X_{2t}$  es la causa de  $X_{1t}$ , si la predicción de  $X_{1t}$  mejora si la información relativa a  $X_{2t}$  se incorpora al análisis (*el término predicción parece preferible en el marco de la causalidad; en efecto, decir que  $Y_t$  causa  $X_t$ , solo significa que es preferible para predecir  $X_t$  conocer  $Y_t$ , que no conocerla*).

Sea el modelo VAR( $p$ ) para el cual las variables  $X_{1t}$  y  $X_{2t}$  son estacionarias:

$$\begin{bmatrix} X_{1t} \\ X_{2t} \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \begin{bmatrix} a_1^1 & b_1^1 \\ a_1^2 & b_1^2 \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \end{bmatrix} + \begin{bmatrix} a_2^1 & b_2^1 \\ a_2^2 & b_2^2 \end{bmatrix} \begin{bmatrix} X_{1t-2} \\ X_{2t-2} \end{bmatrix} + \dots + \\ \dots + \begin{bmatrix} a_p^1 & b_p^1 \\ a_p^2 & b_p^2 \end{bmatrix} \begin{bmatrix} X_{1t-p} \\ X_{2t-p} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix}$$

El bloque de variables ( $X_{2t-1}, X_{2t-2}, \dots, X_{2t-p}$ ) se considera como exógeno en comparación del bloque de variables ( $X_{1t-1}, X_{1t-2}, \dots, X_{1t-p}$ ) si el hecho de añadir el bloque  $X_{2t}$  no mejora significativamente la determinación de las variables  $X_{1t}$ . Se trata de efectuar una prueba de restricciones sobre los coeficientes de las variables  $X_{2t}$  de la representación VAR (Se denotará por RVAR al modelo VAR restringido). La determinación del retardo  $p$  se efectúa por los criterios de Akaike, Shwarz o Hanan-Quinn.

- $X_{2t}$  no causa  $X_{1t}$ , si se acepta la siguiente hipótesis:

$$H_0 : b_1^1 = b_2^1 = \dots = b_p^1 = 0$$

- $X_{1t}$  no causa  $X_{2t}$ , si se acepta la siguiente hipótesis:

$$H_0 : a_1^2 = a_2^2 = \dots = a_p^2 = 0$$

Si se llegan a aceptar las dos hipótesis:  $X_{1t}$  causa a  $X_{2t}$  y viceversa, se habla de efectos de retroalimentación ("*feedback effect*").

Estas pruebas pueden llevarse a cabo con la ayuda de la prueba clásica de Fisher de la nulidad de los coeficientes (prueba de Wald), ecuación por ecuación o bien directamente comparando un modelo VAR sin restricciones (UVAR) y el modelo VAR restringido (RVAR).

Se calcula el siguiente estadístico:

$$L^* = (T - c) (\ln |\Sigma_{RVAR}| - \ln |\Sigma_{UVAR}|) \sim \chi^2(2p)$$

donde:

$\Sigma_{RVAR}$  = matriz de varianzas-covarianzas de los residuos del modelo restringido,

$\Sigma_{UVAR}$  = matriz de varianzas-covarianzas de los residuos del modelo sin restricciones,

$T$  = número de observaciones,

$c$  = número de parámetros estimados de cada ecuación del modelo sin restricciones.

Si  $L^* > \chi^2_{1-\alpha}(2p)$ , entonces se rechaza la hipótesis de la validez de la restricción con un nivel de significación  $\alpha$ .

También se suele utilizar la prueba tradicional de Fisher:

$$F^* = \frac{\frac{SCRR - SCRU}{c}}{\frac{SCRU}{n-k-1}}$$

SRU: es la suma de cuadrados de los residuos del modelo sin restricciones.

SRRR: es la suma de cuadrados de los residuos del modelo restringido.

C= número de restricciones (número de coeficientes que pone a prueba la hipótesis nula).

Si  $F^* > F_{c;n-k-1}^\alpha$  se rechaza la hipótesis nula.

## 5.7.2 La causalidad según Sims

Sims (1980) presenta una especificación de prueba ligeramente diferente. Se considera que si los valores futuros de  $X_{1t}$  permiten explicar los valores presentes de  $X_{2t}$ , entonces  $X_{2t}$  es la causa de  $X_{1t}$ .

Esto se representa de la siguiente manera:

$$X_{1t} = a_1^0 + \sum_{i=1}^p a_{1i}^1 X_{1t-i} + \sum_{i=1}^p a_{1i}^2 X_{2t-i} + \sum_{i=1}^p b_i^2 X_{2t+i} + u_{1t}$$

$$X_{2t} = a_2^0 + \sum_{i=1}^p a_{2i}^1 X_{1t-i} + \sum_{i=1}^p a_{2i}^2 X_{2t-i} + \sum_{i=1}^p b_i^1 X_{1t+i} + u_{2t}$$

- $X_{1t}$  no causa  $X_{2t}$ , si se aceptan la siguiente hipótesis:

$$H_0 : b_1^2 = b_2^2 = \dots = b_p^2 = 0$$

- $X_{2t}$  no causa  $X_{1t}$ , si se aceptan la siguiente hipótesis:

$$H_0 : b_1^1 = b_2^1 = \dots = b_p^1 = 0$$

Se sigue utilizando la prueba clásica de Fisher de nulidad de coeficientes.

**Observación 5.7.** La prueba de Sims, se deja de como referencia dado que no se puede estimar directamente en EViews. Se puede crear un código de programación dentro del programa para poder realizar la estimación, pero esto está fuera del alcance de este documento.

**Ejemplo 5.6.** A partir de la representación VAR (1) estimada en el ejemplo 5.5, se va a proceder a realizar las pruebas de Granger.

### Resolución.

Se procede con una prueba de Fisher, ecuación por ecuación.

### Prueba de Granger

- $H_0$ :  $Y_{2t}$  y  $Y_{3t}$  no causan  $Y_{1t}$

Se han estimado los siguientes modelos:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

$$R^2 = 0,09; \quad n = 48; \quad SCR U = 0,010381 \text{ (sin restricciones)}$$

$$Y_{1t} = -0,0197Y_{1t-1} + 0,025 + \hat{u}_{1t}$$

$$R^2 = 0,0004; \quad n = 48; \quad SCRR = 0,011474 \text{ (restringido)}$$

donde,

SCR U: es la suma de cuadrados de los residuos del modelo sin restricciones.

SCRR: es la suma de cuadrados de los residuos del modelo restringido.

$$F^* = \frac{\frac{SCR R - SCR U}{c}}{\frac{SCR U}{n-k-1}} = \frac{\frac{0,011474 - 0,010381}{1}}{\frac{0,010381}{48-3-1}} = 4,6326$$

c= número de restricciones (número de coeficientes que pone a prueba la hipótesis nula); en este caso c=1, dado que se elimina un coeficiente de cada ecuación.

$F^* > F_{1;48}^{0,05} \approx 4,05$ ; por tanto, se rechaza la hipótesis nula;  $Y_{2t}$  y  $Y_{3t}$  explica significativamente la variable  $Y_{1t}$ ; existe causalidad según Granger.

- Ho:  $Y_{1t}$  y  $Y_{3t}$  no causan  $Y_{2t}$

Se han estimado los siguientes modelos:

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

$$R^2 = 0,0026; \quad n = 48; \quad SCR U = 0,002594 \text{ (sin restricciones)}$$

$$Y_{2t} = 0,1014Y_{2t-1} + 0,0199 + \hat{u}_{2t}$$

$$R^2 = 0,011; \quad n = 48; \quad SCRR = 0,002638 \text{ (restringido)}$$

$$F^* = \frac{\frac{SCR R - SCR U}{c}}{\frac{SCR U}{n-k-1}} = \frac{\frac{0,002638 - 0,002594}{1}}{\frac{0,002594}{48-3-1}} = 0,7463$$

$F^* < F_{1;104}^{0,05} \approx 4,05$ ; por tanto, se acepta la hipótesis nula;  $Y_{1t}$  y  $Y_{3t}$  no explican significativamente la variable  $Y_{2t}$ ; no existe causalidad según Granger.

- Ho:  $Y_{1t}$  y  $Y_{2t}$  no causan  $Y_{3t}$

Se han estimado los siguientes modelos:

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

$$R^2 = 0,0110; \quad n = 48; \quad SCR U = 0,011036 \text{ (sin restricciones)}$$

$$Y_{3t} = -0,1394Y_{3t-1} + 0,0285 + \hat{u}_{3t}$$

$R^2 = 0,011; \quad n = 48; \quad SCRR = 0,011487 \text{ (restringido)}$

$$F^* = \frac{\frac{SCRR - SCRU}{c}}{\frac{SCRU}{n-k-1}} = \frac{\frac{0,011487 - 0,011036}{1}}{\frac{0,011036}{48-3-1}} = 1,7981$$

$F^* < F_{1;104}^{0,05} \approx 4,05$ ; por tanto, se acepta la hipótesis nula;  $Y_{1t}$  y  $Y_{2t}$  no explican significativamente la variable  $Y_{3t}$ ; no existe causalidad según Granger.

El siguiente gráfico muestra la salida del paquete EViews para la Prueba de Granger. Hay que considerar que el paquete considera la prueba de Wald y compara el valor obtenido con el estadístico Chi-cuadrado.

Dependent variable: $Y_{1t}$			
Excluded	Chi-sq	df	Prob.
$Y_{2t}$	3.972118	1	0.0463
$Y_{3t}$	0.007279	1	0.9320
All	4.631706	2	0.0987

Dependent variable: $Y_{2t}$			
Excluded	Chi-sq	df	Prob.
$Y_{2t}$	0.557515	1	0.4553
$Y_{3t}$	0.661477	1	0.4160
All	0.746494	2	0.6885

Dependent variable: $Y_{3t}$			
Excluded	Chi-sq	df	Prob.
$Y_{2t}$	0.236725	1	0.6266
$Y_{3t}$	1.415327	1	0.2342
All	1.799419	2	0.4067

Tabla 5.10: Prueba de causalidad de Granger

Como se puede ver, el p-valor (Prob.) es mayor que 0,05 en todos los casos, esto implica que no existe causalidad entre las variables. Sin embargo, al realizar el cálculo inicial se concluyó que si existe causalidad de  $Y_{3t}$  y  $Y_{2t}$  hacia  $Y_{1t}$ . La diferencia se origina en los algoritmos de cálculo que tienen los paquetes implementados.

### 5.8 Análisis de los “choques”

El análisis de los choques consiste en medir el impacto de la variación de una innovación sobre las variables.

Considérese el modelo estimado del ejemplo 5.5:

$$Y_{1t} = -0,1168Y_{1t-1} + 0,7168Y_{2t-1} - 0,0198Y_{3t-1} + 0,0115 + \hat{u}_{1t}$$

$$Y_{2t} = -0,0575Y_{1t-1} + 0,0674Y_{2t-1} + 0,0900Y_{3t-1} + 0,0197 + \hat{u}_{2t}$$

$$Y_{3t} = 0,1972Y_{1t-1} + 0,4192Y_{2t-1} - 0,3462Y_{3t-1} + 0,0185 + \hat{u}_{3t}$$

Una variación en un instante dado de  $\hat{u}_{1t}$  tiene una consecuencia inmediata sobre  $Y_{1t}$ , y entonces sobre  $Y_{1,t+1}$  y  $Y_{2,t+1}$ ; por ejemplo, si se produce en  $t$  un choque sobre  $\hat{u}_{1t}$  igual a 1 y de orden 0 sobre  $\hat{u}_{2t}$  (las otras variables permanecen iguales que en el tiempo  $(t-1)$ ; para los valores subsiguientes,  $u_{t+i}$  retorna a su valor  $u_{t-1}$ ), se tiene los impactos siguientes:

En el período  $t$ :

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

En el período  $t+1$ :

$$\begin{bmatrix} \Delta \hat{Y}_{1,t+1} \\ \Delta \hat{Y}_{2,t+1} \\ \Delta \hat{Y}_{3,t+1} \end{bmatrix} = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -0,1170 \\ -0,0575 \\ 0,1972 \end{bmatrix}$$

En el período  $t+2$ :

$$\begin{bmatrix} \Delta \hat{Y}_{1,t+2} \\ \Delta \hat{Y}_{2,t+2} \\ \Delta \hat{Y}_{3,t+2} \end{bmatrix} = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix} \begin{bmatrix} -0,1170 \\ -0,0575 \\ 0,1972 \end{bmatrix} = \begin{bmatrix} -0,0315 \\ 0,0206 \\ -0,1154 \end{bmatrix}$$

Por otro lado, utilizando la representación lineal del VAR, se logra realizar el análisis de las funciones de impulso-respuesta, ya que un choque en  $Y_{1t}$  se reflejará como la primera columna de  $M_i = A_1^i$  (recuerde que  $M_0 = I$ ); de forma similar, un impacto sobre  $Y_{2t}$  se reflejará en la segunda columna de la matriz  $M_i$ . La generalización a un VAR con  $k$  variables, es inmediata.

Considere el modelo estimado en el ejemplo 5.5:

$$\hat{A}_1 = \begin{bmatrix} -0,1170 & 0,7168 & -0,0198 \\ -0,0575 & 0,0674 & 0,0900 \\ 0,1972 & 0,4192 & -0,3462 \end{bmatrix}$$

$$\hat{A}_1^2 = \begin{bmatrix} -0,0315 & -0,0437 & 0,0737 \\ 0,0206 & 0,0011 & -0,0240 \\ -0,1154 & 0,0245 & 0,1567 \end{bmatrix}$$

Así, los elementos de  $\hat{A}_1^i$  representan los efectos de los choques unitarios de las variables del sistema luego de  $i$  periodos. Es por esta razón que, se les conoce como **multiplicadores dinámicos o respuestas al impulso**.

