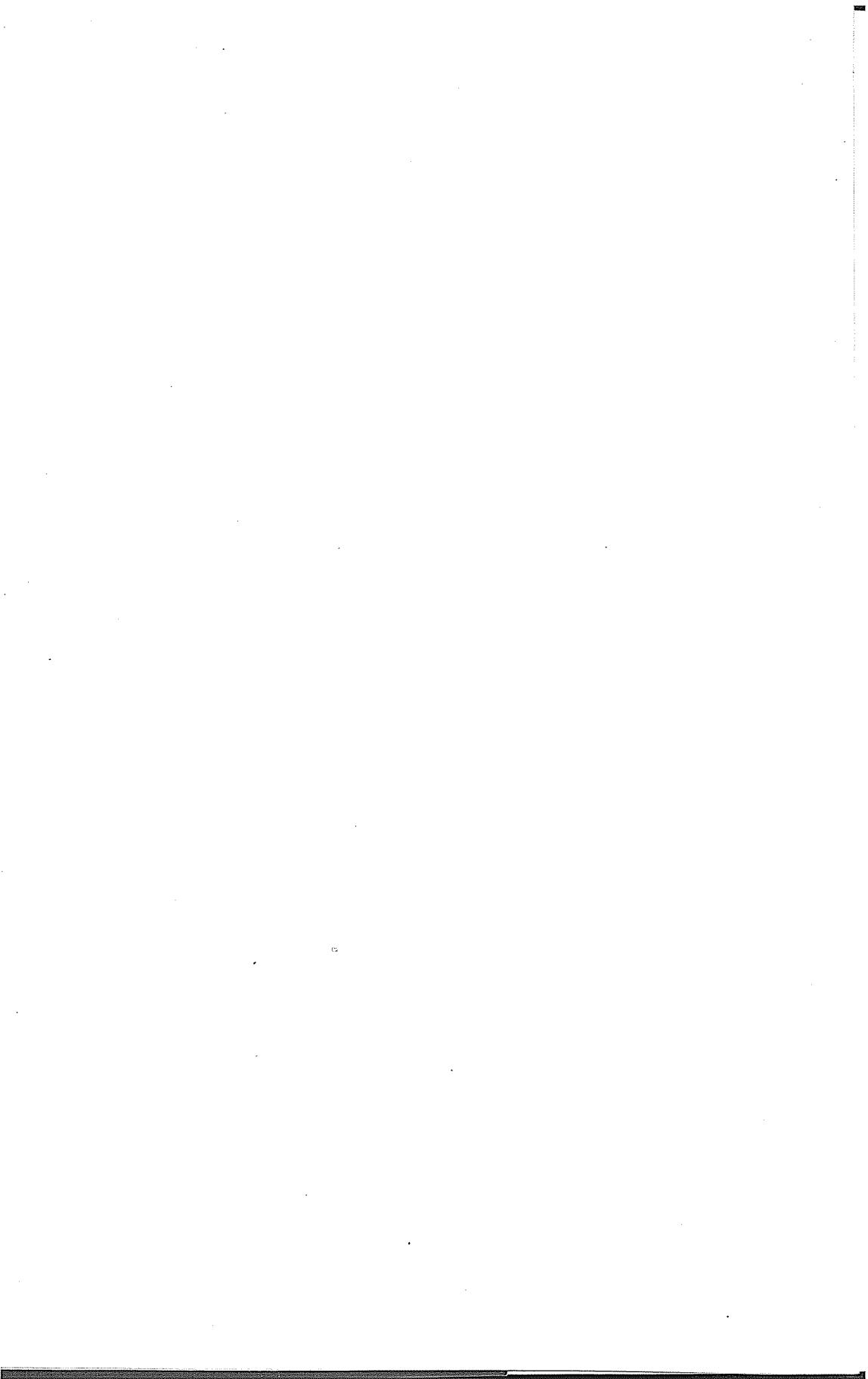


ANÁLISIS DE SERIES DE TIEMPO

Álvaro Montenegro García

Análisis de series de tiempo



Análisis de series de tiempo

Álvaro Montenegro García







Reservados todos los derechos
© Pontificia Universidad Javeriana
© Álvaro Montenegro García

Primera edición: abril de 2011
Bogotá, D.C.

ISBN: 978-958-716-396-4

Número de ejemplares: 300

Impreso y hecho en Colombia
Printed and made in Colombia

Editorial Pontificia Universidad Javeriana
Carrera 7 núm. 37 – 25, piso 13
Edificio Lutaima
Teléfono: 2870691 ext. 4752
www.editorialpuj@javeriana.edu.co
Bogotá, D.C.

Corrección de estilo
Ella Suárez

Diseño de cubierta
Carolina Maya

Diagramación en LATEX
Daniel Contreras Niño

Impresión
Javegraf



Montenegro García, Álvaro

Análisis de series de tiempo / Álvaro Montenegro García. – Bogotá:
Editorial Pontificia Universidad Javeriana, 2010.

398 p. : ilustraciones, gráficas y tablas ; 24 cm.

Incluye referencias bibliográficas.

ISBN: 978-958-716-396-4

1. ANÁLISIS DE SERIES DE TIEMPO. 2. MATEMÁTICAS FINANCIERAS. 3. ESTADÍSTICA MATEMÁTICA. 4. ESTADÍSTICA PARA ADMINISTRADORES. 5. ESTADÍSTICA PARA ECONOMISTAS.
I. Pontificia Universidad Javeriana. Facultad de Ciencias Económicas y Administrativas.

CDD 658.00151 ed. 21

Catalogación en la publicación - Pontificia Universidad Javeriana.
Biblioteca Alfonso Borrero Cabal, S.J.

ech.

Octubre 21 / 2010



Índice general

Prólogo

1. Introducción	13
2. Conceptos y herramientas de análisis	23
Estacionariedad	23
Algunas formas de lograr estacionariedad	25
Ruido blanco	26
Ergodicidad	27
Función de autocovarianza	29
Función de autocorrelación	30
Estimación de $R(\tau)$	30
Estimación de $r(\tau)$	31
Función de autocorrelación en la práctica	33
Función de correlación cruzada	36
Operador de rezago	37
Ecuación de diferencias	38
Raíces imaginarias	43
Operador de diferencia	45
Suavizado aritmético	45
Suavizado exponencial de Holt-Winters	47
Filtro de Hodrick-Prescott	48

Ajuste estacional	49
3. Modelo autorregresivo y modelo de promedio móvil	55
El modelo autorregresivo AR(p)	55
El AR(1)	56
El AR(2)	60
El AR(p)	63
El modelo de promedio móvil MA(q)	65
MA(1)	66
Invertibilidad	67
Teorema de descomposición de Wold	70
Modelo ARMA(p,q)	71
Diagramas de bloque	73
Función de autocorrelación parcial	74
Estacionalidad en los modelos ARMA	81
Respuesta impulso de un AR(p)	83
Predicción utilizando un AR(p)	85
Predicción utilizando un MA(q)	86
Predicción utilizando un ARMA(p,q)	86
Varianza de predicción en modelos ARMA	87
4. Estimación de modelos ARMA(p,q)	91
La función de autocorrelación y la función de autocorrelación parcial	94
Criterios de información de Akaike y Schwarz	94
Estimación de los modelos AR(p)	96
Estimación por el método de momentos de Yule-Walker	96
Estimación por mínimos cuadrados ordinarios	97
Intuición del teorema de Mann-Wald	100
Teorema de Mann-Wald	102
Mínimos cuadrados ordinarios, autorregresiones y autocovarianzas	103
Estimación por máxima verosimilitud condicional	107
Estimación por máxima verosimilitud exacta	109
Estimación de la varianza del error	114
Estimación de los modelos MA(q)	115
Estimación de los modelos ARMA(p,q)	122

Estimación asintótica de un ARMA por mínimos cuadrados ordinarios	124
Intuición de la optimización numérica	125
Valores iniciales para una optimización numérica	130
5. Modelos estacionarios multivariados	133
Motivación para los modelos multivariados	133
El modelo VAR	136
Raíces y eigenvalores	142
El modelo VARMA	145
Intuición para la estabilidad o invertibilidad de un VARMA	146
Respuesta impulso del VAR	146
Predicción con un VAR(p)	157
Varianza de predicción	157
Descomposición de la varianza del error de predicción	159
Vectores autorregresivos estructurales	160
6. Predicción económica	165
Conjunto de información	166
Funciones de pérdida	166
Predicción óptima	171
Intuición de la predicción óptima	172
Errores de predicción	173
Predicción e intuición	174
Combinación de predicciones	176
Consenso <i>vs.</i> incertidumbre	177
La maldición del ganador	178
Criterios para evaluar predicciones	179
Error cuadrado medio	181
Raíz del error cuadrado medio	181
Error medio absoluto	181
Error medio porcentual absoluto	182
Regresión	182
Índice de Theil	183
Descomposición del error cuadrado medio	183
Algunas evaluaciones empíricas	185
Métodos de predicción	186

Indicadores líderes	187
Causalidad	190
Definición de causalidad de Granger	191
Implementación lineal de la prueba de causalidad	193
7. Modelos ARCH	197
Modelo ARCH	199
Modelo GARCH	200
Modelo ARCH-M	203
Identificación del modelo ARCH	204
Estimación del modelo ARCH	205
R^2 negativo	209
Predicción con un ARCH	211
Otras aplicaciones	213
Valor en riesgo	215
8. Procesos estocásticos no estacionarios	217
Paseo aleatorio	218
Regresiones espurias	225
Series integradas	229
9. Raíces unitarias y cointegración bivariada	235
Cointegración bivariada	237
Pruebas de raíz unitaria	238
Estimación de la cointegración bivariada	251
Modelo de corrección de error	253
10. Cointegración multivariada	257
Definiciones	257
Representación del vector autorregresivo y del de corrección de error	263
Vector autorregresivo en diferencias	268
Pruebas de causalidad en un vector de corrección de error . .	268
Estimación por máxima verosimilitud	270
Algoritmo de Johansen	271
Estimativos de los parámetros del vector autorregresivo . .	280
Estadísticos de prueba	281
Cointegración en la práctica	282

Ejercicios	293
Referencias	335
A. Análisis espectral	345
Funciones senoidales	346
Representación senoidal	347
El periodograma	351
La autocovarianza	355
El espectro	356
Propiedades de muestreo	357
Referencias	363
B. Filtro de Kalman	365
Estructura del modelo estado-espacio	366
Ecuaciones del filtro de Kalman	369
Referencias	378
C. Redes neurales artificiales	381
Especificación	384
Estimación	385
Evaluación	386
Predicción	387
Referencias	395

Prólogo

El tema que nos ocupa en este texto es el desarrollo y la aplicación de técnicas estadísticas para modelar y predecir el comportamiento de variables a través del tiempo, específicamente en el área de la economía. Si bien las técnicas son universales, cada disciplina las aplica a sus propias variables. Por ejemplo, en economía se aplican a variables como precios, producción o empleo. En otras disciplinas se utilizan para estudiar la temperatura, la pluviosidad, la tasa de natalidad, la polución, el colesterol, el conteo de peces, el número de asaltos bancarios, la cantidad de manchas solares, etc.

En el análisis de series de tiempo, no obstante, el relieve se pone en los métodos estadísticos y matemáticos que nunca deben perder de vista ni la teoría económica ni el contexto histórico en el cual evolucionan las variables. Más aún, veremos que la práctica aceptada entre la profesión es la de combinar los resultados del modelo matemático con información cualitativa o intuitiva para afinar las predicciones.

Este texto tiene como requisito un manejo adecuado de conceptos de álgebra lineal, probabilidad, estadística y econometría multivariada, como son los de matrices y vectores, variables aleatorias, independencia, condicionalidad, funciones de probabilidad, valor esperado, límites de probabilidad, propiedades de los estimadores, pruebas de hipótesis, así como el modelo clásico de regresión y sus supuestos. En otras palabras, el texto está diseñado para servir como un curso posterior al de econometría tradicional. Los ejemplos y ejercicios pueden ser complementados por el estudiante con talleres que traten aspectos de recolección de datos, depuración, procesamiento, análisis, predicción y evaluación.

Esta edición incluye una revisión general y adición a temas anteriores, con el objeto de completar y actualizar la presentación. Igualmente, se incluyen más ejemplos a lo largo del texto y nuevos problemas en la sección de ejercicios. Se abrió un “Apéndice” con tres temas complementarios: análisis espectral, filtro de Kalman y redes neurales. Agradezco muchísimo a un gran número de estudiantes y profesores que durante varios años han contribuido con sus comentarios y sugerencias a mejorar el contenido del texto.

CAPÍTULO 1

Introducción

La expresión serie de tiempo se utiliza para denotar una secuencia de datos empíricos ordenados en función del tiempo; esto es, que puedan graficarse contra el tiempo. A diferencia de los datos usados en otros estudios estadísticos y econométricos, como aquellos de corte transversal, aquí el ordenamiento en función del tiempo es determinante y es parte integral del análisis.

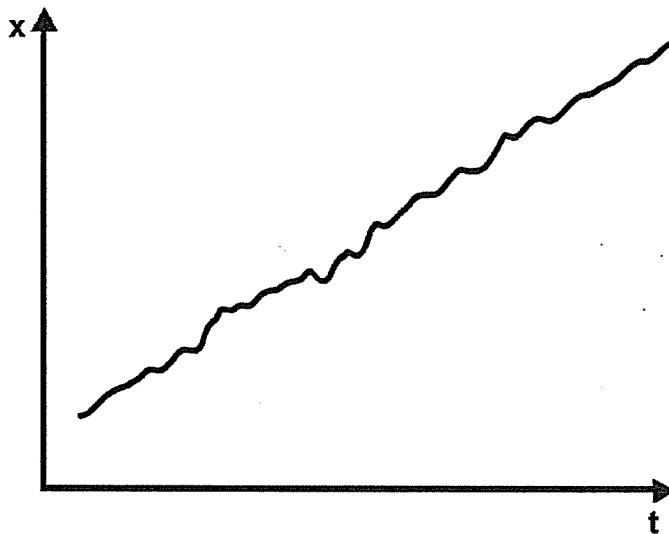
La definición anterior incluye dos tipos de series de tiempo: las que son impredecibles total o parcialmente, porque tienen algún elemento de aleatoriedad (de naturaleza estocástica), y las que son perfectamente predecibles (de naturaleza determinística). Sin embargo, la práctica y la teoría econométrica se ocupan mayormente de las series de naturaleza estocástica, que se modelan mejor incluyendo un término de error. Más adelante precisaremos estos conceptos, al introducir y definir los procesos determinísticos y estocásticos.

Una gráfica es la primera y más importante herramienta en el análisis de series de tiempo. Siempre se deben graficar los datos; esto permite obtener pistas iniciales sobre el comportamiento de las series, y ello sugiere unos procedimientos y modelos y descarta otros.

Las figuras 1.1 a 1.6 ilustran varios conceptos básicos de las series de tiempo. La Figura 1.1 muestra una serie de tiempo x que tiene una

tendencia lineal ascendente o positiva (si es descendente sería negativa). Si, por ejemplo, la serie x fuera el PIB real per cápita de un país, ¿sería bueno vivir en ese país?, ¿depende de la pendiente? La respuesta es no, y no depende de nada. La razón es que el cambio porcentual de x , denotado $\Delta x \%$, o sea el crecimiento en el PIB real per cápita, sería cada vez menor, como se muestra en la Figura 1.2. Sería mejor vivir en un país con un PIB real per cápita con tendencia exponencial, como el de la Figura 1.3, cuyo crecimiento porcentual se muestra en la Figura 1.4.

Figura 1.1: Serie de tiempo con tendencia lineal



La tendencia en una serie describe su comportamiento de largo plazo. Es un movimiento de baja frecuencia, de evolución lenta. En términos coloquiales, responde a ¿para dónde va la serie?, ¿hacia dónde parece moverse en el largo plazo? Aunque el caso más común es cuando la tendencia ascienda o descienda monotónicamente, también puede, en algún momento, cambiar de rumbo; lo que no puede hacer es cambiar de rumbo con frecuencia. La tendencia de una serie puede modelarse de manera determinística o estocástica, o mixta, según convenga.

Figura 1.2: Crecimiento porcentual de la Figura 1.1



Figura 1.3: Serie de tiempo con tendencia exponencial

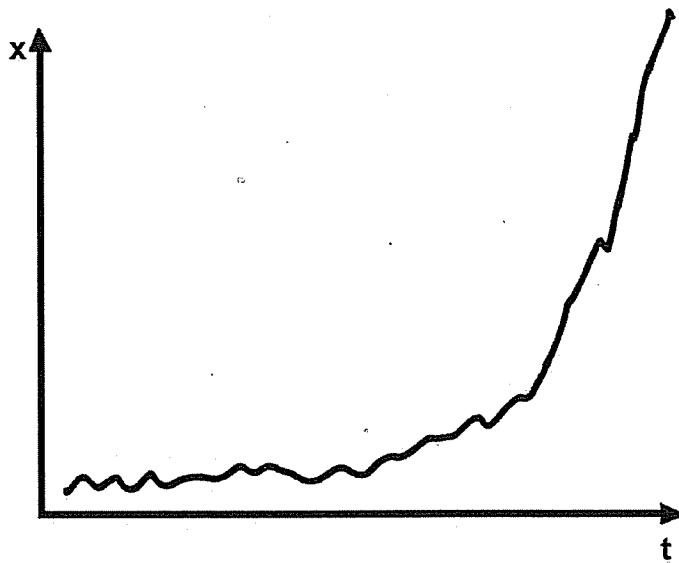
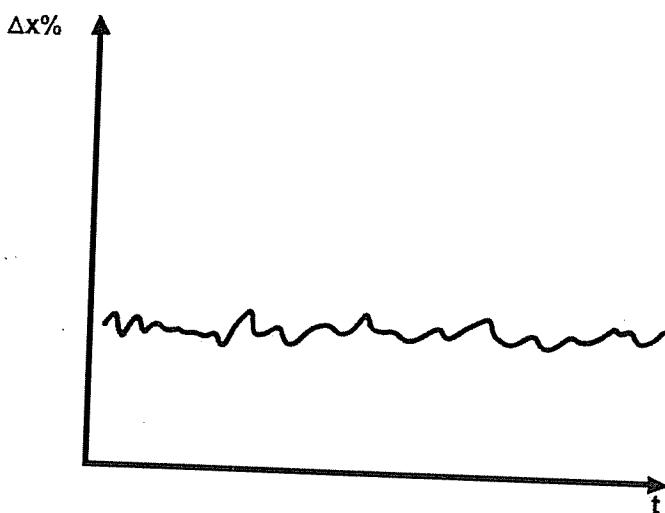


Figura 1.4: Crecimiento porcentual de la Figura 1.3



La media \bar{x} de la serie (el promedio de los datos) no es una medida representativa de las series en las figuras 1.1 y 1.3, pero sí lo es para la serie de la Figura 1.4, ya que en esta última la media está mucho más cerca de los datos porque la varianza es pequeña. El concepto de varianza σ_x^2 tiene poco sentido en series que caen o suben indefinidamente, como en las figuras 1.1 y 1.3.

La media de una serie puede ser positiva, negativa o cero. En muchos casos es deseable o necesario trabajar con series cuya media sea cero; pero si no es cero, simplemente se resta \bar{x} a cada una de las observaciones de la serie, a fin de obtener una nueva serie con media cero, que preserve todas las demás características y el comportamiento de la serie original.

Muchas series en economía, quizás la mayoría, tienen tendencia; por ejemplo, el nivel del PIB, la cantidad de dinero, la población, etc. Sin embargo, también hay series importantes que siempre, o para ciertos intervalos de tiempo, podrían considerarse sin tendencia, como la tasa de desempleo, la tasa de interés real o el crecimiento en el PIB.

La Figura 1.5 muestra una serie que carece de tendencia, pero que representa parte de un ciclo u oscilación de largo plazo. En economía, por la naturaleza estocástica de las series, los ciclos nunca están claramente demarcados ni en amplitud ni en frecuencia (a diferencia de una

onda senoidal pura, por ejemplo), sino que se insinúan. La alternación de épocas de aumento y disminución, como en la historia bíblica de las vacas gordas y flacas, forma los ciclos económicos.

Figura 1.5: Serie de tiempo cíclica

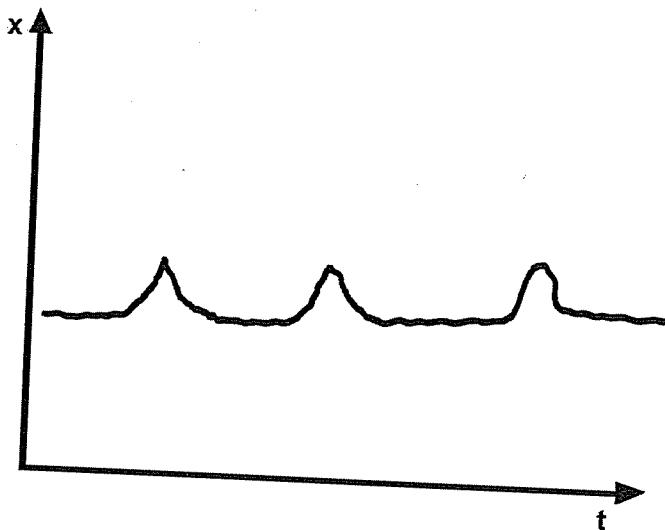


La Figura 1.6 ilustra un comportamiento, relacionado con el de los ciclos, llamado estacional. La estacionalidad es un movimiento similar que se presenta en la misma época del año; por ejemplo, cuando las ventas de juguetes se elevan todos los diciembre o cuando los precios caen en el trimestre de cosecha. El concepto de estacionalidad usualmente se aplica sólo a periodicidades menores de un año (por ejemplo, diario, semanal, mensual, trimestral, etc.). Este concepto de estacionalidad no debe confundirse con el de estacionariedad, que discutiremos más adelante.

En principio, series como las de las figuras 1.1 y 1.3 son relativamente fáciles de predecir, al menos en tendencia, ya que no parecen cambiar de dirección de manera abrupta, propiedad conocida en la física como inerzia en econometría se conoce como correlación serial o autocorrelación positiva. Cero inercia sería como un ovni que puede zigzaguear o quebrar su trayectoria de manera instantánea. Cuando la inerzia es grande, como en el caso de las series de las figuras 1.1 y 1.3, el dato en el tiempo $t - 1$ sirve para predecir el siguiente en t .

En series que no tienen tendencia, como en la Figura 1.4, el interés radica en predecir la variación o el cambio. También en este caso la

Figura 1.6: Comportamiento estacional



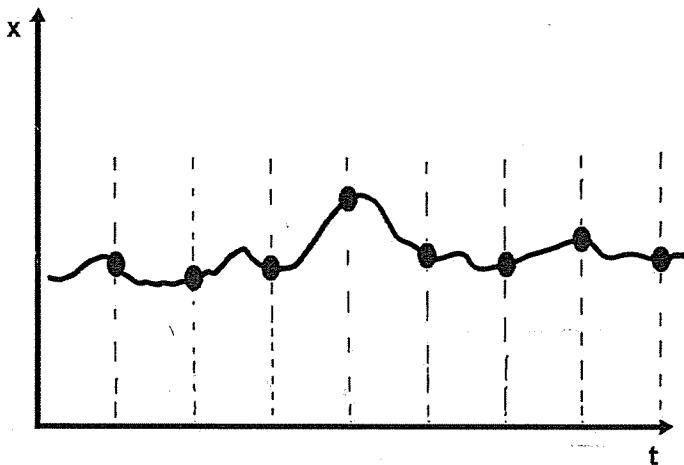
observación en $t - 1$ puede servir para predecir la de t . Para esto es necesario que sepamos si existe algún patrón o relación estable entre observaciones cercanas. Por ejemplo, saber que una observación por encima (por debajo) de la media tiende a ser seguida por otra por encima (por debajo) de la media, o si, al contrario, con cierta regularidad una observación por encima de la media tiende a ser seguida por otra por debajo, o viceversa.

Si en una serie las observaciones pasadas no proveen ninguna información acerca del movimiento de las observaciones futuras, dicha serie será imposible de predecir utilizando la información de su propio pasado, y se denomina *ruido*. Por ejemplo, el número que salga del lanzamiento de un dado no dice nada nuevo acerca de la probabilidad del próximo número; seguirá siendo $1/6$ no importa cuál salió antes. Son eventos independientes, al igual que las observaciones de una serie de ruido.

Desde el punto de vista de la forma como se toman o registran las observaciones, podemos clasificar las series de tiempo en *continuas*, que denotaremos $x(t)$, o en *discretas*, que denotaremos x_t . Serán *continuas* cuando tengamos observaciones o datos para todo instante de tiempo y serán *discretas* cuando tengamos observaciones o datos sólo en ciertos momentos de tiempo, usualmente a intervalos regulares.

En la Figura 1.7 se ilustran ambos conceptos. La curva corresponde a observaciones de una serie continua (pues hay observaciones para todo tiempo t), mientras que los puntos indicados sobre la curva forman una serie discreta (pues sólo hay observaciones en algunos momentos del tiempo). Aunque existen desarrollos teóricos para ambos tipos de series, en la práctica la mayor parte del análisis de series de tiempo en economía se ocupa de series discretas. En principio, la realidad podría ser continua, pero en la mayoría de los casos sólo podemos, o queremos, registrarla de manera discreta.

Figura 1.7: Series de tiempo continuas y discretas



Como mencionamos antes, las series de tiempo pueden originarse en dos tipos de procesos: el estocástico y el determinístico. En otras palabras, el probabilístico y el no probabilístico. Aunque los desarrollos determinísticos ocupan un lugar importante en la literatura, la mayor parte del análisis teórico y práctico de series de tiempo se refiere a los procesos estocásticos; esto es, a procesos de origen probabilístico, como repeticiones de un experimento aleatorio. También es común encontrar series mixtas, con un componente determinístico (una tendencia, una oscilación, etc.) y uno estocástico (un término de error). Formalizaremos estos conceptos a continuación.

Un proceso estocástico $\{x_t\}$ se define como una secuencia de variables aleatorias ordenadas por un índice t , usualmente el tiempo, y puede escribirse como

$$\{x_t\} = \{\dots, x_1, x_2, x_3, \dots\}$$

Un proceso estocástico no es una serie de valores o datos ya conocidos a través del tiempo, esto lo llamamos una serie de tiempo o una realización, sino de variables aleatorias, una para cada momento del tiempo. Nunca podremos predecir con total certeza el resultado para cada momento de tiempo, sólo podremos hablar de probabilidades, y aproximarnos a través de una distribución de probabilidad.

El proceso estocástico es como un experimento aleatorio que consiste en asignar valores a x_t . Por ejemplo, lanzar un dado cien veces e ir graficando el resultado. Cada vez que repitamos el experimento (lanzar el dado cien veces) tendremos una gráfica distinta. Sin embargo, en la práctica es común que no podamos repetir el experimento cuando queramos, sino que dispongamos sólo de los datos que provee la naturaleza o el sistema; esto es, tendremos una de las posibles secuencias de resultados, a la cual nos referiremos como realización o, en general, como serie de tiempo.

Existe cierto paralelismo entre lo que es una realización (o serie de tiempo) y lo que es una muestra tomada de una población. La muestra es una realización de la población con la cual esperamos poder hacer inferencias útiles sobre la población. De manera similar, una serie de tiempo es una realización, una de tantas posibles, del proceso estocástico (que en este caso es como la población), con la cual esperamos hacer inferencias útiles sobre el proceso estocástico.

En algunos textos, el proceso estocástico también es apropiadamente conocido como mecanismo generador de datos (data generating mechanism). En breve, así como existen varios resultados posibles de un experimento aleatorio, llamados eventos, un proceso estocástico puede originar varias posibles series de tiempo, llamadas realizaciones. El conjunto de todas las realizaciones posibles se denomina ensamblaje y corresponde al concepto de la población o espacio muestral en la teoría elemental de probabilidad.

En contraste, la característica principal de un proceso determinístico es que todas las repeticiones son iguales. No importa cuántas veces se repita el experimento, el resultado será el mismo. Obviamente, si todas las repeticiones o realizaciones son idénticas, estaremos tratando con series que teóricamente pueden predecirse sin error, incluidos, por

ejemplo, los experimentos comunes de la física mecánica, como el espacio recorrido en cada instante por un cuerpo en caída libre o el comportamiento de una variable gobernada por una ecuación senoidal. También se incluye la nueva teoría del caos, que trata de procesos que parecen aleatorios, pero que generan series de tiempo perfectamente determinísticas.

En la práctica, el grado de determinismo puede depender de la disponibilidad de datos. Por ejemplo, si sospechamos que un fenómeno es determinístico, pero no disponemos de todas las variables o datos necesarios, posiblemente tengamos que estudiarlo desde el punto de vista estocástico, adicionándole un término aleatorio que ocupe el lugar de la información faltante. Este es precisamente el papel del término de error en cualquier ecuación de regresión clásica.

Utilicemos un ejemplo para ilustrar el concepto de proceso estocástico y el de realizaciones, y la diferencia con un proceso determinístico. Imaginemos una pista de un aeropuerto. Supongamos que queremos graficar la rugosidad de la superficie de la pista al desplazarnos a lo largo de ella, como en la Figura 1.4, sólo que en el eje horizontal no iría el tiempo, sino la distancia recorrida. Si la pista fuera plana como un espejo, la gráfica sería una línea horizontal; pero en la realidad es rugosa. Hacemos varios recorridos paralelos a lo largo de la pista y registraremos cada recorrido como una realización; cada nueva realización (o muestra) será diferente a las demás, pero todas tendrán características probabilísticas comunes, igual media, varianza, etc.

Otro ejemplo podría ser la medición del grosor de un hilo a lo largo de varios segmentos del mismo hilo. Cada segmento producirá una gráfica similar a la Figura 1.4, donde el eje horizontal sería la distancia recorrida. Otros segmentos del hilo producirán gráficas diferentes pero, de nuevo, con características probabilísticas comunes.

Podemos pensar que un proceso estocástico es el ente probabilístico que origina la serie de tiempo. Por ejemplo, los datos de M1 mensual en los últimos 20 años constituyen una serie de tiempo ya realizada; sin embargo, antes de conocer la observación del tiempo t , el valor de M1 era una variable aleatoria x_t de la cual sólo podíamos postular algún comportamiento probable, sin exactitud total.

Supongamos que se construye una serie de tiempo x_t a partir de un experimento consistente en lanzar un dado cada cierto intervalo

de tiempo. El valor que puede mostrar el dado en el primer lanzamiento es x_1 , en el segundo es x_2 , en el tercero es x_3 y así sucesivamente. Antes de lanzar el dado, x_1 tiene una distribución de probabilidad que en este caso es una distribución discreta uniforme con media 3.5 y probabilidad de 1/6 para cada uno de los números 1, 2, 3, 4, 5 y 6. Para el segundo lanzamiento, la variable x_2 tiene también una distribución uniforme con media 3.5 y probabilidad de 1/6 para cada uno de los números 1, 2, 3, 4, 5 y 6. Igual cosa sucede con los lanzamientos siguientes. Podríamos observar la serie de tiempo así realizada graficando en el eje vertical el resultado de cada lanzamiento y en el eje horizontal el tiempo.

¿Qué ocurre si, por ejemplo, el dado está cargado? De todas maneras podemos construir la serie, sólo que ahora cambia su apariencia visual, porque cambia la distribución de probabilidad que gobierna cada lanzamiento; ya no será uniforme con probabilidad 1/6 para cada uno de los números 1, 2, 3, 4, 5 y 6, sino que algún número tendrá mayor probabilidad que los otros. Y ¿qué ocurre si el jugador es hábil y sabe manipular el dado de manera que éste caiga como él quiere? De nuevo, cambia la apariencia visual, pues el jugador tiene la capacidad de moldear la distribución de probabilidad.

Las series económicas de dinero, producción, precios, tasas de cambio, etc., se originan de manera similar. Podemos pensar que las genera un jugador, hábil o no, consciente o inconsciente de lo que está haciendo, que puede ser Dios, el gobierno, la gente, o todo o parte del sistema económico. El análisis de series de tiempo tiene por objeto descubrir, estudiar y utilizar (para política económica y predicción) este mecanismo de generación, llamado formalmente *proceso estocástico*.

Infortunadamente, casi nunca tendremos más información útil que los datos observados en una serie de tiempo (la realización) y debemos aprovechar al máximo la información en ellos contenida.

CAPÍTULO 2

Conceptos y herramientas de análisis

Estacionariedad

El comportamiento de una serie de tiempo, su forma, sus movimientos y su apariencia visual dependen de la distribución de probabilidad que gobierna el mecanismo de generación de datos, o proceso estocástico, que la produce. A su vez, dichas características probabilísticas pueden cambiar o no cambiar en el tiempo. Si cambian en el tiempo será mucho más difícil modelar y más y riesgoso hacer inferencias y predicciones con base en la información provista por los datos de la serie de tiempo disponible. Pero si no cambian, condición que se conoce como *estacionariedad*, será más fácil modelar y hacer predicciones porque podremos confiar en la estabilidad de la especificación y de los parámetros estimados, aun por fuera de la muestra.

Un proceso estocástico $\{x_t\}$ es estrictamente estacionario si sus propiedades estadísticas o probabilísticas no cambian con el tiempo: esto es, si su función de distribución acumulativa es independiente del tiempo. En símbolos, lo anterior quiere decir:

$$F(x_1, x_2, \dots, x_n) = F(x_{1+\tau}, x_{2+\tau}, \dots, x_{n+\tau}) \quad (2.1)$$

para todo n y rezago τ finitos admisibles.

La definición de estacionariedad estricta se da en términos de la función de distribución acumulativa, ya que ésta resume todas las características probabilísticas de cualquier experimento aleatorio o proceso estocástico. También puede expresarse en términos de la función de densidad de probabilidad $f(x_1, x_2, \dots, x_n)$. La definición anterior implica que los momentos de la distribución, si existen y son finitos, deben ser estacionarios. La salvedad sobre los momentos se hace, porque algunos procesos pueden ser estacionarios, como los de Cauchy, cuyos momentos no están definidos.

Una definición menos restrictiva, y menos difícil de probar, es la siguiente (véase Priestley, 1981, p. 105). Se dice que hay estacionariedad de orden m si para cualquier n y rezago τ , todos los momentos conjuntos hasta de orden m de $\{x_1, x_2, \dots, x_n\}$ existen y son iguales a los momentos conjuntos hasta de orden m de $\{x_{1+\tau}, x_{2+\tau}, \dots, x_{n+\tau}\}$, esto es, si:

$$E(x_1^{m_1} x_2^{m_2} \dots x_n^{m_n}) = E(x_{1+\tau}^{m_1} x_{2+\tau}^{m_2} \dots x_{n+\tau}^{m_n}) \quad (2.2)$$

Para todos los enteros no negativos m_1, m_2, \dots, m_n que satisfagan:

$$m_1 + m_2 + \dots + m_n \leq m.$$

Según esta definición, por ejemplo, en el proceso estocástico dado por $\{\dots, x_1, x_2, x_3, \dots\}$ las variables podrían tener distribuciones diferentes siempre y cuando tengan iguales momentos hasta de orden m .

Un caso particular de la anterior definición (2.2), donde $m = 2$, es el conocido como estacionariedad débil o de orden 2 (*covariance stationary*). Se dice que hay estacionariedad débil si solamente los momentos hasta de orden dos (la media, la varianza y las covarianzas) son finitos e independientes del tiempo. Notamos que un proceso gaussiano débilmente estacionario es, a su vez, estrictamente estacionario porque todos los momentos superiores en un proceso gaussiano son funciones de los dos primeros, y al ser los dos primeros estacionarios, los demás también lo serán.

La apariencia visual de un proceso estacionario es más o menos la misma a lo largo del tiempo. Por ejemplo, ninguna de las series de las figuras 1.1 a 1.6 del capítulo anterior, es estacionaria, con excepción de la serie de la Figura 1.4. Aunque a primera vista la Figura 1.6 podría

parecer estacionaria, no lo es porque si tomamos intervalos cortos, n pequeño en (2.1), la media en los picos será diferente de la media en los valles, y éstos ocurren de manera predecible en el tiempo. En capítulos posteriores se estudian procesos no estacionarios y pruebas para determinar la estacionariedad.

Un proceso estacionario es un proceso que está en estado estable (*steady state*), donde si inicialmente hubo transientes, éstos ya desaparecieron. Por ejemplo, cuando un motor recién arranca no se encuentra en estado estable (puede haber movimientos erráticos por algunos segundos) pero cuando "calienta" y entra en operación normal o ralentí, entra en estado estable. El *estado estable* es un concepto de equilibrio y corresponde a la llamada solución particular en ecuaciones diferenciales.

Algunas formas de lograr estacionariedad

Gran parte de los métodos de series de tiempo se han desarrollado para series estacionarias, básicamente porque, al igual que la regresión clásica, éstas incluyen un término aleatorio ε_t de varianza constante, cuyo efecto podría desaparecer con el paso del tiempo si la serie fuera no estacionaria (por ejemplo, si tiene tendencia ascendente o descendente). Sin embargo, muchas series no son estacionarias, de manera que con frecuencia resulta necesario minimizar o remover las no estacionariedades para facilitar la aplicación de técnicas conocidas. A su vez, la no estacionariedad de una serie puede ser de naturaleza determinística, como una tendencia que sigue una ecuación exacta, o puede ser de naturaleza estocástica, como una tendencia cuya ecuación incluye un término aleatorio. Algunas recomendaciones para minimizar la no estacionariedad son:

- a.) Remover la tendencia estimando, por algún método, curvas como:

$$\begin{aligned}\hat{x}_t &= \hat{\alpha} + \hat{\beta} t \\ \hat{x}_t &= e^{\hat{\alpha} + \hat{\beta} t} \\ \hat{x}_t &= \hat{\alpha} + \hat{\beta} t + \hat{\gamma} t^2\end{aligned}\tag{2.3}$$

Y luego restando dichas curvas de la serie original x_t de manera que $x_t - \hat{x}_t$ sea estacionaria. Esto podría hacerse, por ejemplo, para las

figuras 1.1 y 1.3, donde la no estacionariedad parece determinística (una recta, una exponencial).

- (b) También se puede estimar la tendencia por medio de un método o filtro que suavice la serie eliminando las fluctuaciones de corto plazo. Existen muchos en la literatura, entre ellos los de suavizado exponencial, promedio aritmético móvil, Holt-Winters y el filtro de Hodrick-Prescott, reseñados al final de este capítulo.
- (c) Tomar primeras diferencias $x_t - x_{t-1}$ o tomar cambios porcentuales. La diferenciación podría ser indicada para la Figura 1.1 y los cambios porcentuales para la Figura 1.3 porque en esta última las primeras diferencias de todas maneras crecen con el tiempo y no serían estacionarias.
- (d) Si la no estacionariedad está en la varianza, se puede intentar atenuarla tomando logaritmos de los valores de la serie. Si la no estacionariedad está en la media, tomar logaritmos no la elimina, sólo la modifica; quizás la aplana un poco (véase Granger, 1989, p. 66).
- (e) Si la no estacionariedad es causada por la estacionalidad, se pueden emplear métodos para desestacionalizar (como el conocido programa X-11 y otros reseñados al final de este capítulo), o se puede incluir en el modelo utilizando variables *dummy* o rezagos de la variable dependiente a la frecuencia estacional.

Ruido blanco

La serie estocástica más aleatoria posible se denomina *ruido blanco*, y es fundamental para el modelaje de las demás series de tiempo. Formalmente se define el ruido blanco como un proceso estocástico independiente e idénticamente distribuido (*iid*), esto es, x_t es ruido blanco si:

$$x_t \sim iid(\mu, \sigma^2) \quad (2.4)$$

Donde usualmente, aunque no necesariamente, μ es cero. Nótese que en la definición anterior no se especifica la forma de la distribución, tan sólo que sea idéntica de muestreo a muestreo. Ni siquiera tiene que

ser simétrica; podría ser, por ejemplo, una exponencial. Esta definición implica que el ruido blanco es estacionario.

Una serie de tiempo generada por lanzamientos de un dado es una serie de ruido blanco ya que los lanzamientos son independientes unos de otros y su probabilidad no cambia. En el caso del dado cargado es posible que los lanzamientos todavía sean independientes e idénticamente distribuidos y la serie siga siendo ruido blanco, sólo que la distribución de probabilidad ya no es uniforme porque algunas caras del dado tendrán más posibilidad que otras. En el caso del jugador hábil, que pueda manipular la forma en que caigan los dados, es menos claro que la serie sea ruido blanco porque los lanzamientos podrían no ser repeticiones independientes del experimento.

El ruido blanco no es pronosticable a partir de su propio pasado (aunque podría serlo a partir de otras variables) porque, por definición, cada x_t es independiente de los demás; por lo tanto, es pronosticable de manera incondicional, lo cual de poco sirve porque la predicción incondicional es la media. Afortunadamente, la mayoría de las series económicas no son ruido blanco y pueden ser, al menos en algún grado, predichas a partir de su pasado.

Una definición alternativa a (2.4), aceptada por algunos autores, remplaza el requisito de independencia por el menos restrictivo de cero correlación. Obviamente, si hay independencia, hay cero correlación. Además, en el caso de ruido gaussiano, cero correlación implica independencia matemática.

Ergodicidad

Las propiedades de estacionariedad y ergodicidad tienen que ver con el problema de hacer inferencias de la realización al ensamblaje (algo así como de la muestra a la población). Cuando un proceso es tanto estacionario como ergódico, una sola realización finita, que es usualmente lo único que tenemos en la práctica para trabajar, permite inferir todo acerca de la ley de probabilidad del proceso; los momentos estadísticos que se calculan de esta muestra finita convergen hacia los que se calcularían de una muestra infinita (por la estacionariedad), los cuales, con probabilidad 1, son iguales a los momentos de la población (por la ergodicidad).

Si los promedios temporales no cambian de una realización a otra, el proceso es ergódico. Si se tiene un proceso estacionario, y dado que en la realidad sólo se puede observar parte de una realización, lo mejor es asumir el proceso como ergódico ya que no hay una forma operacional de mostrar la diferencia. Sobre este tema, véase Nerlove, Grether y Carvalho (1979, p. 26) o Priestley (1981, p. 340), para mayor información.

La ergodicidad permite usar información de una única realización para estimar promedios de ensamblajes (*ensemble averages*) utilizando como proxies los promedios de tiempo (*time averages*) a lo largo de esa única realización disponible. Es como romper una realización estacionaria en muchísimos segmentos que luego se ponen uno sobre otro y se usan para encontrar los parámetros μ , σ^2 , etc., que describen la distribución de probabilidad del ensamblaje.

Un sistema ergódico tiene una distribución única a la cual converge desde cualquier estado inicial; en otras palabras, su distribución conjunta es independiente de las condiciones iniciales. Por esta razón, la ergodicidad requiere que los valores del proceso suficientemente separados en el tiempo estén casi incorrelacionados, tales que, promediando la serie en el tiempo, se adicione continuamente nueva información útil al promedio (véase Granger y Newbold, 1986, p. 5). Esta condición es necesaria, pero no suficiente para la ergodicidad, y puede expresarse diciendo que la función de autocovarianza o autocorrelación, conceptos que veremos más adelante, tienden a cero a una tasa suficientemente rápida, a medida que aumenta la separación en el tiempo.

Quizás una explicación heurística pueda ayudar. Supongamos un proceso x_t cuya media es $\mu = Ex_t = 0$ (sólo por simplicidad, porque podría ser diferente de cero). En teoría cada realización que genere este proceso puede tener una media diferente de cero; pero, para que se cumpla $Ex_t = 0$ en el ensamblaje, el promedio de todas las medias debe ser cero. Sin embargo, si cada realización puede tener una media diferente de cero, el proceso no será ergódico, ya que la media estimada \bar{x} de la realización disponible nunca convergirá a la del ensamblaje.

Función de autocovarianza

Uno de los objetivos más importantes al analizar series de tiempo es tratar de establecer la relación estadística entre una observación en un tiempo t y otras observaciones cercanas, por ejemplo, en $t - 1$, en $t - 2$, etc. Si existe alguna relación estadística estable entonces podemos utilizar esta información para modelar el proceso o intentar predecir la serie. Una herramienta para este propósito es la función de autocovarianza o, su pariente cercana, la función de autocorrelación (ACF). La de autocovarianza es más utilizada en tratamientos teóricos mientras que la de autocorrelación es más utilizada en la práctica. En el siguiente capítulo veremos otro concepto, el de la función de autocorrelación parcial (PACF) que, junto con la ACF vista aquí, servirá para la especificación de los modelos.

Supongamos una serie de tiempo estacionaria x_t con media μ , y también la misma serie x_t rezagada τ periodos, esto es, $x_{t-\tau}$. Pensemos en aplicar la covarianza o correlación estadística a x_t y a $x_{t-\tau}$ como si fueran dos variables distintas.

Recordemos que la covarianza entre dos variables (x, y) se define como:

$$Cov(x, y) = E(x - \mu_x)(y - \mu_y) \quad (2.5)$$

De manera que aplicando el mismo concepto a x_t y a $x_{t-\tau}$, y recordando que ambas tienen la misma media por ser la serie estacionaria, obtenemos la función autocovarianza $R(\tau)$:

$$R(\tau) = E(x_t - \mu)(x_{t-\tau} - \mu) \quad \text{para } |\tau| = 0, 1, 2, \dots \quad (2.6)$$

Que, como la covarianza tradicional, mide la similitud entre los dos juegos de observaciones de las dos series. Encontramos que cuando $\tau = 0$ la autocovarianza es igual a la varianza:

$$R(0) = E(x_t - \mu)(x_t - \mu) = E(x_t - \mu)^2 = Var(x_t) \quad (2.7)$$

De manera que el $R(\tau)$ máximo será $R(0)$ porque cuando $\tau = 0$ se produce la mayor similitud entre las dos series x_t y $x_{t-\tau}$; de hecho, cuando $\tau = 0$ ambas series son iguales.

Función de autocorrelación

Recordemos que la correlación entre variables se define como:

$$\varphi_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}} \quad (2.8)$$

De manera que aplicando el mismo concepto a x_t y a $x_{t-\tau}$ obtenemos la ACF:

$$r(\tau) = \frac{R(\tau)}{R(0)} \quad \text{para } |\tau| = 0, 1, 2, \dots \quad (2.9)$$

A primera vista, y de acuerdo con (2.8), el denominador debería ser $\sqrt{Var(x_t)Var(x_{t-\tau})}$, pero en razón del supuesto de estacionariedad, tenemos que:

$$\sqrt{Var(x_t)Var(x_{t-\tau})} = \sqrt{Var(x_t)Var(x_t)} = Var(x_t) = R(0)$$

El máximo será entonces:

$$r(0) = \frac{R(0)}{R(0)} = 1 \quad (2.10)$$

En general, $R(\tau)$ y $r(\tau)$ tienden a cero a medida que $\tau \rightarrow \infty$; de otra forma el sistema no es ergódico. Se encuentra también que hay simetría alrededor de τ , de manera que se tiene siempre la propiedad:

$$\begin{aligned} R(\tau) &= R(-\tau) \\ r(\tau) &= r(-\tau) \end{aligned} \quad (2.11)$$

Estimación de $R(\tau)$

En la práctica hay dos formas comunes para estimar la autocovarianza $R(\tau)$ de una serie estacionaria de n observaciones y media \bar{x} . Una es usando la expresión:

$$\hat{R}(\tau) = \frac{1}{n} \sum_{t=1+|\tau|}^n (x_t - \bar{x})(x_{t-|\tau|} - \bar{x}) \quad (2.12)$$

Que es la más popular, y la otra es usando la expresión:

$$\hat{R}(\tau)^* = \frac{1}{n - |\tau|} \sum_{t=1+|\tau|}^n (x_t - \bar{x})(x_{t-|\tau|} - \bar{x}) \quad (2.13)$$

que es la menos popular.

La primera, llamada $\hat{R}(\tau)$, al usar n en el denominador en lugar de $n - |\tau|$, es sesgada ya que la sumatoria sólo tiene $n - |\tau|$ términos, pero en general algunos autores sostienen que su error cuadrado medio resulta menor que el de $\hat{R}(\tau)^*$. De hecho, $\hat{R}(\tau)^*$ también es sesgada, aunque menos que $\hat{R}(\tau)$, porque \bar{x} no es igual a μ ; sin embargo, el sesgo desaparece asintóticamente (véase Priestley, 1981, sección 5.3.3). Para n grande y $|\tau|$ pequeño ambas definiciones arrojan resultados similares, pero para $|\tau|$ muy grande $\hat{R}(\tau)^*$ es muy inestable.

Estimación de $r(\tau)$

También hay dos formas de estimar la autocorrelación $r(\tau)$ para una serie de n observaciones y media \bar{x} . La más popular es usando la expresión:

$$\hat{r}(\tau) = \frac{\hat{R}(\tau)}{\hat{R}(0)} \quad (2.14)$$

Y la menos popular es usando la expresión:

$$\hat{r}(\tau)^* = \frac{\hat{R}(\tau)^*}{\hat{R}(0)^*} \quad (2.15)$$

Mientras que $\hat{r}(\tau)$ cumple con $|\hat{r}(\tau)| \leq 1$ para todo τ , como es de esperar para cualquier coeficiente de correlación, la alternativa $\hat{r}(\tau)^*$ no necesariamente cumple esta condición (Priestley, 1981, p. 331).

Para la expresión más popular $\hat{r}(\tau)$ en (2.14) se puede encontrar (Priestley, 1981, p. 332) que, bajo el supuesto de ser x_t un proceso gaussiano de media cero, y para $\tau > 0$, $\tau + d > 0$ y n grande:

$$\begin{aligned} Cov[\hat{r}(\tau), \hat{r}(\tau + d)] &\doteq \frac{1}{n} \sum_{i=-\infty}^{\infty} [r(i)r(i+d) + r(i+\tau+d)r(i-\tau) \\ &\quad + 2r(\tau)r(\tau+d)r^2(i) - 2r(\tau)r(i)r(i-\tau-d) \\ &\quad - 2r(\tau+d)r(i)r(i-\tau)] \end{aligned} \quad (2.16)$$

Donde el símbolo \doteq quiere decir aproximadamente.

En general, habrá alta correlación entre los $\hat{r}(\tau)$; si el $r(\tau)$ teórico decrece rápido, el $\hat{r}(\tau)$ estimado lo hará menos rápido (véase Priestley, 1981, p. 333). Haciendo $d = 0$ para $\tau > 0$ en (2.16) tenemos la varianza de $\hat{r}(\tau)$:

$$\begin{aligned} Var[\hat{r}(\tau)] \doteq \frac{1}{n} \sum_{i=-\infty}^{\infty} [r^2(i) + r(i+\tau)r(i-\tau) + 2r^2(\tau)r^2(i) \\ - 4r(\tau)r(i)r(i-\tau)] \end{aligned} \quad (2.17)$$

Si, como hipótesis nula, suponemos que el proceso es ruido blanco, entonces sabemos que $r(\tau) = 0$ para $\tau > 0$ porque cada observación es independiente de las otras y no hay correlación; y de (2.17) $Var[\hat{r}(\tau)] \doteq \frac{1}{n}$ asintóticamente. Se puede mostrar que, para ruido blanco gaussiano, $\hat{r}(\tau)$ converge en distribución a una normal con media cero y varianza $1/n$:

$$\hat{r}(\tau) \xrightarrow{d} N(0, 1/n) \quad (2.18)$$

El resultado (2.18) es muy útil para probar la significancia de las autocorrelaciones estimadas. A fin de establecer la intuición del resultado (2.18), es necesario recordar el teorema del límite central que dice que para una variable $y_t \sim (\mu_y, \sigma_y^2)$:

$$\frac{1}{n} \sum y_t \xrightarrow{d} N\left(\mu_y, \frac{\sigma_y^2}{n}\right) \quad (2.19)$$

Pero lo que aquí tenemos en lugar de $\frac{\sum y_t}{n}$ en (2.19) es:

$$\hat{r}(\tau) = \frac{\hat{R}(\tau)}{\hat{R}(0)} = \frac{\hat{R}(\tau)}{\sigma_x^2} = \frac{1}{n} \sum \frac{x_t x_{t-\tau}}{\sigma_x^2}$$

donde $x_t \sim iid N(0, \sigma_x^2)$, de manera que la suma es sobre $\frac{x_t x_{t-\tau}}{\sigma_x^2}$ como variable. Aplicando el teorema del límite central:

$$\frac{1}{n} \sum \frac{x_t x_{t-\tau}}{\sigma_x^2} \xrightarrow{d} N\left[E\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right), \frac{1}{n} Var\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right)\right] \quad (2.20)$$

Hay que encontrar $E\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right)$ y $Var\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right)$:

$$E\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right) = \frac{Ex_t Ex_{t-\tau}}{\sigma_x^2} = 0$$

porque x_t y $x_{t-\tau}$ son independientes (por ser x_t ruido blanco) y tienen media cero, y:

$$\begin{aligned} \text{Var}\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right) &= E\left(\frac{x_t x_{t-\tau}}{\sigma_x^2}\right)^2 = \frac{1}{\sigma_x^4} E(x_t x_{t-\tau})^2 \\ &= \frac{1}{\sigma_x^4} E x_t^2 E x_{t-\tau}^2 = \frac{\sigma_x^2 \sigma_x^2}{\sigma_x^4} = 1 \end{aligned}$$

De nuevo porque x_t y $x_{t-\tau}$ son independientes. Reemplazando en (2.20) obtenemos (2.18):

Otra forma de ver que $\text{Var}[\hat{r}(\tau)] = 1/n$ para $\tau > 0$, es notando que sólo los $r(0) = 1$ quedarán en (2.17) ya que las demás autocorrelaciones valen cero por ser x_t ruido blanco. De (2.17) tendríamos:

$$\text{Var}[\hat{r}(\tau)] = \frac{1}{n} [r^2(0) + 0 + 0 + 0] = \frac{1}{n} [1 + 0 + 0 + 0] = \frac{1}{n}$$

Donde el primer término del paréntesis es uno, porque corresponde a $i = 0$; mientras los siguientes tres valen cero, porque estamos asumiendo $\tau > 0$. Obviamente para $\tau = 0$ la autocorrelación es una constante igual a uno por definición y, en consecuencia, su varianza será cero.

Función de autocorrelación en la práctica

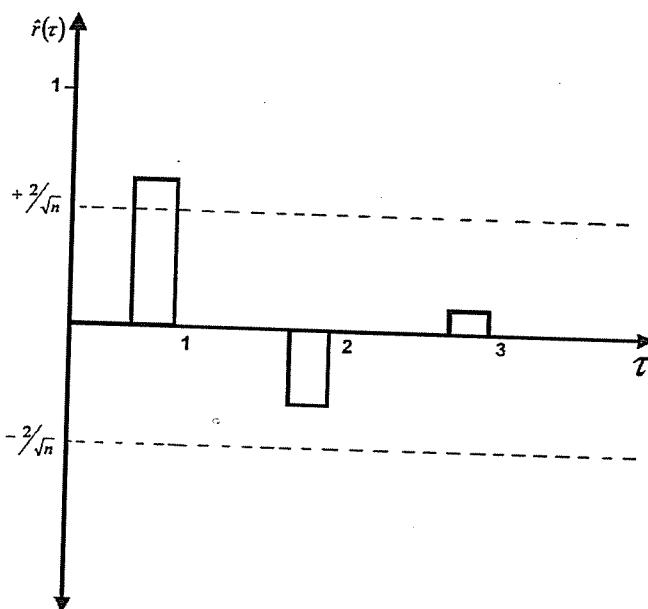
La ACF estimada $\hat{r}(\tau)$ se conoce comúnmente como *correlograma*. En un eje se grafica $\hat{r}(\tau)$ y en el otro τ . Por costumbre:

1. Sólo se muestra el eje positivo de τ debido a la propiedad de simetría [véase (2.11)].
2. Se omite el dato para $\tau = 0$ ya que se sabe que siempre vale uno [véase (2.10)].
3. Se ponen líneas punteadas a ± 2 desviaciones estándar de $\hat{r}(\tau)$, esto es, en $\pm \left(\frac{2}{\sqrt{n}}\right)$, bajo la hipótesis nula de que la serie es ruido blanco (véase (2.18) y la Figura 2.1).

Obviamente, por error estadístico y porque en la práctica trabajamos con muestras finitas, la ACF estimada será un poco diferente de

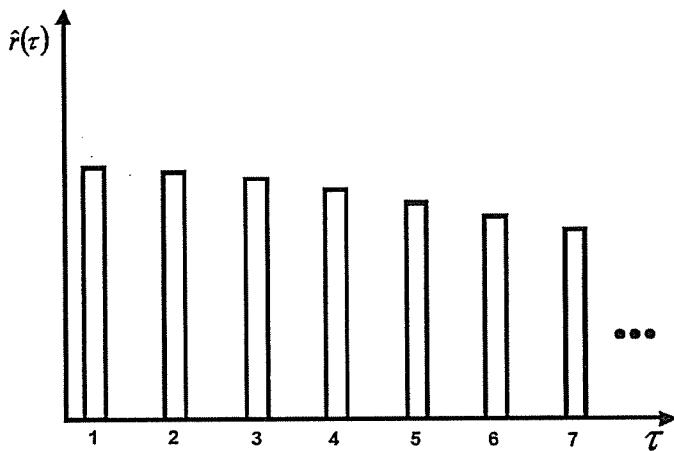
la teórica. Por ejemplo, para ruido blanco la ACF teórica tiene todos los $\hat{r}(\tau)$ iguales a cero, excepto, claro, el de $\tau = 0$; pero en la práctica, las autocorrelaciones estimadas para $\tau \neq 0$ tendrán algún valor, positivo o negativo, aunque usualmente no significativo. Sin embargo, debemos subrayar que una de cada 20 autocorrelaciones estimadas (5 %) podría resultar estadísticamente significativa (salirse de las líneas punteadas a ± 2 desviaciones estándar) a pesar de ser la serie ruido blanco. En cualquier caso, tienen mayor significancia económica las autocorrelaciones que se salgan del intervalo para τ pequeños. Algunos ejemplos de series y sus autocorrelaciones se encuentran en Montenegro (1989b). Una serie que con frecuencia se acerca a ruido blanco es la de cambios porcentuales del precio de las acciones en las bolsas; por eso son tan difíciles de predecir.

Figura 2.1: Ejemplo de ACF estimada



¿Qué pasa si x_t tiene una tendencia y calculamos su ACF? Que la mayoría de los productos $x_t x_{t-\tau}$, para un amplio rango de τ , tienen signo positivo resultando en una ACF que decae lentamente y aporta poca información (véase Figura 2.2). Esto sólo nos dice lo obvio, que hay mucha inercia en la serie.

Figura 2.2: ACF de una serie con tendencia



Cuando hay estacionalidad, como en la Figura 1.6 del capítulo 1, aparecen autocorrelaciones significativas en las frecuencias (rezagos) estacionales.

Las observaciones atípicas (*outliers*) hacen que la serie se vea más como ruido, por lo tanto hacen que $\hat{r}(\tau)$ sea menor. (La gráfica de x_t vs. $x_{t-\tau}$ también va a contener dos puntos extremos por cada *outlier*). Aunque una observación atípica deje todas las demás observaciones al otro lado de la media y a primera vista parezca haber mayor autocorrelación, $\hat{R}(0)$ sube más que proporcionalmente y hace que, en general, $\hat{r}(\tau)$ caiga.

Ejemplo

Encuentre la ACF estimada para la serie $\{2, 3, 1, 0, 1, -1\}$ para $\tau = 0, 1, 2, 3$. La media de los $n = 6$ datos de la serie es $\bar{x} = 1$. Al restar la media a los datos de la serie obtenemos una serie con media cero: $x_t - \bar{x} = \{1, 2, 0, -1, 0, -2\}$. Utilizando la expresión (2.12) estimamos las autocovarianzas:

$$R(0) = \frac{1}{6} \sum (x_t - \bar{x})^2 = \frac{1}{6} (1 + 4 + 0 + 1 + 0 + 4) = \frac{10}{6} = 1.666$$

$$\begin{aligned} R(1) &= \frac{1}{6} \sum (x_t - \bar{x})(x_{t-1} - \bar{x}) = \frac{1}{6} [(1)(2) + (2)(0) + (0)(-1) + (-1)(0) \\ &\quad + (0)(-2)] = \frac{2}{6} = 0.333 \end{aligned}$$

$$R(2) = \sum (x_t - \bar{x})(x_{t-2} - \bar{x}) = \frac{1}{6} [(1)(0) + (2)(-1) + (0)(0) + (-1)(-2)] = 0$$

$$R(3) = \sum (x_t - \bar{x})(x_{t-3} - \bar{x}) = \frac{1}{6} [(1)(-1) + (2)(0) + (0)(-2)] = -\frac{1}{6} = -0.166$$

Y dividiendo por $R(0)$, como en (2.14), obtenemos las autocorrelaciones:

$$r(0) = \frac{R(0)}{R(0)} = 1 \quad r(1) = \frac{R(1)}{R(0)} = 0.2$$

$$r(2) = \frac{R(2)}{R(0)} = 0 \quad r(3) = \frac{R(3)}{R(0)} = -0.1$$

Función de correlación cruzada

Si en lugar de rezagar la misma serie, como se hizo para derivar la ACF, utilizamos dos series distintas y rezagamos una con respecto a la otra, obtenemos la covarianza cruzada cuya expresión es:

$$R_{xy}(\tau) = E(x_t - \mu_x)(y_{t-\tau} - \mu_y) \quad (2.21)$$

De manera similar obtenemos la correlación cruzada:

$$r_{xy}(\tau) = \frac{R_{xy}(0)}{\sqrt{R_x(0)R_y(0)}} \quad (2.22)$$

Donde $R_x(0)$ y $R_y(0)$ son las varianzas (o autocovarianzas para rezago cero) de x y y .

Al igual que la ACF, la función de correlación cruzada también tiene el intervalo aproximado del 95 % en $\pm \frac{2}{\sqrt{n}}$, pero a diferencia de la ACF, no necesariamente tendrá su máximo en $\tau = 0$ ni será simétrica, aunque estará entre -1 y 1 .

Operador de rezago

Una herramienta útil en el análisis de series de tiempo es el operador de rezago, denotado L . Sirve para expresar ecuaciones en diferencia, que es la forma típica de los modelos para series de tiempo discretas. Se define la operación de rezago como

$$L^\tau x_t = x_{t-\tau} \quad (2.23)$$

Donde τ es el rezago que se desea introducir en la variable x_t . Por ejemplo:

$$\begin{aligned} Lx_t &= x_{t-1} \\ L^0 x_t &= x_t \\ L^3 x_t &= x_{t-3} \\ L^{-2} x_t &= x_{t+2} \\ (L^2)^3 &= L^6 \\ Lc &= c \end{aligned} \quad (2.24)$$

Donde c es una constante. Una propiedad muy útil del operador de rezago es:

$$\frac{1}{1 - \alpha L} = [1 + \alpha L + (\alpha L)^2 + (\alpha L)^3 + \dots] \quad \text{para } |\alpha| < 1 \quad (2.25)$$

La propiedad (2.25) corresponde a una expansión de Maclaurin (una serie de Taylor alrededor de cero). Esta igualdad se puede comprobar escribiendo:

$$[1 + \alpha L + (\alpha L)^2 + (\alpha L)^3 + \dots](1 - \alpha L) = 1$$

Y efectuando la multiplicación indicada para obtener:

$$1 - \alpha L + \alpha L - (\alpha L)^2 + (\alpha L)^2 - (\alpha L)^3 + \dots = 1$$

En la expresión anterior se puede ver que todos los términos se cancelen excepto el último que será de la forma $\alpha^n L^n$, el cual tiende a cero cuando n tiende a infinito siempre que $|\alpha| < 1$. Por lo tanto, ambos

lados de la ecuación son iguales ($1=1$) y se demuestra la propiedad. De manera similar se puede demostrar:

$$\frac{1}{1+\alpha L} = 1 - \alpha L + (\alpha L)^2 - (\alpha L)^3 + \dots \quad \text{para } |\alpha| < 1 \quad (2.26)$$

Donde, nótese, los signos alternan. Si $|\alpha| > 1$ entonces (2.25) se puede expresar así:

$$\frac{1}{1-\alpha L} = \frac{1}{-\alpha L \left(1 - \frac{1}{\alpha L}\right)} = (-\alpha L)^{-1} [1 + (\alpha L)^{-1} + (\alpha L)^{-2} + (\alpha L)^{-3} + \dots] \quad (2.27)$$

Tal expresión converge pero con términos futuros, como puede verse de (2.24), y que finalmente queda:

$$\frac{1}{1-\alpha L} = -(\alpha L)^{-1} - (\alpha L)^{-2} - (\alpha L)^{-3} - (\alpha L)^{-4} - \dots \quad (2.28)$$

De manera similar, cuando $|\alpha| > 1$ entonces (2.26) se puede expresar como:

$$\frac{1}{1+\alpha L} = -(\alpha L)^{-1} + (\alpha L)^{-2} - (\alpha L)^{-3} + (\alpha L)^{-4} - \dots \quad (2.29)$$

Ecuación de diferencias

Una ecuación de diferencias como:

$$y_t - \alpha y_{t-1} = f_t \quad (2.30)$$

Tiene una solución (despejar y_t en función de f_t) que es la suma de dos partes: la *homogénea* o complementaria (y_t^h) y la *particular* (y_t^p). En consecuencia, se puede escribir la solución completa como:

$$y_t = y_t^h + y_t^p \quad (2.31)$$

La homogénea debe satisfacer a (2.30) cuando $f_t = 0$. La solución es de la forma $A\alpha^t$, donde A es una constante, de manera que $A\alpha^t - \alpha A\alpha^{t-1} = 0$. Si $|\alpha| < 1$, la solución homogénea $A\alpha^t$ desaparece con el tiempo y sólo quedará la particular o estable. Si, por el contrario,

$|\alpha| > 1$, la solución homogénea crecerá con el tiempo, a menos que sea posible hacer $A = 0$.

Hay que anotar que la solución particular no necesariamente es estacionaria; porque f_t puede no ser estacionaria. En este sentido el *steady state* no necesariamente es un estado estadísticamente estacionario. La solución particular, que depende directamente de la función de excitación del sistema (*forcing function*) f_t , puede hallarse aplicando el operador de rezago a la expresión (2.30):

$$(1 - \alpha L)y_t = f_t \quad (2.32)$$

$$\begin{aligned} y_t &= \frac{f_t}{1 - \alpha L} = [1 + \alpha L + (\alpha L)^2 + (\alpha L)^3 + \dots] f_t \\ y_t^p &= f_t + \alpha f_{t-1} + \alpha^2 f_{t-2} + \alpha^3 f_{t-3} + \dots \end{aligned} \quad (2.33)$$

Donde ponemos la notación y_t^p para mayor claridad. La solución completa será la suma de la homogénea más la particular

$$y_t = y_t^h + y_t^p = A\alpha^t + f_t + \alpha f_{t-1} + \alpha^2 f_{t-2} + \alpha^3 f_{t-3} + \dots$$

El coeficiente A se despeja de las condiciones iniciales dadas (en $t = 0$); esto es, de $y_0 = A + f_0 + \alpha f_{-1} + \alpha^2 f_{-2} + \dots$.

¿Qué pasa si $|\alpha| > 1$? Que la solución particular tendrá una expresión convergente pero hacia el futuro, esto es, de (2.28)

$$\begin{aligned} y_t &= \frac{f_t}{1 - \alpha L} = [-(\alpha L)^{-1} - (\alpha L)^{-2} - (\alpha L)^{-3} - \dots] f_t \\ y_t^p &= -\frac{f_{t+1}}{\alpha} - \frac{f_{t+2}}{\alpha^2} - \frac{f_{t+3}}{\alpha^3} - \dots \end{aligned}$$

La solución completa será la suma de la homogénea más la particular

$$y_t = A\alpha^t - \frac{f_{t+1}}{\alpha} - \frac{f_{t+2}}{\alpha^2} - \frac{f_{t+3}}{\alpha^3} - \dots \quad (2.34)$$

Si bien la solución particular en este caso es convergente, la solución homogénea no lo es porque $|\alpha| > 1$, a menos que la constante A sea igual a cero, lo cual dependerá de las condiciones iniciales. Si y_0 es el valor inicial de y_t entonces en $t = 0$ tendremos:

$$y_0 = A - \frac{f_{+1}}{\alpha} - \frac{f_{+2}}{\alpha^2} - \frac{f_{+3}}{\alpha^3} - \dots$$

De donde vemos que para $A = 0$ sería necesario que:

$$y_0 = -\frac{f_{+1}}{\alpha} - \frac{f_{+2}}{\alpha^2} - \frac{f_{+3}}{\alpha^3} - \dots$$

Si lo anterior se cumple entonces el valor inicial y_0 neutraliza o cancela el efecto del comportamiento futuro de f_t y la solución completa (2.34) será igual a la particular. Este tipo de solución puede darse, por ejemplo, en expectativas racionales con previsión perfecta; pero es menos probable en series de tiempo donde la función f_t suele ser ruido blanco.

Notamos que el coeficiente α , del cual depende si la solución es convergente hacia atrás (*backward looking*) o hacia adelante (*forward looking*), es el inverso de la raíz de la expresión auxiliar $1 - \alpha L$ en (2.32), que es un polinomio en L . La raíz es α^{-1} . Entonces, si esta raíz es > 1 en magnitud, la serie es convergente ya que el inverso será < 1 .

Este importante resultado sobre las raíces y la convergencia es válido, en general, para ecuaciones con un mayor número de rezagos, digamos p rezagos, como en:

$$y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} - \dots - \alpha_p y_{t-p} = f_t \quad (2.35)$$

Aplicamos el operador de rezago y escribimos:

$$(1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p) y_t = f_t \quad (2.36)$$

De donde, despejando, obtenemos:

$$y_t = \frac{f_t}{1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p} \quad (2.37)$$

El denominador del lado derecho de (2.37) es un polinomio en L con máximo p raíces distintas. Si lo factorizamos y llamamos λ_i al inverso de la raíz i -ésima, podemos escribir (excepto por un factor de escala):

$$y_t = \frac{f_t}{(1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)} \quad (2.38)$$

Donde cada término de la forma $1/(1 - \lambda_i L)$ da origen a una expansión como en (2.25), siempre y cuando todos los $|\lambda_i| < 1$. Basta

con que uno solo de los factores incumpla este requisito para que la expansión total resultante sea explosiva, esto es, para que y_t no sea estable y no converja en términos del pasado. Si escribimos la ecuación auxiliar como:

$$1 - \alpha_1 L - \alpha_2 L^2 - \dots - \alpha_p L^p = 0 \quad (2.39)$$

Entonces podremos decir que y_t será estable y convergente en términos del pasado, si todas las raíces (o sea los inversos de las λ_i) de la ecuación auxiliar (2.39) son mayores que 1 en magnitud. En cuanto a la solución homogénea de la ecuación (2.35), ésta será de la forma:

$$y_t = A_1 \lambda_1^t + A_2 \lambda_2^t + \dots + A_p \lambda_p^t$$

Donde, como se anotó antes, los λ_i son los inversos de las raíces de la ecuación auxiliar (2.39).

Además del operador de rezago hay otras formas de encontrar la solución particular de las ecuaciones en diferencias. Una es por el método recursivo, que ilustraremos con la ecuación (2.30). Se trata de hacer reemplazos sucesivos.

$$\begin{aligned} y_t &= \alpha y_{t-1} + f_t \\ y_t &= \alpha [\alpha y_{t-2} + f_{t-1}] + f_t = \alpha^2 y_{t-2} + \alpha f_{t-1} + f_t \\ y_t &= \alpha [\alpha [\alpha y_{t-3} + f_{t-2}] + f_{t-1}] + f_t = \alpha^3 y_{t-3} + \alpha^2 f_{t-2} + \alpha f_{t-1} + f_t \\ &\vdots \\ y_t &= \alpha^n y_{t-n} + \alpha^{n-1} f_{t-(n-1)} + \dots + \alpha^2 f_{t-2} + \alpha f_{t-1} + f_t \end{aligned}$$

Donde notamos que los primeros términos tienden a cero si $|\alpha| < 1$, de manera que, reordenando, podemos escribir:

$$y_t = f_t + \alpha f_{t-1} + \alpha^2 f_{t-2} + \alpha^3 f_{t-3} + \dots$$

que es igual a (2.33).

Ejemplo

Para encontrar la solución particular convergente de:

$$y_t - .2y_{t-1} = f_t,$$

procedemos como sigue:

$$(1 - .2L)y_t = f_t$$

$$\begin{aligned} y_t &= \frac{f_t}{(1 - .2L)} = (1 + .2L + 0.04L^2 + 0.008L^3 + \dots)f_t \\ &= f_t + .2f_{t-1} + .04f_{t-2} + .008f_{t-3} + \dots \end{aligned}$$

Ejemplo

Para encontrar la solución particular convergente de:

$$0.2y_t - y_{t-1} = f_t,$$

procedemos como sigue:

$$(0.2 - L)y_t = f_t$$

$$\begin{aligned} y_t &= \frac{f_t}{(0.2 - L)} = \frac{f_t}{-L(1 - 0.2L^{-1})} \\ &= -f_{t-1} [1 + 0.2L^{-1} + (0.2L^{-1})^2 + (0.2L^{-1})^3 + \dots] \\ &= -f_{t+1} - 0.2f_{t+2} - 0.04f_{t+3} - 0.008f_{t+4} + \dots \end{aligned}$$

Ejemplo

Para polinomios de grado mayor se procede de modo similar, por ejemplo:

$$6y_t + 5y_{t-1} + y_{t-2} = f_t$$

$$(6 + 5L + L^2)y_t = f_t$$

$$(L + 2)(L + 3)y_t = f_t$$

$$y_t = \frac{f_t}{(L + 2)(L + 3)} = \frac{f_t}{6(1 + \frac{L}{2})(1 + \frac{L}{3})}$$

Y se expande cada uno de los factores por separado, primero uno y luego el otro. Como se anotó antes, la solución será convergente en términos del pasado si las raíces de la ecuación auxiliar son mayores que la unidad en magnitud, como en efecto sucede en este ejemplo donde las raíces son -2 y -3 .

Raíces imaginarias

Supongamos la ecuación de diferencias $y_t + 2y_{t-1} + 2y_{t-2} = f_t$. La ecuación auxiliar es $1 + 2L + 2L^2$ y sus raíces son:

$$\frac{-2 \pm \sqrt{4-8}}{4} = -\frac{1}{2} \pm \frac{\sqrt{-4}}{4} = -\frac{1}{2} \pm \frac{\sqrt{-1}}{2}$$

La cantidad $i = \sqrt{-1}$ se define como número imaginario. Las raíces $(-\frac{1}{2} + \frac{i}{2})$ y $(-\frac{1}{2} - \frac{i}{2})$, que incluyen este número, se conocen como raíces imaginarias. Las raíces imaginarias de ecuaciones auxiliares con coeficientes reales vienen siempre en *pares conjugados*, esto es, con signos contrarios en sus partes imaginarias. La factorización de la ecuación auxiliar queda:

$$1 + 2L + 2L^2 = 2 \left(L + \frac{1}{2} - \frac{i}{2} \right) \left(L + \frac{1}{2} + \frac{i}{2} \right)$$

En general, los números que contienen i , incluyendo las raíces del ejemplo anterior, se conocen como números imaginarios o complejos y son de la forma $a + bi$, donde a y b son constantes. En los números reales simplemente la parte imaginaria es cero, esto es, b es cero. Hay propiedades útiles, por ejemplo, que $i^2 = -1$, que $(a + bi)(a - bi) = a^2 + b^2$, y que el inverso de un número complejo también es un número complejo (esto puede mostrarse multiplicando arriba y abajo por el conjugado).

Cualquier número complejo $a + bi$ se puede representar sobre un par de ejes, uno real y uno imaginario, como se muestra en la Figura 2.3. Vemos que:

$$\begin{aligned} a &= m \cos \phi \\ b &= m \operatorname{sen} \phi \\ a \pm bi &= m [\cos \phi \pm i \operatorname{sen} \phi] = m e^{\pm i \phi} \end{aligned} \tag{2.40}$$

Donde $m = \sqrt{a^2 + b^2}$ es la magnitud, ϕ es el ángulo con la horizontal, y la expresión $e^{\pm i \phi} = \cos \phi \pm i \operatorname{sen} \phi$ se conoce como ecuación de Euler.

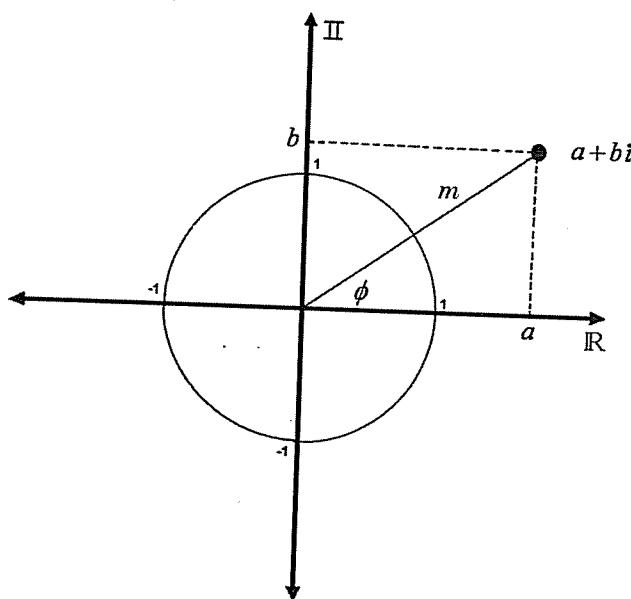
De nuevo, para tener convergencia en términos del pasado, las raíces de la ecuación auxiliar (2.39), sean reales o complejas, deben ser mayores que la unidad en magnitud, esto es, *las raíces deben estar*

por fuera del círculo unitario de la Figura 2.3. Por ejemplo, las raíces anteriores:

$$-\frac{1}{2} + \frac{i}{2} \quad \text{y} \quad -\frac{1}{2} - \frac{i}{2}$$

no están por fuera del círculo unitario y generarían un modelo inestable, ya que su magnitud $\sqrt{.5^2 + .5^2} = .5\sqrt{2} = .707$ es inferior a la unidad.

Figura 2.3: Plano complejo



En resumen, las ecuaciones en diferencias con dos o más rezagos pueden tener raíces reales o complejas. En todo caso, las complejas vienen en pares conjugados. Las ecuaciones en diferencias de un rezago sólo pueden tener una raíz real. Se deja como ejercicio al lector:

a. Mostrar que

$$\frac{1-i}{3+2i} = \frac{1}{13} - \frac{5}{13}i.$$

b. Mostrar que la expresión

$$0.2y_t + 0.8y_{t-1} + y_{t-2} = f_t$$

tiene raíces complejas y que se puede expresar como:

$$(L^2 + 0.8L + 0.2)y_t = (L + 0.4 + 0.2i)(L + 0.4 - 0.2i)y_t = f_t.$$

c. Mostrar que en la expansión de:

$$y_t = \frac{f_t}{(L + 0.4 + 0.2i)(L + 0.4 - 0.2i)}$$

desaparecen las partes imaginarias y quedan sólo coeficientes reales.

Operador de diferencia

Otro operador muy conveniente y muy utilizado para expresar series de tiempo es el operador de diferencia, Δx_t , definido como:

$$\Delta x_t = x_t - x_{t-1} = (1 - L)x_t \quad (2.41)$$

Éste consiste simplemente en tomar una observación y restarle la observación inmediatamente anterior. También puede aplicarse de manera repetitiva; esto es, puede aplicarse de nuevo a una serie que ya haya sido diferenciada.

Ejemplo

$$\begin{aligned}\Delta^2 x_t &= \Delta \Delta x_t = \Delta(x_t - x_{t-1}) = x_t - x_{t-1} - (x_{t-1} - x_{t-2}) \\ &= x_t - x_{t-1} - x_{t-1} + x_{t-2} = x_t - 2x_{t-1} + x_{t-2}\end{aligned}$$

Suavizado aritmético

En muchas aplicaciones es importante estimar la tendencia de una serie, ya sea porque ésta es de interés en sí misma o porque se propone convertir la serie original en estacionaria restándole la tendencia. Los métodos de suavizado remueven el ruido y las fluctuaciones de corto plazo de una serie y dejan sólo su movimiento de largo plazo (tendencia). Uno de tales métodos es el suavizado aritmético, también conocido como suavizado de promedio móvil. Es importante advertir que no deben confundirse con los modelos denotados MA(q), que llevan un

nombre similar, promedio móvil, y que serán tratados en el siguiente capítulo.

El objetivo del procedimiento es obtener una versión suavizada s_t de la serie original x_t . Hay varias alternativas para lograr esto. Por ejemplo, el promedio aritmético en un momento de tiempo t se puede tomar centrado, sobre $2m + 1$ términos, esto es, m hacia adelante de t , m hacia atrás y uno en t , como lo indica la siguiente expresión,

$$s_t = \frac{1}{2m+1} \sum_{i=-m}^m x_{t+i} \quad (2.42)$$

Donde también se puede utilizar x_{t-i} en lugar de x_{t+i} . En este método centrado se pierden m observaciones de s_t al comienzo de la serie original y m al final.

Otra alternativa es el promedio aritmético rezagado, cuya expresión es:

$$s_t = \frac{1}{m+1} \sum_{i=0}^m x_{t-i} \quad (2.43)$$

El cual se toma sobre los $m + 1$ últimos términos (incluido el del tiempo t) y se pierden m observaciones de s_t al principio de la serie. Esta es la versión más utilizada porque generalmente hay mayor interés en tener el dato suavizado más reciente.

También se puede tomar el promedio aritmético hacia adelante, en cuyo caso la expresión es:

$$s_t = \frac{1}{m+1} \sum_{i=0}^m x_{t+i} \quad (2.44)$$

El cual se toma sobre los $m + 1$ términos futuros (incluido el del tiempo t) y se pierden m observaciones de s_t al final de la serie.

En cualquiera de las expresiones anteriores, cuanto mayor sea m , más suave será la serie s_t resultante pero también más lenta será en reflejar (o en dejar de reflejar) los cambios de tendencia a lo largo de x_t . Si la serie x_t tiene tendencia ascendente (descendente), la s_t de un suavizado aritmético rezagado tiende a estar debajo (encima) de x_t . Lo contrario sucede con la serie s_t del suavizado aritmético adelantado.

También es posible hacer suavizado donde el peso dado a los términos de la suma sean diferentes, por ejemplo,

$$s_t = \frac{.5x_{t-1} + x_t + .5x_{t+1}}{2}$$

Suavizado exponencial de Holt-Winters

Este método utiliza ponderaciones que caen exponencialmente para motivar la expresión de la serie suavizada s_t :

$$s_t = Ax_t + (1 - A)Ax_{t-1} + (1 - A)^2Ax_{t-2} + (1 - A)^3Ax_{t-3} + \dots$$

A su vez, esta expresión puede escribirse como:

$$\begin{aligned} &= (1 + (1 - A)L + (1 - A)^2L^2 + (1 - A)^3L^3 + \dots)Ax_t \\ &= \frac{Ax_t}{1 - (1 - A)L} \end{aligned}$$

De donde:

$$s_t - (1 - A)s_{t-1} = Ax_t$$

Para finalmente obtener la expresión de s_t utilizada en el suavizado:

$$s_t = Ax_t + (1 - A)s_{t-1} \quad (2.45)$$

para $0 < A < 1$.

Como puede observarse, el estimado s_t se calcula en cada iteración como un promedio ponderado de x_t y s_{t-1} , algo parecido a un filtro de Kalman, sólo que en este último las ponderaciones también se ajustarían en cada iteración. Cuanto más bajo sea A , más peso se le da al pasado y más suave la serie s_t ; una medida usual de A es cerca de 0.3. Las iteraciones pueden iniciarse haciendo $s_1 = x_1$ y empezando en $t = 2$. El pronóstico de la serie x_t para T unidades de tiempo en el futuro, denotado \hat{x}_{t+T} , será:

$$\hat{x}_{t+T} = s_t \quad (2.46)$$

Para cualquier $T > 0$.

El método anterior puede complementarse de varias maneras. Una de ellas es con la inclusión de la tendencia. Las expresiones quedan:

$$\begin{aligned}s_t &= Ax_t + (1 - A)(s_{t-1} + \Gamma_{t-1}) && \text{para } 0 < A < 1 \\ \Gamma_t &= C(s_t - s_{t-1}) + (1 - C)\Gamma_{t-1} && \text{para } 0 < C < 1\end{aligned}\quad (2.47)$$

Donde Γ_t es el incremento en la tendencia. La iteración puede inicializarse en $t = 3$ con $\Gamma_2 = x_2 - x_1$, $s_2 = x_2$, $s_1 = x_1$. El pronóstico, para todo $T > 0$, será:

$$\hat{x}_{t+T} = s_t + T\Gamma_t \quad (2.48)$$

En cuanto más bajos sean los coeficientes A y C , más estable será la serie s_t , ya que menores valores implican que se otorga más peso al pasado. También existe una versión que incluye factores estacionales.

Filtro de Hodrick-Prescott

Recientemente se ha popularizado el filtro de Hodrick-Prescott para estimar la tendencia de una serie basado en un procedimiento que había sido usado por décadas en las ciencias actuariales (véase Hodrick y Prescott, 1997). El filtro calcula una versión suavizada de la serie, llamada s_t , a partir de la serie original x_t y escoge la secuencia $\{s_t\}$ que combine dos objetivos contradictorios. Por un lado, busca minimizar la varianza (fluctuación) de x_t alrededor de s_t y, por el otro, hace una restricción sobre la segunda derivada de s_t para buscar que el movimiento anterior sea lo más parecido posible al siguiente. En otras palabras, este método de suavizado escoge la serie $\{s_t\}$ que minimiza la siguiente función objetivo:

$$\sum_{t=1}^n (x_t - s_t)^2 + \lambda \sum_{t=2}^{n-1} [(s_{t+1} - s_t) - (s_t - s_{t-1})]^2 \quad (2.49)$$

Donde el parámetro $\lambda > 0$ controla el grado de suavizado; cuanto mayor sea λ , mayor será el suavizado porque le da más peso al hecho de que el movimiento anterior de s_t se parezca al siguiente. Por ejemplo, cuando λ es grande, será deseable que los términos $s_{t+1} - s_t$ y $s_t - s_{t-1}$ se parezcan, a fin de que el segundo sumando en la expresión del

filtro sea pequeño y que la solución s_t , también para minimizar el primer sumando, se acerque a la tendencia lineal $s_t = \alpha + \beta t$ de la serie. Al otro extremo, cuando λ es pequeño, la solución se acercará a $s_t = x_t$. La derivación del filtro genera la expresión (2.49). Para detalles matemáticos de la optimización y condiciones de primer orden, véase Mills (2003, capítulo 4).

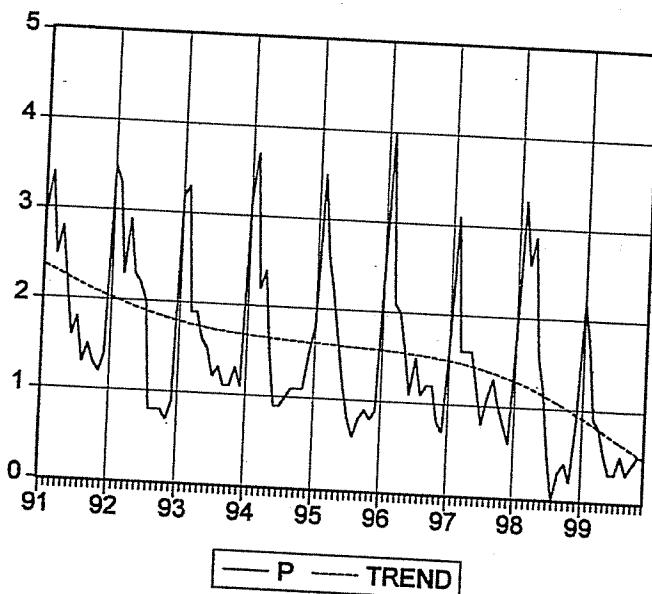
Hodrick y Prescott (1997) utilizan el filtro para descomponer series macroeconómicas x_t en su tendencia de largo plazo (que ellos llaman componente de crecimiento) s_t , y en sus fluctuaciones de corto plazo (que llaman componente cíclico) $x_t - s_t$. Los autores utilizaron un valor λ de 1600 para las series trimestrales analizadas. Para series con otras periodicidades recomiendan valores diferentes, por ejemplo, 100 para anuales y 14400 para mensuales. La Figura 2.4 grafica la serie del cambio mensual en el índice de precios junto con la tendencia estimada por el filtro de Hodrick-Prescott del programa Eviews. Nótese que, a diferencia de otros métodos de suavizado, el filtro de Hodrick-Prescott produce una serie s_t con el mismo número de datos que tiene la serie x_t .

Existen muchos otros filtros para el suavizado o extracción de parte de la señal. Por ejemplo, en contraste con el de Hodrick-Prescott, que es un filtro pasabajos, el filtro de Baxter-King es un pasabanda. Esto quiere decir que deja pasar las frecuencias contenidas en un intervalo de frecuencias, mínima y una máxima, que el usuario puede especificar. El filtro se ha popularizado para extraer las frecuencias asociadas con los ciclos económicos, usualmente consideradas en el rango de año y medio a ocho años. Los filtros prácticos son aproximaciones del filtro ideal (que tiene forma de ventana rectangular), por lo cual pueden distorsionar la señal e, incluso, generar ciclos donde no los hay (véase Woitek, 1998).

Ajuste estacional

Una serie estacional presenta movimientos similares cada año aproximadamente por la misma época. El efecto estacional puede ser causado por variaciones climáticas (temporada de lluvia, sequías y heladas), por normas tributarias (cierres contables, pago de impuestos), por días de fiesta (Navidad, Semana Santa), costumbres sociales (calendario escolar, vacaciones), etc. Con frecuencia nos interesa saber qué

Figura 2.4: Variación porcentual mensual de IPC (P): tendencia calculada por el filtro de Hodrick-Prescott (TREND)



parte del movimiento de una serie se debe al componente estacional y qué parte a la tendencia subyacente. Queremos saber si un movimiento es típico o si realmente representa un aumento o disminución importante en la variable. Los procedimientos de desestacionalización tienen por objeto remover, aislar o repartir el componente estacional de la serie. Para una referencia general sobre el tema, véase Judge *et ál.* (1985) o Pierce (1980).

El movimiento individual de un mes, un trimestre, etc., de una serie estacional no puede ponderarse para obtener un estimativo del total anual porque cada mes o trimestre tiene un comportamiento diferente; por ejemplo, si el primer trimestre suele ser más alto que los demás, no sería acertado multiplicar el resultado del primer trimestre por cuatro para proyectar el año. Sin embargo, esto sí se puede hacer en las series desestacionalizadas.

Una serie estacional muestra barras significativas en las frecuencias estacionales en su ACF; por ejemplo, si la serie estacional es trimestral mostrará autocorrelaciones significativas para rezagos 4, 8, 12, etc., y

si es mensual mostrará autocorrelaciones significativas para rezagos 12, 24, 36, etc. Es importante tener en cuenta que remover el efecto estacional no necesariamente remueve toda la autocorrelación en las frecuencias estacionales, ya que parte de esta autocorrelación puede provenir de factores diferentes a los estacionales. Esto se ve con mayor claridad si imaginamos una serie x_t como la suma del efecto estacional más un proceso estocástico que puede tener autocorrelaciones en varios rezagos incluyendo el estacional; así, remover el efecto estacional no elimina toda la autocorrelación en ese rezago.