En las tablas siguientes se muestran los resultados de las funciones de impulso-respuesta que presenta el paquete EViews para las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$ , que son los cálculos hechos manualmente al inicio de esta sección. Esto también se muestra en los gráficos siguientes:

Período	$Y_{1t}$	$Y_{2t}$	$Y_{3t}$
1	1,000000	0,000000	0,000000
2	-0,193069	-0,080383	0,108038
3	-0,016093	0,019189	-0,090439
4	0,014628	-0,005392	0,036103
5	-0,005860	0,001663	-0,012563
6	0,002041	-0,000533	0,004198
7	-0,000682	0,000173	-0,001384
8	0,000225	-5,66E-05	0,000454
9	-7,38E-05	1,85E-05	-0,000149
10	2,42E-05	-6,06E-06	4,88E-05

**Tabla 5.11:** Respuesta de las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  ante un choque unitario de  $Y_{1t}$

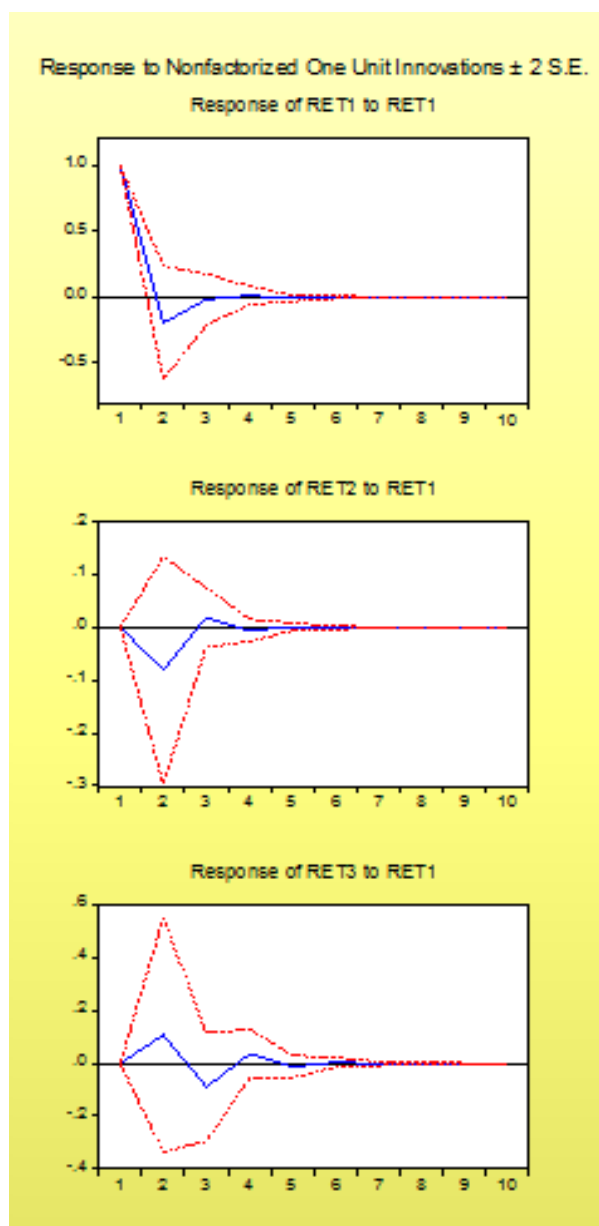


Figura 5.10: Respuesta de las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  ante un choque unitario de  $Y_{1t}$

Período	$Y_{1t}$	$Y_{2t}$	$Y_{3t}$
1	0.000000	1.000000	0.000000
2	0.689275	0.075128	0.424214
3	-0.073293	-0.011642	-0.032978
4	0.005504	0.002053	-0.002027
5	0.000315	-0.000470	0.002131
6	-0.000345	0.000131	-0.000865
7	0.000140	-4.02E-05	0.000303
8	-4.91E-05	1.29E-05	-0.000101
9	1.65E-05	-4.18E-06	3.34E-05
10	-5.43E-06	1.36E-06	-1.10E-05

**Tabla 5.12:** Respuesta de las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  ante un choque unitario de  $Y_{2t}$



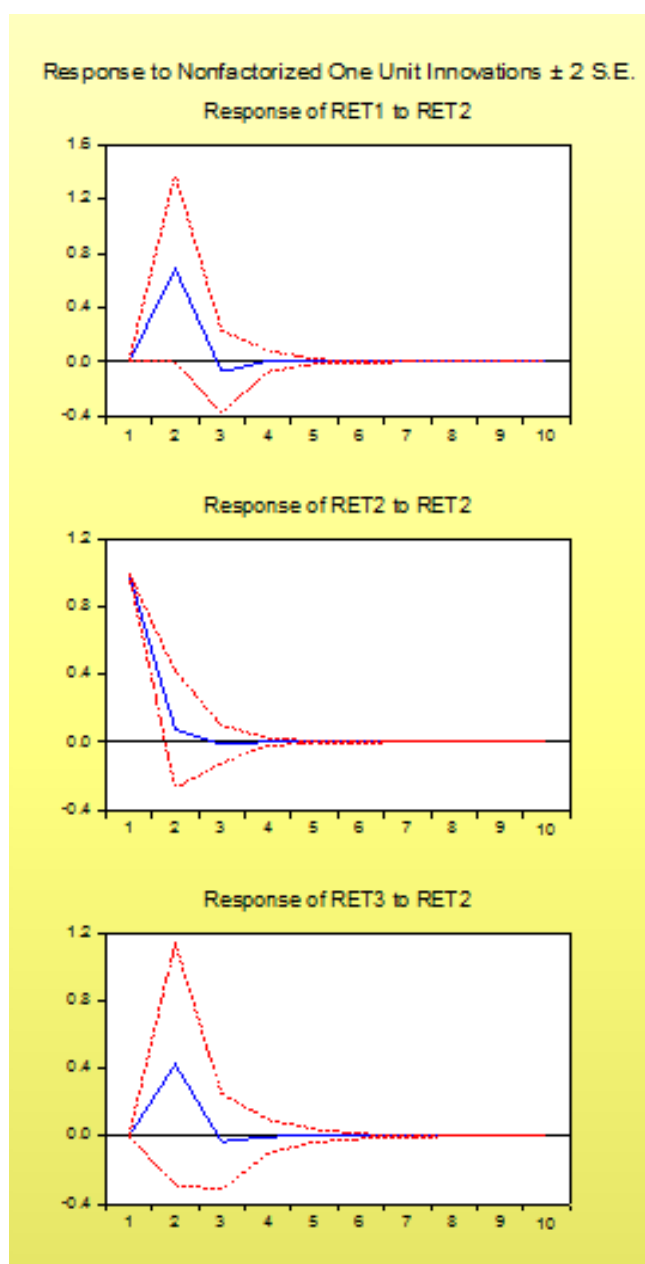
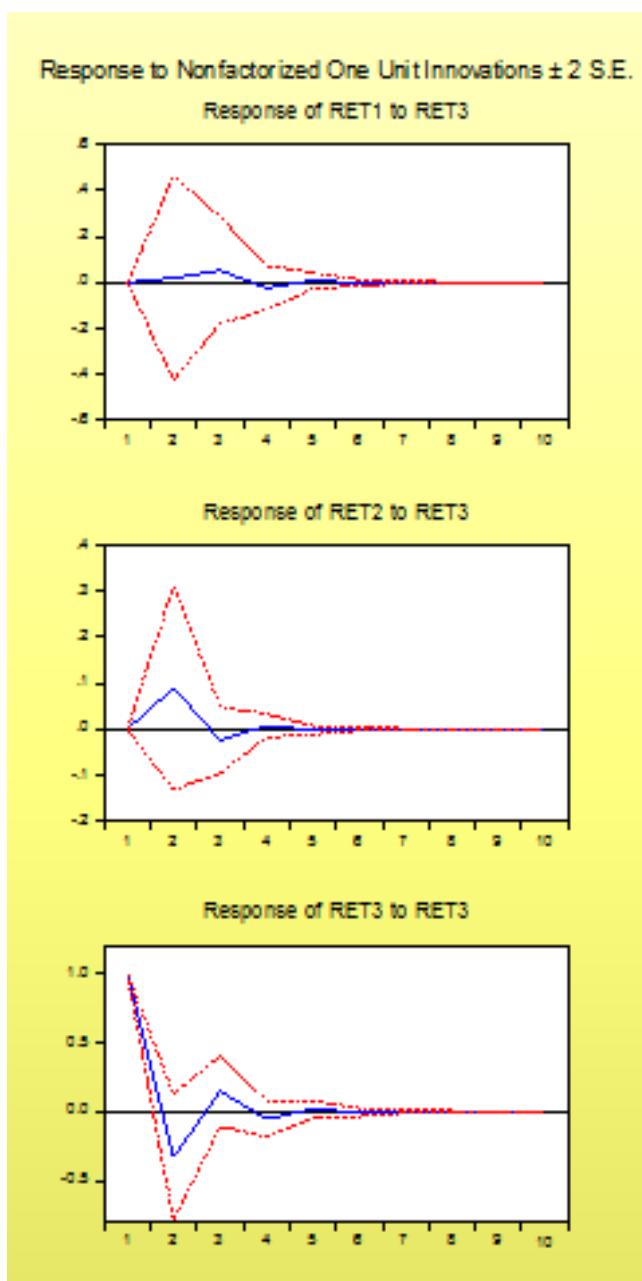


Figura 5.11: Respuesta de las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  ante un choque unitario de  $Y_{2t}$ .

Período	$Y_{1t}$	$Y_{2t}$	$Y_{3t}$
1	0.000000	0.000000	1.000000
2	0.018858	0.089860	-0.328412
3	0.052104	-0.024276	0.148011
4	-0.024001	0.007288	-0.053278
5	0.008653	-0.002311	0.017996
6	-0.002924	0.000748	-0.005955
7	0.000968	-0.000244	0.001957
8	-0.000318	7.98E-05	-0.000642
9	0.000104	-2.61E-05	0.000210
10	-3.42E-05	8.55E-06	-6.88E-05

**Tabla 5.13:** Respuesta de las variables  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$  ante un choque unitario de  $Y_{3t}$



La elección de la dirección del impacto es muy importante y determina los valores obtenidos. Se puede observar que el efecto de la innovación se desvanece con el tiempo; esto caracteriza a un proceso VAR estacionario.

- i. Si las variables están medidas en escalas diferentes es común considerar las innovaciones iguales a su desviación típica en lugar de los choques unitarios; es decir:

$$\hat{u}_{t,0} = \sqrt{\text{Var}(\hat{u}_t)}$$

Esto constituye únicamente un reescalamiento de las funciones de impulso-respuesta.

- Si una variables no causa (en el sentido de Granger) al resto de variables en el sistema, entonces las respuestas al impulso sobre las otras variables serán cero.

## 5.9 Descomposición de la varianza

### 5.9.1 Representación de errores ortogonales

Un problema a considerar en el análisis de las funciones de impulso-respuesta es el de la correlación contemporánea de errores y, por lo tanto, el impacto de un choque sobre una variable puede acompañarse de un impacto en otra variable; ignorarla, puede distorsionar la verdadera relación dinámica entre las variables. Es por esto que se trata de manera general de realizar el análisis a través de la búsqueda de una representación de errores ortogonales.

Considerando la representación lineal del VAR, se puede obtener lo siguiente:

1. Dado que  $\Sigma_u$  es simétrica y definida positiva, entonces existe  $P$  no singular tal que:

$$\Sigma_u = PP'$$

donde,  $P$  es una matriz triangular obtenida a través de la descomposición de Cholesky.

2. Luego, se puede expresar lo siguiente:

$$v_t = Pu_t$$

donde,  $v_t$  es un vector aleatorio con  $E(v_t) = 0$  y  $V(v_t) = I$ .

Entonces,  $v_t$  son las *innovaciones ortogonales* de  $u_t$ . Luego, reemplazando:

$$X_t = \sum_{i=1}^{\infty} \theta_i Pu_{t-i} = \sum_{i=1}^{\infty} M_i u_{t-i}, \quad M_i \equiv \theta_i P$$

Entonces,  $\theta_0 = P$ ; por lo que, salvo el caso que  $\Sigma_u$  sea diagonal,  $\Sigma_v$  no será diagonal y sus elementos recogerán las respuestas inmediatas del sistema de choques unitarios. Es por esto que se los conoce como **multiplicadores de impacto**. Además, el hecho que  $\theta_0 = P$  sea una matriz triangular, implica que el orden de las variables en el vector es importante (órdenes diferentes de descomposición pueden producir funciones de impluso respuesta diferentes).

### 5.9.2 Descomposición de la Varianza

La descomposición de la varianza del error de predicción tiene como objetivo calcular para cada una de las innovaciones su contribución a la varianza del error. Por la técnica matemática de la descomposición de Cholesky de la matriz  $\Sigma_u$  que es simétrica y definida positiva, se puede escribir la varianza del error de predicción para un horizonte  $h$  en función de la varianza del error atribuida a cada una de las variables; es suficiente dividir cada una de estas variaciones por la varianza total, para obtener su peso relativo en porcentaje.

Se retoma el modelo VAR (1) de dos variables  $X_{1t}$  y  $X_{2t}$ . La varianza del error de predicción para  $X_{1,t+h}$  se puede escribir:

$$\sigma_{x_h}^2 = \sigma_{u_1}^2 \left[ m_{11}^2(0) + \dots + m_{11}^2(h-1) \right] + \sigma_{u_2}^2 \left[ m_{22}^2(0) + \dots + m_{22}^2(h-1) \right]$$

Donde los  $m_{ii}^{(j)}$  son los términos de las matrices  $M_i$  (representación lineal del proceso).

Al horizonte  $h$ , la descomposición de la varianza, en porcentaje, de las innovaciones propias de  $X_{1t}$  sobre  $X_{1t}$ , están dadas por:

$$\frac{\sigma_{u_1}^2 \left[ m_{11}^2(0) + \dots + m_{11}^2(h-1) \right]}{\sigma_{X_1}^2(h)} * 100$$

Y la descomposición de la varianza, en porcentaje, de  $X_{2t}$  sobre  $X_{1t}$

$$\frac{\sigma_{u_2}^2 \left[ m_{22}^2(0) + \dots + m_{22}^2(h-1) \right]}{\sigma_{X_1}^2(h)} * 100$$

La interpretación de los resultados es importante:

- Si un choque sobre  $u_{1t}$  no afecta la varianza del error de  $X_{2t}$  independientemente del horizonte de predicción, entonces  $X_{2t}$  puede considerarse como exógena porque  $X_{2t}$  evoluciona independientemente de  $u_{1t}$ .
- En caso contrario, si un choque sobre  $u_{1t}$  afecta fuertemente, en realidad totalmente, la varianza del error de  $X_{2t}$ , entonces  $X_{2t}$  se considera como endógena.

### 5.9.3 Elección del orden de descomposición

Nótese que el problema de la correlación contemporánea de los errores y, por lo tanto, el impacto de un choque sobre una variable, implica una elección de descomposición que proporciona resultados asimétricos en función del orden de las variables. El problema es más complejo si el número de las variables es importante.

**Ejemplo 5.7.** Considérese un modelo VAR con 4 variables  $X_1$ ,  $X_2$ ,  $X_3$  y  $X_4$ .

Supóngase que se elige el siguiente orden para la descomposición de Cholesky:  $X_2 X_3 X_1 X_4$  y esto lleva a obtener la siguiente tabla hipotética:

Ordenación de Cholesky: $X_2 X_3 X_1 X_4$				
Respuesta de $X_2$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	0,000000	4,583291	0,000000	0,000000
2	0,775767	-0,251545	0,815017	-0,905811
Respuesta de $X_3$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	0,000000	2,439203	4,54469	0,000000
2	-0,441257	-0,754324	-3,564595	-0,566602
Respuesta de $X_1$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	4,603662	3,022459	1,802923	0,000000
2	-0,795786	-0,101091	0,388095	-0,255476
Respuesta de $X_4$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	1,568803	0,486351	1,415711	3,191513
2	-0,329818	-4,104756	-0,527459	-3,186928

**Tabla 5.14:** Descomposición hipotética de choques

La interpretación de la tabla 5.14 se haría de la siguiente manera:

- Un choque para el período 1 sobre  $X_2$  tiene un impacto solamente sobre  $X_2$  y ausencia de correlación contemporánea con  $X_3$ ,  $X_1$  y  $X_4$ .
- Un choque para el período 1 sobre  $X_3$  tiene un impacto sobre  $X_2$  y  $X_3$  y ausencia de correlación contemporánea con  $X_1$  y  $X_4$ .
- Un choque para el período 1 sobre  $X_1$  tiene un impacto sobre  $X_2$ ,  $X_3$  y  $X_1$  y ausencia de correlación contemporánea con  $X_4$ .
- Finalmente, un choque para el período 1 sobre  $X_4$  tiene un impacto sobre todas las variables.

Ahora, si se realiza el análisis considerando un orden diferente de las variables para la descomposición, se tendrá:

Ordenación de Cholesky: $X_3X_4X_2X_1$				
Respuesta de $X_3$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	0,000000	0,000000	2,805461	0,000000
2	-0,527459	-3,186928	0,815017	-0,905811
Respuesta de $X_4$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	0,000000	0,000000	-2,231567	1,256043
2	-0,441257	-0,754324	0,388095	-0,255476
Respuesta de $X_2$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	0,000000	-0,022459	1,256923	-2,256123
2	0,815017	-0,905811	0,388095	-0,255476
Respuesta de $X_1$				
Período	$X_1$	$X_2$	$X_3$	$X_4$
1	1,457895	-2,145627	1,711415	3,564281
2	-0,441257	-0,754324	-0,524784	-3,968741

**Tabla 5.15:** Otra descomposición hipotética de choques

La interpretación de la tabla 5.15, se realiza de manera similar a la tabla 5.10.

El orden de descomposición se debería efectuar desde la variable que se supone es más exógena hasta la variable menos exógena. En caso de duda, es necesario realizar diferentes combinaciones del orden de descomposición y analizar la robustez de los resultados.

**Ejemplo 5.8.** A partir de la representación VAR(1) estimada en el ejemplo 5.5, calcular e interpretar las funciones de impulso-respuesta ortogonales, y la descomposición de la varianza.