Hay dos supuestos usuales acerca de la forma como se incorpora el componente estacional a la serie: o de manera *multiplicativa* o *aditiva*. El caso multiplicativo es indicado cuando el efecto estacional varía en proporción al nivel de la serie, mientras que el caso aditivo es indicado cuando el efecto estacional es una cantidad fija que no varía con el nivel de la serie. Por ejemplo, un efecto estacional igual al 4% del nivel de la serie sería multiplicativo, pero un efecto fijo de 5 mil unidades sería aditivo. En la práctica, el efecto estacional rara vez será constante, ya sea en porcentaje o unidades, aunque puede argüirse que el modelo multiplicativo es más compatible con la teoría y el mecanismo generador de datos.

Para ilustrar el procedimiento de desestacionalización, consideremos una serie x_t . El primer paso es obtener una tendencia s_t de la serie que se desea desestacionalizar. La forma preferida de hacer esto es a través de un promedio aritmético centrado como en (2.42), con suficientes términos para cubrir un periodo anual; por ejemplo, con $m = 2$ para cada lado si la serie es trimestral y $m = 6$ para cada lado si la serie es mensual, donde los pesos de los términos extremos son la mitad de los demás para evitar sumar dos veces el efecto de un mismo periodo, por ejemplo:

$$s_t = (.5x_{t-6} + x_{t-5} + \dots + x_t + \dots + x_{t+5} + .5x_{t+6})/12,$$

si la serie es mensual y

$$s_t = (.5x_{t-2} + x_{t-1} + x_t + x_{t+1} + .5x_{t+2})/4,$$

si la serie es trimestral.

Una vez obtenida la serie suavizada s_t , se calculan las diferencias $x_t - s_t$ en el caso aditivo, o los cocientes x_t/s_t si el caso es multiplicativo.

Luego se promedian estas diferencias o cocientes, según sea el caso, para cada mes o trimestre. Esto es, si la serie es mensual se promedian todos los eneros, todos los febreros, etc. Si la serie es trimestral entonces se promedian de todos los primeros trimestres, los segundos trimestres, etc. Estos promedios, después de alguna normalización, se conocen como factores estacionales o de escala, f_i , donde el subíndice i indica el mes o trimestre en cuestión. Los factores f_i representan los movimientos estacionales típicos de cada mes o trimestre.

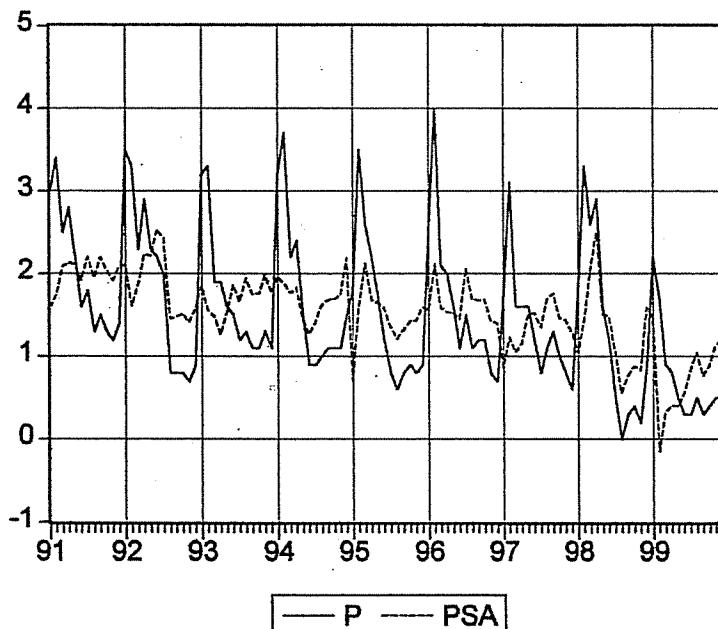
En el caso aditivo, la serie desestacionalizada se obtiene restando estos factores de la serie original para cada mes o trimestre i , y en el caso multiplicativo, dividiendo la serie original por el factor del mes o trimestre correspondiente. Subrayamos que la desestacionalización no debe cambiar el efecto neto anual de la serie; por esta razón, los factores estacionales se normalizan para que tengan media cero en el caso aditivo, o producto unitario en el caso multiplicativo. Esto se hace restando la media de los factores a cada uno de los factores en el caso aditivo, o dividiendo cada factor por el promedio geométrico de los factores en el caso multiplicativo.

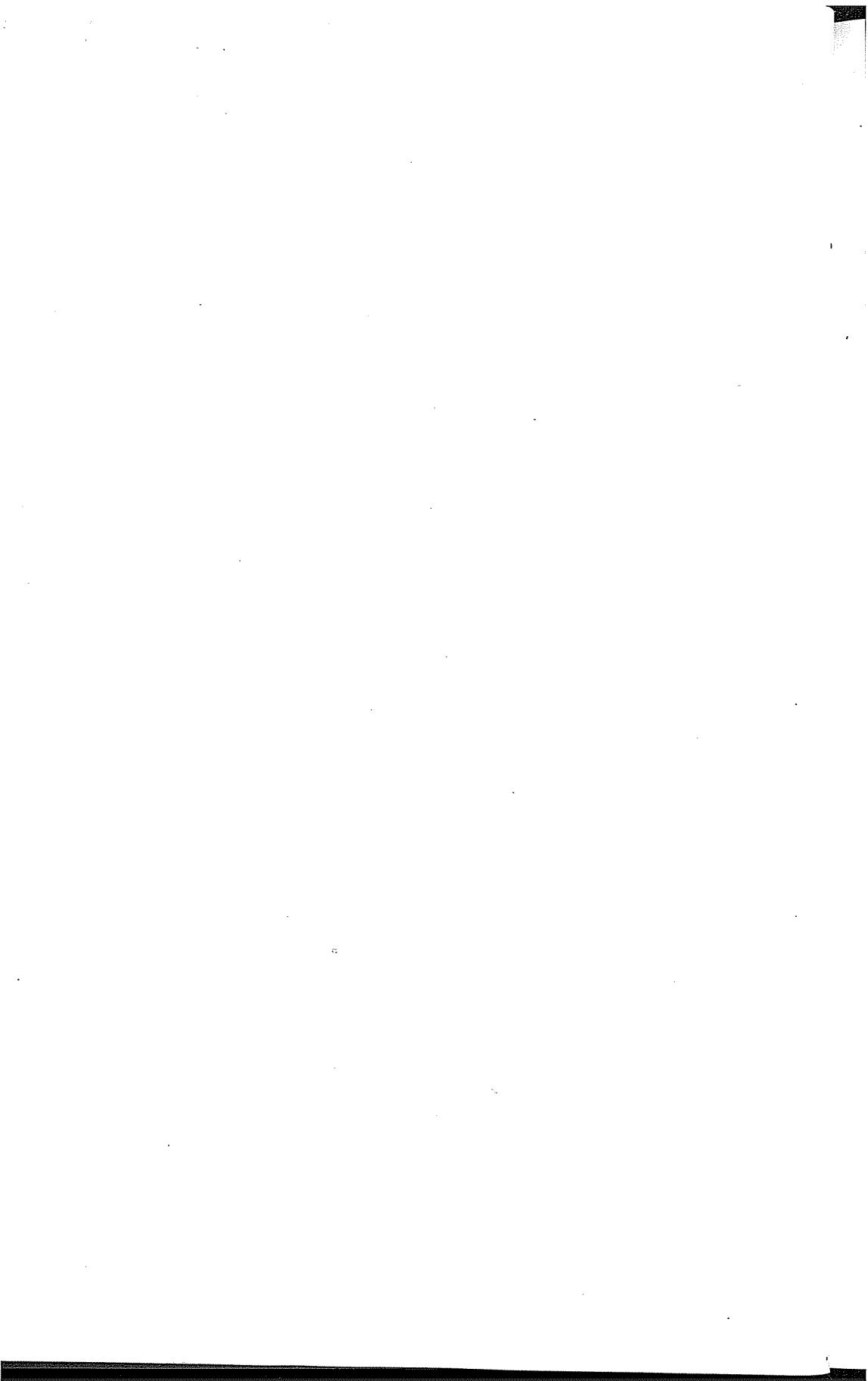
El procedimiento multiplicativo tiene problemas en series que toman valores de cero o negativos. Como ilustración, supongamos una serie mensual donde las cifras de enero suelen ser positivas y 5% más altas que el promedio. Para desestacionalizar, dividimos todas las observaciones de enero por 1.05; así se bajan todas de nivel y se remueve el efecto estacional. Pero si apareciera alguna observación negativa en enero, al dividirla por 1.05 se movería en sentido contrario, con lo cual se subiría de nivel y haría parecer la cifra desestacionalizada menos negativa de lo que en realidad fue. O si algún enero es cero, no se movería al dividir por 1.05. En contraste, en el caso aditivo, siempre restaríamos la misma cantidad a las cifras de enero y se bajarían todas de nivel, sin importar si son positivas, negativas o cero. Por otro lado, si la serie tiene observaciones iguales a cero o negativas podría suceder que la expresión x_t/s_t arroje un factor f_i indefinido con el cual sería imposible dividir la serie original para desestacionalizar.

El procedimiento de mayor difusión actual para realizar ajuste estacional es el programa X-11, desarrollado por el Census Bureau del Departamento de Comercio norteamericano. Otros programas que le compiten son el X-11-ARIMA (una variación del X-11), y el

TRAMO/SEATS popular especialmente en Europa. En general todos estos programas permiten que los factores estacionales o de escala cambien con el tiempo si la naturaleza de la serie así lo exige; también permiten dar menor ponderación a observaciones consideradas atípicas para el periodo respectivo. Pueden tomar en cuenta el efecto del número de días hábiles por mes, así como el efecto de los días de fiesta (por ejemplo, Semana Santa que cambia de un mes a otro o de un trimestre a otro), e incluyen varias pruebas estadísticas para juzgar la significancia de la estacionalidad estimada y su estabilidad. La Figura 2.5 muestra la serie mensual de cambios en el índice de precios junto con la misma serie ajustada estacionalmente por medio del procedimiento aditivo del programa X-11.2, cuya versión se encuentra en Eviews.

Figura 2.5: Variación porcentual del IPC (P): serie con ajuste estacional (PSA)





CAPÍTULO 3

Modelo autorregresivo y modelo de promedio móvil

Un hecho observado en muchos fenómenos temporales es que el comportamiento pasado provee información sobre el comportamiento futuro. El pasado de una serie suele incorporar, y de esa manera reemplazar, la información de otras variables que pueden intervenir en el proceso. La especificación relativamente sencilla, pero efectiva, que resulta de expresar una variable en función de su propio pasado es conocida como *autorregresiva* y es parte del modelaje de Box-Jenkins. Las dos especificaciones principales de esta metodología son los modelos autorregresivos AR(p) y los modelos de promedio móvil MA(q), diseñados para procesos estocásticos estacionarios univariados. Primero veamos el autorregresivo.

El modelo autorregresivo AR(p)

El modelo autorregresivo expresa el valor actual de una serie estacionaria x_t en función de su propio pasado, o sea de sus rezagos x_{t-1}, \dots, x_{t-p} . Un autorregresivo AR(p) tiene la siguiente forma algebraica:

$$x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} = \varepsilon_t + c \quad (3.1)$$

Donde c es una constante, p es el orden o el mayor rezago del

modelo, y ε_t es ruido blanco con media cero y varianza constante σ_ε^2 . En la literatura de series de tiempo, la secuencia $\{\varepsilon_t\}$ se conoce indistintamente como ruido, error, residuos, innovaciones o shocks. Otra presentación usual de (3.1) es con x_t despejada.

El AR(1)

Veamos primero el más sencillo de los modelos autorregresivos, el AR(1), de orden 1, que tiene un solo rezago y se escribe:

$$x_t = c + \alpha x_{t-1} + \varepsilon_t \quad (3.2)$$

O empleando el operador de rezago, como $(1 - \alpha L)x_t = c + \varepsilon_t$.

Este modelo podría representar, por ejemplo, los desempleados en el mes t como una proporción fija α de aquellos desempleados en el mes $t - 1$, mientras que la otra proporción $1 - \alpha$ consiguió empleo, más un nuevo grupo $c + \varepsilon_t$ busca trabajo.

La media del AR(1) se puede encontrar tomando expectativas de (3.2) y recordando que, por el supuesto de estacionariedad, $E x_t = E x_{t-1}$,

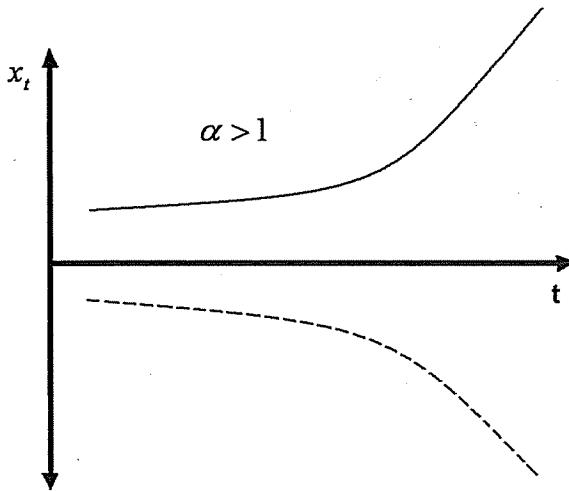
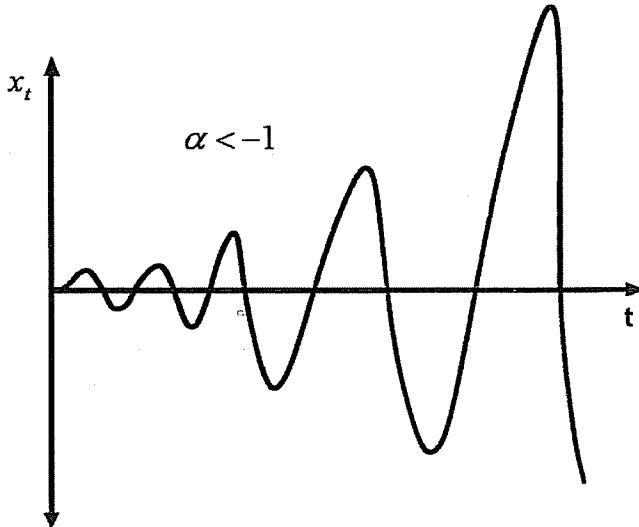
$$E x_t = E(c + \alpha x_{t-1} + \varepsilon_t) = c + \alpha E x_{t-1} + E \varepsilon_t \quad (3.3)$$

De donde, reorganizando, se tiene:

$$E x_t = \frac{c + E \varepsilon_t}{1 - \alpha} = \frac{c}{1 - \alpha} \quad (3.4)$$

porque hemos supuesto que $E \varepsilon_t = 0$.

Es importante observar que en (3.2) es necesario que $|\alpha| < 1$ para que x_t sea estable. Recordamos α que es el inverso de la raíz de la ecuación auxiliar $1 - \alpha L = 0$, y que esta raíz debe ser mayor que uno en magnitud para la estabilidad del sistema. Si, por el contrario, $|\alpha| > 1$ entonces la serie crece sin límite como se muestra en la Figura 3.1. En la Figura 3.1a, con $\alpha > 1$ dependiendo del valor de x_0 y de los ε_t iniciales, x_t seguirá la curva sólida o la línea punteada. En la Figura 3.1b, el coeficiente $\alpha < -1$ induce un comportamiento oscilatorio creciente en la serie.

Figura 3.1a: AR(1) con $\alpha > 1$ Figura 3.1b: AR(1) con $\alpha < -1$ 

Un modelo autorregresivo en tiempo discreto tiene su contraparte en tiempo continuo. Por ejemplo, el AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$ puede escribirse como:

$$x_t - x_{t-1} = \alpha x_{t-1} - x_{t-1} + \varepsilon_t$$

o, simplificando:

$$\Delta x_t = (\alpha - 1)x_{t-1} + \varepsilon_t$$

y, finalmente, puede escribirse como:

$$\frac{\Delta x_t}{\Delta t} = (\alpha - 1)x_{t-1} + \varepsilon_t$$

Esto si recordamos que en tiempo discreto $\Delta t = 1$. La expresión anterior es la versión discreta de la ecuación diferencial lineal continua de coeficientes constantes:

$$\frac{dx(t)}{dt} = (\alpha - 1)x(t) + \varepsilon(t)$$

cuya solución homogénea es:

$$h(t) = Ae^{(\alpha-1)t},$$

la cual vale A para $t = 0$ y después decae exponencialmente hacia cero (si $\alpha - 1 < 0$). En sentido estricto, no puede existir un proceso puramente aleatorio en tiempo continuo, $\varepsilon(t)$, ya que los cambios abruptos en periodos infinitesimales de tal proceso requerirían cero inercia o energía infinita.

Ahora calculemos la función de autocorrelación (ACF) para el modelo AR(1). A fin de simplificar el cálculo supondremos a continuación que la serie x_t tiene media cero, lo que, de (3.4), implica que $c = 0$. En caso contrario, debemos restar el valor de la media a cada x_t . Entonces, para obtener la autocovarianza primero multiplicamos la expresión de x_t por $x_{t-\tau}$ (donde $\tau > 0$) y luego tomamos expectativas:

$$R(\tau) = E(x_t x_{t-\tau}) = \alpha E(x_{t-1} x_{t-\tau}) + E(\varepsilon_t x_{t-\tau}) = \alpha E(x_{t-1} x_{t-\tau}) \quad (3.5)$$

Al reconocerse que el último término de (3.5) es otra autocovarianza, podemos escribir:

$$R(\tau) = \alpha R(\tau - 1) \quad (3.6)$$

Donde, a su vez, el término $R(\tau - 1)$ se puede expresar como otra autocovarianza, y así sucesivamente, de manera que:

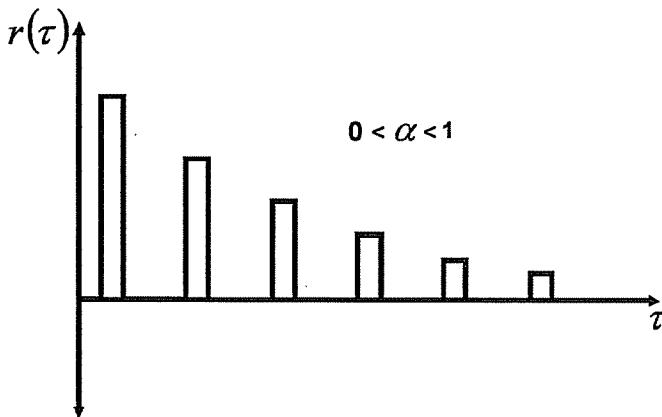
$$R(\tau) = \alpha R(\tau - 1) = \alpha [\alpha R(\tau - 2)] = \dots = \alpha^\tau R(0)$$

Si dividimos ambos lados por $R(0)$ obtenemos la autocorrelación deseada:

$$r(\tau) = \alpha r(\tau - 1) = \alpha [\alpha r(\tau - 2)] = \dots = \alpha^\tau r(0) = \alpha^\tau \quad (3.7)$$

Realmente en (3.7) debe ser $r(\tau) = \alpha^{|\tau|}$, dado que, como vimos en el capítulo anterior, $r(\tau)$ es una función simétrica alrededor τ . La Figura 3.2 muestra la función de autocorrelación para α positivo y α negativo, respectivamente.

Figura 3.2a: ACF de un AR(1) con α positivo



La varianza es conocida indistintamente como $R(0)$, $Var(x_t)$ o σ_x^2 , y es también igual a Ex_t^2 en este caso porque suponemos que la media de x_t es cero. Para encontrar la varianza escribimos:

$$\begin{aligned} R(0) &= Var(x_t) = \sigma_x^2 = Ex_t^2 = \alpha E(x_{t-1}x_t) + E(\varepsilon_t x_t) \\ &= \alpha R(1) + E(\alpha x_{t-1}\varepsilon_t + \varepsilon_t^2) = \alpha [R(0)] + E\varepsilon_t^2 = \alpha^2 \sigma_x^2 + \sigma_\varepsilon^2 \end{aligned}$$

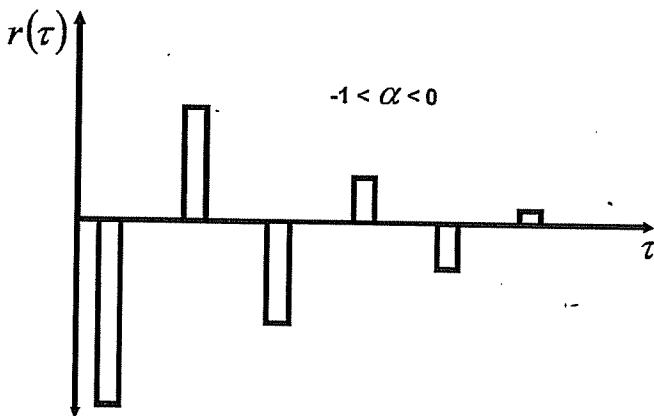
Donde hemos utilizado (3.6). Reorganizando obtenemos:

$$\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1 - \alpha^2} \quad (3.8)$$

Una forma alternativa de encontrar la varianza es escribiendo:

$$\begin{aligned} \sigma_x^2 &= Ex_t^2 = E(\alpha x_{t-1} + \varepsilon_t)^2 = E\alpha^2 x_{t-1}^2 + 2E\alpha x_{t-1}\varepsilon_t + E\varepsilon_t^2 \\ &= \alpha^2 Ex_{t-1}^2 + E\varepsilon_t^2 = \alpha^2 \sigma_x^2 + \sigma_\varepsilon^2 \end{aligned}$$

De donde, reorganizando, se obtiene (3.8).

Figura 3.2b: ACF de un AR(1) con α negativo

El AR(2)

El modelo autorregresivo con dos rezagos (de orden dos) tiene la siguiente expresión:

$$x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} = \varepsilon_t \quad (3.9)$$

Tomando expectativas en (3.9) y suponiendo estacionariedad encontramos que, de manera similar a como se hizo para el AR(1):

$$Ex_t = \frac{E\varepsilon_t}{1 + \alpha_1 + \alpha_2}$$

La expresión (3.9) se puede escribir, utilizando el operador de rezago, como:

$$(1 + \alpha_1 L + \alpha_2 L^2)x_t = \varepsilon_t$$

Para que el proceso sea estable, es necesario que las raíces de la ecuación auxiliar $(1 + \alpha_1 L + \alpha_2 L^2)$ estén por fuera del círculo unitario, ya que de otra forma la expansión expresada en (2.38) resulta explosiva; adicionalmente, la solución homogénea o complementaria, que en este caso es de la forma $A_1\lambda_1^t + A_2\lambda_2^t$, donde λ_1 y λ_2 son los inversos de las raíces de $(1 + \alpha_1 L + \alpha_2 L^2)$, también crecería sin límite.

Para encontrar la función de autocorrelación, multiplicamos ambos lados de la ecuación (3.9) por $x_{t-\tau}$ y tomamos expectativas:

$$E(x_t x_{t-\tau}) + \alpha_1 E(x_{t-1} x_{t-\tau}) + \alpha_2 E(x_{t-2} x_{t-\tau}) = E(\varepsilon_t x_{t-\tau})$$

Donde, para

$$\tau = 0 \text{ tendremos } R(0) [1 + \alpha_1 r(1) + \alpha_2 r(2)] = \sigma_\varepsilon^2 \quad (3.10a)$$

$$\tau = 1 \text{ tendremos } r(1) + \alpha_1 r(0) + \alpha_2 r(1) = 0 \quad (3.10b)$$

$$\tau = 2 \text{ tendremos } r(2) + \alpha_1 r(1) + \alpha_2 r(0) = 0 \quad (3.10c)$$

$$\vdots \qquad \vdots$$

$$r(\tau) + \alpha_1 r(\tau - 1) + \alpha_2 r(\tau - 2) = 0 \quad (3.10d)$$

En general, tenemos en (3.10d) una ecuación en diferencias sin excitación (*sin forcing function*), por lo cual su solución es sólo de la forma homogénea o complementaria

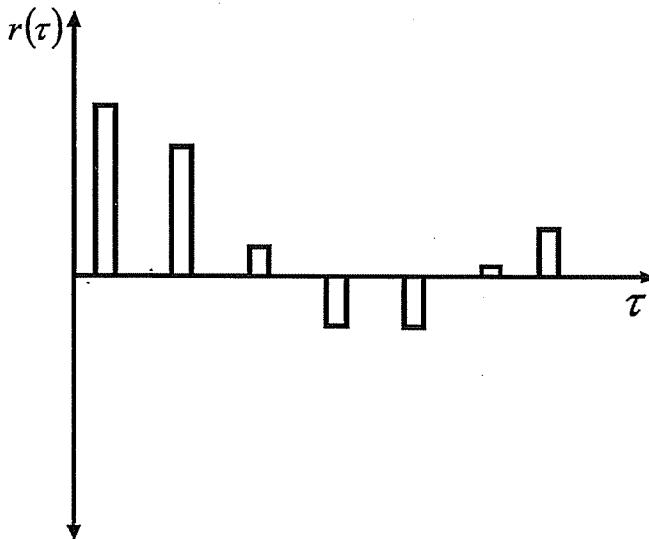
$$r(\tau) = B_1 \lambda_1^\tau + B_2 \lambda_2^\tau \quad (3.11)$$

Donde λ_1 y λ_2 son los inversos de las raíces de la ecuación auxiliar $(1 + \alpha_1 L + \alpha_2 L^2)$. La función de autocorrelación no es igual a la solución homogénea de x_t , ya que las constantes B_1 y B_2 salen de las condiciones iniciales de $r(\tau)$ y no de x_t .

Intuitivamente, podemos observar que si en la expresión $B_1 \lambda_1^\tau + B_2 \lambda_2^\tau$ las raíces son reales, $r(\tau)$ va a caer suavemente (en envolvente) hacia cero siempre y cuando ambas raíces estén por fuera del círculo unitario; pero si las raíces son complejas conjugadas, entonces sus inversos también lo serán y, según (2.40), $B_1 \lambda_1^\tau$ se puede expresar como $B_1 [m e^{i\phi}]^\tau = B_1 m^\tau [\cos \phi \tau + i \operatorname{sen} \phi \tau]$, la conjugada $B_2 \lambda_2^\tau$, como $B_2 [m e^{-i\phi}]^\tau = B_2 m^\tau [\cos \phi \tau - i \operatorname{sen} \phi \tau]$, y podemos inferir que $r(\tau)$ va a tener un comportamiento oscilatorio en función de τ , adicional al decaimiento exponencial, como se muestra en la Figura 3.3, siempre y cuando la magnitud del inverso de las raíces sea menor que uno, esto es, $m < 1$.

No obstante, si las raíces son complejas, ¿cómo desaparece la parte imaginaria de los λ en la expresión (3.11)? Se cancela con parte de B_1 y B_2 , que también corresponden a complejos conjugados. Según Enders (2004, Apéndice 1.1), utilizando varias identidades trigonométricas con los inversos de las raíces $\lambda_1 = m (\cos \phi + i \operatorname{sen} \phi)$ y $\lambda_2 = m (\cos \phi - i \operatorname{sen} \phi)$, y con los coeficientes $B_1 = B (\cos \theta + i \operatorname{sen} \theta)$ y $B_2 = B (\cos \theta - i \operatorname{sen} \theta)$, la autocorrelación será $r(\tau) = 2Bm^\tau \cos(\phi \tau + \theta)$.

Figura 3.3: ACF con decaimiento oscilatorio



Las constantes B_1 y B_2 se calculan remplazando en (3.11) las condiciones iniciales $r(0) = 1$ y $r(1) = \frac{-\alpha_1}{1+\alpha_2}$, donde esta última sale de resolver la ecuación (3.10b). Tendremos así, a partir de (3.11), y dado que λ_1 y λ_2 son conocidas y están en función de α_1 y α_2 , dos ecuaciones para las dos incógnitas B_1 y B_2 , una para $\tau = 0$ y otra para $\tau = 1$:

$$r(0) = 1 = B_1 + B_2$$

$$r(1) = \frac{-\alpha_1}{1+\alpha_2} = B_1\lambda_1 + B_2\lambda_2$$

Para raíces distintas, se puede mostrar que la expresión final para cualquier $\tau \geq 0$ es (Priestley, 1981, p. 128):

$$r(\tau) = \frac{(1 - \lambda_2^2)\lambda_1^{\tau+1} - (1 - \lambda_1^2)\lambda_2^{\tau+1}}{(\lambda_1 - \lambda_2)(1 + \lambda_1\lambda_2)} \quad (3.12)$$

En cuanto a la varianza de la serie, $R(0)$, nótese que las expresiones de $r(1)$ y $r(2)$ pueden obtenerse de (3.10b) y (3.10c); y luego $R(0)$ puede despejarse de (3.10a) y quedar en función de los α y de σ_ε^2 :

$$R(0) = \frac{\sigma_\varepsilon^2}{1 + \alpha_1 r(1) + \alpha_2 r(2)}$$

El AR(p)

La expresión (3.1) del autorregresivo de orden p , en la que se omite la constante sin pérdida de generalidad, se puede escribir en términos del operador de rezago como:

$$(1 + \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_p L^p) x_t = \alpha(L)x_t = \varepsilon_t$$

Al igual que en los casos anteriores, se puede encontrar que la media de x_t es:

$$Ex_t = \frac{E\varepsilon_t}{1 + \alpha_1 + \alpha_2 + \dots + \alpha_p}$$

De nuevo, para obtener estabilidad, las raíces del polinomio $\alpha(L)$, que es la ecuación auxiliar, deben estar por fuera del círculo unitario. Esto se puede ver de la solución general de un AR(p) que es de la forma:

$$x_t = h_t + \alpha(L)^{-1} \varepsilon_t \quad (3.13)$$

Donde h_t es la solución homogénea o complementaria correspondiente a $\alpha(L)x_t = 0$, y es de la forma:

$$h_t = A_1 \lambda_1^t + A_2 \lambda_2^t + \dots + A_p \lambda_p^t \quad (3.14)$$

Tal expresión crecería sin límite si las raíces (iguales al inverso de las λ) no están por fuera del círculo unitario. Adicionalmente, la expansión del segundo término de la derecha de (3.13) no sería convergente en términos del pasado.

Para encontrar la ACF del AR(p) se multiplica la ecuación (3.1) por $x_{t-\tau}$ y, de manera similar a como se hizo para el AR(2), se toma valor esperado para obtener:

$$R(\tau) + \sum_{j=1}^p \alpha_j R(\tau - j) = 0 \quad \tau > 0 \quad (3.15)$$

Para $\tau = 0$ se puede observar, como se hizo en (3.10), que debe adicionarse el término σ_ε^2 . Dividiendo ambos lados de (3.15) por $R(0)$ obtenemos la autocorrelación:

$$r(\tau) + \sum_{j=1}^p \alpha_j r(\tau - j) = 0 \quad \tau > 0 \quad (3.16)$$

Esta también es una ecuación en diferencias sin excitación, cuya solución es de la forma:

$$r(\tau) = B_1 \lambda_1^\tau + B_2 \lambda_2^\tau + \dots + B_p \lambda_p^\tau \quad (3.17)$$

Donde las p constantes B salen de remplazar p condiciones iniciales de $r(0), r(1), r(2), \dots, r(p - 1)$ en (3.17). Estas condiciones iniciales, aparte de $r(0)$ que siempre vale uno, se obtienen resolviendo un número suficiente de ecuaciones (3.16). Los B son, en general, números complejos; esto es, pueden ser reales o imaginarios.

El comportamiento de $r(\tau)$ en (3.17) dependerá de las raíces de la ecuación auxiliar (cuyos inversos son los λ_i). Para un proceso estacionario, las raíces reales generan términos exponenciales amortiguados; las parejas conjugadas de raíces complejas generan senoidales amortiguadas; y múltiples parejas conjugadas generan múltiples comportamientos oscilatorios, pero con diferentes frecuencias (periodos).

Con excepción del AR(1) y el AR(2), es difícil establecer un patrón general de comportamiento para la función de autocorrelación; solo que será la suma algebraica de diferentes exponenciales y senoidales amortiguadas, suma que, inclusive, puede caer abruptamente. La razón es que, especialmente cuando hay muchas raíces, las exponenciales y senoidales pueden cancelarse y mostrar barras no significativas después de algún τ . Como se verá más adelante, la función de autocorrelación de un modelo MA(1) cae abruptamente después del $\tau = 1$; en consecuencia, la función del AR(∞) equivalente también caería abruptamente después de $\tau = 1$. ¿Qué pasa si en un AR(p) la media $E x_t = \mu \neq 0$? En este caso podemos escribir:

$$\begin{aligned} x_t - \mu &= \alpha_1(x_{t-1} - \mu) + \alpha_2(x_{t-2} - \mu) + \dots + \alpha_p(x_{t-p} - \mu) + \varepsilon_t \\ x_t &= (\mu - \alpha_1\mu - \alpha_2\mu - \dots - \alpha_p\mu) + \alpha_1x_{t-1} + \alpha_2x_{t-2} + \dots + \alpha_px_{t-p} + \varepsilon_t \\ x_t &= \mu(1 - \alpha_1 - \alpha_2 - \dots - \alpha_p) + \alpha_1x_{t-1} + \alpha_2x_{t-2} + \dots + \alpha_px_{t-p} + \varepsilon_t \\ x_t &= c + \alpha_1x_{t-1} + \alpha_2x_{t-2} + \dots + \alpha_px_{t-p} + \varepsilon_t \end{aligned} \quad (3.18)$$

Al estimar el modelo AR(p) se puede incluir la constante si la serie no está en desviaciones de la media, o se puede restar la media a todas las observaciones y no incluir la constante. Algunos programas de computador reportan estimativos para la media μ y la constante c por separado con base en la expresión anterior.

El modelo de promedio móvil MA(q)

El segundo modelo dentro de la metodología de Box-Jenkins es el de promedio móvil. Éste expresa la serie x_t en función del presente y el pasado de una serie de ruido blanco $\{\varepsilon_t\}$ con media cero y varianza finita:

$$x_t = c + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} = c + \sum_{s=0}^q \beta_s \varepsilon_{t-s} \quad (3.19)$$

Donde los coeficientes β no tienen restricciones en cuanto a la estabilidad de la serie porque una suma finita de q términos con varianza finita, como lo son los ε_t , no puede ser explosiva. En (3.19) es usual suponer $\beta_0 = 1$. Se ve fácilmente que si $c = 0$ entonces $E x_t = 0$ porque $E \varepsilon_t = 0$, de otra forma $E x_t = c$. La varianza de x_t será $\sigma_\varepsilon^2 \sum_{s=0}^q \beta_s^2$.

Como se mencionó, en el contexto de series de tiempo y especialmente en el caso de los modelos de promedio móvil, la secuencia $\{\varepsilon_t\}$ también se conoce con el nombre de innovaciones, en el sentido de ser información nueva que llega al sistema.

La función de autocovarianza se obtiene multiplicando la ecuación (3.19) por $x_{t-\tau}$ y luego tomando expectativas (se omite la constante sin pérdida de generalidad):

$$\begin{aligned} R(\tau) &= E [x_t x_{t-\tau}] = E \left(\sum_s \beta_s \varepsilon_{t-s} \right) \left(\sum_u \beta_u \varepsilon_{t-\tau-u} \right) \\ &= E \left(\sum_s \sum_u \beta_s \beta_u \varepsilon_{t-s} \varepsilon_{t-\tau-u} \right) = \sum_s \sum_u \beta_s \beta_u E (\varepsilon_{t-s} \varepsilon_{t-\tau-u}) \end{aligned} \quad (3.20)$$

El término $E (\varepsilon_{t-s} \varepsilon_{t-\tau-u})$ en (3.20) será igual a σ_ε^2 cuando se cumpla $t - s = t - \tau - u$, o sea cuando $u = s - \tau$, y será cero en el caso contrario por el supuesto de independencia en los errores, de manera que la autocovarianza queda:

$$R(\tau) = E [x_t x_{t-\tau}] = \sigma_\varepsilon^2 \sum_{s=\tau}^q \beta_s \beta_{s-\tau} \quad (3.21)$$

Donde s va de τ a q . Es necesario que $s \geq \tau$ porque el subíndice $s - \tau$ en $\beta_{s-\tau}$ debe ser mayor o igual a cero, y el límite superior $s \leq q$ se explica porque los β_s para $s > q$ simplemente no existen. Al dividir por $R(0)$ se obtiene la función de autocorrelación cuya expresión es:

$$r(\tau) = \frac{\sum_{s=\tau}^q \beta_s \beta_{s-\tau}}{\sum_{s=0}^q \beta_s^2} = \frac{\beta_\tau + \beta_1 \beta_{\tau+1} + \dots + \beta_{q-\tau} \beta_q}{1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2} \quad (3.22)$$

Para $\tau \leq q$, y cero para $\tau > q$. En otras palabras, la ACF teórica para un $MA(q)$ puede tener autocorrelaciones significativas sólo hasta el rezago q , de ahí en adelante desaparecen. Pero en la práctica, por error estadístico, es posible encontrar alguna significativa después del rezago q .

MA(1)

El modelo $MA(q)$ más sencillo es el que tiene un sólo rezago de ε_t . Este se denota $MA(1)$ y su expresión es:

$$x_t = \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} \quad (3.23)$$

En (3.23) es evidente que $E x_t = 0$ (si hubiera una constante c sería $E x_t = c$). La función de autocovarianza para un $MA(1)$ se encuentra de la fórmula (3.21) o, para ilustrar, multiplicando la ecuación (3.23) por la misma ecuación rezagada τ unidades, y tomando expectativas:

$$\begin{aligned} R(\tau) &= E(x_t x_{t-\tau}) = E(\beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1})(\beta_0 \varepsilon_{t-\tau} + \beta_1 \varepsilon_{t-1-\tau}) \\ &= E(\beta_0 \varepsilon_t \beta_0 \varepsilon_{t-\tau} + \beta_0 \varepsilon_t \beta_1 \varepsilon_{t-1-\tau} + \beta_1 \varepsilon_{t-1} \beta_0 \varepsilon_{t-\tau} + \beta_1 \varepsilon_{t-1} \beta_1 \varepsilon_{t-1-\tau}) \end{aligned} \quad (3.24)$$

Para $\tau \geq 0$ en (3.24) tenemos:

$$\begin{aligned} R(0) &= E(\beta_0^2 \varepsilon_t^2 + \beta_1^2 \varepsilon_{t-1}^2) = \beta_0^2 E \varepsilon_t^2 + \beta_1^2 E \varepsilon_{t-1}^2 = \sigma_\varepsilon^2 (\beta_0^2 + \beta_1^2) \\ R(1) &= E(\beta_0 \beta_1 \varepsilon_{t-1}^2) = \sigma_\varepsilon^2 \beta_0 \beta_1 \\ R(2) &= 0, \quad R(3) = 0, \quad \dots \end{aligned} \quad (3.25)$$

Se deja como ejercicio mostrar que se cumple la propiedad de simetría, por ejemplo, que $R(-1) = E(\beta_0 \varepsilon_t \beta_1 \varepsilon_{t-1+\tau}) = \sigma_\varepsilon^2 \beta_0 \beta_1$. Para

encontrar la función de autocorrelación $r(\tau)$ simplemente dividimos las autocovarianzas anteriores por $R(0)$.

¿Cuál será el máximo $r(1)$ para el MA(1)? Supongamos, por simplicidad, que $\beta_0 = 1$ en (3.23). De acuerdo con la fórmula (3.21) o los resultados (3.25), tenemos:

$$\begin{aligned} R(1) &= \sigma_\varepsilon^2 \beta_1 \\ R(0) &= \sigma_\varepsilon^2 (1 + \beta_1^2) \end{aligned}$$

y, por lo tanto:

$$r(1) = \frac{\beta_1}{1 + \beta_1^2}.$$

Al derivar la expresión anterior con respecto a β_1 e igualando a cero para maximizar, tenemos:

$$\begin{aligned} \beta_1(-(1 + \beta_1^2)^{-2}2\beta_1) + (1 + \beta_1^2)^{-1} &= 0 \\ -2\beta_1^2(1 + \beta_1^2)^{-1} + 1 &= 0 \\ 2\beta_1^2 = (1 + \beta_1^2) &\implies \beta_1^2 = 1 \implies \beta_1 = \pm 1 \end{aligned}$$

De manera que $r(1)$ máximo será:

$$r(1)_{max} = \frac{\beta_1}{1 + \beta_1^2} = \frac{\pm 1}{1 + 1} = \pm \frac{1}{2} \quad (3.26)$$

Positivo o negativo dependiendo del signo de β_1 .

Invertibilidad

Mientras que los β sean finitos, el MA(q) siempre va a ser estacionario puesto que se compone de una suma finita de variables aleatorias con varianza finita. No obstante, si bien la condición de que las raíces de la ecuación auxiliar:

$$1 + \beta_1 L + \beta_2 L^2 + \dots + \beta_q L^q = \beta(L), \quad (3.27)$$

(con $c = 0$ y $\beta_0 = 1$ para simplificar) estén por fuera del círculo unitario, *no* es necesaria para la estabilidad, *sí* lo es para poder expresar el modelo de promedio móvil como un autorregresivo, lo que se conoce como

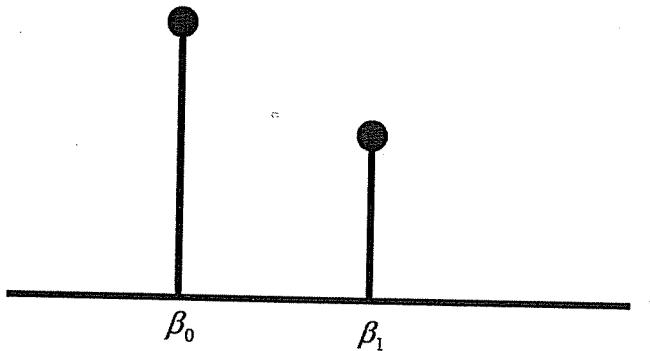
invertibilidad; esto es, despejar ε_t en función de x_t y del pasado de x_t . Por ejemplo, para invertir un MA(1) y convertirlo en un autorregresivo hacemos:

$$\begin{aligned}x_t &= \varepsilon_t + \beta\varepsilon_{t-1} \\x_t &= (1 + \beta L)\varepsilon_t \\ \varepsilon_t &= \frac{x_t}{(1 + \beta L)} = [1 - \beta L + (\beta L)^2 - (\beta L)^3 + \dots] x_t \\ \varepsilon_t &= x_t - \beta x_{t-1} + \beta^2 x_{t-2} - \beta^3 x_{t-3} + \dots\end{aligned}\quad (3.28)$$

Lo anterior se puede hacer sólo si $|\beta| < 1$. Nótese que el MA(1) de este ejemplo se convirtió en un AR(∞).

La invertibilidad impone restricciones a las magnitudes relativas de los coeficientes β_0, \dots, β_q ; en otras palabras, impone restricciones de comportamiento. Implica, por ejemplo, que para cualquier MA(q) el coeficiente $|\beta_0|$ será siempre mayor que $|\beta_q|$. Generalmente suponemos $\beta_0 = 1$ de manera que la condición anterior equivale a $1 > |\beta_q|$. En el caso específico de un MA(1), la condición de invertibilidad implica que la reacción de x_t a la llegada de una nueva innovación ε_t será mayor en t que en el periodo siguiente $t + 1$, como se muestra en la Figura 3.4.

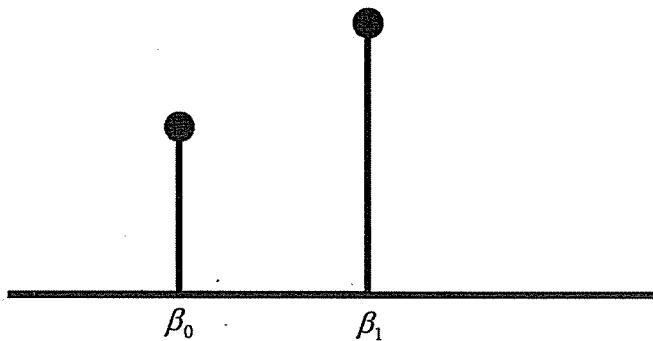
Figura 3.4: Respuesta de un MA(1) invertible



Aunque siempre será posible en un MA(2) que $|\beta_1| > |\beta_0|$, la invertibilidad de todas maneras exige que $|\beta_1| < 2|\beta_0|$ y que $|\beta_0| > |\beta_2|$. Restricciones similares afectan los coeficientes de procesos MA de mayor orden. Al aceptar sólo modelos invertibles estamos, en efecto, pasando

por alto especificaciones alternativas que pueden ser económicamente viables e interesantes. Por ejemplo, un MA(1) no invertible puede tener $|\beta_0| < |\beta_1|$, de manera que la reacción de x_t a una nueva innovación ε_t puede ser mayor después de pasado un periodo, como se ilustra en la Figura 3.5. Para la estimación de modelos MA(q) no invertibles véase Ramsey y Montenegro (1992).

Figura 3.5: Respuesta de un MA(1) no invertible



Vimos en (3.28) que un MA(1) se puede expresar como un AR(∞). En general, un MA(q) se puede expresar como un AR(∞). También un AR(p) se puede expresar como un MA(∞). Ilustremos convirtiendo un AR(1) en un MA(∞):

$$\begin{aligned}x_t &= \alpha x_{t-1} + \varepsilon_t \\(1 - \alpha L)x_t &= \varepsilon_t \\x_t &= \frac{\varepsilon_t}{1 - \alpha L}\end{aligned}$$

donde, para $|\alpha| < 1$

$$x_t = \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha^2 \varepsilon_{t-2} + \alpha^3 \varepsilon_{t-3} + \dots \quad (3.29)$$

Ejemplo

Convierta el siguiente AR en un MA(∞):

$$\begin{aligned}x_t &= \alpha x_{t-2} + \varepsilon_t, \quad |\alpha| < 1 \\x_t - \alpha x_{t-2} &= (1 - \alpha L^2)x_t = \varepsilon_t \\(1 - \sqrt{\alpha}L)(1 + \sqrt{\alpha}L)x_t &= \varepsilon_t\end{aligned}$$

Donde las raíces son $-\frac{1}{\sqrt{\alpha}}, +\frac{1}{\sqrt{\alpha}}$. Podemos despejar x_t y escribir:

$$\begin{aligned}
 x_t &= \frac{\varepsilon_t}{(1 - \sqrt{\alpha}L)(1 + \sqrt{\alpha}L)} \\
 &= [1 + \sqrt{\alpha}L + (\sqrt{\alpha}L)^2 + (\sqrt{\alpha}L)^3 + \dots] \frac{\varepsilon_t}{1 + \sqrt{\alpha}L} \\
 &= [1 - \sqrt{\alpha}L + (\sqrt{\alpha}L)^2 - (\sqrt{\alpha}L)^3 + \dots] \varepsilon_t \\
 &\quad + [\sqrt{\alpha}L - (\sqrt{\alpha}L)^2 + (\sqrt{\alpha}L)^3 - \dots] \varepsilon_t \\
 &\quad + [(\sqrt{\alpha}L)^2 - (\sqrt{\alpha}L)^3 + (\sqrt{\alpha}L)^4 - \dots] \varepsilon_t \\
 &\quad + [(\sqrt{\alpha}L)^3 - (\sqrt{\alpha}L)^4 + (\sqrt{\alpha}L)^5 - \dots] \varepsilon_t + \dots \\
 &= [1 + (\sqrt{\alpha}L)^2 + (\sqrt{\alpha}L)^4 + \dots] \varepsilon_t
 \end{aligned} \tag{3.30}$$

También hubiéramos podido obtener el resultado anterior aplicando desde el principio la propiedad:

$$\frac{1}{1 - \alpha L^2} = [1 + \alpha L^2 + (\alpha L^2)^2 + (\alpha L^2)^3 + \dots]$$

a la expresión $x_t = \varepsilon_t / (1 - \alpha L^2)$.

Teorema de descomposición de Wold

Todo proceso estocástico x_t débilmente estacionario (estacionario de orden 2) con media cero puede ser representado como la suma:

$$x_t = D_t + S_t \tag{3.31}$$

Donde:

- a. D_t y S_t son procesos lineales no correlacionados.
- b. D_t es determinístico y puede ser predicho arbitrariamente bien usando sólo rezagos de su propio pasado.
- c. S_t es estocástico y tiene la forma:

$$S_t = \sum_{\nu=0}^{\infty} \beta_{\nu} \varepsilon_{t-\nu} \tag{3.32}$$

donde $\beta_0 = 1$, $\sum_{\nu=0}^{\infty} \beta_{\nu}^2 < \infty$, y ε_t es un proceso de ruido igual al error de la proyección lineal de x_t sobre el espacio vectorial generado por $x_{t-1}, x_{t-2}, x_{t-3}, \dots$

Existe una generalización de este teorema aplicable a procesos vectoriales. También otra generalización para procesos no estacionarios donde los coeficientes β dependen del tiempo (véase Granger y Newbold, 1986, p. 39). Un recuento de la evolución del pensamiento en series de tiempo, que en cierta forma unifica el trabajo de Wold (1938), se puede consultar en Klein (1997).

Modelo ARMA(p,q)

De acuerdo con el resultado anterior, cualquier proceso débilmente estacionario, con media cero, aun si es generado de manera no lineal o a partir de errores autocorrelacionados, tendrá una representación de Wold. De manera que podemos utilizar los modelos AR(p) y MA(q) como aproximaciones. También podemos utilizar la combinación de ambos modelos y originar el modelo ARMA(p, q), con el cual usualmente se obtiene mayor parsimonia, esto es, emplear un menor número de términos comparado con el AR o el MA individualmente. La expresión del ARMA(p, q) será:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (3.33)$$

O también:

$$x_t = \sum_{j=1}^p \alpha_j x_{t-j} + \sum_{j=0}^q \beta_j \varepsilon_{t-j} \quad (3.34)$$

Donde usualmente $\beta_0 = 1$. Se puede agregar una constante si se considera necesario. El valor esperado de x_t se obtiene, como en los dos modelos anteriores, tomando expectativas a toda la expresión y recordando que la serie es estacionaria; éste será cero si no hay constante y ε_t tiene media cero. La varianza se obtiene de la solución simultánea del conjunto de ecuaciones de autocorrelación similares a (3.15) y (3.16). En términos del operador de rezago escribimos:

$$\alpha(L)x_t = \beta(L)\varepsilon_t$$

Donde $\alpha(L)$ y $\beta(L)$ son polinomios en L .

Vale la pena anotar que p y q se refieren al *último* rezago y no necesariamente al *número* de rezagos de la parte autorregresiva o de la parte promedio móvil, ya que muchos rezagos pueden ser no significativos.

La estacionariedad del modelo ARMA sólo depende de la parte autorregresiva y no de la parte promedio móvil; esto es, de que las raíces de la ecuación auxiliar de la parte autorregresiva estén por fuera del círculo unitario.

La forma de la ACF para un ARMA es más difícil de establecer que en el caso de los modelos anteriores. En general, habrá unas autocorrelaciones iniciales significativas fijadas por la parte promedio móvil seguidas, a partir de $\tau > q$, por un comportamiento similar al de la parte autorregresiva (véase Granger y Newbold, 1986, capítulo 1).

Cuando se adicionan variables exógenas como explicativas, el modelo ARMA se conoce como ARMAX. Este podría ser de la forma:

$$x_t = \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} + \gamma_1 z_{1t} + \dots + \gamma_k z_{kt} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (3.35)$$

Donde z_{1t}, \dots, z_{kt} son variables exógenas. Para no entorpecer las posibilidades predictivas del modelo, es conveniente que las variables exógenas que se incluyan se encuentren rezagadas; de lo contrario (si son contemporáneas) no será posible utilizarlas para predicción, o habría que predecirlas primero.

Ejemplo

Un punto interesante es si en algunos casos la combinación de modelos del mismo orden puede producir un modelo de diferente orden –véase Granger y Newbold (1986, p. 29) para un teorema sobre combinación de modelos ARMA-. Mostramos a continuación que la suma $z_t = x_t + y_t$ es un AR(2), donde x_t y y_t son los AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$ y $y_t = -\alpha y_{t-1} + u_t$; ε_t y u_t procesos de ruido independientes entre ellos con media cero y varianza común $\sigma_\varepsilon^2 = \sigma_u^2 = \sigma^2$.

$$z_t = x_t + y_t = \frac{\varepsilon_t}{1 - \alpha L} + \frac{u_t}{1 + \alpha L}$$

$$(1 - \alpha L)(1 + \alpha L)z_t = (1 + \alpha L)\varepsilon_t + (1 - \alpha L)u_t$$

$$(1 - \alpha^2 L^2)z_t = z_t - \alpha^2 z_{t-2} = \varepsilon_t + \alpha \varepsilon_{t-1} + u_t - \alpha u_{t-1}$$

De manera que obtenemos el AR(2):

$$z_t = \alpha^2 z_{t-2} + k_t$$

Donde se deja como ejercicio mostrar que $k_t = \varepsilon_t + \alpha \varepsilon_{t-1} + u_t - \alpha u_{t-1}$ es ruido blanco con media cero y varianza $2\sigma^2(1 + \alpha^2)$.

Diagramas de bloque

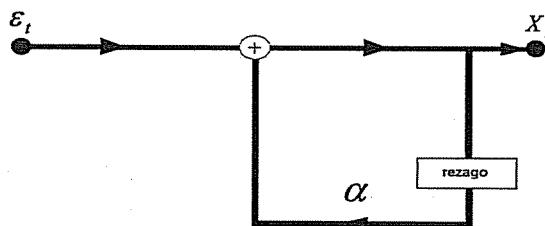
Una forma muy conveniente de ilustrar la interrelación entre los términos de un AR, MA o ARMA (en general, de una ecuación en diferencias), es por medio de los diagramas de bloque, de los cuales mostramos tres ejemplos a continuación.

Para un AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$ véase la Figura 3.6. Se observa que hay *realimentación* (*feedback*). Realimentación es cuando una parte de la salida vuelve a la entrada del sistema; específicamente, cuando un valor anterior de la salida se utiliza para calcular el valor actual de la salida. Puede intuirse que si $|\alpha| > 1$, la salida, x_t en este caso, tiende a crecer sin límite.

Figura 3.6: Diagrama de bloque para un AR(1)

AR(1)

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



Para un MA(1) $x_t = \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1}$ véase la Figura 3.7. Aquí no se observa realimentación; el sistema es inherentemente estable. Para

un ARMA(2, 2) $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$, véase la Figura 3.8. Existen otras formas de representar las ecuaciones en diferencia, o los modelos ARMA, alternativas a las mostradas aquí (para mayores detalles, véase Oppenheim y Willsky, 1983).

Figura 3.7: Diagrama de bloque para un MA(1)

MA(1)

$$X_t = \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1}$$

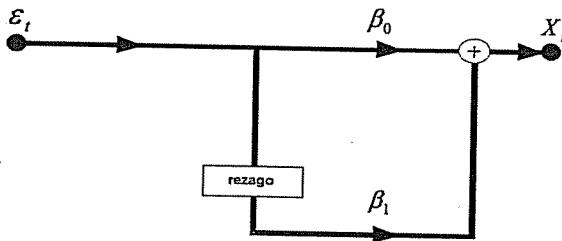
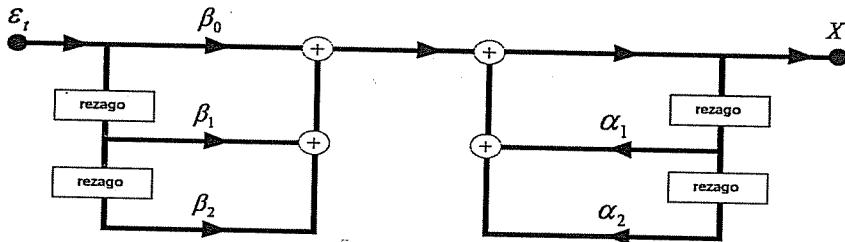


Figura 3.8: Diagrama de bloque para un ARMA(2, 2)

ARMA (2,2)

$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \beta_0 \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2}$$



Función de autocorrelación parcial

Vimos que la ACF mide la relación lineal entre x_t y $x_{t-\tau}$. Sin embargo, parte de esta relación se debe al encadenamiento que hay entre x_t y $x_{t-\tau}$ a través de las variables $x_{t-1}, \dots, x_{t-\tau+1}$. Si quitáramos el efecto del encadenamiento de estas variables intersticiales, ¿cuál sería la correlación restante entre x_t y $x_{t-\tau}$? Esta relación entre x_t y $x_{t-\tau}$, neta del efecto de las variables intersticiales, se conoce como función de autocorrelación parcial (PACF).

El coeficiente de autocorrelación parcial, que denotaremos $r_p(\tau)$, entre x_t y $x_{t-\tau}$ es proporcional a la correlación entre x_t y $x_{t-\tau}$, menos aquella parte que puede ser explicada por las variables intersticiales. La PACF para rezago τ es proporcional a la correlación entre los residuos (variación no explicada) del AR($\tau - 1$) y $x_{t-\tau}$. La función de autocorrelación parcial, $r_p(\tau)$, se define como el *último* coeficiente de un AR(τ); en otras palabras, $r_p(\tau)$ refleja la correlación entre $x_{t-\tau}$ y el término condicional $x_t - Ex_t/x_{t-1}, \dots, x_{t-\tau+1}$; por ejemplo en el AR(τ):

$$x_t = c + \alpha_{\tau,1}x_{t-1} + \alpha_{\tau,2}x_{t-2} + \dots + \alpha_{\tau,\tau-1}x_{t-(\tau-1)} + \alpha_{\tau,\tau}x_{t-\tau} + \varepsilon_t \quad (3.36)$$

La autocorrelación parcial será el *último* coeficiente, $\alpha_{\tau,\tau} = r_p(\tau)$. La ecuación (3.36) se repite para cada τ . Obviamente, la PACF para $\tau = 0$ es uno. Para $\tau = 1$ la ACF y la PACF son iguales, esto es, $r_p(1) = r(1)$, porque no hay variable intersticial. También, como en la ACF, la PACF tiene los intervalos de confianza del 95 % en $\pm 2/\sqrt{n}$. Para la ACF el intervalo corresponde a la hipótesis nula de que la serie es ruido blanco, mientras que para la PACF corresponde a la hipótesis nula de que la serie es realmente un AR(p) para $p < \tau$, porque los $r_p(\tau)$ deben ser cero para $\tau > p$ (véase Hamilton, 1994, p. 112); esto es, que no hay barras significativas después de $\tau = p$.

La PACF tiene un comportamiento contrario a la ACF en los modelos AR y MA. Para un AR(p) la PACF cae a cero abruptamente cuando $\tau > p$, simplemente, porque la expresión:

$$x_t - \alpha_1x_{t-1} - \alpha_2x_{t-2} - \dots - \alpha_px_{t-p}$$

es igual a ε_t por definición para un AR(p), y ε_t carece de covarianza con cualquier $x_{t-\tau}$ para $\tau > p$ porque dicho rezago ni siquiera existe en el modelo original. Sin embargo, a manera de curiosidad, podría argumentarse que, por ejemplo en un AR(2) $x_t = \alpha_1x_{t-1} + \alpha_2x_{t-2} + \varepsilon_t$, las variables x_{t-1} y x_{t-2} sí están correlacionadas con x_{t-3} porque pueden expresarse en función de x_{t-3} , y por lo tanto la covarianza $E[(x_t - \alpha_1x_{t-1} - \alpha_2x_{t-2})(x_{t-3})]$ sería diferente de cero. Esto no es correcto porque al hacer los reemplazos la covarianza con x_{t-3} desaparece y la expresión se iguala a cero,

$$\begin{aligned} E[(x_t - \alpha_1x_{t-1} - \alpha_2x_{t-2})(x_{t-3})] &= Ex_tx_{t-3} - \alpha_1Ex_{t-1}x_{t-3} \\ &\quad - \alpha_2Ex_{t-2}x_{t-3} \end{aligned}$$

$$\begin{aligned}
 &= E(\alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t) x_{t-3} - \alpha_1 E x_{t-1} x_{t-3} - \alpha_2 E x_{t-2} x_{t-3} \\
 &= \alpha_1 E x_{t-1} x_{t-3} + \alpha_2 E x_{t-2} x_{t-3} - \alpha_1 E x_{t-1} x_{t-3} - \alpha_2 E x_{t-2} x_{t-3} = 0
 \end{aligned}$$

Para el MA(q), mientras que su ACF cae a cero abruptamente para $\tau > q$, la PACF decrece monotónicamente o senoidalmente. La intuición de este resultado no es muy evidente; sin embargo, si se recuerda que $r_p(\tau)$ es el último coeficiente de un AR(τ), podemos esperar que, al tener un MA(q) invertible una representación AR infinita con coeficientes eventualmente decrecientes, sus autocorrelaciones parciales también decrezcan monotónicamente o senoidalmente hasta el infinito.

Para un ARMA(p, q), la ACF y la PACF caen exponencialmente (con posibles oscilaciones). La ACF comienza su caída después del rezago q mientras que la PACF comienza su caída después del rezago p (véase Enders, 1995, p. 85). Por ejemplo, para un ARMA(1, 1), tanto la ACF como la PACF tendrán un pico en $\tau = 1$ seguido de un patrón que decae exponencialmente. En cualquier caso, en la práctica estos patrones se desdibujan un poco por error estadístico y por el hecho de que los datos no necesariamente se ajustan de manera exacta a un modelo dado, ni sus coeficientes son inmutables en el tiempo.

Las figuras 3.9 a 3.12 ilustran las autocorrelaciones para cuatro modelos construidos con 1000 observaciones de ε_t . La primera corresponde a un AR(3), donde se observa una caída asintótica en la ACF mientras que la PACF cae después del tercer rezago. La Figura 3.10 corresponde a un MA(3) donde se observa una caída en la ACF después del tercer rezago y una caída asintótica en la PACF. La Figura 3.11 corresponde a un modelo AR(2), sin el primer rezago, cuya ACF muestra autocorrelaciones significativas y exponencialmente decrecientes en los rezagos pares, mientras que la PACF cae después del rezago 2.

La Figura 3.12 corresponde a un modelo MA(2), sin el primer rezago, cuya ACF cae después del rezago 2, mientras que su PACF muestra autocorrelaciones significativas, que alternan de signo, asintóticamente decrecientes en los rezagos pares. Si este MA(2) tuviera signo negativo en lugar de positivo en su rezago, las autocorrelaciones parciales serían todas negativas. El comportamiento de estos signos puede intuirse aproximadamente de la expresión AR equivalente del MA. Decimos de manera aproximada, porque debemos recordar que cada $r_p(\tau)$ es igual al *último* coeficiente de cada AR(τ), y no son exactamente iguales a los coeficientes del AR(∞) equivalente. La diferencia, aunque usual-

mente pequeña, se debe al conocido problema de omisión de variables relevantes en los AR(τ).

Figura 3.9: $X_t = 0.4X_{t-2} + 0.2X_{t-3} + \varepsilon_t$

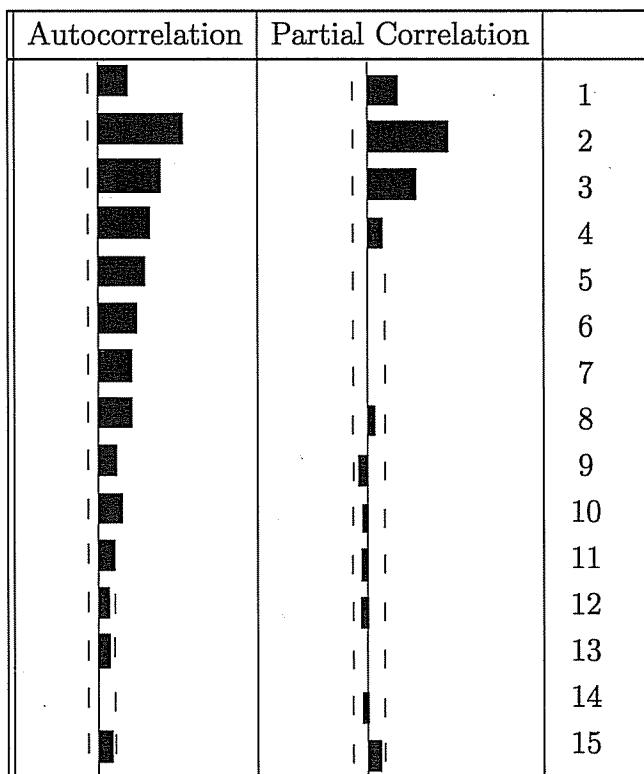


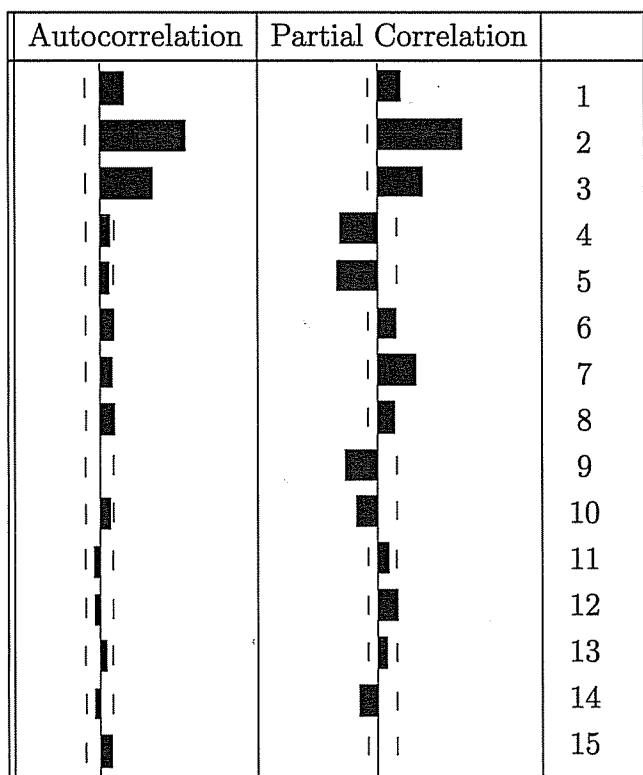
Figura 3.10: $X_t = \varepsilon_t + 0.6\varepsilon_{t-2} + 0.3\varepsilon_{t-3}$ 

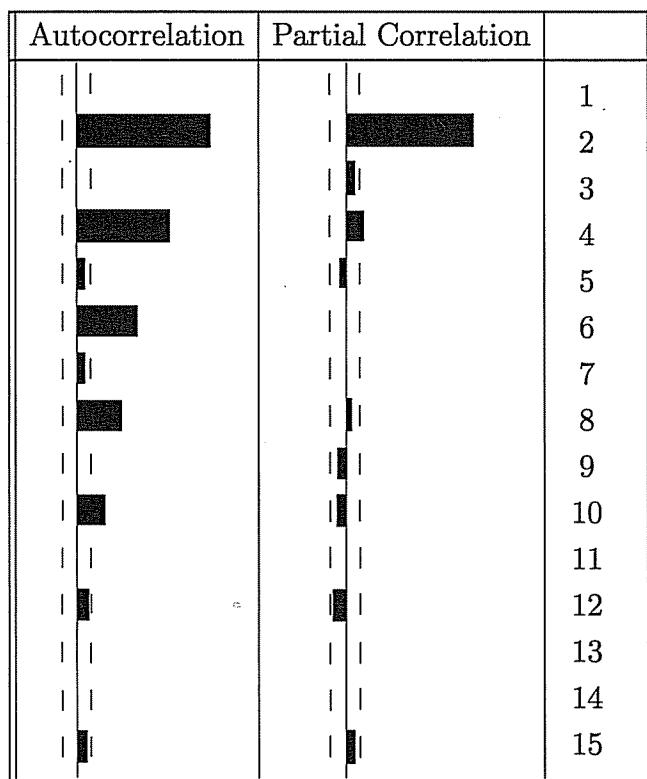
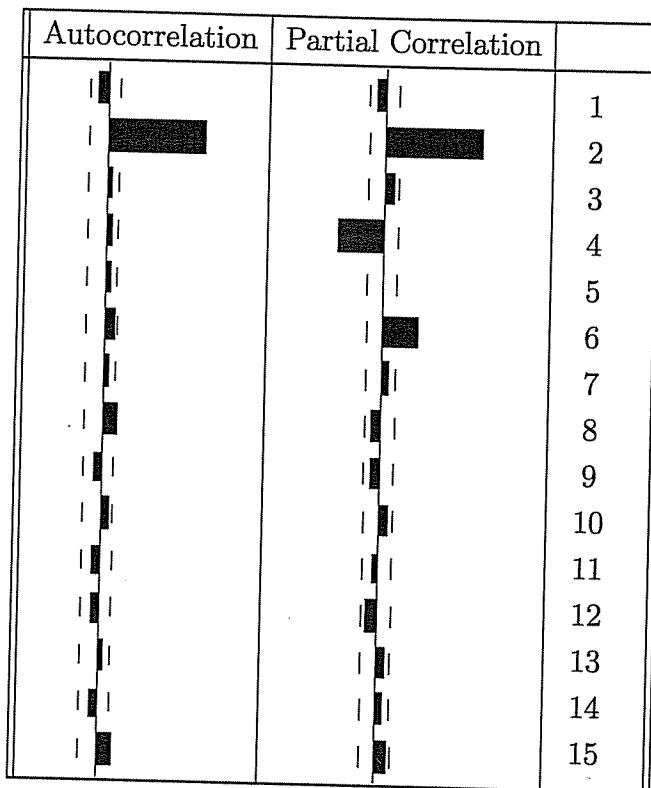
Figura 3.11: $X_t = 0.6X_{t-2} + \varepsilon_t$ 

Figura 3.12: $X_t = \varepsilon_t + 0.6\varepsilon_{t-2}$ *Ejemplo*

Para ilustrar la construcción de la PACF para un promedio móvil, escogemos un MA(1) y escribimos el AR(τ) correspondiente utilizando los términos del MA(1) rezagados:

$$\begin{aligned} x_t &= \varepsilon_t + \beta\varepsilon_{t-1} \doteq \alpha_1(\varepsilon_{t-1} + \beta\varepsilon_{t-2}) + \alpha_2(\varepsilon_{t-2} + \beta\varepsilon_{t-3}) \\ &\quad + \alpha_3(\varepsilon_{t-3} + \beta\varepsilon_{t-4}) + \dots + \alpha_\tau(\varepsilon_{t-\tau} + \beta\varepsilon_{t-\tau-1}) + \varepsilon_t \end{aligned}$$

En la ecuación el símbolo \doteq quiere decir aproximadamente. Vemos que, suponiendo β positivo, el coeficiente estimado $\hat{\alpha}_1$ será tal que haga que la cantidad $\varepsilon_{t-1} + \beta\varepsilon_{t-2}$ sea lo más parecida posible a la

cantidad $\varepsilon_t + \beta\varepsilon_{t-1}$. En consecuencia será *positivo* y cercano a β , pero no podrá evitar que aparezca el término $\alpha_1\beta\varepsilon_{t-2}$. A su turno, el coeficiente estimado $\hat{\alpha}_2$ será el encargado de eliminar en lo posible este efecto indeseado de $\alpha_1\beta\varepsilon_{t-2}$; en consecuencia será *negativo* y cercano a $-\alpha_1\beta$, pero no podrá evitar que aparezca el término $\alpha_2\beta\varepsilon_{t-3}$, por lo cual el siguiente coeficiente estimado $\hat{\alpha}_3$ será *positivo*, y así sucesivamente. Si β es negativo, puede mostrarse que todas las autocorrelaciones parciales serán negativas. Un razonamiento similar explica la PACF del MA(2) de la Figura 3.12.

Estacionalidad en los modelos ARMA

Cuando hay estacionalidad (no confundir con estacionariedad) aparecen en la ACF barras significativas tanto en el rezago estacional como en múltiplos de éste. En la PACF sólo aparece la primera barra estacional (rezago 4 para series trimestrales, 12 para series mensuales) porque, por ejemplo, para $\tau = 24$ en una serie mensual, el rezago $\tau = 12$ intersticial explica casi toda la estacionalidad y deja prácticamente sin correlación a x_t y x_{t-24} .

Hay varias formas de aproximar el efecto de la estacionalidad en los modelos ARMA. Una es usando *dummies* cuando la estacionalidad es determinística y no cambia en el tiempo. Otra es sumando términos estocásticos (caso aditivo) a la expresión del ARMA, ya sean autorregresivos o de promedio móvil, de rezago igual al periodo estacional; y otra es multiplicando la expresión del ARMA por un término estocástico que tenga un rezago igual al anual en términos de los datos usados (caso multiplicativo); esto es, x_{t-4} para datos trimestrales, x_{t-12} para datos mensuales, x_{t-6} para datos bimestrales, etc.

Por ejemplo, para el caso aditivo, la estacionalidad en un AR(1) con datos trimestrales y un término autorregresivo, será:

$$x_t = \alpha_1 x_{t-1} + \alpha_4 x_{t-4} + \varepsilon_t \quad (3.37)$$

La estacionalidad con datos mensuales en un ARMA(2, 1) adicionando un término autorregresivo, sería:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \alpha_{12} x_{t-12} \quad (3.38)$$

O también con un término aditivo de promedio móvil:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_{12} \varepsilon_{t-12} \quad (3.39)$$

En el caso multiplicativo, para introducir estacionalidad en una serie AR(1) de datos trimestrales, se usa:

$$(1 - \alpha_1 L)(1 - \alpha_4 L^4)x_t = \varepsilon_t$$

$$(1 - \alpha_1 L - \alpha_4 L^4 + \alpha_1 \alpha_4 L^5)x_t = \varepsilon_t$$

De donde se obtiene la expresión:

$$x_t = \alpha_1 x_{t-1} + \alpha_4 x_{t-4} - \alpha_1 \alpha_4 x_{t-5} + \varepsilon_t \quad (3.40)$$

Esto contrasta con el caso aditivo anterior en el cual no aparece el término $\alpha_1 \alpha_4 x_{t-5}$. En la versión multiplicativa se realiza la estimación con restricciones, esto es, con $\alpha_5 = \alpha_1 \alpha_4$, usualmente por máxima verosimilitud con restricciones o por el método del Lagrangiano. Cuál forma de aproximar la estacionalidad se prefiera es asunto del investigador, aunque la más intuitiva es la aditiva. El multiplicativo permite tener más rezagos con menos coeficientes estimados pero a costa de restringir la estimación, y la estimación restringida tiene mayor o igual suma de errores al cuadrado, $\sum \hat{\varepsilon}^2$, que la no restringida. No todos los programas de computador incluyen la facilidad para estimar la versión multiplicativa.

Para introducir estacionalidad en un MA(2) con datos trimestrales se podría escribir:

$$x_t = (1 + \beta_1 L + \beta_2 L^2)\varepsilon_t(1 + \beta_4 L^4) \quad (3.41)$$

Donde aparecerá un término de rezago 6. Para introducir el efecto estacional en un ARMA simplemente se multiplica la parte correspondiente.

Si una serie es estacional, digamos mensual, podemos intentar desestacionalizarla corriendo el modelo $x_t = \alpha_1 x_{t-1} + \alpha_{12} x_{t-12} + \varepsilon_t$ (un rezago, x_{t-1} , basta para ilustrar, aunque puede haber otros); la serie desestacionalizada será $x_t - \alpha_{12} x_{t-12}$.

Es de notar que la adición o multiplicación de rezagos sirve para modelar o extraer la estacionalidad de una serie de datos, pero no para

generar la estacionalidad. Por ejemplo, un proceso como $x_t = \alpha x_{t-12} + \varepsilon_t$, donde ε_t es ruido blanco y $0 < \alpha < 1$, es estacionario pero no es estacional; la gráfica de este proceso x_t así generado carece de un comportamiento que se repite periódicamente; para efectos prácticos, su gráfica sería parecida a la de un AR(1) y su ACF caería siguiendo α^{τ} . Generalmente, la estacionalidad se debe a factores exógenos ajenos a la variable x_t , como el clima, las normas tributarias, las costumbres, etc., de manera que la representación autorregresiva de la estacionalidad es una aproximación. Debe notarse, sin embargo, que existe alguna literatura en la cual el comportamiento estacional está relacionado con variables económicas de largo plazo, como los cambios tecnológicos y otras condiciones económicas (por ejemplo, véase Barsky y Miron, 1989).

El hecho de trabajar con series que supuestamente han sido desestacionalizadas por algún método como el X11, TRAMO/SEATS o cualquier otro, no implica que debamos desdenar la posibilidad de modelar la estacionalidad ya que, con frecuencia, por el tamaño de la muestra o las opciones escogidas, queda algo de este efecto en las series. Pasar por alto la estacionalidad residual aumenta la varianza de predicción. Cabe advertir que los términos *aditivo* y *multiplicativo*, utilizados aquí, denotan la forma como se incluye el rezago estacional en la serie, y no deben confundirse con los mismos términos utilizados en estos métodos de ajuste estacional para describir el modo en que se incluye el efecto estacional en la serie.

Respuesta impulso de un AR(p)

Un choque unitario a ε_t en $x_t = \alpha x_{t-1} + \varepsilon_t$ dibujará un patrón de respuesta en x_t llamado *respuesta impulso* (*impulse response*). Por ejemplo, si suponemos el sistema en reposo y hacemos $\varepsilon_0 = 1$ y todos los demás ε_t (pasados y futuros) iguales a cero, tenemos:

$$\begin{aligned} x_0 &= 1 \\ x_1 &= \alpha(1) + 0 = \alpha \\ x_2 &= \alpha^2 \\ x_3 &= \alpha^3 \\ &\vdots \end{aligned} \tag{3.42}$$

Vemos que la secuencia (3.42) es igual a los coeficientes del $MA(\infty)$ equivalente en (3.29), repetido aquí:

$$x_t = \frac{\varepsilon_t}{1 - \alpha L} = \varepsilon_t + \alpha \varepsilon_{t-1} + \alpha^2 \varepsilon_{t-2} + \alpha^3 \varepsilon_{t-3} + \dots$$

También son iguales a su ACF –véase (3.7)– en este caso porque no hay más rezagos x_{t-2}, x_{t-3}, \dots ; pero, para otros casos, no siempre será igual a la ACF ya que la respuesta impulso de un proceso puede ser mayor que uno en algún momento y la ACF no.

Se encuentra que, en general, la respuesta impulso de x_t en cualquier $AR(p)$ a un choque $\varepsilon_0 = 1$ será igual a los coeficientes de su representación $MA(\infty)$ equivalente.

Es importante notar que un $MA(\infty)$ equivalente de un $AR(p)$ estacionario puede tener coeficientes que crecen por encima de β_0 , aunque luego caen.

Ejemplo

Mostrar que los coeficientes del $MA(\infty)$ equivalente al $AR(2)$:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t,$$

pueden tomar valores mayores que uno. Escribimos, en general:

$$\begin{aligned} x_t &= \frac{\varepsilon_t}{(1 - aL)(1 - bL)} = (1 + aL + (aL)^2 + (aL)^3 + \dots) \\ &\quad \times (1 + bL + (bL)^2 + (bL)^3 + \dots) \varepsilon_t \\ &= (1 + bL + (bL)^2 + (bL)^3 + \dots + aL + abL^2 + ab^2L^3 + ab^3L^4 + \dots \\ &\quad + (aL)^2 + a^2bL^3 + a^2b^2L^4 + a^2b^3L^5 + \dots + (aL)^3 + a^3bL^4 \\ &\quad + a^3b^2L^5 + a^3b^3L^6 + \dots) \varepsilon_t \\ &= [1 + (a + b)L + (a^2 + ab + b^2)L^2 + (b^3 + a^2b + ab^2 + a^3)L^3 + \dots] \varepsilon_t \end{aligned} \tag{3.43}$$

Si $a = 0.6$ y $b = 0.6$, entonces el $AR(2)$ es estacionario y tendrá la expresión $MA(\infty)$:

$$(1 + 1.2L + 1.08L^2 + \dots) \varepsilon_t$$

Donde $1.2 = \beta_1 > \beta_0 = 1$.

Predicción utilizando un AR(p)

Sea x_t una serie que sigue un AR(p):

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t$$

Deseamos hacer la predicción de la variable en un momento $t = n$ (al final de la muestra) para $t = n + T$ en el futuro; esto es, para T periodos después del final de la muestra. Como $E\varepsilon_t = 0$, entonces nuestra mejor predicción de ε_t para $t > n$ es $\varepsilon_t = 0$. Llamando \hat{x}_{n+T} la predicción hecha en el tiempo n para $n + T$, tenemos:

$$\begin{aligned}\hat{x}_{n+1} &= c + \alpha_1 x_n + \alpha_2 x_{n-1} + \dots + \alpha_p x_{n+1-p} \\ \hat{x}_{n+2} &= c + \alpha_1 \hat{x}_{n+1} + \alpha_2 x_n + \dots + \alpha_p x_{n+2-p} \\ &\vdots\end{aligned}\tag{3.44}$$

A medida que avanza la predicción hacia el futuro, se utilizan más predicciones anteriores en la predicción misma hasta que, eventualmente, la mejor predicción será la media incondicional de la serie.

Ejemplo

Para el AR(1) $x_t = c + \alpha x_{t-1} + \varepsilon_t$, tendremos las predicciones:

$$\begin{aligned}\hat{x}_{n+1} &= c + \alpha x_n \\ \hat{x}_{n+2} &= c + \alpha \hat{x}_{n+1} = c + \alpha(c + \alpha x_n) = c(1 + \alpha) + \alpha^2 x_n \\ \hat{x}_{n+3} &= c + \alpha [c(1 + \alpha) + \alpha^2 x_n] = c(1 + \alpha + \alpha^2) + \alpha^3 x_n \\ &\vdots \\ \hat{x}_{n+T} &= c(1 + \alpha + \alpha^2 + \alpha^3 + \dots + \alpha^{T-1}) + \alpha^T x_n\end{aligned}$$

Expresión que, para T grande, tiende al valor esperado incondicional de x_t :

$$\hat{x}_{n+T} = \frac{c}{1 - \alpha} = E(x_t)$$

Predicción utilizando un MA(q)

Sea x_t una serie que sigue un MA(q):

$$x_t = c + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q}$$

Deseamos hacer la predicción de la variable en el momento $t = n$ (al final de la muestra) para $t = n + T$ en el futuro. Aquí debemos trabajar con los ε disponibles. Llamando \hat{x}_{n+T} a la predicción hecha en el tiempo n para $n + T$, tenemos:

$$\begin{aligned}\hat{x}_{n+1} &= c + \beta_1 \varepsilon_n + \beta_2 \varepsilon_{n-1} + \dots + \beta_q \varepsilon_{n+1-q} \\ \hat{x}_{n+2} &= c + \beta_2 \varepsilon_n + \beta_3 \varepsilon_{n-1} + \dots + \beta_q \varepsilon_{n+2-q} \\ &\vdots\end{aligned}\tag{3.45}$$

No se pueden utilizar predicciones de ε , porque para ello necesitaríamos conocer la realización en el futuro para poder restar y encontrarla ($x_{n+T} - \hat{x}_{n+T} = \hat{\varepsilon}_{n+T}$), y no tenemos dicha realización. A medida que T se hace mayor, van quedando menos términos de promedio móvil hasta que $\hat{x}_{n+T} = c$ para $T > q$.

Predicción utilizando un ARMA(p,q)

Sea x_t una serie que sigue un modelo ARMA(p, q):

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

Deseamos hacer la predicción de la variable al final de la muestra para $t = n + T$ en el futuro. La forma de predecir en un ARMA es una combinación de lo que se hizo en el AR y el MA. Llamando \hat{x}_{n+T} la predicción de x_{n+T} hecha en el periodo n , tenemos:

$$\begin{aligned}\hat{x}_{n+1} &= c + \alpha_1 x_n + \alpha_2 x_{n-1} + \dots + \alpha_p x_{n+1-p} + \beta_1 \varepsilon_n \\ &\quad + \beta_2 \varepsilon_{n-1} + \dots + \beta_q \varepsilon_{n+1-q} \\ \hat{x}_{n+2} &= c + \alpha_1 \hat{x}_{n+1} + \alpha_2 x_n + \dots + \alpha_p x_{n+2-p} \\ &\quad + \beta_2 \varepsilon_n + \beta_3 \varepsilon_{n-1} + \dots + \beta_q \varepsilon_{n+2-q} \\ &\vdots\end{aligned}\tag{3.46}$$

A medida que avanza la predicción hacia el futuro, quedan menos términos de promedio móvil hasta que éstos desaparecen para $T > q$; a partir de $T > p$ sólo quedan predicciones de x_t en la parte autorregresiva, la cual, como vimos, se aproxima al valor esperado incondicional de x_t para T muy grande.

No debe darse la impresión de que la predicción de un modelo ARMA después de la última observación tiende monotónicamente a la media de x_t . La predicción puede llegar a ser mayor en magnitud que la última observación, antes de converger a la media incondicional de la serie.

Varianza de predicción en modelos ARMA

Como es de esperar, la varianza de predicción tiende a aumentar a medida que el horizonte de predicción se hace más lejano, ya que cuanto más largo sea el plazo de la predicción, más posibilidad hay de acumular errores. Aparentemente, se pueden encontrar excepciones a este principio de que el error tiende a aumentar con el plazo de la predicción. Por ejemplo, hay variables para las cuales se conoce con exactitud (sin error) su nivel en ciertos días del año aunque estén lejanos; la producción de ciertos bienes industriales se puede predecir como cero el día primero de enero; la circulación de prensa escrita también es cero ciertos días cada año. Pero esto realmente no constituye un contraejemplo ya que se refiere a efectos estacionales o predeterminados, de naturaleza determinística.

Cuando el plazo es muy largo, el modelo ya no provee información útil, de manera que la predicción se acerca al valor esperado incondicional de la variable x_t , y la varianza de la predicción a la varianza de x_t .

El siguiente desarrollo se basa parcialmente en Granger (1989, p. 105). Sea $\hat{e}_{n+T} = x_{n+T} - \hat{x}_{n+T}$ el error de la predicción hecha en el periodo n para el periodo $n + T$ y sea el conjunto de información $I_n = x_{n-j}$ para $j = 0, 1, 2, \dots$, esto es, la información disponible para realizar la predicción. Una serie estacionaria x_t puede representarse, según el teorema de Wold, como un promedio móvil:

$$x_t = \sum_{j=0}^{\infty} \beta_j \varepsilon_{t-j} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \beta_3 \varepsilon_{t-3} + \dots \quad (3.47)$$

Con $\beta_0 = 1$ y $E\varepsilon^2 = \sigma_\varepsilon^2$. Suponemos que se conocen los β_j ; de lo contrario, el error de predicción deberá incluir también el efecto de la varianza de la estimación de los β_j en (3.47). La expresión para x_{n+T} será:

$$\begin{aligned} x_{n+T} &= \varepsilon_{n+T} + \beta_1 \varepsilon_{n+T-1} + \beta_2 \varepsilon_{n+T-2} + \dots \\ &\quad + \beta_{T-1} \varepsilon_{n+1} + \beta_T \varepsilon_n + \beta_{T+1} \varepsilon_{n-1} + \dots \\ &= \sum_{j=0}^{T-1} \beta_j \varepsilon_{n+T-j} + \sum_{j=T}^{\infty} \beta_j \varepsilon_{n+T-j} \end{aligned} \tag{3.48}$$

La predicción \hat{x}_{n+T} será el segundo término de la derecha en (3.48), ya que contiene los datos conocidos en $t = n$, y el error de la predicción $\hat{\varepsilon}_{n+T}$ será el primer término de la derecha. En símbolos:

$$\hat{\varepsilon}_{n+T} = x_{n+T} - \hat{x}_{n+T} = x_{n+T} - \sum_{j=T}^{\infty} \beta_j \varepsilon_{n+T-j} = \sum_{j=0}^{T-1} \beta_j \varepsilon_{n+T-j} \tag{3.49}$$

Puede observarse que $\hat{\varepsilon}_{n+T}$ tiene la forma de un MA, esto es, de una combinación lineal de los errores. Por ejemplo, recordando que $\beta_0 = 1$:

$$\begin{aligned} \hat{\varepsilon}_{n+1} &= \varepsilon_{n+1} \\ \hat{\varepsilon}_{n+2} &= \varepsilon_{n+2} + \beta_1 \varepsilon_{n+1} \\ \hat{\varepsilon}_{n+3} &= \varepsilon_{n+3} + \beta_1 \varepsilon_{n+2} + \beta_2 \varepsilon_{n+1} \\ &\vdots \end{aligned}$$

El valor esperado del error de predicción $\hat{\varepsilon}_{n+T}$ es cero y su varianza será:

$$\begin{aligned} \text{var}(\hat{\varepsilon}_{n+T}) &= E \left(\sum_{j=0}^{T-1} \beta_j \varepsilon_{n+T-j} \right)^2 = \sigma_\varepsilon^2 \sum_{j=0}^{T-1} \beta_j^2 \\ &= \sigma_\varepsilon^2 (1 + \beta_1^2 + \beta_2^2 + \dots + \beta_{T-1}^2) \end{aligned} \tag{3.50}$$

Puede verse que:

$$\text{var}(\hat{\varepsilon}_{n+T}) - \text{var}(\hat{\varepsilon}_{n+T-1}) = \sigma_\varepsilon^2 \beta_{T-1}^2 \geq 0 \tag{3.51}$$

De manera que la varianza aumenta, o por lo menos no decrece, a medida que T aumenta. Si suponemos que el error de predicción $\hat{\epsilon}_{n+T}$ tiene una distribución normal, podemos establecer un intervalo de confianza alrededor de la predicción \hat{x}_{n+T} . Para el 95 % este intervalo será:

$$\hat{x}_{n+T} \pm 1.96 \sqrt{var(\hat{\epsilon}_{n+T})} = \hat{x}_{n+T} \pm 1.96\sigma_{\epsilon} \sqrt{1 + \beta_1^2 + \beta_2^2 + \dots + \beta_{T-1}^2} \quad (3.52)$$

Donde los β son los primeros $T - 1$ coeficientes que resultan de estimar o de expresar x_t como un MA. La varianza del error σ_{ϵ}^2 se obtiene de la varianza de los residuos $\hat{\epsilon}_t$ de esta estimación.

Debido a los supuestos y aproximaciones involucradas (básicamente la omisión de los errores de especificación, de estimación y de datos), $var(\hat{\epsilon}_{n+T})$ subestima la verdadera varianza de predicción; sin embargo, es una buena guía.



CAPÍTULO 4

Estimación de modelos ARMA(p,q)

El procedimiento para estimar los modelos autorregresivos AR(p), promedio móvil MA(q), o la combinación de ambos ARMA(p, q), consta básicamente de tres pasos que comprenden: la especificación del modelo (establecer el orden de la parte autorregresiva p y de la parte de promedio móvil q), la estimación propiamente dicha y el diagnóstico y revisión del modelo.

Ocasionalmente, en lugar del ARMA(p, q), se habla del modelo ARIMA(p, d, q), que quiere decir *autorregresivo de promedio móvil integrado* (del inglés *autoregressive integrated moving average*). El ARIMA(p, d, q) incluye la diferenciación de la serie con el objeto de convertirla en estacionaria. Si $d = 1$ quiere decir que fue necesario diferenciarla una vez para volverla estacionaria; si $d = 2$, fue necesario diferenciarla dos veces, etc. La I en la sigla ARIMA se refiere a la palabra integrado, que significa sumado, pues, como el proceso fue diferenciado, hay que sumar o integrar esas variaciones para recuperar la serie original. Granger (1982) sugiere que el ARIMA debió llamarse IARMA, y que el éxito del ARIMA se debe a lo sonoro del acrónimo. En cualquier caso, supongamos que trabajamos con una serie razonablemente estacionaria, de manera que podemos hablar de ARMA(p, q) en general.

En la práctica, los modelos de uso más común son los autorregresivos, con q igual a cero, y en segundo lugar los ARMA con algunos términos p y q . Los modelos menos comunes son los de promedio móvil puro, con p igual a cero. También es usual incluir una constante en el modelo.

Debe tenerse en cuenta que en la práctica las cosas no son tan claras como en la teoría y que no siempre será posible determinar con exactitud el orden p y q . La exploración inicial de los datos suele sugerir dos o tres modelos plausibles, los cuales se pueden estimar para luego escoger aquel que parezca más apropiado. En estos casos de ambigüedad, frecuentes en el trabajo estadístico, cuando las pruebas efectuadas correcta y objetivamente no son concluyentes, es necesario recurrir a la intuición y a cualquier información externa al modelo con el objeto de decidir sobre su especificación. Más aún, el procedimiento para la escogencia del número de rezagos, o de otras características de la especificación, debe estar guiado por alguna base teórica o conocimiento del problema; por ejemplo, considerar el rezago 4 o sus múltiplos cuando se utilicen datos trimestrales, o el rezago 12 cuando se trata de una serie mensual.

Una vez realizada la estimación se utilizan varias herramientas de diagnóstico y chequeo conocidas en la econometría, a fin de ver si el modelo es bueno, comparar y elegir entre modelos u obtener sugerencias sobre modificaciones que puedan mejorar la estimación. Los objetivos usuales son buscar la menor suma de residuos al cuadrado o la mayor verosimilitud, procurar que los coeficientes estimados sean significativos y determinar que no haya problemas de autocorrelación o correlación serial en los residuos \hat{e}_t —por ejemplo, observando la función de autocorrelación (ACF) o la prueba de Ljung-Box—, ya que lo contrario podría indicar falta de alguna variable o rezago en el modelo. Para una serie de tamaño n la prueba de Ljung-Box, también conocida como prueba Q , tiene la siguiente expresión:

$$Q = n(n + 2) \sum_{\tau=1}^s \frac{[\hat{r}(\tau)]^2}{n - \tau} \quad (4.1)$$

Distribuida, bajo la hipótesis-nula de ruido blanco (independencia lineal), como una variable chi-cuadrado χ_s^2 con s grados de libertad, donde s es el número de rezagos que se incluyen en la prueba. Esto se puede ver intuitivamente suponiendo n grande y recordando que

para ruido blanco $\hat{r}(\tau) \xrightarrow{d} N(0, 1/n)$. Si la serie probada corresponde a los residuos de un modelo ARMA, entonces los grados de libertad se ajustan por el número de parámetros estimados, $s - p - q$ sin constante y $s - p - q - 1$ con constante. Es importante que el parámetro alcance a cubrir posibles estacionalidades; por ejemplo, mayor que rezago 4 si los datos son trimestrales, mayor que rezago 12 si los datos son mensuales, etc. Si Q es pequeño es porque las autocorrelaciones son pequeñas e insignificantes y se acepta que la serie bajo prueba es ruido blanco.

En general, estas pruebas de correlación serial, llamadas portmanteau (por el género de cine donde se compilan varios cortos inconexos), deben usarse con precaución. Las conclusiones de la prueba Q pueden verse afectadas cuando las observaciones provienen de distribuciones no estacionarias o muy diferentes de la gaussiana o cuando hay ciertas no linealidades como cambios de distribución a lo largo de la serie de tiempo. En estos casos la χ^2_s no es una buena aproximación a la distribución de Q bajo la hipótesis nula. Para otras advertencias sobre la prueba Q , véase Dezhbakhsh (1990) y Kennedy (1992, p. 261). Por ejemplo, si al probar los residuos éstos no dan ruido blanco, puede ser que el modelo no sea lineal, entonces debe modelarse la no linealidad en lugar de adicionar más rezagos al ARIMA.

Una prueba del tipo multiplicador de Lagrange para detectar autocorrelación en los residuos es la de Breusch-Godfrey. Con ésta se corre la regresión original y se guardan los residuos estimados. Luego se corre una regresión auxiliar de los residuos estimados contra todas las variables explicativas del modelo original (incluida una constante) más h rezagos de los residuos. Bajo la hipótesis nula de no autocorrelación por lo menos hasta de orden h , el estadístico de prueba nR^2 , construido con el R^2 de la regresión auxiliar, estará asintóticamente distribuido como una chi-cuadrado χ^2_h con h grados de libertad. Si el nR^2 excede el valor crítico, se rechaza la hipótesis nula. También se puede emplear un estadístico F para probar la significancia de los h rezagos. Si las variables explicativas de la regresión original son estrictamente exógenas, pueden omitirse de la regresión auxiliar (véase Woolridge, 2009, capítulo 12).

Otra prueba, conocida como *BDS* (véase Brock *et al.*, 1996), se efectúa bajo la hipótesis nula de que x_t es una serie idéntica e independientemente distribuida (independencia total, no sólo lineal). Bajo

para ruido blanco $\hat{r}(\tau) \xrightarrow{d} N(0, 1/n)$. Si la serie probada corresponde a los residuos de un modelo ARMA, entonces los grados de libertad se ajustan por el número de parámetros estimados, $s - p - q$ sin constante y $s - p - q - 1$ con constante. Es importante que el parámetro alcance a cubrir posibles estacionalidades; por ejemplo, mayor que rezago 4 si los datos son trimestrales, mayor que rezago 12 si los datos son mensuales, etc. Si Q es pequeño es porque las autocorrelaciones son pequeñas e insignificantes y se acepta que la serie bajo prueba es ruido blanco.

En general, estas pruebas de correlación serial, llamadas portmanteau (por el género de cine donde se compilan varios cortos inconexos), deben usarse con precaución. Las conclusiones de la prueba Q pueden verse afectadas cuando las observaciones provienen de distribuciones no estacionarias o muy diferentes de la gaussiana o cuando hay ciertas no linealidades como cambios de distribución a lo largo de la serie de tiempo. En estos casos la χ^2 no es una buena aproximación a la distribución de Q bajo la hipótesis nula. Para otras advertencias sobre la prueba Q , véase Dezhbakhsh (1990) y Kennedy (1992, p. 261). Por ejemplo, si al probar los residuos éstos no dan ruido blanco, puede ser que el modelo no sea lineal, entonces debe modelarse la no linealidad en lugar de adicionar más rezagos al ARIMA.

Una prueba del tipo multiplicador de Lagrange para detectar autocorrelación en los residuos es la de Breusch-Godfrey. Con ésta se corre la regresión original y se guardan los residuos estimados. Luego se corre una regresión auxiliar de los residuos estimados contra todas las variables explicativas del modelo original (incluida una constante) más h rezagos de los residuos. Bajo la hipótesis nula de no autocorrelación por lo menos hasta de orden h , el estadístico de prueba nR^2 , construido con el R^2 de la regresión auxiliar, estará asintóticamente distribuido como una chi-cuadrado χ_h^2 con h grados de libertad. Si el nR^2 excede el valor crítico, se rechaza la hipótesis nula. También se puede emplear un estadístico F para probar la significancia de los h rezagos. Si las variables explicativas de la regresión original son estrictamente exógenas, pueden omitirse de la regresión auxiliar (véase Woolridge, 2009, capítulo 12).

Otra prueba, conocida como *BDS* (véase Brock *et al.*, 1996), se efectúa bajo la hipótesis nula de que x_t es una serie idéntica e independientemente distribuida (independencia total, no sólo lineal). Bajo

esta hipótesis, para cualquier par de puntos de la serie, x_t y x_s , donde $t \neq s$, será cierta la siguiente igualdad de probabilidades:

$$\begin{aligned} P(|x_t - x_s| < \delta, |x_{t-1} - x_{s-1}| < \delta, \dots, |x_{t-m+1} - x_{s-m+1}| < \delta) \\ = [P(|x_\epsilon - x_s| < \delta)]^m, \end{aligned}$$

donde $\delta > 0$ y $m > 0$. El estadístico de prueba, que para cada m se calcula sobre $t = 1, \dots, n$, y todo s , es proporcional a la diferencia de las probabilidades del lado derecho y del lado izquierdo de la expresión anterior, y tiene una distribución asintótica Normal estándar. Si el estadístico de prueba es significativamente diferente de cero, se rechaza la hipótesis nula. De manera heurística imaginemos, por ejemplo, que tenemos los eventos A, B, C y D , independientes entre sí e igualmente probables, entonces

$$P(A \cap B \cap C \cap D) = [P(A)]^4.$$

La función de autocorrelación y la función de autocorrelación parcial

Quizás la herramienta más utilizada para la especificación del modelo es la combinación de la ACF y la función de autocorrelación parcial PACF. Como se vio en el capítulo anterior, los estimados de estas funciones tienen características que ayudan a identificar si el modelo es autorregresivo, de promedio móvil, o ambos, y cuáles sus posibles rezagos p y q . La práctica y la experiencia son fundamentales en el manejo de la ACF y la PACF. Algunas pautas, adicionales a las presentadas en el capítulo anterior, se pueden consultar en Box y Jenkins (1970) y en Enders (2004).

Un criterio general para la especificación del modelo ARMA es que entre alternativas similares es preferible un modelo *parsimonioso* (parco, frugal), esto es, el que tenga menos términos, con el objeto de minimizar la pérdida de precisión en la estimación y en la predicción.

Criterios de información de Akaike y Schwarz

Existen otras ayudas para el diagnóstico y la especificación como el criterio de información de Akaike (AIC) y el criterio de información de

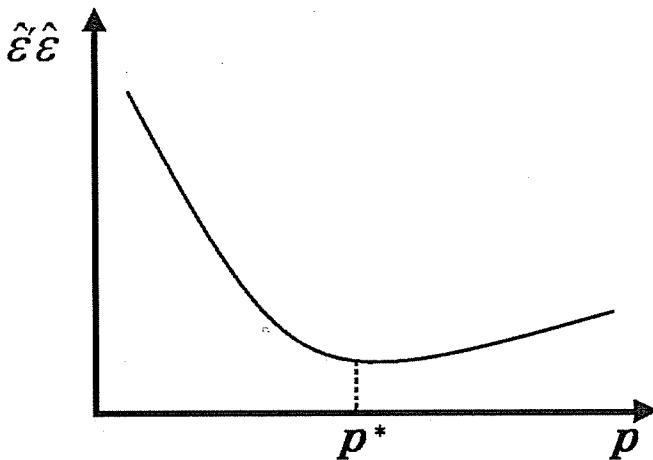
Schwarz (SC), este último también conocido como *criterio de información bayesiana* (BIC). Ambos criterios se basan en el mismo principio. Intuitivamente, es como correr el modelo para diferentes p y q buscando la combinación que dé la más baja suma de errores al cuadrado. Por ejemplo, para un AR se podría ilustrar como se muestra en la Figura 4.1, donde p^* es el rezago escogido porque produce la menor suma de errores al cuadrado. El AIC tiene la siguiente expresión:

$$AIC = \ln \left(\frac{\hat{\varepsilon}' \hat{\varepsilon}}{n} \right) + \frac{2k}{n} \quad (4.2)$$

Y el SC, la siguiente expresión:

$$SC = \ln \left(\frac{\hat{\varepsilon}' \hat{\varepsilon}}{n} \right) + \frac{k \ln(n)}{n} \quad (4.3)$$

Figura 4.1: Suma de errores al cuadrado *versus* rezagos



Donde k es el número agregado de parámetros independientes ensayados que cuentan la constante, si la hay. En otras palabras k es la suma del número de rezagos autorregresivos p y de promedio móvil q ensayados, que son los α y los β , más la constante. Para hacer la prueba hay que perder las primeras observaciones de manera que n es el número mínimo utilizable de observaciones. Es de advertir que pueden

encontrarse variaciones de las expresiones anteriores de un texto a otro, pero el principio sigue siendo el mismo.

La idea es buscar k (una combinación de p y q) para que estos criterios den lo menor posible (incluso pueden ser negativos). Obviamente, los rezagos adicionados deben ser significativos, ya que si agregamos rezagos no significativos al modelo, aumenta k sin que disminuya la suma de errores al cuadrado. Es conocido que el AIC tiende a sobreestimar el verdadero rezago. Para subsanar esto, el SC impone una mayor penalización que en AIC, a medida que se incluyen más rezagos.

Estimación de los modelos AR(p)

En esta sección supondremos que ya se estableció que el modelo correcto es un autorregresivo y que se conoce su orden p . El modelo que se va a estimar es:

$$x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} = \varepsilon_t \quad (4.4)$$

donde ε_t es ruido blanco y se ha omitido la constante por simplicidad. En otras palabras, supondremos que la serie x_t tiene media cero o, en su defecto, que le ha sido restada.

Hay varias formas de realizar la estimación de un AR(p), obviamente unas mejores que otras y unas más complicadas o costosas. Básicamente, los métodos disponibles son: de momentos, mínimos cuadrados ordinarios y máxima verosimilitud.

Estimación por el método de momentos de Yule-Walker

Aunque este es el método menos usado, su explicación es breve y resulta muy ilustrativa. Multiplicando la ecuación (4.4) por $x_{t-\tau}$ y tomando expectativas tendremos, en términos de autocovarianzas estimadas:

$$\hat{R}(\tau) + \hat{\alpha}_1 \hat{R}(\tau-1) + \dots + \hat{\alpha}_p \hat{R}(\tau-p) = 0 \quad \text{para } \tau = 1, 2, \dots, p \quad (4.5)$$

La ecuación anterior también podría expresarse en términos de las autocorrelaciones. El rezago τ varía de 1 hasta p ya que sólo se necesitan p ecuaciones para estimar los p parámetros; sin embargo,

en general, podría ser $\tau > p$. También se podría, si se quiere, utilizar $\tau = 0$ pero en ese caso habría que estimar el término σ_ε^2 que resulta de $E\varepsilon_t^2$.

Se estiman las autocovarianzas y a partir de (4.5) se forman las p ecuaciones simultáneas; luego se resuelve este sistema y se despejan los p coeficientes $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$ que conforman el autorregresivo (4.4).

Ejemplo

Para un AR(1) $x_t + \alpha_1 x_{t-1} = \varepsilon_t$ se requiere una ecuación ($\tau = 1 = p$):

$$\hat{R}(1) + \hat{\alpha}_1 \hat{R}(0) = 0 \implies \hat{\alpha}_1 = -\frac{\hat{R}(1)}{\hat{R}(0)}$$

Para un AR(2) se requieren dos ecuaciones, porque hay dos incógnitas α_1, α_2 :

$$\begin{aligned}\tau = 1 \quad &\hat{R}(1) + \hat{\alpha}_1 \hat{R}(0) + \hat{\alpha}_2 \hat{R}(-1) = 0 \\ \tau = 2 \quad &\hat{R}(2) + \hat{\alpha}_1 \hat{R}(1) + \hat{\alpha}_2 \hat{R}(0) = 0\end{aligned}$$

Los demás términos de cada ecuación son cero, porque $\alpha_3 = \alpha_4 = \dots = 0$.

Estimación por mínimos cuadrados ordinarios

Uno de los métodos más populares para estimar el AR(p) es por mínimos cuadrados ordinarios (MCO). Como siempre, la cantidad que se va a minimizar es la suma de errores al cuadrado, cuya expresión es:

$$\sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p})^2 \quad (4.6)$$

Donde la sumatoria arranca en $p + 1$, porque no podemos calcular los ε entre ε_1 y ε_p . Esto se puede ver de la expresión del AR(p), $\varepsilon_t = x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p}$, donde:

$$\text{para } t = 1 \quad \varepsilon_1 = x_1 + \alpha_1 x_0 + \alpha_2 x_{-1} + \dots + \alpha_p x_{1-p}$$

⋮

$$\text{para } t = p \quad \varepsilon_p = x_p + \alpha_1 x_{p-1} + \alpha_2 x_{p-2} + \dots + \alpha_p x_0$$

Los rezagos de x_t utilizados como variables explicativas estarán parcialmente correlacionados con el vector de errores ε_t , lo cual produce estimadores sesgados, aunque *consistentes* según el teorema de Mann-Wald, que veremos más adelante. Para explicar por qué la aplicación de MCO a un modelo autorregresivo produce estimadores sesgados pero consistentes, podemos utilizar un AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$. Vectorialmente, el AR(1) se puede expresar como:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \alpha + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (4.7)$$

Si definimos $x = (x_1, x_2, \dots, x_n)'$, $X = (x_0, x_1, \dots, x_{n-1})'$ y $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, podemos escribir el modelo en forma compacta como $x = X\alpha + \varepsilon$. La aplicación de MCO produce la siguiente expresión para el vector de coeficientes:

$$\hat{\alpha} = (X'X)^{-1}X'x = (X'X)^{-1}X'X\alpha + (X'X)^{-1}X'\varepsilon = \alpha + (X'X)^{-1}X'\varepsilon \quad (4.8)$$

Por construcción del AR(1), vemos que el error ε_t , en el tiempo t , es independiente del pasado x_{t-1}, x_{t-2}, \dots pero *no* es independiente de x_t ni del futuro x_{t+1}, x_{t+2}, \dots . Por ejemplo, el efecto de ε_1 se nota en el término x_2 a través de x_1 ya que:

$$x_2 = \alpha x_1 + \varepsilon_2 = \alpha(\alpha x_0 + \varepsilon_1) + \varepsilon_2$$

Igualmente en el término x_3 , porque:

$$x_3 = \alpha x_2 + \varepsilon_3 = \alpha(\alpha x_1 + \varepsilon_2) + \varepsilon_3 = \alpha(\alpha(\alpha x_0 + \varepsilon_1) + \varepsilon_2) + \varepsilon_3$$

Y así sucesivamente. En consecuencia, el valor esperado de (4.8) será:

$$E\hat{\alpha} = E[(X'X)^{-1}X'\varepsilon] \neq \alpha \quad (4.9)$$

Esto porque $E(\varepsilon/X) \neq 0$ y viola el supuesto de independencia entre las variables explicativas y el error, lo cual implica que el estimador MCO es sesgado, $E\hat{\alpha} \neq \alpha$ (recordamos que si las variables a y b no

son independientes entonces $Eab \neq EaEb$). De hecho, puede mostrarse (por ejemplo, véase Marriott y Pope, 1954) que el sesgo en el coeficiente estimado es hacia cero; esto es, $E|\hat{\alpha}| < |\alpha|$.

La situación anterior contrasta con el modelo clásico de regresión, en el cual X es tal que se cumple con la independencia $E(\varepsilon/X) = E\varepsilon = 0$, y, por lo tanto, en ese caso:

$$E(X'X)^{-1}X'\varepsilon = (X'X)^{-1}X'E\varepsilon = 0 \quad (4.10)$$

Contrasta también con el caso en el cual X es estocástica, pero independiente de ε , de manera que:

$$E(X'X)^{-1}X'\varepsilon = E(X'X)^{-1}X'E\varepsilon = 0 \quad (4.11)$$

En el caso de los AR(p), X es estocástica pero no es independiente de ε . Veamos con mayor detalle el término que produce el sesgo en (4.9) para este AR(1):

$$\begin{aligned} E(X'X)^{-1}X'\varepsilon &= E \left[(x_0 \ x_1 \dots x_{n-1}) \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{n-1} \end{bmatrix} \right]^{-1} (x_0 \ x_1 \dots x_{n-1}) \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ &= E \left[\frac{1}{x_0^2 + x_1^2 + x_2^2 + \dots + x_{n-1}^2} \right] [x_0\varepsilon_1 + x_1\varepsilon_2 + x_2\varepsilon_3 + \dots + x_{n-1}\varepsilon_n] \end{aligned} \quad (4.12)$$

La ecuación (4.12) representa el valor esperado de un producto. Sería igual al valor esperado del primer factor multiplicado por el valor esperado del segundo si los dos factores fueran independientes, en cuyo caso es evidente que el sesgo sería cero porque el valor esperado del segundo factor es cero debido a que contiene productos de x contra ε del futuro, esto es:

$$E[x_0\varepsilon_1 + x_1\varepsilon_2 + x_2\varepsilon_3 + x_3\varepsilon_4 + \dots + x_{n-1}\varepsilon_n] = 0 \quad (4.13)$$

Mientras menor sea la dependencia entre ambos factores, más cerca estará (4.12) de expresarse como un producto de valores esperados, y más cerca estará de cero.

El caso de un AR(2) $x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t$ sirve como ilustración adicional. La expresión matricial de este modelo sería:

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_0 & x_{-1} \\ x_1 & x_0 \\ \vdots & x_1 \\ x_{n-1} & x_{n-2} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Al igual que se hizo para el AR(1), la estimación del vector de coeficientes es:

$$\hat{\alpha} = \alpha + (X'X)^{-1}X'\varepsilon \quad (4.14)$$

En la anterior estimación, el sesgo es $(X'X)^{-1}X'\varepsilon$, donde:

$$(X'X)^{-1} = \begin{bmatrix} \frac{x_{-1}^2 + \dots + x_{n-2}^2}{\Delta} & \frac{-(x_0x_{-1} + x_1x_0 + \dots + x_{n-1}x_{n-2})}{\Delta} \\ \frac{-(x_0x_{-1} + x_1x_0 + \dots + x_{n-1}x_{n-2})}{\Delta} & \frac{x_0^2 + \dots + x_{n-1}^2}{\Delta} \end{bmatrix} \quad (4.15)$$

$$X'\varepsilon = \begin{bmatrix} x_0\varepsilon_1 + x_1\varepsilon_2 + x_2\varepsilon_3 + \dots + x_{n-1}\varepsilon_n \\ x_{-1}\varepsilon_1 + x_0\varepsilon_2 + x_1\varepsilon_3 + \dots + x_{n-2}\varepsilon_n \end{bmatrix} \quad (4.16)$$

Con el determinante:

$$\Delta = (x_0^2 + \dots + x_{n-1}^2)(x_{-1}^2 + \dots + x_{n-2}^2) - (x_0x_{-1} + x_1x_0 + \dots + x_{n-1}x_{n-2})^2.$$

Las expresiones (4.15) y (4.16) no son independientes, razón por la cual, aunque $EX'\varepsilon$ sea cero, habrá un sesgo en muestras finitas. Afortunadamente, dicho sesgo desaparece asintóticamente, según el teorema de Mann-Wald. También se han hecho intentos por estimar el sesgo con el objeto de mejorar la estimación en muestras finitas (véase, por ejemplo, Patterson, 2000).

Intuición del teorema de Mann-Wald

Supongamos, para simplificar, que tenemos el producto de dos variables ab y queremos encontrar su valor esperado Eab . Si se tratara de la suma $a + b$ simplemente tomamos el valor esperado de cada término, $Ea + Eb$, porque $a + b$ es una operación lineal. Pero ab

no es lineal y, por lo tanto, $Eab \neq EaEb$ a menos que a y b sean independientes. El problema con la aplicación de MCO al modelo autorregresivo es precisamente éste, que los factores, por ejemplo (4.15) y (4.16), no son independientes.

No siempre es intuitivamente obvio por qué $Eab \neq EaEb$. Imaginemos que a está distribuida de manera simétrica con media 10 y que b también está distribuida de manera simétrica pero con media cero. Si fueran independientes $Eab = EaEb = (10)(0) = 0$; sin embargo, supongamos que entre a y b hay correlación positiva; esto es, cuando a está por encima de su media, b también.

Por ejemplo, si tenemos dos pares de observaciones $(a_1, b_1) = (15, 1)$ y $(a_2, b_2) = (5, -1)$, el primer producto será 15 y el segundo -5 , y la media de los productos será 5, lejos de cero. Y aunque adicionemos más observaciones correlacionadas, la media de sus productos seguirá siendo diferente de cero, a menos que la varianza de b se haga cero, en cuyo caso, obviamente, de ahí en adelante todas las observaciones que se tomen de b serán iguales a su media, cero, al igual que los productos ab y la media de éstos, Eab . Esto es precisamente lo que quiere decir límite en probabilidad, $plim$. Entonces, si $plim b = 0$, podremos decir que $plim ab = (plim a)(plim b) = 0$, siempre y cuando el límite $plim a$ también exista (véase, por ejemplo, Greene, 1997, p. 118).

Volviendo a la expresión del sesgo vemos que éste desaparecería, al menos asintóticamente, si $plim (X'X)^{-1}X'\varepsilon = 0$. Al dividir arriba y abajo el segundo término de (4.14) por n obtenemos:

$$\hat{\alpha} = \alpha + \left(\frac{X'X}{n} \right)^{-1} \left(\frac{X'\varepsilon}{n} \right) \quad (4.17)$$

Y luego tomando límite de probabilidad vemos que:

$$\begin{aligned} plim \hat{\alpha} &= \alpha + \left[plim \left(\frac{X'X}{n} \right) \right]^{-1} \left[plim \left(\frac{X'\varepsilon}{n} \right) \right] \\ &= \alpha + \left[plim \left(\frac{X'X}{n} \right) \right]^{-1} (0) = \alpha \end{aligned} \quad (4.18)$$

Se divide arriba y abajo por n porque $plim X'\varepsilon$ no existe. Aunque el valor esperado de $X'\varepsilon$ es cero, su varianza no cae a cero; pero al dividir por n su varianza desaparece cuando $n \rightarrow \infty$, de manera que

el límite existe y es $\text{plim} \left(\frac{X'\varepsilon}{n} \right) = 0$. Adicionalmente, como la matriz $X'X$ podría ser singular, suponemos que $\text{plim} \left(\frac{X'X}{n} \right) = Q$, donde Q es una matriz finita no singular (para mayores detalles, véase Judge et al., 1985, p. 164).

Y si hubiera correlación serial en ε_t ? En este caso tendríamos que $EX'\varepsilon \neq 0$, porque en un autorregresivo los sumandos de $x_0\varepsilon_1 + x_1\varepsilon_2 + x_2\varepsilon_3 + x_3\varepsilon_4 + \dots$ tienden a ser del mismo signo al estar los x del pasado relacionados con los ε del futuro a través de la correlación serial. En consecuencia, $\text{plim} \left(\frac{X'\varepsilon}{n} \right) \neq 0$ y MCO sería sesgado e inconsistente.

Teorema de Mann-Wald

Formalizamos ahora este teorema, debido a Mann y Wald (1943). Para un AR(p) estacionario con $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$, la estimación por MCO es consistente y:

$$\sqrt{n}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \sigma_\varepsilon^2 Q^{-1}) \quad (4.19)$$

Donde el símbolo \xrightarrow{d} quiere decir *converge en distribución*, y donde $Q = \text{plim} \left(\frac{X'X}{n} \right)$. Alternativamente, se puede usar la expresión:

$$\hat{\alpha} \xrightarrow{d} N(\alpha, \sigma_\varepsilon^2 Q^{-1}/n) \quad (4.20)$$

En otras palabras, el estimador MCO de α converge en distribución a una normal, y tanto el estimador como las pruebas de hipótesis serán válidas asintóticamente. En este sentido, un AR(p) puede ser tratado, para n grande, como un caso de regresión clásica. Sin embargo, es importante notar que si el verdadero modelo es un ARMA, lo cual implica autocorrelación en sus errores, estimarlo por MCO como si fuera un AR tiende a producir resultados inconsistentes, especialmente si se utilizan muy pocos rezagos.

Ejemplo

Para un AR(1) estacionario $x_t = \alpha x_{t-1} + \varepsilon_t$, la expresión de la varianza de $\hat{\alpha}$, $\sigma_\varepsilon^2 \text{plim}(X'X)^{-1}$ en (4.20), se construye, para grande (véase Greene, 1997, p. 848):

$$\begin{aligned}\sigma_{\varepsilon}^2 [(x_0^2 + x_1^2 + \dots + x_{n-1}^2)]^{-1} &= \sigma_{\varepsilon}^2 [nVar(x_t)]^{-1} \\ &= \sigma_{\varepsilon}^2 \left(n \frac{\sigma_{\varepsilon}^2}{1 - \alpha^2} \right)^{-1} = \frac{1 - \alpha^2}{n}\end{aligned}$$

Mínimos cuadrados ordinarios, autorregresiones y autocovarianzas

Es ilustrativo mostrar cómo la estimación del vector α de un AR(p) por MCO se puede interpretar en términos de sus autocovarianzas. Sea el AR(p):

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t \quad (4.21)$$

Expresado como:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ x_1 & \cdot & \cdot \\ x_2 & x_1 & \cdot \\ \vdots & \vdots & \vdots \\ x_{n-2} & x_{n-3} & \dots & x_{n-p-1} \\ x_{n-1} & x_{n-2} & \dots & x_{n-p} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \quad (4.22)$$

$$x = X_{-1}\alpha + \varepsilon \implies (X'_{-1}X_{-1})^{-1}X'_{-1}x \quad (4.23)$$

donde definimos X_{-1} como la matriz de datos, también llamada matriz de diseño. Si para n grande el efecto de las observaciones faltantes es despreciable, podemos aproximar los componentes de (4.23) en términos de sus autocovarianzas. El término $X'_{-1}X_{-1}$ lo podemos aproximar por:

$$\begin{aligned}
 X'_{-1}X_{-1} &= \begin{bmatrix} \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} & \dots & \sum x_{t-1}x_{t-p} \\ \sum x_{t-2}x_{t-1} & \sum x_{t-2}^2 & \dots & \sum x_{t-2}x_{t-p} \\ \vdots & & & \vdots \\ \sum x_{t-p}x_{t-1} & \sum x_{t-p}x_{t-2} & \dots & \sum x_{t-p}^2 \end{bmatrix} \\
 &= n \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & & \ddots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix} \tag{4.24}
 \end{aligned}$$

Y el término $X'_{-1}x$, por:

$$X'_{-1}x = \begin{bmatrix} \sum x_{t-1}x_t \\ \sum x_{t-2}x_t \\ \vdots \\ \sum x_{t-p}x_t \end{bmatrix} = n \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \tag{4.25}$$

Por lo tanto, la expresión del vector de coeficientes estimados, en términos de sus autocovarianzas, será, de (4.23):

$$\hat{\alpha} = \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & & \ddots & \vdots \\ R(p-1) & \dots & R(0) \end{bmatrix}^{-1} \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \tag{4.26}$$

Pero, ¿qué ocurre si, a diferencia del caso anterior en (4.21), corremos x_t no contra su pasado, sino contra observaciones futuras? El AR(p) en términos del futuro sería:

$$x_t = \gamma_1 x_{t+1} + \gamma_2 x_{t+2} + \dots + \gamma_p x_{t+p} + u_t \tag{4.27}$$

Expresable como:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_{n-2} \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} x_2 & x_3 & \dots & x_{1+p} \\ x_3 & x_4 & \dots & x_{2+p} \\ \vdots & \vdots & & \vdots \\ & & & x_n \\ x_{n-1} & x_n & & \cdot \\ x_n & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad (4.28)$$

$$x = X_{+1}\gamma + u \implies \hat{\gamma} = (X'_{+1}X_{+1})^{-1}X'_{+1}x \quad (4.29)$$

Donde X_{+1} es la matriz de datos. Si, como en el caso anterior, suponemos que para n grande el efecto de las observaciones faltantes es despreciable, podemos aproximar los componentes de (4.29) en términos de las autocovarianzas. El término $X'_{+1}X_{+1}$ lo podemos aproximar por:

$$\begin{aligned} X'_{+1}X_{+1} &= \begin{bmatrix} \sum x_{t+1}^2 & \sum x_{t+1}x_{t+2} & \dots & \sum x_{t+1}x_{t+p} \\ \sum x_{t+2}x_{t+1} & \sum x_{t+2}^2 & \dots & \sum x_{t+2}x_{t+p} \\ \vdots & & & \vdots \\ \sum x_{t+p}x_{t+1} & \dots & & \sum x_{t+p}^2 \end{bmatrix} \\ &= n \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \ddots & & \vdots \\ R(p-1) & \dots & & R(0) \end{bmatrix} \end{aligned} \quad (4.30)$$

Y el término $X'_{+1}x$, por:

$$X'_{+1}x = \begin{bmatrix} \sum x_{t+1}x_t \\ \sum x_{t+2}x_t \\ \vdots \\ \sum x_{t+p}x_t \end{bmatrix} = n \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \quad (4.31)$$

Por lo tanto, la expresión del vector de coeficientes estimados, en términos de sus autocovarianzas, será, de (4.29):

$$\hat{\gamma} = \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & & \ddots & \vdots \\ R(p-1) & & \dots & R(0) \end{bmatrix}^{-1} \begin{bmatrix} R(1) \\ R(2) \\ \vdots \\ R(p) \end{bmatrix} \quad (4.32)$$

Al comparar (4.26) y (4.32), concluimos que, para n grande, correr x_t contra su pasado o contra su futuro arroja los mismos coeficientes, $\hat{\alpha} = \hat{\gamma}$. En otras palabras, los dos siguientes modelos tienen la misma ACF:

$$\begin{aligned} x_t &= \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t \\ x_t &= \alpha_1 x_{t+1} + \alpha_2 x_{t+2} + \dots + \alpha_p x_{t+p} + u_t \end{aligned}$$

Ejemplo

Para n grande, estímese por MCO los dos AR(1) siguientes y compare los resultados:

$$x_t = \alpha x_{t-1} + \varepsilon_t \quad y \quad x_t = \gamma x_{t+1} + u_t$$

Para el primero, donde x_t está en función del pasado, se encuentra:

$$\hat{\alpha} = \left[\begin{bmatrix} \cdot & & & \\ x_1 & x_2 & \dots & x_{n-1} \end{bmatrix} \begin{bmatrix} \cdot \\ x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{bmatrix} \right]^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$\hat{\alpha} = \frac{x_1 x_2 + x_2 x_3 + \dots + x_{n-1} x_n}{x_1^2 + x_2^2 + \dots + x_{n-1}^2} = \frac{\hat{R}(1)}{\hat{R}(0)}$$

Para el segundo, donde x_t está en función del futuro, se encuentra:

$$\hat{\gamma} = \left[\begin{bmatrix} x_2 & x_3 & \dots & x_n & \cdot \end{bmatrix} \begin{bmatrix} x_2 \\ x_3 \\ \vdots \\ x_n \\ \cdot \end{bmatrix} \right]^{-1} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

$$\hat{\gamma} = \frac{x_2x_1 + x_3x_2 + \dots + x_nx_{n-1}}{x_2^2 + x_3^2 + \dots + x_n^2} = \frac{\hat{R}(1)}{\hat{R}(0)}$$

De manera que, para $n \rightarrow \infty$, los coeficientes estimados serán iguales, $\hat{\alpha} = \hat{\gamma}$.

Estimación por máxima verosimilitud condicional

Como punto de partida, suponemos que los errores ε_t de la expresión autorregresiva AR(p) son ruido blanco gaussiano con media cero y matriz de varianza covarianza ($E\varepsilon\varepsilon' = \sigma_\varepsilon^2 I$). Tal como se explicó para la ecuación (4.6), no podemos construir los primeros p errores $\varepsilon_1, \dots, \varepsilon_p$, ya que no disponemos de las observaciones x_{1-p}, \dots, x_0 necesarias para calcularlos. En consecuencia, suponemos dadas las observaciones x_1, \dots, x_p de manera que podemos escribir la verosimilitud conjunta con los errores disponibles $\varepsilon_{p+1}, \dots, \varepsilon_n$, la cual será:

$$f(\varepsilon_{p+1}, \dots, \varepsilon_n) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{n-p}{2}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^n \varepsilon_t^2 \right) \quad (4.33)$$

En razón de que la variable de interés es x_t , porque está relacionada con los α_i , podemos cambiar las variables $\varepsilon_{p+1}, \dots, \varepsilon_n$ por x_{p+1}, \dots, x_n . Esta transformación se explica a continuación; afortunadamente, resulta sencilla dado que el determinante del jacobiano es unitario. En realidad, podríamos escribir (4.33) para todos los errores $\varepsilon_1, \dots, \varepsilon_n$, pero luego no podríamos hacer la transformación a x_1, \dots, x_n por carecer de x_0, \dots, x_{1-p} . En su lugar, se construye x_{p+1}, \dots, x_n suponiendo dados x_1, \dots, x_p .

Un cambio o transformación de variables de $\varepsilon_{p+1}, \dots, \varepsilon_n$, con distribución $f(\varepsilon_{p+1}, \dots, \varepsilon_n)$, a variables x_{p+1}, \dots, x_n , donde cada una es función única de $\varepsilon_{p+1}, \dots, \varepsilon_n$ tal que la transformación inversa exista, esto es, que a su vez se pueda expresar cada ε_t como una función única de x_{p+1}, \dots, x_n (lo cual necesita que x_1, \dots, x_p sean dados), tendrá la siguiente función de distribución para x_{p+1}, \dots, x_n (una explicación intuitiva se encuentra en Johnston, 1984, p. 535).

$$f_x(x_{p+1}, \dots, x_n) = f[\varepsilon_{p+1}(x_{p+1}, \dots, x_n), \dots, \varepsilon_n(x_{p+1}, \dots, x_n)] |\det J| \quad (4.34)$$

Donde $|\det J|$ es el valor absoluto del determinante del jacobiano J :

$$J = \begin{bmatrix} \frac{\partial \varepsilon_{p+1}}{\partial x_{p+1}} & \frac{\partial \varepsilon_{p+1}}{\partial x_{p+2}} & \dots & \frac{\partial \varepsilon_{p+1}}{\partial x_n} \\ \frac{\partial \varepsilon_{p+2}}{\partial x_{p+1}} & \frac{\partial \varepsilon_{p+2}}{\partial x_{p+2}} & \dots & \frac{\partial \varepsilon_{p+2}}{\partial x_n} \\ \vdots & \ddots & \ddots & \vdots \\ \frac{\partial \varepsilon_n}{\partial x_{p+1}} & \dots & \frac{\partial \varepsilon_n}{\partial x_n} \end{bmatrix} \quad (4.35)$$

En el caso del AR(p) considerado, $\varepsilon_t = x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p}$, el Jacobiano (4.35) tendrá valores unitarios en su diagonal ya que sus elementos son derivadas de ε_i con respecto a x_i y cero arriba de la diagonal. Se sabe que el determinante de una matriz triangular o diagonal es igual al producto de los elementos de su diagonal, por lo tanto, $|\det J| = 1$. En consecuencia, la función de distribución de x_{p+1}, \dots, x_n es igual a la de $\varepsilon_{p+1}, \dots, \varepsilon_n$, sólo que los ε se expresan en términos de los x . Con esta transformación de variables, la verosimilitud (4.33) se reexpresa como:

$$f(x_{p+1}, \dots, x_n) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{n-p}{2}} \exp \left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2 \right) \quad (4.36)$$

Si $p \ll n$ esta verosimilitud (que excluye los primeros x_1, \dots, x_p) será una buena aproximación a la verosimilitud de toda la muestra. La verosimilitud logarítmica (loglikelihood), ℓ , será:

$$\ell = -(n-p) \ln \left(\sigma_\varepsilon \sqrt{2\pi} \right) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2 \quad (4.37)$$

El vector $\hat{\alpha}$ se obtiene de maximizar (4.37), lo cual equivale a minimizar el término:

$$\sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2$$

que es el mismo que se minimiza cuando se aplica MCO. Por lo tanto, el método de máxima verosimilitud arroja los mismos resultados para

â que MCO, también sesgado en muestras finitas, pero consistente; sin embargo, el estimador de la varianza $\hat{\sigma}_\varepsilon^2$ da diferente, como se verá más adelante.

Mientras que la ecuación (4.33) es la probabilidad incondicional conjunta de $\varepsilon_{p+1}, \dots, \varepsilon_n$, la expresión (4.36) no es la probabilidad incondicional conjunta de x_{p+1}, \dots, x_n , sino la probabilidad de x_{p+1}, \dots, x_n condicional a que los valores dados de x_1, \dots, x_p no cambien. En otras palabras, la expresión (4.36) es la probabilidad condicional $f(x_{p+1}, \dots, x_n/x_1, \dots, x_p)$ porque, para construirla, se requieren los valores x_1, \dots, x_p . Así, las distribuciones se multiplican (por ejemplo, véase Lindgren, 1976, p. 40):

$$\begin{aligned} & f(x_{p+1}/x_1, \dots, x_p) f(x_{p+2}/x_1, \dots, x_{p+1}) \dots f(x_n/x_1, \dots, x_{n-1}) \\ &= \frac{f(x_1, \dots, x_n)}{f(x_1, \dots, x_p)} = f(x_{p+1}, \dots, x_n/x_1, \dots, x_p) \end{aligned}$$

Finalmente, para obtener la probabilidad de toda la muestra multiplicamos la anterior por $f(x_1, \dots, x_p)$ y queda:

$$f(x_1, \dots, x_n) = f(x_1, \dots, x_p) f(x_{p+1}, \dots, x_n/x_1, \dots, x_p) \quad (4.38)$$

Estimación por máxima verosimilitud exacta

Con base en la discusión anterior, y siguiendo a Priestley (1981, p. 350), encontramos a continuación la verosimilitud exacta que permite una estimación asintóticamente más eficiente. Recordamos que la expresión general de la distribución de probabilidad gaussiana multivariada es:

$$f(x_1, \dots, x_p) = \frac{1}{(2\pi)^{\frac{p}{2}} [\det \Sigma_p]^{\frac{1}{2}}} \exp \left[-\frac{1}{2}(x - \mu)' \Sigma_p^{-1} (x - \mu) \right] \quad (4.39)$$

Notando que en el caso que nos ocupa hemos supuesto $\mu = E x_t = 0$, podemos expresar la verosimilitud de las primeras p observaciones de x_t , que fueron excluidas de la verosimilitud condicional (4.36), como:

$$f(x_1, \dots, x_p) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{p}{2}} [\det \phi_p^{-1}]^{\frac{1}{2}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} x' \phi_p^{-1} x \right] \quad (4.40)$$

Donde definimos $\sigma_\varepsilon^2 \phi = \Sigma_p$ como la matriz de varianza covarianza de las p observaciones. Notamos que en la práctica se puede factorizar Σ_p en σ_ε^2 y ϕ_p , porque Σ_p es la matriz de varianza covarianza teórica en función de los parámetros. Para una mejor comprensión, Σ_p tiene la forma:

$$\Sigma_p = \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ & & \dots & R(p-2) \\ \vdots & & \ddots & \vdots \\ R(p-1) & \dots & & R(0) \end{bmatrix} \quad (4.41)$$

El determinante de Σ_p será $\sigma_\varepsilon^{2p} [\det \phi_p] = \frac{\sigma_\varepsilon^{2p}}{\det \phi_p^{-1}}$ porque multiplicar una matriz $p \times p$ por alguna constante λ , que en este caso es σ_ε^2 , multiplica su determinante por λ^p (o sea por σ_ε^{2p}). Además $\det A = \frac{1}{\det A^{-1}}$ (véase, por ejemplo, Johnston, 1984, pp. 132 y 134).

En otras palabras, para el caso de las primeras p observaciones de una serie de tiempo, tendremos una matriz de varianza covarianza Σ_p de orden $p \times p$. Si la serie es estacionaria, sus elementos resultarán iguales a las covarianzas $R(\tau)$. La diagonal principal de la matriz de varianza covarianza será, obviamente, $R(0)$, o sea la varianza σ_x^2 de la serie. La siguiente diagonal paralela a ambos lados de la principal será $R(1)$ y la siguiente $R(2)$, y así sucesivamente hasta llegar a $R(p-1)$.

Por ejemplo, para un AR(1) sólo necesitamos la primera observación para formar la verosimilitud (4.40), ya que $p = 1$. De esta manera la matriz de varianza covarianza será un escalar igual a σ_x^2 . El término σ_x^2 se encontró en el capítulo anterior elevando $x_t = \alpha x_{t-1} + \varepsilon_t$ al cuadrado y tomando expectativas:

$$\sigma_x^2 = E x_t^2 = E(\alpha x_{t-1} + \varepsilon_t)^2 = \alpha^2 \sigma_x^2 + \sigma_\varepsilon^2$$

De manera que:

$$\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1 - \alpha^2} = \Sigma = R(0)$$

Para un AR(2) necesitaríamos las dos primeras observaciones x_1 , x_2 a efectos de formar la verosimilitud (4.40); de manera que la matriz de varianza covarianza será de orden 2×2 . Como se hizo en el

capítulo anterior para un AR(2), multiplicamos por $x_{t-\tau}$ y tomamos expectativas:

$$E(x_t x_{t-\tau}) + \alpha_1 E(x_{t-1} x_{t-\tau}) + \alpha_2 E(x_{t-2} x_{t-\tau}) = E(\varepsilon_t x_{t-\tau})$$

Y hacemos $\tau = 0, 1$ y 2 para obtener:

$$\begin{aligned} R(0) + \alpha_1 R(1) + \alpha_2 R(2) &= \sigma_\varepsilon^2 \\ R(1) + \alpha_1 R(0) + \alpha_2 R(1) &= 0 \\ R(2) + \alpha_1 R(1) + \alpha_2 R(0) &= 0 \end{aligned}$$

De la segunda y tercera ecuaciones se tiene:

$$\begin{aligned} R(1) &= \frac{-\alpha_1 R(0)}{1 + \alpha_2} \\ R(2) &= -\alpha_1 R(1) - \alpha_2 R(0) = \frac{\alpha_1^2 R(0)}{1 + \alpha_2} - \alpha_2 R(0) \end{aligned}$$

Y al reemplazar en la primera, se obtiene:

$$\begin{aligned} \sigma_x^2 = R(0) &= \frac{(1 + \alpha_2)\sigma_\varepsilon^2}{(1 - \alpha_2)(1 - \alpha_1 + \alpha_2)(1 + \alpha_1 + \alpha_2)} \\ R(1) &= \frac{-\alpha_1 \sigma_\varepsilon^2}{(1 - \alpha_2)(1 - \alpha_1 + \alpha_2)(1 + \alpha_1 + \alpha_2)} \end{aligned}$$

Nótese que $R(0)$ y $R(1)$ quedan en función de los parámetros α_1 , α_2 y que además tienen el término σ_ε^2 en el numerador, de manera que la matriz de varianza covarianza para la verosimilitud (4.40) de x_1 , x_2 tendrá $R(0)$ en la diagonal principal y $R(1)$ en el resto y, obviamente, será 2×2 :

$$\begin{bmatrix} R(0) & R(1) \\ R(1) & R(0) \end{bmatrix}$$

Nótese que $R(2)$ no entra en esta matriz, ya que se trata de las variables x_1 , x_2 separadas sólo en una unidad de tiempo. En general, para cualquier p se puede factorizar σ_ε^2 de $R(\tau)$ y podemos escribir la matriz de varianza covarianza como se anotó en (4.40), así:

$$\Sigma_p = \sigma_\varepsilon^2 \phi_p$$

Donde ϕ_p es simétrica positiva semidefinida, como toda matriz de varianza covarianza.

Volviendo a la estimación por máxima verosimilitud exacta, multiplicando (4.36) por (4.40), como se indica en (4.38), se obtiene la verosimilitud exacta para la totalidad de las observaciones:

$$f(x_1, \dots, x_n) = \left(\frac{1}{2\pi\sigma_\epsilon^2} \right)^{\frac{n}{2}} [\det\phi_p^{-1}]^{\frac{1}{2}} \times \exp \left[-\frac{\sum_{i=1}^p \sum_{j=1}^p \phi_{ij} x_i x_j + \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2}{2\sigma_\epsilon^2} \right] \quad (4.42)$$

Donde ϕ_{ij} es el elemento i, j de ϕ_p^{-1} y donde el término $\sum_{i=1}^p \sum_{j=1}^p \phi_{ij} x_i x_j = x' \phi_p^{-1} x$ corresponde sólo a las primeras p observaciones, y x_i es la observación i -ésima. La verosimilitud logarítmica exacta será (sin considerar términos constantes):

$$\ell = -\frac{n}{2} \ln \sigma_\epsilon^2 + \frac{1}{2} \ln [\det\phi_p^{-1}] - \frac{\sum_{i=1}^p \sum_{j=1}^p \phi_{ij} x_i x_j + \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2}{2\sigma_\epsilon^2} \quad (4.43)$$

Si hay variables exógenas z_t en el AR(p) por estimar, se afectarán las covarianzas de la matriz Σ_p , así como la expresión para ε_t^2 , que es la última sumatoria.

Naturalmente ϕ_p es función de los parámetros $\alpha_1, \dots, \alpha_p$. Un problema en la estimación es que la derivada del $\ln [\det\phi_p^{-1}]$ es muy complicada. Box y Jenkins (1970, p. 277) muestran que para n grande, el término $\ln [\det\phi_p^{-1}]$ es despreciable comparado con el último término de la ecuación (4.43) (véase Judge et al., 1988, p. 530). De manera que, con esta aproximación al pasar por alto $\ln [\det\phi_p^{-1}]$, la maximización de la verosimilitud exacta será aproximada y equivalente a minimizar

la expresión:

$$\sum_{i=1}^p \sum_{j=1}^p \phi_{ij} x_i x_j + \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p})^2 \quad (4.44)$$

Que es parecido, pero no igual, a MCO o a máxima verosimilitud condicional.

Volviendo a la expresión de la verosimilitud exacta (4.42), ¿por qué no se hace la verosimilitud conjunta $f(x_1, \dots, x_n)$ desde el comienzo, así como se hizo la de $f(x_1, \dots, x_p)$? Esto, en principio, se puede hacer, pero en la práctica resulta (o requiere) una enorme matriz de varianza covarianza $n \times n$ que sería muy difícil de invertir, y cuyos términos pueden ser complicadas funciones de los parámetros. En contraste, para la verosimilitud de x_1, \dots, x_p , en especial si $p \ll n$ como es el caso usual, la matriz de varianza covarianza necesaria para las primeras p observaciones resulta manejable; de 2×2 para un AR(2), de 3×3 para un AR(3), etc.

Ejemplo

Para el AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$, la expresión correspondiente a (4.40) será:

$$f(x_1) = \frac{1}{(2\pi\sigma_x^2)^{\frac{1}{2}}} \exp \left[-\frac{1}{2\sigma_x^2} x_1^2 \right]$$

Como se trata de una sola variable, la matriz de varianza covarianza es un escalar e igual a la varianza $\sigma_x^2 = \frac{\sigma_\varepsilon^2}{1-\alpha^2}$. Al reemplazar tendremos:

$$f(x_1) = \frac{(1-\alpha^2)^{\frac{1}{2}}}{(2\pi\sigma_\varepsilon^2)^{\frac{1}{2}}} \exp \left[-\frac{1}{2\sigma_\varepsilon^2} x_1^2 (1-\alpha^2) \right]$$

Por lo tanto, la verosimilitud exacta será el producto de la anterior con la verosimilitud de $(x_2, \dots, x_n/x_1)$, a la manera de (4.42):

$$f(x_1, \dots, x_n) = \left(\frac{1}{2\pi\sigma_\varepsilon^2} \right)^{\frac{n}{2}} (1-\alpha^2)^{\frac{1}{2}} \times \exp \left[-\frac{1}{2\sigma_\varepsilon^2} \left(x_1^2 (1-\alpha^2) + \sum_{t=2}^n (x_t - \alpha x_{t-1})^2 \right) \right] \quad (4.45)$$

¿Qué pasa si $|\alpha| = 1$? Puede pensarse, viendo la expresión anterior, que el método de máxima verosimilitud deja de funcionar ya que precisamente en $\alpha = 1$ debería ser la máxima probabilidad, pero (4.45) da cero. Realmente no deja de funcionar, sino que la expresión (4.45) no es aplicable, porque no fue derivada para $|\alpha| = 1$; el término $1 - \alpha^2$ que causa el problema es el denominador de la varianza que fue derivada bajo el supuesto de estacionariedad en la serie, $|\alpha| < 1$. Sin embargo, la expresión condicional (4.36) sí es aplicable.

Estimación de la varianza del error

La expresión para $\hat{\sigma}_\epsilon^2$ se puede obtener de varias formas. Si es por MCO, tendremos:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n - 2p} \sum_{t=p+1}^n (x_t + \hat{\alpha}_1 x_{t-1} + \dots + \hat{\alpha}_p x_{t-p})^2 \quad (4.46)$$

Donde el divisor se explica porque hay $n - p$ observaciones utilizables menos p parámetros estimados, esto es, $n - p - p = n - 2p$.

Si $\hat{\sigma}_\epsilon^2$ se obtiene por máxima verosimilitud condicional, entonces tendremos:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n - p} \sum_{t=p+1}^n (x_t + \hat{\alpha}_1 x_{t-1} + \dots + \hat{\alpha}_p x_{t-p})^2 \quad (4.47)$$

Y si se obtiene de la máxima verosimilitud exacta, tendremos:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n} \left[\sum_{i=1}^p \sum_{j=1}^p \hat{\phi}_{ij} x_i x_j + \sum_{t=p+1}^n (x_t + \hat{\alpha}_1 x_{t-1} + \dots + \hat{\alpha}_p x_{t-p})^2 \right] \quad (4.48)$$

Ejemplo de un AR

El Cuadro 4.1 es una salida de Eviews para la serie de primeras diferencias en los datos trimestrales del índice de precios, llamada $P1$. Se muestra la ACF y PACF, junto con el estadístico Q que se computa para las autocorrelaciones; el nivel de significancia (Prob) es cercano

a cero e indica que la serie no es ruido blanco. La ACF estimada (AC en la salida de Eviews) podría interpretarse como osculatoria decreciente a medida que aumenta el número de rezagos, lo cual sugiere un autorregresivo con un asomo de estacionalidad de frecuencia 4. Se observa que la PACF estimada (PAC en el listado) cae después del rezago 4 sugiriendo que la serie podría modelarse como un AR(4). La estimación de este modelo se muestra en el Cuadro 4.2.

Estimación de los modelos MA(q)

Sea x_t un proceso de promedio móvil con especificación q ya conocida y con media cero; esto es, $E x_t = 0$. La expresión del modelo es:

$$x_t = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \beta_2 \varepsilon_{t-2} + \dots + \beta_q \varepsilon_{t-q} \quad (4.49)$$

Si bien es posible invertir el MA(q) anterior y despejar ε_t para escribir la verosimilitud, la expresión es inapropiada porque ε_t quedaría en función de un número infinito de términos. Por otro lado, aunque los $\{\varepsilon_t\}$ sean normales y x_1, \dots, x_n también, escribir la distribución conjunta de x_1, \dots, x_n se dificulta, pues la matriz de varianza covarianza Σ_x es una función complicada de los β , además de ser de orden $n \times n$. En principio se puede utilizar el método de momentos, igualando la expresión de las autocovarianzas teóricas a las estimadas, aunque algunos autores como Hannan (1960, p. 47) consideran que aplicar este método para estimar los β es ineficiente para un MA(q). Tampoco se puede aplicar MCO directamente a (4.49), porque no conocemos los ε_t . No sirve obtener los ε_t de un generador de números aleatorios porque, de acuerdo con el teorema de Wold, éstos deben provenir de un AR infinito. Lo anterior sugiere que, como aproximación, podemos correr un AR de orden grande y usar los rezagos de los residuos para estimar (4.49).

Cuadro 4.1: Correlograma para la serie *P1*

Date: 07/26/99	Partial Correlation	AC	PAC	Q-Stat	Prob
Autocorrelation					
1		1	0.441	0.441	12.431 0.000
2	-	2	0.189	-0.007	14.747 0.001
3		3	0.436	0.441	27.333 0.000
4		4	0.707	0.548	61.014 0.000
5		5	0.348	-0.077	69.346 0.000
6	-	6	0.177	-0.022	71.532 0.000
7		7	0.302	-0.171	78.037 0.000
8		8	0.552	0.178	100.11 0.000
9		9	0.254	-0.127	104.89 0.000
10		10	0.066	-0.094	105.22 0.000
11		11	0.184	-0.079	107.82 0.000
12		12	0.425	0.095	121.95 0.000
13		13	0.154	-0.046	123.85 0.000
14	-	14	-0.012	-0.020	123.86 0.000
15		15	0.087	-0.057	124.49 0.000
16		16	0.282	-0.011	131.28 0.000
17		17	0.082	0.020	131.87 0.000
18		18	-0.080	-0.052	132.44 0.000
19		19	0.000	-0.005	132.44 0.000
20		20	0.149	-0.103	134.54 0.000

Cuadro 4.2: AR estimado para la serie P1

Dependent Variable: P1 Method: Least Squares Date: 06/26/99 Time: 19:09 Sample (adjusted): 1985:1 1999:1 Included observations: 57 after adjusting endpoints				
VARIABLE	COEFFICIENT	Std. Error	t-Statistic	Prob.
P1(-1)	0.317616	0.092902	3.418827	0.0012
P1(-2)	-0.187994	0.101380	-1.854350	0.0693
P1(-3)	0.227195	0.102277	2.221485	0.0306
P1(-4)	0.683438	0.094021	7.269022	0.0000
R-squared	0.741239	Mean dependend var	13.90175	
Ajusted R-squared	0.726592	S.D. dependent var	11.93127	
S.E. of regression	6.238672	Akaike info criterion	6.567003	
Sum squared resid	2062.815	Schwarz criterion	6.710375	
Log likelihood	-183.1596	F-statistic	50.60738	
Durbin-Watson stat	1.923771	Prob (F-statistic)	0.000000	

El método de estimación más popular, ideado por Box y Jenkins (1970), es una solución recursiva sobre ε_t con el fin de aproximar la verosimilitud. Para la recursión suponemos que $\varepsilon_0 = \varepsilon_{-1} = \varepsilon_{-2} = \dots = \varepsilon_{1-q} = 0$ como valores iniciales. Al despejar ε_t de la ecuación (4.49), podemos construir los errores así:

$$\begin{aligned}
 \varepsilon_1 &= x_1 \\
 \varepsilon_2 &= x_2 - \beta_1 \varepsilon_1 \\
 \varepsilon_3 &= x_3 - \beta_1 \varepsilon_2 - \beta_2 \varepsilon_1 \\
 \varepsilon_4 &= x_4 - \beta_1 \varepsilon_3 - \beta_2 \varepsilon_2 - \beta_3 \varepsilon_1 \\
 &\vdots \\
 \varepsilon_q &= x_q - \beta_1 \varepsilon_{q-1} - \beta_2 \varepsilon_{q-2} - \dots - \beta_{q-1} \varepsilon_1 \\
 \varepsilon_{q+1} &= x_{q+1} - \beta_1 \varepsilon_q - \beta_2 \varepsilon_{q-1} - \dots - \beta_{q-1} \varepsilon_2 - \beta_q \varepsilon_1 \\
 &\vdots \\
 \varepsilon_t &= x_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \dots - \beta_q \varepsilon_{t-q}
 \end{aligned} \tag{4.50}$$

Donde la última expresión es para $q < t \leq n$. En cada ecuación se van reemplazando los valores ya conocidos que se obtuvieron en las anteriores. Si hay una constante c se incluye en (4.50) y se trata como cualquier otro coeficiente que se vaya a estimar, o se usa $\bar{x} = c$ después de la estimación.

Elevamos al cuadrado la expresión para cada ε_t y sumamos para obtener la siguiente verosimilitud logarítmica, condicional a $\varepsilon_0 = \dots = \varepsilon_{1-q} = 0$ y, obviamente, bajo el supuesto de normalidad en los errores:

$$\ell = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^n \varepsilon_t^2 \quad (4.51)$$

A primera vista parecería que la optimización necesaria se logra derivando $\sum_{t=1}^n \varepsilon_t^2$ con respecto a los parámetros β_1, \dots, β_q , igualando a cero y despejando de allí los $\hat{\beta}_1, \dots, \hat{\beta}_q$ buscados. Sin embargo, entenderemos mejor la dificultad de hacer esto observando cómo se forma cada uno de los errores $\varepsilon_1, \dots, \varepsilon_n$. Obviamente, al hacer los reemplazos indicados en (4.50) van quedando los β elevados al cuadrado, al cubo, etc., y también multiplicados entre ellos, de manera que será difícil despejar β . Por ejemplo, ε_4 será:

$$\varepsilon_4 = x_4 - \beta_1(x_3 - \beta_1\varepsilon_2 - \beta_2\varepsilon_1) - \beta_2(x_2 - \beta_1\varepsilon_1) - \beta_3x_1$$

Expresión de la cual aparecerán productos $\beta_1\beta_2$. En ε_5 aparecerán productos $\beta_1^2\beta_2$, $\beta_1\beta_2$, $\beta_1\beta_3$ y así con los ε siguientes. De manera que $\Sigma \varepsilon_t^2$ es una función no lineal de los parámetros (o no linealizable).

Dada la dificultad para despejar los β estimados de las derivadas igualadas a cero –lo que contrasta con la facilidad de hacerlo en el caso del modelo clásico de regresión–, aquí se hace necesario ensayar, de manera inteligente, diferentes valores de β hasta encontrar los que maximicen la verosimilitud (4.51). Para esto se utiliza algún método iterativo de optimización numérica, los cuales serán presentados de manera intuitiva más adelante.

El estimador de máxima verosimilitud condicional de la varianza σ_ε^2 tendrá la expresión:

$$\frac{1}{n} \sum_{t=1}^n \hat{\varepsilon}_t^2$$

No obstante, algunos autores (Jenkins y Watts, 1968) prefieren $n - q$ en el denominador en lugar de n . También existen métodos para calcular la verosimilitud exacta del MA(q) (véanse Priestley, 1981, p. 356; Hamilton, 1994, p. 130).

Si x_t tiene raíces cercanas a la unidad, la condición $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{1-q} = 0$ no será trivial, ya que el efecto de este supuesto puede persistir en el tiempo y acentuar el sesgo en la estimación de los β . Una forma de aliviar los efectos del supuesto $\varepsilon_0 = \varepsilon_{-1} = \dots = \varepsilon_{1-q} = 0$, es tratar de estimar el valor de estos errores por medio de una técnica conocida como *backforecast* o *backcast*, lo cual quiere decir predecir hacia atrás, o retropredecir. Al igual que para un AR –véanse (4.21) y (4.27)–, correr un MA con errores pasados o con errores futuros arroja los mismos coeficientes; esto es, la expresión:

$$x_t = u_t + \beta_1 u_{t+1} + \beta_2 u_{t+2} + \dots + \beta_q u_{t+q} \quad (4.52)$$

tiene los mismos coeficientes β_1, \dots, β_q que (4.49), porque tiene su misma ACF, así como se puede ver en (3.22):

$$r(\tau) = \frac{\sum_{s=0}^{q-\tau} \beta_s \beta_{s+\tau}}{\sum_{s=0}^q \beta_s^2} = \frac{\sum_{s=\tau}^q \beta_s \beta_{s-\tau}}{\sum_{s=0}^q \beta_s^2} = \frac{\beta_\tau + \beta_1 \beta_{\tau+1} + \dots + \beta_{q-\tau} \beta_q}{1 + \beta_1^2 + \beta_2^2 + \dots + \beta_q^2}$$

En *backcasting* se hace un procedimiento similar a (4.50), pero comenzando por el final de la muestra, lo cual requiere hacer el supuesto $u_{n+1} = u_{n+2} = u_{n+3} = \dots = u_{n+q} = 0$ sobre los errores futuros. Para iniciar la iteración, se escogen valores iniciales para los coeficientes β_1, \dots, β_q y con ellos se construyen los errores u_n, \dots, u_1 de manera recursiva despejando u_t de la ecuación (4.52); con estos valores se retropredicen $x_0, x_{-1}, \dots, x_{1-q}$. Por ejemplo:

$$x_0 = \beta_1 u_1 + \beta_2 u_2 + \dots + \beta_q u_q$$

⋮

$$x_{1-q} = \beta_q u_1$$

A continuación se utiliza la ecuación (4.49) con estos valores para $x_0, x_{-1}, \dots, x_{1-q}$ y con los mismos β_1, \dots, β_q iniciales para estimar los

errores $\varepsilon_0, \varepsilon_{-1}, \dots, \varepsilon_{1-q}$, igual que se hizo en (4.50), sólo que empezando en $\varepsilon_{1-q} = x_{1-q}$, $\varepsilon_{1-q+1} = x_{1-q+1} - \beta_1\varepsilon_{1-q}, \dots$. Finalmente se construye y evalúa la verosimilitud (4.51). Este proceso de recursión hacia atrás, luego hacia delante y optimización, se repite iniciando cada iteración del método numérico con nuevos valores para β_1, \dots, β_q , hasta que la estimación converja.

Ejemplo de un MA

El Cuadro 4.3 se refiere a la serie trimestral de la tasa de interés de préstamos, llamada de colocación (coloc). A manera de ilustración, se podría interpretar que la ACF cae después del rezago dos mientras que la PACF decrece a medida que aumenta el número de rezagos, lo cual sugiere un modelo MA(2). La estimación de este modelo se muestra en el Cuadro 4.4.

Cuadro 4.3: Correlograma de la series COLOC

		Date: 07/26/99 Time: 17:33				
		Sample: 1987:1 1999:1				
		Included Observations: 49				
Autocorrelation		Partial Correlation	AC	PAC	Q-Stat	Prob
1			1	0.646	0.646	21.715 0.000
2			2	0.426	0.016	31.384 0.000
3			3	0.075	-0.354	31.690 0.000
4			4	0.026	0.204	31.729 0.000
5			5	-0.050	-0.017	31.871 0.000
6			6	-0.140	-0.316	33.017 0.000
7			7	-0.233	-0.022	36.244 0.000
8			8	-0.269	0.043	40.656 0.000
9			9	-0.294	-0.245	46.074 0.000
10			10	-0.285	-0.085	51.274 0.000
11			11	-0.283	0.020	56.538 0.000
12			12	-0.147	0.100	58.003 0.000
13			13	-0.004	0.050	58.004 0.000
14			14	0.146	0.030	59.526 0.000
15			15	0.147	-0.079	61.107 0.000
16			16	0.153	0.054	62.882 0.000
17			17	0.136	0.034	64.330 0.000
18			18	0.166	-0.040	66.554 0.000
19			19	0.144	-0.004	68.271 0.000
20			20	0.134	0.105	69.816 0.000

Cuadro 4.4: MA estimado para la serie COLOC

Dependent Variable: COLOC				
Method: Least Squares				
Date: 06/26/99 Time: 17:38				
Sample: 1987:1 1999:1				
Convergence achieved after 11 iterations				
Backcast: 1986:3 1986:4				
VARIABLE	COEFFICIENT	Std. Error	t-Statistic	Prob.
C	41.18159	1.083439	38.01006	0.0000
MA(1)	0.777503	0.115781	6.715285	0.0000
MA (2)	0.771866	0.122377	6.307303	0.0000
R-squared	0.566656	Mean dependend var	41.21837	
Ajusted R-squared	0.547815	S.D. dependent var	4.513806	
S.E. of regression	3.035296	Akaike info criterion	5.117765	
Sum squared resid	423.7990	Schwarz criterion	5.233591	
Log likelihood	-122.3852	F-statistic	30.07561	
Durbin-Watson stat	2.017039	Prob (F-statistic)	0.000000	
Inverted MA Roots	-.39 -.79i	-.39 + .79i		

Estimación de los modelos ARMA(p,q)

Sea x_t un proceso ARMA con p y q conocidos y media cero. Su expresión es:

$$x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} = \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (4.53)$$

O también:

$$\sum_{j=0}^p \alpha_j x_{t-j} = \sum_{i=0}^q \beta_i \varepsilon_{t-i} \quad (4.54)$$

donde $\alpha_0 = \beta_0 = 1$.

Procedemos de manera similar a como se hizo para el MA(q), bajo el supuesto de hacer cero los ε anteriores a ε_{p+1} ; esto es,

$$\varepsilon_p = \varepsilon_{p-1} = \dots = \varepsilon_{p+1-q} = 0.$$

Escribimos:

$$\begin{aligned}
 \varepsilon_{p+1} &= x_{p+1} + \alpha_1 x_p + \dots + \alpha_p x_1 \\
 \varepsilon_{p+2} &= x_{p+2} + \alpha_1 x_{p+1} + \dots + \alpha_p x_2 - \beta_1 \varepsilon_{p+1} \\
 &\vdots \\
 \varepsilon_{p+q} &= x_{p+q} + \alpha_1 x_{p+q-1} + \dots + \alpha_p x_q - \beta_1 \varepsilon_{p+q-1} - \dots - \beta_{q-1} \varepsilon_{p+1} \\
 \varepsilon_{p+q+1} &= x_{p+q+1} + \alpha_1 x_{p+q} + \dots + \alpha_p x_{q+1} - \beta_1 \varepsilon_{p+q} - \dots \\
 &\quad - \beta_{q-1} \varepsilon_{p+2} - \beta_q \varepsilon_{p+1} \\
 &\vdots \\
 \varepsilon_t &= x_t + \alpha_1 x_{t-1} + \dots + \alpha_p x_{t-p} - \beta_1 \varepsilon_{t-1} - \dots - \beta_q \varepsilon_{t-q}
 \end{aligned} \tag{4.55}$$

Donde la última expresión es para $p + q < t \leq n$.

De nuevo, la estimación por máxima verosimilitud condicional (condicional a $\varepsilon_p = \varepsilon_{p-1} = \dots = \varepsilon_{p+1-q} = 0$ y a los valores iniciales de x_1, \dots, x_p) se obtiene maximizando la expresión:

$$\ell = -(n-p) \ln(\sigma_\varepsilon \sqrt{2\pi}) - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^n \varepsilon_t^2$$

Ello implica minimizar la suma de los errores (4.55) elevados al cuadrado:

$$\sum_{t=p+1}^n \varepsilon_t^2 = \varepsilon_{p+1}^2 + \varepsilon_{p+2}^2 + \dots + \varepsilon_n^2 \tag{4.56}$$

Lo cual es una función no lineal de los parámetros $\alpha_1, \dots, \alpha_p$ y β_1, \dots, β_q . Para esto se utiliza algún método de optimización numérica. El estimador de máxima verosimilitud condicional de σ_ε^2 será:

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-p} \sum_{t=p+1}^n \hat{\varepsilon}_t^2 \tag{4.57}$$

Entre tanto, el de MCO será

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{n-2p-q} \sum_{t=p+1}^n \hat{\varepsilon}_t^2,$$

donde el divisor $n - 2p - q$ se explica porque hay $n - p$ datos utilizables menos $p + q$ parámetros, lo cual da $n - p - p - q = n - 2p - q$.

Al igual que en el caso del autorregresivo, existen métodos para calcular la verosimilitud exacta del modelo ARMA (véase Priestley, 1981, p. 361; Hamilton, 1994, p. 133).

Ejemplo con un AR

El método anterior de solución numérica también se puede aplicar a un AR(p) puro como:

$$x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} = \varepsilon_t .$$

Las ecuaciones al estilo (4.55) para encontrar los ε_t son:

$$\varepsilon_{p+1} = x_{p+1} + \alpha_1 x_p + \alpha_2 x_{p-1} + \dots + \alpha_p x_1$$

$$\varepsilon_{p+2} = x_{p+2} + \alpha_1 x_{p+1} + \alpha_2 x_p + \dots + \alpha_p x_2$$

$$\varepsilon_{p+3} = x_{p+3} + \alpha_1 x_{p+2} + \alpha_2 x_{p+1} + \dots + \alpha_p x_3$$

⋮

De donde la suma que se va a minimizar es:

$$\sum_{t=p+1}^n \varepsilon_t^2 = \sum_{t=p+1}^n (x_t + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p})^2$$

Que es la misma expresión que se obtuvo en la verosimilitud condicional o en MCO para un AR(p).

Estimación asintótica de un ARMA por mínimos cuadrados ordinarios

Otro método para estimar un ARMA o MA utilizando MCO es el indicado por Hannan y Rissanen (1982). En esencia, se generan los ε como los errores de una proyección lineal de x_t contra sus rezagos (cuántos más rezagos mejor). Con estos ε se corre el MA o el ARMA.

Sea x_t de la forma:

$$x_t = \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q}$$

Donde ya se conocen p y q . El primer paso es correr un AR(k), esto es, x_t contra $x_{t-1}, x_{t-2}, \dots, x_{t-k}$ por MCO. El orden k se establece a ojo o graficando $\hat{\epsilon}'\hat{\epsilon}$ para cada k como en la Figura 4.1 o usando los criterios AIC o SC. El segundo paso es utilizar los residuos de la autorregresión anterior:

$$\hat{\epsilon}_t = x_t - \hat{\alpha}_1 x_{t-1} - \hat{\alpha}_2 x_{t-2} - \dots - \hat{\alpha}_k x_{t-k}$$

Esto, a fin de correr una nueva regresión de x_t contra $x_{t-1}, x_{t-2}, \dots, x_{t-p}$, y $\hat{\epsilon}_{t-1}, \dots, \hat{\epsilon}_{t-q}$, sin incluir $\hat{\epsilon}_t$, ya que éste será representado por el error usual de la regresión. De esta regresión obtienen los parámetros estimados:

$$\hat{\alpha}_1, \dots, \hat{\alpha}_p \quad \text{y} \quad \hat{\beta}_1, \dots, \hat{\beta}_q$$

Se puede adicionar un tercer paso, consistente en repetir la regresión del paso dos con los residuos de la ecuación anterior, llamados $\hat{\epsilon}_t$; esto es, corriendo x_t contra $x_{t-1}, x_{t-2}, \dots, x_{t-p}$ y $\hat{\epsilon}_{t-1}, \dots, \hat{\epsilon}_{t-q}$; aunque generalmente este tercer paso no es necesario.

Intuición de la optimización numérica

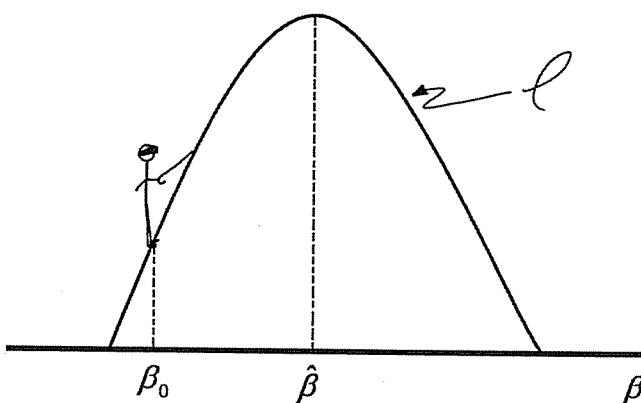
Intentaremos explicar de manera intuitiva, y no rigurosa, la forma como funcionan los algoritmos de optimización que se utilizan para resolver problemas de máxima verosimilitud. Supongamos que se trata de encontrar el valor del parámetro β que coincide con el máximo de una función cóncava (que podría ser una verosimilitud). Es como tratar de llegar a la cima de la colina mostrada en la Figura 4.2.

Llamemos esta colina ℓ . Suponga que usted está inicialmente en el punto β_0 , que está vendado y no puede ver, que tiene un bastón y que debe llegar a la cima de la colina. ¿Cómo lo haría?

Usaría la punta del bastón para saber en qué dirección sube la colina. En otras palabras, comparará la diferencia de nivel entre la punta del bastón y sus pies. Si la punta del bastón está más alta que sus pies, usted dará un paso en esa dirección. Una vez ubicado en la nueva posición se repite el experimento del bastón.

Siempre que la punta del bastón indique desnivel, usted deberá avanzar. Cuando no haya desnivel, o el desnivel sea muy pequeño o indetectable, habrá llegado a la cima. Si se pasa de la cima, el bastón estará más bajo que sus pies y debe devolverse.

Figura 4.2: Verosimilitud por maximizar



El punto importante es éste: usar las diferencias de nivel a lo largo de la colina con el propósito de saber en qué dirección moverse equivale a usar las primeras derivadas con respecto a β de esa curva (la colina l) para saber en qué dirección moverse. Siempre que la primera derivada de la curva de la función que se vaya a maximizar sea positiva, podemos avanzar en la dirección β . Si es negativa, quiere decir que nos pasamos de la cima y debemos retroceder. Cuando sea cero, habremos llegado a la cima.

Entonces el algoritmo iterativo consiste en iniciar en algún punto β_t sobre la curva, evaluar la primera derivada en ese punto y si es positiva dar un paso en la dirección $+\beta$; de lo contrario, se da un paso en la dirección $-\beta$. Una vez se llega a ese nuevo punto, llámese β_{t+1} , se repite el procedimiento hasta cuando la primera derivada sea cero. En símbolos:

$$\beta_{t+1} = \beta_t + \Lambda \frac{d\ell}{d\beta} \quad (4.58)$$

Donde el símbolo Λ es el tamaño del paso. Obsérvese que si $\frac{d\ell}{d\beta} > 0$ estaremos en la subida y entonces $\beta_{t+1} > \beta_t$; de igual forma, cuando $\frac{d\ell}{d\beta} < 0$ estaremos en la bajada y entonces $\beta_{t+1} < \beta_t$. Al acercarnos a la cima $\frac{d\ell}{d\beta}$ será cada vez menor de manera que la diferencia entre β_{t+1} y β_t también será menor.

No siempre se puede obtener una expresión útil y manejable para $\frac{d\ell}{d\beta}$. Cuando sí sea posible, simplemente se evalúa la expresión de la derivada reemplazando el valor de β en ese punto. De lo contrario, se debe utilizar una aproximación numérica a la primera derivada. Por ejemplo, evaluada en el punto β_t , esto es, en $\ell(\beta_t)$, dicha aproximación numérica podría ser:

$$\frac{d\ell(\beta_t)}{d\beta} \approx \frac{\ell(\beta_t + \epsilon) - \ell(\beta_t - \epsilon)}{2\epsilon} \quad (4.59)$$

Donde ϵ es un número pequeño que se suma y resta al punto β_t . La expresión (4.58) es el principio general de los métodos de optimización. Existen innumerables variaciones a partir de este principio, todas con nombres diferentes. Un buen resumen se encuentra en Judge *et al.* (1985), en Greene (1997) o en Hamilton (1994). Obviamente, si se trata de un ejercicio de minimización (y no de maximización como aquí) simplemente maximizamos $-\ell$.

Una de las variantes más populares es el llamado *método de Newton*, del cual intentaremos a continuación un acercamiento intuitivo siguiendo el mismo ejemplo de la colina. Si observamos la Figura 4.2, vemos que cuando más rápido cambia la primera derivada es al acercarse a la cima, que es, precisamente, cuando más cortos deben ser los pasos (para no pasarnos y poder llegar exactamente a $\hat{\beta}$).

No obstante, cuando más rápido está cambiando la primera derivada, es también cuando más grande es la segunda derivada, de manera que si dividimos el paso Λ de la ecuación (4.58) por la segunda derivada, éste se hará muy pequeño cerca de la cima y grande lejos de ella. Como en la Figura 4.2 vemos que la segunda derivada es siempre negativa, será necesario multiplicarla por un signo menos para no cambiar la dirección del paso indicada por el signo de la primera derivada. Haciendo $\Lambda = 1$ por simplicidad en (4.58) tenemos:

$$\beta_{t+1} = \beta_t - \left(\frac{1}{\frac{d^2\ell}{d\beta^2}} \right) \left(\frac{d\ell}{d\beta} \right) \quad (4.60)$$

Donde ambas derivadas se evalúan en el punto β_t . De nuevo, si carecemos de una expresión matemática para las derivadas entonces éstas deberán evaluarse de manera aproximada.

Volviendo al ejemplo de la Figura 4.2, notamos que hemos estado trabajando con una sola variable, β . Y ¿si hubiera dos variables, digamos β_1 y β_2 ? Usando la imaginación vemos que en ese caso habría un eje para β_1 y otro para β_2 y nos estaríamos enfrentando, ya no a una curva, sino a una superficie con el objetivo de llegar a su cima. Al igual que en el caso de un solo eje, las primeras derivadas nos dirán si debemos avanzar o retroceder. Si $\frac{\partial \ell}{\partial \beta_1} > 0$ daremos un paso en la dirección β_1 ; en caso contrario, retrocederemos. Igualmente, si $\frac{\partial \ell}{\partial \beta_2} > 0$ daremos un paso en la dirección β_2 y viceversa. De manera similar al caso de una sola variable, podemos usar las segundas derivadas para disminuir el tamaño de los pasos cuando nos acercamos a la cima o alargarlos lejos de ella.

Generalizando este razonamiento a k variables, la expresión (4.60) del método de Newton quedaría:

$$\beta_{t+1} = \beta_t - \left(\frac{\partial^2 \ell}{\partial \beta \partial \beta'} \right)^{-1} \left(\frac{\partial \ell}{\partial \beta} \right) \quad (4.61)$$

donde los β son vectores de k elementos, $\frac{\partial \ell}{\partial \beta}$ es el vector de las primeras derivadas de ℓ con respecto a cada una de las k variables (o ejes) y $\frac{\partial^2 \ell}{\partial \beta \partial \beta'}$ es la matriz de segundas derivadas de orden $k \times k$, conocida también como matriz hessiana, y donde las derivadas se evalúan en el punto β_t .

Muchas de las variaciones que existen en la optimización numérica surgen con el objeto de mejorar la expresión anterior o para acomodar el hecho de que no siempre la función que se va a optimizar tiene un comportamiento ideal como el de la Figura 4.2. Las superficies pueden tener, por ejemplo, varias cimas y valles o tener segundas derivadas positivas (y no negativas como aquí) fuera de la vecindad del máximo. En cualquier caso, las variaciones utilizadas para optimizar funciones de verosimilitud son intuitivamente similares al método de Newton aquí explicado.

Para encontrar otro uso importante de la segunda derivada pre-guntamos ¿cuál de los dos $\hat{\beta}$ estimados, el de la Figura 4.3a o el de la Figura 4.3b, tendría mayor confiabilidad, significancia o precisión? La respuesta es el de la Figura 4.3b porque se puede ver que un cambio alrededor de $\hat{\beta}$ tendría mucho más impacto sobre el nivel de ℓ que en la

Figura 4.3a. En consecuencia, el parámetro de la Figura 4.3b se puede estimar con mayor precisión; esto es, con menor varianza.

Figura 4.3a: Función por maximizar

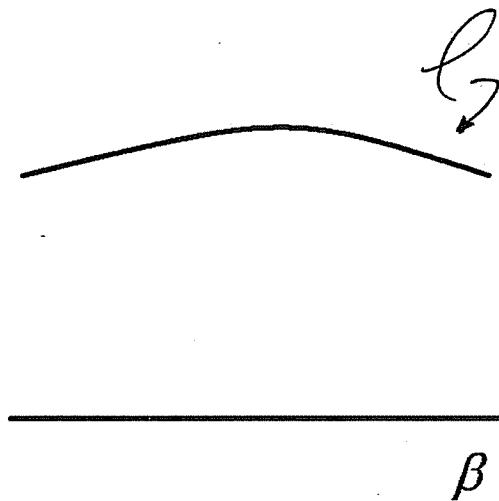
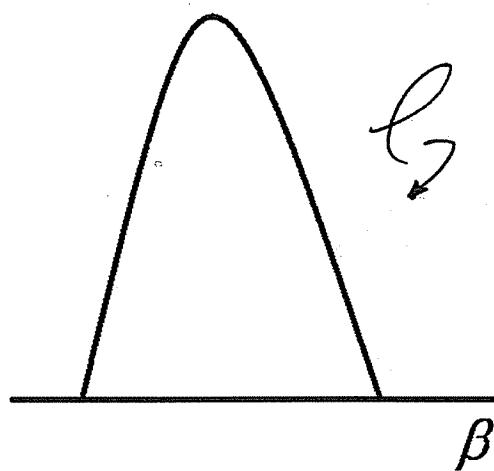


Figura 4.3b: Función por maximizar



¿Cuál de las dos l , el de la Figura 4.3a o el de la Figura 4.3b, tiene mayor segunda derivada (en valor absoluto) cerca de la cima? La

respuesta es el de la Figura 4.3b porque es más puntiaguda, esto es, la primera derivada cambia más rápido en la cima. En consecuencia, intuimos que el inverso de la segunda derivada puede servir para aproximar la varianza de $\hat{\beta}$; obviamente en valor absoluto o con signo negativo para que den positivas las varianzas. A mayor segunda derivada, menor varianza.

Un importante resultado de la estimación por máxima verosimilitud dice que, asintóticamente, el vector estimado $\hat{\beta}$ se distribuye normalmente con media β y varianza igual al inverso del valor esperado de la segunda derivada de la función de verosimilitud (con signo negativo para que la varianza dé positiva):

$$\hat{\beta} \xrightarrow{d} N\left(\beta, \left[-E \frac{\partial^2 \ell}{\partial \beta \partial \beta'}\right]^{-1}\right) \quad (4.62)$$

Donde $\left[-E \frac{\partial^2 \ell}{\partial \beta \partial \beta'}\right]$ se conoce como matriz de información. La matriz de varianza covarianza del vector $\hat{\beta}$ se evalúa en β . Si no se dispone de una expresión matemática para estas segundas derivadas, será necesario aproximarlas numéricamente. Existen varias alternativas en la literatura para remplazar el valor esperado desconocido de la matriz de segundas derivadas. Estas varianzas o desviaciones estándar pueden utilizarse para realizar pruebas de hipótesis sobre los coeficientes $\hat{\beta}$.

Valores iniciales para una optimización numérica

Un problema en algunos ejercicios de optimización numérica tiene que ver con los valores iniciales de los parámetros que se van a estimar, esto es, con el punto de arranque del proceso iterativo. Este es un problema cuando la función que se va a optimizar ℓ no tiene un comportamiento ideal, especialmente cuando no tiene un óptimo global sino varios valles y picos en su geografía. Dependiendo de los valores iniciales, podemos llegar a uno u otro óptimo o quedarnos pegados a la frontera.

Para encontrar valores iniciales para la estimación de un AR(p) se pueden utilizar los que arroja la aplicación MCO que, aunque sesgados en muestras finitas, sabemos que son consistentes. Para un MA(q)

o ARMA(p, q) se puede invertir la expresión y obtener los valores iniciales a partir de una estimación de su representación AR finita (truncada). Por ejemplo, para un MA(q) se corre un AR con muchos rezagos (siguiendo a Wold) y se guardan los residuos; con estos residuos se corre el MA(q), cuyos coeficientes se toman como valores iniciales. También se pueden tomar como valores iniciales las estimaciones del método de Hannan y Rissanen, descrito atrás. Otra posibilidad es usar el método de momentos (Yule-Walker) el cual, aunque ineficiente, sirve para estimar valores iniciales. La mayoría de las rutinas de computador incorporan métodos automáticos para estimar valores iniciales.

Ejemplo de valores iniciales

El ARMA(1, 1) $x_t = \alpha_1 x_{t-1} + \varepsilon_t + \beta_1 \varepsilon_{t-1}$ se puede invertir:

$$\begin{aligned}\varepsilon_t &= \frac{(1 - \alpha_1 L)x_t}{1 + \beta_1 L} = [1 - \beta_1 L + (\beta_1 L)^2 - (\beta_1 L)^3 + (\beta_1 L)^4 - \dots] (1 - \alpha_1 L)x_t \\ &= [(1 - \alpha_1 L) - \beta_1 L(1 - \alpha_1 L) + (\beta_1 L)^2(1 - \alpha_1 L) - (\beta_1 L)^3(1 - \alpha_1 L) \\ &\quad + (\beta_1 L)^4(1 - \alpha_1 L) - \dots] x_t \\ &= [1 - \alpha_1 L - \beta_1 L + \alpha_1 \beta_1 L^2 + (\beta_1 L)^2 - \alpha_1 \beta_1^2 L^3 - (\beta_1 L)^3 \\ &\quad + \alpha_1 \beta_1^3 L^4 + \dots] x_t\end{aligned}\tag{4.63}$$

y se obtiene:

$$\varepsilon_t = x_t - \psi_1 x_{t-1} + \psi_2 x_{t-2} - \psi_3 x_{t-3} + \dots$$

Una vez se estima este AR, obviamente truncando el número de rezagos, se despejan los valores iniciales para α_1 y β_1 a partir de los coeficientes $\hat{\psi}_1$ y $\hat{\psi}_2$. Como puede verse de (4.63), estos coeficientes son función de α_1 y β_1 .



CAPÍTULO 5

Modelos estacionarios multivariados

Los modelos vistos hasta ahora han sido básicamente univariados, donde la variable de interés se explica en función de su propio pasado y de una serie de errores $\{\varepsilon_t\}$. En este capítulo vamos a introducir y estudiar modelos de series de tiempo con más variables, conocidos como ARMA multivariados, de los cuales el más popular es el vector autorregresivo (VAR).

Motivación para los modelos multivariados

Recordemos la expresión del modelo ARMA univariado, la cual en general puede incluir una constante y variables exógenas:

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \gamma_1 z_{1t} + \dots + \gamma_k z_{kt} + \varepsilon_t + \beta_1 \varepsilon_{t-1} + \dots + \beta_q \varepsilon_{t-q} \quad (5.1)$$

Donde z_{1t}, \dots, z_{kt} son variables exógenas y $\gamma_1, \dots, \gamma_k$ sus coeficientes. En el caso particular de un AR con constante y variables exógenas, la expresión es:

$$x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \dots + \alpha_p x_{t-p} + \gamma_1 z_{1t} + \dots + \gamma_k z_{kt} + \varepsilon_t \quad (5.2)$$

La expresión del AR puede verse como un modelo de regresión con variables explicativas estocásticas parcialmente dependientes del error. En razón del Teorema de Mann-Wald, visto en el capítulo 4, este AR puede estimarse por mínimos cuadrados ordinarios (MCO) y obtener resultados consistentes.

Sin embargo, la inclusión de variables exógenas puede causar dificultades. Para que la estimación sea insesgada se requiere que las variables exógenas z_t también sean independientes del error ε_t ; esto es, que sean verdaderamente exógenas. Lo anterior equivale a aseverar que no existe otra ecuación en alguna parte del sistema donde z_t dependa de x_t , ya que, de lo contrario, z_t estará correlacionada con ε_t a través de x_t . En algunos casos se puede suponer que z_t es exógena, en otros no. La teoría no siempre ayuda.

Para recordar porqué esta realimentación (*feedback*) es un problema, usamos el modelo keynesiano sencillo de ecuaciones simultáneas como ilustración:

$$\begin{aligned} C_t &= \alpha + \beta Y_t + \varepsilon_t \\ Y_t &= C_t + I_t \end{aligned} \tag{5.3}$$

Este es el modelo de forma *estructural*, pues las relaciones entre las variables son dictadas por la teoría. El consumo C_t depende del ingreso Y_t pero a su vez Y_t depende de C_t . El coeficiente β es la propensión marginal al consumo. La inversión I_t es exógena y C_t y Y_t son endógenas. Al reemplazar la segunda ecuación en la primera, tenemos:

$$C_t = \alpha + \beta(C_t + I_t) + \varepsilon_t = \alpha + \beta C_t + \beta I_t + \varepsilon_t$$

De donde despejamos C_t y con ello a Y_t , a fin de obtener la forma *reducida* en la cual las variables endógenas están en función de las exógenas:

$$\begin{aligned} C_t &= \frac{\alpha}{1-\beta} + \frac{\beta}{1-\beta} I_t + \frac{\varepsilon_t}{1-\beta} \\ Y_t &= \frac{\alpha}{1-\beta} + \frac{1}{1-\beta} I_t + \frac{\varepsilon_t}{1-\beta} \end{aligned} \tag{5.4}$$

Vemos que Y_t está correlacionado con el error ε_t y que tal dependencia no disminuye cuando $n \rightarrow \infty$. Si aplicamos MCO a la ecuación

$C_t = \alpha + \beta Y_t + \varepsilon_t$, pasando por alto la constante por simplicidad, el valor esperado $E\hat{\beta}$ será:

$$E\hat{\beta} = \beta + E(Y'Y)^{-1}Y'\varepsilon \quad (5.5)$$

cuyo sesgo depende del último término:

$$E(Y'Y)^{-1}Y'\varepsilon = E\left(\frac{Y_1\varepsilon_1 + Y_2\varepsilon_2 + \dots + Y_n\varepsilon_n}{Y_1^2 + Y_2^2 + \dots + Y_n^2}\right)$$

Si este término fuera cero, entonces no habría sesgo; pero de (5.4) vemos que Y_t contiene a ε_t (de manera contemporánea). Por lo tanto, los productos del numerador están correlacionados y tienden a ser del mismo signo, a diferencia de lo que sucedía, por ejemplo, en (4.12). En resumen, aquí no podemos garantizar que $plim(Y'\varepsilon/n) = 0$, de manera que $E(Y'Y)^{-1}Y'\varepsilon \neq 0$ y $\hat{\beta}$ será sesgado e inconsistente.