#### Resolución.

Dado que las variables del ejemplo son variaciones del PIB, el CI y el DFI, es lógico pensar que un choque sobre la variable variación del PIB influya en la variación del CI y el DFI, más que si el choque fuera al revés. Esto sería: una innovación sobre  $Y_{1t}$  (variación del PIB) influya de manera instantánea a  $Y_{2t}$  (variación del CI) y a  $Y_{3t}$  (variación del DFI); por otro lado, una innovación sobre  $Y_{2t}$  o  $Y_{3t}$  no influya de manera contemporánea a  $Y_{1t}$ .

La matriz de varianza-covarianza estimada de los residuos es igual (ejemplo 5.5) a:

$$\hat{\Sigma}_u = \begin{bmatrix} 0,0002 & 0,0001 & 0,0002 \\ 0,0001 & 0,0001 & 0,0001 \\ 0,0002 & 0,0001 & 0,0002 \end{bmatrix}$$

Con el programa EViews, las salidas de las funciones de impulso-respuesta y la des-

composición de varianza serían:

Variance Decomposition of $Y_1$ :				
Período	S.E.	$Y_1$	$Y_2$	$Y_3$
1	0,015360	43,03994	21,09375	35,86631
2	0,016092	40,67934	25,73053	33,59013
3	0,016102	40,63720	25,71162	33,65118
4	0,016104	40,63653	25,70823	33,65524
5	0,016104	40,63654	25,70806	33,65540
6	0,016104	40,63655	25,70805	33,65540
7	0,016104	40,63655	25,70805	33,65540
8	0,016104	40,63655	25,70805	33,65540
9	0,016104	40,63655	25,70805	33,65540
10	0,016104	40,63655	25,70805	33,65540

Variance Decomposition of $Y_2$ :				
Período	S.E.	$Y_1$	$Y_2$	$Y_3$
1	0,007678	0,000000	100,0000	0,000000
2	0,007778	1,084622	98,63426	0,281119
3	0,007784	1,144767	98,54459	0,310640
4	0,007784	1,149508	98,53672	0,313774
5	0,007784	1,149959	98,53593	0,314109
6	0,007784	1,150005	98,53585	0,314145
7	0,007784	1,150010	98,53584	0,314149
8	0,007784	1,150010	98,53584	0,314149
9	0,007784	1,150011	98,53584	0,314149
10	0,007784	1,150011	98,53584	0,314149

Variance Decomposition of $Y_3$ :				
Período	S.E.	$Y_1$	$Y_2$	$Y_3$
1	0,015837	0,000000	34,49091	65,50909
2	0,016226	0,450218	33,21246	66,33733
3	0,016293	0,759358	33,02585	66,21479
4	0,016303	0,808238	33,01068	66,18109
5	0,016304	0,814151	33,00932	66,17653
6	0,016305	0,814812	33,00919	66,17600
7	0,016305	0,814883	33,00917	66,17594
8	0,016305	0,814891	33,00917	66,17594
9	0,016305	0,814892	33,00917	66,17594
10	0,016305	0,814892	33,00917	66,17594
Cholesky Ordering: $Y_2$ $Y_2$ $Y_1$				

Tabla 5.16: Descomposición de la Varianza (Orden de Cholesky  $Y_3$   $Y_2$   $Y_1$ )



La descomposición de la varianza indica que la varianza del error de predicción de  $Y_{1t}$  representa un 40,64 % con sus propias innovaciones, un 25,71 % con las de  $Y_{2t}$  y un 33,66 % con las de  $Y_{3t}$ . La varianza del error de predicción de  $Y_{2t}$  es de un 1,15 % con  $Y_{1t}$ , un 98,54 % con  $Y_{2t}$  y un 0,31 % con  $Y_{3t}$ . Finalmente, la varianza del error de predicción de  $Y_{3t}$  es de 0,81 % con  $Y_{1t}$ , un 33,01 % con  $Y_{2t}$  y un 66,18 % con sus propias innovaciones. Este efecto de asimetría se estudió en la parte de la causalidad (ejemplo 5.5); lo que también se corrobora ahora.

Por último, cabe señalar que la tabla anterior muestra la desviación estándar del error de previsión para  $Y_{1t}$ ,  $Y_{2t}$  y  $Y_{3t}$ , que se calculó de manera tediosa en el ejemplo 5.5.

**Observación 5.8.** La modelización VAR se realiza siempre sobre series estacionarias; sin embargo, el método se puede aplicar a series que se las vuelve estacionarias a través de diferenciación, con la idea de que se debe recuperar la serie original una vez calculadas las predicciones (esto se puede hacer dado que la diferenciación es una transformación lineal). En algunos paquetes econométricos como EViews o Stata, entre otros, el programa da la opción automática de recuperar la serie original de los datos.

## 5.10 Ejemplo Práctico

Se consideran dos series de datos del Ecuador: las variaciones del índice de precios al productor (IPP), denotada por ( $X_{1t}$ ) y del índice de actividad económica (IAE), denotada por ( $X_{2t}$ ). Se dispone de 132 datos mensuales desde enero de 2004 hasta junio de 2015 (Ver Anexo D.2). Para efectos de comparaciones se trabajará únicamente con los datos hasta diciembre de 2014 y se guardarán los del año 2015. Los datos se tomaron de la página oficial del INEC. Se desea estimar un modelo para realizar las predicciones de ambas series.

### Resolución.

#### 1. Matrices de correlación cruzada

##### a) Estadísticos descriptivos de $X_{1t}$ y $X_{2t}$

	Media	Mediana	Máximo	Mínimo	Desv. Est.	Asimetría	Curtosis
$X_{1t}$	0,64	1,16	15,17	-14,86	5,02	-0,40	3,74
$X_{2t}$	0,07	-0,30	15,18	-11,45	4,42	0,51	3,70

##### b) Matrices de correlación cruzada

	retardo 1		retardo 2		retardo 3		retardo 4		retardo 5	
$X_{1t}$	0,17	0,12	0,08	-0,02	0,01	0,08	-0,12	-0,03	-0,03	-0,09
$X_{2t}$	0,21	-0,20	0,05	0,08	0,14	0,02	0,13	-0,09	0,01	0,08

c) Representación simplificada

$$\begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline + & - \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array} \quad \begin{array}{|c|c|} \hline \cdot & \cdot \\ \hline \cdot & \cdot \\ \hline \end{array}$$

Tabla 5.17: Resumen de estadísticas y matrices de correlación cruzada para  $X_{1t}$  y  $X_{2t}$ .

Es fácil ver que las correlaciones cruzadas son significativas al 5 % en el retardo 1 ( $2/\sqrt{T} = 0,1747$  en este caso). Así,  $X_{2t}$  depende de los valores en el primer retardo de  $X_{1t}$  en el primer retardo y del suyo propio. En EViews, se presenta la siguiente salida:

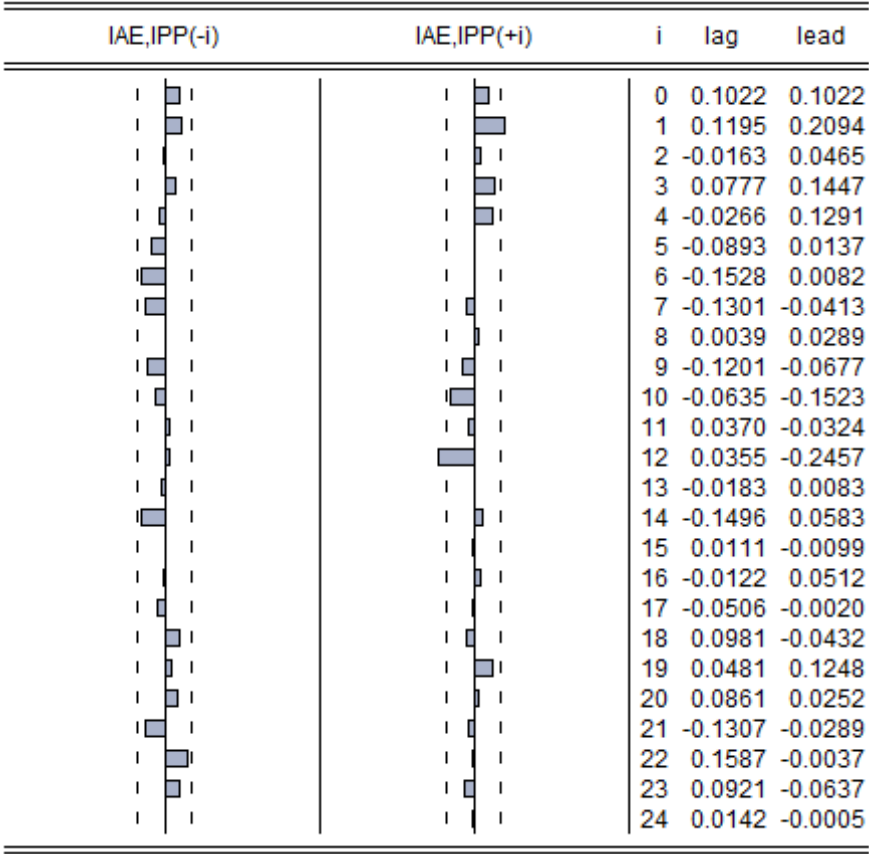


Figura 5.12: Correlaciones cruzadas entre  $X_{1t}$  y  $X_{2t}$

2. Estimación y validación del VAR

Se inicia presentando el gráfico de las series:

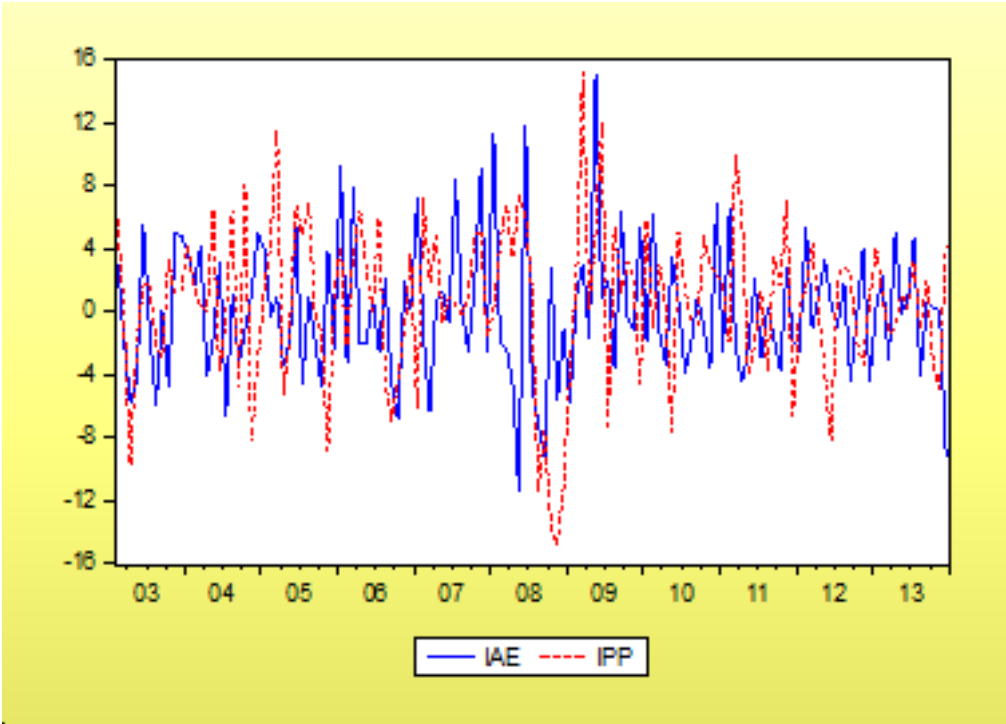


Figura 5.13: Gráfico de secuencia de las series

Antes de realizar los procedimientos para estimar los modelos, se debe verificar si las series a ser analizadas son estacionarias; para ello se realiza la prueba de raíces unitarias para cada serie utilizando el programa EViews:

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.001098	0.0000
Test critical values:	1 % level	-4.037668	
	5 % level	-3.448348	
	10 % level	-3.149326	

\*MacKinnon (1996) one-sided p-values.

Tabla 5.18: Prueba DFA para  $X_{1t}$

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-13.00154	0.0000
Test critical values:	1 % level	-4.037668	
	5 % level	-3.448348	
	10 % level	-3.149326	

\*MacKinnon (1996) one-sided p-values.

Tabla 5.19: Prueba DFA para  $X_{2t}$

Como se puede ver en las figuras 5.18 y 5.19, las dos series son estacionarias.

- a) Se utilizarán los criterios de Akaike, Schwarz y el logaritmo de máxima verosimilitud para determinar el retardo  $p$  entre 1 y 4. Se deben estimar cuatro modelos diferentes y retener aquel que satisfaga la mayor cantidad de criterios óptimos.

Se obtiene lo siguiente:

$$\hat{X}_{1t} = 0,1479X_{1t-1} + 0,2268X_{2t-1} + 0,4928$$

$$\hat{X}_{2t} = 0,1274X_{1t-1} - 0,2187X_{2t-1} - 0,0772$$

Con la ayuda del paquete EViews 7, se realiza la estimación de los 4 modelos. Así, se obtuvieron los siguientes resultados:

Criterio / Retardo	1	2	3	4
Log likelihood	-761,42	-756,05	-746,56	-736,01
Akaike	11,81	11,87	11,88	11,87
Schwarz	11,93	12,10	12,20	12,28

**Tabla 5.20:** Criterios para escoger el retardo del VAR

Como se puede observar en la tabla 5.18, es en el retardo 1 ( $p = 1$ ) donde los criterios de Akaike y Schwarz se minimizan aunque el valor del log de verosimilitud es mínimo (en la práctica, muy pocas veces se utiliza este criterio como decisivo para escoger el retardo del VAR). Por lo tanto se realiza la estimación del VAR(1).

- i. El modelo VAR estimado se escribe:

$$X_{1t} = 0,1479X_{1t-1} + 0,2268X_{2t-1} + 0,4928 + \hat{u}_{1t}$$

$$(1,71) \quad (2,28) \quad (1,14)$$

$R^2 = 0,07$ ;  $n = 130$ ;  $(.)$  = estadístico correspondiente a la distribución t de Student.

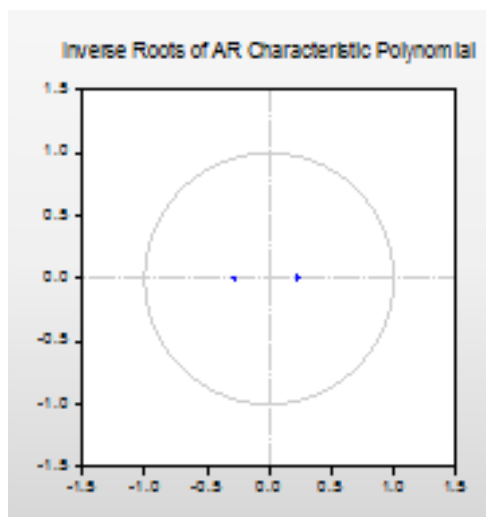
$$X_{2t} = 0,1274X_{1t-1} - 0,2187X_{2t-1} - 0,0772 + \hat{u}_{2t}$$

$$(1,67) \quad (-2,49) \quad (-0,20)$$

$R^2 = 0,06$ ;  $n = 130$ ;  $(.)$  = estadístico correspondiente a la distribución t de Student.

Antes de realizar las predicciones, se debe verificar si el modelo cumple con el criterio de estabilidad. Con ayuda del programa EViews se obtiene:

Raíces	Módulo
-0,285373	0,285373
0,214590	0,214590



**Figura 5.14:** Criterio de estabilidad para el VAR(1) estimado

Analítica y gráficamente, se concluye que las inversas de las raíces del polinomio característico se encuentran dentro del círculo unidad; por lo tanto, se concluye que el modelo es estable y, por tanto, es estacionario.