Ahora veamos qué ocurre en un AR con variable exógena z_t . Usamos el siguiente modelo sencillo para ilustración:

$$x_t = c + \alpha x_{t-1} + \gamma z_t + \varepsilon_t$$

Éste puede expresarse como:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_0 & z_1 \\ x_1 & z_2 \\ \vdots & \vdots \\ x_{n-1} & z_n \end{pmatrix} \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Con su correspondiente estimación por MCO:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} x_0^2 + \dots + x_{n-1}^2 & x_0 z_1 + \dots + x_{n-1} z_n \\ x_0 z_1 + \dots + x_{n-1} z_n & z_1^2 + \dots + z_n^2 \end{pmatrix}^{-1} \begin{pmatrix} x_0 x_1 + \dots + x_{n-1} x_n \\ z_1 x_1 + \dots + z_n x_n \end{pmatrix}$$

De manera similar a (5.5), pero sin tomar el valor esperado, mostramos las ecuaciones para los coeficientes estimados:

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\gamma} \end{pmatrix} = \begin{pmatrix} \alpha \\ \gamma \end{pmatrix} + \begin{pmatrix} \frac{z_1^2 + \dots + z_n^2}{\Delta} & \frac{-x_0 z_1 - \dots - x_{n-1} z_n}{\Delta} \\ \frac{-x_0 z_1 - \dots - x_{n-1} z_n}{\Delta} & \frac{x_0^2 + \dots + x_{n-1}^2}{\Delta} \end{pmatrix} \begin{pmatrix} x_0 \varepsilon_1 + \dots + x_{n-1} \varepsilon_n \\ z_1 \varepsilon_1 + \dots + z_n \varepsilon_n \end{pmatrix} \quad (5.6)$$

[donde $\Delta = (x_0^2 + \dots + x_{n-1}^2)(z_1^2 + \dots + z_n^2) - (x_0 z_1 + \dots + x_{n-1} z_n)^2$ es el determinante]. Las ecuaciones individuales para $\hat{\alpha}$ y para $\hat{\gamma}$ son, respectivamente:

$$\alpha + \left[\frac{(z_1^2 + \dots + z_n^2)(x_0 \varepsilon_1 + \dots + x_{n-1} \varepsilon_n) - (x_0 z_1 + \dots + x_{n-1} z_n)(z_1 \varepsilon_1 + \dots + z_n \varepsilon_n)}{(x_0^2 + \dots + x_{n-1}^2)(z_1^2 + \dots + z_n^2) - (x_0 z_1 + \dots + x_{n-1} z_n)^2} \right]$$

$$\gamma + \left[\frac{-(x_0 z_1 + \dots + x_{n-1} z_n)(x_0 \varepsilon_1 + \dots + x_{n-1} \varepsilon_n) + (x_0^2 + \dots + x_{n-1}^2)(z_1 \varepsilon_1 + \dots + z_n \varepsilon_n)}{(x_0^2 + \dots + x_{n-1}^2)(z_1^2 + \dots + z_n^2) - (x_0 z_1 + \dots + x_{n-1} z_n)^2} \right]$$

El problema de inconsistencia en la estimación se presenta si hay otra ecuación en el sistema (aunque no la conocemos), por ejemplo $z_t = x_t + D_t$ donde D_t es alguna otra variable, de manera que z_t quede correlacionada con ε_t ; así, los productos $z_i \varepsilon_i$ tenderían a ser del mismo signo y los efectos del término $z_1 \varepsilon_1 + \dots + z_n \varepsilon_n$, dividido por n , perdurarían y producirían estimadores sesgados e inconsistentes para γ ; esto es, $\text{plim} \left(\frac{z_1 \varepsilon_1 + \dots + z_n \varepsilon_n}{n} \right) \neq 0$ (nótese que los términos $x_0 \varepsilon_1 + \dots + x_{n-1} \varepsilon_n$ y $x_0 z_1 + \dots + x_{n-1} z_n$, luego de dividir por n , desaparecerían en el límite de probabilidad).

El modelo VAR

Una alternativa que depende menos de que tengamos la teoría correcta para saber cuáles variables son endógenas y cuáles exógenas, es suponer de entrada que existe un sistema de ecuaciones simultáneas y que *todas las variables son endógenas*, esto es, que dependen unas de otras.

Para formalizar esto, supongamos que hay m variables y que x_t es el vector que las reúne:

$$x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ \vdots \\ x_{mt} \end{pmatrix} \quad (5.7)$$

Donde, de nuevo, todas son endógenas. Se trata de no poner el relieve en las ecuaciones estructurales, sino en la forma reducida y dejar que sean los datos, más que la teoría, los que indiquen cuáles variables se deben incluir o excluir de las ecuaciones. En este enfoque,

la teoría no está del todo ausente ya que, por ejemplo, se requiere para escoger las m variables. Recordamos que la forma reducida es la expresión de cada variable endógena en función de las exógenas o predeterminadas; y en nuestro nuevo sistema (5.7) podemos usar los rezagos de las endógenas como predeterminadas y los errores como exógenas.

Por lo tanto, podemos escribir una ecuación reducida para cada variable del vector en función de sus propios rezagos, de rezagos de las otras variables, y del error. Este conjunto de m ecuaciones conforman lo que se llama un vector autorregresivo, VAR(p), cuya expresión es:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t \quad (5.8)$$

Donde A_0 es un vector $m \times 1$ de constantes, A_1, \dots, A_p son matrices de coeficientes de orden $m \times m$ y donde los m elementos del vector ε_t , también llamados shocks o innovaciones, son individualmente ruido blanco, aunque pueden estar correlacionados contemporáneamente; en general, la matriz de varianza covarianza será $E\varepsilon_t\varepsilon_t' = \Sigma$. También se puede adicionar un vector de variables verdaderamente exógenas.

Por ejemplo, un VAR(3) de dos variables ($m = 2$) y una constante tendrá la expresión:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{01} \\ \alpha_{02} \end{pmatrix} + \begin{pmatrix} \alpha_{1,11} & \alpha_{1,12} \\ \alpha_{1,21} & \alpha_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \alpha_{2,11} & \alpha_{2,12} \\ \alpha_{2,21} & \alpha_{2,22} \end{pmatrix} \begin{pmatrix} x_{1t-2} \\ x_{2t-2} \end{pmatrix} + \begin{pmatrix} \alpha_{3,11} & \alpha_{3,12} \\ \alpha_{3,21} & \alpha_{3,22} \end{pmatrix} \begin{pmatrix} x_{1t-3} \\ x_{2t-3} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (5.9)$$

Donde la ecuación para cada variable será:

$$\begin{aligned} x_{1t} &= \alpha_{01} + \alpha_{1,11}x_{1t-1} + \alpha_{1,12}x_{2t-1} + \alpha_{2,11}x_{1t-2} + \alpha_{2,12}x_{2t-2} \\ &\quad + \alpha_{3,11}x_{1t-3} + \alpha_{3,12}x_{2t-3} + \varepsilon_{1t} \\ x_{2t} &= \alpha_{02} + \alpha_{1,21}x_{1t-1} + \alpha_{1,22}x_{2t-1} + \alpha_{2,21}x_{1t-2} + \alpha_{2,22}x_{2t-2} \\ &\quad + \alpha_{3,21}x_{1t-3} + \alpha_{3,22}x_{2t-3} + \varepsilon_{2t} \end{aligned}$$

Obsérvese que las ecuaciones del VAR son ecuaciones de forma reducida y, como no tienen términos de promedio móvil en el error, pueden ser estimadas individualmente por MCO o por máxima verosimilitud.

Esto simplifica la estimación. Debemos notar que la estimación del sistema de ecuaciones del VAR por mínimos cuadrados generalizados es idéntica al MCO cuando, como en este caso, cada ecuación tiene las mismas variables explicativas (véase Judge, 1988, p. 756).

El valor esperado de cada variable en un VAR depende no sólo de las constantes y del valor esperado de sus propios errores, sino del valor esperado de las constantes y errores de las demás ecuaciones. Por ejemplo, para un VAR(1) estacionario, $x_t = c + Ax_{t-1} + \varepsilon_t$, con $m = 2$ y $E\varepsilon_t = 0$, tendremos:

$$E \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} E \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + E \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

$$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \right] E \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

$$E \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 - \alpha_{11} & -\alpha_{12} \\ -\alpha_{21} & 1 - \alpha_{22} \end{pmatrix}^{-1} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{\Delta} \begin{pmatrix} 1 - \alpha_{22} & \alpha_{12} \\ \alpha_{21} & 1 - \alpha_{11} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Donde $\Delta = (1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12}\alpha_{21}$. Los valores esperados serán:

$$Ex_{1t} = \frac{(1 - \alpha_{22})c_1 + \alpha_{12}c_2}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12}\alpha_{21}}$$

$$Ex_{2t} = \frac{\alpha_{21}c_1 + (1 - \alpha_{11})c_2}{(1 - \alpha_{11})(1 - \alpha_{22}) - \alpha_{12}\alpha_{21}}$$

El número de parámetros que se van a estimar en un VAR puede llegar a ser grande. Hay un vector de m elementos de constantes más p matrices de m^2 coeficientes (una por cada rezago) más $m(m+1)/2$ elementos en la matriz Σ . Esto último se debe a que la matriz Σ es simétrica y sólo es necesario estimar los elementos que están sobre la diagonal y a un lado de ésta, lo cual da $m(m+1)/2$ elementos. En resumen, si hay constante, el número de parámetros será $m + pm^2 + m(m+1)/2$. Por ejemplo, si tenemos tres rezagos ($p = 3$) y dos variables ($m = 2$) se requiere estimar: $2 + 3(2 \times 2) + 3 = 17$ parámetros. Si fueran cuatro variables ($m = 4$) y los mismos ($p = 3$) rezagos tendríamos $4 + 3(4 \times 4) + 10 = 62$ parámetros a estimar. En razón de que el

tamaño n de las muestras en economía no es muy grande, los VAR usualmente no pasan de cuatro o cinco variables y tres o cuatro rezagos. Esto es especialmente crítico en la estimación simultánea. Los rezagos lejanos, por ejemplo, para representar la estacionalidad (12, 24, ..., si es mensual; 4, 8, ..., si es trimestral), se pueden incluir como variables exógenas en muchos programas de computador de manera que no se haga inmanejable el número de rezagos del VAR.

Para seleccionar p en un VAR con n observaciones, pueden minimizarse las versiones vectoriales del criterio de información de Akaike (AIC) y del criterio de información de Schwarz (SC), presentados en el Capítulo 4. Estas versiones vectoriales son:

$$\text{AIC} = \ln \left(\det \hat{\Sigma}_p \right) + \frac{2k}{n} \quad (5.10)$$

$$\text{SC} = \ln \left(\det \hat{\Sigma}_p \right) + \frac{k \ln(n)}{n} \quad (5.11)$$

Donde $\hat{\Sigma}_p$ es el estimador de la matriz de varianza covarianza $E \varepsilon_t \varepsilon_t'$ de orden $m \times m$ del vector de errores que se obtiene del VAR de orden p , y donde k es el número agregado de coeficientes que se va a estimar incluyendo las constantes; esto es, pm^2 si no hay constantes y $pm^2 + m$ si hay constantes. Algunas versiones de estos criterios remplazan el término $\ln(\det \hat{\Sigma}_p)$ por el negativo de la verosimilitud conjunta logarítmica estimada (dividida por n).

Adicional a lo anterior, también es importante que los $\hat{\varepsilon}_t$ resultantes de la estimación del VAR sean ruido blanco (que sean independientes o al menos no tengan correlación serial), lo cual debe influir en la escogencia de p . Esto es especialmente importante porque, al igual que en el AR, en el VAR hay variables dependientes rezagadas, y si hay autocorrelación en los errores, puede resultar en una estimación inconsistente, esto es, no aplica el teorema de Mann-Wald.

Se pueden utilizar las pruebas usuales, la prueba Q o la Breusch-Godfrey de multiplicador de Lagrange para autocorrelación. También existen versiones multivariadas de la prueba Q (Lutkepohl, 1992) y de la prueba de Breusch-Godfrey (Johansen, 1995). Sin embargo, como recuerda Lutkepohl (2005, p. 157), este requisito puede ser menos

importante si el objetivo es la predicción, siempre y cuando el modelo prediga bien.

La metodología de modelaje con VAR fue propuesta por Sims (1980) con el objeto de alejarse un poco de las “increíbles restricciones” que se imponen a los parámetros en la estimación tradicional. Aquí todas las variables dependen de todas. En este enfoque ni siquiera se hace hincapié en la eliminación de rezagos no significativos, ya que las pruebas usuales pueden resultar poco confiables por la multicolinealidad que suelen exhibir las variables explicativas del VAR. Más aún, los proponentes de esta metodología sugieren que no es necesario diferenciar, restar tendencias o desestacionalizar las variables no estacionarias, dado que el objetivo primordial es descubrir relaciones entre variables, más que realizar estimaciones precisas de parámetros (véase Sims, 1980; Doan, 1992).

En contraste, el llamado VAR estructural (SVAR), que veremos más adelante, sí impone restricciones y pone el relieve en la estimación de los parámetros individuales para tratar de identificar y recuperar el modelo estructural. A su vez, los defensores del VAR ateórico reaccionan con sorpresa, pues notan que fue precisamente la práctica de imponer restricciones, casi siempre insostenibles, la que originó la metodología VAR.

Ejemplo

Al igual que los AR y MA univariados, sus contrapartes multivariadas también pueden intercambiarse. En otras palabras, un VAR finito puede convertirse en un VMA infinito y viceversa. Ilustremos convirtiendo un VAR(1) de dos variables en un VMA(∞):

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \alpha_{11}L & \alpha_{12}L \\ \alpha_{21}L & \alpha_{22}L \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Y agrupando términos:

$$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \alpha_{11}L & \alpha_{12}L \\ \alpha_{21}L & \alpha_{22}L \end{pmatrix} \right] \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 - \alpha_{11}L & -\alpha_{12}L \\ -\alpha_{21}L & 1 - \alpha_{22}L \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Al tomar el inverso de la matriz en la expresión anterior, su determinante dividirá cada uno de sus elementos o, alternativamente, dividirá

a ε_t que está al lado derecho. Por lo tanto, la estabilidad del sistema dependerá de que las raíces de $(1 - \alpha_{11}L)(1 - \alpha_{22}L) - \alpha_{12}\alpha_{21}L^2 = 0$ estén por fuera del círculo unitario; un sistema con esta estabilidad es estacionario en covarianza (véase Lutkepohl, 2005, capítulo 2). Al continuar, la ecuación queda:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \frac{1-\alpha_{22}L}{\Delta} & \frac{\alpha_{12}L}{\Delta} \\ \frac{\alpha_{21}L}{\Delta} & \frac{1-\alpha_{11}L}{\Delta} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \quad (5.12)$$

Donde $\Delta = (1 - \alpha_{11}L)(1 - \alpha_{22}L) - \alpha_{12}\alpha_{21}L^2$. Siguiendo con las operaciones algebraicas tendremos:

$$\begin{aligned} &= \begin{pmatrix} \frac{1-\alpha_{22}L}{\Delta} \varepsilon_{1t} + \frac{\alpha_{12}L}{\Delta} \varepsilon_{2t} \\ \frac{\alpha_{21}L}{\Delta} \varepsilon_{1t} + \frac{1-\alpha_{11}L}{\Delta} \varepsilon_{2t} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\varepsilon_{1t}}{\Delta} - \frac{\alpha_{22}\varepsilon_{1t-1}}{\Delta} + \frac{\alpha_{12}\varepsilon_{2t-1}}{\Delta} \\ \frac{\alpha_{21}\varepsilon_{1t-1}}{\Delta} + \frac{\varepsilon_{2t}}{\Delta} - \frac{\alpha_{11}\varepsilon_{2t-1}}{\Delta} \end{pmatrix} \end{aligned}$$

Expresión que, después de las expansiones pertinentes (mucha álgebra) causadas por los factores resultantes de la factorización de los denominadores Δ , tomará la forma:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{1,11} & \beta_{1,12} \\ \beta_{1,21} & \beta_{1,22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \begin{pmatrix} \beta_{2,11} & \beta_{2,12} \\ \beta_{2,21} & \beta_{2,22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-2} \\ \varepsilon_{2t-2} \end{pmatrix} + \dots \quad (5.13)$$

Ejemplo

Para ilustrar el papel de las raíces de la ecuación auxiliar (en este caso el determinante) sobre la estabilidad de un VAR, como se mencionó en la sección anterior, consideramos el modelo:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Este se puede factorizar como:

$$\left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2L & L \\ L & 2L \end{pmatrix} \right] \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Del cual se despeja el vector x_t y se obtiene:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \frac{1-2L}{\Delta} & \frac{L}{\Delta} \\ \frac{L}{\Delta} & \frac{1-2L}{\Delta} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Donde $\Delta = (1 - 2L)(1 - 2L) - L^2$. Las raíces del determinante son 1 y $1/3$, lo que implica que el VAR no es estacionario.

Raíces y eigenvalores

La estabilidad de los modelos ARMA o VAR tiene una correspondencia con el problema de los eigenvalores (valores propios) tratado en álgebra lineal. Específicamente, las raíces de la ecuación auxiliar o del determinante, que hemos venido asociando con la estabilidad del sistema, son los inversos de los eigenvalores, λ . De manera que la estabilidad puede juzgarse por la magnitud de las raíces de la ecuación auxiliar o, alternativamente, por la magnitud los eigenvalores. Expresada en términos de los eigenvalores, la regla de estabilidad es que estos sean menores que 1 en magnitud.

Dada una transformación lineal, pueden existir ciertos vectores o señales de entrada que se preservan al ser sometidos a dicha transformación lineal; esto es, el vector de entrada es multiplicado por un escalar al pasar por el sistema. Se trata de vectores que son repetidos por la transformación; la salida del sistema es igual a la entrada, excepto que se encuentra amplificada o atenuada. Los vectores de entrada para los cuales sucede esto se conocen como *eigenvectores*, y los escalares que los multiplican, como *eigenvalores*. La matriz de coeficientes de un VAR es una de tales transformaciones lineales.

Al ilustrar con la matriz del VAR del ejemplo anterior, $A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$, el problema de los eigenvalores se expresa en que $A_1x = \lambda x$ o $(A_1 - \lambda I)x = 0$. Para obtener una solución donde x sea diferente de cero, es necesario que no se pueda invertir la matriz $A_1 - \lambda I$, lo que, a su vez, implica $\det(A_1 - \lambda I) = 0$. Con la matriz A_1 , tendríamos:

$$\det \left[\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right] = \det \left[\begin{pmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{pmatrix} \right] = \lambda^2 - 4\lambda + 3 = 0,$$

cuya solución arroja los eigenvalores 1 y 3 (los inversos de las raíces del ejemplo anterior) e implica que el VAR no es estable.

Para generalizar a un $\text{VAR}(p)$, $x_t = A_1x_{t-1} + A_2x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t$, definimos:

$$y_t = \begin{pmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{pmatrix} \quad A = \begin{pmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_m & 0 & \dots & 0 & 0 \\ 0 & I_m & \dots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \dots & I_m & 0 \end{pmatrix} \quad u_t = \begin{pmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

Donde A , de orden $mp \times mp$, es la llamada *matriz acompañante*. Con estas definiciones, podemos expresar el $\text{VAR}(p)$ como un $\text{VAR}(1)$:

$$y_t = Ay_{t-1} + u_t$$

en cuya expresión el vector que se reproduce es y_t .

La estacionariedad del $\text{VAR}(p)$ dependerá de que los pm eigenvalores de A sean menores que 1 en magnitud. Estos se obtienen de la ecuación $\det(A - \lambda I_{pm}) = 0$, que, como se ilustra más adelante, implica (véase Hamilton, 1994, capítulo 10):

$$\det(\lambda^p I_m - \lambda^{p-1} A_1 - \lambda^{p-2} A_2 - \dots - A_p) = 0 \quad (5.14)$$

Nótese que si remplazamos $\lambda = 1/L$, resulta la ecuación auxiliar del VAR en términos del operador de rezago:

$$\det(I_m - A_1L - A_2L^2 - \dots - A_pL^p) = 0$$

Ejemplo

Para calcular los eigenvalores de un $\text{VAR}(2)$, $x_t = A_1x_{t-1} + A_2x_{t-2} + \varepsilon_t$, restamos λI_{2m} a la matriz acompañante y escribimos:

$$\det(A - \lambda I_{2m}) = \det \begin{bmatrix} A_1 - \lambda I_m & A_2 \\ I_m & -\lambda I_m \end{bmatrix} = 0$$

La expresión anterior representa el determinante de una matriz en bloques (*partitioned matrix*). Para el determinante y otras fórmulas

relacionadas con estas matrices, véase, por ejemplo, Johnston (1984, capítulo 4) o Lutkepohl (2005, Apéndice). El determinante de cualquier matriz en bloques es:

$$\det \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = \det(M_{22}) \det(M_{11} - M_{12}M_{22}^{-1}M_{21})$$

Donde M_{11} y M_{22} son matrices cuadradas invertibles. Al aplicar esta fórmula tenemos:

$$\det \begin{bmatrix} A_1 - \lambda I_m & A_2 \\ I_m & -\lambda I_m \end{bmatrix} = \det(-\lambda I_m) \det(A_1 - \lambda I_m - A_2(-\lambda I_m)^{-1}I_m)$$

Expresión, que al igualar a cero, da:

$$\det(\lambda^2 I_m - \lambda A_1 - A_2) = 0$$

Y, así, corresponde a la expresión (5.14) con dos rezagos, a partir de la cual se pueden calcular los eigenvalores.

Para un VAR(3), $x_t = A_1x_{t-1} + A_2x_{t-2} + A_3x_{t-3} + \varepsilon_t$, escribimos:

$$\begin{aligned} \det(A - \lambda I_{3m}) &= \det \begin{bmatrix} A_1 - \lambda I_m & A_2 & A_3 \\ I_m & -\lambda I_m & 0 \\ 0 & I_m & -\lambda I_m \end{bmatrix} = \\ &\det \begin{bmatrix} -\lambda I_m & 0 \\ I_m & -\lambda I_m \end{bmatrix} \det \left(A_1 - \lambda I_m - (A_2 \quad A_3) \begin{bmatrix} -\lambda I_m & 0 \\ I_m & -\lambda I_m \end{bmatrix}^{-1} \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right) \\ &= (-\lambda)^{2m} \det \left(A_1 - \lambda I_m - (A_2 \quad A_3) \begin{bmatrix} -\lambda^{-1} I_m & 0 \\ -\lambda^{-2} I_m & -\lambda^{-1} I_m \end{bmatrix} \begin{pmatrix} I_m \\ 0 \end{pmatrix} \right) \\ &= (-\lambda)^{2m} \det(A_1 - \lambda I_m + \lambda^{-1} A_2 + \lambda^{-2} A_3) \end{aligned}$$

Al igualar la expresión anterior a cero y al simplificar podemos obtener los eigenvalores a partir de:

$$\det(\lambda^3 I_m - \lambda^2 A_1 - \lambda A_2 - A_3) = 0$$

El modelo VARMA

Una generalización del modelo VAR consiste en adicionar rezagos del error con lo cual se conforma la expresión de un ARMA vectorial, VARMA:

$$x_t = A_1 x_{t-1} + \dots + A_p x_{t-p} + \varepsilon_t + B_1 \varepsilon_{t-1} + \dots + B_q \varepsilon_{t-q} \quad (5.15)$$

Donde x_t y ε_t son vectores de m variables cada uno y A_1, \dots, A_p y B_1, \dots, B_q son matrices de coeficientes de tamaño $m \times m$. El vector de errores ε_t tendrá matriz de varianza covarianza $E\varepsilon_t\varepsilon_t' = \Sigma$. También es posible, por ejemplo, agregar un vector de constantes o un vector de variables verdaderamente exógenas. Obviamente, si q es cero, tendremos un VAR(p). Igualmente, si p es cero, la expresión sólo tendrá la parte de promedio móvil y se llamará VMA(q).

Para familiarizarnos con estos modelos, escribamos la expresión de un vector ARMA(1, 1) con $m = 2$ y constante:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix}$$

Donde las ecuaciones de cada variable serán:

$$\begin{aligned} x_{1t} &= c_1 + \alpha_{11} x_{1t-1} + \alpha_{12} x_{2t-1} + \varepsilon_{1t} + \beta_{11} \varepsilon_{1t-1} + \beta_{12} \varepsilon_{2t-1} \\ x_{2t} &= c_2 + \alpha_{21} x_{1t-1} + \alpha_{22} x_{2t-1} + \varepsilon_{2t} + \beta_{21} \varepsilon_{1t-1} + \beta_{22} \varepsilon_{2t-1} \end{aligned} \quad (5.16)$$

Un vector ARMA se puede estimar ecuación por ecuación por máxima verosimilitud o por el método aproximado de Hannan-Rissanen. Recientemente se ha utilizado el filtro de Kalman para derivar la función de verosimilitud de este proceso (por ejemplo, véase Hamilton, 1994, capítulo 13). También se puede hacer la estimación de manera simultánea por máxima verosimilitud (*full information maximum likelihood*). En cuanto al método de regresiones aparentemente no relacionadas (SUR), éste no será de beneficio mientras todas las ecuaciones tengan las mismas variables, sin importar si $E\varepsilon_{it}\varepsilon_{jt} \neq 0$. Si hay restricciones que excluyan algunas variables de algunas ecuaciones entonces el método SUR será conveniente.

Usando el operador de rezago, vemos que la ecuación del VARMA (5.15) puede escribirse como:

$$(I - A_1 L - A_2 L^2 - \dots - A_p L^p) x_t = (I + B_1 L + B_2 L^2 + \dots + B_q L^q) \varepsilon_t$$

$$A(L)x_t = B(L)\varepsilon_t \quad (5.17)$$

Donde $A(L)$ y $B(L)$ son polinomios de matrices en L .

El modelo será estacionario, no explosivo, si las raíces del determinante de $A(L)$ están por fuera del círculo unitario, y será invertible si las raíces del determinante de $B(L)$ están por fuera del círculo unitario. Esto se ilustra a continuación.

Intuición para la estabilidad o invertibilidad de un VARMA

Supongamos un vector ARMA(1, 1):

$$\begin{aligned} x_t &= A_1x_{t-1} + \varepsilon_t + B_1\varepsilon_{t-1} \\ (I - A_1L)x_t &= \varepsilon_t + B_1\varepsilon_{t-1} \\ x_t &= (I - A_1L)^{-1}(\varepsilon_t + B_1\varepsilon_{t-1}) \end{aligned} \quad (5.18)$$

La condición de estacionariedad (poder expresar x_t en función del pasado de ε_t de manera convergente) dependerá de las raíces del polinomio que divide el término $\varepsilon_t + B_1\varepsilon_{t-1}$ en (5.18). Este polinomio provendrá del denominador de los términos de $(I - A_1L)^{-1}$. Y el denominador de los términos de esta matriz inversa, como cualquier matriz inversa de cualquier orden, es el determinante de la matriz $I - A_1L$. De manera que si las raíces de $\det(I - A_1L)$ o, en general, de $\det A(L)$ están fuera del círculo unitario, habrá estacionariedad. Obviamente, esta condición aplica tanto a un VAR como a un VARMA.

Un argumento similar con un MA(1) vectorial sirve para generalizar. Si las raíces de $\det B(L)$ están fuera del círculo unitario, el proceso VMA(q) será invertible.

A diferencia del caso univariado, donde la estabilidad o invertibilidad de una variable depende únicamente de los coeficientes que multiplican sus rezagos o los del error, en el caso multivariado, donde las variables son endógenas, estas propiedades dependen, en general, de coeficientes de todas las ecuaciones.

Respuesta impulso del VAR

Recordemos que en el caso del AR(p) univariado, un choque unitario $\varepsilon_0 = 1$ generaba una respuesta a través del tiempo en x_t , como se

mostró en (3.42), llamada *respuesta de impulso*. Intuitivamente, hacer $\varepsilon_0 = 1$, y de ahí en adelante todos los demás ε_t igual a cero, es como darle un golpe instantáneo (impulso) al sistema y sentarse a ver cuál es la respuesta a ese impulso. Algo similar se puede hacer para el caso multivariado, al dar un choque unitario a uno de los m errores del vector ε_t . Este ejercicio sirve también para estudiar la estabilidad del sistema. Si el sistema es estable, el efecto del choque decae con el tiempo.

Para ilustrar, usamos el VAR(1) de dos variables ($m = 2$), donde hacemos $\varepsilon_{20} = 1$ y mantenemos en cero todos los demás ε_t , incluido $\varepsilon_{10} = 0$. Adicionalmente, suponemos que $x_t = 0$ para $t < 0$ (en realidad para $t < 0$ se pone $E x_t$, pero en este caso es cero). Podemos seguir la evolución del sistema paso a paso para $t \geq 0$:

$$\begin{aligned} \begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} &= \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \begin{pmatrix} x_{11} \\ x_{21} \end{pmatrix} &= \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \end{pmatrix} \\ \begin{pmatrix} x_{12} \\ x_{22} \end{pmatrix} &= \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} \alpha_{12} \\ \alpha_{22} \end{pmatrix} = \begin{pmatrix} \alpha_{11}\alpha_{12} + \alpha_{12}\alpha_{22} \\ \alpha_{21}\alpha_{12} + \alpha_{22}^2 \end{pmatrix} \\ &\vdots \end{aligned} \tag{5.19}$$

También se acostumbra, en lugar de hacer ε_{j0} (donde j es un dígito entre 1 y m) igual a la unidad, usar ε_{j0} igual a una desviación estándar de ε_{jt} , y mantener, como siempre, los demás ε_{jt} en cero.

Por la interrelación de variables que hay en un VAR, el interés radica no sólo en la respuesta a un choque en una misma variable, sino en la respuesta de una variable a un choque en otra. Es común dar un choque a una variable y graficar la respuesta de ésa y de las demás variables a través del tiempo.

La respuesta impulso también se puede obtener si recordamos que un VAR(p) estacionario tiene una representación VMA(∞), la cual podemos aprovechar. Al igual que en el caso univariado, los coeficientes del MA(∞) son iguales a la respuesta impulso si se utiliza un impulso unitario, o proporcionales si se utiliza un impulso igual a la desviación estándar. Por ejemplo, considerando (5.13), la aplicación de un choque

a uno solo de los errores, digamos $\varepsilon_{20} = 1$, permite observar la respuesta de x_{1t} y x_{2t} a través del tiempo.

El entusiasmo inicial que esta técnica de respuesta impulso generó entre los econométristas ha disminuido recientemente, aunque los algoritmos para calcular la respuesta impulso continúan siendo parte de los programas computacionales para el análisis de series de tiempo. Ocurre que, con frecuencia, ε_{1t} y ε_{2t} están correlacionados contemporáneamente de manera que resulta contradictorio suponer, por ejemplo, que $\varepsilon_{20} = 1$ mientras ε_{10} se mantiene en cero (algo así como *ceteris paribus*); si hay correlación y ε_{20} aumenta de valor entonces ε_{10} tiende a moverse. Por lo tanto, si hay correlación entre los errores, no es claro qué significado puede asociarse a la respuesta impulso. Este problema se ha tratado de solucionar convirtiendo la matriz de varianza covarianza Σ del error en una matriz diagonal eliminando así la correlación entre los errores. Veamos.

Al generalizar (5.13), un AR(p) tiene representación MA(∞) así:

$$x_t = \varepsilon_t + B_1\varepsilon_{t-1} + B_2\varepsilon_{t-2} + \dots \quad (5.20)$$

Donde las B_i son las matrices de coeficientes. Sabemos que $\Sigma = E\varepsilon\varepsilon'$ es simétrica positiva definida, como toda matriz de varianza covarianza. También sabemos que para toda matriz simétrica positiva definida existe una matriz P no singular tal que $PP' = \Sigma$, de donde $P^{-1}\Sigma P^{-1'} = I$ (por ejemplo, véase Johnston, 1984, p. 152). Sin alterar la igualdad (5.20) escribimos:

$$\begin{aligned} x_t &= (P)(P^{-1}\varepsilon_t) + (B_1P)(P^{-1}\varepsilon_{t-1}) + (B_2P)(P^{-1}\varepsilon_{t-2}) + \dots \\ &= \psi_0\nu_t + \psi_1\nu_{t-1} + \psi_2\nu_{t-2} + \dots \end{aligned} \quad (5.21)$$

Donde $\psi_i = (B_iP)$ y $\nu_i = (P^{-1}\varepsilon_i)$. De manera que ν_i es el nuevo término de error y ψ_i la nueva matriz de coeficientes. La matriz de varianza covarianza de ν_i será diagonal porque:

$$\Sigma_\nu = E\nu_i\nu_i' = EP^{-1}(\varepsilon_i\varepsilon_i')P^{-1'} = P^{-1}\Sigma P^{-1'} = I \quad (5.22)$$

donde la varianza es unitaria y se ha eliminado la correlación cruzada.

Es de notar que la matriz P no es única y que al cambiarla tendremos variaciones en la respuesta impulso. Una opción para obtener

P es a partir de los vectores característicos de Σ . Otra alternativa, bastante popular, es encontrar la matriz P por medio de la llamada *descomposición de Choleski*, que produce una matriz P triangular inferior, con diagonal positiva, tal que $PP' = \Sigma$. Sin embargo, como se verá más adelante, surge un problema adicional porque P es triangular y convierte ε_i en $\nu_i = P^{-1}\varepsilon_i$ de manera que la respuesta impulso podría depender del orden en que se pongan las variables x_1, \dots, x_m en el VAR. Recordamos que si P es triangular inferior, P^{-1} también lo será. Es de notar que como las varianzas en (5.22) son unitarias, un impulso unitario en ν_i automáticamente es un impulso de una desviación estándar.

Ilustremos la discusión anterior con un ejemplo sencillo para $m = 2$. Sea:

$$P = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix}$$

Y la expresión VMA:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t-1} \\ \varepsilon_{2t-1} \end{pmatrix} + \dots$$

Los nuevos errores serán, llamando

$$P^{-1} = \begin{pmatrix} \bar{P}_{11} & 0 \\ \bar{P}_{21} & \bar{P}_{22} \end{pmatrix},$$

$$\begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} = \begin{pmatrix} \bar{P}_{11} & 0 \\ \bar{P}_{21} & \bar{P}_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} = \begin{pmatrix} \bar{P}_{11}\varepsilon_{1t} \\ \bar{P}_{21}\varepsilon_{1t} + \bar{P}_{22}\varepsilon_{2t} \end{pmatrix}$$

Ahora sí se puede incrementar un solo ν_{it} porque ν_{1t} y ν_{2t} no están correlacionados. De (5.21) podemos escribir:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \nu_{1t-1} \\ \nu_{2t-1} \end{pmatrix} + \dots \quad (5.23)$$

Nótese que la respuesta impulso, el conjunto de los ψ de la ecuación (5.21), depende de cuál variable se ponga primero. Si en el ejemplo anterior cambiamos y ponemos $\begin{pmatrix} x_{2t} \\ x_{1t} \end{pmatrix}$, la respuesta impulso puede alterarse. Se observa del primer término a la derecha en (5.23) que

ν_{1t} , el error asociado con x_{1t} , afecta tanto a x_{1t} como a x_{2t} de manera contemporánea pero el error asociado con x_{2t} no afecta a x_{1t} de manera contemporánea. Se dice entonces que x_{1t} tiene precedencia sobre x_{2t} , lo cual se refleja en la forma triangular de la matriz P . En otras palabras, para efectos de la respuesta impulso, deberá colocarse primero en el VAR aquella variable con precedencia; esto es, aquella que parece tener un efecto contemporáneo sobre las otras. Hay que admitir que escoger un ordenamiento implica inevitablemente establecer una restricción paramétrica, la cual puede apoyarse en la teoría o en la evidencia o, en su defecto, ser algo arbitraria.

Estamos atribuyendo toda la variación común en los errores ε_{it} a la variable puesta de primero en el VAR. La respuesta de los diferentes ordenamientos converge en el tiempo si el sistema es estable. También debe tenerse presente que si la correlación entre ε_{1t} y ε_{2t} es insignificante, el ordenamiento será irrelevante, y que cuanto más alta la correlación, mayor será el efecto del cambio en el ordenamiento.

Existen otras alternativas de ortogonalización que tratan de independizarse del orden de las variables en el VAR (por ejemplo, véase Pesaran y Shin, 1998). También los VAR estructurales ofrecen una alternativa de ortogonalización.

Ejemplo

Ilustración de cómo el orden cambia la respuesta impulso. Supongamos que tenemos el sistema:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{pmatrix} \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} + \dots$$

Si damos un *shock* a x_2 haciendo $\nu_{20} = 1$ y se mantienen los demás errores ν_{jt} en cero, entonces para el primer periodo tendremos:

$$\begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} = \begin{pmatrix} 0 \\ P_{22} \end{pmatrix}$$

Pero si invertimos el orden de las variables tendremos:

$$\begin{pmatrix} x_{2t} \\ x_{1t} \end{pmatrix} = \begin{pmatrix} d_{11} & 0 \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} u_{2t} \\ u_{1t} \end{pmatrix} + \dots$$

Donde la nueva matriz puede diferir de la anterior aunque sigue siendo triangular inferior. Si ahora damos un *shock* a x_2 haciendo $u_{20} = 1$ y los demás $u_{jt} = 0$, tenemos:

$$\begin{pmatrix} x_{20} \\ x_{10} \end{pmatrix} = \begin{pmatrix} d_{11} \\ d_{21} \end{pmatrix}$$

De manera que x_{10} ya no arranca, para un *shock* a x_2 , en cero como en el ordenamiento anterior.

Ejemplo

Ilustración de la descomposición de Choleski. Sea:

$$E\epsilon\epsilon' = \Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}$$

La descomposición de Choleski encuentra una matriz P triangular inferior, tal que $PP' = \Sigma$. En este caso, una opción es:

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix} = PP'$$

Otra posibilidad es:

$$\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & -2 \end{pmatrix} = PP'$$

Ambas sirven desde el punto de vista matemático, aunque esta última cambiaría el signo del efecto de las innovaciones; para la descomposición de Choleski se toma la de diagonal positiva porque P es similar a una raíz cuadrada.

Siguiendo con este ejemplo, mostremos cómo los nuevos errores ν_{1t} y ν_{2t} resultan sin correlación.

Para $E\epsilon\epsilon' = \Sigma = \begin{pmatrix} 1 & 2 \\ 2 & 8 \end{pmatrix}$ y $P = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}$ tenemos que $P^{-1} = \begin{pmatrix} 1 & 0 \\ -1 & \frac{1}{2} \end{pmatrix}$, de manera que:

$$\begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} = \begin{pmatrix} \epsilon_{1t} \\ -\epsilon_{1t} + \frac{\epsilon_{2t}}{2} \end{pmatrix}$$

Con lo cual:

$$E\nu_{1t}\nu_{2t} = E \left[\varepsilon_{1t} \left(-\varepsilon_{1t} + \frac{\varepsilon_{2t}}{2} \right) \right] = -E\varepsilon_{1t}^2 + E \frac{\varepsilon_{1t}\varepsilon_{2t}}{2}$$

De Σ vemos que $E\varepsilon_{1t}^2 = 1$ y $E\varepsilon_{1t}\varepsilon_{2t} = 2$; por lo tanto, $E\nu_{1t}\nu_{2t} = -1 + \frac{2}{2} = 0$, y no hay correlación. Por otro lado, $E\nu_{1t}^2 = E\varepsilon_{1t}^2 = 1$ (de Σ) y para ν_{2t} :

$$E\nu_{2t}^2 = E \left(-\varepsilon_{1t} + \frac{\varepsilon_{2t}}{2} \right)^2 = E \left(\varepsilon_{1t}^2 - \varepsilon_{1t}\varepsilon_{2t} + \frac{\varepsilon_{2t}^2}{4} \right) = 1 - 2 + \frac{8}{4} = 1$$

De manera que $\Sigma_\nu = P^{-1}E(\varepsilon\varepsilon')P^{-1'} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$.

La respuesta impulso para este ejemplo, dando un *shock* a x_{2t} (esto es, haciendo $\nu_{20} = 1$ y $\nu_{10} = 0$) comenzará en:

$$\begin{pmatrix} x_{10} \\ x_{20} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix} .$$

Ahora supongamos que se cambia el orden de $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ por $\begin{pmatrix} x_2 \\ x_1 \end{pmatrix}$, de manera que, en consecuencia, $\Sigma = \begin{pmatrix} 8 & 2 \\ 2 & 1 \end{pmatrix}$. La nueva descomposición podría ser:

$$\begin{pmatrix} 8 & 2 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{8} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} \sqrt{8} & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{pmatrix} = PP'$$

Donde vemos que P , de todas maneras, es una matriz triangular inferior. Encontramos que la respuesta impulso es diferente con este ordenamiento, porque al dar un *shock* a x_{2t} tendremos que comienza en:

$$\begin{pmatrix} x_{20} \\ x_{10} \end{pmatrix} = \begin{pmatrix} \sqrt{8} & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \sqrt{8} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

Ejemplo

El Cuadro 5.1 muestra la estimación de un VAR(4) con las variables trimestrales tasa de interés de préstamos (*coloc*) y tasa de interés de captación (*capta*). La Figura 5.1 muestra la respuesta impulso para el ordenamiento *coloc* y *capta*; mientras que la Figura 5.2, la respuesta impulso para el ordenamiento contrario. Cada figura muestra la respuesta de la variable endógena indicada a innovaciones de una desviación estándar, tanto en esa misma variable como en la otra. La correlación entre residuos es de 0.87, por lo cual la respuesta impulso es sensible al ordenamiento. Para el cálculo de intervalos de confianza, véase Lutkepohl (2005, capítulo 3). También es posible estimar intervalos de confianza utilizando la técnica *bootstrap* con los residuos originales para construir VAR y respuestas impulso (Enders, 2004, p. 277).

Cuadro 5.1: Estimación de un VAR

Date: 09/22/99 Time: 11:36		
Sample(adjusted): 1988:1 1999:1		
Included observations: 45 after adjusting endpoints		
Standard errors & t-statistics in parentheses		
	COLOC	CAPTA
COLOC(-1)	0.062296 (0.32953) (0.189050)	-0.214216 (0.32141) (-0.666490)
COLOC(-2)	0.456099 (0.340090) (1.341120)	0.577207 (0.331710) (1.740110)
COLOC(-3)	-1.032818 (0.353140) (-2.924690)	-1.093478 (0.344440) (-3.174690)

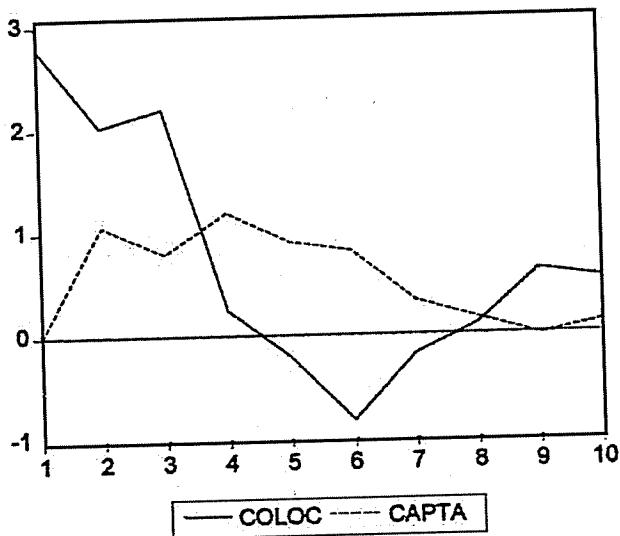
Continúa

Continuación

COLOC(-4)	0.040294 (0.432020) (0.093270)	0.142068 (0.421380) (0.337150)
CAPTA(-1)	0.784830 (0.330460) (2.374970)	1.083511 (0.322320) (3.361840)
CAPTA(-2)	-0.314300 (0.382310) (-0.822100)	-0.432458 (0.372890) (-1.159750)
CAPTA(-3)	0.378475 (0.384680) (0.983870)	0.377804 (0.375200) (1.006940)
CAPTA(-4)	0.255666 (0.408340) (0.626110)	0.260519 (0.398280) (0.654120)
C	26.494730 (9.066980) (2.922110)	15.152780 (8.843540) (1.713430)
R-squared	0.642930	0.673129
Adj. R-squared	0.563581	0.600491
Sum sq. Resids	348.670700	331.698000
S. E. equation	3.112121	3.035430
Log likelihood	-109.920200	-108.797400
Akaike AIC	5.285342	5.235439
Schwarz SC	5.646675	5.596772
Mean dependent	41.222220	30.860000
S. D. dependent	4.710910	4.802386
Determinant Residual Covariance		14.05698
Log Likelihood		-187.1746
Akaike Information Criteria		9.118873
Schwarz Criteria		9.841538

Figura 5.1: Respuesta impulso con el ordenamiento COLOC, CAPTA

Response of COLOC to One S.D. Innovations



Response of CAPTA to One S.D. Innovations

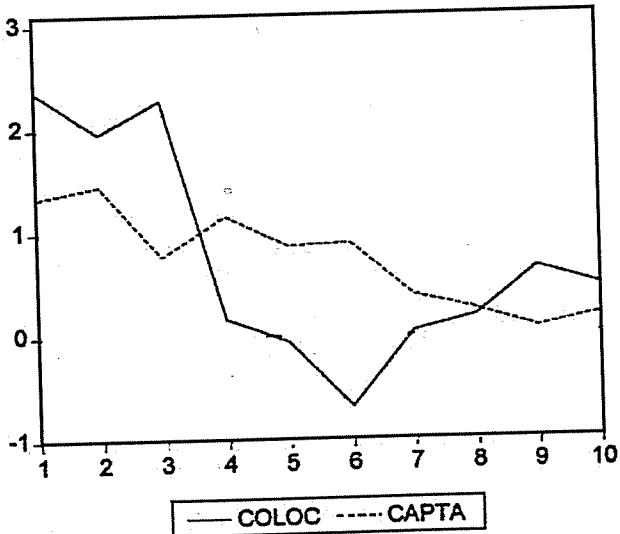
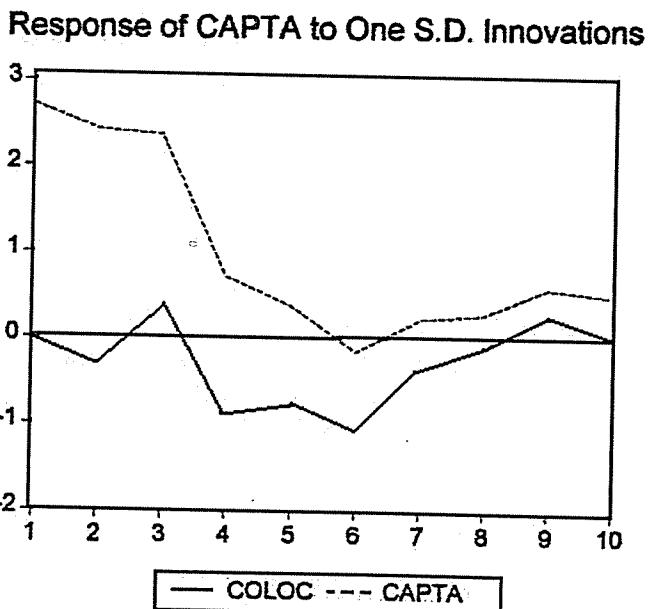
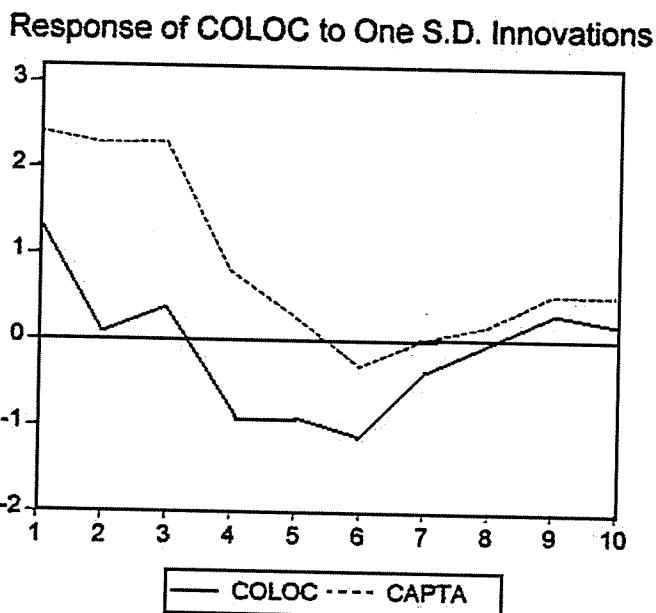


Figura 5.2: Respuesta impulso con el ordenamiento CAPTA, COLOC



Predicción con un VAR(p)

La predicción en un VAR es similar a la predicción utilizando un AR. En el VAR:

$$x_t = c + A_1 x_{t-1} + \dots + A_p x_{t-p} + \varepsilon_t \quad (5.24)$$

es usual postular que $E\varepsilon_t = 0$, por lo cual se supone $\varepsilon_t = 0$ para $t > n$. Al llamar \hat{x}_{n+T} la predicción de x_{n+T} hecha en el tiempo n para el tiempo $n+T$, tenemos:

$$\begin{aligned} \hat{x}_{n+1} &= c + A_1 x_n + A_2 x_{n-1} + \dots + A_p x_{n+1-p} \\ \hat{x}_{n+2} &= c + A_1 \hat{x}_{n+1} + A_2 x_n + \dots + A_p x_{n+2-p} \\ &\vdots \end{aligned} \quad (5.25)$$

Ejemplo

Para un VAR(1) con $m = 2$:

$$\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} + \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$$

Al usar las ecuaciones individuales para la predicción, tendríamos:

$$\begin{aligned} \hat{x}_{1,n+1} &= c_1 + \alpha_{11} x_{1,n} + \alpha_{12} x_{2,n} \\ \hat{x}_{2,n+1} &= c_2 + \alpha_{21} x_{1,n} + \alpha_{22} x_{2,n} \\ \hat{x}_{1,n+2} &= c_1 + \alpha_{11} \hat{x}_{1,n+1} + \alpha_{12} \hat{x}_{2,n+1} \\ \hat{x}_{2,n+2} &= c_2 + \alpha_{21} \hat{x}_{1,n+1} + \alpha_{22} \hat{x}_{2,n+1} \\ &\vdots \end{aligned}$$

Varianza de predicción

Expresando el VAR como su contraparte VMA(∞):

$$x_t = \sum_{j=0}^{\infty} B_j \varepsilon_{t-j} \quad (5.26)$$

Donde $E\epsilon\epsilon' = \Sigma$ y donde $B_0 = I$. Para el periodo $n+T$ escribimos x_{n+T} en dos partes:

$$x_{n+T} = \sum_{j=0}^{T-1} B_j \epsilon_{n+T-j} + \sum_{j=T}^{\infty} B_j \epsilon_{n+T-j} \quad (5.27)$$

Vemos que la predicción \hat{x}_{n+T} será igual al segundo término del lado derecho, porque sus datos son conocidos para $j \geq T$, de manera que el error de predicción para x_{n+T} es el primer término del lado derecho:

$$\hat{\epsilon}_{n+T} = x_{n+T} - \hat{x}_{n+T} = \sum_{j=0}^{T-1} B_j \epsilon_{n+T-j} \quad (5.28)$$

Para ilustrar la expresión anterior, supongamos un sistema de dos variables x_{1t} , x_{2t} . El error de predicción para, digamos x_{1t} , será:

$$\begin{aligned} \hat{\epsilon}_{1,n+T} = & \epsilon_{1,n+T} + \beta_{1,11}\epsilon_{1,n+T-1} + \beta_{1,12}\epsilon_{2,n+T-1} + \dots + \beta_{T-1,11}\epsilon_{1,n+1} \\ & + \beta_{T-1,12}\epsilon_{2,n+1} \end{aligned} \quad (5.29)$$

Al elevar al cuadrado esta expresión y al tomar el valor esperado, obtendremos la varianza del error de la predicción de $\hat{x}_{1,n+T}$ en función de, o con contribución de, los errores propios ϵ_1 , así como los errores ϵ_2 de la otra variable del sistema y de los productos cruzados contemporáneos de ambos errores. Esta varianza o desviación se puede utilizar para formar intervalos de confianza de predicción.

De vuelta al caso general, se puede mostrar a partir de (5.28) que la matriz de varianza covarianza del error de predicción Σ_{n+T} es:

$$\begin{aligned} \Sigma_{n+T} &= E\hat{\epsilon}_{n+T}\hat{\epsilon}'_{n+T} \\ &= E(\epsilon_{n+T} + B_1\epsilon_{n+T-1} + \dots + B_{T-1}\epsilon_{n+1})(\epsilon'_{n+T} + \epsilon'_{n+T-1}B'_1 \\ &\quad + \dots + \epsilon'_{n+1}B'_{T-1}) \\ &= \Sigma + B_1\Sigma B'_1 + B_2\Sigma B'_2 + \dots + B_{T-1}\Sigma B'_{T-1} \end{aligned} \quad (5.30)$$

La expresión anterior supone que los coeficientes de las matrices B_j son conocidos. Una expresión más completa de la varianza introduciría el efecto adicional del error de estimación usual de los parámetros del modelo.

Descomposición de la varianza del error de predicción

Es posible descomponer la expresión de la varianza del error de predicción de una variable de manera que sepamos cuál es la contribución del error ϵ de la misma variable y de las otras variables. En otras palabras, queremos saber qué porcentaje de la varianza de predicción de x_{it} se debe a su propio ϵ_{it} y qué porcentaje se debe al error de cada una de las otras variables. Este ejercicio de descomposición de la varianza del error de predicción se incluye regularmente en los programas computacionales.

La descomposición se basa en una ortogonalización similar a la aplicada en (5.21) para obtener la respuesta impulso y dependerá de la matriz P . Esto es necesario para separar la contribución de cada error, ya que en (5.30) se encuentran mezclados en las tres matrices que componen $B_i \Sigma B_i'$, las cuales convertiremos en las dos matrices $\psi_k \psi_k'$, como se verá a continuación. Para descomponer la varianza transformamos la expresión (5.30) recordando que $PP' = \Sigma$:

$$\begin{aligned}\Sigma_{n+T} &= PP' + B_1 PP' B_1' + \dots + B_{T-1} PP' B_{T-1}' \\ &= \psi_0 \psi_0' + \psi_1 \psi_1' + \dots + \psi_{T-1} \psi_{T-1}'\end{aligned}\tag{5.31}$$

Donde $\psi_k = B_k P$. El i -ésimo elemento de la diagonal de $\psi_k \psi_k'$ es igual a la suma de los cuadrados de los elementos de la fila i -ésima de ψ_k . La suma de todos los i -ésimos elementos de las diagonales de las matrices, $\psi_0 \psi_0', \dots, \psi_{T-1} \psi_{T-1}'$, es la varianza del error de predicción de la i -ésima variable. La contribución de los errores de la variable j -ésima a la varianza i -ésima es:

$$\psi_{0,ij}^2 + \psi_{1,ij}^2 + \dots + \psi_{T-1,ij}^2\tag{5.32}$$

Donde $\psi_{k,ij}$ es el elemento de la fila i columna j de la matriz ψ_k . De esta manera, es posible asignar la contribución de los errores de todas las variables a la varianza de la i -ésima variable.

Los temas de respuesta impulso y descomposición de la varianza se conocen, en general, con el nombre de contabilidad de innovaciones (*innovation accounting*).

Vectores autorregresivos estructurales

En los vectores autorregresivos estructurales (SVAR) se permite la interacción contemporánea entre variables; esto es, una variable puede depender no sólo de sus propios rezagos y de los rezagos de las demás variables (como en un VAR convencional) sino, del valor contemporáneo de las demás variables. Los primeros SVAR fueron desarrollados por Sims (1986) y Bernanke (1986) y ofrecían una alternativa de ortogonalización de los errores para elaborar respuestas impulsivas.

Supongamos que la realidad se puede modelar a través del siguiente VAR estructural:

$$B_0 x_t = B_1 x_{t-1} + B_2 x_{t-2} + \dots + B_p x_{t-p} + \omega_t \quad (5.33)$$

Donde x_t es un vector de m variables estacionarias. La matriz B_0 , de orden $m \times m$, permite que las variables de x_t tengan dependencia contemporánea. Utilizando un SVAR con $m = 2$ y $p = 1$ para ilustrar, tenemos que (5.33) queda:

$$\begin{pmatrix} b_{0,11} & b_{0,12} \\ b_{0,21} & b_{0,22} \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} b_{1,11} & b_{1,12} \\ b_{1,21} & b_{1,22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix}$$

Donde, por ejemplo, puede observarse que x_{1t} es función contemporánea de x_{2t} y viceversa. Si omitimos variables contemporáneas, siendo relevantes, los errores recogerán este efecto y pueden generar correlación en los errores de una ecuación a otra.

Un supuesto fundamental de los SVAR es que la matriz de varianza covarianza del vector de errores estructurales, $E(\omega_t \omega_t')$ = Ω , es diagonal. Estos errores estructurales ortogonales ω_t se relacionan con los errores ε_t de la forma reducida a través de:

$$\omega_t = B_0 \varepsilon_t \quad (5.34)$$

De cierta forma, aquí B_0 cumple el papel de P^{-1} en la descomposición de Choleski; es la matriz que ortogonaliza los errores de la forma reducida, aunque en esta formulación no está restringida a ser triangular inferior. Con los errores ε_t , estimados de la forma reducida, y con una estimación de B_0 , podemos recuperar los errores estructurales

ortogonales ω_t y realizar respuestas impulso. Infortunadamente, como en cualquier otro método de ortogonalización, la estimación de B_0 depende de la imposición de restricciones.

La expresión (5.33) representa un sistema de ecuaciones simultáneas (donde x_{1t} y x_{2t} son funciones de ambos errores ω_{1t} y ω_{2t}) por lo que su estimación por MCO resulta sesgada e inconsistente, y es preferible estimar la forma reducida. La forma reducida se obtiene premultiplicando el SVAR por el inverso de B_0 , esto es:

$$x_t = B_0^{-1}B_1x_{t-1} + B_0^{-1}B_2x_{t-2} + \dots + B_0^{-1}B_p x_{t-p} + B_0^{-1}\omega_t \quad (5.35)$$

O, en nuestra notación usual (donde $A_i = B_0^{-1}B_i$):

$$x_t = A_1x_{t-1} + A_2x_{t-2} + \dots + A_p x_{t-p} + \varepsilon_t \quad (5.36)$$

De manera que los errores de la forma reducida, ε_t , con matriz de varianza covarianza $E(\varepsilon_t\varepsilon_t')$ = Σ , son una combinación lineal de los errores de la forma estructural, esto es:

$$\varepsilon_t = B_0^{-1}\omega_t \quad (5.37)$$

Si bien los errores de la forma estructural se suponen ortogonales, los de la forma reducida, al ser función de los estructurales, estarán correlacionados entre ellos. En otras palabras, para realizar un análisis estructural o para encontrar la respuesta impulso a partir de los errores estructurales ortogonales, será necesario identificar (estimar) los errores ω_t a partir de la información contenida en la estimación de la forma reducida; esto es, será necesario estimar los elementos de la matriz B_0 .

De (5.37) obtenemos que la matriz de varianza covarianza de ε_t se puede expresar como:

$$\Sigma = B_0^{-1}\Omega B_0^{-1'} \quad (5.38)$$

Ello nos genera un sistema de m^2 ecuaciones donde los m^2 elementos de Σ son conocidos (estimables de la forma reducida). Las incógnitas serán los m^2 elementos de B_0 más los m elementos de la matriz diagonal Ω . Pero en razón de que ambos lados de la ecuación son matrices simétricas, sólo tenemos $m(m+1)/2$ ecuaciones independientes; de manera que será necesario hacer $m^2 + m - m(m+1)/2 = m(m+1)/2$

restricciones adicionales para lograr la identificación de B_0 . Notamos que, en general, las ecuaciones de la expresión (5.38) son no lineales en los elementos de B_0^{-1} , por lo que una solución puede no existir.

Al ilustrar con el ejemplo bivariado, (5.38) queda:

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{pmatrix} b_{0,11} & b_{0,12} \\ b_{0,21} & b_{0,22} \end{pmatrix}^{-1} \begin{bmatrix} \Omega_{11} & 0 \\ 0 & \Omega_{22} \end{bmatrix} \begin{pmatrix} b_{0,11} & b_{0,12} \\ b_{0,21} & b_{0,22} \end{pmatrix}^{-1'}$$

En la literatura hay dos tipos de restricciones usuales para la identificación de B_0 : de corto plazo y de largo plazo. Las de corto plazo se realizan directamente sobre los elementos de B_0 . Una posibilidad es normalizar la diagonal de B_0 haciendo que sea unitaria, lo cual representa m restricciones. Así, en cada ecuación queda una variable en tiempo t con coeficiente uno (en el ejemplo bivariado, $b_{0,11} = b_{0,22} = 1$). Alternativamente, podríamos normalizar los errores ω_t haciendo iguales a uno los m elementos de la diagonal de Ω , pero en este caso habría que dejar la diagonal de B_0 sin normalizar. No se pueden hacer ambas restricciones simultáneamente, porque hay dependencia entre la varianza de ω_t y los coeficientes de B_0 .

Para completar las $\frac{m(m+1)}{2} - m = \frac{m(m-1)}{2}$ restricciones que faltan se pueden hacer cero algunos de los elementos de B_0 , ya sea de manera arbitraria o con base en la teoría pertinente, o podemos apelar a restricciones de largo plazo.

Las restricciones de largo plazo consisten en suponer neutralidad, esto es, algunos elementos de ω_t no afectan algunas de las variables de x_t de manera acumulativa en el largo plazo (Blanchard y Quah, 1989). Por ejemplo, es común suponer que los shocks a las variables nominales no afectan a las variables reales en el largo plazo. Se trata de que la respuesta impulso acumulativa sea cero, esto es, que para un impulso dado en el tiempo $t = 0$ se cumpla:

$$\sum_{t=0}^{\infty} x_t = 0 \quad (5.39)$$

Utilizando el VMA equivalente del SVAR (5.35), podemos escribir:

$$\begin{aligned} x_t &= B_0^{-1} [B_1 L + B_2 L^2 + \dots + B_p L^p] x_t + B_0^{-1} \omega_t \\ [I - B_0^{-1} B(L)] x_t &= B_0^{-1} \omega_t \\ x_t &= [I - B_0^{-1} B(L)]^{-1} B_0^{-1} \omega_t \end{aligned} \quad (5.40)$$

Donde $B(L) = B_1 L + B_2 L^2 + \dots + B_p L^p$.

Si se ilustra con el ejemplo de dos variables y un rezago, la expresión (5.40) quedaría:

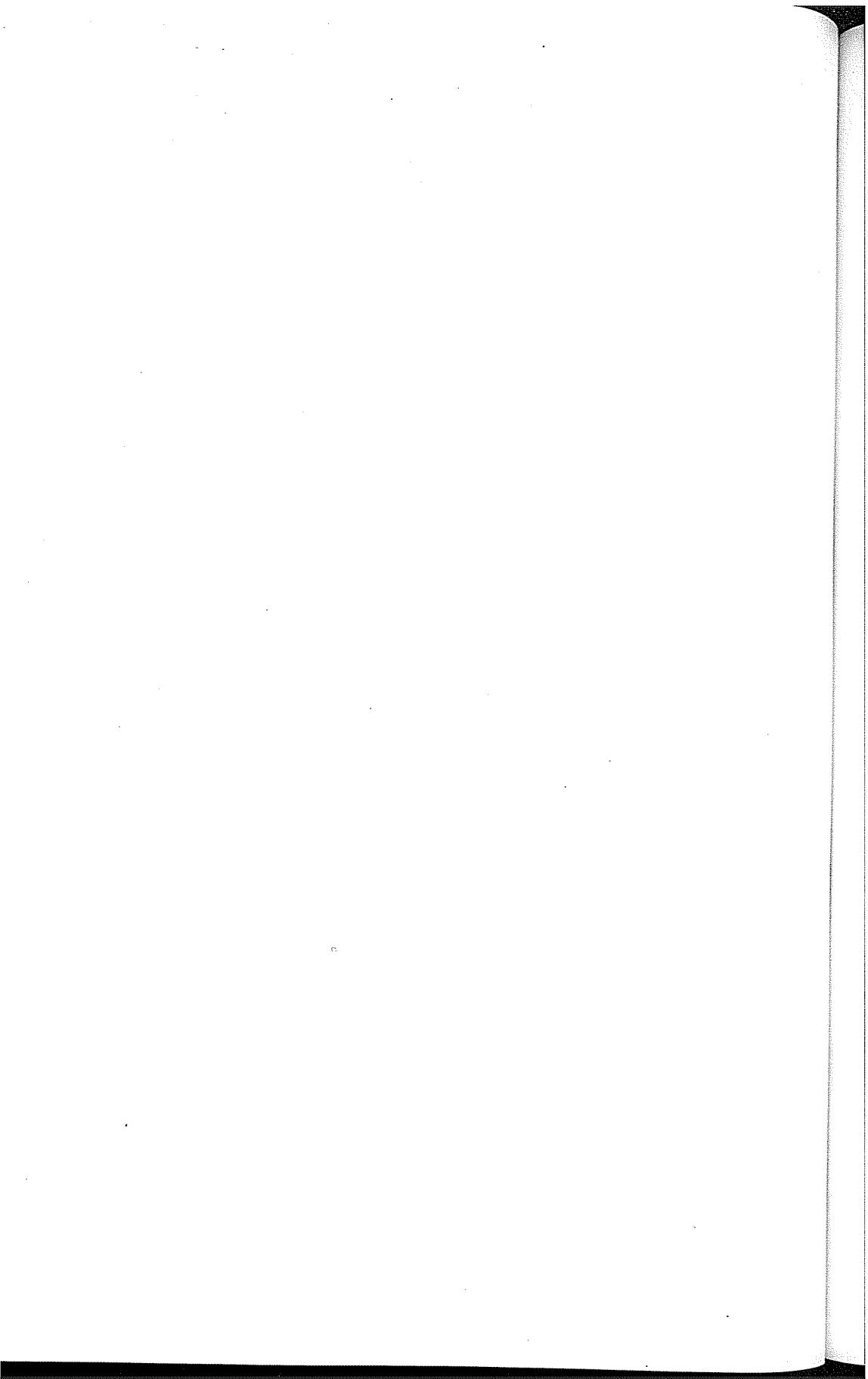
$$\begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} b_{0,11} & b_{0,12} \\ b_{0,21} & b_{0,22} \end{bmatrix}^{-1} \begin{bmatrix} b_{1,11}L & b_{1,12}L \\ b_{1,21}L & b_{1,22}L \end{bmatrix}^{-1} \begin{bmatrix} b_{0,11} & b_{0,12} \\ b_{0,21} & b_{0,22} \end{bmatrix}^{-1} \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \end{bmatrix}$$

De (5.40) vemos que el efecto acumulativo sobre x_t de un impulso propinado a un elemento de ω_t estará dado por el término correspondiente de la matriz:

$$[I - B_0^{-1}B(1)]^{-1} B_0^{-1} \quad (5.41)$$

Con $L = 1$, porque el efecto acumulativo es la suma de los coeficientes, y donde $B_0^{-1}B(1) = B_0^{-1}B_1 + B_0^{-1}B_2 + \dots + B_0^{-1}B_p = A_1 + A_2 + \dots + A_p$ es una matriz conocida a partir de la estimación de la forma reducida. En otras palabras, la restricción consiste en igualar a cero el elemento ij de la matriz $[I - B_0^{-1}B(1)]^{-1} B_0^{-1}$, lo que equivale a decir que la suma acumulada de las respuestas de la variable x_{it} a un impulso en el error ω_{jt} es cero. La decisión sobre cuáles errores estructurales no tendrán efecto de largo plazo y sobre cuáles variables, puede basarse en la teoría o ser subjetiva.

Con restricciones de este tipo completamos las ecuaciones necesarias para la identificación de B_0 . Así, podemos utilizar la forma reducida (5.36), remplazando ε_t de (5.37), o el VMA de (5.40) para generar respuestas a partir de impulsos a los elementos ortogonales de ω_t .



CAPÍTULO 6

Predicción económica

Los capítulos anteriores trataron principalmente el tema de la especificación y estimación de modelos para series de tiempo y centraron su interés en describir el comportamiento de la serie o la estructura de su mecanismo generador. Sin embargo, estos modelos en los cuales se expresa x_t en función de su propio pasado o del pasado de otras series también resultan apropiados para ejercicios de predicción; esto es, para intentar adivinar el valor que tendrá la variable dependiente en el futuro. Hemos tocado algunos temas puntuales, pero en este capítulo hacemos un tratamiento más general.

Antes de seguir adelante, es conveniente subrayar que no todas las variables económicas fueron creadas para ser predecibles. Aun así, lo que hemos logrado en materia de predicción económica es notable y funciona mejor en el corto plazo que en el largo.

Las mejores predicciones son para uno o dos períodos adelante, un mes un trimestre, etc., dependiendo de la unidad de tiempo utilizada, porque los errores se acumulan y la estructura del sistema cambia con rapidez relativa y hace que el modelo pierda vigencia. Obviamente, se pueden reportar excepciones en las cuales la predicción de largo plazo funciona perfectamente bien –por ejemplo, cuando se predice que en

el largo plazo todos estaremos muertos–; sin embargo, esas predicciones no son realmente de naturaleza estocástica.

Hay series, como el ruido blanco, que no se pueden predecir de su propio pasado, pero que podrían, en principio, y dependiendo de cómo fueron construidas, predecirse con la ayuda de otras variables. Otro caso particular son las llamadas series caóticas que parecen ruido (por correlograma, por pruebas Q , etc.); no obstante, son generadas de manera determinística no lineal. De este modo, descubierto el mecanismo generador, se podrían predecir sin error alguno. El problema es que la no linealidad de los mecanismos caóticos amplifica exponencialmente cualquier imprecisión en la medición del valor inicial, de manera que una diferencia infinitesimal en el valor inicial puede generar divergencias abismales en los valores futuros de la serie caótica, haciendo inmanejable o inútil la predicción.

Conjunto de información

Toda predicción está basada en el conjunto de información disponible en el tiempo t para realizar el pronóstico (o la estimación). El conjunto de información, denotado I_t , puede incluir el pasado y el presente de la serie en cuestión y el pasado y el presente de otras variables. Puede incluir, además, otras cosas como los supuestos del modelo, la teoría pertinente, y hasta rumores. Un I_t más amplio no hace daño, excepto porque la recolección de información y su procesamiento serán más costosos.

Por ejemplo, un conjunto de información que incluya el presente y el pasado de una serie hasta n períodos atrás, como el que se utiliza para construir un AR, sería:

$$I_t = \{x_{t-j} | j = 0, 1, 2, \dots, n\}$$

Funciones de pérdida

Usualmente no pronosticamos en el vacío, donde no haya consecuencias. En la vida real se toman decisiones con base en las predicciones hechas. Un error de predicción puede acarrear costos (pérdidas) para quien hace la predicción (en términos de credibilidad, prestigio, poder,

autoestima; o en términos laborales, monetarios, etc.) y también puede acarrear costos (pérdidas) para quien contrata o utiliza el pronóstico en la toma de decisiones.

Definimos el error de una predicción hecha en el tiempo t para T unidades de tiempo adelante como:

$$\begin{aligned}\hat{\varepsilon}_{t+T} &= \text{realidad} - \text{pronóstico} \\ \hat{\varepsilon}_{t+T} &= x_{t+T} - \hat{x}_{t+T}\end{aligned}\quad (6.1)$$

Donde x_{t+T} es la variable que se va a pronosticar y \hat{x}_{t+T} es el pronóstico hecho en el tiempo t para el tiempo T .

En términos operacionales, podríamos postular que el costo o la pérdida asociada a la predicción sea una función del error de predicción, llamada *función de pérdida (loss function)* y denotada $c(\hat{\varepsilon}_{t+T})$. La función de pérdida nos dice cuánto le puede doler al pronosticador cometer errores. En lenguaje de microeconomía, es como su función de utilidad, de la cual dependerá la forma como llegue a su predicción. En este sentido, la función de pérdida es un supuesto de comportamiento. El objetivo será encontrar una forma de pronosticar que minimice esta función particular.

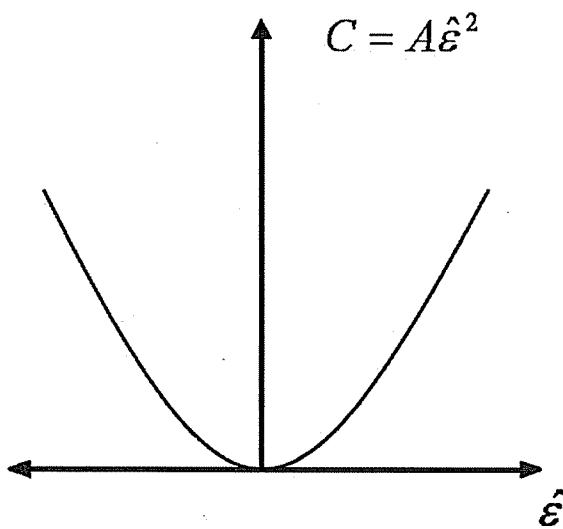
Como en la práctica no es fácil conocer la función de pérdida exacta, usamos un modelo o una aproximación y la función más comúnmente utilizada es de la forma cuadrática, escogida porque es simétrica, razonable y fácil de manejar:

$$c(\hat{\varepsilon}) = A\hat{\varepsilon}^2 \quad (6.2)$$

Realmente, la constante $A > 0$ no importa en el ejercicio de minimización y puede fijarse en la unidad. La función $A\hat{\varepsilon}^2$ hace que el costo de incurrir en errores extremadamente grandes (*outliers*) sea particularmente grave. Al contrario, una alternativa en valor absoluto, $A|\hat{\varepsilon}|$, asigna igual costo a todos los errores; es proporcional a cualquier tamaño de error. Otra función de costos es la llamada *todo o nada (hit or miss)*, que tolera errores pequeños pero que impone un costo grande igual para todos los demás errores cuyo valor absoluto excede el mínimo determinado (Schwartz y Shaw, 1975, p. 292). También puede suceder que la predicción no sea por una sola vez (*one-shot*), sino repetida, en cuyo caso la función de pérdida podría depender de los errores anteriores y habría que introducir algún tipo de dependencia en la toma

de decisiones. La Figura 6.1 muestra la función de pérdida cuadrática, $c(\hat{\varepsilon}) = A\hat{\varepsilon}^2$, donde se nota que los errores grandes tienen un costo más que proporcional. La Figura 6.2 muestra la función de pérdida de la forma $c(\hat{\varepsilon}) = A|\hat{\varepsilon}|$, donde todos los errores tienen proporcionalmente el mismo costo.

Figura 6.1: Función de pérdida cuadrática

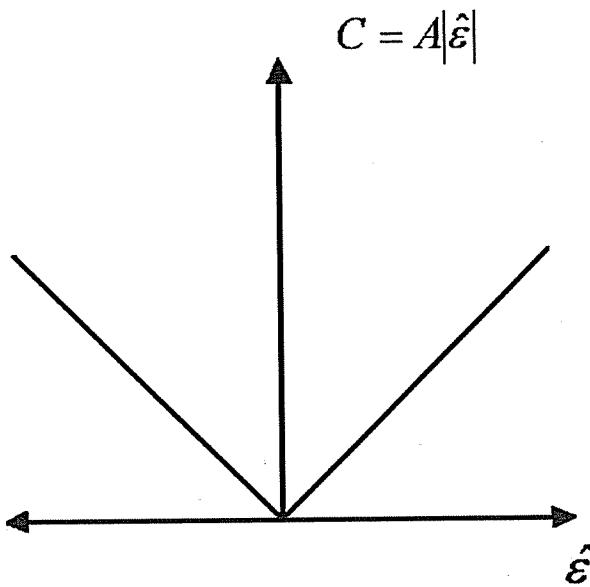


Ejemplo

La Figura 6.3 ilustra una posible función de pérdida para el caso de un empresario que posee una fábrica de buñuelos y produce diariamente lo que se venderá mañana. Debe predecir las ventas (no la demanda, sino las ventas) de mañana. Si la venta es mayor que la predicción, $\hat{\varepsilon} > 0$, pierde ganancias por los buñuelos que hubiera podido vender. Si la venta es menor que la predicción, $\hat{\varepsilon} < 0$, entonces pierde los buñuelos que hizo y no vendió. Supongamos que el costo es de \$30/buñuelo y el precio de venta es de \$50/buñuelo. La pérdida de sobreproducción es de \$30/buñuelo; mientras que la de subproducción es de \$20/buñuelo (\$50-\$30). Si la distribución de la variable ventas es simétrica alrededor

de su media, el productor de buñuelos probablemente preferirá sesgar hacia abajo su predicción y subproducir.

Figura 6.2: Función de pérdida proporcional



Ejemplo

El encargado de poner combustible en la última entrada a pits de un auto de carreras debe predecir la variable aleatoria *número de litros de combustible por kilómetro*. Si pone demasiado combustible perderá valioso tiempo en los pits y en la pista; pero si pone muy poco, el auto se puede quedar sin combustible antes de llegar a la meta. La función de pérdida para el encargado de poner el combustible puede ser como en la Figura 6.4. Nótese que se tolera un pequeño error positivo, porque no sería grave si el combustible se agota unas milésimas de segundo antes de la meta.

Figura 6.3: Ejemplo de función de pérdida

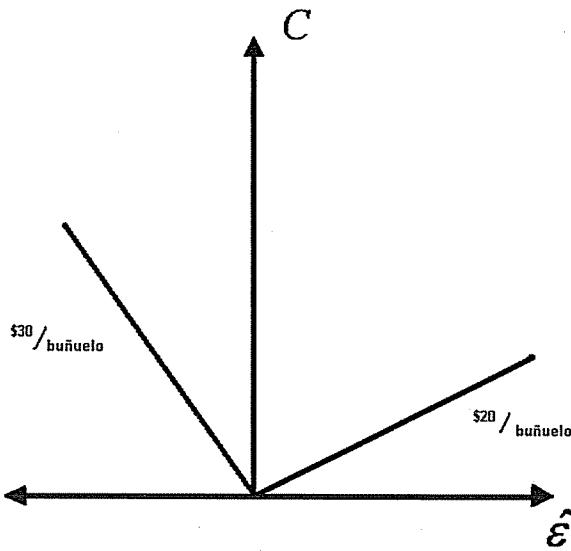
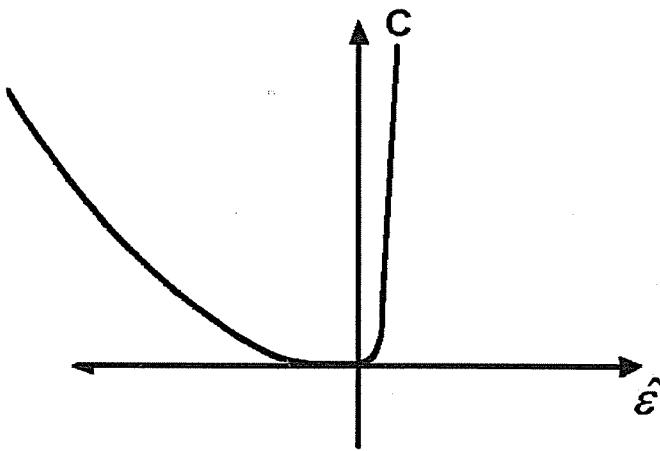


Figura 6.4: Ejemplo de función de pérdida



Predicción óptima

Supongamos que tenemos a disposición un conjunto de información $I_t = \{x_t, x_{t-1}, \dots, x_{t-n}\}$ y queremos pronosticar la variable x_t para T unidades de tiempo en el futuro. Denotamos este pronóstico \hat{x}_{t+T} . Como todo lo que sabemos es I_t , entonces no nos queda otra salida que hacer que el pronóstico \hat{x}_{t+T} sea una función del conjunto de información I_t . El objetivo general es encontrar una forma de hacer la predicción de \hat{x}_{t+T} condicional a I_t tal que se minimice la expectativa del costo o pérdida, cuya expresión es:

$$\int_{-\infty}^{\infty} c(x_{t+T} - \hat{x}_{t+T}) f(x_{t+T}/I_t) dx_{t+T} \quad (6.3)$$

Donde c es la función de pérdida. Si utilizamos la popular función de pérdida cuadrática (6.2), encontraremos que la predicción óptima es $\hat{x}_{t+T} = E(x_{t+T}/I_t)$ y que, suponiendo *gaussianidad*, \hat{x}_{t+T} será una función lineal en I_t .

Derivemos este resultado minimizando $E(A\hat{\varepsilon}^2)$ o, sin perder generalidad, minimizando $E(\hat{\varepsilon}^2)$, condicional a I_t . No olvidemos que \hat{x}_{t+T}/I_t ya no es una variable estocástica sino un número, porque los datos de I_t están dados y \hat{x}_{t+T} es una función determinística de I_t . Para hallar la forma de la función de predicción, dado I_t , escribimos:

$$\begin{aligned} \min E(\hat{\varepsilon}_{t+T}/I_t) &= \min E((x_{t+T} - \hat{x}_{t+T})^2/I_t) \\ &= \min E(x_{t+T}^2/I_t) - 2\hat{x}_{t+T}E(x_{t+T}/I_t) + \hat{x}_{t+T}^2 \end{aligned} \quad (6.4)$$

Al derivar con respecto a \hat{x}_{t+T} , obtenemos:

$$-2E(x_{t+T}/I_t) + 2\hat{x}_{t+T} = 0 \quad (6.5)$$

De manera que la expresión del error cuadrático medio se minimiza cuando:

$$\hat{x}_{t+T} = E(x_{t+T}/I_t) = \int_{-\infty}^{\infty} x_{t+T} f(x_{t+T}/I_t) dx_{t+T} \quad (6.6)$$

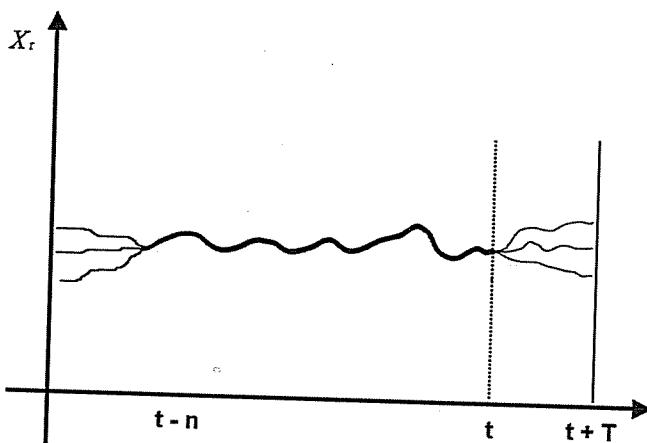
En otras palabras, bajo el criterio del mínimo error cuadrado medio, nuestra mejor predicción (o el mejor predictor) es igual al valor esperado

de x_{t+T} dado I_t . Un importante resultado dado en Granger y Newbold (1986, p. 125) muestra que este mejor predictor derivado bajo el criterio del mínimo error cuadrado medio aplica no sólo cuando se trata de funciones de pérdida cuadráticas como $c(\varepsilon) = A\varepsilon_{t+T}^2$, sino (con algunas excepciones) cuando la función de pérdida sea simétrica alrededor de $\hat{\varepsilon}_{t+T} = 0$ y la función de probabilidad de x_{t+T} dado I_t sea simétrica alrededor de su media $E(x_{t+T}/I_t)$.

Intuición de la predicción óptima

La siguiente explicación sigue a Priestley (1981, p. 728) y se basa en la Figura 6.5, en la cual se muestra el ensamblaje de un proceso estocástico. Nuestra tarea es observar una realización hasta el tiempo t y luego predecirla para $t + T$.

Figura 6.5: Predicción óptima condicional



Consideremos todas las posibles realizaciones de x_t , en teoría infinitas en número, y pensemos que algunas coinciden en el intervalo $[t - n, t]$, donde n es el número de observaciones de nuestra realización. Removamos todas las realizaciones que no coincidan con las observaciones en ese intervalo, ya que ninguna de éstas podría ser la realización sobre la cual estamos tratando de hacer la predicción. Cada una de las que queda produce un valor diferente de x_{t+T} y, dado que no sabemos

cuál de todos estos valores es el correcto (el que correspondería a nuestra realización), lo mejor que podemos hacer es tomar el promedio de todos estos x_{t+T} generados por esta subcolección de realizaciones que coinciden en el intervalo $[t - n, t]$. Este promedio será nuestro mejor predictor de x_{t+T} . Obsérvese que el valor promedio de x_{t+T} sobre esta subcolección es precisamente el valor esperado condicional de x_{t+T} dado $x_t, x_{t-1}, \dots, x_{t-n}$, o la expresión (6.6). Si no hay diferencia entre esta subcolección y el ensamblaje, entonces la mejor predicción será igual a la predicción incondicional.

Adicionalmente, si la distribución conjunta de $x_{t+T}, x_t, x_{t-1}, \dots, x_{t-n}$ es gaussiana, entonces se puede demostrar (después de algo de álgebra) que la distribución marginal $x_{t+T}/x_t, x_{t-1}, \dots, x_{t-n}$ también será gaussiana y que la mejor predicción será una función lineal de los datos (por ejemplo, véase Hamilton, 1994, p. 102; Priestley, 1981, p. 85):

$$E(x_{t+T}/x_t, x_{t-1}, \dots, x_{t-n}) = \hat{x}_{t+T} = a_0 x_t + a_1 x_{t-1} + \dots + a_n x_{t-n} \quad (6.7)$$

Lo cual resulta conveniente y fácil de manejar.

Errores de predicción

Los errores de predicción definidos en (6.1) pueden surgir de varias fuentes: (1) de un modelo equivocado o mal especificado (lo cual incluye cambios estructurales); (2) de errores de estimación de los parámetros (varianza y método); (3) de errores de medición de las variables en I_t sobre las cuales se basa la predicción, y, obviamente, (4) de suponer que los términos de error del modelo en el futuro, hasta ε_{t+T} , son cero porque no se conoce su verdadero valor.

Hay que subrayar la importancia que para la predicción tiene el minimizar los errores tipo (1) y (2) listados arriba, los de especificación y estimación, en lo cual la teoría y la práctica serán de gran ayuda. Parafraseando a Kennedy (1992, p. 270), es claro que los dos principales objetivos de los economistas, obtener buenos parámetros estimados y hacer buenas predicciones, están íntimamente ligados. También es conveniente minimizar los errores de medición, aunque no siempre el analista puede influir en este aspecto. Puede suceder, como se afirma con frecuencia, que los agentes económicos reaccionan a los datos

publicados, tengan o no tengan error de medición, de manera que son éstos los que hay que utilizar.

Es sano tener una mente abierta sobre el tema de la predicción en economía (o en cualquier disciplina), ya que es un tema que genera una gran variedad de controversia y apuntes curiosos. Por ejemplo:

- Los economistas hacen predicciones para mostrarle al mundo que tienen buen sentido del humor.
- Cuando tenga alguna duda, prediga que la tendencia actual continuará.
- Los economistas han pronosticado nueve de las últimas cinco recesiones.
- Las predicciones de largo plazo son la peor forma de llegar al futuro. Un economista es un experto que sabrá mañana por qué las cosas que predijo ayer no sucedieron hoy. (Atribuido a Lawrence Peter).
- Si un economista es alguien que adivina mal sobre la economía, un econométrista es alguien que usa un computador para adivinar mal sobre la economía.
- Una adivina comenta a otra adivina, “El futuro ya no es lo que solía ser”.
- Si torturamos los datos lo suficiente, ellos confiesan.
- Las predicciones son bastante difíciles, particularmente aquellas acerca del futuro. (Neils Bohr)
- Todos los modelos están mal, pero algunos sirven. (George Box)

Predicción e intuición

El último apunte, que todos los modelos están mal, simplemente reconoce que no es posible modelar perfectamente la realidad. La posibilidad de errores de predicción –introducidos por equivocaciones, imprecisiones o limitaciones en la especificación– es la razón por la cual

muchos pronosticadores (*forecasters*) complementan las predicciones que salen de los modelos con su propio juicio, especulación, intuición o instinto económico, antes de publicar las cifras.

Hay algo bayesiano en todo esto. No olvidemos que un analista o pronosticador experimentado, de todas maneras, tiene en su cabeza algún tipo de modelo, seguramente no lineal, que ha ido armando, probando y revisando a lo largo de su ejercicio profesional, quizás basado en un conjunto de información no expresable en lenguaje matemático o cuyas series son de diversas periodicidades.

En cualquier caso, involucrar el juicio y la intuición es inevitable ya que éstos intervienen, así sea de manera inconsciente, en el ejercicio de la especificación, pruebas, selección final del modelo, y hasta en la forma como se presentan los resultados.

Error de especificación quiere decir simplemente que el modelo no coincide con la realidad, ya sea porque el investigador escoge mal la forma funcional o las variables, o porque, habiendo escogido bien, la realidad cambia de manera inmodelable; por ejemplo, cuando el cambio es muy reciente y aún no se tienen datos o experiencia suficiente para modelar adecuadamente. En estos casos de mala o incierta especificación es cuando seguramente hay que apelar al instinto económico.

Otra justificación para apelar al instinto y al juicio es la no descartable posibilidad de que factores extraestadísticos temporales induzcan un sesgo predecible en el procesamiento y cálculo de las cifras de algunas variables, especialmente de aquellas sensibles a la presión política del momento o a intereses creados. Es de notar que si el sesgo predecible es permanente, éste debe considerarse dentro de la especificación del modelo y no a través del juicio.

En la praxis usual de predicción se combinan el modelo econométrico con la especulación y el juicio del experto. Estudios realizados a través de encuestas y entrevistas, han establecido que entre los pronosticadores profesionales lo común es dar un peso de 60 %-70 % al modelo econométrico, 20 %-30 % a la especulación o juicio del pronosticador y un 10 % al análisis de los datos o sucesos más recientes, incluida la interacción de última hora con otros expertos. Por ejemplo, si la ecuación contiene datos trimestrales, pero algunas de las variables están disponibles mensualmente, entonces se pueden utilizar, o al menos mirar, esos datos.

Mahmoud (1984) estudia el desempeño de una amplia gama de métodos de predicción y encuentra que aquellos cuyo mayor componente es cuantitativo le ganan a los métodos cuyo mayor componente es cualitativo o subjetivo. Encuentra también que los métodos sencillos funcionan tan bien como los complicados, y confirma algo que se sabía desde antes: combinar las predicciones de diversos métodos o pronosticadores resulta mejor o igual que usar una sola predicción.

Combinación de predicciones

Combinar predicciones tiende a mejorar o superar la predicción de métodos individuales. Podemos trabajar con combinaciones de resultados de métodos o combinaciones de predicciones de expertos, aunque éstas no necesariamente se basen en modelos explícitos. Es razonable esperar que las predicciones originadas en métodos diferentes, basadas en conjuntos de información diferentes y con la contribución del instinto económico o intuición de diferentes pronosticadores, pueden mejorar la predicción cuando se combinan.

Una forma sencilla de hacer esta combinación es tomando el promedio aritmético. Por ejemplo, para N predicciones de la misma variable:

$$\text{predicción combinada} = \frac{1}{N} (\text{predicción}_1 + \text{predicción}_2 + \dots + \text{predicción}_N) \quad (6.8)$$

También se pueden asignar pesos o ponderaciones diferentes a cada predicción aunque, obviamente, en esto entrará algún elemento de subjetividad. Como alternativa, se puede hacer la combinación aplicando mínimos cuadrados ordinarios (MCO) (véase Granger, 1989, p. 191). Para lo anterior se requieren series suficientemente largas que permitan correr la regresión de las observaciones contra las predicciones:

$$\begin{aligned} \text{observación} = & \beta_0 + \beta_1 \text{predicción}_1 + \beta_2 \text{predicción}_2 + \dots \\ & + \beta_N \text{predicción}_N + \varepsilon_t \end{aligned} \quad (6.9)$$

Una vez se tengan los coeficientes estimados, se utilizan para combinar las predicciones con el objeto de predecir las próximas observaciones.

Un método relacionado con este enfoque donde se combinan predicciones es el llamado *método de Delfos*, en el cual se trata de extraer de un grupo de expertos o personas inteligentes e informadas, a través de un mecanismo iterativo de consenso, la mejor predicción posible sobre algún asunto. El método fue diseñado para pronosticar avances tecnológicos pero su aplicación es general. Se pregunta, por ejemplo, ¿cuándo habrá drogas para aumentar la inteligencia?, o ¿cuándo será posible producir animales inteligentes, como simios para reemplazar la mano de obra menos calificada? Igual se puede preguntar por cualquier variable económica, por ejemplo, ¿cuánta será la inflación, cuánto el crecimiento o cuánto el desempleo este año?

El método de Delfos sigue un proceso iterativo. La primera ronda de respuestas de los expertos es tabulada y su promedio y varianza (o alguna medida de tendencia central y dispersión) se comunican a los expertos (a manera de realimentación) para que la tomen en cuenta al revisar su predicción en la segunda ronda. El proceso se repite hasta que la combinación de predicciones converja o hasta que se alcance algún criterio aceptable de finalización.

En algunos ejercicios de predicción el problema puede ser que se tiene demasiada información. Por ejemplo, cuando el investigador dispone de 100 o más variables, con sus rezagos, para incluir en el modelo pero el tamaño de la muestra no lo permite. Una solución es disminuir el número de variables y sustituirlas por unos pocos componentes principales. Otra solución es estimar varios modelos, con diferentes subconjuntos de las variables disponibles, y luego promediar las predicciones de estos modelos para obtener la predicción final. Para una discusión de este tema, véase Eklund y Karlsson (2007).

Consenso vs. incertidumbre

Cuando examinamos varias predicciones sobre el mismo evento debemos tener en cuenta que una cosa es el *consenso* y otra es la *incertidumbre*. El hecho de que haya un gran consenso entre pronosticadores –por ejemplo, que todos coincidan en que la inflación del año entrante será de X% o que el crecimiento del producto interno bruto (PIB) será Y%– no quiere decir que la predicción tenga baja incertidumbre. El consenso es el grado de coincidencia de las predicciones puntuales

individuales, mientras que la incertidumbre está dada por la varianza de las predicciones individuales. Puede haber gran consenso con enorme incertidumbre o con baja incertidumbre. O puede haber poco consenso con enorme incertidumbre o con baja incertidumbre (para una ampliación de este tema, véase Zarnowitz y Lambros, 1987).

La maldición del ganador

Un caso que ilustra la importancia de la distribución de probabilidad de las predicciones y la relación entre su media y varianza es conocido en la literatura sobre el tema como la *maldición del ganador* (*the winner's curse*). Consiste en que, por lo general, el ganador de una licitación (o subasta) pierde dinero o le va peor de lo que esperaba, porque ofrece más de lo que vale el premio que se lleva. La maldición se da aun si cada participante es inteligente y se asesora de expertos para formular su oferta.

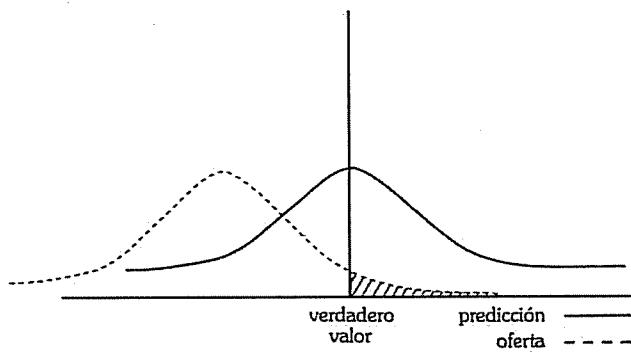
Imaginemos una licitación o subasta para adjudicar un bien o contrato que tendría el mismo valor para todos los participantes; pero donde ese valor no es conocido por ellos con exactitud al momento de licitar –por ejemplo, el valor de un pozo de petróleo, de un futbolista, del derecho de publicar un posible *best seller* o de un frasco lleno de monedas de oro–. En esta subasta, las ofertas se presentan en sobre cerrado y gana quien ofrezca el precio más alto.

Los pasos que siguen los participantes son más o menos así: cada uno hace en privado su mejor predicción del valor del bien subastado y después, naturalmente, presenta como oferta un precio inferior a su predicción. Como todos los participantes son buenos en su oficio, podemos suponer que el valor esperado del agregado de todas las predicciones sea igual al verdadero valor del bien. Sin embargo, por error estadístico habrá diferencias entre las predicciones individuales, algunas de las cuales estarán por encima y otras por debajo del verdadero valor del bien.

En otras palabras, la distribución (o el histograma) de las predicciones estará centrada en el valor del bien como se muestra por la curva sólida en la Figura 6.6. Como cada participante le resta a su predicción la cantidad que piensa ganar en el negocio, la curva de oferta (curva punteada) estará centrada a la izquierda del verdadero valor, esto

es, tendrá un promedio menor. Sin embargo, aunque todos ofrezcan un precio inferior a su predicción, muy seguramente quedará algún participante (uno de los que, por error estadístico, tenía una predicción muy elevada) cuyo precio de oferta todavía supera el valor del bien; este participante gana la subasta pero pierde dinero. La probabilidad de que esto ocurra corresponde al área sombreada en la Figura 6.6.

Figura 6.6: Distribución de probabilidad de predicciones y ofertas en la maldición del ganador



Fuente: tomada de Montenegro (1995).

La maldición se hace más notoria a medida que aumenta el número de participantes porque también aumenta la probabilidad de que alguien tenga una predicción desatinada (en una muestra grande cualquier cosa puede suceder). Por ello, obtener información confidencial (*inside information*) ayuda a acertar en la predicción pero no ayuda a ganar la subasta. Para una ampliación de este tema desde el punto de vista de la economía experimental, véase Montenegro (1995).

Criterios para evaluar predicciones

Evaluar predicciones es una de las tareas más difíciles. Depende, entre otras cosas, de las características de la variable que se va a predecir, de la anticipación en el tiempo con que se haga la predicción, del conjunto de información disponible, del número de veces que se

repite el experimento y de otras circunstancias impredecibles que rodean la predicción. En el fondo, de lo que se trata es de evaluar y comparar métodos de predicción. Las más fáciles de evaluar son predicciones del tipo binario sobre la ocurrencia de algún suceso; si ocurre, acierta; si no ocurre, falla.

Este tipo de predicción binaria se evalúa registrando aciertos sobre el total de predicciones e indicando el porcentaje de aciertos. Por ejemplo, ¿la inflación, el desempleo, el crecimiento, serán más altos este año que el pasado? Las preguntas anteriores, sobre si sube o baja una variable, son extremadamente útiles para predecir el curso o tendencia de la actividad. Sin embargo, por alguna razón, la gente prefiere predicciones numéricas para la mayoría de las variables económicas; ¿cuánto será la inflación, el desempleo, el crecimiento?

La mayoría de los métodos de evaluación requiere una serie de predicciones y una serie de observaciones para poder comparar las dos. Infortunadamente, casi nunca hay una serie de observaciones suficientemente larga diferente de la muestra utilizada para estimar el modelo. Por esta razón, una vez se tiene estimado el modelo, es común usarlo para ver qué tan bien hubiera predicho las observaciones de la muestra. Sin embargo, como al efectuar la estimación el modelo se ajusta lo mejor posible a la muestra, entonces probar su capacidad predictiva en la misma muestra tiende a producir resultados demasiado optimistas. Por desgracia, esta advertencia es pasada por alto con frecuencia. El método de *validación cruzada (cross validation)* consiste en utilizar una parte de la muestra para estimar el modelo y otra parte para evaluarlo; por ejemplo, usar el 75 % de las observaciones para estimar y el 25 % restante para probar y evaluar.

Una condición que siempre debe tenerse en cuenta es buscar que la serie de los errores de predicción $\hat{\epsilon}_{t+T}$ sea ruido blanco; de lo contrario, la predicción es subóptima, porque si los errores de predicción están autocorrelacionados, quiere decir que se podrían predecir de su propio pasado, y que debimos haber aprovechado esa información para mejorar la predicción, en lugar de suponer ciegamente que $E\hat{\epsilon}_{t+T} = 0$. Predicción óptima es lo mismo que expectativa racional, donde se usa eficientemente toda la información disponible y no se cometen errores sistemáticos.

A continuación se describen varios indicadores comúnmente utilizados para evaluar predicciones. Aunque no hay consenso sobre cuál es mejor, los más usados son la raíz del error cuadrado medio, el error medio absoluto y el error medio absoluto porcentual. Enders (2004, capítulo 2) presenta algunas pruebas estadísticas asociadas a los indicadores de predicción. En la práctica, será obvio que el econométrista utilice su recién estimado modelo para hacer predicciones más allá del último dato de la muestra, pero deberá recordar que la evaluación que reporta el computador se hace dentro de la muestra, ya que estos criterios o indicadores no se pueden calcular para predicciones por fuera de la muestra sencillamente porque no se tienen observaciones para x_{t+T} . De nuevo, hacemos hincapié en que los indicadores calculados sobre la misma muestra utilizada en la estimación del modelo, incluido el R^2 , el error estándar del residuo, la verosimilitud, los criterios de información de Akaike (AIC) y Schwarz (SC), suelen ser mejores que los calculados por fuera de ésta.

Error cuadrado medio

El error cuadrado medio (*Mean Square Error* [MSE]) está dado en unidades al cuadrado de la variable que se va a predecir:

$$\text{MSE} = E(x_{t+T} - \hat{x}_{t+T})^2 \quad (6.10)$$

Raíz del error cuadrado medio

La raíz del error cuadrado medio (*Root Mean Square Error* [RMSE]) es simplemente la raíz cuadrada del MSE, dado en unidades de la variable que se va a predecir.

Error medio absoluto

El error medio absoluto (*Mean Absolute Error* [MAE]) está dado en unidades de la variable que se va a predecir:

$$\text{MAE} = E|x_{t+T} - \hat{x}_{t+T}| \quad (6.11)$$

Es importante tener en cuenta que una disminución (aumento) a través del tiempo en cualquiera de los tres índices anteriores no

necesariamente implica que la predicción haya mejorado (empeorado); podría indicar una disminución (aumento) en la magnitud de la variable x_t a través del tiempo.

Error medio porcentual absoluto

El error medio porcentual absoluto (*Mean Absolute Percentage Error [MAPE]*) es adimensional, dado en porcentaje. Este indicador no tiene mucho sentido cuando la variable muestra valores cercanos a cero:

$$\text{MAPE} = E \left| \frac{x_{t+T} - \hat{x}_{t+T}}{x_{t+T}} \right| \quad (6.12)$$

Regresión

La regresión se trata simplemente de correr por MCO la serie de observaciones, como variable dependiente, contra la serie de predicciones, como variable explicativa:

$$x_{t+T} = \alpha + \beta \hat{x}_{t+T} + u_{t+T} \quad (6.13)$$

Si la predicción es buena, entonces debemos encontrar que $\alpha = 0$ y $\beta = 1$. Un problema con este tipo de regresión es que, comúnmente, x_t es una variable autocorrelacionada y si algo falta en el predictor \hat{x}_t entonces el error u_t también estará autocorrelacionado e invalidará las pruebas de hipótesis usuales.

Dado que en muchos casos hacer la predicción de una variable en niveles tiene poco mérito, algunos autores sugieren correr la regresión en diferencias o en cambios porcentuales:

$$\Delta x_{t+T} = \alpha + \beta \Delta \hat{x}_{t+T} + u_{t+T} \quad (6.14)$$

Lo cual establece un criterio más exigente que las regresiones en niveles.

Índice de Theil

El índice de Theil es como un MSE, pero normalizado:

$$\text{Theil} = \frac{\sqrt{E(x_{t+T} - \hat{x}_{t+T})^2}}{\sqrt{Ex_{t+T}^2 + E\hat{x}_{t+T}^2}} \quad (6.15)$$

Corregir

Donde x_t y \hat{x}_t tienen media cero o se les ha restado la media. Este índice es adimensional y va de cero a uno; cero es una predicción perfecta, y uno, pésima. Nótese que el término bajo el radical en el numerador es igual a:

$$E(x_{t+T}^2 - 2x_{t+T}\hat{x}_{t+T} + \hat{x}_{t+T}^2) \quad (6.16)$$

El cual será igual a cero cuando $x_{t+T} = \hat{x}_{t+T}$, e igual al término bajo el radical en el denominador cuando $Ex_{t+T}\hat{x}_{t+T} = 0$, esto es, cuando la predicción es pésima.

Nótese que $Ex_{t+T}\hat{x}_{t+T} = \text{Cov}(x_{t+T}, \hat{x}_{t+T})$ porque x_t y \hat{x}_t tiene media cero o se les ha restado. La expresión del Theil se modifica si x_t o \hat{x}_t no tienen media cero o si ésta no se ha restado previamente. Si las medias de x_t y \hat{x}_t son μ_x y $\mu_{\hat{x}}$, respectivamente, el término bajo el radical del numerador de la ecuación (6.15) se puede escribir como:

$$E(x_{t+T} - \hat{x}_{t+T})^2 = Ex_{t+T}^2 - 2(\text{Cov}(x_{t+T}, \hat{x}_{t+T}) + \mu_x\mu_{\hat{x}}) + E\hat{x}_{t+T}^2 \quad (6.17)$$

Si la predicción es perfecta entonces $E(x_{t+T} - \hat{x}_{t+T})^2$ seguirá siendo cero. Pero si la predicción \hat{x}_{t+T} es pésima entonces no tiene covarianza con la serie x_{t+T} y la expresión $E(x_{t+T} - \hat{x}_{t+T})^2$ será:

$$E(x_{t+T} - \hat{x}_{t+T})^2 = Ex_{t+T}^2 + E\hat{x}_{t+T}^2 - 2\mu_x\mu_{\hat{x}} \quad (6.18)$$

De manera que el término $2\mu_x\mu_{\hat{x}}$ debe restarse de la expresión bajo el radical en el denominador de (6.15) para mantener el Theil en el rango de cero a uno. Algunos autores prefieren esta formulación.

Descomposición del error cuadrado medio

El MSE en (6.10) ocasionalmente se presenta separado por componentes de sesgo, varianza y covarianza como proporción del MSE total.

Para mostrar esto escribimos:

$$\text{MSE} = E(x_{t+T} - \hat{x}_{t+T})^2 = Ex_{t+T}^2 - 2Ex_{t+T}\hat{x}_{t+T} + E\hat{x}_{t+T}^2 \quad (6.19)$$

Si recordamos de la estadística básica que:

$$Ex^2 = E(x - \mu_x)^2 + \mu_x^2 = \sigma_x^2 + \mu_x^2$$

y que

$$Ex\hat{x} = E(x - \mu_x)(\hat{x} - \mu_{\hat{x}}) + \mu_x\mu_{\hat{x}} = \rho\sigma_x\sigma_{\hat{x}} + \mu_x\mu_{\hat{x}}$$

Donde ρ es el coeficiente de correlación entre x y \hat{x} , podemos reescribir (6.19) de la siguiente manera:

$$\begin{aligned} \text{MSE} &= \sigma_x^2 + \mu_x^2 - 2(\rho\sigma_x\sigma_{\hat{x}} + \mu_x\mu_{\hat{x}}) + \sigma_{\hat{x}}^2 + \mu_{\hat{x}}^2 \\ &= (\mu_x - \mu_{\hat{x}})^2 + (\sigma_x - \sigma_{\hat{x}})^2 + 2\sigma_x\sigma_{\hat{x}}(1 - \rho) \end{aligned} \quad (6.20)$$

Donde, dividiendo los tres términos anteriores de (6.20) por la expresión del MSE = $E(x_{t+T} - \hat{x}_{t+T})^2$, siempre y cuando $\text{MSE} \neq 0$, obtenemos las siguientes proporciones:

$$\text{proporción de sesgo} = \frac{(\mu_x - \mu_{\hat{x}})^2}{E(x_{t+T} - \hat{x}_{t+T})^2} \quad (6.21)$$

$$\text{proporción de varianza} = \frac{(\sigma_x - \sigma_{\hat{x}})^2}{E(x_{t+T} - \hat{x}_{t+T})^2} \quad (6.22)$$

$$\text{proporción de covarianza} = \frac{2\sigma_x\sigma_{\hat{x}}(1 - \rho)}{E(x_{t+T} - \hat{x}_{t+T})^2} \quad (6.23)$$

La proporción de sesgo nos dice qué tan lejos está la media de la predicción de la media de las observaciones; la proporción de la varianza dice qué tan lejos está la varianza de la predicción de la varianza de la serie; y la proporción de la covarianza mide el residuo o los errores no explicados por las dos anteriores. Estas proporciones suman uno. Si el MSE = 0 estas proporciones no están definidas. En general, un buen pronóstico debe tener proporciones de sesgo y varianza pequeñas.

Algunas evaluaciones empíricas

La evaluación de la capacidad predictiva de los economistas se dificulta por varias razones. Muchas veces no se dispone de series de predicciones suficientemente largas para hacer una comparación confiable con la realidad; la mayoría son predicciones discontinuas y esporádicas, o que cambian de horizonte con frecuencia, de manera que no son comparables.

Otro problema es el cambio en la definición o metodología de medición de la variable predicha; o también el hecho de que múltiples entidades publiquen series de mediciones de la misma variable (observaciones) no coincidentes o contradictorias. A veces olvidamos que los modelos se estiman y calibran, obviamente, con los datos publicados, y se usan para predecir el próximo dato publicado, que no necesariamente coincide con la variable teórica.

Por ejemplo, si la entidad que recoge, procesa y publica las cifras mide mal o tiene equivocado el concepto teórico, las cifras publicadas (la tasa de desempleo, la inflación, etc.) podrían no coincidir con las teóricas. Sobre el tema de la capacidad predictiva de los economistas se pueden consultar, entre otros, a Wallis (1989), quien hace una buena revisión de la literatura; Jansen y Kishan (1996), quienes encuentran sesgos en las predicciones que hace la Reserva Federal, lo que implica que comete errores sistemáticos corregibles, y a Swidler y Ketcher (1990), que encuentran evidencia de insesgamiento en las predicciones económicas, pero aceptan que no hay consenso sobre esto.

La Universidad de Michigan (proyecto RSQE del Departamento de Economía) ha venido realizando una predicción del cambio en el PNB real estadounidense; esto lo hace cada noviembre para el año siguiente, de manera continua desde 1953 (McNees, 1988, p. 16). La evaluación arroja un error medio absoluto (MAE) de 1.3%, cuando el crecimiento promedio del PNB es de 3%. Nótese que el MAE está en porcentaje porque esa es la unidad de la variable que se predice. McNees concluye que las predicciones anuales del PNB han mejorado con el paso del tiempo (desde 1953), en especial cuando se compara con los métodos ingenuos o poco sofisticados. El proyecto RSQE es valioso por la longitud y continuidad de sus datos y puede ser consultado directamente en <http://rsqe.econ.lsa.umich.edu./record.html>.

Un estudio de Artis (1996) examina el desempeño de las predicciones hechas por el Fondo Monetario Internacional (FMI) y publicadas en el *World Economic Outlook* en el periodo 1971-1994. En cuanto al crecimiento económico en países industrializados, el MAE ha sido de 1% anual, lo cual debe compararse con un crecimiento promedio de 2.75%. En cuanto a la inflación de los países industrializados, el MAE es de 0.75% comparado con una inflación promedio observada de 5.75% anual. Por otro lado, Artis divide la muestra en dos submuestras y encuentra que los errores de predicción parecen de la misma magnitud en ambas submuestras, lo cual implica que la técnica no parece haber mejorado en el periodo estudiado. Para el autor lo más difícil de predecir son los puntos de quiebre (*turning points*), cuya predicción es el tema de los indicadores líderes que se verá más adelante.

Métodos de predicción

Para realizar predicciones existe una gran variedad de métodos que van desde los muy primitivos hasta los muy sofisticados. Los primitivos, también llamados ingenuos (o naive), son, por ejemplo, hacer $\hat{x}_{n+1} = c$, donde c es una constante; esto es, siempre predecir lo mismo sin mirar alrededor o sin usar la información disponible. También primitivo, aunque menos, es predecir $\hat{x}_{n+1} = x_n$, por ejemplo, “el próximo año será igual que el actual”. Los llamados filtros de suavizado, ya sean del tipo aritmético o exponencial como, “el próximo año será igual que el promedio aritmético de los últimos n años”, son también algo elementales.

Entre los métodos sofisticados están el autorregresivo (AR), el de promedio móvil (MA), el ARMA (o el ARIMA), que se han convertido en un estándar; también, los multivariados vectores autorregresivos (VAR) y los vectores de corrección de error (VEC); la regresión por MCO; los sistemas de ecuaciones estructurales simultáneas; el filtro de Kalman, y otros modelos no lineales como los STAR y las redes neurales artificiales.

Los AR funcionan bien en predicciones de corto plazo y los sistemas multiecuacionales un poco mejor a largo plazo. A un AR se le pueden adicionar variables exógenas pero es necesario tener en cuenta que éstas también deben predecirse, por lo cual usualmente se adicionan

rezagadas. El método conocido como filtro de Kalman es en esencia una especie de regresión que ajusta los parámetros (incluida la varianza del error) con cada nueva observación. El filtro funciona bien en teoría; no obstante, depende crucialmente de los supuestos que se hagan para su modelaje –para una explicación, véase Montenegro (1991) o el Apéndice B–.

Un método que no es claro si debe ser clasificado como primitivo o como sofisticado, consiste en hacer predicciones sin modelo estadístico alguno, con base sólo en la intuición, la especulación y el juicio del pronosticador; en últimas, con base en el modelo que tiene el pronosticador en la cabeza.

Lo que parece funcionar mejor en la práctica, y es el método adoptado por la mayoría de las instituciones profesionales que publican predicciones económicas, es la combinación de algún modelo econométrico con el juicio y la intuición de los economistas pronosticadores. El modelo produce su predicción y luego ésta es matizada con el juicio y la intuición, antes de ser publicada. Como se mencionó, lo usual en estas instituciones es dar al modelo un peso o ponderación del 70%-80% y el otro 30%-20% al juicio o información de última hora.

Infortunadamente, no existen aún estudios de evaluación del desempeño de equipos de economistas que predigan sin modelo explícito alguno. La evaluación se dificulta porque casi ningún grupo de economistas hace predicción sistemática y continua sin modelo. Sin embargo, ésta sería una información interesante como estándar de comparación con otros métodos.

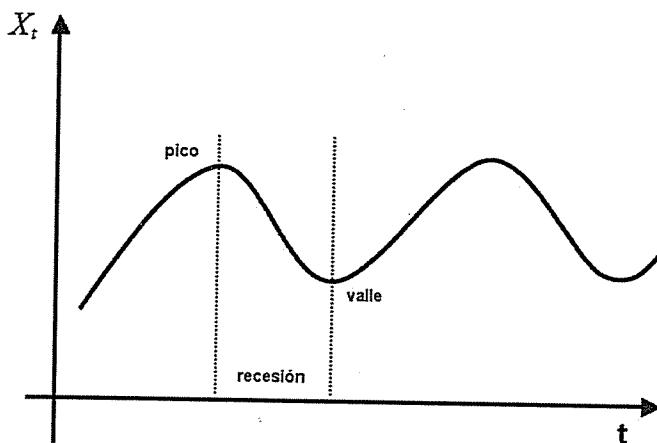
Un excelente recuento del desarrollo de la predicción económica se encuentra en Faulhaber y Baumol (1988), y de la perspectiva de la predicción en Diebold (1998).

Indicadores líderes

El método de indicadores líderes (*leading indicator*) se diseñó para la difícil e importante tarea de predecir puntos de quiebre en las series de tiempo económicas. Es de recalcar que aquí el relieve no está puesto en predecir el nivel de una serie, sino su giro o quiebre; esto es, sus valles y picos, como se muestran en la Figura 6.7. El objetivo es predecir o anticipar el momento cuándo empieza o termina una recesión o

expansión. Incluso, el interés puede restringirse aún más, a predecir sólo quiebres hacia arriba o sólo quiebres hacia abajo. La serie cuyos valles o picos nos interesa anticipar usualmente es el PIB, aunque puede ser cualquier otra.

Figura 6.7: Quiebres en el ciclo económico



La idea fundamental es encontrar otras series cuyos giros o quiebres ocurran antes de los giros y quiebres de la serie de interés de manera que los puedan anticipar, y que estas series, por sí solas o combinadas, sirvan como indicador líder.

Existen también indicadores coincidentes (*coincident indicator*) y rezagados (*lagging indicator*), los cuales se componen, respectivamente, de series cuyos quiebres coinciden con los de la serie en cuestión y de series cuyos quiebres se rezagan u ocurren después de los quiebres de la serie en cuestión. Puede parecer extraño que alguien se interese por indicadores rezagados, pero si nos situamos en un momento t , sin el beneficio de conocer la información futura, es difícil saber si la economía puede estar saliendo de una recesión o expansión. Los indicadores rezagados ayudan a confirmar la ocurrencia o ausencia del supuesto quiebre. En cualquier caso, la mayor atención se concentra en los indicadores líderes.

No existe una metodología preestablecida para encontrar las series líderes ni sobre cómo combinarlas, aunque hay sugerencias (véase

Granger, 1989, capítulo 7). Entre las series más comúnmente utilizadas para formar parte de un indicador líder están: cambios en el empleo, volumen de pedidos de bienes de capital, número de nuevas empresas, licencias de construcción residencial, cambio en inventarios, precio de las acciones, ganancias empresariales y crédito nuevo. Sin embargo, además de las anteriores, hay unas más aptas para anticipar reactivaciones, por ejemplo, la utilización de la capacidad instalada, aumento de cartera, tasas de interés, etc.; y otras más aptas para anticipar recesiones, por ejemplo, producción industrial, dificultades de cartera, etc.

Además de las anteriores, Watson (1991) sugiere variables como el índice de la tasa de cambio real, la diferencia o brecha entre la tasa de interés del ahorro en papeles sin riesgo y con riesgo, o entre corto y largo plazos. Otras sugerencias prácticas se pueden encontrar en Klein y Moore (1983).

Para el indicador coincidente se sugiere, entre otros, el índice de producción industrial y el de las ventas del comercio. Para el indicador rezagado se sugiere, entre otros, la duración promedio del desempleo y la relación de inventarios a ventas.

Usualmente, por facilidad de manejo o restricción de recursos, se escoge sólo un subconjunto de todas las series posibles para formar el indicador. Existen varios criterios, algunos obvios, para aceptar o descartar series. Por ejemplo, el comportamiento estadístico de la serie con respecto a los picos y valles de la serie que se va a predecir; si los anticipa, coincide o rezaga, dependiendo de si lo que se desea es un indicador líder, coincidente o rezagado, y con qué confiabilidad.

La serie que se va a incluir debe tener, en lo posible, significancia económica, ya que esto le dará mayor credibilidad al indicador. La regularidad, calidad, oportunidad y continuidad de publicación de los datos de la serie es otro factor fundamental; si los datos están sujetos a frecuentes revisiones, si se atrasan, interrumpen o cambian de metodología con frecuencia, será muy difícil utilizar la serie en el indicador. Darle un puntaje a cada criterio puede servir no sólo para escoger series, sino para ponderar las series dentro del índice.

Si buscamos series líderes con ojo de estadístico, seguramente vamos a investigar con métodos como la correlación cruzada o regresión, que estudian los rezagos y adelantos con respecto a todos los puntos del tiempo; pero para el propósito de anticipar giros o puntos de quiebre sólo

nos interesa el comportamiento anticipatorio estadístico directamente relacionado con éstos, usualmente justo antes del giro o quiebre, y no en el resto del tiempo. Por esta razón se aceptan los métodos ad hoc.

En general, las ponderaciones de cada serie para formar el índice son arbitrarias aunque se pueden modificar y ajustar según sugieran las evaluaciones del desempeño del indicador. Es necesario rotar 180° sobre su tendencia (poniéndole una ponderación negativa) cualquier serie que se mueva en dirección contraria a la serie que se va a predecir, pues de otra forma restaría de las demás. También es conveniente normalizar las series, a fin de que su variabilidad sea similar; en caso de no hacerlo, los coeficientes de ponderación deberán ocuparse de reflejar las diversas escalas o unidades de cada variable.

Algunos autores sugieren que se puede construir un indicador líder dividiendo el indicador coincidente por el indicador rezagado (Hoel *et al.*, 1983, p. 62). Se argumenta que los quiebres en este nuevo indicador líder anticipan (o acentúan) los quiebres en el indicador coincidente si, por ejemplo, al llegar el fin de la expansión (pico) el indicador coincidente comienza a desacelerarse, mientras que el rezagado continua subiendo.

Causalidad

Cuando se corre una regresión por MCO, es común suponer que las variables independientes "causan" la dependiente; incluso, éstas también se llaman variables explicativas porque "explican" la dependiente. Sin embargo, tal causalidad no está garantizada. El solo hecho de que una variable se encuentre despejada en una expresión no implica que sea causada por las otras. Por ejemplo, en álgebra es frecuente escribir:

$$y = f(x)$$

Sin que ello implique que x cause a y . Igualmente, en teoría económica, es común utilizar la función de producción:

$$Q = f(K, L)$$

Sin que K o L necesariamente causen el producto Q , aunque a veces implícitamente pensemos lo contrario. Más aún, en la teoría del acelerador, por ejemplo, la causalidad va de Q a K .

Como anotan Simon e Iwasaki (1988, p. 150): "La función de producción define un vínculo simétrico entre los fenómenos, ya que cualquiera de las variables puede ser escrita como la dependiente y las otras dos como las independientes. Deben suministrarse supuestos adicionales para convertir este vínculo en un ordenamiento causal" (traducción del autor).

La definición filosófica de causalidad (atribuida a Herbert Feigl) es: predicción de acuerdo con una ley o un conjunto de leyes. Por lo tanto, la mera correlación o incluso el buen desempeño predictivo no es suficiente para establecer causalidad si falta una explicación o justificación teórica.

La correlación observada entre dos variables puede ser, o bien porque una variable causa a la otra $x_1 \rightarrow x_2$, o porque hay una tercera x_3 que causa ambas x_1, x_2 .

Definición de causalidad de Granger

La causalidad de acuerdo con ley o leyes tiene atractivo intuitivo, pero es una definición poco útil en economía mientras no tengamos una teoría más desarrollada. Granger (1969) ofrece una definición mucho más operacional aunque menos filosófica, la cual seguimos. Comenzamos afirmando que el futuro no puede causar el pasado y estableciendo alguna notación. Sean x_t y z_t procesos estocásticos estacionarios:

1. \bar{x}, \bar{z} son el pasado de x y z , respectivamente.
2. $\bar{\bar{x}}, \bar{\bar{z}}$ son el pasado y el presente de x y z , respectivamente.
3. $E(x/\bar{x})$ es el predictor insesgado de mínimos cuadrados de x usando \bar{x} (el predictor que derivamos en (6.6)).
4. $\hat{\varepsilon}(x/\bar{x}) = x - E(x/\bar{x})$ es la serie de errores de predicción.
5. $\sigma^2(x/\bar{x})$ es la varianza de $\hat{\varepsilon}(x/\bar{x})$.

Con lo anterior se definen varios tipos de causalidad:

1. Se dice que z causa x si $\sigma^2(x/\bar{x}, \bar{z}) < \sigma^2(x/\bar{x})$.

2. Se dice que z causa x instantáneamente si $\sigma^2(x/\bar{x}, \bar{z}) < \sigma^2(x/\bar{x}, z)$.
3. Se dice que hay realimentación (*feedback*) o causalidad bidireccional si: $\sigma^2(x/\bar{x}, \bar{z}) < \sigma^2(x/\bar{x})$ y $\sigma^2(z/\bar{x}, \bar{z}) < \sigma^2(z/\bar{z})$.
4. Se dice que hay independencia si: $\sigma^2(x/\bar{x}, \bar{z}) = \sigma^2(x/\bar{x}, z) = \sigma^2(x/\bar{x})$ y $\sigma^2(z/\bar{x}, \bar{z}) = \sigma^2(z/\bar{x}, z) = \sigma^2(z/\bar{z})$.

¿Por qué se supone que sean series estacionarias? Porque $\sigma^2(x/\bar{x}, \bar{z}) < \sigma^2(x/\bar{x})$ sólo es generalizable para todo el proceso estocástico si la serie es estacionaria. De otra forma será válido únicamente en un momento t específico. Y aun para cada momento, se requiere la estacionariedad para poder calcular σ^2 .

La causalidad de Granger incorpora el concepto de *predicción* de la definición filosófica; pero sustituye *leyes* por la correlación estadística entre variables rezagadas. En otras palabras: se dice que la variable z causa x en el sentido de Granger (ayuda a predecir x) cuando, en presencia del pasado de x , el pasado de z ayuda a explicar x . Debe entenderse en todo esto que una variable que no cambia (que es una constante o que no varía en la muestra) no puede causar nada ni ser causada.

Charemza y Deadman (1992, p. 189) anotan que, filosóficamente, la causalidad instantánea no existe porque siempre hay una diferencia, así sea infinitesimal, entre acciones independientes; pero que en la práctica se habla de este tipo de causalidad simplemente porque no disponemos de observaciones en tiempo continuo o con mayor frecuencia de muestreo que permitan separar causa y resultado. En otras palabras, la causalidad instantánea o bidireccional puede ser espuria y deberse sólo a que los intervalos de muestreo no son suficientemente pequeños como para identificar la causalidad unidireccional.

En general, la agregación temporal puede generar causalidad espuria o esconder la verdadera causalidad. La agregación temporal puede ocurrir cuando los datos no se recogen con la frecuencia suficiente como para capturar el movimiento de las variables económicas. Si las decisiones de los agentes no coinciden con el intervalo de muestreo, pueden generarse relaciones de causalidad espurias.

Algunas aplicaciones económicas han estudiado la causalidad entre variables, como el gasto en publicidad y el cambio en las ventas, el

salario mínimo y la inflación, la inversión residencial y el PIB (Montenegro, 1989a). También se han propuesto otras pruebas de causalidad basadas en conceptos estadísticos similares, como la prueba de Sims (véase Granger y Newbold, 1986).

Implementación lineal de la prueba de causalidad

En la práctica, la prueba de causalidad de Granger se hace con las ecuaciones de un VAR(p). Por ejemplo, para dos variables tendremos las dos ecuaciones siguientes:

$$\begin{aligned} z_t &= a + \sum_{i=1}^{\infty} b_i z_{t-i} + \sum_{j=1}^{\infty} c_j x_{t-j} + \varepsilon_t \\ x_t &= d + \sum_{i=1}^{\infty} f_i z_{t-i} + \sum_{j=1}^{\infty} g_j x_{t-j} + u_t \end{aligned} \quad (6.24)$$

Obviamente, en la práctica el número de rezagos no será infinito sino que se fija de manera que sean significativos y adecuados para que los errores estimados sean ruido blanco, lo cual se puede revisar utilizando el correlograma (la función de autocorrelación estimada) o la prueba Q . Como la causalidad depende de que una variable (la causa) provea información adicional al pasado de la otra variable (la causada), es importante utilizar un número suficiente de rezagos de manera que los rezagos de la variable causada aporten toda la información posible. En una prueba de causalidad más rezagos son mejores que menos. En este sentido se recomienda el criterio AIC sobre el SC para escoger el número de rezagos del VAR.

Se estudia una ecuación a la vez. Veamos la segunda, la de x_t como variable dependiente, donde se prueba la causalidad de z hacia x . Se corre la regresión sin restringir, tal como aparece en (6.24), y otra restringida haciendo $f_i = 0$ para todo i . La hipótesis nula, $f_i = 0$, significa que no hay causalidad de z a x ; esto es, que z no causa a x . Como estadístico de prueba se puede usar la F , cuya expresión es:

$$\frac{(\text{suma de cuadrados restringida} - \text{suma de cuadrados no restringida}) / N^{\circ} \text{ de restricciones}}{\text{suma de cuadrados no restringida} / (N^{\circ} \text{ de observaciones utilizables} - N^{\circ} \text{ de parámetros estimados})} \quad (6.25)$$

Donde “ N^o de restricciones” es igual al número de coeficientes eliminados para obtener la ecuación restringida (f_i en el ejemplo) y donde “ N^o de observaciones utilizables” es igual a n menos el número de rezagos. Si la F calculada es significativa, entonces se rechaza la hipótesis de no causalidad y se acepta que sí hay causalidad de z a x . Como estadístico de prueba también se puede utilizar la prueba de Wald, la cual es asintóticamente chi-cuadrado (χ^2) con grados de libertad igual al número de restricciones.

De igual manera se procede con la primera ecuación, en la cual se probaría causalidad de x hacia z . Se corre la regresión sin restringir, tal como aparece en (6.24), y la regresión restringida haciendo $c_j = 0$ para todo j . La hipótesis nula $c_j = 0$, significa que no hay causalidad de x hacia z ; esto es, que x no causa a z .

Si en lugar de dos variables hay m , igualmente se construye un VAR. Cada ecuación contendrá los rezagos de las m variables. La prueba se hace formando la regresión restringida, retirando todos los rezagos de la variable cuya causalidad se desea probar (dejando los de las otras $m - 1$ variables) y comparando, como se hizo arriba, la suma de errores al cuadrado de esta regresión restringida con los de la regresión no restringida.

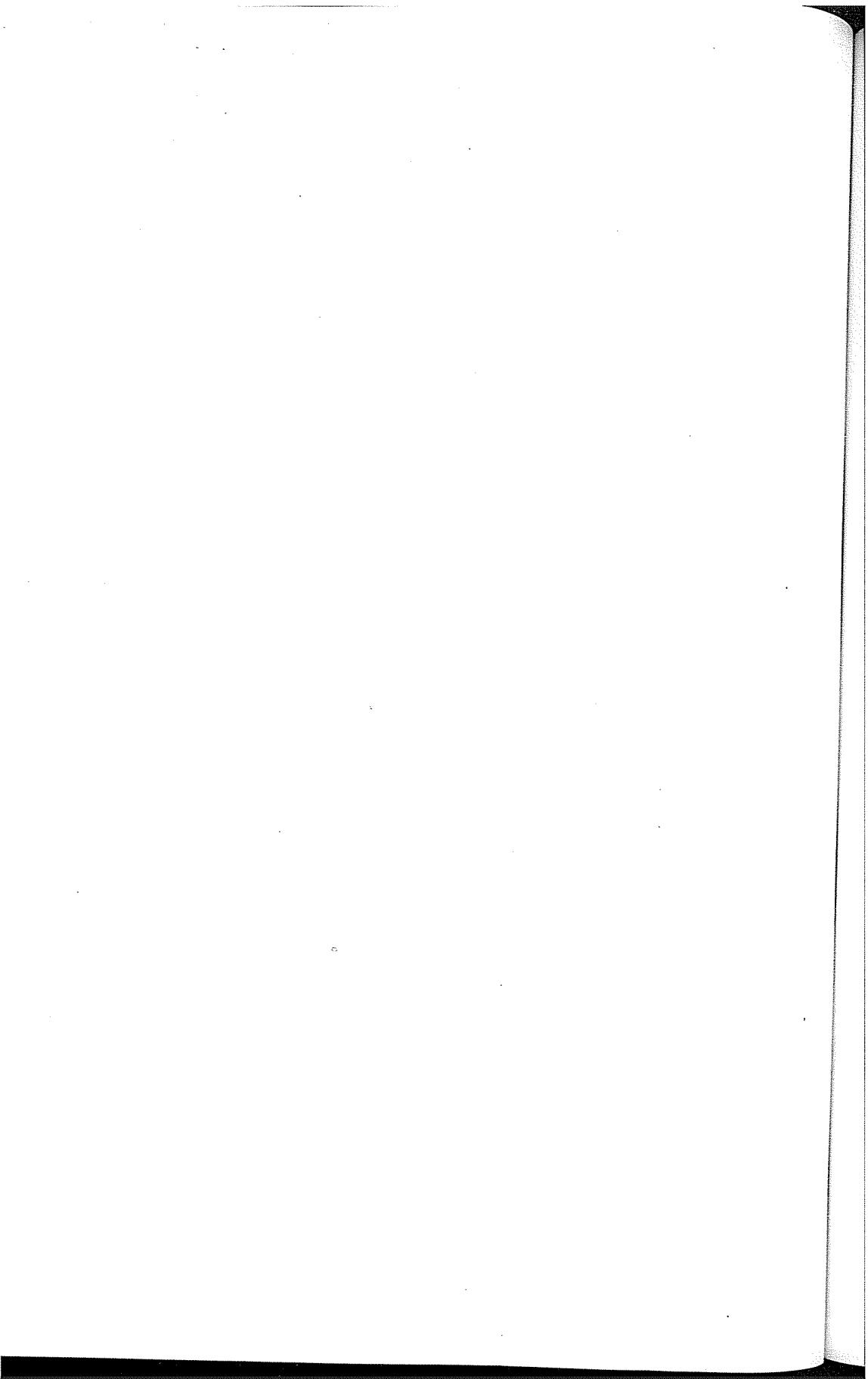
Es importante notar que el hecho de que el movimiento en una variable preceda el de otra, no implica que la cause, aunque es una condición necesaria para la causalidad. Adicionalmente, la información contenida en la variable causante debe ayudar a predecir la causada por encima de otra información relevante. Así, no es cierto, como se afirma con frecuencia, que las ventas de tarjetas (o adornos) de Navidad causen la Navidad en el sentido de Granger. La fecha de Navidad es fija, y una vez considerada esta información o incluido el rezago anual de la serie de Navidad, debe desaparecer el efecto predictivo de las ventas (al respecto, ver Atukeren, 2007). Otros casos son menos claros. Hamilton (1994) nota que el hecho de que el precio de las naranjas cause el clima en la Florida, sugiere cierta capacidad predictiva por parte de los compradores que parecen anticipar el mal clima y suben los precios. En otro estudio, Thurman y Fisher (1988) concluyen que el huevo causa la gallina.

Ejemplo

La prueba de causalidad de Granger aplicada al VAR del Cuadro 5.1, por medio del programa Eviews, se muestra en el Cuadro 6.1. No es posible rechazar la hipótesis de que la tasa de interés de captación (*capta*) no cause la tasa de interés de colocación (*coloc*); pero es más fácil rechazar la hipótesis de que *coloc* no causa *capta*. En consecuencia, podríamos aceptar, con el 96.6 % de confiabilidad (1–0.03371), que la tasa de interés de los préstamos (*coloc*) sí causa la tasa de interés del ahorro (*capta*).

Cuadro 6.1: Prueba de causalidad

Pairwise Granger Causality Tests				
Date: 09-22-1999 Time: 18:32				
Sample: 1987:1-1999:1				
Lags: 4				
Null Hypothesis		Obs.	F-Statistic	Probability
CAPTA does not Granger cause COLOC		45	1.80379	0.14959
COLOC does not Granger cause CAPTA			2.93550	0.03371



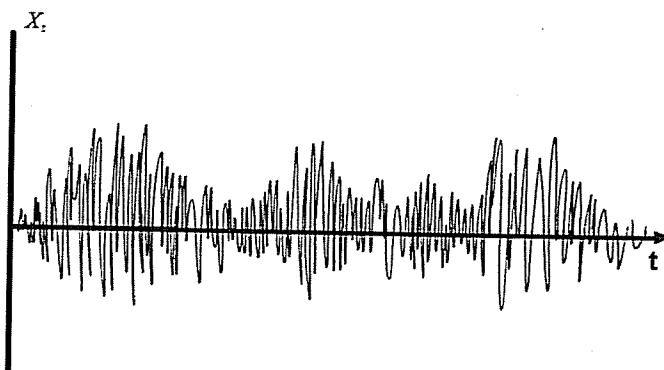
CAPÍTULO 7

Modelos ARCH

Un supuesto usual en econometría es el de la homoscedasticidad; esto es, que la varianza del error es constante. La violación de este supuesto se conoce como heteroscedasticidad, y se ha asociado con frecuencia a estimaciones de corte transversal. Sin embargo, en ocasiones encontramos series de tiempo en las cuales se presentan periodos de mayor varianza que en otros, como brotes de mayor agitación o incertidumbre seguidos por periodos de relativa calma o de menor varianza. No se trata de que una observación aislada alcance mayor magnitud, ya que esto es común en cualquier proceso estocástico, sino que una serie consecutiva de observaciones alcance mayor magnitud en promedio (sin importar el signo).

Se trata entonces de situaciones donde parece haber cierta autocorrelación en la varianza de la serie, como se ilustra en la Figura 7.1. Muchas variables macroeconómicas, especialmente en el área financiera, pueden tener este tipo de comportamiento; la volatilidad de la inflación, la estructura intertemporal de las tasas de interés, la valoración de activos y acciones con respecto a su riesgo y, en general, precios especulativos de alta frecuencia (datos diarios, o cada hora o minuto, etc.).

Figura 7.1: Serie con autocorrelación en la varianza



La técnica ARCH, que quiere decir heteroscedasticidad condicional autorregresiva, permite el modelaje condicional de la varianza, en lugar de la tradicional estimación incondicional que se hace bajo el supuesto de homoscedasticidad, donde la varianza del error es una constante σ^2 .

La hipótesis de expectativas racionales postula que los agentes no desdeñan ni malgastan información útil. En muchos casos, como en las negociaciones salariales o la inversión en acciones, las decisiones que se toman dependen de la varianza o incertidumbre predicha para el periodo futuro relevante y no para el futuro en general. Otra aplicación tiene que ver con la estimación del valor en riesgo (*value at risk*), una medida de la pérdida posible para un portafolio de inversión en el peor escenario. En situaciones como éstas, que involucran o necesitan la predicción de la incertidumbre futura, será beneficioso el modelaje condicional de la varianza.

Engle (1982) propuso modelar la varianza en función de los ε_t del pasado. En otras palabras, σ^2 no se trata como una constante en el tiempo, sino que depende de los errores del pasado, lo cual origina el nombre del modelo ARCH. Partimos del modelo clásico de regresión:

$$y_t = \gamma' x_t + \varepsilon_t, \quad (7.1)$$

Donde y_t es la variable dependiente, γ' es el vector de parámetros y x_t es el vector de variables explicativas. Engle propuso una modificación a los supuestos del error en el sentido de que el error ε_t , condicionado al conjunto de información I_{t-1} en el tiempo $t-1$ (que incluye rezagos del

error), tenga una distribución normal cuya varianza no sea constante, sino un proceso aleatorio estacionario σ_t^2 , esto es:

$$\varepsilon_t / I_{t-1} \sim N(0, \sigma_t^2)$$

Sin embargo, para los errores incondicionales sigue siendo cierto que $\varepsilon_t \sim (0, \sigma^2)$, donde el tipo de distribución no se especifica porque, como se discutirá más adelante, la distribución no necesariamente es normal. En otras palabras, $E(\varepsilon_t^2 / I_{t-1}) = \sigma_t^2$ será la varianza condicional y $E(\varepsilon_t^2) = \sigma^2$ será la varianza incondicional. Nótese que, por la ley de expectativas iterativas, $E(\sigma_t^2) = E[E(\varepsilon_t^2 / I_{t-1})] = E(\varepsilon_t^2) = \sigma^2$.

Modelo ARCH

En la formulación del ARCH(q), el cuadrado de los errores, ε_t^2 , se postula como un autorregresivo:

$$\varepsilon_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \varepsilon_{t-2}^2 + \dots + \beta_q \varepsilon_{t-q}^2 + \nu_t \quad (7.2)$$

Donde ν_t es ruido blanco. Tomando el valor esperado condicional (dado $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$) de la expresión anterior, obtenemos la varianza condicional σ_t^2 :

$$\begin{aligned} \sigma_t^2 &= E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = E(\varepsilon_t^2 / I_{t-1}) \\ &= \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \varepsilon_{t-2}^2 + \dots + \beta_q \varepsilon_{t-q}^2 \end{aligned} \quad (7.3)$$

Nótese que σ_t^2 está en función de datos del pasado ya conocidos, por lo cual puede considerarse la predicción de ε_t^2 , esto es, $E(\varepsilon_t^2 / I_{t-1})$. Una condición suficiente para que la varianza condicional σ_t^2 sea positiva es que los coeficientes β_0, \dots, β_q sean mayores que cero (estrictamente β_0 debe ser mayor que cero y β_1, \dots, β_q mayores o iguales a cero). A primera vista (7.3) parece un promedio móvil (MA); sin embargo, esto no es cierto porque, como se discutirá más adelante, los ε_t^2 tienen correlación serial aunque los ε_t sean ruido blanco.

Puede verse de (7.1) que la varianza de y_t , dado I_{t-1} , es proporcional a la varianza de ε_t dado I_{t-1} . De esta manera, la técnica ARCH permite modelar períodos de tranquilidad y volatilidad en la variable dependiente y_t .

Mientras la varianza condicional en (7.3) no es constante, pues depende del tiempo, la varianza incondicional sí es constante. Tomando valor esperado incondicional de cada término en la expresión (7.2) y, bajo el supuesto de estacionariedad, podemos escribir:

$$\sigma^2 = \beta_0 + \beta_1\sigma^2 + \beta_2\sigma^2 + \dots + \beta_q\sigma^2,$$

ya que, incondicionalmente, todas las varianzas son iguales e independientes del tiempo. De la expresión anterior llegamos a:

$$Var(\varepsilon_t) = \sigma^2 = \frac{\beta_0}{1 - \beta_1 - \beta_2 - \dots - \beta_q} \quad (7.4)$$

La anterior ecuación es la varianza incondicional (o predicción de largo plazo de la varianza). Recalcamos que la expresión (7.4) no es válida si $\beta_1 + \dots + \beta_q \geq 1$ o si (7.2) no es estacionaria.

En la ecuación (7.1) x_t puede incluir variables exógenas, dummies, rezagos de y_t o términos de promedio móvil. En razón de que la metodología ARCH se desarrolla en el contexto del modelo clásico de regresión, preferimos utilizar y_t en este capítulo para denotar la serie de tiempo que actúa como variable dependiente, en lugar de x_t , que venía siendo la notación utilizada en los capítulos anteriores.

La volatilidad también se puede expresar en términos de valores absolutos de los errores en lugar de los cuadrados (véase Schwert, 1989), como $|\varepsilon_t| = \beta_0 + \beta_1|\varepsilon_{t-1}| + \dots + \beta_q|\varepsilon_{t-q}| + \nu_t$, ecuación que provee predicciones de la desviación estándar condicional.

Modelo GARCH

Bollerslev (1986) generalizó el modelo ARCH de manera que la varianza condicional siga un modelo parecido a un ARMA. Este modelo se conoce como GARCH(p, q), o ARCH generalizado, donde la varianza condicional tiene la expresión:

$$\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \alpha_2\sigma_{t-2}^2 + \dots + \alpha_p\sigma_{t-p}^2 + \beta_1\varepsilon_{t-1}^2 + \beta_2\varepsilon_{t-2}^2 + \dots + \beta_q\varepsilon_{t-q}^2 \quad (7.5)$$

Donde p es el número de rezagos de la varianza y q es el número de rezagos de los errores al cuadrado. Aquí la volatilidad no sólo depende de las observaciones anteriores, sino de las volatilidades anteriores. Para

un GARCH(1,1), por ejemplo, la varianza condicional tendría la forma $\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta \varepsilon_{t-1}^2$. Si $\alpha_1 = \beta = 0$, la varianza condicional sería igual a la incondicional y ambas, a su vez, igual a α_0 .

Si la suma $\alpha_1 + \beta$ es cercana a la unidad, quiere decir que la volatilidad sería muy persistente, ya que el efecto de cualquier *shock* tardaría muchos periodos en desaparecer (Engle, 2001). Al reescribir este modelo como $\sigma_t^2 = \alpha_0 + (\alpha_1 + \beta)\sigma_{t-1}^2 + \beta(\varepsilon_{t-1}^2 - \sigma_{t-1}^2)$, donde $\varepsilon_{t-1}^2 - \sigma_{t-1}^2$ es ruido o error de predicción –porque σ_t^2 es $E(\varepsilon_t^2 | I_{t-1})$, la mejor predicción de ε_t^2 –, notamos que el coeficiente $\alpha_1 + \beta$ define el comportamiento de σ_t^2 a través del tiempo.

La popularidad del GARCH radica en que permite modelar con menos términos que el ARCH; esto es, con mayor parsimonia. Usando el operador de rezago L , la ecuación (7.5) queda:

$$\sigma_t^2 = \alpha_0 + A(L)\sigma_t^2 + B(L)\varepsilon_t^2 \quad (7.6)$$

Donde $A(L)$ y $B(L)$ son polinomios en L . Despejando σ_t^2 de (7.6) se puede ver que:

$$\sigma_t^2 = \frac{\alpha_0 + B(L)\varepsilon_t^2}{1 - A(L)} \quad (7.7)$$

Donde, para lograr estacionariedad, las raíces de $1 - A(L) = 0$ deben estar por fuera del círculo unitario. Adicionalmente, como se muestra más adelante en (7.8), se espera que $A(1) + B(1) < 1$, ya que de lo contrario la varianza incondicional daría negativa o indefinida; lo anterior también se requiere para que los momentos de orden 4, 6, etc. de la serie de los errores sean estables. El hecho de que σ_t^2 sea estable no necesariamente implica que los momentos de orden superior también lo sean (Greene, 1997, p. 861).

De nuevo, como en el ARCH, la varianza incondicional es constante ya que, tomando expectativas en (7.5) o (7.6), tenemos:

$$Var(\varepsilon_t) = \sigma^2 = \frac{\alpha_0}{1 - A(1) - B(1)} = \text{constante} \quad (7.8)$$

Donde, por la ley de expectativas iterativas $E(\sigma_t^2) = E[E(\varepsilon_t^2 | I_{t-1})] = E(\varepsilon_t^2) = \sigma^2$. Nótese que $L = 1$ en los polinomios porque las varianzas son iguales e independientes del tiempo en razón de la estacionariedad, de manera que $1 - A(1) - B(1)$ es simplemente la suma algebraica de todos los coeficientes.

Un modelo GARCH estacionario puede expresarse como un ARCH con infinitos rezagos. Por ejemplo el modelo $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2$ puede escribirse como:

$$(1 - \alpha_1 L)\sigma_t^2 = \alpha_0 + \beta\varepsilon_{t-1}^2$$

De donde obtenemos el ARCH:

$$\sigma_t^2 = \frac{\alpha_0}{1 - \alpha_1} + \beta [\varepsilon_{t-1}^2 + \alpha_1\varepsilon_{t-2}^2 + \alpha_1^2\varepsilon_{t-3}^2 + \dots]$$

Cuando haya más rezagos del error al cuadrado, por ejemplo en $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta_1\varepsilon_{t-1}^2 + \beta_2\varepsilon_{t-2}^2$, es posible despejar ε_{t-1}^2 o ε_t^2 en función de infinitos rezagos de σ_t^2 .

La expresión de σ_t^2 en (7.6) se puede adicionar con variables exógenas, por ejemplo:

$$\sigma_t^2 = \alpha_0 + A(L)\sigma_t^2 + B(L)\varepsilon_t^2 + \pi z_t \quad (7.9)$$

E, incluso, *dummies* o rezagos de la variable dependiente de la regresión y_t (Longstaff y Schwartz, 1992). Es de notar que en (7.9) no hay garantía de que σ_t^2 sea positiva. Se sugiere escoger los z de manera que sean siempre positivos (por ejemplo, en valor absoluto), lo que minimiza la posibilidad de tener $\sigma_t^2 < 0$.

En adelante, nos referiremos a estos modelos de varianza condicional indistintamente como ARCH o GARCH.

Una generalización son los GARCH multivariados (MGARCH), similares a los vectores autorregresivos (VAR), donde la varianza condicional de una variable se expresa en función de sus propios rezagos, de rezagos de las varianzas de otras variables y de rezagos de las covarianzas. Por ejemplo, un MGARCH(1, 1) para dos variables:

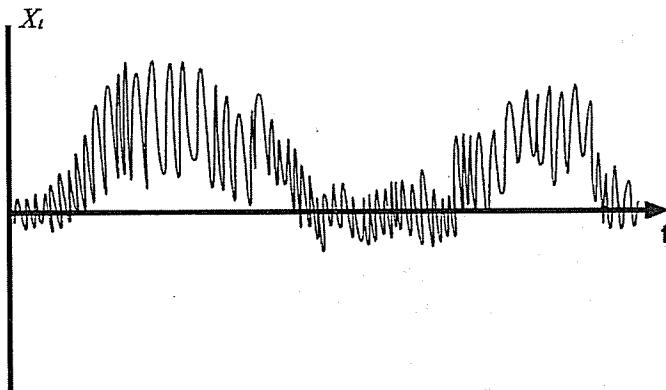
$$\begin{bmatrix} \sigma_{1t}^2 \\ \sigma_{2t}^2 \\ \sigma_{12t} \end{bmatrix} = \begin{bmatrix} \alpha_{01} \\ \alpha_{02} \\ \alpha_{03} \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{bmatrix} \begin{bmatrix} \sigma_{1t-1}^2 \\ \sigma_{2t-1}^2 \\ \sigma_{12t-1} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1}^2 \\ \varepsilon_{2t-1}^2 \\ \varepsilon_{1t-1}\varepsilon_{2t-1} \end{bmatrix}$$

Donde $\sigma_{12t} = \sigma_{21t}$ representa la covarianza condicional de los errores.

Modelo ARCH-M

Otra extensión del modelo se conoce como ARCH-M, debida a Engle, Lilien y Robins (1987). Se trata, como se muestra en la Figura 7.2, de modelar el hecho de que la media condicional de muchas series puede estar relacionada con la varianza; por ejemplo, la rentabilidad exigida de una acción y su varianza suelen estar correlacionadas positivamente. Este enfoque ha despertado gran interés en estudios financieros sobre la prima de riesgo, esto es, la compensación entre retorno esperado y riesgo, como en los modelos de valoración de activos de capital (CAPM). También se ha utilizado para estudiar la relación entre la inflación y su varianza.

Figura 7.2: Efecto ARCH en la media de la serie



Una forma sencilla de implementar este comportamiento sería modificando el modelo clásico de regresión, de manera que incluya la varianza condicional como variable explicativa:

$$y_t = \gamma_0 + \gamma_1 \sigma_t^2 + \varepsilon_t \quad (7.10)$$

Donde en general σ_t^2 es un proceso GARCH como en (7.6); sin embargo, en la práctica se prefiere la más simple especificación ARCH (7.3), porque facilita la convergencia en la estimación de este modelo. La variable y_t podría ser, por ejemplo, el retorno adicional exigido (*premium*), proporcional al nivel de incertidumbre, por invertir en un activo. Si r_t es el retorno del activo y r_f es el retorno libre de riesgo,

entonces podríamos postular que $y_t = r_t - r_f = \gamma_0 + \gamma\sigma_t^2 + \varepsilon_t$, donde la media condicional de y_t es una función de la varianza condicional (Chou, 1988).

La ecuación de regresión (7.10) puede contener variables adicionales, por ejemplo:

$$y_t = \gamma_0 + \gamma_1\sigma_t^2 + \gamma_2z_t + \varepsilon_t \quad (7.11)$$

Aunque esta formulación tiende a alejarse de la derivación del modelo de prima de riesgo cuando $y_t = r_t - r_f$ (Enders, 2004, capítulo 3). También en (7.10) se puede usar la desviación estándar σ_t , o la raíz de σ_t , en lugar de la varianza σ_t^2 .

Identificación del modelo ARCH

La identificación es quizás la parte más difícil de la estimación de los modelos ARCH. Como una primera aproximación se puede correr la regresión $y_t = \gamma'x_t + \varepsilon_t$, especificada en (7.1), guardar $\hat{\varepsilon}_t$, y luego utilizar $\hat{\varepsilon}_t^2$ para observar su función de autocorrelación en busca de pistas que indiquen autocorrelación en los errores al cuadrado. Si la función de autocorrelación (ACF) de $\hat{\varepsilon}_t^2$ corresponde a ruido blanco, no habría sospecha de ARCH.

Existe una prueba del tipo multiplicador de Lagrange (LM) que consiste en probar un ARCH(q) contra la no existencia de ARCH. En esta prueba LM se corre la regresión usual $y_t = \gamma'x_t + \varepsilon_t$ y se guardan los residuos $\hat{\varepsilon}_t$ para luego correr $\hat{\varepsilon}_t^2$ contra una constante y q rezagos de $\hat{\varepsilon}_t^2$, así:

$$\hat{\varepsilon}_t^2 = \alpha_0 + \beta_1\hat{\varepsilon}_{t-1}^2 + \beta_2\hat{\varepsilon}_{t-2}^2 + \dots + \beta_q\hat{\varepsilon}_{t-q}^2 + \eta_t \quad (7.12)$$

Donde η_t es el término de error de la regresión.

Si no hay efecto ARCH, entonces debe esperarse que solamente el coeficiente α_0 sea significativo y positivo. Bajo la hipótesis nula de que no hay ARCH hasta de orden q , se encuentra que el número de observaciones n multiplicado por el R^2 (que tiende a cero bajo la hipótesis nula) tiene una distribución asintótica chi-cuadrado con q grados de libertad (χ_q^2). Si nR^2 es mayor que el valor crítico escogido, entonces se puede concluir que sí hay efecto ARCH, al menos hasta de orden q ; y si la prueba es positiva, entonces es muy posible que la especificación

GARCH, siendo más general, sirva para lograr parsimonia. También, de manera aproximada, se puede utilizar la prueba F de la regresión (7.12); aquí el estadístico F estimado no tiene la distribución usual, porque el error η_t sólo puede ser aproximadamente normal en razón de que está limitado para garantizar $\hat{\varepsilon}_t^2 > 0$.

Como siempre, las pruebas son válidas si la expresión $y_t = \gamma'x_t + \varepsilon_t$ está bien especificada, ya que una mala especificación puede inducir autocorrelación en los errores al cuadrado. En otras palabras, el hecho de que las pruebas rechacen la hipótesis de ausencia de ARCH, no necesariamente implica la existencia de ARCH en el modelo.

Para no excederse innecesariamente en el uso del modelaje ARCH, es importante notar que este efecto con frecuencia desaparece cuando se incluyen rezagos de la variable dependiente como variables explicativas en la regresión. En este caso la reespecificación del modelo estaría indicada.

Para un ARCH-M, los $\hat{\varepsilon}_t$ se obtienen de correr y_t contra la constante γ_0 , la cual (como se sabe de MCO) resultará siendo \bar{y}_t , de donde $\hat{\varepsilon}_t = y_t - \bar{y}_t$. Luego se corre $\hat{\varepsilon}_t^2$ contra una constante y su pasado y se aplica la prueba LM. Si hay evidencia de heteroscedasticidad condicional autorregresiva, entonces el modelo $y_t = \gamma_0 + \varepsilon_t$ podría estar mal especificado y se sugiere probar con $y_t = \gamma_0 + \gamma_1\sigma_t^2 + \varepsilon_t$.

Estimación del modelo ARCH

El modelo que se va a estimar está formado por la ecuación de regresión $y_t = \gamma'x_t + \varepsilon_t$ y la ecuación de varianza $\sigma_t^2 = \alpha_0 + A(L)\sigma_t^2 + B(L)\varepsilon_t^2$.

Los errores ε_t cumplen con las condiciones de la regresión clásica en el sentido de tener media cero, varianza incondicional constante y carecer de correlación serial entre ellos. En consecuencia, los coeficientes de la regresión pueden estimarse por mínimos cuadros ordinarios (MCO), ya que esta técnica continúa siendo el mejor estimador lineal insesgado (*blue*). Los parámetros de la ecuación de varianza (7.3) de la especificación ARCH pueden estimarse con los residuos de la regresión, usando $\hat{\varepsilon}_t^2$ como *proxy* de σ_t^2 . Sin embargo, la estimación no lineal por máxima verosimilitud es más eficiente que por MCO (Greene, 1997, p. 571) y es la utilizada de preferencia.

Adicionalmente, los errores estándar de la estimación por MCO pueden requerir ajuste para tomar en cuenta la heteroscedasticidad (Bollerslev, Engle y Nelson, 1994, p. 2981). Un estimador por mínimos cuadrados generalizados se puede consultar en Judge *et al.* (1985, p. 443), y uno por el método generalizado de momentos en Hamilton (1994, p. 664).

Formar la verosimilitud con los errores incondicionales sería inútil, porque éstos tienen varianza constante, cuando precisamente estamos tratando de estimar los parámetros de σ_t^2 . Más aún, Bollerslev, Engle y Nelson (1994) sugieren que la distribución incondicional de ε_t ni siquiera es normal, sino *leptokúrtica*. Esto puede intuirse porque los errores incondicionales son un promedio de los condicionales (que a veces tienen varianza pequeña y a veces mayor), de manera que su distribución tendrá un exceso de valores cerca y lejos del promedio.

En cuanto a los errores condicionales podemos decir que, así no haya correlación serial entre ellos (lo cual es una propiedad lineal), no son independientes porque están conectados a través de sus cuadrados (una relación no lineal). Por ejemplo, un ε_t grande en valor absoluto sugiere una distribución de probabilidad con varianza grande en $t + 1$; un ε_t^2 grande tiende a ser seguido por un ε_{t+1}^2 grande. En otras palabras, saber ε_t modifica la asignación de probabilidades para ε_{t+1} . Curiosamente, la ausencia de correlación entre los errores condicionales implica que si su verosimilitud es normal, ésta se podrá expresar como el producto de las densidades normales individuales, lo cual implica independencia, al menos desde el punto de vista de la definición matemática de independencia. En consecuencia, la verosimilitud normal conjunta para los errores condicionales, cuya matriz de varianza covarianza es:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_n^2 \end{bmatrix},$$

Será de la forma:

$$\begin{aligned}
 f(\varepsilon_1, \dots, \varepsilon_n) &= \frac{1}{(2\pi)^{n/2}(\det\Sigma)^{1/2}} \exp\left[-\frac{1}{2}(\varepsilon'\Sigma^{-1}\varepsilon)\right] \\
 &= \frac{1}{(2\pi)^{n/2}(\sigma_1^2\sigma_2^2\dots\sigma_n^2)^{1/2}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_1^2}{\sigma_1^2} + \frac{\varepsilon_2^2}{\sigma_2^2} + \dots + \frac{\varepsilon_n^2}{\sigma_n^2}\right)\right] \\
 &= f_1(\varepsilon_1)f_2(\varepsilon_2)\dots f_n(\varepsilon_n)
 \end{aligned} \tag{7.13}$$

Para hacer el cambio de variables a y_1, \dots, y_n , debemos multiplicar la densidad anterior por el valor absoluto del determinante del jacobiano [véase (4.34)] y, además, remplazar en la densidad las expresiones ε_t por su equivalente $y_t - \gamma'x_t$. El jacobiano del problema anterior es una matriz diagonal (o triangular si hay rezagos de y_t) de orden $n \times n$ que contiene las derivadas de ε_t con respecto a y_t , esto es, con unos en su diagonal. El determinante del jacobiano, siendo el producto de los términos de su diagonal, será unitario. Así, la verosimilitud conjunta de las observaciones queda:

$$f(y_1, \dots, y_n) = \frac{1}{(2\pi)^{n/2}(\sigma_1^2\sigma_2^2\dots\sigma_n^2)^{1/2}} \exp\left[-\frac{1}{2}\sum_{t=1}^n\left(\frac{y_t - \gamma'x_t}{\sigma_t}\right)^2\right] \tag{7.14}$$

Con la correspondiente verosimilitud logarítmica:

$$\ell = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\sum_{t=1}^n \ln \sigma_t^2 - \frac{1}{2}\sum_{t=1}^n \left(\frac{y_t - \gamma'x_t}{\sigma_t}\right)^2 \tag{7.15}$$

También hubiéramos podido trabajar con los errores estandarizados ε_t/σ_t que tienen media cero y varianza condicional constante y unitaria; esto los hace idénticamente distribuidos, de manera que conocer ε_t/σ_t no modifica la asignación de probabilidades para $\varepsilon_{t+1}/\sigma_{t+1}$; en otras palabras, los errores estandarizados, además de no estar correlacionados, son independientes. En consecuencia, bajo este supuesto de que los errores estandarizados se distribuyen *iid* normales con media cero y varianza uno, construimos la verosimilitud:

$$f\left(\frac{\varepsilon_1}{\sigma_1}, \dots, \frac{\varepsilon_n}{\sigma_n}\right) = \frac{1}{(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\left(\frac{\varepsilon_1^2}{\sigma_1^2} + \frac{\varepsilon_2^2}{\sigma_2^2} + \dots + \frac{\varepsilon_n^2}{\sigma_n^2}\right)\right]$$

En ella, al hacer el cambio de variables reemplazando las expresiones ε_t/σ_t por su equivalente $(y_t - \gamma'x_t)/\sigma_t$ y multiplicando la verosimilitud por el valor absoluto del determinante del jacobiano, que en este caso es $\sigma_1^{-1}\sigma_2^{-1}\dots\sigma_n^{-1} = (\sigma_1^2\sigma_2^2\dots\sigma_n^2)^{-1/2}$, obtenemos la misma expresión (7.14).

En caso de sospechar falta de normalidad en la distribución de los errores, se puede utilizar un método estadísticamente consistente para estimar los parámetros; este método, llamado cuasi o seudomáxima verosimilitud, consiste en maximizar la verosimilitud normal como si fuera la correcta y luego hacer un ajuste en las varianzas estimadas (Greene, 2003, capítulo 11). En cualquier caso, el método de máxima verosimilitud se puede aplicar con otras distribuciones.

En la optimización de (7.15), σ_t^2 se reemplaza por su expresión (7.5). Faltarán algunas observaciones en razón de los rezagos, por lo cual la estimación de máxima verosimilitud de todas maneras es aproximada. Por ejemplo, los términos $\left(\frac{y_t - \gamma'x_t}{\sigma_t}\right)^2$ en (7.15) serán entonces de la forma:

$$\frac{(y_t - \gamma'x_t)^2}{\alpha_0 + \alpha_1\sigma_{t-1}^2 + \alpha_2\sigma_{t-2}^2 + \dots + \alpha_p\sigma_{t-p}^2 + \beta_1\varepsilon_{t-1}^2 + \beta_2\varepsilon_{t-2}^2 + \dots + \beta_q\varepsilon_{t-q}^2} \quad (7.16)$$

El término σ_t^2 , que también aparece en el segundo término de (7.15), se puede construir de manera similar a como se hizo el procedimiento iterativo en la estimación del ARMA en el Capítulo 4. Para ilustrar tomamos un GARCH(1,1), cuya expresión es $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta_1\varepsilon_{t-1}^2$, donde vemos que la varianza condicional podría construirse de la siguiente manera:

$$\sigma_2^2 = \alpha_0 + \beta_1\varepsilon_1^2 = \alpha_0 + \beta_1(y_1 - \gamma'x_1)^2$$

$$\sigma_3^2 = \alpha_0 + \alpha_1\sigma_2^2 + \beta_1\varepsilon_2^2 = \alpha_0 + \alpha_1\left(\alpha_0 + \beta_1(y_1 - \gamma'x_1)^2\right) + \beta_1(y_2 - \gamma'x_2)^2$$

⋮

Bollerslev (1986, p. 316) sugiere remplazar los primeros valores desconocidos de σ_t^2 y de ε_t^2 en las expresiones anteriores, por $\hat{\sigma}^2$, donde $\hat{\sigma}^2 = n^{-1} \sum (y_t - \hat{\gamma}'x_t)^2$. Otra opción para estimar estos valores desconocidos es intentar alguna técnica de *backcasting*, similar a la descrita en el Capítulo 4 para la estimación de modelos ARMA.

La maximización de la verosimilitud logarítmica puede hacerse por algún método iterativo apropiado, por ejemplo el de Scoring, que se encuentra incluido en la mayoría de los programas de computador. Por lo complejo de la expresión, la optimización de los ARCH no siempre converge fácilmente. Los valores iniciales requeridos por la optimización numérica para γ se pueden obtener del estimador $\hat{\gamma}$, que resulta de aplicar MCO a la ecuación de regresión. Adicionalmente, la varianza estimada s^2 de los errores de esta regresión se puede usar como valor inicial para α_0 en la ecuación de la varianza condicional, mientras que los parámetros de los demás rezagos en $A(L)$ y $B(L)$ pueden inicialmente fijarse en cero. Una alternativa para los parámetros de $B(L)$ es tomarlos de la regresión de $\hat{\varepsilon}_t^2$ contra una constante y q rezagos de $\hat{\varepsilon}_t^2$. Para el modelo ARCH-M cambiamos $\varepsilon_t = y_t - \gamma'x_t$ por $\varepsilon_t = y_t - \gamma_0 - \gamma_1\sigma_t^2$.

Muchos programas de computador incluyen la opción de realizar pruebas de diagnóstico con los errores estandarizados ε_t/σ_t . Por ejemplo, es de esperar que si el modelo está bien especificado, la ACF de estos errores estandarizados elevados al cuadrado denote ruido blanco y que, de manera similar, el estadístico de la prueba LM sea insignificante por la ausencia de autocorrelación entre estos errores estandarizados al cuadrado.

R^2 negativo

En el caso de la estimación ARCH, ocasionalmente aparece un R^2 negativo. Revisando (7.15), vemos que no estamos minimizando la suma de errores al cuadrado (que es el objetivo usual) sino que, en adición al segundo término de (7.15), se trata de expresiones mucho más complicadas, como (7.16) con ε_t^2 tanto en el numerador como en el denominador. Por ejemplo, si corremos el modelo $y_t = \gamma_0 + \varepsilon_t$, la estimación $\hat{\gamma}_0$ puede diferir de la media de la variable dependiente, \bar{y} , con el riesgo de producir $\sum \hat{\varepsilon}_i^2 = \sum (y_i - \hat{\gamma}_0)^2 > \sum (y_i - \bar{y})^2$, con lo cual el R^2 sería negativo de acuerdo con la fórmula $R^2 = 1 - (\sum \hat{\varepsilon}_i^2 / \sum (y_i - \bar{y})^2)$.

Son conocidos otros casos de R^2 negativo en la literatura, ya no relativos a la estimación ARCH, pero ilustrativos. Por ejemplo, es bien sabido que omitir la constante en una regresión puede producir un $\sum \hat{\varepsilon}_i^2$ mayor que $\sum (y_i - \bar{y})^2$. La razón básica es que al omitir la constante,

la sumatoria de errores $\sum \hat{\varepsilon}_i$ ya no necesariamente es igual a cero, de manera que la expresión cuadrática $\sum \hat{\varepsilon}_i^2$ sólo podrá ser mayor (o igual) que cuando se incluye la constante en la regresión. Recordamos que en el modelo clásico de regresión (con la notación matricial más usual) $y = x\beta + \varepsilon$, tenemos que:

$$x'\hat{\varepsilon} = x'(y - x\hat{\beta}) = x'y - x'x(x'x)^{-1}x'y = x'y - x'y = 0$$

Donde, si la primera columna de la matriz x es la columna de unos de la constante, se garantiza que $\sum \hat{\varepsilon}_i = 0$.

Otro caso de posible R^2 negativo es cuando, aunque se incluya la constante, se hace la estimación por variable instrumental o por un método que utilice variable instrumental, como 2SLS. El método de variable instrumental no fue derivado, como si lo fue MCO, para minimizar $\sum \hat{\varepsilon}_i^2$, por lo cual (aunque $\sum \hat{\varepsilon}_i = 0$) puede suceder que $\sum \hat{\varepsilon}_i^2 > \sum (y_i - \bar{y})^2$ resultando en un $R^2 < 0$. De todas maneras, y al igual que en MCO, en variable instrumental también se cumple $\sum \hat{\varepsilon}_i = 0$ si se incluye la constante porque, siendo z la matriz de instrumentos con unos en la primera columna:

$$z'\hat{\varepsilon} = z'(y - x(z'x)^{-1}z'y) = z'y - z'x(z'x)^{-1}z'y = z'y - z'y = 0$$

Ejemplo

El Cuadro 7.1 presenta un ejercicio ARCH. En este se utiliza la misma serie de primeras diferencias de los datos trimestrales del índice de precios, llamada $P1$, que fue utilizada en el ejemplo del Cuadro 4.2, donde se estimaba un modelo $AR(4)$. Aquí, se estima un modelo GARCH(1, 1). La ecuación de la varianza es $\sigma_t^2 = 2.318 + .876\varepsilon_{t-1}^2 + .26\sigma_{t-1}^2$.

Cuadro 7.1: Ejemplo de estimación ARCH

Dependent Variable: P1				
Method: ML-ARCH				
Date: 07/26/99 Time: 19:11				
Sample(ajusted): 1985:1 1999:1				
Included observations: 57 after adjusting endpoints				
Convergence achieved after 26 iterations				
	Coefficient	Std. Error	z-Statistic	Prob.
P1(-1)	0.109497	0.067531	1.621441	0.1049
P1(-2)	-0.031264	0.056662	-0.551755	0.5811
P1(-3)	0.117566	0.046946	2.504283	0.0123
P1(-4)	0.987691	0.055065	17.93668	0.0000
Variance Equation				
C	2.318445	1.521412	1.523877	0.1275
ARCH(1)	0.876320	0.376273	2.328948	0.0199
GARCH(1)	0.260737	0.138492	1.882679	0.0597
R-squared	0.658884	Mean dependent var	13.90175	
Adjusted R-squared	0.617950	S.D. dependent var	11.93127	
S.E. of regression	7.374742	Akaike info criterion	5.821614	
Sum squared resid	2719.341	Schwarz criterion	6.072515	
Log likelihood	-158.9160	F-statistic	16.09627	
Durbin-Watson stat	1.280362	Prob(F-statistic)	0.00000	

Predicción con un ARCH

La predicción con los modelos ARCH tiene dos partes igualmente importantes. La predicción de la media condicional, esto es, de la variable dependiente y_t de la regresión, y la predicción de la varianza condicional σ_t^2 . La predicción de y_t es relativamente directa y se basa en la ecuación de regresión estimada y en el conjunto de información

disponible (véase el Capítulo 3, en caso de tratarse de la predicción de una ecuación ARMA). La predicción de σ_t^2 se basa en la expresión condicional de la varianza:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \alpha_2 \sigma_{t-2}^2 + \dots + \alpha_p \sigma_{t-p}^2 + \beta_1 \varepsilon_{t-1}^2 + \beta_2 \varepsilon_{t-2}^2 + \dots + \beta_q \varepsilon_{t-q}^2.$$

Llamando $\hat{\sigma}_{n+T}^2$ la predicción hecha en n para $n+T$, tenemos:

$$\begin{aligned}\hat{\sigma}_{n+1}^2 &= \alpha_0 + \alpha_1 \sigma_n^2 + \alpha_2 \sigma_{n-1}^2 + \dots + \alpha_p \sigma_{n+1-p}^2 \\ &\quad + \beta_1 \varepsilon_n^2 + \beta_2 \varepsilon_{n-1}^2 + \dots + \beta_q \varepsilon_{n+1-q}^2 \\ \hat{\sigma}_{n+2}^2 &= \alpha_0 + \alpha_1 \hat{\sigma}_{n+1}^2 + \alpha_2 \sigma_n^2 + \dots + \alpha_p \sigma_{n+2-p}^2 \\ &\quad + \beta_1 \hat{\sigma}_{n+1}^2 + \beta_2 \varepsilon_n^2 + \dots + \beta_q \varepsilon_{n+2-q}^2 \\ \hat{\sigma}_{n+3}^2 &= \alpha_0 + \alpha_1 \hat{\sigma}_{n+2}^2 + \alpha_2 \hat{\sigma}_{n+1}^2 + \dots + \alpha_p \sigma_{n+3-p}^2 \\ &\quad + \beta_1 \hat{\sigma}_{n+2}^2 + \beta_2 \hat{\sigma}_{n+1}^2 + \dots + \beta_q \varepsilon_{n+3-q}^2 \\ &\vdots\end{aligned}\tag{7.17}$$

Llegará un momento, para T grande, cuando el modelo ya no provea información y entonces la mejor predicción de $\hat{\sigma}_{n+T}^2$ será su valor esperado incondicional σ^2 . La convergencia hacia σ^2 al aumentar T no necesariamente se presenta de manera monótona. Las expresiones (7.17) pueden utilizarse para determinar la varianza de predicción de y_t condicional al conjunto de información disponible, $(\hat{y}_{n+T}/y_n, y_{n-1}, y_{n-2}, \dots)$. Por ejemplo, en el modelo:

$$y_t = \gamma_0 + \gamma_1 y_{t-1} + \varepsilon_t + \gamma_2 \varepsilon_{t-1},$$

la varianza de la predicción $\hat{y}_{n+1} = \gamma_0 + \gamma_1 y_n + \gamma_2 \varepsilon_n$ será $\hat{\sigma}_{n+1}^2$, dado que aquí se hace cero el término ε_{n+1} , lo cual introduce error en la predicción.

En general, para cualquier horizonte $n+T$ sabemos, del Capítulo 3, que el error de la predicción óptima será de la forma $\varepsilon_{n+T} + \beta_1 \varepsilon_{n+T-1} + \beta_2 \varepsilon_{n+T-2} + \dots + \beta_{T-1} \varepsilon_{n+1}$, de manera que la varianza de predicción se puede escribir como $\sigma_{n+T}^2 + \beta_1^2 \sigma_{n+T-1}^2 + \beta_2^2 \sigma_{n+T-2}^2 + \dots + \beta_{T-1}^2 \sigma_{n+1}^2$, donde los términos $\sigma_{n+T}^2, \sigma_{n+T-1}^2, \dots, \sigma_{n+1}^2$ pueden obtenerse de un modelo ARCH.

Ejemplo

Para el GARCH(1, 1) $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2$, con varianza incondicional $\sigma^2 = \alpha_0/(1 - \alpha_1 - \beta)$, tendremos las predicciones:

$$\hat{\sigma}_{n+1}^2 = \alpha_0 + \alpha_1\sigma_n^2 + \beta\varepsilon_n^2$$

$$\begin{aligned}\hat{\sigma}_{n+2}^2 &= \alpha_0 + \alpha_1\hat{\sigma}_{n+1}^2 + \beta\hat{\sigma}_{n+1}^2 \\ &= \alpha_0 + (\alpha_1 + \beta)\alpha_0 + (\alpha_1 + \beta)[\alpha_1\sigma_n^2 + \beta\varepsilon_n^2]\end{aligned}$$

$$\begin{aligned}\hat{\sigma}_{n+3}^2 &= \alpha_0 + (\alpha_1 + \beta)\hat{\sigma}_{n+2}^2 \\ &= \alpha_0 + (\alpha_1 + \beta)\alpha_0 + (\alpha_1 + \beta)^2\alpha_0 + (\alpha_1 + \beta)^2[\alpha_1\sigma_n^2 + \beta\varepsilon_n^2]\end{aligned}$$

⋮

$$\hat{\sigma}_{n+T}^2 = \alpha_0 [1 + (\alpha_1 + \beta) + \dots + (\alpha_1 + \beta)^{T-1}] + (\alpha_1 + \beta)^{T-1} [\alpha_1\sigma_n^2 + \beta\varepsilon_n^2]$$

Expresión que, para $\alpha_1 + \beta < 1$ y T grande, tiende al valor esperado incondicional de σ_t^2 :

$$= \frac{\alpha_0}{1 - \alpha_1 - \beta} + 0 = \sigma^2$$

Otras aplicaciones

En la práctica, muchas series económicas no son estacionarias y, en particular, son crecientes. Un modelo autorregresivo o ARMA de bajo orden, aunque no estacionario, podría proponerse para estas series. Sin embargo, un supuesto del modelaje tradicional ARMA es que la varianza del error es constante, lo cual no parece adecuado en estos casos en que la intuición sugiere que la varianza crece con el nivel de la serie. Ilustremos suponiendo un *AR*(1) no estacionario para modelar la serie que llamamos x_t :

$$x_t = \alpha_1 x_{t-1} + \varepsilon_t \tag{7.18}$$

Donde $\alpha > 1$. Por ejemplo, α podría ser 1.05 e indicar que la serie x_t tiende a crecer al 5% por periodo. Con varianza del error constante en (7.18), x_t podría predecirse sin error en el límite, ya que ε_t se torna despreciable comparado con el nivel creciente de x_t . Un escenario más

realista sería suponer que, aunque los errores no estén correlacionados, su varianza cambia con el nivel de la serie. Una especificación como (7.18) con varianza del error variable podría modelarse como ARCH y mejorar la eficiencia de la estimación. En este caso, la varianza se estimaría como no estacionaria con tendencia estocástica creciente.

En la literatura se encuentran muchas variaciones del modelo ARCH que se han diseñado con el objeto de adaptarlo a las peculiaridades de los fenómenos económicos y financieros. Hay ARCH asimétricos, llamados AARCH, TARCH o EGARCH, donde la varianza condicional depende no sólo de la magnitud de los shocks del pasado (como en GARCH), sino también de su signo; por ejemplo, donde los movimientos hacia la baja en el mercado son seguidos por mayor volatilidad que en el caso de los movimientos al alza, como en los modelos de apalancamiento financiero. En el TARCH se utiliza la ecuación:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta \varepsilon_{t-1}^2 + \varphi \varepsilon_{t-1}^2 d_{t-1} \quad (7.19)$$

Donde $d_t = 1$ si $\varepsilon_t < 0$ y $d_t = 0$ en caso contrario; por ejemplo, $\varepsilon_t < 0$ podría representar malas noticias. Obviamente, si φ resulta ser cero, entonces no hay asimetría. Cuando se forman los términos $\varepsilon_t = y_t - \gamma' x_t$ en la verosimilitud, y al escoger γ en cada iteración, se sabe si el error es mayor o menor que cero, con lo cual se determina d_t . También pueden adicionarse variables exógenas o más rezagos del término asimétrico en el modelo.

Un problema con el TARCH es que la varianza podría ser negativa si $\varphi < 0$. El EGARCH resuelve este problema utilizando logaritmos o exponentiales, según la expresión:

$$\ln \sigma_t^2 = \alpha_0 + \alpha_1 \ln \sigma_{t-1}^2 + \beta \left| \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \right| + \varphi \frac{\varepsilon_{t-1}}{\sigma_{t-1}} \quad (7.20)$$

Donde $\varphi \neq 0$ implica asimetría. Igualmente, se pueden adicionar variables exógenas o más rezagos del término asimétrico en este modelo.

Es más difícil predecir σ_t^2 en un EGARCH, ya que eventualmente los términos $\hat{\varepsilon}_{t+T}$ en el numerador de los cocientes de la expresión (7.20) quedarían reemplazados, a su vez, por su mejor predicción que es σ_{t+T} , igual al denominador de los cocientes. Para un TARCH, en muchas ocasiones, se puede suponer $E(d_t) = .5$ (véase Enders, 2004, p. 142).

Otros modelos, como los SWARCH, o ARCH de *suicheo*, suponen que existen diversos regímenes ARCH y que el sistema alterna entre ellos. En el modelo GARCH de componentes se permite que, de cierta forma, la constante varíe con el tiempo en la ecuación de la varianza condicional. El IGARCH o GARCH integrado (algo similar a la I en el ARIMA) es básicamente un GARCH con raíces unitarias. También hay, naturalmente, extensiones al ARCH multivariado, que incluyen un concepto parecido al de la cointegración (que veremos más adelante) entre las diversas varianzas, llamado *copersistencia*. Para una revisión de esta literatura, referencias adicionales y ejemplos, véase Bollerslev, Engle y Nelson (1994) o Franses y Van Dijk (2003).

Valor en riesgo

Un ejercicio común en la práctica de las finanzas es el de proyectar la rentabilidad de un activo o portafolio para un periodo futuro determinado. Para esto se necesita conocer la distribución de probabilidad condicional de la rentabilidad en ese periodo. Con esta información podemos encontrar la probabilidad de que la rentabilidad sea mayor o menor que cierto nivel, o proyectar la pérdida máxima del portafolio para un nivel de confianza dado. Este es el tema del enfoque conocido como valor en riesgo. La abreviatura popularizada es VaR, del inglés *Value at Risk*, la cual no debe confundirse con la del vector autorregresivo (VAR).

El VaR se construye con el valor crítico de la cola izquierda de la distribución de retornos, para un nivel de confianza dado. Con frecuencia el VaR se expresa en moneda; simplemente multiplicando el valor crítico por el valor del portafolio. Por ejemplo, un VaR de \$5 millones con un nivel de confianza del 99% (o de significancia del 1%) quiere decir que la probabilidad de perder más de \$5 millones en el próximo periodo es de 1%.

Una primera aproximación para este cálculo puede hacerse con una distribución normal, extrapolando la media y la desviación estándar a partir de los promedios históricos. También se puede utilizar el histograma estimado de la información histórica junto con los valores críticos observados.

Otra alternativa de cálculo del VaR consiste en proyectar la media y la varianza condicional a través de un modelo ARCH (Engle, 2001; Berkowitz y O'Brien, 2001). Por ejemplo, supongamos que el retorno r_t de un portafolio se modela como:

$$r_t = \gamma_0 + \gamma_1 r_{t-1} + \varepsilon_t + \gamma_2 \varepsilon_{t-1} \quad (7.21)$$

Y la ecuación de la varianza de ε_t como:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta \varepsilon_{t-1}^2 \quad (7.22)$$

Adicionalmente, supongamos normalidad y un nivel de confianza del 1% (lo cual equivale a un valor crítico de -2.33 en una normal estándar). Así, el VaR en términos del retorno para el próximo periodo será:

$$\hat{r}_{t+1} - 2.33 \hat{\sigma}_{t+1} \quad (7.23)$$

O, en términos de moneda, será $(\hat{r}_{t+1} - 2.33 \hat{\sigma}_{t+1})$ multiplicado por el valor del portafolio en el tiempo t . Los términos \hat{r}_{t+1} y $\hat{\sigma}_{t+1}$ son las predicciones hechas en el tiempo t para el tiempo $t + 1$ a partir de las ecuaciones (7.21) y (7.22).