- ii. Ahora, se necesita verificar que los residuos del modelo sean ruidos blancos; en general, se prueba la independencia. Para ello, se utilizará el paquete EViews para obtener las pruebas sobre los residuos que se describieron anteriormente. Así se obtiene:

### Prueba Portmanteau

Lags	Q-Stat	Prob	Adj Q-Stat	Prob	df
1	0,050722	NA*	0,051115	NA*	NA*
2	1,429068	0,9641	1,450997	0,9627	6
3	6,825196	0,7418	6,974593	0,7278	10
4	14,04034	0,4467	14,41879	0,4190	14
5	15,45813	0,6303	15,89329	0,6000	18
6	24,47227	0,3230	25,34360	0,2809	22
7	26,60556	0,4302	27,59829	0,3785	26
8	34,51039	0,2609	36,02148	0,2074	30
9	37,10211	0,3279	38,80597	0,2619	34
10	51,10065	0,0760	53,97106	0,0447	38
11	53,74575	0,1057	56,86065	0,0627	42
12	68,78477	0,0164	73,42907	0,0062	46
13	72,41862	0,0207	77,46668	0,0076	50
14	76,83463	0,0223	82,41566	0,0077	54
15	77,52667	0,0443	83,19797	0,0167	58
16	80,19277	0,0599	86,23825	0,0226	62
17	82,43977	0,0832	88,82330	0,0321	66
18	86,47531	0,0883	93,50741	0,0318	70
19	90,20916	0,0968	97,88039	0,0330	74
20	93,29485	0,1141	101,5271	0,0380	78

**Tabla 5.21:** Prueba de autocorrelación Pormanteau

En la tabla 5.21 se observa que los p-valores (Prob.) para los retardos de 1 al 9 son no significativos; sin embargo, a partir del retardo 10 se vuelven significativos, esto sugiere que los residuos están autocorrelacionados.

## Prueba LM

Lags	LM-Stat	Prob
1	0,600602	0,9630
2	1,492304	0,8280
3	5,496276	0,2401
4	7,679129	0,1041
5	1,468612	0,8322
6	9,278191	0,0545
7	2,213578	0,6965
8	8,222604	0,0838
9	2,633746	0,6209
10	14,83753	0,0051
11	2,909367	0,5731
12	16,83962	0,0021
13	4,069454	0,3967
14	4,565627	0,3348
15	0,731412	0,9474
16	2,790990	0,5934
17	2,346395	0,6723
18	4,123520	0,3895
19	3,978227	0,4090
20	3,335910	0,5033

Tabla 5.22: Prueba LM

Por los valores en la columna Prob. (ver tabla 5.20), se puede concluir que existe autocorrelación entre los residuos (retardos 10 y 12). Esto confirma que se hay que reformular el modelo planteado.

### Reformulación del modelo

Se puede ver en las figuras 5.3 y 5.4 que a partir del retardo 10 existe autocorrelación de los residuos, por lo que se prueba un nuevo modelo VAR (10) para corregir este inconveniente.

Se estimó un VAR(10); sin embargo, al retardo 12 y 6 se tenía autocorrelación de residuos. Luego, se agregaron los retardos de orden 6 al VAR pero se encontró autocorrelación al retardo 12. Finalmente, se agregó el retardo 12 y se consiguió que los residuos no estén correlacionados, pero el problema era que no seguían una distribución normal multivariante.

Luego, se procedió a agregar un retardo de orden 14 y a quitar el retardo 12; con esto se consiguió que los residuos sean ruidos blancos, aunque presenta algo de correlación en el retardo 12, se decidió conservar este modelo ya que es el que cumple con más pruebas de independencia de residuos. Se debe mencionar que solamente se retienen los coeficientes significativos. Así, se obtuvo el siguiente modelo:

$$X_{1t} = 0,1416X_{1t-1} - 0,2403X_{1t-6} + 0,0889X_{1t-10} - 0,1047X_{1t-14} + 0,2241X_{2t-1} \\ + 0,0326X_{2t-6} - 0,1393X_{2t-10} + 0,1082X_{2t-14} + 0,7716 + \hat{u}_{1t}$$

$$X_{2t} = 0,0703X_{1t-1} - 0,1787X_{1t-6} - 0,1133X_{1t-10} - 0,1893X_{1t-14} - 0,2875X_{2t-1} \\ - 0,0605X_{2t-6} - 0,2330X_{2t-10} - 0,0422X_{2t-14} + 0,3525 + \hat{u}_{2t}$$

iii. La predicción calculada por el modelo, de manera recurrente es:

$$\hat{X}_{1t} = 0,1416X_{1t-1} - 0,2403X_{1t-6} + 0,0889X_{1t-10} - 0,1047X_{1t-14} + 0,2241X_{2t-1} \\ + 0,0326X_{2t-6} - 0,1393X_{2t-10} + 0,1082X_{2t-14} + 0,7716$$

$$\hat{X}_{1,15:1} = 0,1416 * 4,15 - 0,2403 * (3,26) + 0,0889 * (-1,34) - 0,1047 * (-3,37) \\ + 0,2241 * (-9,23) + 0,0326 * (4,71) - 0,1393 * (-3,12) + 0,1082 * (4,03) \\ + 0,7716$$

$$\hat{X}_{1,15:1} = -0,235$$

$$\hat{X}_{2t} = 0,0703X_{1t-1} - 0,1787X_{1t-6} - 0,1133X_{1t-10} - 0,1893X_{1t-14} - 0,2875X_{2t-1} \\ - 0,0605X_{2t-6} - 0,2330X_{2t-10} - 0,0422X_{2t-14} + 0,3525$$

$$\hat{X}_{2,15:1} = 0,0703 * 4,15 - 0,1787 * (-3,26) - 0,1133 * (-1,34) - 0,1893 * (-3,37) \\ - 0,2875 * (-9,23) - 0,0605 * (4,71) - 0,2330 * (-3,12) - 0,0422 * (4,03) \\ + 0,3525$$

$$\hat{X}_{2,15:1} = 3,777$$

donde,  $\hat{X}_{i,15:j}$  significa, la previsión de la variable  $X_i$  para el mes  $j$  del año 2015 (15:j).

De la misma manera se obtiene:

$$\hat{X}_{1,15:2} = 0,400$$

$$\hat{X}_{2,15:2} = -1,281$$

$$\hat{X}_{1,15:3} = -0,221$$

$$\hat{X}_{2,15:3} = -0,536$$

Para calcular la varianza del error de predicción, si fuera un VAR (1) se tiene:

$$M_1 = \hat{A}_1; \quad M_2 = \hat{A}_1 M_1 = \hat{A}_1^2; \quad \text{etc} \dots$$



Dado que el modelo es un VAR (14) con coeficientes 1, 6, 10 y 14, se tendrán las matrices  $\hat{A}_1$ ,  $\hat{A}_6$ ,  $\hat{A}_{10}$ ,  $\hat{A}_{14}$  y estarán compuestas por los coeficientes de los retardos de las variables analizadas; así, se obtiene:

$$\hat{A}_1 = \begin{bmatrix} 0,1416 & 0,2241 \\ 0,0703 & -0,2871 \end{bmatrix}; \quad \hat{A}_6 = \begin{bmatrix} -0,2403 & 0,0326 \\ -0,1787 & -0,0605 \end{bmatrix}$$

$$\hat{A}_{10} = \begin{bmatrix} 0,0889 & -0,1393 \\ -0,1133 & -0,2330 \end{bmatrix}; \quad \hat{A}_{14} = \begin{bmatrix} -0,1047 & 0,1082 \\ -0,1893 & -0,0422 \end{bmatrix}$$

La matriz de varianza covarianza estimada de la predicción, para el horizonte  $h = 1$ , es:

$$\hat{\Sigma}_T(1) = \hat{\Sigma}_u = \begin{bmatrix} 23,821 & 0,422 \\ 0,422 & 16,602 \end{bmatrix}$$

Así, la varianza del error de predicción para  $\hat{X}_{1,15:1}$  es igual a 23,821 y la varianza del error de predicción para  $\hat{X}_{2,15:1}$  es igual a 16,602.

Los intervalos de confianza para  $X_{1,15:1}$  y  $X_{2,15:1}$  vienen dados, respectivamente, por:

$$-0,235 \pm 1,96 * \sqrt{23,821} = [-9,80; 9,33]$$

$$3,777 \pm 1,96 * \sqrt{16,602} = [-4,21; 11,76]$$

Para los horizontes  $h = 2$ ,  $h = 3$ , se utilizan las siguientes fórmulas:

$$\begin{aligned} \hat{\Sigma}_T(2) &= \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \hat{A}_6 \hat{\Sigma}_u \hat{A}_6' + \hat{A}_{10} \hat{\Sigma}_u \hat{A}_{10}' + \hat{A}_{14} \hat{\Sigma}_u \hat{A}_{14}' \\ &= \begin{bmatrix} 27,78 & 1,51 \\ 0,93 & 20,44 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \hat{\Sigma}_T(3) &= \hat{\Sigma}_u + \hat{A}_1 \hat{\Sigma}_u \hat{A}_1' + \dots + \hat{A}_{12} \hat{\Sigma}_u \hat{A}_{12}' + \hat{A}_1^2 \hat{\Sigma}_u \hat{A}_1^{2'} + \dots + \hat{A}_{14}^2 \hat{\Sigma}_u \hat{A}_{14}^{2'} \\ &= \begin{bmatrix} 27,92 & 1,25 \\ 0,96 & 20,72 \end{bmatrix} \end{aligned}$$

Entonces los intervalos de confianza son:

$$IC(X_{1,15:2}) = -1,394 \pm 1,96 * 5,43 = [-9,93; 10,73]$$

$$IC(X_{2,15:2}) = 0,268 \pm 1,96 * 4,59 = [-10,14; 7,58]$$

$$IC(X_{1,15:3}) = -0,645 \pm 1,96 * 5,45 = [-10,58; 20,51]$$

$$IC(X_{2,15:3}) = -0,832 \pm 1,96 * 4,62 = [-9,46; 17,99]$$

### Comparación con modelos univariantes

La teoría VAR sugiere que las predicciones logradas son de mejor calidad que si se realiza la modelación de las series de manera univariante. Para comprobar esto, se realizó un modelo univariante para cada una de las series analizadas en este ejemplo. Así, se encontró que para el IPP el modelo univariante es  $X_{1t} = 0,18X_{1t-1} - 0,24X_{1t-6} + \hat{u}_t$ ; mientras que, para el IAE es  $X_{2t} = -0,25X_{2t-1} - 0,27X_{2t-10} + \hat{u}_t - 0,32\hat{u}_{t-8}$ . Con estos modelos se realizaron las predicciones para el año 2014 y se obtiene lo siguiente:

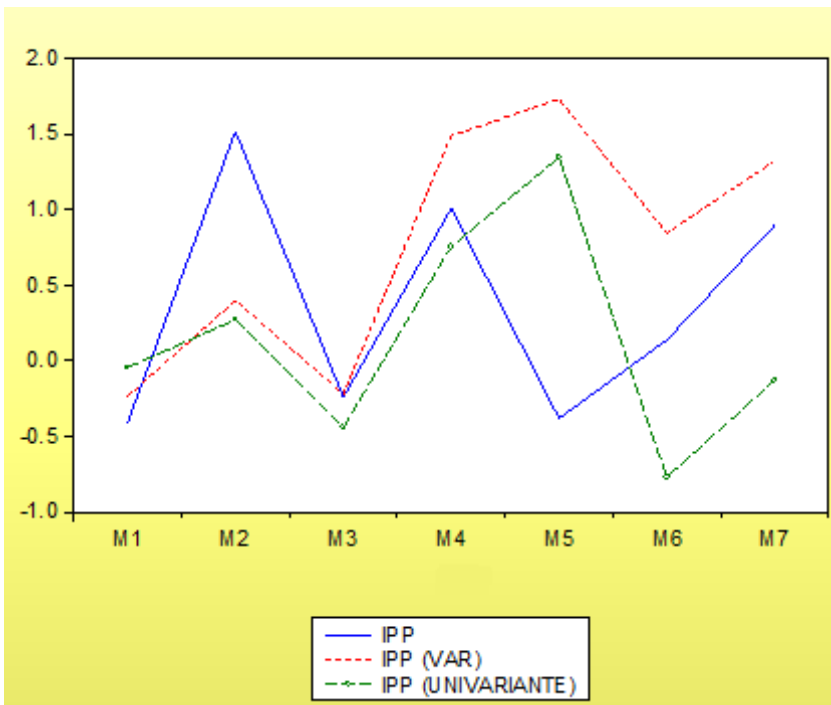
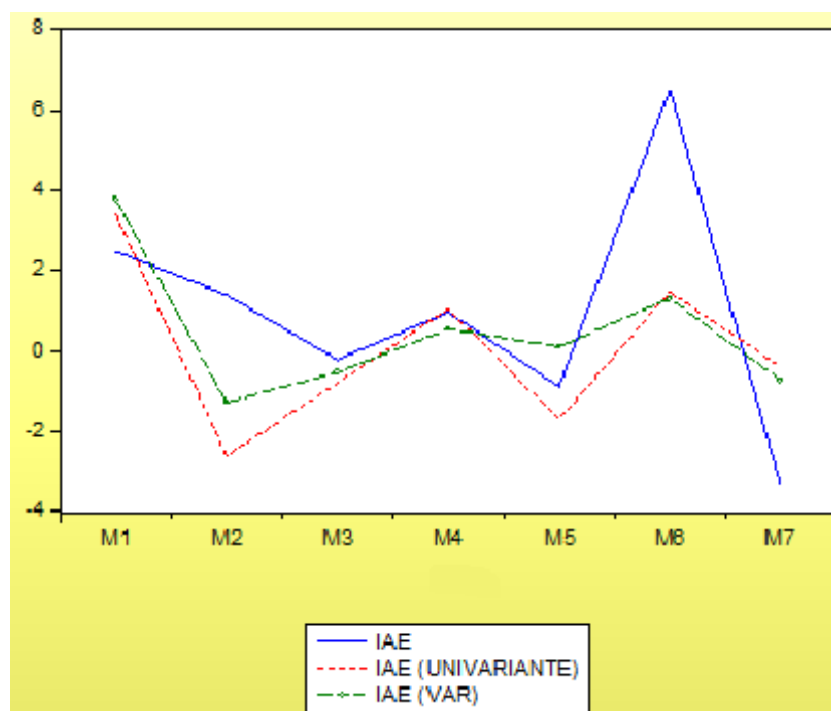


Figura 5.15: Comparación de las predicciones VAR y UNIVARIANTE para  $X_{1t}$  (IPP)



**Figura 5.16:** Comparación de las predicciones VAR y UNIVARIANTE para  $X_{2t}$  (IAE)

Se puede observar que el ajuste que tienen las predicciones del modelo VAR para el IPP es un tanto mejor que las del modelo univariante. Por otro lado, en el caso del IAE, las predicciones parecen bastante similares entre los dos modelos. Para poder determinar esto, se realiza el cálculo del error cuadrático medio para determinar el mejor ajuste. Así se tiene:

	IPP	IAE
VAR	0,95	6,29
UNIVARIANTE	0,95	7,41

**Tabla 5.23:** Error Medio Cuadrático estimado para los modelos VAR y UNIVARIANTE

Como se puede observar el modelo VAR es mejor en el caso del IAE; sin embargo, no lo es para el IPP. Se realizó una prueba t de medias para determinar si existe diferencia estadística entre las medias de los errores cuadráticos generados por los modelos y se comprobó que en el caso del IAE el error medio cuadrático es diferente estadísticamente entre los modelos (el modelo VAR ajusta mejor los datos); en el caso del IPP se determinó que no existe diferencia significativa entre los errores cuadráticos. Como conclusión, el modelo VAR predice de mejor manera que el modelo univariante.

### 3. La causalidad

El siguiente gráfico muestra la salida del paquete EViews para la Prueba de Granger.

Dependent variable: IPP			
Excluded	Chi-Square	df	Prob.
IAE	6.971411	4	0.1374
All	6.971411	4	0.1374

Dependent variable: IAE			
Excluded	Chi-Square	df	Prob.
IPP	11.44772	4	0.0220
All	11.44772	4	0.0220

**Tabla 5.24:** Prueba de causalidad de Granger

Como se puede ver, el p-valor (Prob.) es menor que 0,05 en el segundo caso y mayor que 0,05 en el primero; por lo que se concluye que  $X_{1t}$  (IPP) explica significativamente la variable  $X_{2t}$  (IAE), pero  $X_{2t}$  no explica significativamente la variable  $X_{1t}$ .

#### 4. Análisis de los “choques”

En las tablas siguientes se muestran los resultados de las funciones de impulso-respuesta que presenta el paquete EViews para las variables  $X_{1t}$  y  $X_{2t}$ ; también se muestra los gráficos correspondientes:

Período	IPP	IAE
1	1,000000	0,000000
2	0,144733	0,078893
3	0,042822	-0,010459
4	0,003298	0,006279
5	0,002218	-0,001481
6	-8,96E-05	0,000586
7	-0,198874	-0,156816
8	-0,100462	0,001894
9	-0,022618	-0,014730
10	-0,007966	0,001527

**Tabla 5.25:** Respuesta de las variables  $X_{1t}$  y  $X_{2t}$  ante un choque unitario de  $X_{1t}$

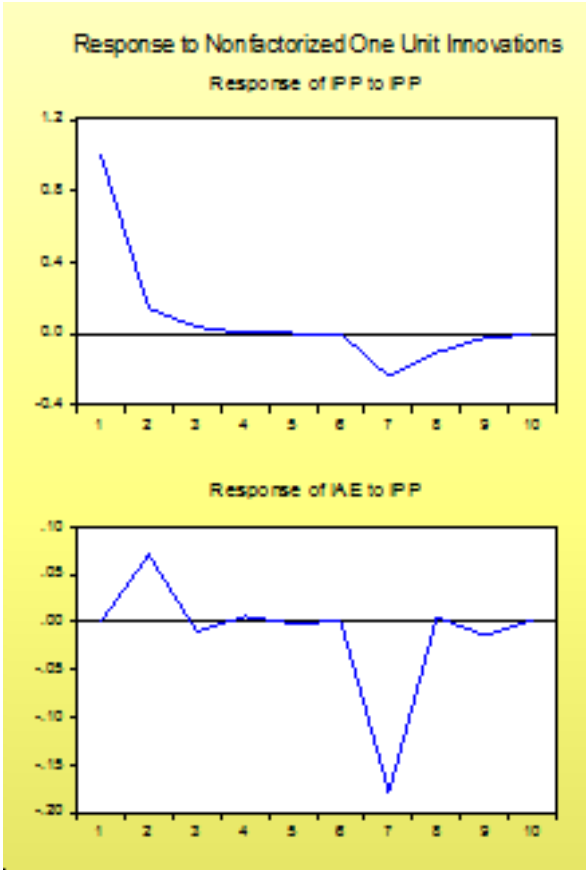


Figura 5.17: Respuesta de las variables  $X_{1t}$  y  $X_{2t}$  ante un choque unitario de  $X_{1t}$

Período	IPP	IAE
1	0.000000	1.000000
2	0.224126	-0.287467
3	-0.032697	0.098387
4	0.017422	-0.030581
5	-0.004387	0.010015
6	0.001624	-0.003187
7	0.032155	-0.059519
8	-0.072018	-0.003284
9	0.000135	-0.004230
10	-0.006113	-3.68E-05

Tabla 5.26: Respuesta de las variables  $X_{1t}$  y  $X_{2t}$  ante un choque unitario de  $X_{2t}$

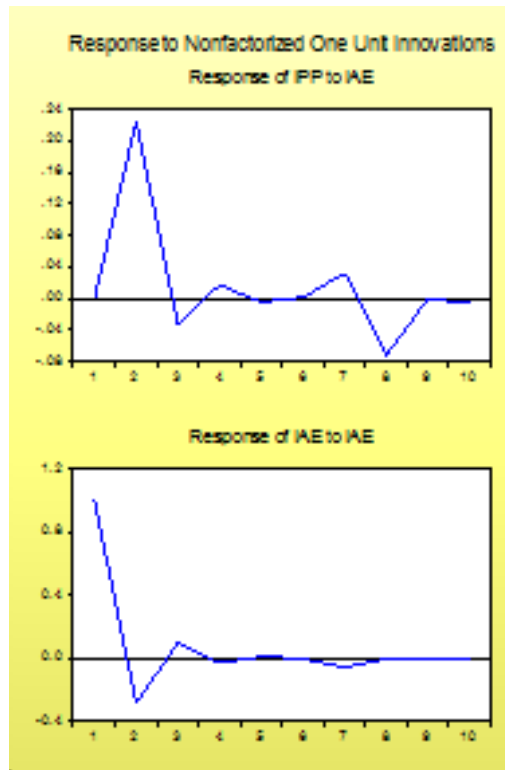


Figura 5.18: Respuesta de las variables  $X_{1t}$  y  $X_{2t}$  ante un choque unitario de  $X_{2t}$

La elección de la dirección del impacto es muy importante y determina los valores obtenidos. Se puede observar que el efecto de la innovación se desvanece con el tiempo; esto caracteriza a un proceso VAR estacionario.

### 5. Descomposición de la varianza

A partir de la representación  $VAR(14)$  estimada se calcula la descomposición de la varianza.

Dado que las variables del ejemplo son variaciones del IPP y el IAE, es lógico pensar que un choque sobre la variable variación del IPP inflencie la variación del IAE más que si el choque fuera al revés. Esto sería: una innovación sobre  $X_{1t}$  (variación del IPP) influencia de manera instantánea a  $X_{2t}$  (variación del IAE); por otro lado, una innovación sobre  $X_{2t}$  no inflencia de manera contemporánea a  $X_{1t}$ .

La matriz de varianza-covarianza estimada de los residuos es igual (ejemplo 5.5) a:

$$\hat{\Sigma}_u = \begin{bmatrix} 23,82 & 0,42 \\ 0,42 & 16,60 \end{bmatrix}$$

Con el programa EViews, las salidas de las funciones de impulso-respuesta y la descomposición de varianza serían:

Descomposición de la varianza de X1:				Descomposición de la varianza de X2:			
Periodo	S.E.	X1	X2	Periodo	S.E.	X1	X2
1	4,880657	100,0000	0,000000	1	4,074543	0,045030	99,95497
2	5,015877	96,68675	3,313246	2	4,251402	0,601299	99,39870
3	5,020588	96,62258	3,377415	3	4,270454	0,605403	99,39460
4	5,021112	96,60331	3,396688	4	4,272338	0,608029	99,39197
5	5,021150	96,60210	3,397905	5	4,272537	0,608160	99,39184
6	5,021154	96,60193	3,398072	6	4,272558	0,608181	99,39182
7	5,157177	96,71430	3,285697	7	4,368618	4,623096	95,37690
8	5,191890	96,43880	3,561199	8	4,368704	4,625945	95,37406
9	5,193082	96,44043	3,559566	9	4,369329	4,651664	95,34834
10	5,193258	96,43838	3,561625	10	4,369338	4,652042	95,34796
Orden de Cholesky: X2 X1							

**Tabla 5.27:** Descomposición de la Varianza (Orden de Cholesky X1 X2)

La descomposición de la varianza indica que la varianza del error de predicción de  $X_{1t}$  representa un 96,63 % con sus propias innovaciones y un 3,56 % con las de  $X_{2t}$ . La varianza del error de predicción de  $X_{2t}$  es de un 4,65 % con  $X_{1t}$  y un 95,34 % con  $X_{2t}$ . Este efecto de asimetría se estudió en la parte de la causalidad; lo que también se corrobora ahora.

Por último, cabe señalar que la tabla anterior muestra la desviación estándar del error de previsión para  $X_{1t}$  y  $X_{2t}$ .





## Capítulo 6

# La Cointegración

El análisis de cointegración fue tratado por Granger (1983) y por Engle y Granger (1987); este concepto se considera uno de los más importantes dentro del análisis de series temporales.

La cointegración aparece cuando dos o más series presentan una relación de movimiento conjunto (tendencia) a largo plazo y las diferencias entre ellas son estables.

### 6.1 Propiedades del orden de Integración de una serie

Sean dos series afectadas de una tendencia (ver los gráficos 6.1 y 6.2); intuitivamente se puede decir lo siguiente:

- En el primer caso, las dos series tienen una tendencia de evolución similar en un primer lapso y después una tendencia de evolución divergente en un segundo período; entonces las series no están cointegradas.
- En el segundo caso, las dos series tienen una evolución similar en todo el período de análisis; las series están cointegradas, si existe una evolución a largo plazo similar entre las series.

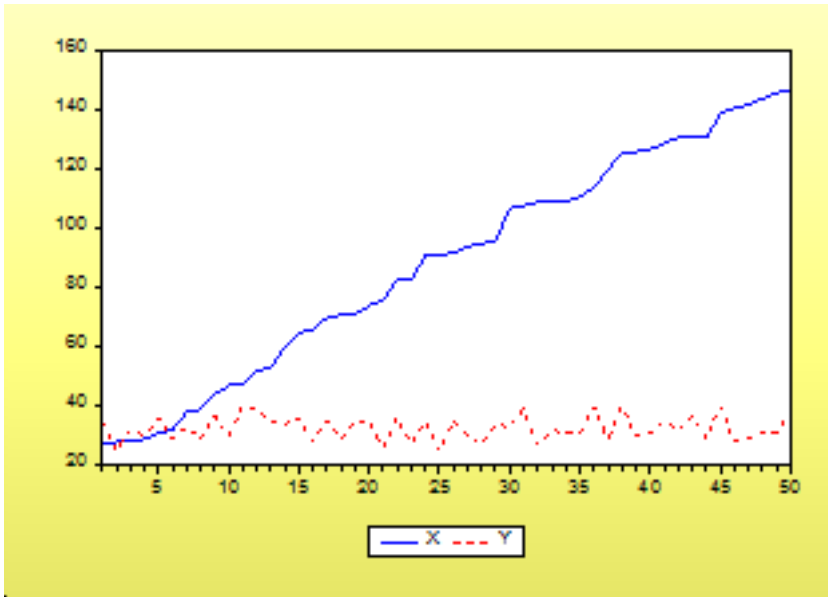


Figura 6.1: Las variables  $X_{1t}$  y  $X_{2t}$  no están cointegradas

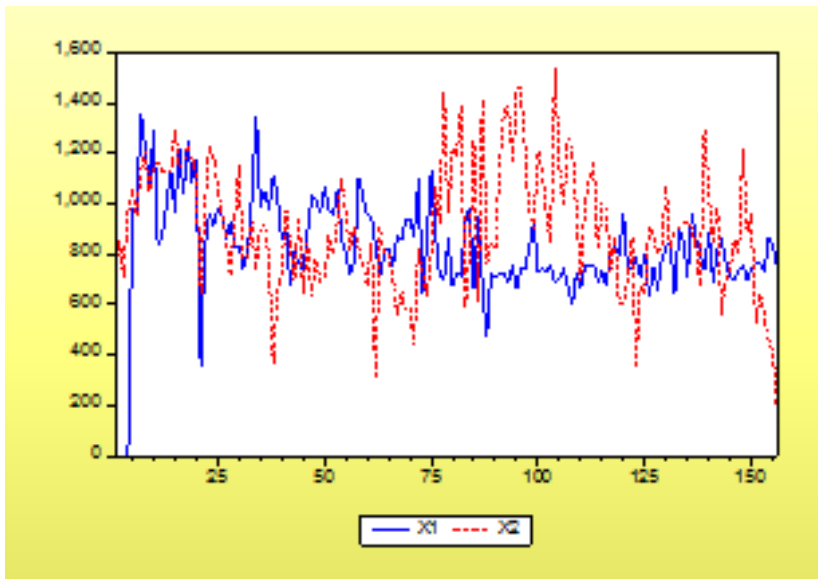


Figura 6.2: Las variables  $X_{1t}$  y  $X_{2t}$  están cointegradas

Recuérdese que una serie es integrada de orden  $d$ , si hay que diferenciarla  $d$  veces para volverla estacionaria.

Sea una serie  $X_{1t}$  estacionaria y una serie  $X_{2t}$  integrada de orden 1; entonces se obtiene:

$$\begin{aligned} X_{1t} &\rightarrow I(0) \\ X_{2t} &\rightarrow I(1) \end{aligned} \Rightarrow X_{1t} + X_{2t} \rightarrow I(1)$$

La serie  $X_t = X_{1t} + X_{2t}$  no es estacionaria ya que es la suma de una serie afectada de una tendencia y una serie estacionaria.

En general:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d') \end{matrix}, d \neq d' \Rightarrow X_{1t} + X_{2t} \rightarrow I(?)$$

Es imposible llegar a una conclusión con respecto a la suma de dos series de orden de integración diferente.

Incluso si dos series  $X_{1t}$  y  $X_{2t}$  son integradas de orden  $d$ , en general se obtiene:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \end{matrix} \Rightarrow X_{1t} + X_{2t} \rightarrow I(?)$$

También, para la combinación lineal  $\alpha X_{1t} + \beta X_{2t} \rightarrow I(?)$ .

En efecto, el resultado depende de los signos de los coeficientes  $\alpha$ ,  $\beta$  y de la existencia de una dinámica no estacionaria común.

## 6.2 Condiciones de cointegración

Dos series  $X_{1t}$  y  $X_{2t}$  son cointegradas si se satisfacen las siguientes condiciones:

- Ambas están afectadas por una tendencia del mismo orden de integración  $d$ .
- Una combinación lineal de estas series permite reducir a una serie de orden de integración inferior.

Es decir, si:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \end{matrix}$$

Tal que,  $\alpha X_{1t} + \beta X_{2t} \rightarrow I(d - b)$ , donde  $d \geq b > 0$ .

Se denota:  $X_{1t}, X_{2t} \rightarrow CI(d, b)$ , con  $[\alpha \quad \beta]'$  como el vector de cointegración.

En el caso general ( $k$  variables), se tiene:

$$\begin{matrix} X_{1t} \rightarrow I(d) \\ X_{2t} \rightarrow I(d) \\ \dots \\ X_{kt} \rightarrow I(d) \end{matrix}$$

denótese por  $X'_t = [X_{1t} \quad X_{2t} \dots X_{kt}]$ .

Si existe un vector de cointegración  $\beta' = [\beta_1 \ \beta_2 \dots \beta_k]$ , tal que  $\beta' X_t \rightarrow I(d - b)$ , entonces las  $k$  variables están cointegradas y el vector de cointegración es  $\beta$ . Se denota  $X_t \rightarrow CI(d - b)$  con  $b > 0$ .

### Observación 6.1.

1. No se diferencian las series individualmente ya que, en general, esto produciría que se sobrediferencie el sistema. Es decir, cuando se encuentra un vector de cointegración ocurre que se elimina la tendencia común que existe entre las series (lo que correspondería a una diferenciación); si se diferencian individualmente se estaría realizando diferenciaciones adicionales a las que se logra con el vector de cointegración.
2. En la práctica, el caso más considerado es cuando  $d = b = 1$ , que es lo que se asume en este documento.

**Ejemplo 6.1.** Se consideran tres series económicas de un país sudamericano: Producto Interno Bruto, Consumo Interno y la Demanda Final Interna denotadas por  $X_{1t}$ ,  $X_{2t}$ ,  $X_{3t}$ , respectivamente. Los datos son trimestrales en un período que va desde noviembre 2010 hasta junio de 2015 (56 observaciones), los que se utilizarán para analizar una posible cointegración entre las variables; para efectos de ejemplo, se trabajará únicamente con 50 datos y se dejarán 6 (enero a junio de 2015) para poder realizar las comparaciones con las predicciones (ver anexo D.1). Se pide analizar una eventual cointegración entre las variables.