CAPÍTULO 8

Procesos estocásticos no estacionarios

Los procesos estacionarios discutidos en capítulos anteriores son aquellos cuyas propiedades estadísticas no cambian en el tiempo; en otras palabras, su función de distribución conjunta es la misma en t y en $t + \tau$. Los procesos de ruido blanco son estacionarios, al igual que los procesos ARMA(p, q), cuando las raíces de su ecuación auxiliar están por fuera del círculo unitario.

En general, los procesos estocásticos estacionarios se caracterizan, por que: (1) tienen un mecanismo que mantiene su magnitud dentro de cierto límite, esto es, muestran una tendencia a volver a su media, la cual es constante; (2) su varianza es finita, y (3) su función de autocorrelación ACF cae relativamente rápido a medida que τ aumenta.

En contraste, un proceso no estacionario (1) tiende a moverse en una sola dirección, nunca o rara vez cruzando el eje, vagando sin estar anclado a su media, sin tendencia a revertir a la media; (2) su varianza tiende a infinito cuando el tiempo tiende a infinito, y (3) su ACF estimada cae muy lentamente a medida que aumenta τ , como en la Figura 2.2 del Capítulo 2.

Aunque la no estacionariedad en media (en tendencia) es la que tradicionalmente ha recibido mayor atención, también hay procesos que, siendo estacionarios en media no son estacionarios en varianza o

covarianza. En el Capítulo 2 se mencionaron varias posibilidades para tratar de convertir una serie no estacionaria en estacionaria, entre las cuales están estimar y restar la tendencia, diferenciar o tomar cambios porcentuales. En la literatura, *tendencia* es un movimiento de largo plazo en el proceso estocástico (o, en términos espectrales, de muy baja frecuencia), y se habla de dos tipos: la *determinística* y la *estocástica*.

Siendo algo tautológicos, la determinística es una tendencia que se puede estimar como una función exacta o determinística del tiempo t ; mientras que la estocástica contiene un término de error en su expresión, no se puede modelar como una función exacta de t y evoluciona lentamente sin rumbo predecible. Una implicación importante de la diferencia entre estos dos tipos de tendencia es que los *shocks* o innovaciones $\{\varepsilon_t\}$ tienen un impacto pasajero en la variable si la tendencia es determinística; pero tienen un impacto persistente si es estocástica ya que, en este último caso, los *shocks* pasan a formar parte de la tendencia.

Obviamente, también puede presentarse una combinación de ambos tipos de tendencia en un proceso estocástico. Las series con tendencia determinística suelen llamarse *estacionarias en tendencia* (*trend stationary*), y las de tendencia estocástica, *estacionarias en diferencia* (*difference stationary*). Las primeras son estacionarias alrededor de su tendencia y las segundas se tornan estacionarias al diferenciarlas, como se discutirá a continuación.

Paseo aleatorio

El modelo más sencillo que exhibe, o que es en sí mismo, una tendencia estocástica es el llamado *paseo aleatorio* (*random walk*). Este modelo consiste en hacer $\alpha = 1$ en la expresión del modelo autorregresivo AR(1) para obtener:

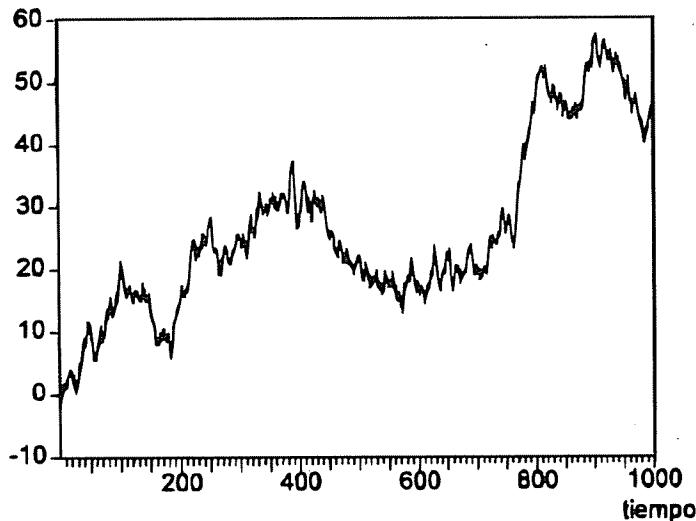
$$x_t = x_{t-1} + \varepsilon_t \quad (8.1)$$

donde ε_t es $iid(0, \sigma_\varepsilon^2)$.

De la expresión (8.1) es claro que el paseo aleatorio debe su nombre a que avanza dando pasos al azar. A partir del punto donde se encuentre el proceso en un momento dado, digamos x_{t-1} , dará un paso ε_t para llegar a x_t ; allí dará un paso ε_{t+1} para llegar a x_{t+1} , y así sucesivamente.

A partir de un x_0 inicial, el paseo aleatorio vagará sin rumbo y con muy pocos o ningún cruce por el eje cero. La Figura 8.1 muestra el comportamiento típico de un paseo aleatorio.

Figura 8.1: Paseo aleatorio



Si bien el paseo aleatorio no es estacionario (la raíz de su ecuación auxiliar $1 - L = 0$ es unitaria), debe notarse que su primera diferencia sí lo es, ya que resulta en una serie de ruido (la cual es estacionaria) según la expresión:

$$\Delta x_t = x_t - x_{t-1} = \varepsilon_t \quad (8.2)$$

La diferencia del paseo aleatorio también se puede escribir como $\Delta x_t = (x_{t-1} + \varepsilon_t) - (x_{t-2} + \varepsilon_{t-1}) = \Delta x_{t-1} + \Delta \varepsilon_t$, donde, por la forma de la expresión, podría pensarse que Δx_t es otro paseo aleatorio. Tal conclusión sería incorrecta por dos razones; porque dentro de la expresión anterior (y de [8.1]) tenemos que $x_{t-1} - x_{t-2} = \varepsilon_{t-1}$, por lo cual $\Delta x_t = \varepsilon_t$, que es estacionario, y porque $\Delta \varepsilon_t$ no es ruido, sino $\varepsilon_t - \varepsilon_{t-1}$, un promedio móvil MA(1).

Volviendo a la ecuación (8.1), se pueden sumar los ε_t ocurridos desde un x_0 inicial para llegar al valor de la serie en el tiempo t , de manera que podemos escribir:

$$\begin{aligned}x_1 &= x_0 + \varepsilon_1 \\x_2 &= x_0 + \varepsilon_1 + \varepsilon_2 \\x_3 &= x_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 \\x_4 &= x_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4 \\x_5 &= x_0 + \varepsilon_1 + \varepsilon_2 + \varepsilon_3 + \varepsilon_4 + \varepsilon_5 \\&\vdots \\x_t &= x_0 + \sum_{s=1}^t \varepsilon_s\end{aligned}\tag{8.3}$$

La expectativa $E x_t$ del paseo aleatorio (8.3), dado un número inicial x_0 , será el mismo x_0 para cualquier tiempo t ; esto es, la media será igual al punto de arranque, porque el movimiento a partir de allí es puramente aleatorio; es una suma de términos de ruido.

De igual forma, dado cualquier valor x_t , el movimiento futuro de la serie a partir de allí será aleatorio y dicho movimiento consistirá en una suma de términos de ruido:

$$x_{t+\tau} = x_t + \sum_{s=1}^{\tau} \varepsilon_{t+s}.\tag{8.4}$$

De manera que, tomando expectativa condicional de (8.4), tenemos:

$$E(x_{t+\tau} | x_t) = x_t\tag{8.5}$$

Esto es, la expectativa condicional dado x_t será idéntica a ese valor conocido. Lo anterior se puede generalizar anotando que $E(x_{t+\tau} | x_t, x_{t-1}, \dots, x_0) = x_t$, el último valor conocido de la serie.

Para encontrar la varianza del paseo aleatorio utilizamos (8.3), de donde:

$$var(x_t) = var\left[x_0 + \sum_{s=1}^t \varepsilon_s\right] = var(\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_t) = t\sigma_{\varepsilon}^2\tag{8.6}$$

La cual vemos que depende del tiempo. De manera similar encontramos la varianza en el tiempo $t - \tau$, dado x_0 :

$$\begin{aligned} \text{var}(x_{t-\tau}) &= \text{var}\left[x_0 + \sum_{s=1}^{t-\tau} \varepsilon_s\right] = \text{var}(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{t-\tau}) \\ &= (t - \tau)\sigma_\varepsilon^2 \end{aligned} \quad (8.7)$$

Mientras que la media condicional a x_0 del paseo aleatorio es constante, la varianza condicional a x_0 no lo es, sino que depende del tiempo, como puede verse de la expresión (8.6). Esto muestra que el paseo aleatorio es un proceso no estacionario, ya que su varianza cambia con el tiempo t . La razón intuitiva del crecimiento en la varianza es que a partir de x_0 la serie puede tomar cualquier rumbo, hacia arriba o hacia abajo, abriendo su rango de posibles valores como un abanico a medida que pasa el tiempo, de manera que en el momento t puede estar en cualquier parte de un rango mayor que en el momento $t - \tau$.

Para τ positivo, la autocovarianza del proceso será:

$$\begin{aligned} R(\tau) &= E(x_t - x_0)(x_{t-\tau} - x_0) \\ &= E[(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_t)(\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_{t-\tau})] \\ &= E(\varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_{t-\tau}^2) = (t - \tau)\sigma_\varepsilon^2 \end{aligned} \quad (8.8)$$

Donde, como se mencionó atrás, x_0 es su valor esperado. Lo anterior da igual a la varianza, $\text{var}(x_{t-\tau})$, en el tiempo $t - \tau$. Como la serie es no estacionaria en varianza, para encontrar la ACF debemos dividir la autocovarianza (8.8) por las desviaciones estándar en el tiempo t y en el tiempo $t - \tau$:

$$r(\tau) = \frac{(t - \tau)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2(t - \tau)\sigma_\varepsilon^2}} = \frac{t - \tau}{\sqrt{t(t - \tau)}} = \sqrt{\frac{t - \tau}{t}} \leq 1 \quad (8.9)$$

La expresión (8.9) indica que si t es grande (lo que equivale a una muestra grande), entonces las autocorrelaciones estarán cerca de uno durante muchos rezagos τ ; esto quiere decir que caerán muy lentamente. La función de autocorrelación parcial PACF para el paseo aleatorio es igual a la ACF para rezago 1, y 0 para los demás rezagos. Esto se puede intuir escribiendo un $\text{AR}(\tau)$ para x_t , pero remplazando x_t por su expresión (8.3).

Como es sabido (Capítulo 4), la aplicación de mínimos cuadrados ordinarios MCO a un AR(1) estacionario genera una estimación sesgada hacia cero en muestras finitas, aunque consistente y asintóticamente normal. Sin embargo, para $x_t = \alpha x_{t-1} + \varepsilon_t$ con $\alpha = 1$, como es el caso del paseo aleatorio, puede demostrarse que MCO produce un estimador sesgado hacia abajo en muestras finitas ($E\hat{\alpha} < 1$), pero donde la distribución de $(\hat{\alpha} - 1)/(error\ est\'andar)$ no es t de Student ni asintóticamente normal (Fuller, 1976; Greene, 1997, p. 848).

Bajo la hipótesis nula de $\alpha = 1$, la varianza de $\hat{\alpha}$ será menor que la usual [de orden $O(1/n^2)$ en lugar de $O(1/n)$], lo cual, aunado al sesgo $E\hat{\alpha} < 1$ en muestra finita, genera una tendencia a rechazar la hipótesis nula correcta $\alpha = 1$ usando, por ejemplo, la prueba t de Student. Para resolver este problema de estimación con raíz unitaria se utilizan tablas especiales de probabilidad, como se verá más adelante. De todas maneras el estimador MCO sigue siendo consistente; más aún, se dice que es *superconsistente* porque converge más rápido cuando $\alpha = 1$ que cuando $|\alpha| < 1$.

Si trabajamos con un τ negativo será necesario modificar el numerador de la expresión (8.9), esto es, la expresión (8.8), para obtener:

$$r(\tau) = \frac{t\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2(t-\tau)\sigma_\varepsilon^2}} = \frac{t}{\sqrt{t(t-\tau)}} = \sqrt{\frac{t}{t-\tau}} \leq 1$$

Lo cual, para t y τ dados, es mayor que (8.9), pero prácticamente igual a (8.9) si t es grande (no debe olvidarse que τ es negativo). Además de la varianza, la autocorrelación también aumenta en un paseo aleatorio a medida de avanza el tiempo, esto es, hacia adelante (Cryer, 1986, pp. 11-13), lo cual ocurre precisamente cuando τ es negativo y comparamos t contra $t - \tau$. Estrictamente, aquí no se cumple la propiedad de simetría, $r(\tau) = r(-\tau)$, como en los ARMA. En la práctica la autocorrelación estimada $\hat{r}(\tau)$ diferirá de la teórica $r(\tau)$ por los problemas de muestra finita; por ejemplo, \bar{x}_t difícilmente coincidirá con x_0 .

Una variación al paseo aleatorio (8.1) es el llamado *paseo aleatorio con corrimiento (drift)*, el cual sigue la expresión:

$$x_t = x_{t-1} + c + \varepsilon_t \quad (8.10)$$

Donde c es una constante. En muestras finitas no es fácil distinguir entre un paseo aleatorio con corrimiento y sin éste. Dado un x_0 inicial, podemos ver que se introduce una tendencia lineal:

$$x_1 = x_0 + c + \varepsilon_1$$

$$x_2 = x_0 + c + \varepsilon_1 + c + \varepsilon_2$$

$$x_3 = x_0 + c + \varepsilon_1 + c + \varepsilon_2 + c + \varepsilon_3$$

$$x_4 = x_0 + c + \varepsilon_1 + c + \varepsilon_2 + c + \varepsilon_3 + c + \varepsilon_4$$

$$x_5 = x_0 + c + \varepsilon_1 + c + \varepsilon_2 + c + \varepsilon_3 + c + \varepsilon_4 + c + \varepsilon_5$$

⋮

$$x_t = x_0 + ct + \sum_{s=1}^t \varepsilon_s$$

Su media será $E x_t = x_0 + ct$ y su varianza, $t\sigma_\varepsilon^2$. Si bien en este modelo x_t es una serie no estacionaria, aun en media, su primera diferencia $\Delta x_t = c + \varepsilon_t$ es estacionaria. Se obtiene el mismo resultado si en lugar de la constante en (8.10) suponemos que $E \varepsilon_t = c$; en este caso, la varianza se obtiene de:

$$\begin{aligned} var(x_t) &= E \left[x_0 + \sum_{s=1}^t \varepsilon_s - x_0 - ct \right]^2 = E \left(\sum_{s=1}^t \varepsilon_s \right)^2 - 2(ct)^2 + (ct)^2 \\ &= var \left(\sum_{s=1}^t \varepsilon_s \right) + \left(E \sum_{s=1}^t \varepsilon_s \right)^2 - (ct)^2 = t\sigma_\varepsilon^2 + (ct)^2 - (ct)^2 \\ &= t\sigma_\varepsilon^2 \end{aligned}$$

Volviendo a la ecuación del paseo aleatorio (8.1), ¿qué ocurre si $\alpha = -1$? En este caso se introduce una fuerte autocorrelación negativa en el modelo que hace menos útil y realista su comportamiento, aunque sigue siendo interesante para efectos ilustrativos. El modelo quedaría:

$$x_t = -x_{t-1} + \varepsilon_t. \quad (8.11)$$

Suponiendo un x_0 inicial tendremos:

$$\begin{aligned}
 x_1 &= -x_0 + \varepsilon_1 \\
 x_2 &= x_0 - \varepsilon_1 + \varepsilon_2 \\
 x_3 &= -x_0 + \varepsilon_1 - \varepsilon_2 + \varepsilon_3 \\
 x_4 &= x_0 - \varepsilon_1 + \varepsilon_2 - \varepsilon_3 + \varepsilon_4 \\
 x_5 &= -x_0 + \varepsilon_1 - \varepsilon_2 + \varepsilon_3 - \varepsilon_4 + \varepsilon_5 \\
 &\vdots \\
 x_t &= (-1)^t x_0 + \sum_{s=1}^t (-1)^{t+s} \varepsilon_s
 \end{aligned} \tag{8.12}$$

El valor esperado de (8.12) condicional a x_0 es $(-1)^t x_0$, de manera que la varianza de (8.12) será:

$$var(x_t) = E\left(\sum_{s=1}^t (-1)^{t+s} \varepsilon_s\right)^2 = t\sigma_\varepsilon^2 \tag{8.13}$$

Donde los signos negativos desaparecen porque los términos se elevan al cuadrado. La autocovarianza será:

$$\begin{aligned}
 R(\tau) &= E\left(\left(\sum_{s=1}^t (-1)^{t+s} \varepsilon_s\right)\left(\sum_{u=1}^{t-\tau} (-1)^{t-\tau+u} \varepsilon_u\right)\right) \\
 &= (-1)^\tau (t - \tau) \sigma_\varepsilon^2
 \end{aligned} \tag{8.14}$$

Donde los productos cruzados desaparecen y sólo los primeros $t - \tau$ términos de $\{\varepsilon_t\}$ podrán elevarse al cuadrado. Estos términos de $\{\varepsilon_t\}$ elevados al cuadrado corresponden a $s = u$ y tendrán signo negativo si τ es impar. Por ejemplo, para $t = 4$ y $\tau = 1$ tendremos de (8.14) que:

$$\begin{aligned}
 R(1) &= E\left(\sum_{s=1}^4 (-1)^{4+s} \varepsilon_s\right)\left(\sum_{u=1}^3 (-1)^{3+u} \varepsilon_u\right) \\
 &= E[(-\varepsilon_1 + \varepsilon_2 - \varepsilon_3 + \varepsilon_4)(\varepsilon_1 - \varepsilon_2 + \varepsilon_3)] \\
 &= -E\varepsilon_1^2 - E\varepsilon_2^2 - E\varepsilon_3^2 = -3\sigma_\varepsilon^2
 \end{aligned}$$

Siguiendo con este ejemplo, para $\tau = 2$ tendremos:

$$R(2) = E[(-\varepsilon_1 + \varepsilon_2 - \varepsilon_3 + \varepsilon_4)(-\varepsilon_1 + \varepsilon_2)] = E\varepsilon_1^2 + E\varepsilon_2^2 = 2\sigma_\varepsilon^2$$

La autocorrelación teórica $r(\tau)$ se obtiene dividiendo la autocovarianza (8.14) por las desviaciones estándar en t y en $t - \tau$, así:

$$r(\tau) = \frac{(-1)^\tau(t - \tau)\sigma_\varepsilon^2}{\sqrt{t\sigma_\varepsilon^2(t - \tau)\sigma_\varepsilon^2}} = (-1)^\tau \sqrt{\frac{t - \tau}{t}} \quad (8.15)$$

La cual, como puede observarse, es similar a la expresión (8.9) y también decrece lentamente en magnitud a medida que τ aumenta, pero alternando de signo.

Regresiones espurias

Supongamos dos paseos aleatorios independientes entre sí:

$$x_t = x_{t-1} + u_t \quad (8.16)$$

$$y_t = y_{t-1} + v_t \quad (8.17)$$

Donde:

$$u_t \sim iid(0, \sigma_u^2), \quad v_t \sim iid(0, \sigma_v^2), \quad y \quad Eu_t v_s = 0 \quad \forall t, s.$$

En otras palabras, x_t y y_t no están correlacionados.

Supongamos que corremos una regresión entre estas dos series x_t y y_t :

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \quad (8.18)$$

Obviamente, esperamos que, en promedio, el resultado de este experimento arroje β_1 igual a cero y un R^2 insignificante, ya que x_t y y_t son linealmente independientes por construcción. Podemos apelar al método de Monte Carlo, que consiste en realizar y tabular muchísimas repeticiones del experimento (de la regresión), construyendo x_t y y_t en cada repetición con nuevas realizaciones de u_t y de v_t tomadas de un generador de números aleatorios. Encontraremos que, con demasiada frecuencia, en el 75 % de los casos (véase Banerjee *et al.*, 1993, p. 74), ocurre lo inesperado, que $\hat{\beta}_1 \neq 0$ (al 5 % de significancia) y R^2 estimado sea alto en la regresión (8.18). En contraste, si x_t y y_t fueran procesos independientes estacionarios, encontraríamos que $plim \hat{\beta}_1 = 0$.

Puede demostrarse (Banerjee *et al.* 1993, capítulo 3) que MCO aplicado a (8.18) produce estimadores inconsistentes y altera las distribuciones de probabilidad de las pruebas t y F , de manera que más datos no arreglan el problema (véanse Phillips, 1986; Hamilton, 1994, capítulo 18); el estadístico t crece a una tasa de \sqrt{t} , por lo cual a medida que aumenta el tamaño de la muestra, es más fácil encontrar $\hat{\beta}_1 \neq 0$.

Básicamente, el problema es que, en una regresión típica $y = X\beta + \varepsilon$, donde las variables X son paseos aleatorios, $\text{plim} \frac{X'X}{n}$ no existe porque $X'X$ tiende a crecer más rápido que n , invalidando la normalidad asintótica o la consistencia. Se recordará del Capítulo 4 que $\text{plim} \frac{X'X}{n} = Q$ es requisito para el teorema de Mann-Wald. Este problema no es sólo para la estimación por MCO, sino que se puede extender a mínimos cuadrados no lineales, máxima verosimilitud, o método generalizado de momentos, que involucran matrices similares a $X'X$.

Es importante notar que, si en razón de la independencia lineal de los paseos aleatorios x_t y y_t cedemos a la tentación y hacemos $\beta_1 = 0$ en (8.18), y queda:

$$y_t = \beta_0 + \varepsilon_t, \quad (8.19)$$

estariamos diciendo que y_t es ruido blanco, lo cual también es falso. Entonces es falso que $\beta_1 \neq 0$, porque en (8.18) x_t y y_t son independientes, y también es falso que $\beta_1 = 0$ porque en (8.19) y_t no es ruido blanco.

En efecto, al correr la regresión lo que se estaría pidiendo al computador es que escoja entre dos modelos falsos: el modelo $y_t = \beta_0 + \varepsilon_t$ (donde y_t es ruido blanco) y el modelo $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$ (donde se acepta relación entre x_t y y_t). Como MCO minimiza la suma de estos errores estimados al cuadrado, no sorprende que con frecuencia prefiera (8.18) a (8.19). En estas expresiones puede intuirse que el error estimado $\hat{\varepsilon}_t$ no puede ser ruido blanco, ni siquiera estacionario, básicamente porque debe reflejar la parte no explicada de la variable y_t .

Estas relaciones se conocen con el nombre de *regresiones espurias*, porque suelen tener un R^2 satisfactorio, aunque no tengan sentido. También suelen acompañarse de muy bajos estadísticos de Durbin-Watson en razón de la autocorrelación positiva inducida en el error estimado $\hat{\varepsilon}_t$ por la omisión del rezago y_{t-1} (para mayores detalles

sobre este tema, véase Granger y Newbold, 1974). El modelo correcto está dado por (8.17), donde el rezago y_{t-1} explica y_t , de manera que las especificaciones (8.18) y (8.19) tienen el problema de omisión de variables al no contemplar el rezago y_{t-1} , lo cual hace inconsistentes los estimadores e induce autocorrelación positiva en ε_t .

La importancia de la discusión anterior radica en el hecho de que regresiones como la (8.18) entre series no estacionarias pueden ser comunes en econometría, por ejemplo, entre variables en niveles como precios, dinero, producto interno bruto (PIB), población, etc. Sin embargo, el problema de las regresiones espurias no cabe conceptualmente en los modelos estructurales, ya que las relaciones entre sus variables son dictadas directamente por la teoría; siempre y cuando, claro, dicha teoría se considere correcta. Tampoco se presenta en los vectores auto-regresivos (VAR), porque cada ecuación contiene rezagos de todas las variables, incluyendo los de la variable dependiente, siempre y cuando los errores sean estacionarios.

Una regresión espuria con variables de tendencia estocástica no se corrige apelando al método tradicional de estimar y restar una tendencia determinística a la serie; si la tendencia de la serie es un paseo aleatorio, restarle una tendencia determinística, especialmente la tendencia lineal usual, puede simplemente producir otro paseo aleatorio. Tampoco se soluciona incluyendo la variable t en la regresión (8.18), esto es:

$$y_t = \beta_0 + \beta_1 x_t + \gamma t + \varepsilon_t \quad (8.20)$$

Esto, porque, de acuerdo con el teorema de Frisch–Waugh (Banerjee *et al.*, 1993, p. 70), restar la tendencia temporal a cada serie antes de usarla en una regresión es equivalente a incluir el tiempo t como una de las variables explicativas. En la regresión (8.20) el coeficiente $\hat{\gamma}$ estimado tiende a cero ($\text{plim } \hat{\gamma} = 0$), pero β_1 no.

Sin embargo, es importante recalcar que las regresiones espurias también se pueden presentar entre variables con tendencia determinística; por ejemplo, Charemza y Deadman (1992, p. 124) muestran

cómo con las variables $y_t = \{1, 2, 3, \dots, 30\}$ y $x_t = \{1, 4, 9, \dots, 900\}$ se obtiene la siguiente regresión estimada:

$$\begin{aligned}y_t &= 5.92 + 0.03x_t \\&= (9.9) \quad (21.2)\end{aligned}$$

$$R^2 = 0.94 \quad DW = 0.06$$

Cuyos estadísticos t se muestran en paréntesis. Con este tipo de variables, incluir la tendencia determinística como variable explicativa sí puede corregir el problema de regresión espuria.

Nelson y Kang (1984) hicieron un experimento similar corriendo paseos aleatorios (generados por Monte Carlo) con $n = 100$ contra el tiempo, esto es, corrieron:

$$x_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Y así encontraron que β_1 era significativo en el 87% de los casos. Según esto, x_t podría predecirse a partir de t , lo cual es espurio. El experimento se repitió con $n = 20$ y se halló que β_1 aún era significativo en el 73% de los casos. Se concluiría falsamente que x_t tiene una tendencia lineal.

En el fondo, el problema de la regresión espuria se presenta cuando se especifica mal el modelo y tratamos de juzgar la significancia de los parámetros estimados con las pruebas tradicionales del modelo lineal clásico (estadísticos t , R^2 , etc.) cuando éstas no son válidas. Cuando en una regresión con variables no estacionarias sea imposible estimar los coeficientes de manera que el error sea estacionario, existe la posibilidad de regresión espuria. Por ejemplo, en (8.18) el error de la regresión no cumple con los supuestos del modelo, porque no hay combinación de β_0 y β_1 que produzca un error estacionario.

Incluso podría presentarse una regresión espuria entre variables estacionarias si el error no es ruido. Davidson y MacKinnon (2004, capítulo 14) presentan un estudio de Monte Carlo con dos AR(1) estacionarios, donde, no obstante ser independientes entre sí, la regresión al estilo (8.18) da significativa más veces de lo esperado (aunque no tanto como en la regresión con series no estacionarias). En este caso, al ser β_1 igual a cero, el residuo ε_t en (8.19) es un AR(1) y queda

autocorrelacionado e invalida la prueba t . En resumen, el problema de la regresión espuria se resuelve mejorando la especificación, usualmente agregando a la regresión rezagos de las variables.

Series integradas

Una serie no estacionaria x_t que pueda convertirse en un ARMA estacionario diferenciándola d veces se define como integrada de orden d y se denota $x_t \sim I(d)$. (Granger y Newbold, 1986).

Una serie no estacionaria será $I(1)$ si al diferenciarla una vez se vuelve estacionaria. Para que sea $I(1)$ no es necesario que crezca de manera aproximadamente lineal, como a veces se cree; el paseo aleatorio vaga sin rumbo y es $I(1)$. Una serie no estacionaria será $I(2)$ si al diferenciarla dos veces se vuelve estacionaria. Una serie $I(0)$ es una serie que no hay que diferenciarla, porque ya es estacionaria y tiene representación ARMA. La suma de una $I(0)$ y una $I(1)$ será siempre una $I(1)$.

De la experiencia se encuentra que la mayoría de las series económicas son hasta $I(2)$. La mayor parte de los trabajos empíricos tratan con series $I(0)$ o $I(1)$, razón por la cual nos concentraremos en éstas en adelante.

El paseo aleatorio es una típica $I(1)$, ya que al diferenciarla una vez queda estacionaria:

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t \\ (1 - L)x_t &= \varepsilon_t \sim I(0) \end{aligned} \tag{8.21}$$

Vemos de (8.21) que el paseo aleatorio tiene una raíz unitaria (de la ecuación auxiliar $1 - L = 0$). Diferenciar una serie es aplicarle el operador $(1 - L)$ que, de hecho, tiene una raíz unitaria. Supongamos que podemos escribir:

$$x_t = \frac{B(L)\varepsilon_t}{(1 - L)\Phi(L)} \tag{8.22}$$

donde $\frac{B(L)\varepsilon_t}{\Phi(L)}$ es un ARMA estacionario. Al diferenciar una vez se tiene:

$$\Delta x_t = x_t - x_{t-1} = \frac{B(L)\varepsilon_t}{(1 - L)\Phi(L)} - \frac{LB(L)\varepsilon_t}{(1 - L)\Phi(L)} = \frac{B(L)\varepsilon_t}{\Phi(L)} \tag{8.23}$$

expresión que es estacionaria. De manera similar, la expresión:

$$x_t = \frac{B(L)\varepsilon_t}{(1-L)^d\Phi(L)} \quad (8.24)$$

donde $\frac{B(L)\varepsilon_t}{\Phi(L)}$ es un ARMA estacionario, habría que diferenciarla d veces para que sea estacionaria.

De hecho, se puede considerar una primera diferencia de la forma $1 - L^s$, donde s es un entero positivo, por ejemplo, al diferenciar una serie estacional. El proceso:

$$x_t = \frac{B(L)\varepsilon_t}{(1 - L^s)\Phi(L)}$$

podría volverse estacionario al aplicar la diferencia:

$$\Delta^s x_t = x_t - x_{t-s} = \frac{B(L)\varepsilon_t}{(1 - L^s)\Phi(L)} - \frac{L^s B(L)\varepsilon_t}{(1 - L^s)\Phi(L)} = \frac{B(L)\varepsilon_t}{\Phi(L)}$$

La definición de serie integrada está en el contexto de diferenciar para obtener estacionariedad, más que en términos de raíz unitaria. Por ejemplo, el proceso $x_t = -x_{t-1} + \varepsilon_t$ contiene una raíz unitaria, $L = -1$, pero no se convierte en estacionario al diferenciarlo (para esto habría que sumar $x_t + x_{t-1} = \varepsilon_t$).

Ejemplo

Muestre que $x_t = 2x_{t-1} - x_{t-2} + \varepsilon_t$ es $I(2)$. [Notamos que la expresión anterior es igual a $(1 - L)(1 - L)x_t = \varepsilon_t$]. Para que una serie sea $I(2)$ debe convertirse en estacionaria después de diferenciarla dos veces. La primera diferencia es:

$$\Delta x_t = 2x_{t-1} - x_{t-2} + \varepsilon_t - x_{t-1} = x_{t-1} - x_{t-2} + \varepsilon_t = \Delta x_{t-1} + \varepsilon_t$$

de donde vemos que la segunda diferencia es:

$$\Delta\Delta x_t = \Delta x_t - \Delta x_{t-1} = \varepsilon_t$$

lo cual es $I(0)$.

Ejemplo

Aunque a primera vista parezca lo contrario, el modelo $x_t = x_{t-1} + \varepsilon_t - \varepsilon_{t-1}$, donde ε_t es ruido blanco, no requiere diferenciación, porque ya es estacionario. Al reescribirlo tenemos:

$$x_t = \frac{(1-L)\varepsilon_t}{1-L} = \varepsilon_t.$$

La anterior es una solución estacionaria o particular. La estacionariedad de este modelo también se puede ver haciendo reemplazos sucesivos, así:

$$\begin{aligned} x_t &= x_{t-2} + \varepsilon_{t-1} - \varepsilon_{t-2} + \varepsilon_t - \varepsilon_{t-1} = x_{t-2} - \varepsilon_{t-2} + \varepsilon_t \\ &= x_{t-3} + \varepsilon_{t-2} - \varepsilon_{t-3} + \varepsilon_t - \varepsilon_{t-2} = x_{t-3} - \varepsilon_{t-3} + \varepsilon_t \\ &= x_{t-4} + \varepsilon_{t-3} - \varepsilon_{t-4} + \varepsilon_t - \varepsilon_{t-3} = x_{t-4} - \varepsilon_{t-4} + \varepsilon_t \\ &\vdots \\ &= x_{t-n} - \varepsilon_{t-n} + \varepsilon_t \\ &\vdots \\ &= x_0 - \varepsilon_0 + \varepsilon_t \end{aligned}$$

Expresión que es estacionaria por ser igual a una constante ($x_0 - \varepsilon_0$) más ruido (ε_t), y es la solución general. Lo anterior contrasta con un paseo aleatorio:

$$\begin{aligned} x_t &= x_{t-1} + \varepsilon_t \\ &\vdots \\ &= x_0 + \sum_{s=1}^t \varepsilon_s \end{aligned}$$

el cual, como vimos, no es estacionario.

El modelo AR(1) con $\alpha > 1$

La serie AR(1)

$$x_t = \alpha x_{t-1} + \varepsilon_t, \quad (8.25)$$

donde $\alpha > 1$ no es estacionaria, ya que tiene una tendencia de crecimiento exponencial. Por ejemplo, si $\alpha = 1.05$, la serie tiende a crecer al 5 % por periodo.

La expresión general para el valor esperado de x_t , partiendo de un valor inicial x_0 , se encuentra haciendo las recursiones:

$$\begin{aligned}x_1 &= \alpha x_0 + \varepsilon_1 \\x_2 &= \alpha(\alpha x_0 + \varepsilon_1) + \varepsilon_2 = \alpha^2 x_0 + \alpha \varepsilon_1 + \varepsilon_2 \\&\vdots \\x_t &= \alpha^t x_0 + \alpha^{t-1} \varepsilon_1 + \alpha^{t-2} \varepsilon_2 + \cdots + \alpha \varepsilon_{t-1} + \varepsilon_t\end{aligned}$$

Y aplicando valor esperado a esta última expresión para obtener:

$$Ex_t = \alpha^t x_0$$

Si tomamos primeras diferencias de (8.25), tendremos:

$$\Delta x_t = x_t - x_{t-1} = \alpha x_{t-1} - x_{t-1} + \varepsilon_t = (\alpha - 1)x_{t-1} + \varepsilon_t \quad (8.26)$$

Donde vemos que la nueva serie Δx_t tampoco es estacionaria, porque será una fracción fija, $\alpha - 1 > 0$, del nivel creciente de x_{t-1} . Con $\alpha > 1$ podemos diferenciar varias veces y, en principio, la serie no se convierte en estacionaria. Sin embargo, puede ocurrir que las diferencias alcancen una magnitud pequeña comparadas con el ruido ε_t , de manera que la segunda o tercera diferenciación parecerá estacionaria en una muestra finita; o, en el lenguaje de teoría de la información, la relación señal a ruido cae.

A veces se sugiere que para convertir la serie (8.25) en estacionaria en tendencia se tomen cambios porcentuales; esto es, dividimos (8.26) por x_{t-1} :

$$\frac{\Delta x_t}{x_{t-1}} = (\alpha - 1) + \frac{\varepsilon_t}{x_{t-1}}. \quad (8.27)$$

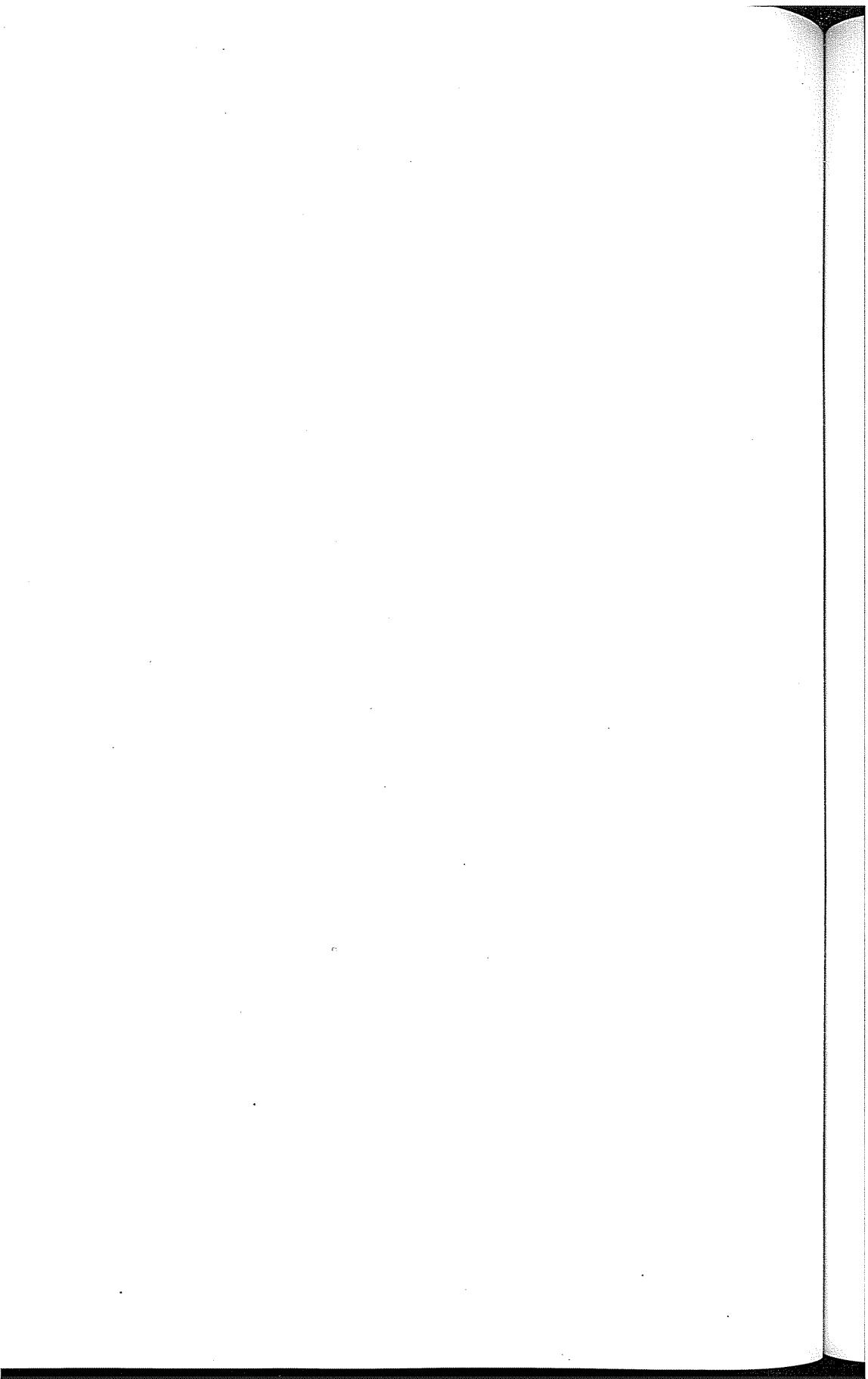
donde se observa que la media, $E\left(\frac{\Delta x_t}{x_{t-1}}\right) = \alpha - 1$, es ahora una constante independiente del tiempo. Sin embargo, de todas maneras la serie no será estacionaria en varianza, porque dado que σ_ε^2 se supone finita:

$$var\left(\frac{\Delta x_t}{x_{t-1}}\right) = var\left(\frac{\varepsilon_t}{x_{t-1}}\right) = E\left(\frac{\varepsilon_t}{x_{t-1}}\right)^2 = \frac{E\varepsilon_t^2}{Ex_{t-1}^2} = \frac{\sigma_\varepsilon^2}{Ex_{t-1}^2}$$

Esta última depende del tiempo ya que x_{t-1} crece, y:

$$\text{plim var}\left(\frac{\Delta x_t}{x_{t-1}}\right) = 0.$$

Ahora, si suponemos que σ_ϵ^2 también crece con el nivel de x_{t-1} , como podría parecer adecuado en muchos casos de la vida real, entonces los cambios porcentuales podrían arrojar estacionariedad en varianza. Muchas series macroeconómicas no estacionarias pueden presentarse como un proceso con crecimiento porcentual (aunque no necesariamente constante en el tiempo), más un término de error que crece con el tiempo (en particular con el nivel de la serie).

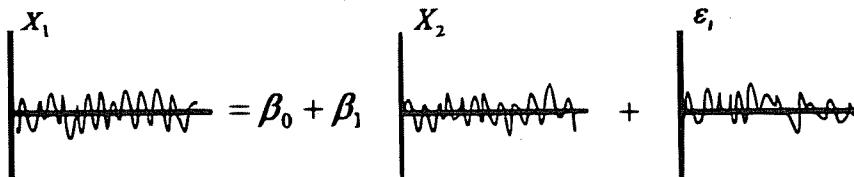


CAPÍTULO 9

Raíces unitarias y cointegración bivariada

Es común en econometría correr regresiones con series de tiempo sin indagar primero si las variables involucradas son compatibles entre sí en una relación lineal. El punto es que la linealidad impone restricciones al comportamiento de las variables a través del tiempo. Es posible que físicamente no se pueda conformar la regresión; esto es, en muchos casos la combinación lineal de las variables explicativas, incluido el error (y el supuesto de homoscedasticidad acerca del error), simplemente no puede producir el tipo de comportamiento que tiene la variable dependiente.

Figura 9.1: Regresión factible

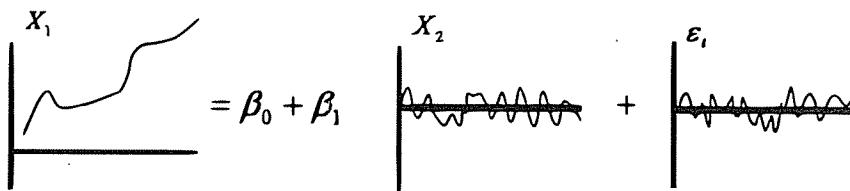


Por ejemplo, en la Figura 9.1 se ilustra una regresión factible donde se expresa una variable dependiente estacionaria x_{1t} en función de una constante β_0 y de dos series estacionarias que son la variable

explicativa x_{2t} y el error ε_t . Esta regresión es admisible, porque es físicamente posible que series estacionarias sumadas a una constante puedan generar otra serie estacionaria.

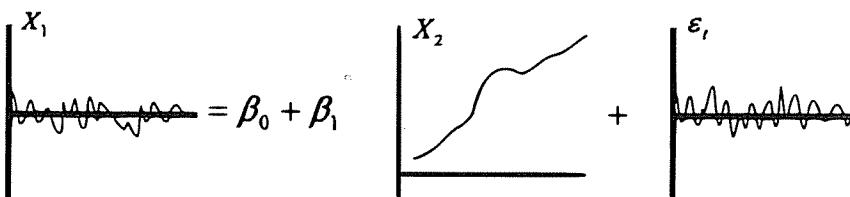
En contraste, la Figura 9.2 muestra una regresión inadmisible, donde hay una variable dependiente no estacionaria y una variable explicativa estacionaria. No hay forma posible de que una constante β_0 sumada a una serie estacionaria, multiplicada ésta por cualquier coeficiente β_1 constante, genere una serie no estacionaria como la ilustrada. En esta regresión el error estimado no sería ruido, sino que tendría un comportamiento no estacionario similar al de la variable dependiente.

Figura 9.2: Regresión inadmisible



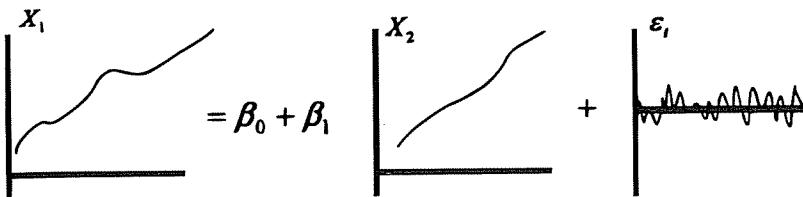
La Figura 9.3 muestra una regresión inútil, ya que es imposible que la variable explicativa no estacionaria pueda explicar la variable dependiente estacionaria.

Figura 9.3: Regresión inútil



La Figura 9.4 muestra un tipo de relación lineal muy común en econometría, donde proponemos una $I(1)$ para tratar de explicar otra $I(1)$. Esta regresión podría ser físicamente posible si ambas variables tienen un comportamiento similar a través del tiempo (tendencia común); en caso contrario, la regresión será espuria y la serie del error ε_t tendrá que recoger la divergencia creciente entre x_{1t} y x_{2t} , haciendo que el error sea no estacionario.

Figura 9.4: Relación lineal



Cointegración bivariada

La regresión de la Figura 9.4 es de mucho interés. Es admisible sólo si existe algún mecanismo que mantenga las dos variables, x_{1t} y x_{2t} , cerca la una de la otra a través del tiempo, esto es, que no permita que se separen demasiado aunque ambas individualmente sean no estacionarias. Debe existir algún *mecanismo de equilibrio de largo plazo* entre las dos. Aunque en el corto plazo las series pueden diferir un poco, algo las obliga a acercarse en caso de que se estén alejando demasiado, de manera que su diferencia sea estacionaria.

Variables que pueden estar atadas por este tipo de relación de equilibrio son, por ejemplo, los precios de un producto agrícola en zonas geográficas cercanas, las tasas de interés de corto y largo plazo, los precios de una acción en diferentes bolsas de valores del mismo país, los ingresos y los gastos del hogar. Es importante, desde el punto de vista económico, poder descubrir y estimar estas relaciones de equilibrio que rigen los movimientos en los mercados.

En otras palabras, *aunque las series no sean estacionarias, su combinación lineal sí lo es*. Formalicemos estos conceptos utilizando la terminología de series integradas y concentrándonos, por ahora, en el caso bivariado con variables $I(1)$.

Podemos decir que cuando existe un mecanismo de equilibrio, la combinación lineal de dos variables no estacionarias puede producir una estacionaria $I(0)$. Cuando esto ocurre se llama *cointegración*, y se dice que las dos variables no estacionarias están cointegradas. En general, para que haya cointegración entre dos variables, éstas deben tener el mismo orden de integración; de otra forma, una cambia más

rápido que la otra y será imposible que su combinación lineal dé un $I(0)$.

Definición: dos variables x_{1t} y x_{2t} están cointegradas si ambas son $I(1)$ y se puede escribir:

$$x_{1t} - ax_{2t} = z_t \quad (9.1)$$

Donde $a \neq 0$ es una constante y z_t es $I(0)$.

Aquí z_t es como el residuo de la regresión entre las dos variables. Sin embargo, es crucial resaltar que si bien z_t debe ser $I(0)$ para que haya equilibrio, *esto no implica que deba ser ruido blanco*.

Una manera sencilla de investigar la cointegración es correr una regresión de x_{1t} contra x_{2t} como en la Figura 9.4. Si la variable z_t , que es el residuo (los errores) de esta regresión, es $I(0)$, se dice que hay cointegración. Los errores z_t serán estacionarios sólo si x_{1t} y x_{2t} están cointegradas. Si x_{1t} y x_{2t} no están cointegradas, entonces z_t no será estacionaria sino que, por el contrario, también será $I(1)$.

El procedimiento para estudiar la cointegración consiste, primero, en determinar si las series individualmente vistas son $I(1)$; segundo, en examinar si hay o no hay cointegración, y, tercero, en caso afirmativo, incorporar la cointegración en las ecuaciones de regresión. Las pruebas estadísticas para determinar el grado de integración de una serie, en particular si son $I(1)$, se conocen como pruebas de raíz unitaria.

Pruebas de raíz unitaria

Lo que buscamos establecer con estas pruebas es si una serie es estacionaria, esto es, $I(0)$, o no lo es. Se recomienda graficar primero la serie en cuestión para tener una idea de su comportamiento. Según se discutió antes, en magnitud una serie estacionaria tiende a quedarse dentro de ciertos límites superior e inferior como se muestra en la Figura 9.5. En contraste, la Figura 9.6 muestra una serie no estacionaria. Al observarlas, inferimos que hay algo que tiende a mantener una serie estacionaria dentro de sus límites (líneas punteadas); no así en el caso de la no estacionaria.

Intuitivamente, cuando una serie estacionaria se acerca a su límite superior o inferior, debe haber algún mecanismo que tienda a devolverla, haciendo que el próximo *movimiento* o *cambio* sea hacia su media;

Figura 9.5: Serie estacionaria sin raíz unitaria

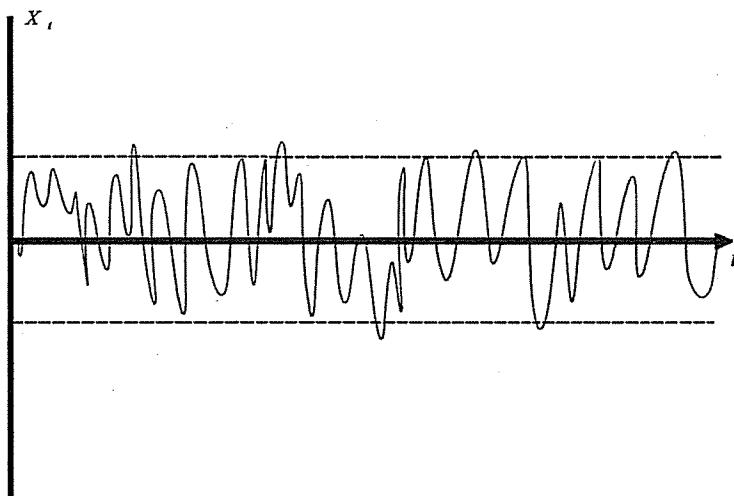


Figura 9.6: Serie no estacionaria con raíz unitaria

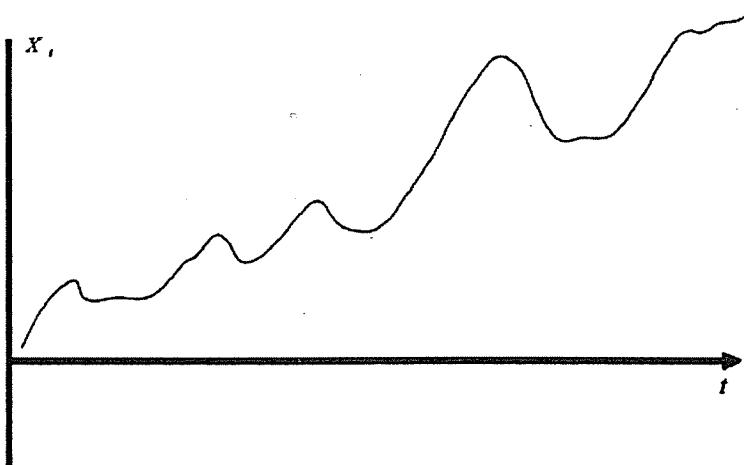


Figura 9.5: Serie estacionaria sin raíz unitaria

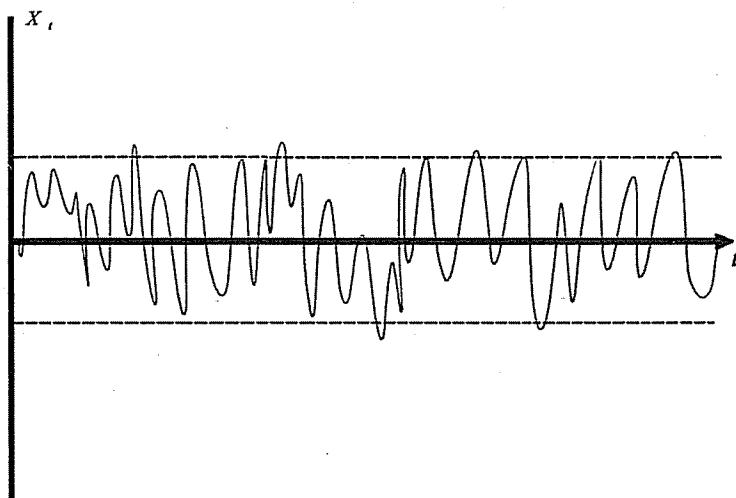
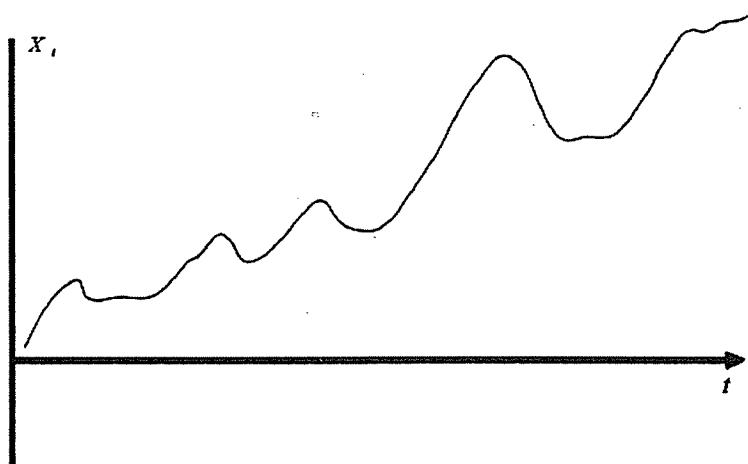


Figura 9.6: Serie no estacionaria con raíz unitaria



esta reversión a la media será más fuerte cuanto más lejos esté la serie de la media. Pruebas de raíz unitaria, como la Dickey-Fuller y la Phillips-Perron, incorporan este principio de reversión a la media.

Prueba de Dickey-Fuller aumentada

Para realizar la prueba de Dickey-Fuller aumentada (ADF) a una serie x_t se corre por mínimos cuadrados ordinarios MCO la siguiente regresión;

$$\Delta x_t = \varphi x_{t-1} + \sum \gamma_i \Delta x_{t-i} + \varepsilon_t \quad (9.2)$$

donde el número de rezagos de Δx_t puede escogerse corriendo (9.2) inicialmente con muchos rezagos y luego disminuyéndolos hasta que el último sea significativo (véase Enders, 2004, capítulo 4) y aseguremos que ε_t sea ruido blanco, según la función de autocorrelación ACF, la prueba Q , etc. También podría usarse el coeficiente Durbin-Watson DW, pero con cautela. En razón de que el DW se sesga hacia 2 cuando se usan variables dependientes rezagadas como explicativas, si este estadístico es muy pequeño, podemos decir que la autocorrelación debe ser notoria para sacarlo de su sesgo y disminuirlo. En otras palabras, si el DW es pequeño, entonces hay indicios de autocorrelación, y si es cercano a 2 no quiere decir que no la haya.

Supongamos que el verdadero modelo es un AR(p), $x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$, pero que, en su lugar, corremos un AR(1), $x_t = c + \alpha_1 x_{t-1} + u_t$. Encontramos que el error del AR(1), $u_t = \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t$, tendrá autocorrelación. Esta es la razón por la cual agregar rezagos de la variable dependiente es usualmente la primera recomendación para disminuir la autocorrelación en los errores.

Más al caso, supongamos que el verdadero modelo es un AR(2), $x_t = c + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \varepsilon_t$. Podemos escribirlo como $x_t = c + (\alpha_1 + \alpha_2)x_{t-1} - \alpha_2(x_{t-1} - x_{t-2}) + \varepsilon_t$, de donde, restando x_{t-1} a ambos lados, obtenemos la prueba ADF $\Delta x_t = c + \varphi x_{t-1} + \gamma \Delta x_{t-1} + \varepsilon_t$. Vemos que si el modelo es un AR(2), debe incluirse un rezago de Δx_t en la ecuación de la prueba ADF. Y, en general, si la variable x_t se ajusta a un AR(p), deben incluirse $p - 1$ rezagos en la ecuación de la prueba ADF.

En (9.2) $\varphi = 0$ implica que Δx_t no depende del nivel anterior de la serie, de manera que nada tiende a devolver la serie cuando se aleja de su media; y en este caso x_t será $I(1)$. Para que x_t sea $I(0)$, se espera que $\varphi < 0$ de manera que tienda a devolver la serie si ésta se aleja demasiado por arriba o por abajo.

Como se mencionó en el capítulo anterior, las pruebas de hipótesis usuales no siempre funcionan cuando involucran raíces unitarias, porque las distribuciones de probabilidad pueden no ser asintóticamente normales o simétricas (véase Banerjee *et al.*, 1993, p. 100). En otras palabras, las tablas usuales de la prueba t no serían apropiadas para probar hipótesis sobre φ . En su lugar, se utilizan tablas especiales para el estadístico $\hat{\varphi}/(\text{error estándar})$. Si este estadístico es menor que el valor crítico (esto es, más negativo que el valor crítico), se acepta $\varphi \neq 0$; de lo contrario $\varphi = 0$ y la serie no dependerá de su nivel anterior. Si el estadístico es mayor que cero, entonces el proceso es aún más explosivo. Consideremos el AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$. Restando x_{t-1} a ambos lados tenemos:

$$\Delta x_t = (\alpha - 1)x_{t-1} + \varepsilon_t = \varphi x_{t-1} + \varepsilon_t$$

Nótese que, en esencia, este es el meollo de la prueba de Dickey-Fuller, y constituye una versión simple de ésta. La prueba mostrada antes en (9.2) se conoce como "aumentada", porque incluye rezagos de Δx_t , cuyo objetivo es mejorar la aproximación de los residuos al ruido blanco. En cualquier caso, si $\varphi = 0$ concluimos que x_t es $I(1)$.

A la ecuación (9.2) también se le puede agregar una constante (drift) o una tendencia determinística βt si se desea (φx_{t-1} ya representa la tendencia estocástica), pero hay que usar tablas apropiadas para estos casos; afortunadamente, la mayoría de los programas de computador ya las incorporan automáticamente. La gráfica de x_t sugerirá si conviene incluir la constante, la tendencia o ninguna. Si x_t tiene la expresión $x_t = c + \alpha x_{t-1} + \varepsilon_t$, la prueba de Dickey-Fuller será de la forma $\Delta x_t = c + (\alpha - 1)x_{t-1} + \varepsilon_t$ y conviene incluir la constante; o si x_t tiene la expresión $x_t = c + \beta t + \alpha x_{t-1} + \varepsilon_t$, la prueba tendrá la forma $\Delta x_t = c + \beta t + (\alpha - 1)x_{t-1} + \varepsilon_t$ y conviene adicionar la tendencia.

La adición de regresores determinísticos, como la constante o la tendencia, tiende a hacer más negativos los valores críticos de la distribución, y a disminuir el poder estadístico de la prueba (esto es, se

reduce la probabilidad de rechazar H_0 cuando es falsa). Una alternativa que parece mejorar el poder estadístico en presencia de regresores determinísticos, conocida como ADF-GLS (véase Elliott, Rothenberg y Stock, 1996), estima y remueve el efecto de los regresores determinísticos antes de realizar la prueba ADF. El nombre GLS se da porque las variables, dependientes e independientes, se transforman de manera similar a como se hace en mínimos cuadrados generalizados, antes de correr la regresión.

Un resultado general de Sims, Stock y Watson (1990), dice que los estadísticos t y F tienen distribuciones asintóticas t y F si las pruebas se realizan sobre coeficientes de variables $I(0)$ con media cero (simultáneamente $I(0)$ en el caso de la F). Por ejemplo, en la ADF de la ecuación (9.2), las pruebas t y F sobre los γ_i son válidas asintóticamente; no así la t sobre φ , que, como se mencionó, requiere valores críticos especiales.

Prueba de Phillips-Perron

La prueba de Phillips-Perron (PP) es una variante de la prueba de Dickey-Fuller en la cual se permite (o se acepta) que los errores ε_t no sean exactamente ruido blanco. Para efectuar la prueba se requiere una estimación del espectro del error ε_t (*i.e.* la estimación de las autocorrelaciones del error). Para la prueba de PP se corre la de Dickey-Fuller simple:

$$\Delta x_t = \varphi x_{t-1} + \varepsilon_t \quad (9.3)$$

Donde ε_t puede estar autocorrelacionado. El estadístico de esta prueba no es $\hat{\varphi}/(\text{error estándar})$, como en la ADF, sino que tiene otra expresión en la cual $\hat{\varphi}/(\text{error estándar})$ es ajustado por el efecto de la autocorrelación de ε_t . Para determinar este efecto se involucran las autocorrelaciones estimadas de ε_t , ponderadas a través de una ventana (*kernel*) que cubra un número de rezagos suministrado por el usuario. No obstante, a veces, dependiendo del método (Newey-West, por ejemplo), el programa de computador puede sugerir el número de rezagos de truncamiento. Asintóticamente los valores críticos son los mismos que los de la ADF. También se le puede agregar la constante (*drift*) y la tendencia determinística β_t .

Práctica y problemas de las pruebas de raíz unitaria

En general, se recomienda recurrir a varias pruebas de raíz unitaria, como la ADF y la PP, para determinar la estacionariedad de las series. Obviamente, pueden dar resultados contradictorios o inconclusos y quedar al criterio del analista cómo dirimirlos. La discusión que se hace a continuación puede ayudar en este sentido.

A fin de establecer si hay más de una raíz unitaria, algunos autores sugieren pruebas secuenciales. Si se encuentra raíz unitaria en la serie original, pero no en la serie en diferencias, se puede concluir que la serie tiene una sola raíz unitaria, y es $I(1)$. Si la serie en primeras diferencias tiene raíz unitaria, mas no en segundas diferencias, se puede concluir que tiene dos raíces unitarias, y que es $I(2)$.

Sin embargo, algunos autores como Banerjee *et al.* (1993, p. 119) o Maddala y Kim (1998, capítulo 11) sostienen que probar la raíz unitaria en niveles (sin diferenciar) y luego probar por una segunda raíz unitaria en la serie diferenciada, es un procedimiento que tiende a subestimar el número de raíces unitarias, por lo cual debe ser tomado como aproximación. Enders (2004, capítulo 4) sugiere una alternativa secuencial para realizar estas pruebas.

Debemos precisar un poco más cuál es la hipótesis nula de estas pruebas. Por la forma como se deriva la expresión de Dickey-Fuller, parecería ser que $H_0: \varphi = 0$, esto es, que sí hay raíz unitaria. Sin embargo, desde el punto de vista heurístico, lo que detecta la prueba sobre $\varphi = 0$ (o aun $\varphi \geq 0$) en la expresión de Dickey-Fuller es simplemente si la serie carece de una tendencia a regresar a su media. De manera que podríamos decir:

H_0 : No estacionariedad en media.

H_A : Estacionariedad en media.

Vista así, la hipótesis nula H_0 es simplemente que la serie no es estacionaria en media, lo cual puede suceder por una o más raíces unitarias o por algún otro tipo de no estacionariedad que ni siquiera tenga que ver con raíces unitarias, como un cambio estructural en el sistema. En general, estas pruebas no detectan cualquier tipo de no

estacionariedad, por ejemplo, la de la ecuación $1 + L$, que implica la raíz -1 .

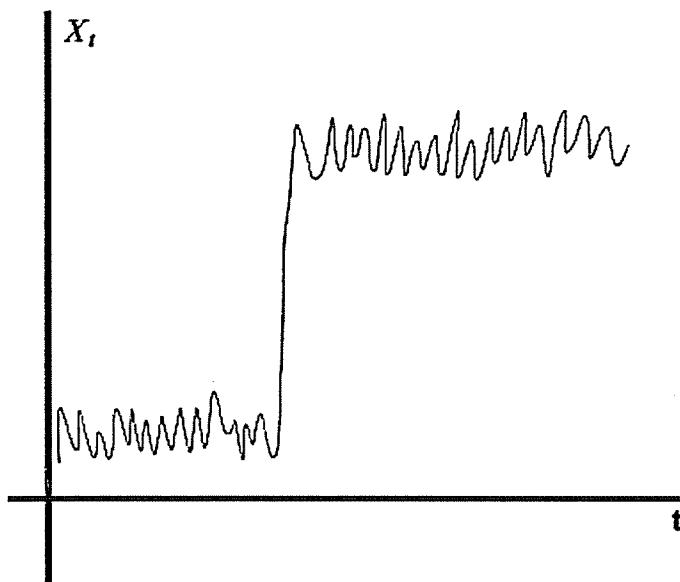
Las pruebas de raíz unitaria tienen varios problemas: uno, es que tienen bajo poder estadístico [1-probabilidad de error tipo II], esto es, baja probabilidad de rechazar H_0 cuando es falsa, y tienen dificultad para poder distinguir entre un proceso de raíz unitaria y uno estacionario con raíz cercana a 1; esto es, en muestras finitas una serie estacionaria en tendencia puede confundirse con una serie de raíz unitaria y viceversa. La consecuencia es que la prueba acepta con demasiada (e incorrecta) frecuencia que la serie tiene raíz unitaria; esto es, tiende a aceptar la hipótesis falsa $\varphi = 0$ con demasiada frecuencia. Por lo anterior, en casos de indefinición de la prueba, se recomienda al analista considerar el rechazo de la hipótesis. Según Kennedy (1992, p. 267), el poder estadístico depende más de la longitud del periodo estudiado que del número de datos.

Otro problema que se presenta en estas pruebas es que una serie estacionaria, excepto por un cambio estructural como en la Figura 9.7, puede parecer como si tuviera una raíz unitaria. La prueba indicaría raíz unitaria cuando lo que hay que modelar es el cambio estructural en la economía; una apertura comercial, un cambio en el régimen político, etc. Siempre se recomienda graficar la serie y conocer algo de la historia pertinente. Perron (1989) diseñó una prueba para detectar raíz unitaria cuando hay cambio estructural pero requiere conocer la fecha exacta de tal cambio; entre tanto, Perron y Vogelsang (1992) modificaron esta prueba para cuando la fecha del cambio estructural sea incierta.

También podría presentarse cambio estructural junto con un proceso de raíz unitaria o con uno de tendencia determinística. En este último caso puede pensarse que la serie tiene tendencia estocástica en lugar de modelarla correctamente como una tendencia determinística. Por otro lado, Leybourne, Hills y Newbold (1998) muestran que si el verdadero proceso es $I(1)$, pero con un quiebre temprano en la serie, entonces las pruebas tienden a rechazar la hipótesis de raíz unitaria.

Ocasionalmente, el analista puede separar la muestra en dos, antes del cambio y después del cambio, y aplicar la prueba en ambas submuestras para ver si el comportamiento permanece inalterado de un régimen a otro. El problema es que el tamaño (grados de libertad)

Figura 9.7: Serie estacionaria con cambio estructural



de cada submuestra puede ser insuficiente para llegar a conclusión alguna; en otras palabras, la reducción de grados de libertad disminuye rápidamente el poder estadístico de la prueba.

Cuando sea posible, una forma de aumentar el tamaño de la muestra y el poder estadístico de las pruebas es utilizando datos panel (corte transversal a través del tiempo). Maddala y Kim (1998) discuten diversas pruebas de raíz unitaria para datos panel; véase también Baltagi (2005).

Otra dificultad es que el resultado de la prueba (las tablas de valores críticos) depende crucialmente de si se incluye constante o tendencia en la ecuación de ADF o PP. Si la prueba se especifica mal, su poder estadístico puede caer a casi cero. Se recomienda al analista observar el comportamiento gráfico de la serie para decidir entre estas opciones. De nuevo, en caso de indefinición se recomienda considerar el rechazo de la hipótesis $H_0: \varphi = 0$.

Granger y Swanson (1997) diseñaron un modelo llamado *Stochastic Unit Root* (STUR), el cual corresponde a un proceso cuyas raíces no son constantes, sino estocásticas y varían alrededor de la unidad. Estos procesos pueden ser estacionarios en algunos períodos y no estacio-

narios en otros, dependiendo del valor que vayan tomando sus raíces. Obviamente, las pruebas usuales de raíz unitaria tienen dificultades para detectar este tipo de modelos.

Ejemplo

Los cuadros 9.1 y 9.2 muestran la aplicación de las pruebas ADF y PP, respectivamente, a la serie de la tasa porcentual de desempleo trimestral para el periodo 1984-2000 mostrada en la Figura 9.8. En ninguna de las dos el estadístico de prueba (*test statistic*) es significativamente diferente de cero ya que nunca llega a ser más negativo que los valores críticos. En consecuencia, ambas pruebas sugieren que la serie puede tratarse como no estacionaria.

Cuadro 9.1: Prueba de raíz unitaria Dickey-Fuller aumentada

ADF Test Statistic	0.567467	1 %	Critical Value*	-2.5994
		5 %	Critical Value	-1.9456
		10 %	Critical Value	-1.6185

*MacKinnon critical values for rejection of hypothesis of a unit root

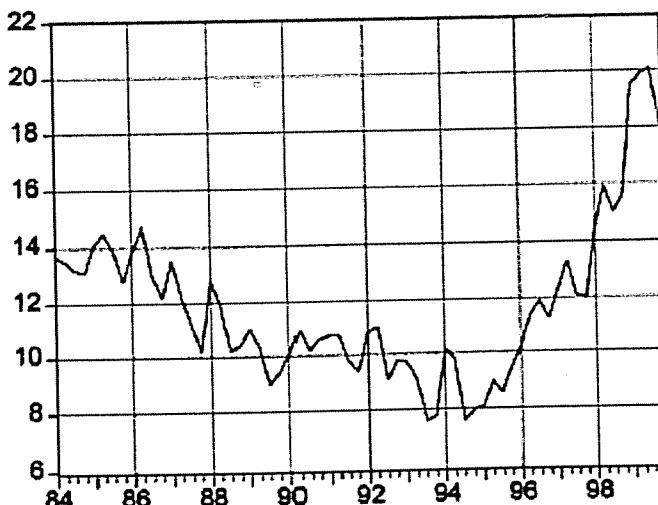
Augmented Dickey-Fuller
 Dependent Variable: D(U)
 Method: Least Squares
 Date: 04/29/02 Time: 13:35
 Sample(adjusted): 1985:2 2000:4
 Included observations: 63 after adjusting endpoints

Variable	Coefficient	Std. Error	t-Statistic	Prob.
U(-1)	0.005987	0.010550	0.567467	0.5726
D(U(-1))	-0.088716	0.115835	-0.765883	0.4469
D(U(-2))	-0.199537	0.116492	-1.712871	0.0921
D(U(-3))	-0.013861	0.115320	-0.120193	0.9047
D(U(-4))	0.529751	0.117619	4.503964	0.0000
R-squared	0.393182	Mean dependent var		0.088889
Adjusted R-squared	0.351332	S.D. dependent var		1.207251
S.E of regression	0.972319	Akaike info criterion		2.857773
Sum squared resid	54.83347	Schwarz criterion		3.027863
Log likelihood	-85.01986	Durbin-Watson stat		2.101460

Cuadro 9.2: Prueba de raíz unitaria Phillips-Perron

PP Test Statistic	0.885554	1 %	Critical Value*	-2.5973
		5 %	Critical Value	-1.9452
		10 %	Critical Value	-1.6183
*MacKinnon critical values for rejection of hypothesis of a unit root				
Lag truncation for Bartlett kernel: 3	(Newey-West suggests: 3)			
Residual variance with no correction			1.369030	
Residual variance with correction			0.695813	
Phillips-Perron Test Equation Dependent Variable: D(U) Method: Least Squares Date: 04/29/02 Time: 13:37 Sample(adjusted): 1984:2 2000:4 Included observations: 67 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
U(-1)	0.005074	0.011405	0.444834	0.6579
R-squared	-0.002868	Mean dependent var	0.089552	
Adjusted R-squared	-0.002868	S.D. dependent var	1.177199	
S.E. of regression	1.178886	Akaike info criterion	3.181830	
Sum squared resid	91.72500	Schwarz criterion	3.214738	
Log likelihood	-105.5913	Durbin-Watson stat	2.101416	

Figura 9.8: Tasa de desempleo (porcentual trimestral)



Otras pruebas de raíz unitaria

Se han desarrollado otras variantes de la prueba de raíz unitaria para acomodar diferentes especificaciones; por ejemplo, para cuando el error ε_t no está autocorrelacionado, pero sigue un modelo GARCH (véase Byeongseon, 1999), lo cual altera la distribución de probabilidad de las pruebas.

Kwiatkowski *et al.* (1992) desarrollaron una prueba tipo multiplicador de Lagrange conocida como KPSS en la cual se prueba la hipótesis nula de que la serie x_t es estacionaria alrededor de una tendencia frente a la hipótesis alterna de que la serie es estacionaria en diferencias (*i. e.* tiene raíz unitaria). El estadístico de prueba es proporcional a la suma acumulada del cuadrado de los residuos de una regresión de la variable x_t contra variables determinísticas, como una constante o tendencia. Un estadístico de prueba más grande que los valores críticos (provistos por los autores) es indicio de no estacionariedad en tendencia (rechazo de H_0).

Otra prueba de estacionariedad es la conocida como relación de varianza (*variance ratio test*), propuesta por Cochrane (1988). La hipótesis nula es que la serie es un paseo aleatorio mientras que la hipótesis alterna es que es estacionaria. La relación de varianza $VR(\tau)$ tiene la siguiente expresión:

$$VR(\tau) = \frac{\text{var}(x_{t+\tau} - x_t)}{\tau \text{var}(x_{t+1} - x_t)}$$

para rezago τ . Bajo la hipótesis nula, el numerador es $\tau\sigma_\varepsilon^2$ y el denominador también, de manera que $VR(\tau)$ tiende a uno para cualquier τ . Si es estacionaria, la relación de varianza cae por debajo de 1 a medida que τ aumenta. Se puede mostrar que bajo la hipótesis nula $VR(\tau)$ tiende asintóticamente a una Normal con media uno y varianza $4\tau/3n$ (véase Maddala y Kim, 1998).

Estacionalidad

La presencia de estacionalidad complica las pruebas tradicionales de raíz unitaria porque tiende a introducir dependencia entre el nivel de la serie y el cambio en ésta. Por ejemplo, cuando estamos en un

pico estacional, hay una tendencia a bajar (de otra forma no se dibuja la estacionalidad). De manera que la no estacionariedad de una serie podría alterarse cuando se le suma un patrón estacional. En contraste con el caso no estacional en el cual la estacionariedad se alcanzaba con la diferencia, $x_t - x_{t-1}$, la presencia de estacionalidad puede requerir diferenciación a rezagos estacionales para lograr estacionariedad; por ejemplo, $x_t - x_{t-4}$ para el caso trimestral o $x_t - x_{t-12}$ para el caso mensual.

Si la estacionalidad es determinística, se puede eliminar utilizando variables *dummy*; aplicando luego las pruebas de raíz unitaria sobre la serie desestacionalizada de la manera usual. Por ejemplo, si hay estacionalidad trimestral, para lo cual bastan tres *dummies* más intercepto, tendríamos:

$$x_t = \hat{\gamma}_0 + \hat{\gamma}_1 D_1 + \hat{\gamma}_2 D_2 + \hat{\gamma}_3 D_3 + x_t^d$$

Donde D_t es uno para el trimestre correspondiente y x_t^d es la serie x_t desestacionalizada. Las pruebas se aplican sobre la serie x_t^d .

Cuando la estacionalidad no es determinística, hay que tener cierta precaución con la aplicación de la prueba usual de raíz unitaria a una serie diferenciada estacionalmente $(1 - L^s)x_t$ ya que, además de eliminar la estacionalidad, también puede atenuar la tendencia, que es lo que estamos buscando con la prueba. En otras palabras, tomar primera diferencia estacional, especialmente en términos porcentuales $(1 - L^s)x_t/x_{t-s}$, puede convertir una serie no estacionaria (como el M1 mensual) en estacionaria.

Existen varias modificaciones de la prueba para cuando la raíz unitaria sea de la forma $1 - L^s$; una de las cuales es la desarrollada por Hylleberg *et al.* (1990). Sin embargo, es posible que una modificación de la prueba ADF aplicada a la serie $(1 - L^s)x_t$ dé resultados satisfactorios sin necesidad de recurrir al más elaborado método de Hylleberg. En efecto, la ADF modificada, en el sentido de usar rezago estacional, se aplica sobre la expresión:

$$\Delta_s x_t = (1 - L^s)x_t = \varphi z_{t-s} + \sum_{i=1}^p \gamma_i \Delta_s x_{t-i} + \varepsilon_t. \quad (9.4)$$

Donde z_t es una versión de x_t desestacionalizada (véase Charemza y Deadman, 1992, p. 137). Tablas especiales de valores críticos fueron

calculadas por Dickey, Hasza y Fuller (1984). De todas maneras, se puede efectuar una prueba estacional *aproximada* utilizando x_{t-s} en lugar de z_{t-s} en la expresión anterior, así:

$$\Delta_s x_t = \varphi x_{t-s} + \sum_{i=1}^p \gamma_i \Delta_s x_{t-i} + \varepsilon_t \quad (9.5)$$

Los valores críticos de esta prueba son los mismos de Dickey, Hasza y Fuller (1984) y pueden ser consultados, por ejemplo, en la Tabla 3 de Charemza y Deadman (1992).

Algunos autores desestacionalizan las series, por ejemplo, con el X-11 o el TRAMO/SEATS, antes de someterlas a las pruebas de raíz unitaria; sin embargo, Ghysels y Perron (1993) encuentran que los procedimientos para desestacionalizar series tienden a disminuir el poder estadístico (probabilidad de rechazar H_0 cuando es falsa) de las pruebas. Es mejor trabajar con las series originales, sin ajuste estacional.

Una raíz estacional realmente representa varias raíces unitarias en diferentes frecuencias. En general, una raíz estacional $1 - L^s$ puede escribirse como:

$$(1 - L^s) = (1 - L)(1 + L + L^2 + \cdots + L^{s-1})$$

Un ejemplo de estacionalidad trimestral podría ser:

$$\begin{aligned} (1 - L^4) &= (1 - L)(1 + L + L^2 + L^3) \\ &= (1 - L)(1 + L)(1 + L^2) \\ &= (1 - L)(1 + L)(1 - iL)(1 + iL) \end{aligned}$$

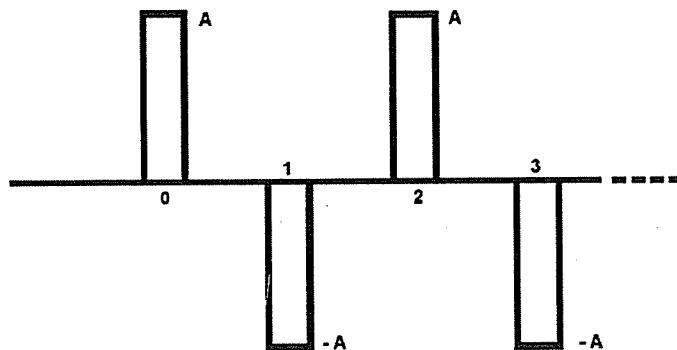
Como cualquier polinomio, $1 - L^4$ tiene hasta cuatro raíces y al ser un polinomio real, entonces si acaso algunas de las raíces son imaginarias, éstas vendrán en parejas conjugadas. Todas estas raíces tienen magnitud unitaria. Para ver la relación con la frecuencia, sigamos con la ilustración trimestral escribiendo:

$$(1 - L^4)x_t = (1 - L)(1 + L)(1 - iL)(1 + iL)x_t = 0$$

Dentro de la expresión anterior, tomemos el término $(1 + L)x_t = 0$, de donde $x_t = -x_{t-1}$. Esta ecuación dibuja un ciclo cada dos unidades

de tiempo (frecuencia 0.5 por unidad de tiempo), ya que si iniciamos con $x_0 = A$, entonces $x_1 = -A$ y $x_2 = A$ de nuevo, como se muestra en la Figura 9.9.

Figura 9.9: Ciclo cada dos unidades de tiempo



Tomemos ahora el término $(1 - iL)x_t = 0$. En este caso $x_t = ix_{t-1}$, y si iniciamos con $x_0 = A$, entonces $x_1 = iA$, $x_2 = i(iA) = -A$, $x_3 = i(-A) = -iA$ y, finalmente, $x_4 = i(-iA) = -i^2A = A$, de manera que para volver a A se necesitan cuatro unidades de tiempo (frecuencia 0.25 por unidad de tiempo). La frecuencia del término $(1 + iL)x_t = 0$ es igual.

Tomemos el término $(1 - L)x_t = 0$ donde $x_t = x_{t-1}$. Puede verse que x_t tendrá siempre el mismo valor, y no habrá ciclos, de manera que este término corresponde a una frecuencia de cero. Un método para probar estas raíces se encuentra en Enders (2004, capítulo 4).

Estimación de la cointegración bivariada

Una vez se establece que las series x_{1t} y x_{2t} son $I(1)$, entonces el siguiente paso es determinar si están cointegradas. Para ello, en el siguiente procedimiento desarrollado por Engle y Granger (1987), se corre la regresión de una variable contra la otra, digamos:

$$x_{1t} = a_0 + a_1 x_{2t} + u_t \quad (9.6)$$

Si efectivamente hay cointegración, debe encontrarse que u_t es $I(0)$. Estimar una regresión entre variables cointegradas, esto es,

encontrar \hat{a}_1 es parecido a sintonizar un radio antiguo; giramos la perilla que controla los posibles valores de \hat{a}_1 y vamos escuchando el ruido, que en este caso es u_t . Sólo cuando \hat{a}_1 le pegue a a_1 lograremos identificar el preciso equilibrio entre las variables, y el ruido (su varianza) caerá drásticamente porque u_t pasa casi instantáneamente de ser una $I(1)$ a $I(0)$. Esto hace que la estimación de MCO entre variables cointegradas sea mucho más sensible en la vecindad de a_1 que la estimación entre variables no cointegradas, razón por la cual se dice que la estimación con variables cointegradas es *superconsistente* (véase, por ejemplo, Davidson y MacKinnon, 2004).

En la literatura se le da el nombre de *vector de cointegración* a los coeficientes de la relación lineal que multiplican las variables en (9.6), incluyendo la constante, esto es, a $(1 - a_0 - a_1)$.

Nótese que correr (9.6) produce un \hat{a}_1 que no es simplemente el inverso de \hat{b}_1 en la regresión

$$x_{2t} = b_0 + b_1 x_{1t} + v_t \quad (9.7)$$

excepto cuando $R^2 \rightarrow 1$, o cuando $n \rightarrow \infty$ (véase Verbeek, 2004, capítulo 9). Por esta razón se sugiere que los resultados que apuntan a la cointegración sean válidos sólo si la regresión muestra alguna significancia. Hacer igual a uno el coeficiente de la variable dependiente es arbitrario. Ambas regresiones anteriores (9.6) o (9.7) pueden ser escogidas para la prueba de cointegración, y pueden dar resultados contradictorios en la práctica. Algunos de los métodos de máxima verosimilitud con información completa encuentran los vectores de cointegración sin necesidad de escoger variable dependiente (Johansen, 1988; Stock y Watson, 1988; Bierens, 1997; Quintos, 1998). Volveremos sobre este tema más adelante cuando tratemos la cointegración multivariada.

Recordamos que en la literatura los residuos de la regresión de cointegración, en este caso u_t o v_t en (9.6) o (9.7), se denominan z_t , ya que no necesariamente deben ser ruido blanco, sino $I(0)$. Una prueba que nos permite ver si el residuo de la regresión de cointegración es $I(0)$ es la de Sargan y Bhargava (1983) que consiste en observar el DW. Un DW de z_t muy pequeño denota alta correlación serial y es evidencia de raíz unitaria. Por ende, de manera aproximada, si DW es menor a 0.5 o, según algunos autores (Granger y Newbold, 1974) si $DW < R^2$,

se concluye que el residuo de la regresión de cointegración es $I(1)$, se rechaza la hipótesis de cointegración y se acepta la de regresión espuria. Existen tablas especiales para el DW utilizado de esta manera como prueba de cointegración.

Sin embargo, la prueba usual de cointegración bivariada es la misma de raíz unitaria ADF, o la de Perron, aplicada a los residuos z_t . Aunque algunos autores utilizan las mismas tablas que se usan para la prueba de raíz unitaria, otros notan que probar cointegración no es lo mismo que probar estacionariedad. La razón es que MCO minimiza la suma de los residuos al cuadrado, escogiendo el parámetro $\hat{\alpha}$, de manera que x_{1t} y x_{2t} tengan la mayor probabilidad de parecer cointegradas en la muestra, esto es, de moverse juntas.

Por esta razón, cuando se aplican pruebas como la ADF o la PP para probar cointegración, se sugiere trabajar con valores críticos más exigentes (más negativos) o utilizar tablas especialmente calculadas para ello (véase Banerjee *et al.*, 1993, p. 9; Davidson y Mackinnon, 1993, Tabla 20.2). En el caso de la ADF se prueban los residuos, por ejemplo, de la regresión (9.6), así:

$$\Delta \hat{u}_t = \varphi \hat{u}_{t-1} + \sum_{i=1}^p \gamma_i \Delta \hat{u}_{t-i} + \varepsilon_t. \quad (9.8)$$

Si el estadístico de prueba $\hat{\varphi}/(\text{error estándar}) = 0$ estadísticamente, entonces \hat{u}_t es $I(1)$ y no hay cointegración. En caso contrario, la regresión entre las variables x_{1t} y x_{2t} , como en (9.6), no sería espuria. Sin embargo, para que sea una regresión apropiada, no basta con que u_t sea $I(0)$, sino que es requisito que u_t sea ruido; para lo cual puede ser necesario agregar rezagos de x_{1t} o x_{2t} como variables explicativas. Cuando x_{1t} y x_{2t} están cointegradas, hay otra alternativa que se describe a continuación.

Modelo de corrección de error

Si las dos variables x_{1t} y x_{2t} están cointegradas, entonces debe existir un mecanismo de corrección de error, una especie de estabilizador automático que preserve el equilibrio. Este resultado se conoce como *teorema de representación de Granger* (Granger, 1984; Johansen, 1991).

El mecanismo de corrección de error consiste en una regresión que combina la información de largo plazo contenida en la cointegración o relación de equilibrio, $z_t = x_{1t} - ax_{2t}$, con la información de corto plazo contenida en las variaciones de x_{1t} y x_{2t} .

La expresión de un mecanismo de corrección de error es similar a un VÁR (que ahora se llama VEC, por vector de corrección de error), donde hay una ecuación para cada una de las variables. La variable dependiente en cada ecuación es la primera diferencia de la variable correspondiente, ya sea x_{1t} o x_{2t} , expresada en función de sus propios rezagos en diferencias, los rezagos de la otra variable en diferencias y el término de cointegración rezagado; por ejemplo:

$$\begin{aligned}\Delta x_{1t} &= \beta_{10} + \beta_{11}\Delta x_{1t-1} + \beta_{12}\Delta x_{2t-1} + \beta_{13}z_{t-1} + \varepsilon_{1t} \\ &= \beta_{10} + \beta_{11}\Delta x_{1t-1} + \beta_{12}\Delta x_{2t-1} \\ &\quad + \beta_{13}(x_{1t-1} - ax_{2t-1}) + \varepsilon_{1t} \\ \Delta x_{2t} &= \beta_{20} + \beta_{21}\Delta x_{1t-1} + \beta_{22}\Delta x_{2t-1} + \beta_{23}z_{t-1} + \varepsilon_{2t} \\ &= \beta_{20} + \beta_{21}\Delta x_{1t-1} + \beta_{22}\Delta x_{2t-1} \\ &\quad + \beta_{23}(x_{1t-1} - ax_{2t-1}) + \varepsilon_{2t}\end{aligned}\tag{9.9}$$

lo cual en notación de VEC quedaría:

$$\begin{bmatrix} \Delta x_{1t} \\ \Delta x_{2t} \end{bmatrix} = \begin{bmatrix} \beta_{10} \\ \beta_{20} \end{bmatrix} + \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{bmatrix} \begin{bmatrix} \Delta x_{1t-1} \\ \Delta x_{2t-1} \end{bmatrix} + \begin{bmatrix} \beta_{13} \\ \beta_{23} \end{bmatrix} (x_{1t-1} - ax_{2t-1}) + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}\tag{9.10}$$

Es de notar que todas las variables explicativas utilizadas en estas regresiones, incluido el término de cointegración $z_t = x_{1t} - ax_{2t}$, son $I(0)$, por lo cual son admisibles. Se pueden agregar más rezagos de ambas variables para aumentar el poder explicativo o para asegurarse de que ε_t sea ruido blanco. Los coeficientes β_{13} y β_{23} que multiplican la relación de cointegración z_{t-1} se conocen como coeficientes de ajuste. Su magnitud indica el grado de ajuste con respecto a desviaciones del equilibrio de largo plazo.

En la primera ecuación de (9.9) puede observarse que el término $\beta_{13}(x_{1t-1} - ax_{2t-1})$ ejerce la corrección necesaria sobre Δx_{1t} con el

objetivo de mantener el equilibrio entre x_{1t} y x_{2t} ; para lo cual debe cumplirse que $\beta_{13} < 0$. Si en un momento dado $x_{1t-1} > ax_{2t-1}$, entonces $\beta_{13}(x_{1t-1} - ax_{2t-1})$ será negativo y actuará de manera que tienda a disminuir la siguiente observación de la serie x_{1t} y la acerque de nuevo hacia x_{2t} . Lo contrario sucede si $x_{1t-1} < ax_{2t-1}$. Un mecanismo similar opera en el caso de la segunda ecuación de (9.9) para $\beta_{23} > 0$. Sin embargo, si x_{1t} y x_{2t} se mueven en sentido contrario, la relación de cointegración será $x_{1t} + ax_{2t}$, y la corrección se obtiene con $\beta_{13} < 0$ y $\beta_{23} < 0$.

Debe recordarse que, en general, uno de los dos coeficientes β_{13} y β_{23} puede ser cero, de manera que las desigualdades anteriores deben entenderse como menor o igual que cero, o mayor o igual que cero, según sea el caso. Si ambos coeficientes de ajuste son cero, entonces cada variable se mueve sin referencia a la otra y no habría cointegración. Una exposición intuitiva de la cointegración y del papel que desempeña el término de corrección de error puede consultarse en Murray (1994).