### Resolución.

Para iniciar, se presentan los gráficos de las series temporales a analizarse:

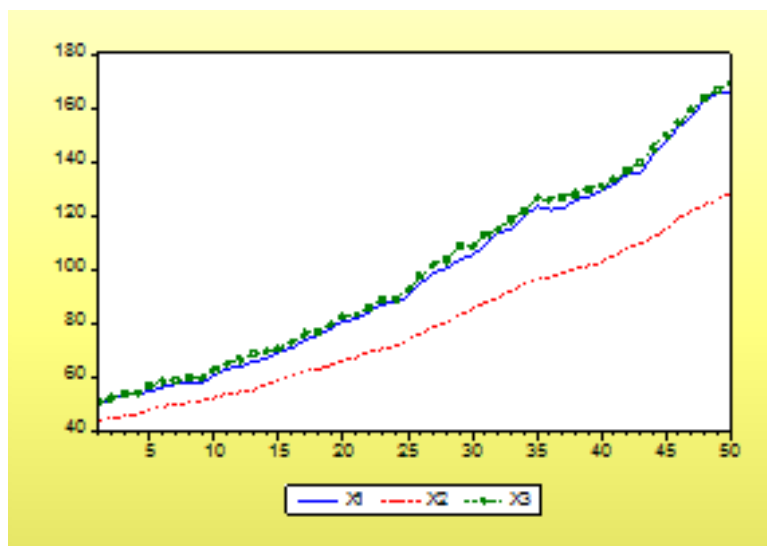


Figura 6.3: Series consideradas en el ejemplo

Ahora, para determinar si las series son estacionarias es necesario verificar si existe una raíz unitaria en las series. Se realiza la prueba DFA a nivel, para las tres series:

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.317447	0.8718
Test critical values:		
1 % level	-4.156734	
5 % level	-3.504330	
10 % level	-3.181826	

**Tabla 6.1:** Prueba de raíces unitarias para las  $X_{1t}$ 

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-0.813922	0.9572
Test critical values:		
1 % level	-4.156734	
5 % level	-3.504330	
10 % level	-3.181826	

**Tabla 6.2:** Prueba de raíces unitarias para las  $X_{2t}$ 

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-1.054405	0.9263
Test critical values:		
1 % level	-4.156734	
5 % level	-3.504330	
10 % level	-3.181826	

**Tabla 6.3:** Prueba de raíces unitarias para las  $X_{3t}$ 

Se concluye que las tres series tienen raíz unitaria. A continuación, se presentan las pruebas de raíces unitarias de las series en primeras diferencias:

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-6.222194	0.0000
Test critical values:		
1 % level	-4.161144	
5 % level	-3.506374	
10 % level	-3.183002	

**Tabla 6.4:** Prueba de raíces unitarias para  $X_{1t}$  en primera diferencia

	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-5.620775	0.0001
Test critical values:		
1 % level	-4.161144	
5 % level	-3.506374	
10 % level	-3.183002	

**Tabla 6.5:** Prueba de raíces unitarias para  $X_{2t}$  en primera diferencia

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-7.038579	0.0000
Test critical values:	1 % level	-4.161144	
	5 % level	-3.506374	
	10 % level	-3.183002	

**Tabla 6.6:** Prueba de raíces unitarias para  $X_{3t}$  en primera diferencia

De las pruebas se concluye que se necesita realizar solamente una diferenciación no estacional para volverlas estacionarias; por lo tanto, las series son integradas de orden 1,  $I(1)$ .

## Capítulo 7

# Modelo de Corrección del error (MCE)

El modelo de corrección del error (MCE) recibe su nombre debido a que mediante este, se corrigen los errores o desviaciones que las variables presentan en el corto plazo con respecto a su relación de equilibrio a largo plazo.

Del párrafo anterior se puede determinar que el MCE está relacionado con la cointegración de las variables; si las variables no están cointegradas, no se puede plantear el MCE.

Al utilizar este método, se considera el comportamiento dinámico de las series; además, presenta ventajas con respecto a los métodos en los que se utiliza la diferenciación de las series para convertirlas en estacionarias, pues pueden dar lugar a casos de sobrediferenciación y pérdida de información.

### 7.1 Planteamiento del MCE

Se considera el modelo VAR(p) de  $k$  variables, donde por simplicidad se omiten las posibles componentes deterministas:

$$X_t = A_1 X_{t-1} + A_2 X_{t-2} + \dots + A_p X_{t-p} + u_t$$

con  $u_t$  ruido blanco.

Considérese,

$$A(B) = I - A_1 B - \dots - A_p B^p$$

Recuérdese que si las raíces del polinomio característico están fuera del círculo unidad,  $X_t$  es estacionario; es decir, es  $I(0)$ . Si  $|A(1)| = 0$ , se dice que  $X_t$  tiene una raíz

unitaria. Por simplicidad de la exposición, se asume que  $X_t$  es integrada de orden 1 ( $I(1)$ ); esto significa que:

$$(1 - B)X_t$$

es estacionaria, si  $X_t$  no lo es.

Un modelo de corrección del error (MCE) para un proceso  $VAR(p)$  es:

$$\Delta X_t = \Omega X_{t-1} + A_1^* \Delta X_{t-1} + \dots + A_{p-1}^* \Delta X_{t-p+1} + u_t$$

donde,

$$A_j^* = - \sum_{i=j+1}^p A_i, \quad j = 1, \dots, p-1$$

$$\Omega = \alpha \beta' = A_p + A_{p-1} + \dots + A_1 - I = -A(1)$$

$\alpha$  y  $\beta$  son matrices de dimensión  $k \times r$  ( $r$  es el número de factores de cointegración y  $r < k$ ), de rango completo.

Al término  $\Omega X_{t-1}$  se lo conoce como el **término de corrección del error**, que desempeña un papel clave en el estudio de cointegración. La existencia del término  $\beta' X_{t-1}$  es natural en la representación de corrección del error; además, es estacionario. Las columnas de  $\beta$  son los vectores de cointegración de  $X_t$ .

La estacionariedad de  $\beta' X_{t-1}$  puede ser justificada de la siguiente manera: La teoría de series con raíces unitarias muestra que el coeficiente de correlación simple entre una serie no estacionaria con raíz unitaria y una estacionaria tiende a 0 cuando el tamaño de la muestra tiende a infinito ( $T \rightarrow \infty$ ). En un MCE  $X_{t-1}$  es no estacionario, pero  $\Delta X_t$  es estacionario; por lo tanto, solo existe una forma en que  $\Delta X_t$  puede relacionarse significativamente con  $X_{t-1}$ : a través de la serie estacionaria  $\beta' X_{t-1}$ .

Para poder recuperar  $A_i$  a partir de MCE se tiene:

$$A_1 = I + \Omega + A_j^*$$

$$A_i = A_i^* - A_{i-1}^*, \quad i = 2, \dots, p$$

donde  $A_p^* = 0$ , la matriz cero. En lo que sigue se supone que  $X_t$  es al menos  $I(1)$ . Por lo tanto, se consideran tres casos en el MCE:

- Si  $rg(\Omega) = 0$ , implica que  $\Omega = 0$  y las  $X_t$  son no cointegradas. Con esto, se estima un modelo VAR clásico en primeras diferencias a fin de eliminar la tendencia.
- Si  $rg(\Omega) = k$ , implica que  $X_t$  no contiene raíces unitarias; esto es,  $X_t$  es  $I(0)$ . El modelo de corrección del error no es informativo y se debe estudiar directamente  $X_t$ .
- Si  $0 < rg(\Omega) = r < k$ , existen  $r$  vectores (fila) linealmente independientes que recogen  $r$  relaciones de cointegración diferentes; entonces, se puede descomponer



la matriz  $\Omega$  en el producto de dos submatrices  $\alpha$  y  $\beta$ , de orden  $(k \times r)$  de forma tal que:

$$\Omega = \alpha\beta'$$

donde  $\beta'$  contiene los coeficientes de los vectores de cointegración y  $\alpha$  los parámetros de la velocidad de ajuste. Sustituyendo la matriz  $\Omega$  en la expresión general del modelo, se tiene:

$$\Delta X_t = \Omega X_{t-1} + A_1^* \Delta X_{t-1} + \dots + A_{p-1}^* \Delta X_{t-p+1} + u_t$$

donde, el producto  $\beta' X_{t-1}$  es estacionario.

## 7.2 Prueba de Johansen

Para determinar el número de relaciones de cointegración Johansen (1988) propuso dos pruebas fundamentadas en los valores propios de la matriz  $\Omega$ .

En la práctica, el procedimiento de aplicación de la prueba de Johansen se realiza mediante las siguientes etapas:

1. Determinación del orden óptimo del modelo VAR con las series que son integradas del mismo orden. Esta etapa es muy importante, ya que los resultados del contraste son muy sensibles frente a una mala especificación de retardos. Posteriormente, se realizará un ajuste con respecto a la tendencia común.  
En algunos paquetes estadísticos como el EViews, se escoge el retardo del VAR pero con las variables diferenciadas una vez. Es decir, el paquete realiza una diferenciación en el VAR de manera automática.
2. Estimación del modelo. Para realizar la descomposición de la matriz  $\Omega$  en las submatrices  $\alpha$  y  $\beta$ , se utiliza el método de máxima verosimilitud propuesto por Johansen (1988).
3. Determinación del rango de cointegración, que vendrá dado por el propio rango de la matriz  $\Omega$ . Teniendo en cuenta que se trata de una matriz de coeficientes estimados, y por lo tanto aleatorios, el rango de dicha matriz no puede determinarse de forma absoluta, sino que habrá que definirlo dentro de un entorno probabilístico.

Se definen dos estadísticos alternativos, basados en los valores propios  $\lambda_i$  de la matriz  $\hat{\Omega}$  y formulados como:

a. **Traza**

$$\lambda_{\text{traza}}(r) = -T \sum_{i=r+1}^k \ln(1 - \lambda_i)$$

donde:

$\lambda_i$  = es el  $i$ -ésimo valor propio de la matriz  $\hat{\Omega}$  (en orden creciente),  
 $r$  = rango de la matriz  $\hat{\Omega}$ ,  
 $k$  = número de variables,  
 $T$  = número de observaciones.

Con este estadístico se contrasta la hipótesis nula de existencia de un número de vectores de cointegración menor o igual a  $r$  frente a la hipótesis alternativa de existencia de más de  $r$  relaciones de cointegración.

#### b. Máximo valor propio

$$\lambda_{\max}(r, r+1) = -T \ln(1 - \lambda_{r+1})$$

donde:

$\lambda_i$  = es el  $i$ -ésimo valor propio de la matriz  $\hat{\Omega}$  (en orden creciente),  
 $r$  = rango de la matriz  $\hat{\Omega}$ ,  
 $k$  = número de variables,  
 $T$  = número de observaciones.

Este estadístico contrasta la hipótesis nula de existencia de  $r$  vectores de cointegración frente a la hipótesis alternativa de existencia de  $r+1$  vectores de cointegración.

Estos estadísticos tienen una ley de probabilidad (similar a una  $\chi^2$ ) tabulada mediante simulaciones por Johansen y Juselius (1990).

**Ejemplo 7.1.** A partir del ejemplo 6.1, se pide analizar una eventual cointegración entre las variables y estimar un modelo VAR o un modelo VEC si fuese el caso.

#### Resolución

- a. **Primera etapa:** Determinar el número de retardos de la representación VAR.

Para iniciar, se presentan los gráficos de las series temporales a analizarse:

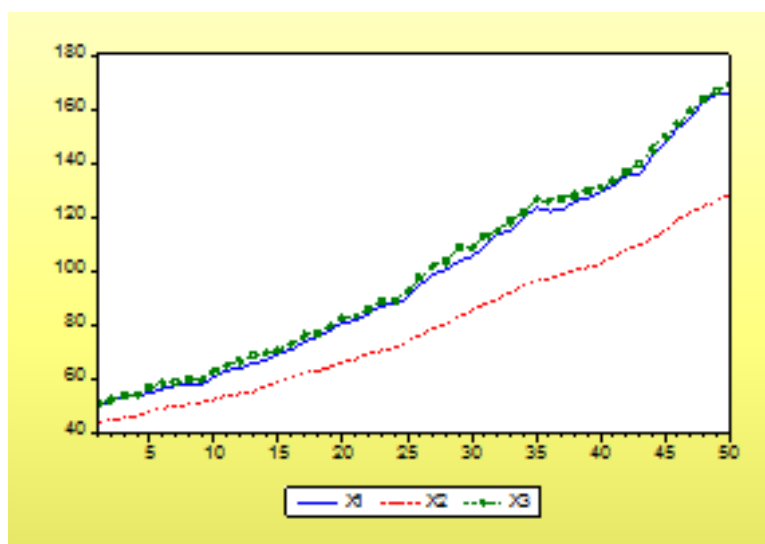


Figura 7.1: Series consideradas en el ejemplo

Como se vio en el ejemplo 6.1, las series son cointegradas de orden 1. Ahora, se determina el mejor retardo para el modelo VAR (considerando diferentes criterios):

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-170.1455	NA	1.441464	8.879257	9.007223*	8.925170*
1	-158.4805	20.93714	1.260025*	8.742591	9.254456	8.926244
2	-150.7736	12.64724	1.359430	8.808904	9.704668	9.130296
3	-146.6524	6.129033	1.787069	9.059096	10.33876	9.518228
4	-143.2875	4.486496	2.493999	9.348077	11.01164	9.944949
5	-124.5157	22.14106*	1.627928	8.846961	10.89442	9.581572
6	-119.0699	5.585502	2.196024	9.029224	11.46058	9.901575
7	-116.5909	2.161115	3.663332	9.363638	12.17890	10.37373
8	-98.07554	13.29311	2.931050	8.875669	12.07483	10.02350
9	-88.11999	5.615947	4.160914	8.826666	12.40972	10.11224
10	-76.43760	4.792776	6.776135	8.689108*	12.65606	10.11242

Tabla 7.1: Elección del retardo del VAR

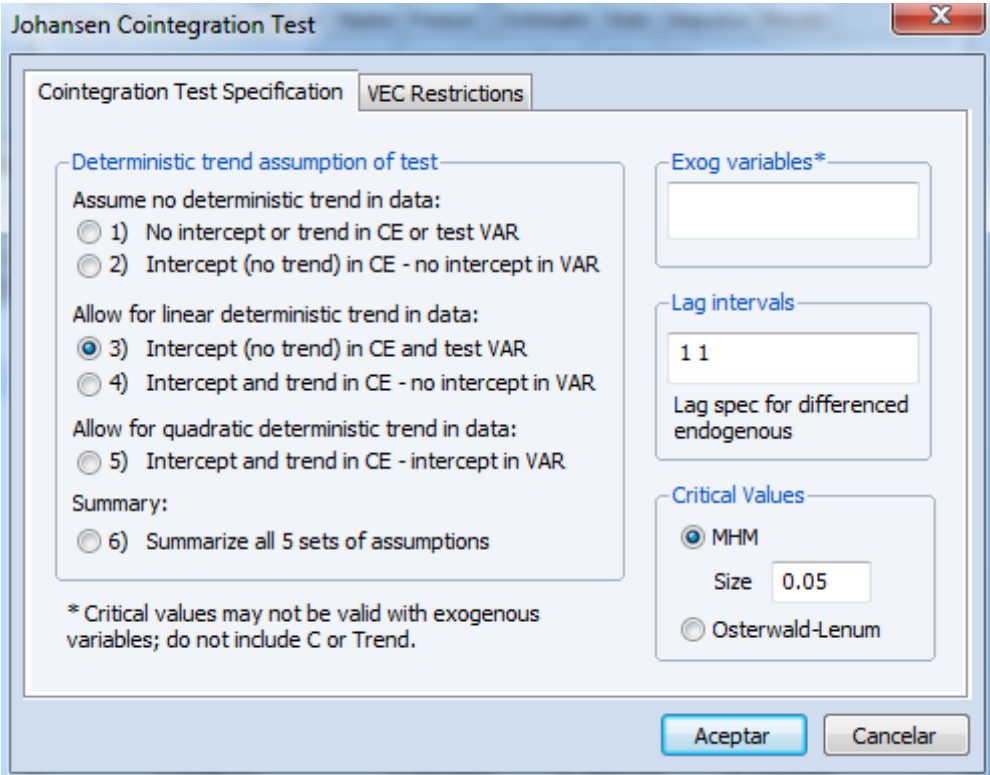
Como se puede observar, los criterios difieren de cuál es el mejor retardo del VAR a estimar (cabe considerar que en este caso se consideran las series en primeras diferencias). Los valores dados en la tabla se pueden considerar como cotas superiores del retardo del VAR. Debido a que se tienen pocos datos se desechará el retardo 10, por lo que, se puede estimar un VAR(5).