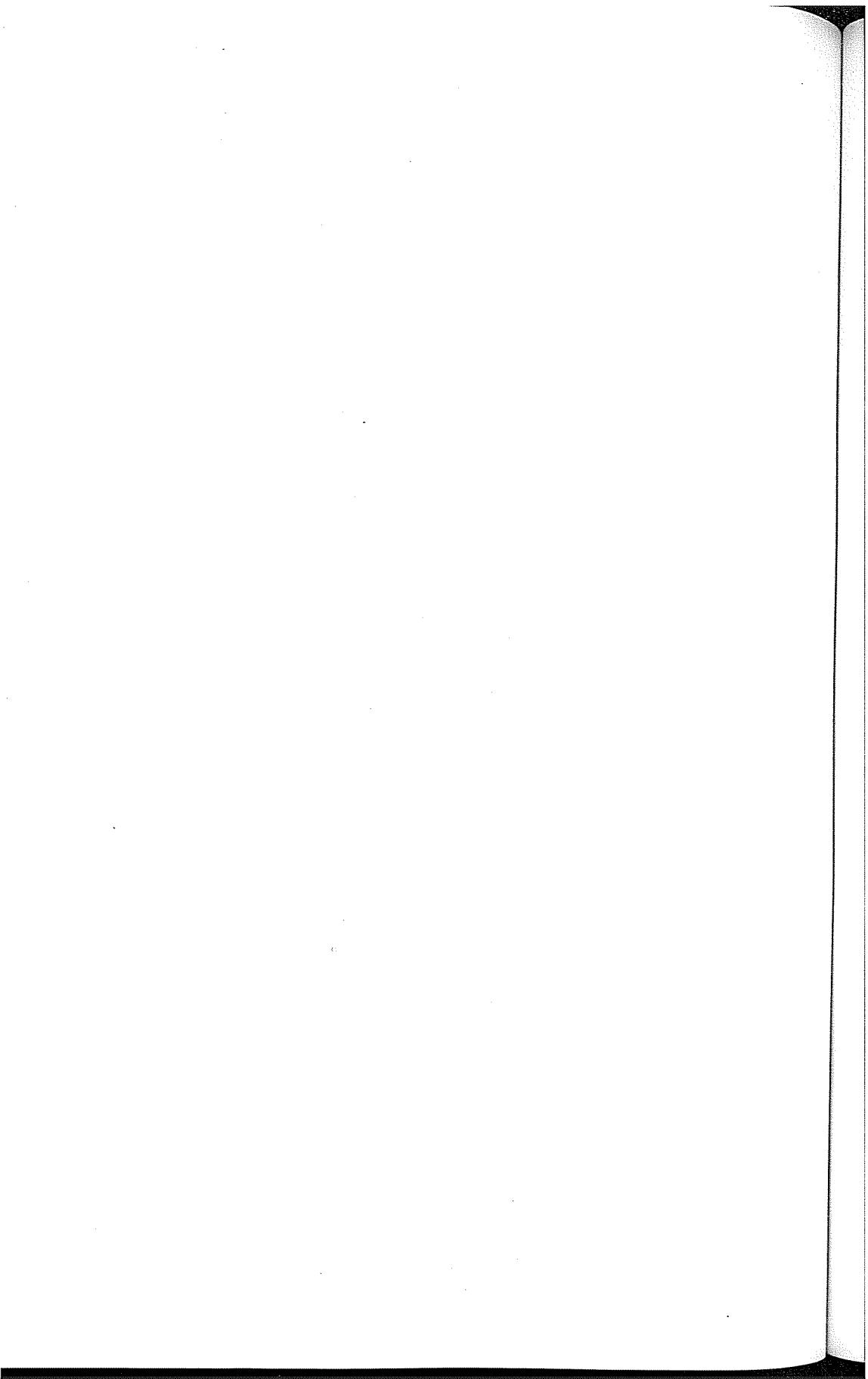
Algunos autores sugieren que, una vez establecida la cointegración, se puede correr (9.9) pero sin forzar el término z_{t-1} , así:

$$\begin{aligned}\Delta x_{1t} = & \beta_{10} + \beta_{11}\Delta x_{1t-1} + \beta_{12}\Delta x_{2t-1} \\ & + \beta_{14}x_{1t-1} + \beta_{15}x_{2t-1} + \varepsilon_{1t}\end{aligned}\tag{9.11}$$

$$\begin{aligned}\Delta x_{2t} = & \beta_{20} + \beta_{21}\Delta x_{1t-1} + \beta_{22}\Delta x_{2t-1} \\ & + \beta_{24}x_{1t-1} + \beta_{25}x_{2t-1} + \varepsilon_{2t}\end{aligned}$$

y confiar en que los signos sean los apropiados. Es importante subrayar que cuando existe una relación de cointegración entre variables, existe un mecanismo de corrección de error que las mantiene en equilibrio. Una regresión como (9.9) en diferencias no debe correrse sin el término de equilibrio z_{t-1} cuando hay cointegración, ya que se incuraría en el problema de omisión de variable relevante. Como se verá en el siguiente capítulo, el mecanismo de corrección de error incluye en su derivación el término z_{t-1} .

La estimación bivariada expuesta aquí, tanto de la relación de cointegración como del mecanismo de corrección de error, puede igualmente efectuarse por alguno de los métodos que se han desarrollado para la estimación de la cointegración multivariada (más de dos variables).



CAPÍTULO 10

Cointegración multivariada

Entre dos variables $I(1)$ sólo puede haber un vector de cointegración, una sola forma de combinar ambas variables para obtener una serie $I(0)$; pero cuando se consideran tres o más variables el análisis se complica ya que pueden presentarse múltiples vectores de cointegración. Como se verá, puede haber desde cero vectores de cointegración, hasta un máximo de $m - 1$ vectores de cointegración independientes, donde m es el número de variables involucradas. También es posible que haya variables con distinto orden de integración y que sus combinaciones estén cointegradas (situación que en la literatura se conoce como multicointegración); por ejemplo, dos variables pueden ser $I(2)$ y una tercera $I(1)$, y las dos $I(2)$ combinarse para producir una $I(1)$ que, al combinarse con la otra $I(1)$, producen una $I(0)$.

Definiciones

Decimos que un vector x_t de m variables está cointegrado si las variables son $I(1)$ y se puede encontrar al menos un vector a tal que $a'x_t$ sea una serie estacionaria $I(0)$, que llamaremos z_t ; esto es, $a'x_t = z_t$. El vector a se llama *vector de cointegración*. Obviamente, si multiplicamos a por un escalar γ , la expresión $\gamma a'x_t = \gamma z_t$ seguirá siendo estacionaria

$I(0)$. La razón es que una variable $I(0)$ multiplicada por una constante sigue siendo $I(0)$, lo cual implica que $\gamma a'$ será un nuevo vector de cointegración válido. Alguna literatura reciente define la cointegración en el sentido de que $a'x_t$ sea estacionaria en tendencia, de manera que, por ejemplo, si $a'x_t$ es igual a una serie $I(0)$ más una tendencia β_t , multiplicar $a'x_t$ por γ genera otra serie estacionaria en tendencia. Esta definición más amplia se acepta en consideración a que dicha tendencia determinística puede ser fácilmente incluida en el modelo.

Se trata entonces de que sea posible expresar una de las series del vector x_t en función de las otras $m-1$ series de manera que la diferencia (la serie de los residuos) sea $I(0)$. La cointegración es una relación de equilibrio de largo plazo entre variables, donde cualquiera de ellas se puede expresar como una combinación lineal de las demás, más un término de error z_t estacionario (no necesariamente ruido blanco). Así, se acostumbra normalizar el coeficiente de alguna de las variables del vector x_t (usualmente x_{1t}) para que valga uno.

Es conveniente que, tanto la relación del equilibrio representada por la expresión $a'x_t = z_t$, como las variables propuestas e involucradas en x_t , tengan algún sentido y explicación teórica. Aunque es posible trabajar y analizar los datos con poca teoría, la aceptación y utilidad práctica de los resultados obtenidos será mayor en la medida en que cuenten con un apoyo teórico.

Las m series individuales que componen x_t pueden verse como vectores, cada uno de n datos (dimensiones), y la cointegración como el hecho de poder expresar uno de esos vectores (una de las series) como una combinación lineal de los otros $m-1$ vectores. Por ejemplo, para el caso de dos variables, x_{1t} y x_{2t} , la cointegración implicaría una relación de equilibrio expresable de manera general como $a_{11}x_{1t} + a_{12}x_{2t} = z_t$, donde el vector de cointegración a , de orden 2×1 , será $\begin{bmatrix} a_{11} \\ a_{12} \end{bmatrix}$.

Para tres variables, x_{1t} , x_{2t} y x_{3t} , es posible tener hasta dos relaciones de equilibrio, por ejemplo:

$$a_{11}x_{1t} + a_{12}x_{2t} + a_{13}x_{3t} = z_{1t}$$

$$a_{21}x_{1t} + a_{22}x_{2t} + a_{23}x_{3t} = z_{2t}$$

Donde los dos vectores de cointegración son

$$a_1 = \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \end{bmatrix} \text{ y } a_2 = \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \end{bmatrix}.$$

Es intuitivamente claro que la combinación lineal de dos variables $I(0)$ produce otra variable $I(0)$, de manera que, por ejemplo para cualquier par de escalares, β_1 y β_2 , tenemos $\beta_1 z_{1t} + \beta_2 z_{2t} \sim I(0)$, expresión que se puede escribir como:

$$\beta_1 a'_1 x_t + \beta_2 a'_2 x_t = (\beta_1 a'_1 + \beta_2 a'_2) x_t = (\beta_1 a_1 + \beta_2 a_2)' x_t \sim I(0) \quad (10.1)$$

Lo anterior implica que existe un número infinito de vectores de cointegración construidos a partir de combinaciones lineales de los vectores a_1 y a_2 , porque se puede utilizar una infinidad de β_1 y β_2 para combinar a_1 y a_2 como en (10.1) y obtener una serie $I(0)$, donde $\beta_1 a_1 + \beta_2 a_2$ será un nuevo vector de cointegración. En consecuencia, podemos identificar el espacio de los vectores de cointegración, pero no los vectores individuales; si queremos identificar vectores individuales dentro de este espacio vectorial, tendremos que adoptar algún mecanismo arbitrario de identificación (volveremos sobre este tema más adelante).

Un resultado importante, que mostramos a continuación de manera intuitiva, es que, para m variables, puede haber máximo $m - 1$ vectores de cointegración *independientes*, cada uno de orden $m \times 1$. En otras palabras, para m variables, puede haber cero, una, dos, ..., hasta $m - 1$ relaciones de equilibrio independientes. Para entender la razón por la cual sucede esto debemos considerar las siguientes dos propiedades:

Propiedad 1. *Todo vector de cointegración debe tener, por lo menos, dos elementos diferentes de cero.* Si tuviera un solo elemento diferente de cero, entonces al multiplicar dicho vector de cointegración por el vector de variables x_t estaríamos multiplicando una sola de las m variables por ese elemento y diciendo que tal producto es $I(0)$, cuando sabemos que individualmente las m variables son $I(1)$.

Propiedad 2. *Cualquier combinación lineal de vectores de cointegración de x_t es un vector de cointegración de x_t .* Como se explicó, para un vector de cointegración a y un escalar γ , la expresión $\gamma a'$ también

es un vector de cointegración. Lo mismo sucede con la combinación lineal de varios vectores de cointegración, como en (10.1).

Para ilustrar, utilicemos un vector x_t con tres variables, $m = 3$. Esto quiere decir que los vectores de cointegración, si los hay, tendrán tres elementos (ejes o dimensiones), uno para cada variable; en otras palabras estarían en un espacio \mathbb{R}^3 , un volumen.

Una posibilidad es que no haya ningún vector de cointegración entre las tres variables. Otra posibilidad es que haya un solo vector de cointegración, caso en el cual ese vector podría estar en cualquier parte del volumen, excepto sobre uno de los ejes, porque se violaría la Propiedad 1 al ser cero los elementos de los otros dos ejes.

Si hay dos vectores de cointegración independientes, obviamente de orden 3×1 , éstos podrían estar en cualquier parte del volumen excepto en un plano que contenga alguno de los ejes porque dos vectores independientes en un plano generan todos los vectores del plano y por la Propiedad 2 todos esos vectores serían a su vez vectores de cointegración, incluido el eje, lo cual violaría la Propiedad 1, al ser cero los elementos de los otros dos ejes.

No obstante, si hubiera tres vectores de cointegración independientes, no podrían estar en ninguna parte porque tres vectores independientes, obviamente de orden 3×1 , generan (*span*) todos los posibles vectores del volumen y todas estas combinaciones lineales serían, a su vez, vectores de cointegración por la Propiedad 2. Obviamente, entre todas estas combinaciones algunas tendrían un solo elemento diferente de cero, porque coincidirían con alguno de los ejes, lo cual contradice la Propiedad 1. Por lo tanto, con $m = 3$, no puede haber tres vectores de cointegración independientes. Tampoco puede haber cuatro, ni cinco, ni seis, etc., porque más de tres vectores en un volumen ya no son independientes.

Generalizando: si hay m variables en x_t , los vectores de cointegración (si los hay) estarían en un espacio \mathbb{R}^m y serían obviamente de orden $m \times 1$. Si hubiera m vectores de cointegración independientes, sus combinaciones lineales generarían la totalidad del espacio \mathbb{R}^m , incluidos sus ejes, los cuales también serían vectores de cointegración según la Propiedad 2 pero en violación de la Propiedad 1. Por ello se deduce que no puede haber m vectores de cointegración independientes; aunque si tenemos *máximo* $m - 1$ vectores de cointegración independientes,

esta contradicción no se presenta ya que será posible que generen subespacios que no incluyan los ejes.

Denotamos con el símbolo M el número de vectores de cointegración independientes que, por la discusión anterior, tiene que cumplir $M \leq m - 1$. Podemos recoger en una matriz A , de orden $m \times M$, los M vectores de cointegración:

$$A = [a_1 \quad a_2 \quad a_3 \quad \cdots \quad a_M]$$

Donde cada a_i es un vector de m elementos, con lo cual $A'x_t$ será un vector de orden $M \times 1$ de procesos $I(0)$; esto es, estacionarios. Al igual que en el caso de un solo vector, multiplicar cada uno de los M vectores a_i por un escalar sigue produciendo un proceso $I(0)$. En otras palabras, si β' es un vector $1 \times M$ que reúne estos escalares, entonces $\beta'A'x_t = \beta'z_t$ seguirá siendo un proceso $I(0)$, pero sólo uno.

La combinación lineal de vectores de cointegración $\beta'A'$ es, a la vez, un vector de cointegración de m elementos (al igual que cada a_i), como se anotó en la Propiedad 2 arriba.

Cuando hay M vectores de cointegración y corremos una variable contra las otras –con lo cual ya se está seleccionando una de las m variables para que sea la dependiente–, el método de mínimos cuadrados ordinarios MCO estima, de entre las M posibles relaciones de cointegración, aquella relación cuyo vector de errores $\hat{\epsilon}_t$ sea mínimo y ortogonal al espacio generado por las demás variables usadas como independientes (nótese que hablamos de $\hat{\epsilon}_t$ y no de z_t).

Por ejemplo, si corremos x_{1t} contra x_{2t}, \dots, x_{mt} tenemos, vectorialmente, que \hat{x}_{1t} será la proyección de x_{1t} sobre el espacio generado por x_{2t}, \dots, x_{mt} y que $\hat{\epsilon}_t$ será ortogonal a este espacio. Sin embargo, cuando no existe relación de cointegración alguna y corremos x_{1t} contra x_{2t}, \dots, x_{mt} , simplemente tenemos el problema de regresión espuria descrito en el Capítulo 8. También puede suceder que exista cointegración pero que no involucre la variable escogida como dependiente, sino las independientes; entonces, la regresión sería espuria con el agravante de que el análisis de sus residuos no detectaría la cointegración existente entre las variables independientes (véase Enders, 2004, capítulo 6). Por ejemplo, si x_{1t} , x_{2t} y x_{3t} son $I(1)$ y sólo hay cointegración entre x_{2t} y x_{3t} , una regresión al estilo del procedimiento de Engle-Granger,

descrita en el capítulo anterior, en la cual se corre x_{1t} contra x_{2t} y x_{3t} , sería espuria y no serviría para detectar dicha cointegración.

Ejemplo

Si a_1 y a_2 son vectores de cointegración de orden 3×1 para tres variables x_{1t} , x_{2t} y x_{3t} , podemos escribir las dos posibles relaciones de cointegración independientes como:

$$\begin{aligned} a'_1 x_t &= a_{11}x_{1t} + a_{12}x_{2t} + a_{13}x_{3t} = z_{1t} \sim I(0) \\ &\quad (10.2) \end{aligned}$$

$$a'_2 x_t = a_{21}x_{1t} + a_{22}x_{2t} + a_{23}x_{3t} = z_{2t} \sim I(0)$$

O, en forma compacta, como:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = \begin{bmatrix} a'_1 \\ a'_2 \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{bmatrix} = A' x_t = \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = z_t \quad (10.3)$$

Donde z_t es el *vector de procesos estacionarios resultantes*. En este ejemplo, $A' x_t = z_t$ será un vector 2×1 de procesos estacionarios z_{1t} y z_{2t} . Por otro lado, cualquier combinación lineal de los anteriores, $\beta_1 z_{1t} + \beta_2 z_{2t}$, seguirá siendo $I(0)$. De manera que:

$$[\beta_1 \quad \beta_2] \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \beta_1 z_{1t} + \beta_2 z_{2t} = \beta' z_t = \beta' A' x_t \sim I(0) \quad (10.4)$$

donde β' es $1 \times M$, A' es $M \times m$ y x_t es $m \times 1$.

Ejemplo

Para ilustrar la forma como puede haber varias relaciones de cointegración (equilibrios) en la práctica, supongamos que r_{mt} es la tasa de interés nominal del mercado, r_{ft} es la tasa de interés nominal libre de riesgo y π_t es la inflación, y que las tres variables son $I(1)$. Una relación de cointegración entre estas variables puede ser la que se conoce como prima de riesgo, que involucra la tasa de interés de mercado y la tasa de interés libre de riesgo:

$$a_{10} + a_{11}r_{mt} + a_{12}\pi_t + a_{13}r_{ft} = z_{1t}$$

Donde se espera que $a_{12} = 0$. Otra relación podría ser la del efecto de Fisher, que involucra la tasa de interés libre de riesgo con la inflación:

$$a_{20} + a_{21}r_{mt} + a_{22}\pi_t + a_{23}r_{ft} = z_{2t}$$

Se espera entonces que $a_{21} = 0$.

Representación del vector autorregresivo y del de corrección de error

Supongamos que x_t es un vector de m variables que puede modelarse como un vector autorregresivo VAR(p):

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_p x_{t-p} + \varepsilon_t \quad (10.5)$$

Donde α_0 es un vector $m \times 1$ de interceptos y los demás α_i son matrices $m \times m$ de coeficientes. Puede mostrarse (véase ejemplo ilustrativo más adelante) que la expresión del VAR(p) en (10.5) es equivalente a:

$$x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_2 \Delta x_{t-2} + \cdots + \zeta_{p-1} \Delta x_{t-p+1} + \varphi x_{t-1} + \varepsilon_t \quad (10.6)$$

Donde $\varphi = \alpha_1 + \alpha_2 + \cdots + \alpha_p$ y $\zeta_i = -(\alpha_{i+1} + \alpha_{i+2} + \cdots + \alpha_p)$ para $i = 1, 2, \dots, p-1$. Por ejemplo, para $i = p-1$ tendríamos $\zeta_{p-1} = -(\alpha_p)$. Al restar el vector x_{t-1} a ambos lados de (10.6) tenemos:

$$\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_2 \Delta x_{t-2} + \cdots + \zeta_{p-1} \Delta x_{t-p+1} + \zeta_0 x_{t-1} + \varepsilon_t \quad (10.7)$$

donde $\zeta_0 = \varphi - I_m$.

Nótese que la forma de (10.7) es similar a una prueba de raíz unitaria del tipo de Dickey-Fuller aumentada (ADF), pero multivariada. Lo importante para el análisis que sigue es la matriz ζ_0 , que es asimétrica de orden $m \times m$.

Si todas las m variables del vector x_t son $I(1)$, y no están cointegradas, la matriz ζ_0 estaría llena de ceros (rango cero), porque variables $I(1)$ no explican Δx_t que es $I(0)$. Al igual que sucede en la prueba ADF, el nivel de las variables no influye sobre Δx_t . El VAR estaría correctamente especificado en diferencias; esto es, tal como queda cuando $\zeta_0 = 0$ en (10.7). También puede correrse el VAR en niveles como en (10.5), caso en el cual la estimación será consistente y no habrá

problema de regresión espuria, porque se incluyen (no se omiten) los rezagos de cada variable dependiente como variables explicativas.

Si todas las m variables del vector x_t son $I(0)$, la diagonal de ζ_0 tendría elementos negativos como en la ADF y diferentes por fuera de la diagonal de manera que todas las filas serían linealmente independientes y el rango de ζ_0 sería m , y se puede hacer la regresión tal como está especificada en (10.5) o (10.7).

Si todas las m variables son $I(1)$ y existen $M < m$ vectores de cointegración independientes, habrá hasta $M < m$ filas independientes (rango M) en la matriz ζ_0 . La única forma en que x_{t-1} , que es $I(1)$, puede servir para explicar la variable dependiente Δx_t , que es $I(0)$, es si las variables del vector x_{t-1} están cointegradas. Siendo así, los coeficientes de ζ_0 deben hacer que las combinaciones lineales de las m series sean $I(0)$.

Siempre y cuando haya M relaciones de cointegración, cada fila de $\zeta_0 x_{t-1}$, digamos la fila i -ésima, puede ser remplazada por expresiones de la forma $\beta_i' A' x_{t-1}$, o sea $\beta_{i1} z_{1t-1} + \beta_{i2} z_{2t-1} + \dots + \beta_{iM} z_{Mt-1}$. Si ahora consideramos no una fila sino *todas* las m filas, podemos apilar todos los m vectores β_i' en una matriz B de orden $m \times M$ de manera que la expresión:

$$\zeta_0 x_{t-1} = BA' x_{t-1} = Bz_{t-1} \quad (10.8)$$

represente m combinaciones lineales de los M procesos (elementos) del vector estacionario z_{t-1} . Por ejemplo, para $m = 3$ y $M = 2$, el término $BA' x_{t-1}$ será de la forma:

$$BA' x_{t-1} = \begin{bmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \\ \beta_{31} & \beta_{32} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_{1t-1} \\ x_{2t-1} \\ x_{3t-1} \end{bmatrix}$$

Si hay M relaciones de cointegración podemos escribir $\zeta_0 = BA'$, donde las tres matrices son de rango M . Si no hubiera cointegración, ζ_0 debería tener rango cero (en cuyo caso no vale separarla, porque está llena de ceros) o rango completo m . Dado que el rango de A es M , el rango máximo de BA' sería M , menos que m , de manera que, sin cointegración, ζ_0 no podría separarse en BA' , pues el rango de un producto de matrices no puede superar el rango de cualquiera de ellas.

Al reemplazar $\zeta_0 = BA'$ en (10.7) tenemos:

$$\begin{aligned}\Delta x_t &= \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_2 \Delta x_{t-2} + \cdots \\ &\quad + \zeta_{p-1} \Delta x_{t-p+1} + BA' x_{t-1} + \varepsilon_t \\ &= \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_2 \Delta x_{t-2} + \cdots \\ &\quad + \zeta_{p-1} \Delta x_{t-p+1} + B z_{t-1} + \varepsilon_t\end{aligned}\tag{10.9}$$

Tal expresión se conoce como *vector de corrección de error VEC*. El VEC de (10.9) es la generalización de la expresión bivariada del mecanismo de corrección de error mostrado en el capítulo anterior. Segundo el Teorema de Representación de Granger (1984), cuando existe cointegración entre variables también existe un mecanismo, como (10.9), que corrige las desviaciones y mantiene las variables dentro de la relación de equilibrio. Cada elemento del vector $B z_{t-1}$ provee la corrección necesaria para mover la variable x_t correspondiente, a través de Δx_t , en la dirección que asegure el mantenimiento del equilibrio.

Tomando expectativas de la ecuación (10.9) vemos que:

$$(I_m - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1}) E \Delta x_t = \alpha_0 + B E z_{t-1}$$

De donde el valor esperado de Δx_t será:

$$E \Delta x_t = (I_m - \zeta_1 - \zeta_2 - \cdots - \zeta_{p-1})^{-1} (\alpha_0 + B E z_{t-1}) \tag{10.10}$$

Encontramos que $E \Delta x_t = 0$ si $\alpha_0 = -B E z_{t-1}$, de otra forma tendríamos *drift* o corrimiento en las series, lo cual puede ser deseable o no dependiendo de la realidad que estemos modelando. El hecho de que un proceso tenga $E \Delta x_t = 0$ no necesariamente implica que x_t sea estacionario, ya que puede tratarse, por ejemplo, de un paseo aleatorio, el cual tiene $E \Delta x_t = E \varepsilon_t = 0$. En general, el uso de la constante es similar al de la prueba ADF donde, por ejemplo, estaría indicada si la variable muestra una tendencia lineal (*drift*).

Recordamos que $E z_{t-1} = E z_t$ ya que z_t es $I(0)$; esto es, estacionaria. Se recomienda incluir el vector de constantes α_0 en el VEC (10.9), especialmente si z_t puede tener $E z_t \neq 0$ (por ejemplo, si el vector de cointegración no incluye la constante). El *default* usual en los programas de estimación es incluir la constante α_0 en el VEC y también una constante en el vector de cointegración.

Aunque la especificación más popular para manejar la cointegración es a través de los VAR, también se puede manejar en modelos VARMA, que combinan una parte autorregresiva y una de promedio móvil (véase Lutkepohl y Claessen, 1997). Por ejemplo, si en lugar de (10.5) hubiéramos empezado con:

$$\begin{aligned} x_t = & \alpha_0 + \alpha_1 x_{t-1} + \cdots + \alpha_p x_{t-p} + \gamma_1 y_{1t} \cdots \\ & + \gamma_k y_{kt} + \varepsilon_t + d_1 \varepsilon_{t-1} + \cdots + d_q \varepsilon_{t-q} \end{aligned}$$

Donde y_{1t}, \dots, y_{kt} son variables exógenas, la expresión (10.9) quedaría igual pero con la adición del término $\gamma_1 y_{1t} + \cdots + \gamma_k y_{kt} + d_1 \varepsilon_{t-1} + \cdots + d_q \varepsilon_{t-q}$.

Ejemplo

Para ilustrar la transformación del VAR al VEC –esto es, para pasar de la ecuación (10.5) a la (10.9)–, usaremos dos variables, $x'_t = (w_t, y_t)$, y dos rezagos. Supongamos el siguiente VAR:

$$\begin{bmatrix} w_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_{01} \\ g_{01} \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ g_{11} & g_{12} \end{bmatrix} \begin{bmatrix} w_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} c_{21} & c_{22} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} w_{t-2} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \quad (10.11)$$

Sumamos y restamos al lado derecho del VAR el término:

$$\begin{bmatrix} c_{21} & c_{22} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} w_{t-1} \\ y_{t-1} \end{bmatrix}$$

para obtener:

$$\begin{bmatrix} w_t \\ y_t \end{bmatrix} = \begin{bmatrix} c_{01} \\ g_{01} \end{bmatrix} - \begin{bmatrix} c_{21} & c_{22} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} w_{t-1} - w_{t-2} \\ y_{t-1} - y_{t-2} \end{bmatrix} + \left\{ \begin{bmatrix} c_{11} & c_{12} \\ g_{11} & g_{12} \end{bmatrix} + \begin{bmatrix} c_{21} & c_{22} \\ g_{21} & g_{22} \end{bmatrix} \right\} \begin{bmatrix} w_{t-1} \\ y_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

lo cual, en forma compacta, se puede escribir como:

$$x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \varphi x_{t-1} + \varepsilon_t \quad (10.12)$$

Este último corresponde a la ecuación (10.6). Al restar x_{t-1} a ambos lados de la ecuación (10.12) llegamos a la (10.7):

$$\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_0 x_{t-1} + \varepsilon_t \quad (10.13)$$

Donde $\zeta_0 = \varphi - I_2$. Finalmente, remplazando $\zeta_0 x_{t-1} = BA'x_{t-1} = Bz_{t-1}$ arribamos a la ecuación (10.9) del VEC:

$$\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + Bz_{t-1} + \varepsilon_t \quad (10.14)$$

El procedimiento es el mismo si adicionamos un término $I(0)$ exógeno, ya que este nuevo término no se alteraría en la transformación anterior. Al observar la expresión (10.9), vemos que en general la i -ésima ecuación tendrá términos de corrección de error de la forma:

$$\beta_{i1} z_{1t-1} + \beta_{i2} z_{2t-1} + \cdots + \beta_{iM} z_{Mt-1}$$

adicionales a los términos autorregresivos $\Delta x_{t-1}, \Delta x_{t-2}, \dots, \Delta x_{t-p+1}$. Sin embargo, cuando se trata de dos variables (w_t, y_t) como en este ejemplo, donde sólo es posible *una relación de cointegración* z_t , las cosas se simplifican y quedan las dos ecuaciones del VEC de la siguiente manera:

$$\begin{aligned} \Delta w_t &= c_{10} + \zeta_{1,11} \Delta w_{t-1} + \zeta_{1,12} \Delta y_{t-1} + \zeta_{2,11} \Delta w_{t-2} + \zeta_{2,12} \Delta y_{t-2} + \cdots \\ &\quad + \zeta_{p-1,11} \Delta w_{t-p+1} + \zeta_{p-1,12} \Delta y_{t-p+1} + \beta_1 z_{t-1} + \varepsilon_{1t} \\ \Delta y_t &= g_{10} + \zeta_{1,21} \Delta w_{t-1} + \zeta_{1,22} \Delta y_{t-1} + \zeta_{2,21} \Delta w_{t-2} + \zeta_{2,22} \Delta y_{t-2} + \cdots \\ &\quad + \zeta_{p-1,21} \Delta w_{t-p+1} + \zeta_{p-1,22} \Delta y_{t-p+1} + \beta_2 z_{t-1} + \varepsilon_{2t} \end{aligned} \quad (10.15)$$

Ejemplo

Mostramos cómo pasar de un VAR en niveles (10.5) con tres rezagos al VEC (10.9). Podemos escribir en forma compacta:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3} + \varepsilon_t$$

Sumamos y restamos $\alpha_3 x_{t-2}$ y obtenemos:

$$x_t = \alpha_0 + \alpha_1 x_{t-1} - \alpha_3 (x_{t-2} - x_{t-3}) + (\alpha_2 + \alpha_3) x_{t-2} + \varepsilon_t$$

A continuación sumamos y restamos $(\alpha_2 + \alpha_3)x_{t-1}$ y, así, obtenemos:

$$\begin{aligned}x_t &= \alpha_0 - (\alpha_2 + \alpha_3)(x_{t-1} - x_{t-2}) - \alpha_3(x_{t-2} - x_{t-3}) \\&\quad + [\alpha_1 + (\alpha_2 + \alpha_3)]x_{t-1} + \varepsilon_t\end{aligned}$$

Ello corresponde a la ecuación (10.6) y se reescribe con el operador de diferencias:

$$\begin{aligned}x_t &= \alpha_0 - (\alpha_2 + \alpha_3)\Delta x_{t-1} - \alpha_3\Delta x_{t-2} + (\alpha_1 + \alpha_2 + \alpha_3)x_{t-1} + \varepsilon_t \\&= \alpha_0 + \zeta_1\Delta x_{t-1} + \zeta_2\Delta x_{t-2} + \varphi x_{t-1} + \varepsilon_t\end{aligned}$$

De este modo, al restar x_{t-1} a ambos lados llegamos a (10.7):

$$\Delta x_t = \alpha_0 + \zeta_1\Delta x_{t-1} + \zeta_2\Delta x_{t-2} + \zeta_0x_{t-1} + \varepsilon_t$$

Finalmente, reemplazando $\zeta_0x_{t-1} = Bz_{t-1}$, llegamos al VEC (10.9):

$$\Delta x_t = \alpha_0 + \zeta_1\Delta x_{t-1} + \zeta_2\Delta x_{t-2} + Bz_{t-1} + \varepsilon_t$$

Vector autorregresivo en diferencias

Si bien un VAR en diferencias no es consistente con un sistema cointegrado (porque estaríamos omitiendo la variable relevante z_{t-1}) ni con un sistema de variables $I(0)$ (porque estaríamos omitiendo la variable relevante x_{t-1}), un VAR en niveles sí puede serlo (Hamilton, 1994, p. 579). Sin embargo, si tomamos diferencias en (10.5) obtenemos:

$$\Delta x_t = \alpha_1\Delta x_{t-1} + \alpha_2\Delta x_{t-2} + \cdots + \alpha_p\Delta x_{t-p} + \varepsilon_t - \varepsilon_{t-1} \quad (10.16)$$

De manera que un $\text{VAR}(p)$ en niveles puede escribirse como un $\text{VARMA}(p, 1)$ no invertible en diferencias, el cual puede ser compatible con un sistema cointegrado. Aunque la expresión (10.16) carece del vector de constantes, se sugiere incluirlo en la estimación.

Pruebas de causalidad en un vector de corrección de error

Engle y Granger (1987) demostraron que cuando hay cointegración entre variables, entonces también hay causalidad, al menos en una

dirección; esto es, cointegración implica causalidad (aunque no al contrario). En un VEC, como en (10.9), es necesario modificar un poco la forma como se hace la prueba de causalidad, ya que en el VEC aparece el término de corrección de error, el cual vincula todas las variables. Aunque la tradición ha sido la de utilizar series estacionarias para estas pruebas, no existe consenso acerca de este punto.

Sin embargo, algunos autores como Demetriades y Hussein (1996) sostienen que las pruebas estadísticas de causalidad no son válidas en VAR con variables no estacionarias a menos que las variables se encuentren cointegradas, lo que implica que las ecuaciones de un VAR en niveles pueden ser utilizadas para probar causalidad en un sistema cointegrado. Granger (1988) sugiere utilizar el término de cointegración en las pruebas de causalidad dentro de un VEC. Zapata y Rambaldi (1997) también prefieren probar causalidad utilizando un VEC.

Para ilustrar la prueba en un sistema cointegrado, volvamos al VEC de dos variables w_t , y_t en (10.15) y veamos la primera ecuación. Para que y_t no cause a w_t se requiere que los rezagos de Δy_t no contribuyan a explicar Δw_t y que el término z_{t-1} tampoco contribuya; esto es, se espera que $\beta_1 = 0$, lo cual implica que w_t no responde a desviaciones del equilibrio de largo plazo con y_t . En otras palabras, los coeficientes $\zeta_{i,12}$ para $i = 1, 2, \dots, p - 1$ y el coeficiente β_1 deben ser cero.

De manera similar, si tuviéramos tres variables, digamos a_t , b_t y c_t , y quisieramos investigar si c_t causa a b_t , eliminaríamos los rezagos de c_t en la ecuación de b_t junto con todos los términos de corrección de error que haya en esa ecuación. Para ver si a_t causa a b_t , eliminaríamos los rezagos de a_t junto con todos los términos de corrección de error en la ecuación de b_t . Procedemos de la misma forma para determinar causalidad en las otras ecuaciones hasta poder afirmar algo acerca de la dirección de causalidad entre todas las posibles parejas de variables. Las sumas de residuos al cuadrado de las ecuaciones restringidas y no restringidas se pueden comparar a través de pruebas F o chi-cuadrado χ^2 , como se indicó en el Capítulo 6.

Toda y Yamamoto (1995) extendieron la prueba convencional de causalidad, descrita en el Capítulo 6, para cuando se trate de variables no estacionarias o integradas de orden diferente. La extensión consiste en aumentar el número de rezagos del VAR en d , donde d es el máximo

orden de integración de las variables del VAR; en otras palabras, si p es el número adecuado de rezagos (según AIC, SC, etc.) para el VAR en niveles, entonces se corre con $p + d$ rezagos. La prueba se realiza sobre los p rezagos (pasando por alto los d rezagos) aplicando un estadístico de Wald, el cual tiene validez asintótica.

Estimación por máxima verosimilitud

Sea x_t un vector de m variables $I(1)$ el cual se puede expresar como en la ecuación (10.7), la cual se repite aquí por conveniencia:

$$\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_2 \Delta x_{t-2} + \cdots + \zeta_{p-1} \Delta x_{t-p+1} + \zeta_0 x_{t-1} + \varepsilon_t$$

Donde ε_t es un vector de ruido blanco gaussiano con $E\varepsilon_t = 0$ y $E\varepsilon_t \varepsilon'_t = \Sigma$ para $t = \tau$, y cero para $t \neq \tau$.

Como se discutió anteriormente, si entre las m variables de x_t existen $M < m$ relaciones independientes de cointegración, ζ_0 se puede descomponer en dos matrices, ambas de rango M , tal que $\zeta_0 = BA'$, donde B es de orden $m \times M$ y A' es de orden $M \times m$.

La verosimilitud de x_1, \dots, x_n se puede escribir como se muestra a continuación, condicional a las p observaciones $x_{-p+1}, x_{-p+2}, \dots, x_0$, que se suponen conocidas (también se podría escribir desde $t = p + 1$ hasta n condicional a x_1, \dots, x_p):

$$\begin{aligned} l(\alpha_0, \zeta_1, \zeta_2, \dots, \zeta_{p-1}, \zeta_0, \Sigma) = & -\frac{nm}{2} \log 2\pi - \frac{n}{2} \log \det \Sigma \\ & - \frac{1}{2} \sum_{t=1}^n (\Delta x_t - \alpha_0 - \zeta_1 \Delta x_{t-1} - \zeta_2 \Delta x_{t-2} - \cdots - \zeta_{p-1} \Delta x_{t-p+1} \\ & - \zeta_0 x_{t-1})' \Sigma^{-1} (\Delta x_t - \alpha_0 - \zeta_1 \Delta x_{t-1} \\ & - \zeta_2 \Delta x_{t-2} - \cdots - \zeta_{p-1} \Delta x_{t-p+1} - \zeta_0 x_{t-1}) \end{aligned} \quad (10.17)$$

Donde el objeto es maximizar l escogiendo $\alpha_0, \zeta_1, \zeta_2, \dots, \zeta_{p-1}, \zeta_0, \Sigma$, sujeto a la condición $\zeta_0 = BA'$, donde B es de orden $m \times M$ y A' es de orden $M \times m$, y el rango de ζ_0 , B y A es M .

Podríamos estimar válidamente el VAR en (10.7) sin sujetarnos a la condición $\zeta_0 = BA'$, y obtendríamos $\hat{\zeta}_0$, lo cual estaría bien si, por ejemplo, nuestro objetivo es la predicción más que determinar

las expresiones de equilibrio de largo plazo. Sin embargo, cuando hay interés en descubrir y estudiar estos equilibrios de cointegración, es necesario utilizar métodos de estimación que separen ζ_0 en las matrices B y A' , donde A representa estas relaciones de equilibrio, y B , el peso que cada relación tiene como variable explicativa dentro del VAR.

Existen varios métodos para estimar el rango $M < m$ de la matriz ζ_0 en (10.7) y, por lo tanto, el número de relaciones de cointegración. Para este fin, algunos autores trabajan directamente con la matriz estimada $\hat{\zeta}_0$ para tratar de establecer su rango o con la matriz simétrica $\hat{\zeta}_0'\hat{\zeta}_0$, cuyo rango es igual al de ζ_0 , y cuyos eigenvalores son reales, más sencillos de calcular, y generan eigenvectores ortogonales.

El método de Johansen, presentado a continuación, estima los eigenvalores de una versión normalizada de la matriz $\zeta_0'\zeta_0$ y tiene la ventaja de garantizar que las distribuciones de probabilidad para las pruebas de hipótesis necesarias sean simples y manejables, por lo cual estas distribuciones resultan ser generalizaciones de las distribuciones de la prueba ADF para raíz unitaria (Johansen, 1991).

Algoritmo de Johansen

El algoritmo de Johansen (1988 y 1995) es una de las formas más populares de realizar la optimización de la función de verosimilitud logarítmica (10.17) sujeta a $\zeta_0 = BA'$. En esta exposición seguiremos parcialmente a Hamilton (1994, capítulo 20). El primer paso es simplificar la verosimilitud l y dejarla en función de un menor número de parámetros, lo que se conoce como *concentrar la verosimilitud*. Para tal fin, supondremos inicialmente que ζ_0 y Σ están dados y maximizamos l con respecto a los demás parámetros; de esta manera obtenemos una nueva verosimilitud cuyos parámetros quedarán en función de ζ_0 y Σ . Por ejemplo, dado ζ_0 , los residuos del último término de (10.17) se podrán reescribir como:

$$(\Delta x_t - \zeta_0 x_{t-1}) - \alpha_0^* - \zeta_1^* \Delta x_{t-1} - \zeta_2^* \Delta x_{t-2} - \cdots - \zeta_{p-1}^* \Delta x_{t-p+1} \quad (10.18)$$

Donde los parámetros marcados con asterisco resultan de la concentración y están en función de ζ_0 . Estos residuos tienen media cero y no están correlacionados con las variables explicativas $\Delta x_{t-1}, \dots, \Delta x_{t-p+1}$.

La expresión (10.18) se puede reescribir de manera que sea aún más fácil trabajarla. Para ello se estiman dos regresiones auxiliares, cuyos residuos habrán de desempeñar un papel importante en el desarrollo que sigue. La primera expresa Δx_t en función de una constante y de $\Delta x_{t-1}, \dots, \Delta x_{t-p+1}$:

$$\Delta x_t = \hat{\pi}_0 + \hat{\pi}_1 \Delta x_{t-1} + \hat{\pi}_2 \Delta x_{t-2} + \dots + \hat{\pi}_{p-1} \Delta x_{t-p+1} + \hat{u}_t \quad (10.19)$$

Donde $\hat{\pi}_0$ es un vector $m \times 1$ de constantes y las demás $\hat{\pi}_i$ son matrices de orden $m \times m$; entre tanto, \hat{u}_t es un vector de residuos $m \times 1$. La segunda expresa a x_{t-1} en función de una constante y de las mismas variables $\Delta x_{t-1}, \dots, \Delta x_{t-p+1}$:

$$x_{t-1} = \hat{\psi}_0 + \hat{\psi}_1 \Delta x_{t-1} + \hat{\psi}_2 \Delta x_{t-2} + \dots + \hat{\psi}_{p-1} \Delta x_{t-p+1} + \hat{v}_t \quad (10.20)$$

Donde $\hat{\psi}_0$ es un vector $m \times 1$ de constantes y las demás $\hat{\psi}_i$ son matrices de $m \times m$ y \hat{v}_t es un vector de residuos $m \times 1$. No debe esperarse gran significancia en la regresión anterior porque se trata de una variable $I(1)$ como dependiente y es explicada por variables $I(0)$; en cualquier caso \hat{v}_t será parecida a x_{t-1} .

Obsérvese que en el VAR (10.7), la matriz ζ_0 refleja el grado de correlación entre Δx_t y x_{t-1} , *neto del efecto de los demás rezagos en diferencias*. Nótese que en las regresiones auxiliares anteriores (10.19) y (10.20), el residuo \hat{u}_t contiene Δx_t y el residuo \hat{v}_t contiene a x_{t-1} , *netos* de los rezagos $\Delta x_{t-1}, \dots, \Delta x_{t-p+1}$. Con estos residuos \hat{u}_t y \hat{v}_t procedemos entonces a reescribir la expresión (10.18) así:

$$\begin{aligned} \hat{u}_t - \zeta_0 \hat{v}_t &= (\Delta x_t - \hat{\pi}_0 - \hat{\pi}_1 \Delta x_{t-1} - \hat{\pi}_2 \Delta x_{t-2} - \dots \\ &\quad - \hat{\pi}_{p-1} \Delta x_{t-p+1}) \\ &\quad - \zeta_0 (x_{t-1} - \hat{\psi}_0 - \hat{\psi}_1 \Delta x_{t-1} - \hat{\psi}_2 \Delta x_{t-2} - \dots \\ &\quad - \hat{\psi}_{p-1} \Delta x_{t-p+1}) \\ &= (\Delta x_t - \zeta_0 x_{t-1}) - (\hat{\pi}_0 - \zeta_0 \hat{\psi}_0) - (\hat{\pi}_1 - \zeta_0 \hat{\psi}_1) \Delta x_{t-1} \\ &\quad - (\hat{\pi}_2 - \zeta_0 \hat{\psi}_2) \Delta x_{t-2} - \dots - (\hat{\pi}_{p-1} - \zeta_0 \hat{\psi}_{p-1}) \Delta x_{t-p+1} \end{aligned}$$

Donde son evidentes los términos correspondientes con la ecuación (10.18); por ejemplo, $\hat{\pi}_0 - \zeta_0 \hat{\psi}_0 = \alpha_0^*$, $\hat{\pi}_1 - \zeta_0 \hat{\psi}_1 = \zeta_1^*$, etc. Al reemplazar

en (10.17), obtenemos la verosimilitud concentrada como:

$$\begin{aligned} l(\zeta_0, \Sigma) &= l(\alpha_0^*, \zeta_1^*, \zeta_2^*, \dots, \zeta_{p-1}^*, \zeta_0, \Sigma) \\ &= -\frac{nm}{2} \log 2\pi - \frac{n}{2} \log(\det \Sigma) \\ &\quad - \frac{1}{2} \sum_{t=1}^n (\hat{u}_t - \zeta_0 \hat{v}_t)' \Sigma^{-1} (\hat{u}_t - \zeta_0 \hat{v}_t) \end{aligned} \quad (10.21)$$

El asterisco indica parámetros en función de aquellos sin asterisco. La verosimilitud (10.21) se puede optimizar con respecto a ζ_0 y Σ .

Para concentrar todavía más la verosimilitud y dejarla sólo en función de ζ_0 , recordamos que en la estimación por máxima verosimilitud la matriz $m \times m$ de varianza covarianza que optimiza $l(\zeta_0, \Sigma)$ dado ζ_0 , es:

$$\hat{\Sigma}^*(\zeta_0) = \frac{1}{n} \sum_{t=1}^n (\hat{u}_t - \zeta_0 \hat{v}_t)(\hat{u}_t - \zeta_0 \hat{v}_t)' \quad (10.22)$$

Al reemplazar (10.22) en (10.21) intuimos (y se puede demostrar) que los términos $\hat{u}_t - \zeta_0 \hat{v}_t$ de la sumatoria en (10.21) desaparecen (se cancelan) y queda (10.21) como:

$$\begin{aligned} l(\zeta_0) &= l(\alpha_0^*, \zeta_1^*, \zeta_2^*, \dots, \zeta_{p-1}^*, \hat{\Sigma}^*, \zeta_0) \\ &= -\frac{nm}{2} \log 2\pi - \frac{n}{2} \log(\det \hat{\Sigma}^*) - \frac{nm}{2} \\ &= -\frac{nm}{2} \log 2\pi - \frac{nm}{2} \\ &\quad - \frac{n}{2} \log \left(\det \left[\frac{1}{n} \sum_{t=1}^n (\hat{u}_t - \zeta_0 \hat{v}_t)(\hat{u}_t - \zeta_0 \hat{v}_t)' \right] \right) \end{aligned} \quad (10.23)$$

Esta expresión, como puede observarse, se maximiza minimizando el término:

$$\det \left[\frac{1}{n} \sum_{t=1}^n (\hat{u}_t - \zeta_0 \hat{v}_t)(\hat{u}_t - \zeta_0 \hat{v}_t)' \right] \quad (10.24)$$

con respecto a ζ_0 y sujeto a la restricción de que hay M relaciones de cointegración, esto es, que se puede escribir $\zeta_0 = BA'$.

La expresión $\hat{u} - \zeta_0 \hat{v}$ es un vector donde el i -ésimo elemento representa los residuos de una regresión de \hat{u}_{it} (que contiene Δx_{it}) como

variable dependiente contra los elementos de \hat{v}_t (que contienen x_{t-1}) como variables independientes. La expresión (10.24), sin tomar aún el determinante, es la matriz de varianza covarianza de estos residuos y, en particular, su diagonal muestra los residuos al cuadrado promedio de la regresión correspondiente.

En consecuencia, intuimos que para minimizar (10.24) debemos lograr regresiones con la mayor variación explicada, pero como \hat{u}_{it} contiene a Δx_{it} , que es $I(0)$, y \hat{v}_t contiene las variables x_{t-1} , que son $I(1)$, sólo nos sirven las M relaciones de cointegración, que son $I(0)$. Lo que sigue ahora es estimar esas M relaciones de cointegración minimizando el determinante (10.24).

Aunque poco intuitivo, se encuentra que el método de correlación canónica es apropiado para manejar este problema de estimación. La correlación canónica es un método multivariado propuesto por Hotelling (1935), comparable con otros métodos como el análisis de factores o componentes principales, útil para estudiar todas las posibles correlaciones entre combinaciones lineales de variables, especialmente cuando hay múltiples variables dependientes, como en el caso que nos ocupa.

Con la ayuda de este método simplificaremos la matriz en (10.24) y la vamos a convertir, como se verá a continuación, en una matriz diagonal cuyo determinante (como en toda matriz diagonal) se calcula fácilmente multiplicando los términos de la diagonal, y los términos de la diagonal estarán entre cero y uno por construcción.

El método de correlación canónica transforma \hat{u}_t y \hat{v}_t a través de las combinaciones lineales:

$$\Gamma \hat{u}_t = \eta_t, \quad \Pi \hat{v}_t = \xi_t$$

Donde Γ y Π son matrices $m \times m$, y η_t y ξ_t son vectores $m \times 1$ cuyos elementos tienen media cero (porque \hat{u}_t y \hat{v}_t tienen media cero) y las matrices de varianza covarianza resultan unitarias por construcción:

$$E(\eta_t \eta_t') = I_m, \quad E(\xi_t \xi_t') = I_m$$

Los coeficientes de las combinaciones lineales entre \hat{u}_t y \hat{v}_t , esto es, las matrices Γ y Π , se escogen de manera que la combinación i -ésima η_i tenga correlación positiva r_i con la combinación i -ésima ξ_i : pero

correlación cero con las demás combinaciones. El método construye las combinaciones lineales de manera que la correlación de la primera pareja η_1, ξ_1 sea mayor que para la segunda η_2, ξ_2 y ésta, a su vez, mayor que para la tercera η_3, ξ_3 , y así sucesivamente. En consecuencia, la matriz de correlaciones entre η_t y ξ_t será diagonal, como se muestra a continuación:

$$E(\eta_t \xi'_t) = \begin{bmatrix} r_1 & 0 & \cdots & 0 \\ 0 & r_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & r_m \end{bmatrix} \quad (10.25)$$

donde $1 > r_1 \geq \cdots \geq r_m \geq 0$.

Las filas de la matriz Π que realiza la transformación de \hat{v}_t se obtienen de los m eigenvectores (vectores característicos) de una matriz de orden $m \times m$ construida de manera que parezca un gran coeficiente de correlación al cuadrado entre \hat{u}_t y \hat{v}_t . Veremos más adelante que algunos de estos eigenvectores servirán para construir los vectores de cointegración que estamos buscando. La expresión de esta matriz es:

$$\hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vu} \hat{\Sigma}_{uu}^{-1} \hat{\Sigma}_{uv} \quad (10.26)$$

Donde:

$$\hat{\Sigma}_{vv} = \frac{1}{n} \sum_{t=1}^n \hat{v}_t \hat{v}'_t, \quad \hat{\Sigma}_{uu} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t \hat{u}'_t, \quad \hat{\Sigma}_{uv} = \frac{1}{n} \sum_{t=1}^n \hat{u}_t \hat{v}'_t = \hat{\Sigma}'_{vu}.$$

De manera similar, las filas de la matriz Γ que realiza la transformación de \hat{u}_t se obtienen de los eigenvectores de otra matriz, cuya expresión es $\hat{\Sigma}_{uu}^{-1} \hat{\Sigma}_{uv} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{vu}$.

El siguiente paso es reemplazar los términos $\hat{u}_t = \Gamma^{-1} \eta_t$ y $\hat{v}_t = \Pi^{-1} \xi_t$ en la expresión (10.24), con lo cual, después de cierta álgebra, se obtiene:

$$\Psi \det \left[\frac{1}{n} \sum_{t=1}^n (\eta_t - \Phi \xi_t)(\eta_t - \Phi \xi_t)' \right] \quad (10.27)$$

Donde $\Phi = \Gamma \zeta_0 \Pi^{-1}$ y es una matriz $m \times m$. El término $\Psi = \det \hat{\Sigma}_{uu}$ es una constante de proporcionalidad que resulta de factorizar $\Gamma^{-1} \Gamma'^{-1}$, expresión que puede escribirse como $\Gamma^{-1} \Gamma'^{-1} = \Gamma^{-1} E(\eta \eta') \Gamma'^{-1} =$

$E(\Gamma^{-1}\eta\eta'\Gamma'^{-1}) = \Sigma_{uu}$. De manera similar a la discusión en torno a (10.24), podemos decir que la expresión $\eta_t - \Phi\xi_t$ representa el vector de residuos de la regresión de η_{it} como variable dependiente contra ξ_t como vector de variables independientes. Los elementos de la diagonal de la matriz en (10.27) contienen términos de la forma $(\eta_{it} - \Phi_i\xi_t)^2$, donde Φ_i representa la i -ésima fila de Φ , mientras que los elementos fuera de la diagonal contienen términos de la forma $(\eta_{it} - \Phi_i\xi_t)(\eta_{jt} - \Phi_j\xi_t)$, para $i \neq j$.

En la regresión $\eta_{it} - \Phi_i\xi_t$, la única variable explicativa que tiene correlación con η_{it} será ξ_{it} , como puede verse de (10.25), de manera que será la única variable del vector ξ_t que tendrá un coeficiente diferente de cero dentro de Φ_i . Por construcción, el i -ésimo elemento de η_t no está correlacionado con el j -ésimo elemento de ξ_t , de manera que la matriz en (10.27) será diagonal y mostrará en su diagonal la varianza del residuo (el residuo al cuadrado promedio) para la regresión $\eta_{it} - \Phi_i\xi_t$.

Nótese que el coeficiente para la regresión $\eta_{it} - \Phi_i\xi_t$, que corresponde a la posición i -ésima en la diagonal, será:

$$\frac{\sum \eta_{it}\xi_{it}}{\sum \xi_{it}^2} = \frac{\sum \eta_{it}\xi_{it}}{n} = r_i$$

Donde se utiliza el hecho de que las varianzas de ξ_t y η_t son unitarias ($\sum \xi_{it}^2/n = 1$). Con esto podemos calcular el residuo al cuadrado promedio, que ocupará la posición i -ésima de la diagonal, como:

$$\begin{aligned} \frac{1}{n} \sum (\eta_{it} - r_i\xi_{it})^2 &= \frac{1}{n} (\eta_{it}^2 - 2r_i\eta_{it}\xi_{it} + r_i^2\xi_{it}^2) \\ &= 1 - 2r_i^2 + r_i^2 = 1 - r_i^2 \end{aligned} \quad (10.28)$$

Para los elementos por fuera de la diagonal, digamos el elemento (i,j) , tendremos:

$$\frac{1}{n} \sum (\eta_{it} - r_i\xi_{it})(\eta_{jt} - r_j\xi_{jt})$$

Esto, como ya se dijo, valdrá cero porque todos los productos cruzados resultantes de la expresión anterior tienen correlación cero.

En resumen, el término $(\eta_t - \Phi\xi_t)(\eta_t - \Phi\xi_t)'$ de la expresión (10.27) se convierte en:

$$\begin{bmatrix} \eta_{1t} - r_1\xi_{1t} \\ \eta_{2t} - r_2\xi_{2t} \\ \vdots \\ \eta_{mt} - r_m\xi_{mt} \end{bmatrix} \begin{bmatrix} \eta_{1t} - r_1\xi_{1t} & \eta_{2t} - r_2\xi_{2t} & \cdots & \eta_{mt} - r_m\xi_{mt} \end{bmatrix}$$

Recordamos que nuestra misión es minimizar (10.27) manteniendo la restricción $\zeta_0 = BA'$, la cual conlleva la existencia de M relaciones de cointegración. En otras palabras, la restricción $\zeta_0 = BA'$ implica que ζ_0 podrá (y deberá) descomponerse en dos matrices, la primera de orden $m \times M$ y la segunda $M \times m$. Al observar el término $\Phi = \Gamma\zeta_0\Pi^{-1}$ en (10.27) y al notar que ambas Γ y Π^{-1} son $m \times m$, vemos que ζ_0 podrá descomponerse en estas dos matrices si y sólo si Φ también puede descomponerse en dos matrices, de orden $m \times M$ la primera y $M \times m$ la segunda.

Por el orden que tiene esta segunda matriz, sólo podremos escoger M combinaciones lineales de los m elementos disponibles en ξ_t para correlacionar con η_t . No obstante, por construcción, y de la discusión anterior, sabemos que los primeros elementos de ξ_t , individualmente y no en combinación, son los que tienen la mayor correlación con los elementos correspondientes de η_t ; y como nuestro objetivo es minimizar la expresión (10.27) minimizando los residuos al cuadrado $(\eta_{it} - \Phi_i\xi_{it})^2$ escogeremos, en consecuencia, los primeros M elementos de ξ_t como variables explicativas.²

Los primeros M elementos de ξ_t tendrán correlación r_i con los primeros M elementos de η_t ; entonces para $i \leq M$ los residuos serán $\eta_{it} - r_i\xi_{it}$ mientras que para $M < i \leq m$ los residuos serán simplemente η_{it} , ya que la correlación es cero. En (10.27) se escriben primero estos M elementos de η_t y ξ_t , seguidos por los otros $m - M$ elementos. En consecuencia, los primeros M elementos de la diagonal corresponderán a las regresiones con menor residuo al cuadrado promedio (o mayor variación explicada), seguidos por $m - M$ elementos con mayor residuo, ya que no hay correlación entre los últimos $m - M$ elementos de η_t y los primeros M elementos de ξ_t que fueron incluidos como regresores.

Como se vio en (10.28), los elementos de la diagonal corresponden en general a $1 - r_i^2$. Por lo tanto, los primeros M elementos pueden

escribirse como $1 - r_i^2$, mientras que los últimos $m - M$ elementos serán uno, porque para estos el r_i^2 es cero. Con estas consideraciones, la matriz en (10.27) toma una forma diagonal y su determinante se reduce al producto de los elementos de la diagonal, como se muestra a continuación:

$$\begin{aligned} \Psi \det & \left[\frac{1}{n} \sum_{t=1}^n (\eta_t - \Phi \xi_t)(\eta_t - \Phi \xi_t)' \right] \\ & = \Psi \det \begin{bmatrix} 1 - r_1^2 & \dots & & \dots & 0 \\ 0 & 1 - r_2^2 & & & \vdots \\ \vdots & & \ddots & & \\ & & & 1 - r_M^2 & \\ & & & & 1 & \vdots \\ \vdots & & & & & \ddots & 0 \\ 0 & \dots & & \dots & 0 & 1 \end{bmatrix} \\ & = \Psi \prod_{i=1}^M (1 - r_i^2) \quad (10.29) \end{aligned}$$

De la técnica de correlación canónica se encuentra que los r_i^2 son iguales a los eigenvalores (valores característicos) λ_i de la matriz (10.26), los cuales estarán entre cero y uno (véase Hamilton, 1994, p. 653), esto es, los que salen de la expresión del problema típico de eigenvalores $\det(\hat{\Sigma}_{vv}^{-1}\hat{\Sigma}_{vu}\hat{\Sigma}_{uu}^{-1}\hat{\Sigma}_{uv} - \lambda I_m) = 0$. En resumen, tenemos que:

$$\begin{bmatrix} r_1^2 & 0 & \dots & 0 \\ 0 & r_2^2 & & \\ \vdots & \ddots & & \\ 0 & \dots & r_m^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \\ \vdots & & \ddots & \\ 0 & & \dots & \lambda_m \end{bmatrix}.$$

Donde los eigenvalores se han ordenado de mayor a menor de manera que $1 > \lambda_1 > \lambda_2 > \dots > \lambda_m \geq 0$.

El número M de correlaciones diferentes de cero será igual al número M de eigenvalores diferentes de cero de la matriz (10.26) (e igual a su rango). Recordamos que los primeros M elementos de ξ_t se formaron (por el método de correlación canónica) multiplicando las primeras M

filas de la matriz Π por \hat{v}_t (que esencialmente es x_{t-1}) y que estas filas, a su vez, corresponden a los primeros M eigenvectores de la matriz (10.26). En otras palabras, estos M eigenvectores (después de cierta normalización) serán los M vectores de cointegración que forman la matriz A .

Como los λ_i resultan iguales a los r_i^2 , entonces podemos usar $1 - \lambda_i$ para remplazar las expresiones $1 - r_i^2$ en (10.29). Sólo contribuirán a maximizar la verosimilitud (10.17) o, lo que es lo mismo, a minimizar la expresión (10.24) o (10.29), los λ_i diferentes de cero, que son M . Al volver a la expresión (10.23), tomando logaritmos de (10.29), reordenando términos, y escribiéndola en función de los eigenvalores estimados, se tiene el siguiente valor *máximo* de la verosimilitud concentrada sujeta a la restricción de que existen M relaciones de cointegración:

$$l^*(\lambda_i) = -\frac{nm}{2} \log 2\pi - \frac{nm}{2} - \frac{n}{2} \log \Psi - \frac{n}{2} \sum_{i=1}^M \log(1 - \hat{\lambda}_i) \quad (10.30)$$

En consecuencia, al observar el último término en (10.30) vemos que la verosimilitud depende de los términos $1 - \hat{\lambda}_i$. Como en general hay m eigenvalores λ_i y como se pueden ordenar de mayor a menor, $1 > \lambda_1 > \lambda_2 > \dots > \lambda_m \geq 0$, entonces los primeros λ_i , más cercanos a uno, elevarán la verosimilitud de manera significativa, porque cuando:

$$\lambda_i \rightarrow 1 \Rightarrow \log(1 - \lambda_i) \rightarrow -\infty$$

Ello hace que el último término de (10.30) sea positivo por estar precedido de un signo negativo. Para maximizar la verosimilitud basta con encontrar el número M de eigenvalores diferentes de cero.

En otras palabras, las consideraciones sobre el término $\sum \log(1 - \hat{\lambda}_i)$ son suficientes para establecer el número M de relaciones de cointegración, que será igual al número de eigenvalores λ_i significativos de la matriz (10.26). De estos eigenvalores significativos se obtienen los M vectores de cointegración, o por lo menos el espacio vectorial generado por estos vectores de cointegración. Obviamente, en la práctica, se tomarán como eigenvalores diferentes de cero aquellos que sean *estadísticamente* diferentes de cero. Esta es la base de las pruebas estadísticas de cointegración descritas más adelante.

filas de la matriz II por \hat{v}_t (que esencialmente es x_{t-1}) y que estas filas, a su vez, corresponden a los primeros M eigenvectores de la matriz (10.26). En otras palabras, estos M eigenvectores (después de cierta normalización) serán los M vectores de cointegración que forman la matriz A .

Como los λ_i resultan iguales a los r_i^2 , entonces podemos usar $1 - \lambda_i$ para remplazar las expresiones $1 - r_i^2$ en (10.29). Sólo contribuirán a maximizar la verosimilitud (10.17) o, lo que es lo mismo, a minimizar la expresión (10.24) o (10.29), los λ_i diferentes de cero, que son M . Al volver a la expresión (10.23), tomando logaritmos de (10.29), reordenando términos, y escribiéndola en función de los eigenvalores estimados, se tiene el siguiente valor *máximo* de la verosimilitud concentrada sujeta a la restricción de que existen M relaciones de cointegración:

$$l^*(\lambda_i) = -\frac{nm}{2} \log 2\pi - \frac{nm}{2} - \frac{n}{2} \log \Psi - \frac{n}{2} \sum_{i=1}^M \log(1 - \hat{\lambda}_i) \quad (10.30)$$

En consecuencia, al observar el último término en (10.30) vemos que la verosimilitud depende de los términos $1 - \hat{\lambda}_i$. Como en general hay m eigenvalores λ_i y como se pueden ordenar de mayor a menor, $1 > \lambda_1 > \lambda_2 > \dots > \lambda_m \geq 0$, entonces los primeros λ_i , más cercanos a uno, elevarán la verosimilitud de manera significativa, porque cuando:

$$\lambda_i \rightarrow 1 \Rightarrow \log(1 - \lambda_i) \rightarrow -\infty$$

Ello hace que el último término de (10.30) sea positivo por estar precedido de un signo negativo. Para maximizar la verosimilitud basta con encontrar el número M de eigenvalores diferentes de cero.

En otras palabras, las consideraciones sobre el término $\sum \log(1 - \hat{\lambda}_i)$ son suficientes para establecer el número M de relaciones de cointegración, que será igual al número de eigenvalores λ_i significativos de la matriz (10.26). De estos eigenvalores significativos se obtienen los M vectores de cointegración, o por lo menos el espacio vectorial generado por estos vectores de cointegración. Obviamente, en la práctica, se tomarán como eigenvalores diferentes de cero aquellos que sean *estadísticamente* diferentes de cero. Esta es la base de las pruebas estadísticas de cointegración descritas más adelante.

Estimativos de los parámetros del vector autorregresivo

Sean $\hat{a}_1, \dots, \hat{a}_M$ los eigenvectores de la matriz (10.26) asociados con los M mayores eigenvalores significativos; éstos forman una base para el espacio de vectores de cointegración. Como se discutió antes, cualquier vector de cointegración puede escribirse en función de ellos:

$$b_1\hat{a}_1 + b_2\hat{a}_2 + \cdots + b_M\hat{a}_M \quad (10.31)$$

Donde b_1, \dots, b_M son escalares. En el método de Johansen estos vectores \hat{a}_i se normalizan de manera que $\hat{a}_i'\hat{\Sigma}_{vv}\hat{a}_i = 1$. Una vez normalizados, se reúnen los eigenvectores en la matriz $\hat{A} = (\hat{a}_1, \dots, \hat{a}_M)$. Ya teniendo la matriz \hat{A} (y recordando que $\zeta_0\hat{v}_t = BA'\hat{v}_t$) podemos correr una regresión de \hat{u}_t contra $\hat{A}'\hat{v}_t$ y obtener la estimación de B . Esta estimación \hat{B} estará dada por:

$$\hat{B} = \hat{\Sigma}_{uv}\hat{A} \quad (10.32)$$

La de $\zeta_0 = BA'$ por:

$$\hat{\zeta}_0 = \hat{B}\hat{A}' = \hat{\Sigma}_{uv}\hat{A}\hat{A}' \quad (10.33)$$

La estimación de máxima verosimilitud de los otros ζ_i estará dada por –véase la discusión que llevó a la expresión (10.21)–:

$$\hat{\zeta}_i = \hat{\pi}_i - \hat{\zeta}_0\hat{\psi}_i \quad (10.34)$$

El estimado de la constante α_0 por:

$$\hat{\alpha}_0 = \hat{\pi}_0 - \hat{\zeta}_0\hat{\psi}_0 \quad (10.35)$$

Y el estimador de máxima verosimilitud de la matriz de varianza covarianza del error, $\hat{\Sigma}$, como se mencionó en (10.22), estará dado por:

$$\hat{\Sigma} = \frac{1}{n} \sum_{t=1}^n (\hat{u}_t - \hat{\zeta}_0\hat{v}_t)(\hat{u}_t - \hat{\zeta}_0\hat{v}_t)' \quad (10.36)$$

Estadísticos de prueba

Volviendo a la ecuación (10.30) de la verosimilitud $l^*(\lambda_i)$, recordamos que los λ_i cercanos a uno contribuirán más a la maximización que aquellos cercanos a cero y que los hemos ordenado de mayor a menor.

La verosimilitud que llamaremos l_M^* , con la restricción de que sólo hay M combinaciones lineales de x_t como regresores, naturalmente será menor o igual a la verosimilitud *no restringida* que llamaremos l_m^* , donde puede haber hasta m combinaciones.

Si los λ_i más pequeños no contribuyen es porque sus $m - M$ relaciones lineales son insignificantes. En otras palabras, si:

$$l_m^* - l_M^* = -\frac{n}{2} \sum_{i=M+1}^m \log(1 - \hat{\lambda}_i) \quad (10.37)$$

es muy pequeño, es porque los λ_i que siguen después de los primeros M no contribuyen significativamente y habremos de aceptar que hay *máximo M relaciones de cointegración*. Una formalización de esta prueba de cointegración es del tipo *likelihood ratio* y se llama estadística de traza (*trace statistic*), λ_{traza} :

$$\lambda_{traza} = 2(l_m^* - l_M^*) = -n \sum_{i=M+1}^m \log(1 - \hat{\lambda}_i) \quad (10.38)$$

Si λ_{traza} excede el valor crítico, se rechaza la hipótesis de que hay *máximo M relaciones de cointegración independientes* (recuérdese que los λ 's están ordenados de mayor a menor). En la práctica, la prueba se hace de manera *escalonada*. Primero para $M = 0$; si esta hipótesis no se puede rechazar entonces se acepta que no hay cointegración. Si la hipótesis $M = 0$ se rechaza, entonces se intenta $M = 1$, y si ésta no se puede rechazar entonces se acepta que hay máximo una relación. Si se rechaza $M = 1$, entonces se prueba si hay máximo dos, $M = 2$, y así sucesivamente hasta aceptar la hipótesis de turno. Si llegamos a la hipótesis de que hay máximo $M = m - 1$ relaciones de cointegración y la rechazamos, tendríamos que aceptar $M = m$, lo cual es imposible en cointegración y probablemente indica que todas las variables son

$I(0)$. También existe otra prueba llamada λ_{max} , cuya expresión es:

$$\lambda_{max} = -n \log(1 - \hat{\lambda}_{M+1}) \quad (10.39)$$

Para su aplicación se utiliza el mismo procedimiento escalonado de λ_{traza} . Si λ_{max} excede el valor crítico es porque λ_{M+1} sí contribuye a mejorar la verosimilitud y se rechaza la hipótesis de que hay máximo M relaciones de cointegración independientes; luego se prueba con $M + 1$, y así sucesivamente.

Los valores críticos para ambas pruebas, λ_{traza} y λ_{max} , han sido tabulados por varios autores, entre ellos Johansen y Juselius (1990) y Osterwald (1992), y se encuentran incorporados en la mayoría de los programas de computador.

En la estimación también se pueden imponer y probar restricciones sobre los elementos de la matriz A de cointegración y sobre los coeficientes de ajuste de la matriz B ; por ejemplo, que algún coeficiente valga uno, o cero, o que sea igual a otro.

Cointegración en la práctica

En razón de que la cointegración investiga el comportamiento de largo plazo entre variables, es conveniente contar para su estimación con un horizonte de tiempo amplio. Pero qué es *largo plazo* depende del fenómeno bajo estudio. Por ejemplo, largo plazo puede ser horas o días cuando se trata de la transmisión de ciertas características genéticas, inmunológicas o de hábitos entre generaciones de una población de organismos unicelulares, o puede ser siglos para una población de organismos multicelulares. Para la mayoría de los fenómenos económicos estaríamos hablando en términos de meses o años. Tener muchas observaciones no es lo mismo que tener un adecuado cubrimiento temporal del comportamiento; diez mil observaciones parece una muestra grande pero puede ser una muestra inútil para establecer cointegración si cubre sólo unas pocas horas o días de algún comportamiento financiero, como el de las tasas de interés, la tasa de cambio o los precios de acciones. En cualquier caso, se percibe cierta contradicción entre la necesidad de que las observaciones cubran el largo plazo y el hecho muy probable de que en el largo plazo cambie la relación estructural de equilibrio que estamos tratando de identificar. De nuevo, como todo en

estadística y econometría, los modelos son aproximaciones a la realidad y sus resultados una guía, y se recomienda reestimarlos con alguna frecuencia, no sólo porque la estructura del sistema evoluciona sino, además, porque las entidades que producen las cifras las revisan con frecuencia.

Por lo anterior, y porque a veces las pruebas de raíz unitaria no son concluyentes, se acepta alguna flexibilidad en cuanto al grado de integración de las m variables bajo consideración. En la práctica no todas serán claramente de orden $I(1)$. Una $I(2)$ se puede diferenciar, y una que parezca estar cerca de $I(1)$, ya sea por arriba o por abajo, de todas maneras se podría incluir en el vector x_t , en especial si hay alguna justificación teórica para pensar que no es estacionaria. Para una discusión de cointegración con variables $I(2)$, véase Maddala y Kim (1998, capítulo 11), Enders (2004) o Johansen (2006).

Recordamos que un VEC comparte la misma filosofía de un VAR, esto es el énfasis en la forma reducida y en permitir que los datos, más que la teoría, sugieran las relaciones entre las variables. Al igual que en los VAR, es conveniente mirar los signos y la significancia de algunos coeficientes individuales con cierta flexibilidad o tolerancia en razón de la frecuente multicolinealidad entre las variables explicativas.

Si nuestro interés radica en hacer predicción y no en estimar las diferentes relaciones de cointegración, podemos correr un VAR en diferencias con las mismas especificaciones de la prueba de cointegración, pero adicionando el vector x_{t-1} como variable exógena; esto es, el VAR de la ecuación (10.7). Los coeficientes significativos serán similares en ambos VEC y VAR. Al comparar entre los dos, es necesario hacer los ajustes correspondientes; por ejemplo, la constante del VAR deberá ahora incluir tanto la constante del VEC como la constante del vector de cointegración multiplicada por el parámetro estimado de dicho vector. Ajustes similares se deben hacer con los demás coeficientes del vector x_{t-1} .

Finalmente, no debemos olvidar que la cointegración puede implementarse de manera sencilla, como en el Capítulo 9, en una sola ecuación sin que necesariamente tenga que estar incorporada en un sistema como el VEC, a veces menos flexible.

Ejemplo con dos variables

Los cuadros 10.1 y 10.2 son salidas de Eviews y muestran un ejemplo de cointegración con dos variables trimestrales: la tasa de interés de préstamos, llamada de colocación COLOC, y la inflación INF. En una etapa anterior, no mostrada aquí, se aplicaron las pruebas de raíz unitaria y se determinó que las variables pueden tratarse como $I(1)$. En el Cuadro 10.1 se hace la prueba de cointegración utilizando un VAR con cuatro rezagos, con constante en el vector de cointegración, pero sin constante (sin tendencia determinística) en el VAR. El estadístico λ_{traza} , identificado en la salida como *likelihood ratio*, rechaza al 5 % de nivel de significancia la hipótesis de que haya máximo cero relaciones de cointegración (*cointegrating equations [CE]*), pero no rechaza que haya máximo una relación.

Cuadro 10.1: Prueba de cointegración de Johansen

Date: 11/10/96 Time: 20:15				
Sample: 1991:1 1998:2				
Included observations: 30				
Test assumption: No deterministic trend in the data				
Series: COLOC INF				
Lags interval: 1 to 4				
Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.385664	20.96664	19.96	24.60	None *
0.190767	6.350070	9.24	12.97	At most
*(**) denotes rejection of the hypothesis at 5%(1%) significance level				
L.R. test indicates 1 cointegrating equation(s) at 5% significance level				
Unnormalized Cointegration Coefficients				
COLOC	INF	C		
-0.046633	0.020749	1.575102		
0.037603	0.024141	-2.042360		
Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)				
COLOC	INF	C		
1.000000	-0.444937	-33.77647		
	(0.25521)	(6.17760)		
Log likelihood	-107.5024			

En razón de que sólo puede haber máximo una relación con dos variables, se acepta que hay una. Los dos eigenvectores correspondientes a los dos eigenvalores (0.385664 y 0.190767) se muestran en el cuadro como *coeficientes de cointegración sin normalizar (unnormalized cointegrating coefficients)*. Al final del cuadro se muestra el eigenvector escogido de entre los anteriores, por ser el que corresponde al único eigenvalor significativo (0.385664). Éste se normaliza dividiendo los coeficientes del vector por -0.046633 y se presenta bajo el nombre *coeficientes de cointegración normalizados*. El vector de cointegración simplemente indica un equilibrio en el cual la tasa de interés de colocación es igual al 0.4449 por la inflación más 33.78.

El Cuadro 10.2 muestra el VEC, que debe tener las mismas especificaciones que se usaron en la prueba de cointegración, pues incluye el mismo vector de cointegración escogido en el Cuadro 10.1. Este se identifica en el Cuadro 10.2 como CointEq1, el cual corresponde al término $A'x_{t-1}$ en la ecuación (10.9). En la segunda parte del cuadro se muestran las dos ecuaciones del VEC, cada una en una columna, empezando por los coeficientes estimados para la relación de cointegración, CointEq1, que son -0.284419 y 0.091290, los cuales forman la matriz B (que en este caso es un vector porque $M = 1$). Siguen los demás coeficientes de los rezagos de las dos variables en diferencias, DCOLOC y DINF.

Ejemplo con tres variables

El siguiente ejemplo ilustra la ecuación cuantitativa del dinero. La Figura 10.1 grafica las series anuales (en logaritmos) de los medios de pago LM, el producto nacional LY y el índice de precios LP, cuyas cifras se muestran en el Cuadro 10.3. Las pruebas de raíz unitaria (ADF con constantes), no mostradas aquí, determinan que las variables pueden ser tratadas como $I(1)$, de manera que se puede proceder a probar la cointegración.

En el Cuadro 10.4 se realiza la prueba de cointegración empleando el estadístico λ_{traza} . Se rechaza la hipótesis de que haya cero relaciones de cointegración, pero no se puede rechazar la hipótesis de que haya máximo una relación de cointegración; en otras palabras, se encuentra que solamente uno de los tres eigenvalores (0.737604) es significativamente diferente de cero.

Cuadro 10.2: Vector de corrección de error estimado

Date: 11/10/96 Time: 20:17 Sample: 1991:1 1998:2 Included observations: 30 Standard errors & t-statistics in parentheses		
Cointegrating Eq:	CointEq1	
COLOC(-1)	1.000000	
INF(-1)	-0.444937 (0.25521) (-1.74341)	
C	-33.77647 (6.17760) (-5.46757)	
Error Correction	D(COLOC)	D(INF)
CointEq 1	-0.284419 (0.13240) (-2.14811)	0.091290 (0.05536) (1.64900)
D(COLOC(-1))	0.947002 (0.25836) (3.66550)	-0.033958 (0.10802) (-0.31435)
D(COLOC(-2))	1.59111 (0.29397) (0.54123)	0.093348 (0.12291) (0.75946)
D(COLOC(-3))	-0.421843 (0.29049) (-1.45217)	0.068137 (0.12146) (0.56098)
D(COLOC(-4))	0.390199 (0.28942) (1.34823)	-0.086195 (0.12101) (-0.71229)
D(INF(-1))	0.853287 (0.41681) (2.04717)	0.130947 (0.17428) (0.75136)
D(INF(-2))	0.106015 (0.44379) (0.23888)	-0.153368 (0.18556) (-0.82651)
D(INF(-3))	0.054748 (0.42417) (0.12907)	0.014106 (0.17736) (0.07954)
D(INF(-4))	-0.012499 (0.41702) (-0.02997)	-0.588574 (0.17436) (-3.26084)
R-squared	0.564644	0.364709
Adj. R-squared	0.398794	0.122693

En cualquier caso, los tres eigenvalores generan tres eigenvectores, que se muestran como coeficientes de cointegración sin normalizar en tres filas ordenadas de manera que correspondan a los eigenvalores de mayor a menor. Sabemos que con tres variables, uno o dos de estos eigenvectores pueden constituirse en vectores de cointegración. En caso de existir un solo vector de cointegración, sus coeficientes serían los mostrados normalizados como *1 Cointegrating Equation*, esto es, (1 - 1.45 - 0.86 12.2). Esta relación de cointegración dice que el logaritmo del dinero es igual a 1.45 por el logaritmo del producto nacional más 0.86 por el logaritmo del índice de precios menos una constante igual a 12.2.

En caso de existir dos vectores de cointegración, sus coeficientes serían los mostrados normalizados como *2 Cointegrating Equations*, donde en el primer vector se hace unitaria la variable LM y en el segundo se hace unitaria la variable LY. Aunque existen muchas alternativas para normalizar los vectores de cointegración, la alternativa particular utilizada por Eviews, donde se forma una matriz diagonal haciendo unitario el coeficiente de LM y cero el de LP en el primer vector, y viceversa para el otro vector, es la descrita por Phillips (1991). Por ejemplo, partimos de los primeros dos coeficientes de cointegración sin normalizar:

$$\begin{array}{ccc} -6.301427 & 9.144648 & 5.443426 \\ 6.191408 & -0.381224 & -6.838767 \end{array}$$

Y dividimos la primera fila por -6.301427 para obtener:

$$\begin{array}{ccc} 1 & -1.451203 & -0.863840 \\ 6.191408 & -0.381224 & -6.838767 \end{array}$$

Se multiplica la primera fila anterior por 6.191408 y se resta de la segunda fila:

$$\begin{array}{ccc} 1 & -1.451203 & -0.863840 \\ 0 & 8.603766 & -1.490381 \end{array}$$

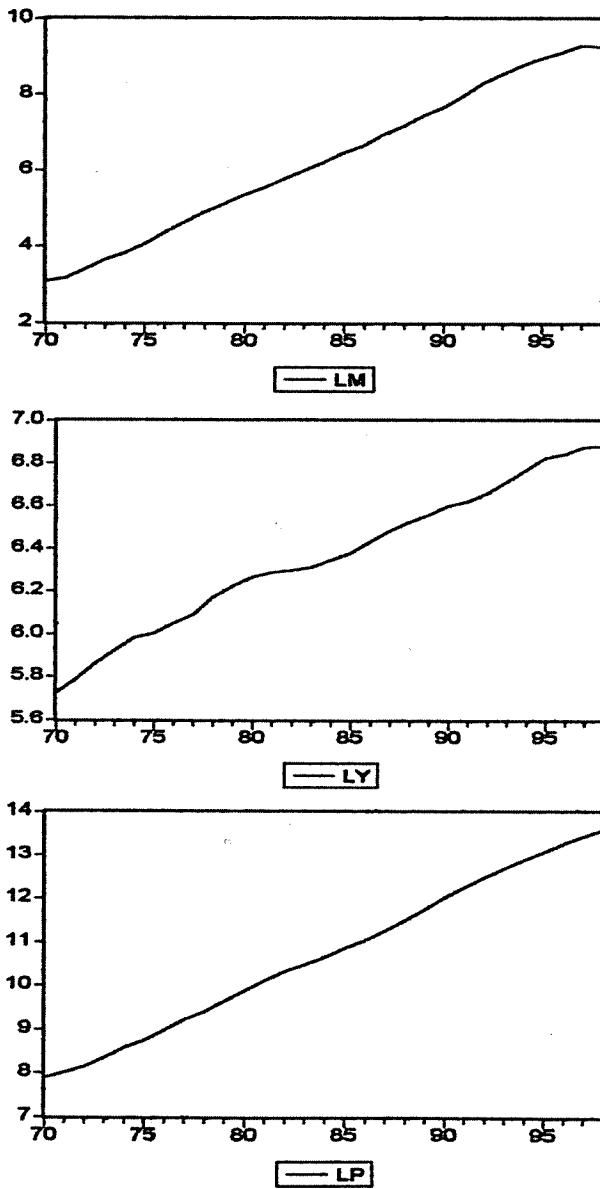
Se divide la segunda fila anterior por 8.603766:

$$\begin{array}{ccc} 1 & -1.451203 & -0.863840 \\ 0 & 1 & -0.173224 \end{array}$$

Se multiplica la segunda fila por -1.451203 y se resta de la primera:

$$\begin{array}{ccc} 1 & 0 & -1.115224 \\ 0 & 1 & -0.173224 \end{array}$$

Figura 10.1: Medios de pago, *PIB* y precios (cifras anuales en logaritmos)



Cuadro 10.3: Medios de pago, *PIB* y precios
 (cifras anuales en logaritmos)

Obs	LM	LY	LP
1970	3.091042	5.726848	7.887209
1971	3.178054	5.786897	8.014997
1972	3.401197	5.860786	8.146130
1973	3.663562	5.924256	8.357494
1974	3.828641	5.981414	8.588769
1975	4.077537	6.003887	8.751791
1976	4.369448	6.049733	8.980424
1977	4.644391	6.091310	9.230143
1978	4.905275	6.171701	9.402117
1979	5.123964	6.224558	9.655218
1980	5.361292	6.265301	9.886039
1981	5.560682	6.287859	10.11993
1982	5.774552	6.297109	10.33527
1983	6.006353	6.311735	10.48919
1984	6.216606	6.345636	10.65709
1985	6.464588	6.376727	10.85963
1986	6.669498	6.432940	11.04981
1987	6.954639	6.484635	11.26509
1988	7.184629	6.525030	11.51293
1989	7.439559	6.558198	11.74499
1990	7.669028	6.599870	12.02539
1991	7.952263	6.620073	12.26302
1992	8.297294	6.659294	12.48726
1993	8.541886	6.711740	12.69101
1994	8.767018	6.768493	12.89477
1995	8.951311	6.824374	13.07292
1996	9.104202	6.843750	13.26849
1997	9.300912	6.874198	13.43145
1998	9.261604	6.880384	13.58589

Cuadro 10.4: Prueba de cointegración de Johansen

Date: 04/29/99 Time: 15:43

Sample: 1970 1998

Included observations: 27

Test assumption: Linear deterministic trend in the data

Series: LM LY LP

Lags interval: 1 to 1

Eigenvalue	Likelihood Ratio	5 Percent Critical Value	1 Percent Critical Value	Hypothesized No. of CE(s)
0.737604	46.43850	29.68	35.65	None **
0.259753	10.31520	15.41	20.04	At most 1
0.078058	2.194378	3.76	6.65	At most 2

*(**) denotes rejection of the hypothesis at 5 %(1%) significance level

L.R. test indicates 1 cointegrating equation(s) at 5 % significance level

Unnormalized Cointegrating Coefficients

LM	LY	LP
-6.301427	9.144648	5.443426
6.191408	-0.381224	-6.838767
1.841503	3.096764	-2.745505

Normalized Cointegrating Coefficients: 1 Cointegrating Equation(s)

LM	LY	LP	C
1.000000	-1.451203 (0.18870)	-0.863840 (0.03434)	12.21604
Log likelihood			195.9462

Normalized Cointegrating Coefficients: 2 Cointegrating Equation(s)

LM	LY	LP	C
1.000000	0.000000	-1.115224 (0.00852)	5.670129
0.000000	1.000000	-0.173224 (0.00621)	-4.510677
Log likelihood			200.0066

Y, finalmente, se resta la media de la z_t correspondiente a cada ecuación anterior construida con los datos disponibles; esto es, a la primera ecuación se le resta la media de la serie $LM - 1.115224LP = -5.670129$ y a la segunda la media de la serie $LY - 0.173224LP = 4.510677$.

El VEC, mostrado en el Cuadro 10.5, tiene la misma especificación de la prueba de cointegración del Cuadro 10.4, esto es, un rezago, constante en el VAR y constante en el vector de cointegración. En el primer recuadro está la relación de cointegración, identificada como CointEq1, y en el segundo, en columnas, las ecuaciones correspondientes a cada variable. Por ejemplo, el término de corrección de error para la ecuación en diferencias $D(LM)$ es 0.697CointEq1 , el de la segunda ecuación $D(LY)$ es 0.425CointEq1 y el de la tercera $D(LP)$ es -0.23CointEq1 , donde el término CointEq1 se remplaza por la relación de cointegración.

Los demás coeficientes en cada columna corresponden a rezagos de las variables en diferencia, como en la ecuación (10.9). Si hubiera dos vectores de cointegración, éstos se denominarían CointEq1 y CointEq2. Todas las variables dentro del vector de cointegración parecen significativas a juzgar por la prueba t en paréntesis debajo de los coeficientes de $LY(-1)$ y $LP(-1)$; sin embargo, el vector CointEq1 no parece significativo en todas las ecuaciones del VEC. Por ejemplo, en la ecuación de los precios $D(LP)$ su estadístico t es sólo de -1.29 .

Advertimos que, aun cuando es usual que los programas de computador muestren las pruebas t , éstas deben verse como aproximaciones. Se recomienda como ejercicio calcular las predicciones de las variables LM , LY , y LP para 1999, que son, en su orden, 9.35, 6.85 y 13.68. Obviamente, esto se hace con las ecuaciones individuales del VEC y con los datos del Cuadro 10.3.

Cuadro 10.5: Vector de corrección de error estimado

Date: 04/29/99 Time: 15:45 Sample (adjusted): 1972 1998 Included observations: 27 after adjusting endpoints Standard errors & t-statistics in parentheses			
Cointegrating Eq:	CointEq1		
LM(-1)	1.000000		
LY(-1)	1.451203 (0.18870) (-7.69070)		
LP(-1)	-0.863840 (0.03434) (-25.1529)		
C	12.21604		
Error Correction	D(LM)	D(LY)	D(LP)
CointEq1	0.697063 (0.40348) (1.72761)	0.424935 (0.05415) (7.84691)	-0.230200 (0.17809) (-1.29261)
D(LM(-1))	-0.183034 (0.36599) (-0.50011)	-0.129889 (0.04912) (-2.64429)	0.504351 (0.16154) (3.12215)
D(LY(-1))	-0.757852 (0.88263) (-0.85863)	-0.088254 (0.11846) (-0.72812)	0.874480 (0.38958) (2.24470)
D(LP(-1))	0.943424 (0.47411) (1.98988)	0.185517 (0.06363) (2.91545)	0.063359 (0.20926) (0.30277)
C	0.105893 (0.08830) (1.19925)	0.035944 (0.01185) (3.03299)	0.040160 (0.03897) (1.03044)
R-squared	0.224041	0.825814	0.492333
Adj. R-squared	0.082958	0.794144	0.400030
Sum sq. resids	0.090198	0.001625	0.017572
S.E. equation	0.064030	0.008594	0.028262
Log likelihood	38.66008	92.88465	60.74197
Akaike AIC	-2.493339	-6.509974	-4.129035
Schwarz SC	-2.253368	-6.270004	-3.889065
Mean dependent	0.225317	0.040500	0.206329
S.D. dependent	0.066884	0.018941	0.036487
Determinant Residual Covariance	9.98E - 11		
Log Likelihood	195.9462		
Akaike Information Criteria	-13.18120		
Schwarz Criteria	-12.31731		

Ejercicios

1. Si y_t es una serie de ruido formada con el último dígito de los números del directorio telefónico y x_t es una serie iid $N(-5, 5)$, el valor esperado de la suma de ambas series, $E(y_t + x_t)$, será:

a) Menor que cero

b) Igual a cero

c) Mayor que cero

d) Ninguna de las anteriores

$$y_t \sim iid(\mu_y, \sigma_y^2) \quad \mu_y = \frac{0+1+2+3+4+\dots+9}{10}$$

$$x_t \sim iid N(-5, 5)$$

$$\mu_x = -5$$

$$E(y_t + x_t) = E(y_t) + E(x_t)$$

$$= \mu_y - 5$$

$$= 4.5 - 5$$

$$= -0.5$$

2. Para convertir una serie de tiempo x_t distribuida $N(0, 1)$ en otra serie distribuida $N(-5, 9)$:

$$(x_t - \mu) \sigma \sim N(-\mu, \sigma^2)$$

a) Primero se multiplica por 9 y luego se resta 5

b) Primero se resta 5 y luego se multiplica por 9

c) Primero se multiplica por 3 y luego se resta 5

d) Primero se resta 5 y luego se multiplica por 3

e) Ninguna de las anteriores

3. Si z se define como $z = \frac{x - \mu_x}{2\sigma_x}$, ¿cuál será la varianza de z ?

a) 2

- b) 1
- c) $\frac{1}{2}$
- d) $\frac{1}{4}$
- e) $\frac{1}{2} \text{ var}(x)$
- f) Ninguna de las anteriores

4. ¿Cuál es el valor esperado al lanzar un dado cargado en el cual los números impares tiene el doble de probabilidad de salir que los pares?