Esto da lugar a 2 posibles modelos:

1. VAR(5), considerando solamente el retardo 5.
2. VAR(5), considerando los retardos del 1 al 5.

b. **Segunda Etapa:** Prueba de Johansen

Como ya se mencionó, las series son  $I(1)$ . Se realizan las pruebas sobre los supuestos de la tendencia de los datos (éstas están implementadas en EViews):



**Figura 7.2:** Prueba de cointegración de Johansen

Nótese, que esta prueba es válida solo para series que no son estacionarias. Además, los retardos se fijan para la series en primeras diferencias (ver figura 9.3, *lag intervals*) y **no en niveles**. Por ejemplo, si se escribe “1 2” la prueba hace la regresión de  $\Delta X_t$  sobre  $\Delta X_{t-1}$ ,  $\Delta X_{t-2}$  y sobre las variables exógenas especificadas, de ser el caso. Para  $X_t$  tendría 3 retardos. Si se desea correr la prueba con un retardo en niveles, se debe escribir “0 0”.

Resumen de los 5 conjuntos de supuestos		
	CE	VAR
No tendencia determinística en los datos		
1	Ni intercepto ni Tendencia	Ni intercepto ni Tendencia
2	Intercepto-no Tendencia	No intercepto
Tendencia determinística lineal en los datos		
3	Intercepto-no Tendencia	Intercepto-no Tendencia
4	Intercepto y Tendencia	No Tendencia
Tendencia determinística cuadrática en los datos		
5	Intercepto y Tendencia	Tendencia lineal

La prueba de Johansen propone 5 modelos del VEC que se describen de la siguiente manera:

- Es el modelo más restrictivo: No admite ni constante ni tendencia. Es decir, los datos de nivel  $X_t$  no tienen tendencia determinística y las ecuaciones de cointegración no tienen interceptos.
- Incluye una constante en el vector de cointegración; pero, los datos de nivel no tienen tendencia determinística.
- Incluye una constante en el vector de cointegración y además una tendencia lineal en las componentes de  $X_t$ .
- Incluye una constante en el vector de cointegración, una tendencia lineal en las componentes de  $X_t$ , y además, una tendencia lineal en el vector de cointegración.
- Finalmente el último modelo, incluye una constante en el vector de cointegración, una tendencia lineal en las componentes de  $X_t$ , una tendencia lineal en el vector de cointegración y una tendencia cuadrática en las variables en nivel. Este es el modelo menos restrictivo.

En la figura siguiente se muestran los supuestos y las relaciones de cointegración, para las pruebas de la traza y del máximo valor absoluto para los dos modelos descritos anteriormente:

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	2	3	3	2	0
Max-Eig	2	1	0	0	0

**Tabla 7.2:** Resumen de la prueba de cointegración de Johansen para VAR(5), con solo el retardo 5

Data Trend:	None	None	Linear	Linear	Quadratic
Test Type	No Intercept No Trend	Intercept No Trend	Intercept No Trend	Intercept Trend	Intercept Trend
Trace	0	0	0	0	0
Max-Eig	0	0	0	0	0

**Tabla 7.3:** Resumen de la prueba de cointegración de Johansen para VAR(5), con los retardos desde el 1 al 5

Como se puede observar, dependiendo del número de retardos y la forma de ingresarlos se tienen diferentes relaciones de cointegración. En el modelo VAR(5) con los retardos del 1 al 5 no se tienen relaciones de cointegración. Ahora, se tiene un solo modelo posible:

- Modelo VAR(5) solo con el retardo 5, con dos relación de cointegración sin intercepto ni tendencia.

**Primera prueba:** El rango de la matriz  $\Omega$  igual a 0. Sea  $H_0 : r = 0$  contra  $H_1 : r > 0$ .

Los valores propios de la matriz  $\Omega$ , estimados por máxima verosimilitud, son iguales a:  $\lambda_3 = 0,4451$ ,  $\lambda_2 = 0,2734$  y  $\lambda_1 = 0,0257$ .

Se calcula el estadístico de Johansen:

$$\lambda_{traza} = -T \sum_{i=r+1}^k \ln(1 - \lambda_i); \quad \text{para } r = 0$$

$$\begin{aligned} \lambda_{traza} &= -T * \{ \ln(1 - \lambda_1) + \ln(1 - \lambda_2) \} \\ &= -50 * \{ \ln(1 - 0,4451) + \ln(1 - 0,2734) + \ln(1 - 0,0257) \} \\ &= 41,11 \end{aligned}$$

También, se calcula el del máximo valor propio:

$$\begin{aligned} \lambda_{max}(r, r+1) &= -T \ln(1 - \lambda_{r+1}) \quad \text{para } r = 0 \\ \lambda_{max}(0, 1) &= -T \ln(1 - \lambda_2) = -50 * \ln(1 - 0,2734) = 25,91 \end{aligned}$$

El valor crítico para la traza es igual a 24,27 para un nivel de significación del 5 %. Por lo tanto para el caso se la traza, se rechaza la hipótesis nula; el rango de la matriz no es 0 (las series no son estacionarias). Mientras que, la prueba del máximo valor propio tiene un valor crítico de 25,91, se rechaza la hipótesis nula; el rango de la matriz no es 0.

Se acepta que la hipótesis de que existe al menos una relación de cointegración.

**Segunda prueba:** El rango de la matriz  $\Omega$  igual a 1. Sea  $H_0 : r = 1$  contra  $H_1 : r > 1$ .

$$\begin{aligned} \lambda_{traza} &= -15,19 \\ \lambda_{max}(1, 2) &= 14,05 \end{aligned}$$

El valor crítico es igual a 12,32 y de 11,22 para la prueba de la traza y del máximo valor propio, respectivamente, con un nivel de significación del 5 %. Por lo tanto, se rechazar  $H_0$ ; se considera que el rango de la matriz no es 1. Se tiene al menos 2 relaciones de cointegración.

**Tercera prueba:** El rango de la matriz  $\Omega$  igual a 2. Sea  $H_0 : r = 2$  contra  $H_1 : r > 2$ .

$$\begin{aligned} \lambda_{traza} &= 1,14 \\ \lambda_{max}(1, 2) &= 1,14 \end{aligned}$$

El valor crítico es igual a 4,12 tanto para la prueba de la traza como para la del máximo valor propio con un nivel de significación del 5 %. Por lo tanto, no se rechaza  $H_0$ ; se considera que el rango de la matriz es 2.

Se acepta la hipótesis que existen dos relaciones de cointegración.

**Observación 7.1.** A pesar de que la prueba de Johansen está implementada en algunos paquetes estadísticos, tiene algunos problemas al momento de realizar las predicciones. Hay observaciones sobre la especificación de los modelos derivados de la prueba de Johansen (Ver, Tsay 2008).

c. **Tercera Etapa:** Estimación de un modelo vectorial de corrección del error (MCE)

Se presentan los resultados del modelo encontrado. La estimación final de un VEC con 50 observaciones es el siguiente:

Cointegrating Eq:	CointEq1	CointEq2	
X1(-1)	1.000000	0.000000	
X2(-1)	0.000000	1.000000	
X3(-1)	-1.087335	-0.349631	
	(0.01946)	(0.07942)	
	[-55.8619]	[-4.40212]	
Error Correction:	D(X1)	D(X2)	D(X3)
CointEq1	-0.141775	-0.068674	0.323563
	(0.24049)	(0.09656)	(0.22752)
	[-0.58952]	[-0.71123]	[ 1.42214]
CointEq2	0.026182	0.019842	0.139864
	(0.06048)	(0.02428)	(0.05721)
	[0.43293]	[0.81719]	[2.44458]
D(X1(-5))	-0.640939	-0.117720	-0.159909
	(0.24748)	(0.09936)	(0.23413)
	[-2.58981]	[-1.18474]	[-0.68298]
D(X2(-5))	0.375685	0.330682	0.512122
	(0.55345)	(0.22221)	(0.52359)
	[0.67880]	[1.48817]	[0.97809]
D(X3(-5))	0.219208	-0.074681	-0.280249
	(0.26134)	(0.10493)	(0.24724)
	[0.83877]	[-0.71173]	[-1.13349]

Las expresiones de cointegración tienen la siguiente forma:

$$\begin{aligned} X_{1,t-1} - 1,087 * X_{3,t-1} \\ X_{1,t-2} - 0,350 * X_{3,t-1} \end{aligned}$$

Las ecuaciones de corrección del error son:

$$\begin{aligned} \Delta X_{1t} &= -0,641\Delta X_{1,t-5} + 0,376\Delta X_{2,t-5} + 0,219\Delta X_{3,t-5} \\ &\quad - 0,142(3X_{1,t-1} - 1,087 * X_{3,t-1}) + 0,026(X_{1,t-2} - 0,350 * X_{3,t-1}) \\ \Delta X_{2t} &= -0,118\Delta X_{1,t-5} + 0,331\Delta X_{2,t-5} - 0,075\Delta X_{3,t-5} \\ &\quad - 0,069(3X_{1,t-1} - 1,087 * X_{3,t-1}) + 0,020(X_{1,t-2} - 0,350 * X_{3,t-1}) \\ \Delta X_{3t} &= -0,160\Delta X_{1,t-5} + 0,512\Delta X_{2,t-5} - 0,280\Delta X_{3,t-5} \\ &\quad + 0,324(3X_{1,t-1} - 1,087 * X_{3,t-1}) + 0,140(X_{1,t-2} - 0,350 * X_{3,t-1}) \end{aligned}$$

#### d. Diagnóstico y verificación del modelo

Para verificar el modelo, se consideran los mismos estadísticos que se utilizaron en el caso del VAR.

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	7.902984	NA*	8.086774	NA*	NA*
2	24.71026	NA*	25.69440	NA*	NA*
3	29.93520	NA*	31.30165	NA*	NA*
4	32.89659	NA*	34.55918	NA*	NA*
5	35.08583	NA*	37.02909	NA*	NA*
6	41.49298	0.7349	44.44789	0.6192	48
7	52.58789	0.6411	57.64185	0.4513	57
8	63.54760	0.5627	71.03705	0.3137	66
9	71.85129	0.5817	81.47597	0.2849	75
10	79.78680	0.6099	91.74545	0.2639	84
11	86.84222	0.6601	101.1527	0.2644	93
12	90.95700	0.7751	106.8105	0.3527	102
13	96.73747	0.8306	115.0150	0.3779	111
14	101.9794	0.8817	122.7032	0.4144	120
15	107.4967	0.9161	131.0742	0.4325	129

**Tabla 7.4:** Prueba Portmanteau

La prueba expresa que no existe autocorrelación de los residuos.

#### Prueba de Jarque-Bera



Component	Jarque-Bera	df	Prob.
1	0.994354	2	0.6082
2	1.961766	2	0.3750
3	5.545940	2	0.0625
Joint	8.502061	6	0.2036

**Tabla 7.5:** Prueba de Jarque-Bera

La prueba de Jarque-Bera concluye que los residuos tienen una distribución normal, tanto de manera marginal como global. Finalmente, se muestra la prueba LM de correlación serial:

Lags	LM-Stat	Prob
1	7.935467	0.5407
2	16.71692	0.0533
3	5.199048	0.8166
4	2.892506	0.9684
5	13.83981	0.1281
6	8.614968	0.4735
7	14.98006	0.0915
8	12.22423	0.2010
9	13.94869	0.1242
10	11.32514	0.2541
11	10.29224	0.3273
12	7.225709	0.6136
13	10.47207	0.3136
14	7.876060	0.5467
15	9.297943	0.4102

**Tabla 7.6:** Prueba LM

Como se puede observar, no existe correlación serial en los residuos.

#### e. Predicciones a partir del MCE

Las predicciones obtenidas a partir del modelo son las siguientes:

	X <sub>1</sub>		X <sub>2</sub>		X <sub>3</sub>	
obs	Real	Predicción	Real	Predicción	Real	Predicción
51	165,071	167,8575	130,579	130,9032	169,377	171,4889
52	169,493	171,6238	132,612	133,0139	171,449	174,5759
53	172	173,1461	133,298	134,601	175,665	176,6129
54	175,099	176,1222	136,343	136,5069	177,327	178,8983
55	178,819	180,4314	138,931	138,8625	184,029	182,4962
56	180,759	183,9694	140,683	141,5225	183,299	186,2786

**Tabla 7.7:** Predicciones para las tres series utilizando el MCE

A continuación, se presenta gráficamente la comparación con los datos reales guardados:

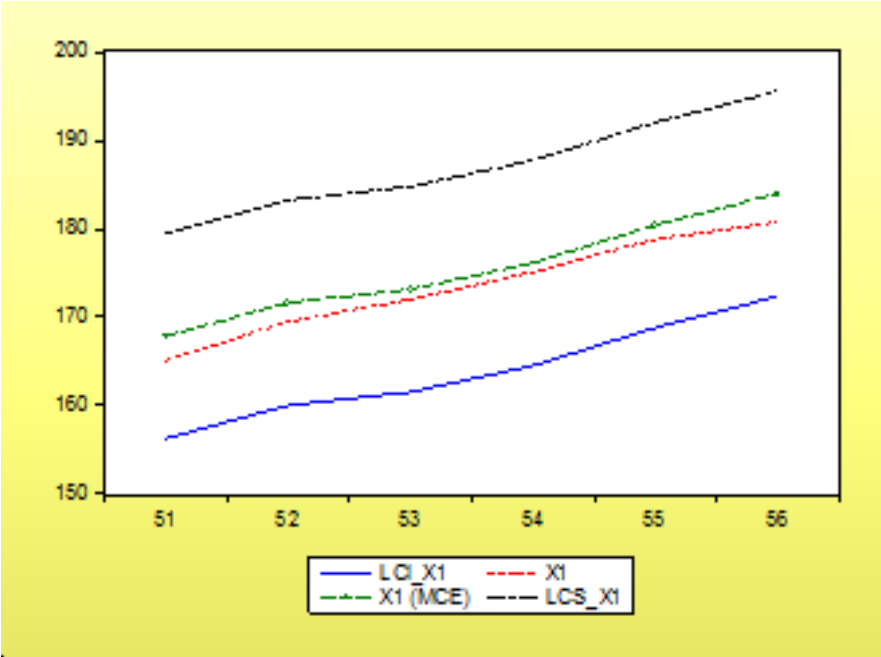


Figura 7.3: Predicción para  $X_{1t}$

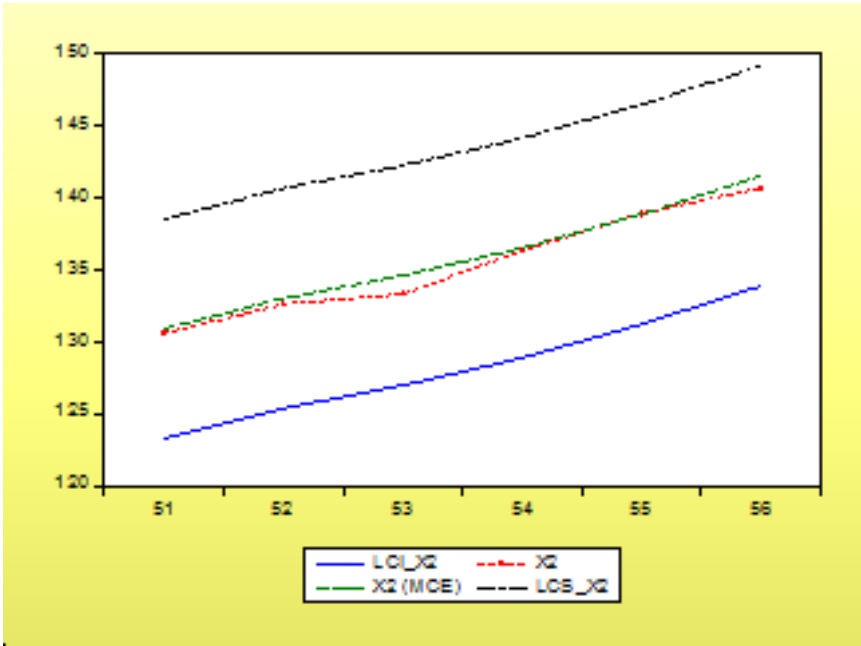


Figura 7.4: Predicción para  $X_{2t}$

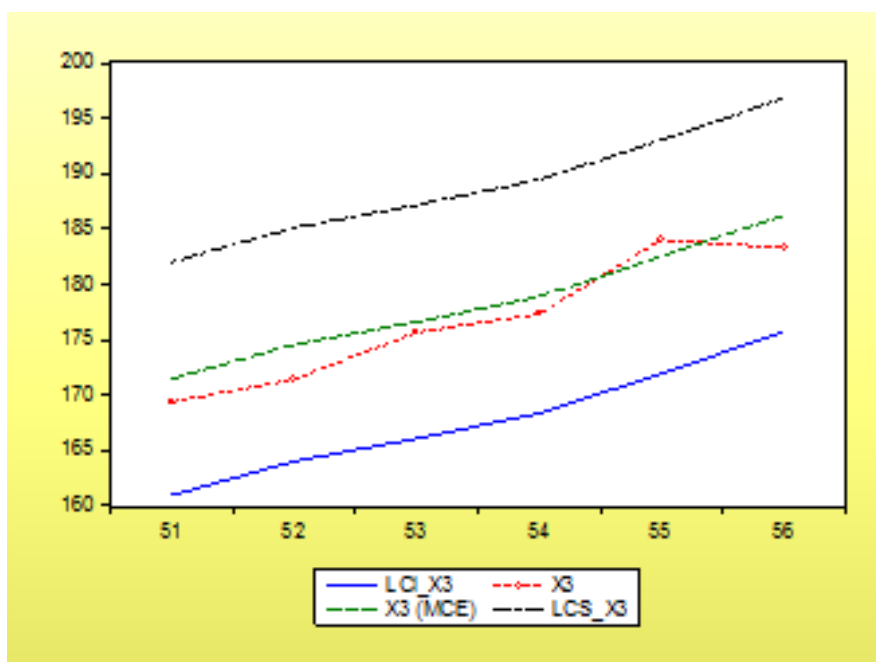


Figura 7.5: Predicción para  $X_{3t}$

En los gráficos las leyendas significan:

- LCI: Límite de confianza inferior.
- $X_i$ : Serie real de datos,  $i=1,2,3$ .
- $X_i$  (MCE): Predicción de la serie  $X_i$ ,  $i=1,2,3$ .
- LCS: Límite de confianza superior.

Adicionalmente, se presentan las predicciones de los modelos univariantes y se realiza la comparación con las predicciones del modelo de corrección del error. Así, para  $X_{1t}$  se estimó el modelo:  $\Delta \hat{X}_{1t} = 0,92X_{1t-1} + \hat{u}_{1t} - 0,96u_{1t-1} + 3,05$ . Para  $X_{2t}$  se obtuvo:  $\Delta \hat{X}_{2t} = 0,50X_{1t-1} + \hat{u}_{1t} + 1,75$  y; finalmente, para  $X_{3t}$  se encontró:  $\Delta \log(\hat{X}_{3t}) = 0,78X_{1t-6} + \hat{u}_{1t}$ .

A continuación, se presentan los gráficos con la comparación de las predicciones:

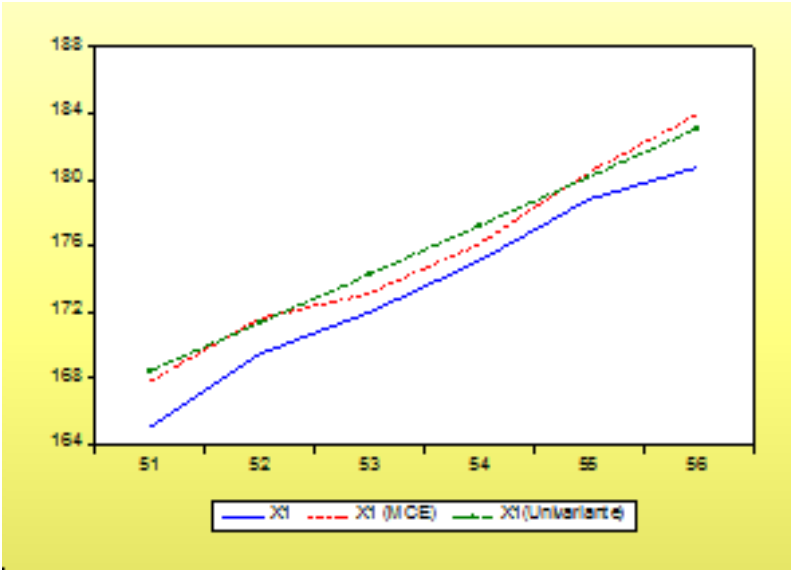


Figura 7.6: Comparación de las predicciones para  $X_{1t}$  entre el MCE y el modelo univariante.

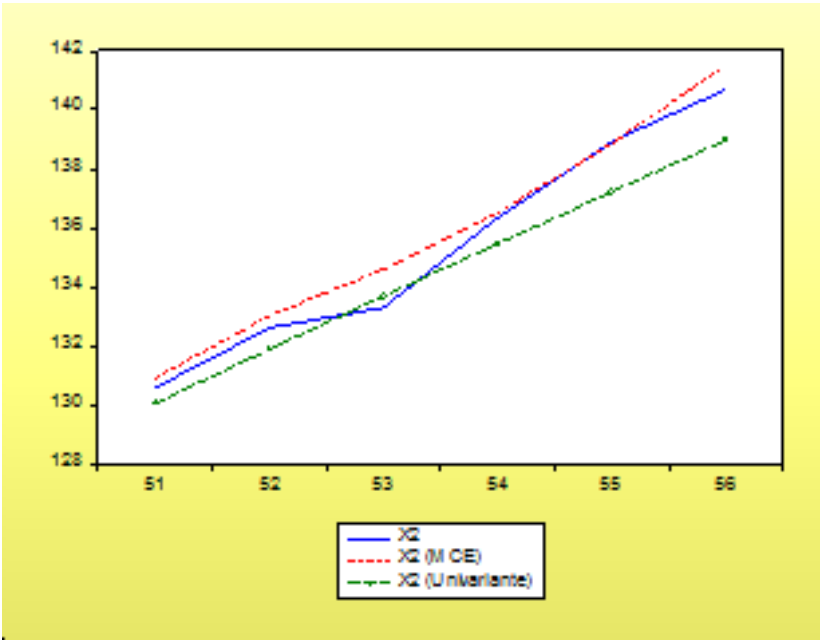
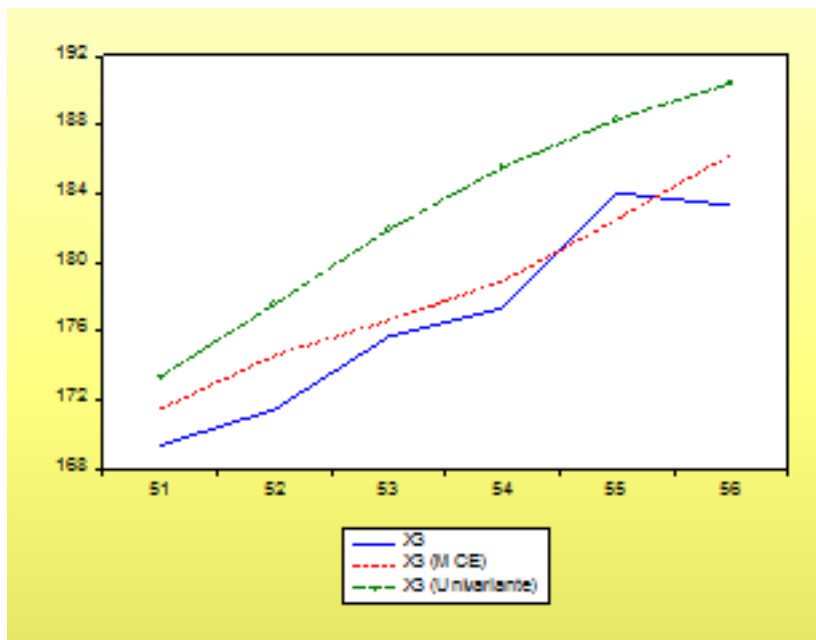


Figura 7.7: Comparación de las predicciones para  $X_{2t}$  entre el MCE y el modelo univariante.



**Figura 7.8:** Comparación de las predicciones para  $X_{3t}$  entre el MCE y el modelo univariante.

Se puede observar, en las tres series, que el ajuste de las predicciones realizadas por el MCE son mejores que las que se obtienen con el modelo univariante.

### 7.3 Prueba de Exogeneidad Débil

Esta prueba busca ver si las variables son endógenas o no. Esta prueba se refiere al coeficiente  $\alpha$  de la fuerza de empuje (si existe una sola relación de cointegración) o los coeficientes  $\alpha$  cuando existe más de una relación de cointegración. Realizar una prueba sobre  $\alpha$  comprueba si la relación de cointegración está presente en todas las ecuaciones del modelo. Esta prueba se realiza sobre la matriz  $\pi = \alpha' \beta$ , donde  $\beta$  son las relaciones de cointegración y  $\alpha$  los pesos de estas relaciones en cada una de las ecuaciones del sistema.

Por ejemplo, considere un MCE con dos relaciones de cointegración:

$$\begin{pmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \end{pmatrix} = A \begin{pmatrix} \Delta X_{1,t-1} \\ \Delta X_{2,t-1} \\ \Delta X_{3,t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{pmatrix} \begin{pmatrix} \beta_{11} & \beta_{21} & \beta_{31} \\ \beta_{12} & \beta_{22} & \beta_{32} \end{pmatrix} \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ X_{3,t-1} \end{pmatrix}$$

Para probar la exogeneidad de la variable  $X_{2t}$ , se debe realizar la prueba de hipótesis:

$$\begin{cases} H_0 : & \alpha_{21} = \alpha_{22} = 0 \\ H_a : & \alpha_{21}, \alpha_{22} \neq 0 \end{cases}$$

Si se acepta la hipótesis nula, esto significa que la fuerza de empuje  $\alpha$  no interviene en ninguna relación de cointegración y que la variable  $X_{2t}$  es débilmente exógena. En

este caso, se estima un MCE imponiendo la restricción  $\alpha_{21} = \alpha_{22} = 0$ .

## 7.4 Cointegración Estacional

Los modelos descritos anteriormente no cubren la variación estacional de las series de tiempo. Sin embargo, los resultados básicos de invertir los polinomios matriciales pueden ser extendidas para cubrir este caso también, con el resultado que se obtiene de una formulación de corrección del error y la posibilidad de calcular el rango de cointegración sobre diversas frecuencias complejas. Las asíntotas son, aproximadamente, las mismas que para el modelo usual, pero implican movimientos Brownianos complejos, como consecuencia de permitir raíces en las frecuencias complejas. El lector puede indagar más sobre el tema, por ejemplo, en "Likelihood analysis of seasonal cointegration" (Johansen and Schaumburg, 1998) o en "Cointegration: an overview" (Johansen, 2004).

# Índice alfabético

- Procesos estacionarios
  - Débilmente, 5
  - Fuertemente, 5
- ARCH
  - Definición, 93
  - Restricciones, 94
- Caminata aleatoria, 82
- Causalidad, 132
  - Granger, 133
  - Sims, 134
- Cointegración
  - Condiciones, 167
  - Definición, 165
  - Estacional, 186
- Dependencia lineal multivariante, 105
- Descomposición de Cholesky, 144
- Descomposición de Wold, 21
- Diferenciación, 81
  - Definición, 43
  - Tendencia cuadrática, 43
  - Tendencia exponencial, 44
  - Tendencia lineal, 43
  - Tendencia polinomial, 43
  - Estacional, 86
- EGARCH
  - Definición, 96
- Estimación, 3
  - Notación, 3
- Factor de impacto, 114
- Función
  - De autocorrelación, 28
  - De autocorrelación parcial, 28
  - De autocovarianza, 6
- GARCH
  - Definición, 94
- IGARCH
  - Definición, 95
- Innovación, 29
  - Ortogonal, 144
- Integral Estocástica, 16
- Matriz
  - Correlaciones cruzadas, 104
  - De correlación residual, 118
  - De covarianza residual, 118
- Media móvil, 8
- Medida espectral, 12
- Medida estocástica
  - De Wiener, 13
- Metodología de Box y Jenkins
  - Identificación a posteriori , 65
  - Verificación, 57, 65
- Metodología de Box y Jenkins, 57
  - Estimación de los coeficientes, 57
  - Estimación de los coeficientes , 60
  - Identificación a posteriori, 58
  - Identificación a priori, 57, 58
- Modelos MCE, 171
  - Diagnóstico y verificación, 180
  - Planteamiento, 171
  - Termino de corrección del error, 172
- Modelos VAR, 109
  - Bivariante, 109
  - Diagnóstico y validación, 118
  - Estabilidad, 111
  - Estimación de los coeficientes, 116
  - Forma estándar, 110
  - Forma estructural, 110
  - Forma matricial, 110
  - Predicción, 121
  - Representación de VAR(1), 112
  - Representación general, 111

- Representación lineal, 113
- Modelos VARMA, 115
- Modelos VMA
  - Orden  $q$ , 113
- Multiplicadores
  - De impacto, 144
  - Dinámicos, 137
- Operador
  - Avance, 27
  - Retardo, 26
- PARCH
  - Definición, 96
- Período de la estacionalidad, 85
- Predicción, 3
  - Modelo ARIMA, 72
  - Estabilidad estructural, 78
  - Funciones de, 77
  - Intervalos, 76
  - Lineal, 20
  - Modelos VAR, 121
  - No lineal, 20
- Proceso
  - Determinista, 21
  - Equivalentes, 23
  - Gaussiano, 22
- Proceso ARIMA, 43
  - Definición, 43
  - Orden  $(p, d, q)$ , 43
- Proceso lineal
  - Definición, 25
  - Filtro lineal, 26
  - Invertible, 26
- Proceso no estacionario
  - Media no constante, 9
  - Tendencia no constante y varianza no homogénea, 10
  - Varianza no homogénea, 9
- Procesos AR, 29
  - Coeficiente de autocorrelación, 32
  - Coeficiente de autocorrelación parcial, 33
  - Orden  $p$ , 29
  - Orden 1, 29
  - Reconocer y determinar, 33
- Procesos ARIMA
  - Representación AR, 45
  - Representación MA, 45
- Procesos ARMA, 40
  - Coeficiente de autocorrelación, 41
  - Coeficiente de autocorrelación parcial, 41
  - Orden  $(p, q)$ , 40
  - Representación canónica, 41
- Procesos MA
  - Coeficiente de autocorrelación, 39
  - Coeficiente de autocorrelación parcial, 39
  - Orden  $q$ , 38
  - Reconocer y determinar, 40
- Procesos SARIMA
  - Orden  $[(p, d, q)(P, D, Q)]$ , 85
- Prueba
  - Exogeneidad débil, 185
  - Jarque-Bera multivariante, 120
  - Johansen, 173
  - Multiplicador de Lagrange multivariante, 119
  - Portmanteau multivariante, 119
- Prueba de Chow
  - Predicción, 78
- Raíz unitaria, 81
  - Prueba de Dickey-Fuller, 82
  - Prueba de Dickey-Fuller Aumentada, 83
- Relaciones en retardo, 104
- Ruido blanco, 6
- TARCH
  - Definición, 96
- Tendencia, 9
- Teorema
  - Ergódico débil, 18
  - Ergódico fuerte, 19
- Teorema Central de límite, 20
- Transformación de datos
  - Box-Cox, 81
  - Logarítmica, 79