- a) 2.333
- b) 2.5
- c) 3.333
- d) 3.5

- e) 4.333
- f) 4.75

- g) Ninguna de las anteriores

$$\begin{aligned}
 P_{\text{normal}} &= \frac{1}{6} & \xrightarrow{\frac{1}{2} \text{ por}} \\
 P_{\text{par}} &= \frac{1}{3}/\frac{1}{3} = \frac{1}{9} & \xrightarrow{\frac{1}{2} \text{ impares}} \\
 P_{\text{impares}} &= \frac{2}{3}/\frac{1}{3} = \frac{2}{9} & \xrightarrow{1 - P_{\text{par}}} \\
 E(x) &= 1 \cdot \frac{2}{9} + 2 \cdot \frac{1}{9} + 3 \cdot \frac{2}{9} + 4 \cdot \frac{1}{9} + 5 \cdot \frac{2}{9} + 6 \cdot \frac{1}{9} \\
 &= 3.83 & \xrightarrow{\text{agrupan pares e impares para facilitar la operación}}
 \end{aligned}$$

normalizamos
 para que
 sumen uno

5. Sea x una variable aleatoria. ¿En cuáles unidades se mide $f(x)$, la función de densidad de probabilidad de x ?

- a) En las mismas unidades de x
- b) En las unidades de $1/x$
- c) En las unidades de x^2
- d) Es adimensional
- e) Ninguna de las anteriores

6. Si los ingresos de una firma rentable aumentan 10% y sus costos aumentan 10% durante el mismo periodo, la ganancia:

- a) Aumenta 10%
- b) Aumenta 5%
- c) Aumenta menos que 5%
- d) No cambia

- e) Cae 5% aproximadamente
 f) Ninguna de las anteriores
 7. El cambio porcentual en la tendencia de la serie $x_t = t + \varepsilon_t$ es:

a) Constante y positivo

$$\frac{\partial x_t}{\partial t} = 1 \rightarrow \text{constante}$$

b) Oscilante

c) Lineal creciente

d) Cero

e) Escalonado creciente

f) Ninguna de las anteriores

8. Si $x_t > 0$ en $t = 0$ y el cambio porcentual es $\Delta x_t \% = -5\%$ para $t > 0$, entonces:

a) x_t eventualmente se torna negativa

$$\Delta x_{t+1} = \frac{x_t - x_{t-1}}{x_{t-1}} = -5\%$$

b) x_t nunca se torna negativa

\downarrow constante

c) x_t será constante

d) x_t cae cada vez más rápido

e) Ninguna de las anteriores

9. Considere la serie de tiempo $x_t = t + t^2 + \varepsilon_t$. La primera diferencia de x_t , definida como $\Delta x_t = x_t - x_{t-1}$:

$$\frac{(t-1)(t-1)}{t^2 - 2t + 1}$$

a) Será estacionaria

$$\Delta x_t = x_t - x_{t-1}$$

b) Tendrá una tendencia exponencial

$$= (t + t^2 + \varepsilon_t) - [(t-1) + (t-1)^2 + \varepsilon_{t-1}]$$

c) Tendrá una tendencia lineal

$$= t + t^2 + \varepsilon_t - t + 1 - t^2 + 2t - 1 - \varepsilon_{t-1}$$

d) Tendrá una tendencia cuadrática

$$\Delta x_t = 2t - 1 + u_t$$

e) Ninguna de las anteriores

$$\downarrow \varepsilon_t - \varepsilon_{t-1}$$

Tendencia lineal

10. La autocovarianza $\hat{R}(2)$ para la serie $\{1, 0, 3, 0\}$ es:

a) $10/4$

$$\text{Cov}(X_t, X_{t-2}) = E(X_t - \bar{x})(X_{t-2} - \bar{x})$$

$$= E[X_t X_{t-2} - X_t \bar{x} - X_{t-2} \bar{x} + \bar{x}^2]$$

b) $6/4$

$$\text{Cov}(X_t, X_{t-2}) = E(X_t X_{t-2}) - \bar{x}E(X_t) - \bar{x}E(X_{t-2}) + \bar{x}^2$$

$$= E(X_t X_{t-2}) - \bar{x}^2 - \bar{x}^2$$

$$= \frac{1}{4}(1 \cdot 3 + 0 \cdot 0) - 1^2 = \frac{3}{4} - 1 = -\frac{1}{4}$$

- c) $\frac{3}{4}$
d) $\frac{2}{4}$
e) $\frac{1}{4}$
f) 0

$$\begin{aligned} \text{Cov}(X_t, X_{t-2}) &= \frac{1}{n} \sum (X_t - \bar{x})(X_{t-2} - \bar{x}) \\ &= \frac{1}{4} [(1-1)(3-1)^2 + (0-1)(0-1)] \\ &= \frac{1}{4} \end{aligned}$$

\hookrightarrow Se pierden dos

- g) Ninguna de las anteriores

11. La autocovarianza para el primer rezago, $R(1)$, de la serie $\{\dots, -1, 1, -1, 1, -1, 1, \dots\}$ será: $\bar{x} = 0$

- a) 1
b) -1
c) 0.5
d) -0.5
e) 0

$$\begin{aligned} \text{Cov}(X_t, X_{t-1}) &= \frac{1}{n} \sum (X_t - \bar{x})(X_{t-1} - \bar{x}) \\ &= \frac{1}{n} [(1-0)(-1-0) + (-1-0)(1-0) + \dots] \\ &= \frac{1}{n} (-1 - 1 - 1 - \dots) \\ &= -1 \end{aligned}$$

- f) Ninguna de las anteriores
 \hookrightarrow por infinito

12. El coeficiente de autocorrelación $\hat{r}(1)$ para la serie $\{5, 7, 9, 2, 7\}$ es:

- a) -0.5
b) 0.5
c) -0.15
d) 0.1
e) 0.34
f) -0.34

$$\begin{aligned} \bar{x} &= \frac{5+7+9+2+7}{5} = 6 \\ \text{Var}(X_t) = R(0) &= \frac{1}{n} \sum (X_t - \bar{x})^2 = \frac{1}{5} [(5-6)^2 + (7-6)^2 + (9-6)^2 + (2-6)^2 + (7-6)^2] \\ &= \frac{1}{5} [5.6] \\ \text{Cov}(X_t, X_{t-1}) = R(1) &= \frac{1}{n} \sum (X_t - \bar{x})(X_{t-1} - \bar{x}) \\ &= \frac{1}{5} [(5-6)(7-6) + (7-6)(9-6) + (9-6)(2-6) + (2-6)(7-6)] \\ &\quad \hookrightarrow \text{se pierde uno} \\ \hat{r}(1) &= \frac{R(1)}{R(0)} = \frac{-2.8}{5.6} = -0.5 \end{aligned}$$

13. Las autocorrelaciones estimadas $\hat{r}(1)$ y $\hat{r}(2)$ para la serie $\{1, 4, 2, 5\}$ son, respectivamente:

- a) 0.5 y -0.5
b) -0.4 y 0.2
c) 0.4 y -0.2
d) -0.5 y 0.4
e) Ninguna de las anteriores

$$\begin{aligned} \bar{x} &= \frac{1+4+2+5}{4} = 3 \\ \text{Var}(X_t) = R(0) &= \frac{1}{n} \sum (X_t - \bar{x})^2 = \frac{1}{4} [(1-3)^2 + (4-3)^2 + (2-3)^2 + (5-3)^2] \\ &= 2.5 = \frac{10}{4} \\ \text{Cov}(X_t, X_{t-1}) = R(1) &= \frac{1}{n} \sum (X_t - \bar{x})(X_{t-1} - \bar{x}) \\ &= \frac{1}{4} [(1-3)(4-3) + (4-3)(2-3) + (2-3)(5-3)] \\ &= -\frac{5}{4} = -1.25 \\ \text{Cov}(X_t, X_{t-2}) = R(2) &= \frac{1}{n} \sum (X_t - \bar{x})(X_{t-2} - \bar{x}) \end{aligned}$$

\hookrightarrow se pierde uno

$$\hat{r}(1) = \frac{R(1)}{R(0)} = \frac{-5/4}{10/4} = -\frac{5}{10} = -\frac{1}{2}$$

$$\hat{r}(2) = \frac{R(2)}{R(0)} = \frac{-1}{10/4} = -\frac{1}{10} = -\frac{1}{2}$$

$$= \frac{1}{4} \sum [(1-3)(2-3) + (4-3)(5-3)]$$

14. Para las series $x_t = \{1, 4, 2, 5\}$ y $y_t = \{4, 2, 4, 2\}$ encuentre (aproximadamente) la correlación cruzada para rezago uno, $\hat{r}_{xy}(1)$:

a) -0.33333

$$\bar{x} = \frac{1+4+2+5}{4} = \frac{12}{4} = 3$$

b) 0.63

$$\text{Var}(x_t) = R(0) = \frac{1}{n} \sum (x_t - \bar{x})^2 = \frac{1}{4} [(1-3)^2 + (4-3)^2 + (2-3)^2 + (5-3)^2] = \frac{20}{4} = 5$$

c) 0.16

$$\text{Var}(y_t) = R(0) = \frac{1}{n} \sum (y_t - \bar{y})^2 = \frac{1}{4} [(4-3)^2 + (2-3)^2 + (4-3)^2 + (2-3)^2] = \frac{4}{4} = 1$$

d) -0.47

$$\begin{aligned} R_{xy}(1) &= E(x_t - \bar{x})(y_{t-1} - \bar{y}) = \frac{1}{n} \sum (x_t - \bar{x})(y_{t-1} - \bar{y}) \\ &= \frac{1}{4} [(1-3)(2-3) + (4-3)(4-3) + (2-3)(2-3)] = -1 \end{aligned}$$

e) Ninguna de las anteriores

$$\hat{r}_{xy}(1) = \frac{R_{xy}(1)}{\sqrt{R(0)R(0)}} = \frac{-1}{\sqrt{5 \cdot 1}} = -\frac{1}{\sqrt{5}}$$

15. La serie $x_t = \sin\left(\frac{\pi t}{4}\right) + \varepsilon_t$, donde ε_t es ruido blanco, tiene $r(\tau) = 0$ para:

a) $\tau = 0$

b) $\tau = 1$

c) $\tau = 2$

d) $\tau = 4$

e) Ninguna de las anteriores

16. Si en la ecuación $x_t = 0.3x_{t-1} + \varepsilon_t$, se sabe que $x_0 = 5$, y que $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\} = \{-5, 2, 4\}$, entonces x_3 está más cerca de:

a) 0

$$X_0 = 5$$

b) 4.3

$$X_1 = 0.3X_0 + \varepsilon_1 = 0.3(5) + (-5) = -3.5$$

c) -1.7

$$X_2 = 0.3X_1 + \varepsilon_2 = 0.3(-3.5) + 2 = 0.95$$

d) -3

$$X_3 = 0.3X_2 + \varepsilon_3 = 0.3(0.95) + 4 = 4.285$$

e) 2.3

f) -2.7

g) 0.3

17. Utilizando el modelo $x_t = 5 - x_{t-2} + \varepsilon_t + \varepsilon_{t-2}$, encuentre x_3 si $\{\varepsilon_1, \varepsilon_2, \varepsilon_3\} = \{-10, 3, -8\}$ y $x_t = \varepsilon_t = 0$ para $t \leq 0$.

a) -8

$$X_0 = 5 - X_{-2} + \varepsilon_0 + \varepsilon_{-2} = 5$$

b) -18

$$X_1 = 5 - X_{-1} + \varepsilon_1 + \varepsilon_{-1} = 5 - 10 = -5$$

$$X_2 = 5 - X_0 + \varepsilon_2 + \varepsilon_{-2} = 5 - 5 + 3 = 3$$

$$X_3 = 5 - X_1 + \varepsilon_3 + \varepsilon_{-1} = 5 - 5 - 8 - 10 = -18$$

- c) -5
d) -21
e) Ninguna de las anteriores
18. Una serie de ruido blanco:

- a) Debe tener media cero \rightarrow No necesariamente
b) Se puede predecir utilizando su propio pasado
 c) Es estacionaria
d) Nunca tiene autocorrelaciones estimadas $\hat{r}(\tau)$ significativas
e) Debe tener valores positivos y negativos \rightarrow La media puede no ser cero
f) Debe tener una distribución de probabilidad simétrica
g) Ninguna de las anteriores
19. Si el estadístico $Q = n(n + 2) \sum_{\tau=1}^s \frac{\hat{r}^2(\tau)}{n-\tau}$ se aplica a una serie y_t , y Q resulta significativamente diferente de cero, se infiere que:

- a) La varianza de las autocorrelaciones es unitaria
b) La media de la serie y_t es 0 Cuando $Q \geq 0$ significa que y_t es ruido
c) La serie y_t es ruido blanco Cuando $Q > 0$ y_t tiene autocorrelación
d) La media de la serie y_t es negativa
 e) Ninguna de las anteriores

20. Considere la prueba Q , cuya expresión es: $Q = n(n + 2) \sum_{\tau=1}^s \frac{\hat{r}^2(\tau)}{n-\tau}$.
Para n grande, el valor esperado de la prueba Q es:

- a) 0
b) $1/n$
c) $1/\sqrt{n}$
d) s/n
e) s
f) 1

$$\hat{r}^2(\tau) = \frac{\hat{R}(\tau)}{\hat{R}(0)} = \frac{\frac{1}{n} \sum (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\frac{1}{n} \sum (x_t - \bar{x})^2}$$

En general, $\hat{R}(\tau) < \hat{R}(0)$, a mayor n la diferencia es menor, por lo que $\hat{r}^2(\tau) \xrightarrow{d} 0$ cuando $n \rightarrow \infty$

$Q \xrightarrow{d} 0$ cuando $n \rightarrow \infty$

- g) Ninguna de las anteriores
pida la correlación serial en los errores
21. Considere el modelo $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$. Para la prueba de Breusch-Godfrey de autocorrelación, podríamos correr (donde u es un error):
residuos vs explicativas + residuos retrasados
- $\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_t$
 - $|\hat{\varepsilon}_t| = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_t$
 - $\hat{\varepsilon}_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + u_t$
 - $\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 \hat{\varepsilon}_{t-1}^2 + u_t$
 - $|\hat{\varepsilon}_t| = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 |\hat{\varepsilon}_{t-1}| + u_t$
 - $\hat{\varepsilon}_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 \hat{\varepsilon}_{t-1} + u_t$
 - $\hat{\varepsilon}_t^2 = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{1t}^2 + \alpha_4 x_{2t}^2 + \alpha_5 x_{1t} x_{2t} + u_t$
 - h) Ninguna de las anteriores
22. La expresión compleja $\frac{1}{i}$ es equivalente a:
multiplicando por el conjugado
- $-i$
 - -1
 - i
 - 1
 - Ninguna de las anteriores
23. La expresión compleja $\frac{1}{(a+bi)i}$ es equivalente a:
- $\frac{-b-ai}{b^2-a^2}$
 - $\frac{-b-ai}{b^2+a^2}$
 - $\frac{b+ai}{b^2+a^2}$
 - $\frac{1}{b^2-a^2}$
 - Ninguna de las anteriores
24. La expresión $L^2(-2 - L^{-1}x_t) + L^{-1}$, es igual a:
 $-2 - X_{t-1}$
- x_{t+1}
 - $L^2(-2 - L^{-1}X_t) + L^{-1}$
 - $L^2(-2 - X_{t+1})$
 - $-2 - L^2 X_{t+1}$

- b) $1 - x_{t+1}$
 c) $2 + x_{t+1}$
 d) $1 - x_{t+2}$
 e) $-1 - x_{t-1}$
 f) x_{t-1}
 g) Ninguna de las anteriores

25. La expresión $\frac{1}{4}(L + 2i)(L - 2i)x_t = (1 + L)\varepsilon_t$, donde $i = \sqrt{-1}$, corresponde a:

- a) $x_t = -x_{t-1} + 4\varepsilon_t + 4\varepsilon_{t-1}$
 b) $x_t = 0.2x_{t-2} + 4\varepsilon_t + 4\varepsilon_{t-1}$
 c) $x_t = 4x_{t-1} - 2x_{t-2} + \varepsilon_t + \varepsilon_{t-1}$
 d) $x_t = -0.25x_{t-2} + \varepsilon_t + \varepsilon_{t-1}$
 e) Ninguna de las anteriores

$$\begin{aligned} & [L^2 - 2^2 i^2] x_t \\ & L^2 x_t + 4 x_t \\ & x_{t-2} + 4 x_t \\ & x_{t-2} + 4 x_t = 4\varepsilon_t + 4\varepsilon_{t-1} \\ & x_t = -0.25 x_{t-2} + \varepsilon_t + \varepsilon_{t-1} \end{aligned}$$

26. La solución particular de x_t en el modelo $x_t = 0.2x_{t-1} + \varepsilon_t$ será:

- a) $x_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \dots$
 b) $x_t = \varepsilon_t - 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} - \dots$
 c) $x_t = \varepsilon_t + 0.2\varepsilon_{t-1} + 0.04\varepsilon_{t-2} + \dots$
 d) $x_t = \varepsilon_{t-1} + 0.2\varepsilon_{t-2} + 0.04\varepsilon_{t-3} + \dots$
 e) Ninguna de las anteriores

27. La solución particular de $2y_t = 0.2y_{t-1} + \varepsilon_t$ es:

- a) $y_t = \frac{\varepsilon_t}{2} + \frac{\varepsilon_{t-1}}{20} + \frac{\varepsilon_{t-2}}{200} + \dots$
 b) $y_t = \frac{\varepsilon_t}{2} + \frac{\varepsilon_{t-1}}{4} + \frac{\varepsilon_{t-2}}{16} + \dots$
 c) $y_t = \frac{\varepsilon_t}{2} + \frac{\varepsilon_{t-1}}{4} + \frac{\varepsilon_{t-2}}{8} + \dots$
 d) $y_t = \varepsilon_t + \frac{\varepsilon_{t-1}}{2} + \frac{\varepsilon_{t-2}}{4} + \dots$
 e) Ninguna de las anteriores

28. La expresión $\varepsilon_t = \frac{x_t}{(1-0.5L)^2}$ corresponde a:

Tiene dos raíces pero generan la misma expansión

a) $\varepsilon_t = x_t + 0.5x_{t-1} + 0.25x_{t-2} + \dots$

b) $\varepsilon_t = x_t + x_{t-1} + 0.75x_{t-2} + \dots$

c) $\varepsilon_t = x_{t-2} + 0.5x_{t-4} + 0.25x_{t-6} + \dots$ $\frac{X_t}{(1-0.5L)} = X_t + 0.5X_{t-1} + 0.25X_{t-2} + \dots$ $(1-0.5L) \rightarrow 1 - 0.5L$

d) $\varepsilon_t = x_{t-2} + 2x_{t-3} + 4x_{t-4} + \dots$

e) Ninguna de las anteriores

29. La expresión autorregresiva equivalente del MA(1) $x_t = \varepsilon_t + 0.4\varepsilon_{t-1}$ es:

a) $x_t = -2.5x_{t-1} + \varepsilon_t$

$$\begin{aligned} X_t &= \varepsilon_t + 0.4\varepsilon_{t-1} \\ X_t &= \varepsilon_t + 0.4(X_{t-1} - 0.4\varepsilon_{t-2}) \\ X_t &= \varepsilon_t + 0.4X_{t-1} - 0.16\varepsilon_{t-2} \\ &\vdots \\ &\hookrightarrow AR(oo) \end{aligned}$$

b) $x_t = 2.5x_{t-1} + 6.25x_{t-2} + \dots + \varepsilon_t$

c) $x_t = 0.4x_{t-1} + 6.25x_{t-2} + \dots + \varepsilon_t$

d) $x_t = -0.4x_{t-1} + \varepsilon_t$

e) $x_t = 0.4x_{t-1} - 0.16x_{t-2} + \dots + \varepsilon_t$

f) Ninguna de las anteriores

30. La autocorrelación $r(1)$ para el modelo $x_t = \varepsilon_t - 0.5\varepsilon_{t-1}$ es:

a) 0.5 $r(z) \rightarrow ACF \rightarrow q's sgnificativas$

b) -0.5 $r(z) = 0$

c) -0.4

d) -0.25

e) 0

f) Ninguna de las anteriores

31. Considere el modelo $x_t = 0.6x_{t-1} - 0.2x_{t-2} + \varepsilon_t$. La autocorrelación para el primer rezago $r(1)$ es:

a) 0.2 $X_t = 0.6[0.6X_{t-2} - 0.2X_{t-3} + \varepsilon_{t-2}] - 0.2X_{t-2} + \varepsilon_t$

b) 0.3 $X_t = 0.36X_{t-2} - 0.12X_{t-3} + 0.6\varepsilon_{t-1} - 0.2X_{t-2} + \varepsilon_t$

c) 0.4 $X_t = 0.56X_{t-2} - 0.12X_{t-3} + \varepsilon_t + 0.6\varepsilon_{t-1}$

d) 0.5

- e) 0.6
 f) 0.8
 g) Ninguna de las anteriores

32. Con respecto al modelo $x_t = 0.5x_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$, se puede decir que:

- a) x_t no es estacionaria
 b) x_t es ruido blanco
 c) $x_t \rightarrow 0$ cuando $t \rightarrow \infty$
 d) La función de autocorrelación de x_t decae exponencialmente
 e) Ninguna de las anteriores

33. El modelo $x_t = 1 + 0.5x_{t-1} + \varepsilon_t$, es equivalente a:

- a) $x_t = 0.5 + \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \dots$ $X_t = 1 + 0.5(1 + 0.5X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$
 b) $x_t = 1 + \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \dots$ $X_t = 1.5 + 0.25(1 + 0.5X_{t-2} + \varepsilon_{t-1}) + 0.5\varepsilon_{t-1} + \varepsilon_t$
 c) $x_t = 2 + \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \dots$ $X_t = 1.75 + 0.125(1 + 0.5X_{t-2} + \varepsilon_{t-1}) + 0.25\varepsilon_{t-1} + 0.5\varepsilon_{t-2} + \varepsilon_t$
 d) $x_t = 1 + 0.5\varepsilon_t + 0.25\varepsilon_{t-1} + 0.125\varepsilon_{t-2} + \dots$ $X_t = 1.5 + 0.5\varepsilon_t + 0.25\varepsilon_{t-1} + 0.125\varepsilon_{t-2} + \dots$ Cambie
 e) $x_t = \varepsilon_t + 0.5\varepsilon_{t-1} + 0.25\varepsilon_{t-2} + \dots$ $X_t = 1.75 + 0.5\varepsilon_t + 0.25\varepsilon_{t-1} + 0.125\varepsilon_{t-2} + \dots$
 f) Ninguna de las anteriores

34. El modelo $x_t = \alpha x_{t-1} + \varepsilon_t$ se puede expresar como:

- a) $x_t = \alpha^{-1}\varepsilon_1 + \alpha^{-2}\varepsilon_2 + \alpha^{-3}\varepsilon_3 + \dots$ $X_t = \alpha X_{t-1} + \varepsilon_t$
 b) $x_t = \alpha^{-1}x_1 + \alpha^{-2}x_2 + \alpha^{-3}x_3 + \dots + \varepsilon_t$ $X_t = \varepsilon_t + \alpha\varepsilon_{t-1} + \alpha^2\varepsilon_{t-2} + \dots + \alpha^{t-2}\varepsilon_2 + \alpha^{t-1}x_1$
 c) $x_t = \varepsilon_t + \alpha^{-1}\varepsilon_{t-1} + \alpha^{-2}\varepsilon_{t-2} + \alpha^{-3}\varepsilon_{t-3} + \dots$
 d) $x_t = \alpha^t + \alpha^{t-2}\varepsilon_2 + \alpha^{t-3}\varepsilon_3 + \dots + \varepsilon_t$
 e) $x_t = \alpha^{t-1}x_1 + \alpha^{t-2}\varepsilon_2 + \alpha^{t-3}\varepsilon_3 + \dots + \varepsilon_t$
 f) Ninguna de las anteriores

35. Si el operador de diferencia, Δ , se define como $\Delta x_t = x_t - x_{t-1}$, entonces aplicarlo dos veces, $\Delta(\Delta x_t)$, dará:

$$\begin{aligned} \Delta x_t &= x_t - x_{t-1} \\ \Delta^2 x_t &= \Delta(\Delta x_t) = (x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) \\ &= x_t - 2x_{t-1} + x_{t-2} \end{aligned}$$

- b) $x_t - x_{t-2}$
- c) $x_t - x_{t-1} - x_{t-2}$
- d) $x_t - x_{t-1} + 0.5x_{t-2}$
- e) Ninguna de las anteriores

36. Sea x_t una serie estacionaria con $E x_t = \mu$, y sea $s_{1t} = \frac{1}{5} \sum_{i=-2}^2 x_{t-i}$

un promedio aritmético de x_t y $s_{2t} = \frac{1}{7} \sum_{i=-3}^3 x_{t-i}$ otro promedio aritmético de x_t . ¿Cuál será $E(s_{1t} + s_{2t})$?

- a) 0
- b) μ
- c) $\frac{2}{3}\mu$
- d) $\frac{3}{2}\mu$
- e) 2μ
- f) 3μ
- g) Ninguna de las anteriores

37. Si aplicamos un suavizado rezagado de dos términos {donde $m = 1$ en $s_t = \frac{1}{m+1} \sum_{i=0}^1 x_{t-i}$ } a otro suavizado rezagado de dos términos, el resultado será:

- a) $s_t = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{4}$
- b) $s_t = \frac{x_t + 2x_{t-1} + x_{t-2}}{4}$
- c) $s_t = x_t + x_{t-1}$
- d) $s_t = \frac{x_t + x_{t-1} + x_{t-2} + x_{t-3}}{2}$
- e) Ninguna de las anteriores

38. Si aplicamos el suavizado $s_t = \frac{1}{3} \sum_{i=0}^2 x_{t-i}$ a la serie $x_t = t/3$, el resultado será:

- a) $s_t = \frac{t}{3} - 1$

- b) $s_t = \frac{t}{9}$
- c) $s_t = 3t - 3$
- d) $s_t = \frac{t}{3} - \frac{1}{3}$
- e) $s_t = t - 1$
- f) Ninguna de las anteriores
39. Un método para suavizar una serie x_t es por medio de la expresión $s_t = Ax_t + A(1-A)x_{t-1} + A(1-A)^2x_{t-2} + A(1-A)^3x_{t-3} + \dots$, donde $0 < A < 1$ y s_t es la serie suavizada. Esta expresión es equivalente a:
- a) $s_t = (x_t + x_{t-1})/(1 - A)$
- b) $s_t = Ax_t - A^2x_{t-1} + A^3x_{t-2} - A^4x_{t-3} + \dots$
- c) $s_t = (x_{t+1} + x_t + x_{t-1})/(1 - A)$
- d) $s_t = Ax_t + s_{t-1} - s_{t-2}$
- e) $s_t = Ax_t + A(1 - A)s_{t-1}$
- f) $s_t = x_t + As_{t-1}$
- g) $s_t = x_t + (1 - A)s_{t-1}$
- h) $s_t = Ax_t + (1 - A)x_{t-1}$
- i) $s_t = Ax_t + (1 - A)s_{t-1}$
- j) $s_t = Ax_t + (1 - A)(s_{t-1} + s_{t+1})$
- k) Ninguna de las anteriores
40. Si estimamos correctamente la tendencia determinística \hat{x}_t de la serie $x_t = \alpha t + \varepsilon_t$, entonces $\frac{x_t}{\hat{x}_t}$ será:
- a) Estacional
- b) Estacionaria
- c) Imaginaria
- d) Constante
- e) Ninguna de las anteriores

41. Si una serie mensual vale 0 en todos los meses, excepto en diciembre, cuando vale 12, los factores estacionales serán:
- 1 en todos los meses
 - 12 en diciembre y 0 en los demás meses
 - 11 en diciembre y -1 en los demás meses
 - 1 en diciembre y 0 en los demás meses
 - Ninguna de las anteriores
42. Si la producción diaria es siempre la misma, entonces los factores estacionales de la producción mensual:
- Serán siempre iguales en el tiempo
 - Aumentarán linealmente en el tiempo
 - Aumentarán exponencialmente en el tiempo
 - Disminuyen linealmente en el tiempo
 - Ninguna de las anteriores
43. Con respecto a la serie trimestral $\{\dots, 0, 1, 0, 1, 0, 1, 0, 1, 0, \dots\}$, los cuatro factores estacionales serían:
- $0, 1/2, 0, 1/2$
 - $-1/2, 1/2, -1/2, 1/2$
 - $0, 3/5, 0, 3/5$
 - $-2/5, 3/5, -2/5, 3/5$
 - $0, -1, 0, -1$
 - Ninguna de las anteriores
44. Suponga una variable mensual x_t mayor que 0 que siempre crece al 2% por mes. La serie x_t desestacionalizada será:
- Igual a x_t para todo t
 - Menor que x_t para todo t
 - 0 para todo t

- d) Mayor que x_t para todo t
 e) Igual a 0.02 para todo t
 f) Igual a -0.02 para todo t
 g) Ninguna de las anteriores
45. En el modelo $x_t = D_1 + 2D_2 + 3D_3 - 2D_4 + 0.5x_{t-1} + \varepsilon_t$, donde los D_i son *dummies* trimestrales que valen 1 para el trimestre i -ésimo y 0 para otros trimestres, el valor esperado $E x_t$ será:
- a) 0
 b) 2
 c) 4
 d) 6
 e) 8
 f) -4
 g) -8
 h) Ninguna de las anteriores
46. El espectro de la variable x_t , que sigue el modelo $x_t = \varepsilon_t + \varepsilon_{t-1}$, es:
 [la expresión del espectro: $s(\omega) = \frac{R(0)}{2\pi} + \frac{1}{\pi} \sum_{\tau=1}^{\infty} R(\tau) \cos(\omega\tau)$]
- a) 1
 b) $\frac{\sigma_\varepsilon^2}{2\pi}$
 c) $\frac{\sigma_\varepsilon^2}{\pi} + \frac{\sigma_\varepsilon^2}{\pi} \cos(\omega)$
 d) $\frac{\sigma_\varepsilon^2}{\pi} + \frac{1}{\pi} \cos(\pi)$
 e) $\frac{1}{2} + \frac{1}{\pi} \cos(\pi)$
 f) $\frac{1}{2} + \frac{1}{2\pi} \cos(2\omega)$
 g) $\frac{\sigma_\varepsilon^2}{\pi}$
 h) $\frac{1}{2\pi}$
 i) Ninguna de las anteriores

7. Sea un plano cartesiano con x_{t-1} en el horizontal y x_t en el eje vertical, donde se han graficado los puntos $(1, -1)$, $(-1, 0)$ y $(0, 1)$. Si la primera observación de la serie $\{x_t\}$ es 0, ¿la cuarta observación debe ser?:
- 0
 - 1
 - 1
 - $\sqrt{\pi}$
 - $\sqrt{2}$
 - Ninguna de las anteriores
8. El valor esperado de la serie $x_t = \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \alpha_3 x_{t-3} + \varepsilon_t$, es:
- α_0
 - $\alpha_0(1 - \alpha_1 - \alpha_2 - \alpha_3)$
 - $\alpha_0 / (1 - \alpha_1 - \alpha_2 - \alpha_3)$
 - $1 + \alpha_1 + \alpha_2 + \alpha_3$
 - 0
 - Ninguna de las anteriores
9. Sea $x_t = 0.5x_{t-1} + \varepsilon_t$ donde $\varepsilon_t \sim (0, \sigma_\varepsilon^2)$ es ruido blanco. La varianza de x_t será:
- $2\sigma_\varepsilon^2$
 - $4\sigma_\varepsilon^2/3$
 - $3\sigma_\varepsilon^2/2$
 - $2\sigma_\varepsilon^2/3$
 - $\sigma_\varepsilon^2/3$
 - Ninguna de las anteriores
10. Si $E x_t = \mu$, entonces c en $x_t = c + \alpha x_{t-1} + \varepsilon_t$ es:
- $\mu(1 - \alpha)$

- b) μ
- c) $1 + \alpha$
- d) $\frac{\mu}{1-\alpha}$
- e) α
- f) Ninguna de las anteriores

51. Sea $x_t = c + \alpha x_{t-1} + \varepsilon_t + \beta \varepsilon_{t-1}$. ¿Cuál es $E x_t$?

- a) 0
- b) $\frac{c+\beta\sigma_\varepsilon^2}{1-\alpha}$
- c) $c + \alpha \sigma_\varepsilon^2$
- d) $\frac{c}{1-\alpha}$
- e) $c + \alpha + \beta \sigma_\varepsilon^2$
- f) Ninguna de las anteriores

52. En el modelo ARMA $x_t = 0.4x_{t-1} + \varepsilon_t + 2\varepsilon_{t-1}$, donde $E \varepsilon_t = c$, ¿cuál será $E x_t$?

- a) c
- b) $2c$
- c) $3c$
- d) $4c$
- e) $5c$
- f) Ninguna de las anteriores

53. El modelo $x_t = \varepsilon_t$ es:

- a) Un AR(0)
- b) Un MA(0)
- c) Un ARMA(0, 0)
- d) Ruido
- e) Todas las anteriores
- f) Ninguna de las anteriores

54. Sea $x_t = \alpha x_{t-1} + \varepsilon_t$ una serie construida con ε_t , donde ε_t es ruido con media 0. Se puede decir que x_t es:
- a) Ruido
 - b) Un MA finito
 - c) Estacional
 - d) Ninguna de las anteriores
55. Sea $x_t = \alpha x_{t-1} + \varepsilon_t$ donde ε_t es $idd(0, 1)$, la expresión $E(\varepsilon_t x_{t+1})$ es:
- a) 0
 - b) 1
 - c) α
 - d) α^2
 - e) $1 - \alpha$
 - f) Ninguna de las anteriores
56. Utilizando el modelo $x_t = 1 + 0.4x_{t-1} + 0.1x_{t-2} + \varepsilon_t$, ¿cuál será la mejor predicción de x_t para el futuro lejano?
- a) 0.5
 - b) 1
 - c) 1.5
 - d) 2
 - e) Ninguna de las anteriores
57. Sea $x_t = c + \alpha x_{t-1} + \varepsilon_t$. La predicción \hat{x}_{n+2} hecha en el tiempo n (al final de la muestra) para el tiempo $n + 2$, será:
- a) $c(1 + \alpha)$
 - b) c
 - c) $c + \alpha x_n$
 - d) $c + \alpha \hat{x}_n$
 - e) $c(1 + \alpha) + \alpha^2 x_n$

f) Ninguna de las anteriores

58. ¿Cuál de los siguientes modelos pudo producir las predicciones $\{\hat{x}_{n+1}, \hat{x}_{n+2}, \hat{x}_{n+3}\} = \{6, 5, 4.5\}$

a) $x_t = 6 - 0.5x_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$

b) $x_t = 6 + 0.5x_{t-1} + \varepsilon_t + \varepsilon_{t-1}$

c) $x_t = 6 + 6x_{t-1} + \varepsilon_t + 6\varepsilon_{t-1}$

d) $x_t = 6 + x_{t-1} + \varepsilon_t$

e) $x_t = 2 + 6x_{t-1} + \varepsilon_t$

f) $x_t = 2 + 0.5x_{t-1} + \varepsilon_t + \varepsilon_{t-1}$

g) Ninguna de las anteriores

59. Usted sabe que en una tómbola hay 333 balotas blancas y 667 negras. El experimento consiste en sacar 50 balotas secuencialmente. Su tarea es predecir el color de cada balota antes de que salga. La predicción óptima será:

a) Siempre predecir blanca

b) Siempre predecir negra

c) En la tercera parte de las veces predecir blanca

d) Alternar entre predecir negra y predecir blanca

e) Ninguna de las anteriores

60. Sea un nuevo indicador construido dividiendo el indicador coincidente por el indicador rezagado. Este nuevo indicador tendería a servir como:

a) Indicador líder

b) Indicador coincidente

c) Indicador rezagado

d) Ruido

e) Ninguna de las anteriores

61. Si al evaluar la predicción de una serie x_t encontramos que el Error Medio Porcentual Absoluto (MAPE) es mayor que 100%, podemos decir que:
- La serie x_t no es estacionaria
 - La series x_t es estacionaria
 - La serie x_t no tiene raíz unitaria
 - La serie x_t está cointegrada
 - La predicción es pésima
 - Ninguna de las anteriores
62. Sea \hat{x} la predicción de la inflación. ¿Cuáles son las unidades del Error Medio Absoluto (MAE)? [La expresión del MAE es $E|x_{t+T} - \hat{x}_{t+T}|$]
- Porcentaje
 - $(\text{Porcentaje})^2$
 - 1/porcentaje
 - Es adimensional
 - Ninguna de las anteriores
63. Si en la ecuación $x_t = -0.6x_{t-1} + \varepsilon_t$, se sabe que $x_1 = -2$, la predicción de x_4 estará más cerca de:
- 2
 - 1.9
 - 1
 - 1.2
 - 0.7
 - 0.4
 - 0
64. Sea $x_t = \varepsilon_t + \beta\varepsilon_{t-3}$. El error de predicción para la predicción \hat{x}_{n+3} es:
- $\varepsilon_{n+1} + \varepsilon_{n+2} + \varepsilon_{n+3}$

- b) ε_{n+2}
 c) ε_{n+3}
 d) $\varepsilon_{n+3} + \beta\varepsilon_{n+1}$
 e) $\beta\varepsilon_{n+3}$
 f) Ninguna de las anteriores
65. El valor máximo positivo que puede tomar la autocorrelación $r(1)$ de un proceso MA(1), $x_t = \varepsilon_t + \beta_1\varepsilon_{t-1}$, es:
- a) 0
 b) 0.2666
 c) $2\sqrt{2}$
 d) 0.5
 e) 0.6666
 f) 1
 g) Ninguna de las anteriores
66. ¿Cuál será (aproximadamente) la estimación por mínimos cuadrados ordinarios (MCO) de α en el AR(1) $x_t = \alpha x_{t-1} + \varepsilon_t$, donde $x_t = \{1, 4, 2, 5\}$?
- a) 1.05
 b) -0.4
 c) 1.42
 d) -0.6
 e) -0.85
67. Suponga que, para n grande, se corren por mínimos cuadrados ordinarios las siguientes dos regresiones: $x_t = \alpha x_{t-1} + \varepsilon_t$ y $x_t = \gamma x_{t+1} + \nu_t$ donde ε_t y ν_t son ruido blanco. Con respecto a los coeficientes estimados, se encontrará que:
- a) $\hat{\alpha} = \hat{\gamma}$
 b) $\hat{\alpha} > \hat{\gamma}$

- c) $\hat{\alpha} < \hat{\gamma}$
d) $\hat{\alpha} = 1/\hat{\gamma}$
e) Ninguna de las anteriores
68. Para estimar β en el MA(1) $x_t = \varepsilon_t + \beta\varepsilon_{t-1}$, con una muestra de tres observaciones ($n = 3$), la cantidad que se va a minimizar ($\sum \hat{\varepsilon}_t^2$) será:
- a) $\sum \hat{\varepsilon}_t^2 = (x_1 - \beta)^2 + (x_2 - \beta)^2 + (x_3 - \beta)^2$
b) $\sum \hat{\varepsilon}_t^2 = (\beta x_1)^2 + \beta(x_1 - \beta x_2)^2 + \beta^2(x_1 - \beta x_2 + \beta^2 x_3)^2$
c) $\sum \hat{\varepsilon}_t^2 = x_1^2 + x_2^2 + x_3^2 + \beta^2$
d) $\sum \hat{\varepsilon}_t^2 = x_1^2 + (x_2 - \beta x_1)^2 + (x_3 - \beta x_2 + \beta^2 x_1)^2$
e) Ninguna de las anteriores
69. Considere el modelo $x_t = \alpha x_{t-2} + \varepsilon_t$. ¿Su autocovarianza $R(1)$ será?
- a) $\sigma_\varepsilon^2 / (1 - \alpha^2)$
b) α
c) α^2
d) 1
e) 0
f) Ninguna de las anteriores
70. Para el modelo $x_t = \alpha_1 x_{t-1} + \alpha_3 x_{t-3} + \varepsilon_t$, la expresión de Yule-Walker correspondiente es $E(x_t x_{t-\tau}) = \alpha_1 E(x_{t-1} x_{t-\tau}) + \alpha_3 E(x_{t-3} x_{t-\tau}) + E(\varepsilon_t x_{t-\tau})$. Para estimar α_1 y α_3 se necesitan las ecuaciones de Yule-Walker para:
- a) $\tau = 0, \tau = 1$
b) $\tau = 0, \tau = 2$
c) $\tau = 1, \tau = 2$
d) $\tau = 0, \tau = 3$
e) Ninguna de las anteriores

71. En el VAR $\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.5 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$, donde x_{1t} y x_{2t} son estacionarias y ε_{1t} y ε_{2t} tienen media 0, ¿cuál es $E x_{1t}$?
- 0
 - 1/2
 - 1
 - 2
 - 2.5
 - Ninguna de las anteriores
72. El VAR(1), $\begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ c & 1 \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$, será estacionario si las raíces de la siguiente expresión están por fuera del círculo unitario:
- $1 - cL - (2 + c)L^2 = 0$
 - $1 - 2cL + cL^2 = 0$
 - $1 - 2L + (1 - 2c)L^2 = 0$
 - $1 - 2cL + (2 + c)L^2 = 0$
 - Ninguna de las anteriores
73. En el VAR(1) del punto anterior, la predicción de x_{2t} hecha en el tiempo n para el tiempo $n + 3$, llamada \hat{x}_{2n+3} , será:
- $c\hat{x}_{1n+2} + 2\hat{x}_{2n+2}$
 - $x_{1n} + cx_{2n}$
 - $c\hat{x}_{1n+2} + \hat{x}_{2n+2}$
 - $cx_{1n} + x_{2n}$
 - Ninguna de las anteriores
74. Considere las regresiones: $z_t = \alpha_0 + \alpha_1 z_{t-1} + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \varepsilon_t$ y $z_t = \beta_0 + \beta_1 z_{t-1} + \beta_3 y_{t-1} + u_t$. Con estas regresiones se puede probar:
- Si z causa y

- b) Si z causa x
 c) Si x causa y
 d) Si x causa z
 e) Si y causa z
 f) Si y causa x
 g) Ninguna de las anteriores
75. En el VAR(1) del punto anterior, con $c = 3$, la respuesta impulso de la variable x_{2t} a un shock $\varepsilon_{10} = 1$ será:
- a) 1, 9, 27, 81, ...
 b) 0, 1, 3, 6, ...
 c) 1, 3, 6, 9, ...
 d) 0, 3, 6, 27, ...
 e) Ninguna de las anteriores
76. Sea $\nu_t = P^{-1}\varepsilon_t$ donde $\nu_t = \begin{pmatrix} \nu_{1t} \\ \nu_{2t} \end{pmatrix}$, $\varepsilon_t = \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}$, y P^{-1} es el inverso de la matriz de descomposición de Choleski, tal que $PP' = \Sigma$, donde Σ es la matriz de varianza covarianza del vector ε_t . Si $\nu_{1t} = \varepsilon_{1t}$ y $\nu_{2t} = \frac{-2\varepsilon_{1t}}{3} + \frac{\varepsilon_{2t}}{3}$, encuentre el valor que falta en la matriz de varianza covarianza $\Sigma = \begin{pmatrix} 1 & 2 \\ 2 & ? \end{pmatrix}$.
- a) 5
 b) 8
 c) 13
 d) 15
 e) 7
 f) Ninguna de las anteriores
77. Suponga que, con la descomposición de Cholesky, transformamos ε_{1t} y ε_{2t} y obtenemos $\nu_{1t} = \frac{\varepsilon_{1t}}{2}$ y $\nu_{2t} = -\frac{\varepsilon_{1t}}{4} + \frac{\varepsilon_{2t}}{2}$. ¿Cuál es la covarianza de ε_{1t} y ε_{2t} ?

- a) 0
 b) 1
 c) 2
 d) 4
 e) $1/2$
 f) Ninguna de las anteriores
78. La matriz de varianza covarianza Σ_{n+T} de los errores de predicción de un VAR(p) con m variables y n observaciones, es de orden:
- a) $n \times n$
 b) $(n + T) \times (n + T)$
 c) $m \times m$
 d) $(m + T) \times (m + T)$
 e) $(n + p) \times (n + p)$
 f) Ninguna de las anteriores
79. Considere el VAR estructural

$$\begin{pmatrix} 0.6 & 0.2 \\ 0.2 & 0.4 \end{pmatrix} \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} \begin{pmatrix} x_{1t-1} \\ x_{2t-1} \end{pmatrix} + \begin{pmatrix} \omega_{1t} \\ \omega_{2t} \end{pmatrix},$$

donde ω_t es el vector de errores estructurales no correlacionados entre sí. Encuentre la covarianza de los errores de la forma reducida, $Cov(\varepsilon_1, \varepsilon_2) = ?$

- a) 0
 b) $Var(\omega_1) + Var(\omega_2)$
 c) $-Var(\omega_1) - Var(\omega_2)$
 d) $0.2Var(\omega_1) + 0.2Var(\omega_2)$
 e) $0.04Var(\omega_1) + 0.04Var(\omega_2)$
 f) $-0.4Var(\omega_1) + 0.6Var(\omega_2)$
 g) $-4Var(\omega_1) - 9Var(\omega_2)$

h) $4Var(\omega_1) - 6Var(\omega_2)$

i) $-2Var(\omega_1) - 3Var(\omega_2)$

j) -1

k) Ninguna de las anteriores

80. Considere el SVAR, $B_0x_t = B_1x_{t-1} + B_2x_{t-2} + \dots + B_px_{t-p} + \omega_t$,

donde $B_0 = \begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}$. El error ε_{2t} en el VAR equivalente será:

a) $\varepsilon_{2t} = 0$

b) $\varepsilon_{2t} = \omega_{1t}$

c) $\varepsilon_{2t} = -0.5\omega_{1t}$

d) $\varepsilon_{2t} = \omega_{2t}$

e) $\varepsilon_{2t} = -0.5\omega_{1t} + \omega_{2t}$

f) $\varepsilon_{2t} = \omega_{1t} - 0.5\omega_{2t}$

g) $\varepsilon_{2t} = -0.5\omega_{2t}$

h) Ninguna de las anteriores

81. Con respecto al VAR estructural $B_0x_t = B_1x_{t-1} + B_2x_{t-2} + \dots + B_px_{t-p} + \omega_t$, con 3 variables ($m = 3$), donde $E(\omega_t\omega_t') = \Omega$, ¿cuántas incógnitas hay en la expresión $B_0^{-1}\Omega B_0^{-1'}$?

a) 3

b) 6

c) 9

d) 12

e) 15

f) 18

g) 24

h) 27

i) Ninguna de las anteriores

82. Considere la siguiente prueba de causalidad de Granger:

Null Hypothesis:	F-Statistic	Probability
A no causa a B	1.39184	0.28638
B no causa a A	1.40255	0.26950

De la prueba podemos concluir que:

- a) A no causa a B
- b) B no causa a A
- c) A causa a B y B causa a A
- d) No hay causalidad entre A y B

83. Considere el siguiente modelo estructural: $x_t = c + \varepsilon_{1t}$, $y_t = x_{t-1} + \varepsilon_{2t}$, $z_t = y_{t-1} + \varepsilon_{3t}$, donde c es una constante y ε_{1t} , ε_{2t} , ε_{3t} son series de ruido independientes entre sí. En el sentido de Granger podemos decir que:

- a) x no causa a y
- b) x causa a z
- c) y no causa a z
- d) z causa a y
- e) Ninguna de las anteriores

84. En la ecuación de la varianza del modelo ARCH, $\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2$, se incluye el parámetro β_0 para:

- a) Asegurar que la suma de los errores ε_t sea 0
- b) Asegurar que la varianza incondicional sea diferente de 0
- c) Asegurar que la varianza condicional sea estacionaria
- d) Evitar la heteroscedasticidad condicional en la varianza
- e) Ninguna de las anteriores

85. En un modelo ARCH:

- a) Si ε_t es positivo, ε_{t+1} tiende a ser positivo
b) Si ε_t es negativo, ε_{t+1} tiende a ser negativo
c) Hay autocorrelación en los errores $\{\varepsilon_t\}$
d) Todas las anteriores
e) Ninguna de las anteriores
86. En el modelo GARCH $\sigma_t^2 = 2 + 0.4\sigma_{t-1}^2 + 0.4\varepsilon_{t-1}^2$, el valor esperado incondicional de la varianza, $E\sigma_t^2$, es:
- a) 2
b) 40
c) 10
d) 2.8
e) 4
f) 20
g) Ninguna de las anteriores
87. El modelo ARCH es:
- a) Estacionario en media y no estacionario en varianza
b) Estacionario en media y estacionario en varianza
c) No estacionario en media y no estacionario en varianza
d) No estacionario en media y estacionario en varianza
e) Ninguna de las anteriores
88. En el modelo ARCH los errores estandarizados ε_t/σ_t :
- a) Tienen autocorrelación
b) Son independientes
c) Tienen varianza σ_t^2
d) Son heteroscedásticos
e) Todas las anteriores
f) Ninguna de las anteriores

89. El modelo GARCH, $\sigma_t^2 = 3 + 0.7\sigma_{t-1}^2 + 0.1\varepsilon_{t-1}^2$, puede escribirse como:

- a) $\sigma_t^2 = 10 + 0.1\varepsilon_{t-1}^2 + 0.07\varepsilon_{t-2}^2 + \dots$
- b) $\sigma_t^2 = 10 + 0.1\varepsilon_{t-1}^2 + 0.03\varepsilon_{t-2}^2 + \dots$
- c) $\sigma_t^2 = 10 + 0.7\sigma_{t-1}^2 + 0.7^2\sigma_{t-2}^2 + \dots$
- d) $\sigma_t^2 = \frac{30}{9} + 0.1\varepsilon_{t-1}^2 + 0.01\varepsilon_{t-2}^2 + \dots$
- e) $\sigma_t^2 = 3 + 0.7\sigma_{t-1}^2 + 0.7^2\sigma_{t-2}^2 + \dots$
- f) $\sigma_t^2 = 3 + 0.8\sigma_{t-1}^2$
- g) 15
- h) Ninguna de las anteriores

90. El modelo GARCH $\sigma_t^2 = \alpha_0 + \alpha_1\sigma_{t-1}^2 + \beta\varepsilon_{t-1}^2$ puede escribirse como [donde u_t es ruido y γ_i son coeficientes]:

- a) $\sigma_t^2 = \gamma_0 + \gamma_1\sigma_{t-1}^2 + \gamma_2\sigma_{t-2}^2 + \gamma_3\sigma_{t-3}^2 + \dots$
- b) $\sigma_t^2 = \gamma_0 + \gamma_1\varepsilon_{t-1}^2 + u_t$
- c) $\sigma_t^2 = \gamma_0 + \gamma_1\sigma_{t-1}^2 + u_t$
- d) $\sigma_t^2 = \gamma_0 + \gamma_1(\sigma_{t-1}^2 + \varepsilon_{t-1}^2)$
- e) $\sigma_t^2 = \gamma_0$
- f) $\sigma_t^2 = \gamma_0 + u_t$
- g) Ninguna de las anteriores

91. En el modelo $\sigma_t = 1 + 0.6|\varepsilon_{t-1}| - 0.6\varepsilon_{t-1}$, la desviación estándar σ_t será constante para:

- a) $\varepsilon_{t-1} > 0$
- b) $|\varepsilon_{t-1}| < 1$
- c) $\varepsilon_{t-1} < 0.6$
- d) $\varepsilon_{t-1} < 0$
- e) $|\varepsilon_{t-1}| > 0.6$
- f) $|\varepsilon_{t-1}| > 1.6$
- g) Ninguna de las anteriores

92. Si $E\Delta x_t = 0$, entonces la serie x_t puede ser:
- Estacionaria
 - No estacionaria
 - Un paseo aleatorio
 - Ruido blanco
 - Todas las anteriores
 - Ninguna de las anteriores
93. Si en el paseo aleatorio $x_t = x_{t-1} + \varepsilon_t$ tenemos que $E\varepsilon_t = c$, donde c es una constante, entonces $E x_t$ dado x_0 será:
- c
 - $x_0 + c$
 - $t(x_0 + c)$
 - $x_0 + ct$
 - Ninguna de las anteriores
94. Sea $x_t = x_{t-1} + \varepsilon_t^2$, donde ε_t está distribuida $iid(0, \sigma_\varepsilon^2)$. El valor esperado condicional $E(x_t/x_0)$, donde x_0 es un valor inicial, será:
- x_0
 - tx_0
 - $x_0 + t\sigma_\varepsilon^2$
 - $x_0 + \sigma_\varepsilon^2$
 - $x_0 + t$
 - ∞
 - Ninguna de las anteriores
95. Considere el modelo $x_t = x_{t-1} + \alpha + \varepsilon_t + \nu_t$, donde α es una constante y ε_t y ν_t son series de ruido independientes entre sí con media 0 y varianzas σ_ε^2 y σ_ν^2 . El valor esperado de x_t dado x_0 es:
- $x_0 + \alpha t + (\sigma_\varepsilon^2 + \sigma_\nu^2)t$

- b) $x_0 + \alpha t$
 c) $x_0 + \alpha$
 d) $x_0 + \alpha(\sigma_\varepsilon^2 + \sigma_\nu^2)$
 e) $x_0 + \frac{\sigma_\varepsilon^2 + \sigma_\nu^2}{1-\alpha}$
 f) Ninguna de las anteriores
96. Sea el modelo de regresión $y_t = \gamma_1 x_{1t} + \gamma_2 x_{2t} + \varepsilon_t^2$, donde ε_t tiene una distribución Normal $iidN(0, 1)$. Puede decirse que:
- a) Hay efecto ARCH
 b) $Ey_t = 0$
 c) x_{1t} y x_{2t} están cointegradas
 d) y_t es $I(1)$
 e) y_t es $I(2)$
 f) Ninguna de las anteriores
97. El proceso $x_t = \varepsilon_t + \varepsilon_{t-1}$ es:
- a) $I(0)$
 b) $I(1)$
 c) $I(2)$
 d) $I(3)$
 e) Ninguna de las anteriores
98. Sean los procesos $\omega_t = \omega_{t-1} + \varepsilon_t$, $y_t = 2\omega_t + \nu_t$ y $x_t = \omega_t + u_t$, donde ε_t , u_t y ν_t son procesos de ruido independientes entre ellos. Podemos afirmar que:
- a) x_t es $I(0)$
 b) y_t es $I(2)$
 c) $y_t - x_t$ es $I(0)$
 d) x_t y y_t están cointegradas
 e) $x_t + y_t$ es $I(3)$

- f) Todas las anteriores
- g) Ninguna de las anteriores
99. En la regresión espuria $x_{1t} = \beta_0 + \beta_1 x_{2t} + \varepsilon_t$, donde x_{1t} y x_{2t} son $I(1)$, $\hat{\varepsilon}_t$ será:
- a) $I(0)$
 - b) $I(1)$
 - c) $I(2)$
 - d) $I(3)$
 - e) Ninguna de las anteriores
100. Sea x_t una serie de ruido blanco. Si le aplicamos la prueba de Dickey–Fuller aumentada (ADF) podemos decir que:
[la expresión de la prueba ADF es $\Delta x_t = \varphi x_{t-1} + \sum_{i=1}^p \gamma_i \Delta x_{t-i} + \varepsilon_t$]
- a) Los coeficientes φ y γ_i serán 0
 - b) El coeficiente φ será 0
 - c) El coeficiente φ será positivo
 - d) El coeficiente φ será negativo
 - e) Ninguna de las anteriores
101. Considere el modelo $x_t = 0.6x_{t-1} + \varepsilon_t$. Al aplicarle la prueba de raíz unitaria en la cual se corre Δx_t contra x_{t-1} , el coeficiente de x_{t-1} será:
- a) -1.6
 - b) -1
 - c) -0.6
 - d) -0.4
 - e) 0
 - f) 0.4
 - g) 0.6

h) 1

i) Ninguna de las anteriores

102. A medida que τ crece, la razón de varianza, cuya expresión es $VR(\tau) = \frac{\text{VAR}(x_{t+\tau} - x_t)}{\tau \text{VAR}(x_{t+1} - x_t)}$, tiende:

- a) a 0 si x_t es un paseo aleatorio
- b) a 0 si x_t es estacionaria
- c) a 1 si x_t es estacionaria
- d) a -1 si x_t es un paseo aleatorio
- e) a τ si x_t es un paseo aleatorio
- f) a τ si x_t es estacionaria
- g) Ninguna de las anteriores

103. Si en la regresión $x_{1t} = \alpha_0 + \alpha_1 x_{2t} + \varepsilon_t$, donde x_{1t} y x_{2t} son $I(1)$, el estadístico de Durbin-Watson es muy bajo, entonces:

- a) x_{1t} y x_{2t} están cointegradas
- b) x_{1t} y x_{2t} no están cointegradas
- c) Hay autocorrelación negativa
- d) Ninguna de las anteriores

104. Se dice que hay cointegración cuando las series x_{1t} y x_{2t} son $I(1)$ y la expresión $z_t = x_{1t} - ax_{2t}$ es:

- a) Ruido blanco
- b) $I(0)$
- c) $I(1)$
- d) $I(2)$
- e) Ninguna de las anteriores

105. Suponga que x_{1t} , x_{2t} , y x_{3t} son variables $I(1)$. Si corremos una regresión de x_{1t} contra x_{2t} y x_{3t} , y encontramos que los residuos son $I(1)$, existe la posibilidad de que:

- a) x_{2t} y x_{3t} estén cointegradas
- b) x_{1t} y x_{2t} estén cointegradas
- c) x_{1t} y x_{3t} estén cointegradas
- d) x_{1t} , x_{2t} y x_{3t} estén cointegradas
- e) Haya dos relaciones de cointegración
- f) Todas las anteriores
- g) Ninguna de las anteriores
06. Si aplicamos la prueba de cointegración traza a cuatro variables ($m = 4$) y rechazamos, en su orden, las hipótesis de que haya máximo cero, una, dos y tres relaciones de cointegración, podemos concluir de esta prueba que el número de relaciones de cointegración será:
- a) 0
- b) 1
- c) 2
- d) 3
- e) 4
- f) Ninguna de las anteriores
07. Sea un sistema cointegrado de m variables al cual se le aplica la prueba *max eigen*, cuya expresión es $-n \log(1 - \hat{\lambda}_{H+1})$. Si la prueba excede el valor crítico:
- a) Hay una relación de cointegración
- b) Hay m relaciones de cointegración
- c) Hay al menos H relaciones de cointegración
- d) Hay al menos $H + 1$ relaciones de cointegración
- e) Hay al menos $H + 2$ relaciones de cointegración
- f) Ninguna de las anteriores
08. Considere el modelo $\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_0 x_{t-1} + \varepsilon_t$, donde $x_t = \begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix}$, todas $I(1)$. Si $\zeta_0 = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 4 & 8 & 12 \end{pmatrix}$, ¿cuántos vectores de cointegración independientes existen?

- a) 0
 b) 1
 c) 2
 d) 3
 e) Ninguna de las anteriores
109. Considere el modelo $\Delta x_t = \alpha_0 + \zeta_1 \Delta x_{t-1} + \zeta_0 x_{t-1} + \varepsilon_t$. Si $\zeta_0 = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix}$, ¿cuántos vectores de cointegración independientes existen?
- a) 4
 b) 3
 c) 2
 d) 1
 e) 0
 f) Ninguna de las anteriores
110. En el mecanismo de corrección de error $\Delta x_{1t} = \alpha_0 + \alpha_1 \Delta x_{1t-1} + \alpha_2 \Delta x_{2t-1} + \alpha_3 \Delta x_{3t-1} - \beta(\alpha_4 + x_{1t-1} - \alpha_5 x_{2t-1} + \alpha_6 x_{3t-1}) + \varepsilon_t$, donde todos los coeficientes son mayores que 0, puede decirse que en equilibrio:
- a) x_{1t} es proporcional a x_{2t}
 b) x_{1t} es proporcional a x_{3t}
 c) x_{1t} es igual a una constante
 d) x_{1t} se mueve inversamente a x_{2t}
 e) Ninguna de las anteriores
111. Sea el mecanismo de corrección de error $\Delta x_1 = -4(x_{t-1} - y_{t-1}) + \varepsilon_t$. Si las predicciones para el siguiente periodo después del final de la muestra son $\Delta \hat{x}_{n+1} = 8$ y $\hat{x}_{n+1} = 7$, ¿cuál es la última observación de la muestra y_n ?
- a) 0

- b) 1
 c) 2
 d) -1
 e) -2
 f) -3
 g) Ninguna de las anteriores
112. Dado el VEC estimado $\begin{pmatrix} \Delta x_t \\ \Delta y_t \end{pmatrix} = \begin{pmatrix} 0.7 & -0.3 \\ -0.4 & 0.6 \end{pmatrix} \begin{pmatrix} \Delta x_{t-1} \\ \Delta y_{t-1} \end{pmatrix} + \begin{pmatrix} -1 \\ 0.5 \end{pmatrix}$
 $(x_{t-1} - y_{t-1})$, y dados los datos $\{x_{1997}, x_{1998}\} = \{1, 2\}$ y $\{y_{1997}, y_{1998}\} = \{2, 1\}$, ¿cuál es la predicción para x_{1999} ?
 a) -1
 b) 0
 c) 1
 d) 2
 e) 6
 f) Ninguna de las anteriores
113. Si x_t es un vector con cuatro variables $I(1)$, el número máximo de relaciones de cointegración que puede haber es?
 a) 1
 b) 2
 c) 3
 d) 4
 e) Ninguna de las anteriores
114. Si en la expectativa condicional $E(y/x) = E(y)$, entonces $E(xy) = 0$.

Verdadero Falso

115. Cuando el estadístico Q de una serie es insignificante (muy pequeño) se acepta que la serie es ruido blanco.

Verdadero Falso

16. Las observaciones atípicamente grandes en magnitud (*outliers*) tienden a disminuir la significancia de las autocorrelaciones.

Verdadero Falso

17. Un electrocardiograma representa una serie estacional.

Verdadero Falso

18. La presencia significativa de la señal $\sin\left(\frac{\pi t}{600}\right)$ en una serie de tiempo, donde t está medido en meses, sugiere un ciclo de 8 años de duración.

Verdadero Falso

19. Un estimador insesgado, $\hat{\beta}$, sirve para conocer el valor exacto de β .

Verdadero Falso

20. Un estimador sesgado $\hat{\beta}_S$ puede tener mayor probabilidad de acertar que un estimador insesgado $\hat{\beta}_I$ si $var(\hat{\beta}_S) < var(\hat{\beta}_I)$.

Verdadero Falso

21. Una serie generada por lanzamientos de un dado cargado puede ser *iid*.

Verdadero Falso

22. Si ε_t es una serie de ruido, entonces $\Delta\varepsilon_t$ también es una serie de ruido.

Verdadero Falso

23. Si $x_t = \beta + \alpha x_{t-1} + \varepsilon_t$ es una serie estacionaria, donde ε_t es ruido con media 0, podemos decir $E(\Delta x) = 0$.

Verdadero Falso

24. El proceso $x_t = 0.5x_{t-1} + 0.5x_{t-2} + \varepsilon_t$ es estacionario.

Verdadero Falso

125. Si x_t es una serie mensual estacional, una forma de eliminar la estacionalidad es tomando el logaritmo de las primeras diferencias, esto es, tomando $\log(x_t - x_{t-1})$.

Verdadero Falso

126. La estimación de un modelo autorregresivo por mínimos cuadrados ordinarios es insesgada en muestras finitas.

Verdadero Falso

127. Las ecuaciones de un $\text{VAR}(p)$ se pueden estimar individualmente por mínimos cuadrados ordinarios.

Verdadero Falso

128. La respuesta impulso de un $\text{VAR}(p)$ muestra el comportamiento del sistema a choques externos.

Verdadero Falso

129. Un VAR de tres variables tiene seis coeficientes por cada rezago.

Verdadero Falso

130. Si en un $\text{VAR}(p)$ tenemos que $E\epsilon\epsilon' = \Sigma = \begin{pmatrix} 4 & 2 \\ 2 & 5 \end{pmatrix}$, la matriz de Cholesky podría ser $\begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$.

Verdadero Falso

131. El supuesto que se hace en un vector autorregresivo (VAR) cuando se utiliza una matriz triangular inferior de Cholesky (con ceros por encima de la diagonal) es que los errores de la primera variable no afectan las demás variables de manera contemporánea.

Verdadero Falso

132. Una disminución a través del tiempo en el error cuadrado medio de predicción, $E(x_{t+T} + \hat{x}_{t+T})^2$, indica una mejora en la predicción o una disminución en la magnitud de x_t a través del tiempo.

Verdadero Falso

133. En el sentido de la causalidad de Granger, las ventas de adornos de Navidad causan la Navidad.

Verdadero Falso

134. En el modelo ARCH los errores ε_t no son independientes aunque no están autocorrelacionados.

Verdadero Falso

135. En la ecuación de regresión del modelo ARCH-M, $y_t = \gamma_0 + \gamma_1 \sigma_t^2 + \varepsilon_t$, el coeficiente γ_1 debe ser siempre positivo.

Verdadero Falso

136. Cuando se estima correctamente un modelo ARCH, los residuos al cuadrado de esta estimación deben ser ruido.

Verdadero Falso

137. De acuerdo con el modelo $\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \beta \varepsilon_{t-1}^2 + \gamma D_t$, donde $D_t = 1$ antes del 11 de septiembre 2001, y 0 después de esa fecha, puede decirse que los ataques terroristas aumentan la varianza σ_t^2 si $\gamma > 0$.

Verdadero Falso

138. En una serie con tendencia estocástica los *shocks* o *innovaciones* $\{\varepsilon_t\}$ tienen efecto permanente.

Verdadero Falso

139. En un VAR con variables $I(1)$ corremos el riesgo de tener regresiones espurias.

Verdadero Falso

140. Puede haber regresión espuria entre variables con tendencia determinística.

Verdadero Falso

141. Si en la prueba de raíz unitaria, $\Delta x_t = \varphi x_{t-1} + \sum_i \gamma_i \Delta x_{t-i} + \varepsilon_t$, encontramos que $\varphi = 0$, podríamos concluir que x_t es una serie no estacionaria.

Verdadero Falso

142. Considerando la variable x_t del modelo $x_t = 2x_{t-1} - x_{t-2} + \varepsilon_t$, si aplicamos la prueba de raíz unitaria de Phillips-Perron en la forma $\Delta^2 x_t = \varphi \Delta x_{t-1} + u_t$, obtendremos $\varphi < 0$.

Verdadero Falso

143. La primera diferencia del proceso $x_t = x_{t-1} + \varepsilon_t^2$ es estacionaria.

Verdadero Falso

144. La segunda diferencia del proceso $x_t = t^2 + \varepsilon_t$ es estacionaria.

Verdadero Falso

145. El proceso $x_t = 1.05x_{t-1} + \varepsilon_t$ es $I(1)$.

Verdadero Falso

146. El proceso $x_t = 2x_{t-1} + \varepsilon_t$ es $I(2)$.

Verdadero Falso

147. La cointegración entre dos variables sólo puede darse si ambas están integradas del mismo orden.

Verdadero Falso

148. En la relación de cointegración $\alpha_1 x_{1t} + \alpha_2 x_{2t} = z_t$, donde x_{1t} y x_{2t} son variables $I(1)$ y z_t es $I(0)$, el coeficiente α_2 nunca puede ser cero.

Verdadero Falso

149. La suma de dos series $I(1)$ siempre resulta ser una serie $I(0)$.

Verdadero Falso

150. Las series $x_t = 5$ y $y_t = 5 + \varepsilon_t$, donde ε_t es ruido blanco, están cointegradas.

Verdadero Falso

151. Si $\Delta x_{2t} = \beta_0 + \beta_1 \Delta x_{1t-1} + \beta_2 \Delta x_{2t-1} + \beta_3(x_{1t-1} - ax_{2t-1})$ es la ecuación del mecanismo de corrección de error para la variable x_{2t} , se puede afirmar que el signo del coeficiente β_3 debe ser negativo.

Verdadero Falso

152. En el vector de corrección de error (VEC) $\Delta x_t = \alpha_0 + \alpha_1 \Delta x_{t-1} + BA'x_{t-1} + \varepsilon_t$, el valor esperado $E(BA'x_{t-1})$ es independiente del tiempo.

Verdadero Falso

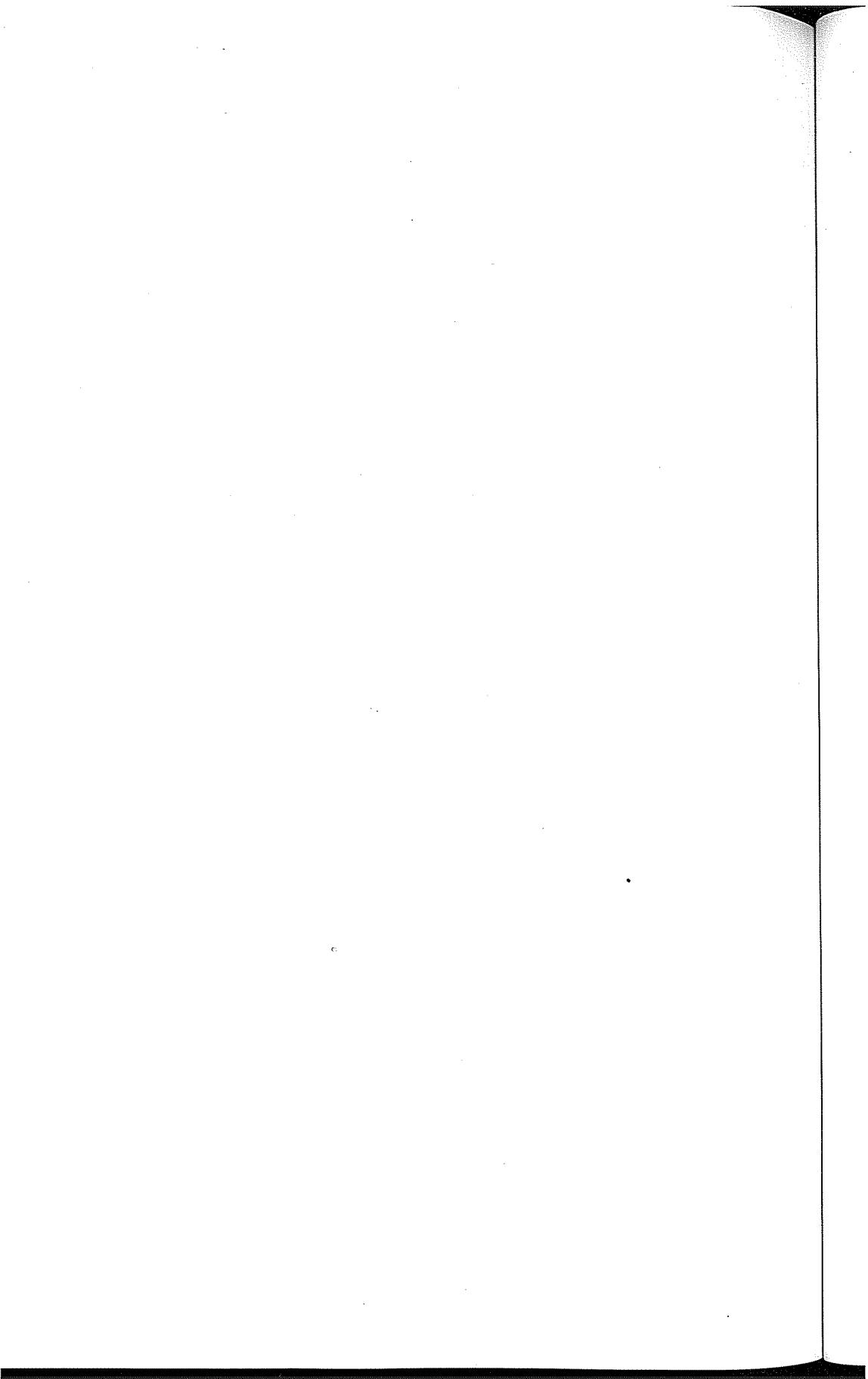
153. La hipótesis nula del estadístico $\lambda_{max} = -n \log(1 - \hat{\lambda}_{M+1})$ es que hay $M + 1$ relaciones de cointegración.

Verdadero Falso

Respuestas

1	a	41	c	81	d	121	V
2	c	42	e	82	d	122	F
3	d	43	b	83	e	123	V
4	c	44	d	84	b	124	F
5	b	45	b	85	e	125	F
6	a	46	c	86	c	126	F
7	f	47	a	87	b	127	V
8	b	48	c	88	b	128	V
9	c	49	b	89	a	129	F
10	e	50	a	90	c	130	V
11	b	51	d	91	a	131	F
12	a	52	e	92	e	132	V
13	d	53	e	93	d	133	F
14	b	54	d	94	c	134	V
15	c	55	c	95	b	135	F
16	b	56	d	96	f	136	F
17	a	57	e	97	a	137	F
18	c	58	f	98	d	138	V
19	e	59	b	99	b	139	F
20	e	60	a	100	d	140	V
21	f	61	f	101	d	141	V
22	a	62	a	102	b	142	F
23	b	63	f	103	b	143	V
24	e	64	c	104	b	144	V
25	d	65	d	105	a	145	F
26	c	66	a	106	a	146	F
27	a	67	a	107	d	147	V
28	b	68	d	108	b	148	V

29	e	69	e	109	e	149	F
30	c	70	c	110	a	150	F
31	d	71	e	111	b	151	F
32	b	72	c	112	d	152	V
33	c	73	c	113	c	153	F
34	e	74	d	114	F		
35	a	75	d	115	V		
36	e	76	c	116	V		
37	b	77	c	117	V		
38	d	78	c	118	F		
39	i	79	i	119	F		
40	e	80	e	120	V		



Referencias

- Artis, M. (1996). "How Accurate are the IMF's Short Term Forecasts?: Another Examination of WEO", IMF Working Paper 96/89, Washington.
- Atukeren, E. (2008). "Christmas Cards, Easter Bunnies, and Granger Causality", *Quality and Quantity*, 42 (6), 835-844.
- Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*, New York, John Wiley.
- Banerjee, A.; Dolado, J. J.; Galbraith, J. y Hendry, D. (1993). *Cointegration, Error Correction, and the Econometric Analysis of Time Series*, Oxford, Oxford University Press.
- Barsky, R. B. y Miron, J. A. (1989). "The Seasonal Cycle and the Business Cycle", *Journal of Political Economy*, 97, 503-535.
- Berkowitz, J. y O'Brien, J. (2001). "How Accurate are Value at Risk Models at Commercial Banks?", mimeo, Irvine, Graduate School of Management, University of California.
- Bernanke, B. (1986). "Alternative Explanations of the Money-Income Correlation", *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.
- Bierens, H. (1997). "Nonparametric Cointegration Analysis", *Journal of Econometrics*, 77 (2), 379-404.

- Blanchard, O. J. y Quah, D. (1989). "The Dynamic Effects of Aggregate Demand and Supply Disturbances", *American Economic Review*, 79, 655-673.
- Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31, 307-327.
- Bollerslev, T.; Engle, R. y Nelson, D. (1994). "ARCH Models", *Handbook of Econometrics* (vol. 4), Amsterdam, North-Holland.
- Box, G. y Jenkins, G. (1970). *Time Series analysis, Forecasting and Control*, San Francisco, Holden-Day.
- Brock, W.; Dechert, D.; Scheinkman, J. A. y LeBaron, B. (1996). "A Test for Independence based on the Correlation Dimension", *Econometric Reviews*, 15 (3), 197-235.
- Byeongseon, S. (1999). "Distribution Theory for Unit Root Tests with Conditional Heteroscedasticity", *Journal of Econometrics*, 91, 113-144.
- Charemza, W. y Deadman, D. (1992). *New Directions in Econometric Practice*, New York, Edward Elgar.
- Chou, R. (1988). "Volatility, Persistence and Stock Valuation: Some Empirical Evidence using Garch", *Journal of Applied Economics*, 3, 279-294.
- Cochrane, J. H. (1988). "How Big is the Random Walk in GNP?", *Journal of Political Economy*, 96, 893-920.
- Cryer, J. D. (1986). *Time Series Analysis*, Boston, Duxbury Press.
- Davidson, R. y MacKinnon, J. (1993). *Estimation and Inference in Econometrics*, Oxford, Oxford University Press.
- (2004). *Econometric Theory and Methods*, Oxford, Oxford University Press.
- Demetriades, P. y Hussein, K. (1996). "Does Financial Development Cause Economic Growth?: Time Series Evidence from 16 Countries", *Journal Development Economics*, 51, 387-411.
- Dezhbakhsh, H. (1990). "The Inappropriate use of Serial Correlation Tests in Dynamic Linear Models", *Review of Economics and Statistics*, 72 (1), 126-132.

- Dickey, D.; Hasza, D. y Fuller, W. (1984). "Testing for Unit Roots in Seasonal Time Series", *Journal of the American Statistical Association*, 79, 355-367.
- Diebold, F. (1998). "The Past, Present, and Future of Macroeconomic Forecasting", *Journal of Economic Perspectives*, 12 (2), 175-192.
- Doan, T. (1992). *Regression Analysis of Time Series, Rats User's Manual*, Evanston (Il), Estima.
- Eklund, J. y Karlsson, S. (2007). "An Embarrassment of Riches: Forecasting using Large Panels", Working Paper 34, Economics Department, Central Bank of Iceland.
- Elliott, G.; Rothenberg, T. y Stock, J. (1996). "Efficient Tests for an Autoregressive Unit Root", *Econometrica*, 64, 813-836.
- Enders, W. (2004). *Applied Econometric Time Series*, New York, John Wiley & Sons.
- Engle, R. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation", *Econometrica*, 50, 987-1008.
- (2001). "Garch 101: The Use of Arch/Garch Models in Applied Econometrics", *Journal of Economic Perspectives*, 15, 4, 157-168.
- y Granger, Clive (1987). "Cointegration and Error Correction: Representation, Estimation and Testing", *Econometrica*, 35, 251-276.
- Engle, R.; Lilien, D. y Robins, R. (1987). "Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model", *Econometrica*, 55, 391-407.
- Faulhaber, G. y Baumol, W. (1988). "Economists as Innovators", *Journal of Economic Literature*, 26, 577-600.
- Franses, P. y van Dijk, D. (2003). *Non-linear Times Series Models in Empirical Finance*, Cambridge University Press.
- Fuller, W. A. (1976). *Introduction to Statistical Time Series*, New York, John Wiley.
- Ghysels, E. y Perron, P. (1993). "The Effects of Seasonal Adjustment Filters on Tests for a Unit Root", *Journal of Econometrics*, 55, 57-98.

- Granger, C. (1969). "Investigating Causal Relationships by Econometric Models and Cross-Spectral Models", *Econometrica*, 37, 135-145.
- (1981). "Some Properties of Time Series Data and their use in Econometric Model Especification", *Journal of Econometrics*, 16, 121-130.
- (1982). "Acronyms in Time Series Analysis", *Journal of Time Series Analysis*, 3, 103-107.
- (1984). "Cointegrated Variables and Error Correction Models", Working Paper, San Diego, Dept. of Economics, University of California.
- (1988). "Some Recent Developments in a Concept of Causality", *Journal of Econometrics*, 39, 199-211.
- (1989). *Forecasting in Business and Economics*, Academic Press.
- Granger, C. y Newbold, P. (1974). "Spurious Regressions in Econometrics", *Journal of Econometrics*, 2, 111-120.
- (1986). *Forecasting Economic Time Series*, Orlando (FL), Academic Press.
- Granger, C. y Swanson, N. (1997). "An Introduction to Stochastic Unit Root Processes", *Journal of Econometrics*, 80 (1), 35-62.
- Greene, W. (2003). *Econometric Analysis*, New Jersey, Prentice Hall.
- Hamilton, J. (1994). *Time Series Analysis*, New Jersey, Princeton University press.
- Hannan, E. J. (1960). *Time Series Analysis*, London, Methuen.
- y Rissanen, J. (1982). "Recursive Estimation of Mixed Autorre gressive-Moving Average Order", *Biometrika*, 69, 81-94.
- Hodrick, R. J. y Prescott, E. C. (1997). "Postwar U.S. Business Cycles: An Empirical Investigation", *Journal of Money, Credit, and Banking*, 29 (1), 1-16.
- Hoel, A.; Clarkson, K. y Miller, R. (1983). *Economics Sourcebook of Government Statistics*, Lexington (MA), D.C. Heath and Company.
- Hotelling, H. (1935). "The Most Predictable Criterion", *Journal of Educational Psychology*, 26, 139-142.

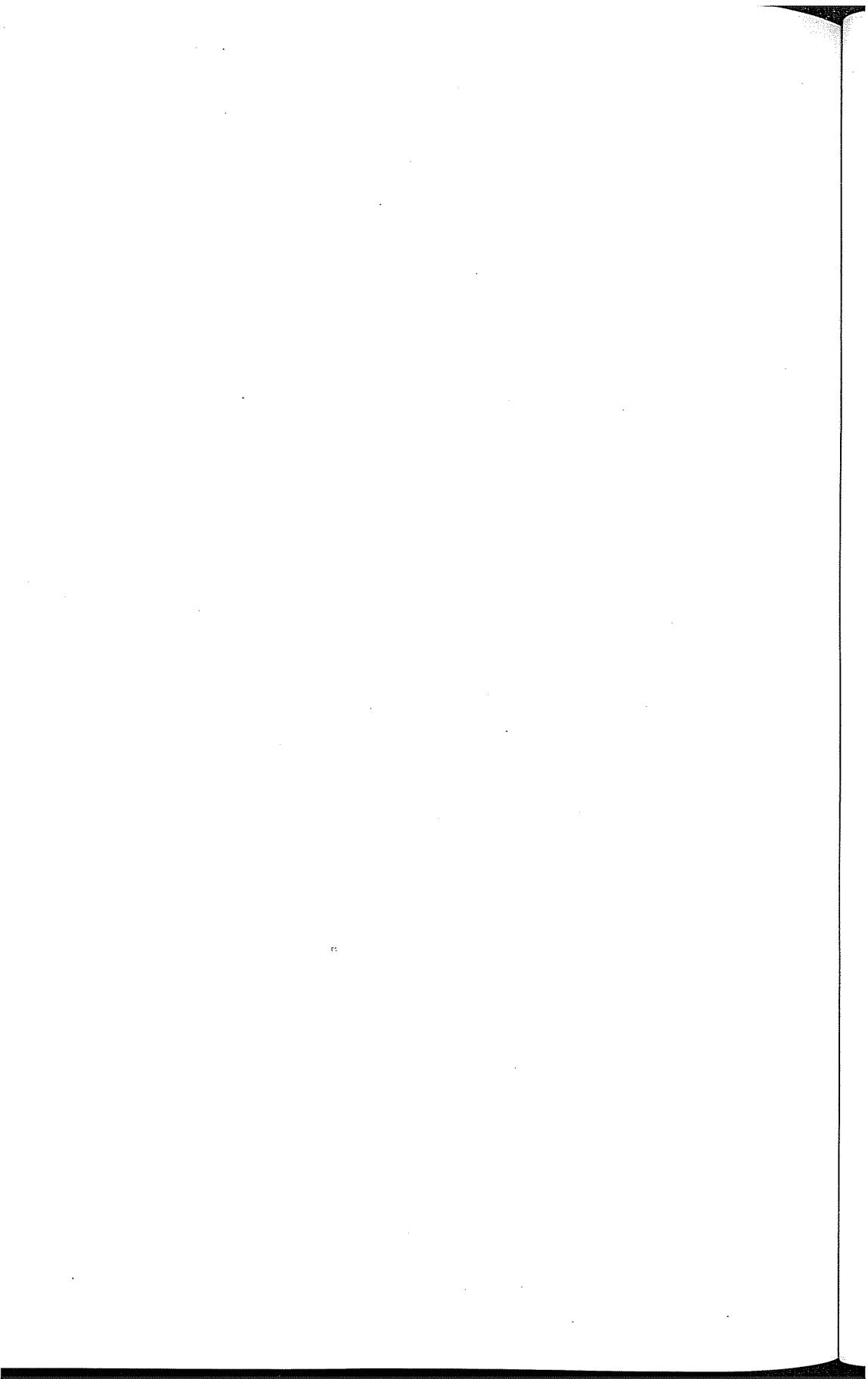
- Hylleberg, S.; Engle, R.; Granger, C. y Yoo, B. (1990). "Seasonal Integration and Cointegration", *Journal of Econometrics*, 44, 215-238.
- Jansen, D. y Kishan, R. (1996). "An Evaluation of Federal Reserve Forecasting", *Journal of Macroeconomics*, 18 (1), 89-109.
- Jenkins, G. M. y Watts, D. G. (1968). *Spectral Analysis and its Applications*, San Francisco, Holden-Day.
- Johansen, S. (1988). "Statistical Analysis of Cointegration Vectors", *Journal of Economic Dynamics and Control*, 12, 231-254.
- (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models", *Econometrica*, 59 (6), 1551-1580.
- (1995). *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford, Oxford University Press.
- (2006). "Statistical Analysis of Hypotheses on the Cointegrating Relations in the I(2) Model", *Journal of Econometrics*, 132, 81-115.
- y Juselius, K. (1990). "Maximum Likelihood Estimation and Inferences on Cointegration-with Applications to the Demand for Money", *Oxford Bulletin of Economics and Statistics*, 52, 169-210.
- Johnston, J. (1984). *Econometric Methods*, New York, McGraw-Hill.
- Judge, G.; Griffiths, W.; Hill, C.; Lutkepohl, H. y Lee, T. (1985). *The Theory and Practice of Econometrics*, New York, John Wiley and Sons.
- (1988). *Introduction to the Theory and Practice of Econometrics*, New York, John Wiley and Sons.
- Kennedy, P. (1992). *A Guide to Econometrics*, Boston, The MIT Press.
- Klein, J. (1997). *Statistical Versions in Time: a History of Time Series Analysis*, 1662-1938, Cambridge, Cambridge University Press.
- Klein, P. y Moore, G. (1983). "The Leading Indicator Approach to Economic Forecasting", *Journal of Forecasting*, 2, 119-135.
- Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P. y Shin, Y. (1992). "Testing the Null Hypothesis of Stationary against the Alternative of a Unit Root", *Journal of Econometrics*, 54, 159-178.

- Leybourne, S.; Hills, T. y Newbold, P. (1998). "Spurious Rejections by Dickey-Fuller Tests in the Presence of a Break under the Null", *Journal of Econometrics*, 87 (1), 191-203.
- Lindgren, B. (1976). *Statistical Theory*, New York, Macmillan.
- Longstaff, F. A. y Schwartz, E. S. (1992). "Interest Rate Volatility and the Term Structure: a Two Factor General Equilibrium Model", *Journal of Finance*, 4, 1259-1282.
- Lutkepohl, H. (1995). *Introduction to Multiple Time Series Analysis*, Berlin, Springer-Verlag.
- (2005). *New Introduction to Multiple Time Series Analysis*, Berlin, Springer.
- y Claessen, H. (1997). "Analysis of Cointegrated VARMA Processes", *Journal of Econometrics*, 80 (2), 223-239.
- Maddala, G. S. y Kim, In-M. (1998). *Unit Roots, Cointegration and Structural Change*, Cambridge, Cambridge University Press.
- Mahmoud, E. (1984). "Accuracy in Forecasting: A Survey", *Journal of Forecasting*, 3, 139-159.
- Mann, H. B. y Wald, A. (1943). "On the Statistical Treatment of Linear Stochastic Difference Equations", *Econometrica*, 11, 173-220.
- Marriott, F. H. C. y Pope, J. A. (1954). "Bias in the Estimation of Autoregressions", *Biometrika*, 41, 390-402.
- McNees, S. K. (1988). "How Accurate are Macroeconomic Forecasts", *New England Economic Review*, julio-agosto, 15-36.
- Mills, T. (2003). *Modelling Trends and Cycles in Economic Time Series*, New York, Palgrave MacMillan.
- Montenegro, Á. (1989a). "Inversión y PNB: relaciones de causalidad", *Desarrollo y Sociedad*, 24, 51-62.
- (1989b). "La función de autocorrelación y su empleo en el análisis de series de tiempo", *Desarrollo y Sociedad*, 23, 117-132.
- (1991). "El filtro Kalman: derivación y práctica", *Desarrollo y Sociedad*, 28, 79-91.

- (1995). *Introducción a la economía experimental*, Bogotá, Ediciones Uniandes-Ecoe.
- Murray, M. (1994). "A Drunk and Her Dog: An Illustration of Cointegration and Error Correction", *The American Statistician*, 48 (1), 37-39.
- Nelson, C. y Kang, H. (1984). "Pitfalls in the Use of Time as an Explanatory Variable in Regression", *Journal of Business and Economic Statistics*, 2 (1), 73-82.
- Nerlove, M.; Grether, D. y Carvalho, J. (1979). *Analysis of Economic Time Series*, New Jersey, Academic Press.
- Oppenheim, A. y Willsky, A. (1983). *Signals and Systems*, New Jersey, Prentice Hall.
- Osterwald-Lenum, M. (1992). "A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics", *Oxford Bulletin of Economics and Statistics*, 54, 461-472.
- Patterson, K. D. (2000). "Bias Reduction in Autoregressive Models", *Economics Letters*, 68 (2), 135-142.
- Perron, P. (1989). "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis", *Econometrica*, 57, 1361-1401.
- y Vogelsang, T. (1992). "Nonstationary and Levels Shifts with an Application to Purchasing Power Parity", *Journal of Business and Economic Statistics*, 10, 301-320.
- Pesaran, H. y Shin, Y. (1998). "Impulse Response Analysis in Linear Multivariate Models", *Economics Letters*, 58, 17-29.
- Phillips, P. (1986). "Understanding Spurious Regressions in Econometrics", *Journal of Econometrics*, 33, 311-40.
- (1991). "Optimal Inference in Cointegrated Systems", *Econometrica*, 59, 283-306.
- Pierce, D. A. (1980). "A Survey of Recent Developments in Seasonal Adjustments", *American Statistician*, 34, 125-134.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*, New Jersey, Academic Press.

- Quintos, C. (1998). "Analysis of Cointegration using GMM Approach", *Journal of Econometrics*, 85 (1), 155-188.
- Ramsey, J. y Montenegro, Á. (1992). "Identification and Estimation of Noninvertible Non-Gaussian MA(q) Processes", *Journal of Econometrics*, 54 (1-3), 301-320.
- Sargan, J. D. y Bhargava, A. (1983). "Testing Residuals from Least Squares Regressions for being generated by the Gaussian Random Walk", *Econometrica*, 51, 153-174.
- Schwarz, M. y Shaw, L. (1975). *Signal Processing, Discrete Spectral Analysis, Detection and Estimation*, New York, McGraw-Hill.
- Schwert, G. W. (1989). "Why does Stock Market Volatility Change Over Time?", *Journal of Finance*, 44 (5), 1115-1153.
- Simon, H. y Iwasaki, Y. (1988). "Causal Ordering, Comparative Statics and Near Decomposability", *Journal of Econometrics*, 39, 150.
- Sims, C. (1980). "Macroeconomics and Reality", *Econometrica*, 48, 1-48.
- (1986). "Are forecasting Models Usable for Public Policy Analysis?", *Federal Reserve Bank of Minneapolis Quarterly Review*, 1-16.
- Stock, J. y Watson, M. (1990). "Inference in Linear Time Series Models with some Unit Roots", *Econometrica*, 58, 113-144.
- Stock, J. y Watson, M. (1988). "Testing for Common Trends", *Journal of the American Statistical Association*, 83, 1097-1107.
- Swidler, S. y Ketcher, D. (1990). "Economic Forecasts, Rationality and the Processing of new Information over Time", *Journal of Money Credit and Banking*, 22 (1), 65-76.
- Thurman, W. y Fisher, M. (1988). "Chickens, Eggs, and Causality, or Which Came First?", *American Journal of Agricultural Economics*, 70 (2), 237-238.
- Toda, H. y Yamamoto, T. (1995). "Statistical Inferences in Vector Autoregressions with Possibly Integrated Processes", *Journal of Econometrics*, 66, 225-250.
- Verbeek, M. (2004). *A Guide to Modern Econometrics*, 2a ed., New York, John Wiley.

- Wallis, K. (1989). "Macroeconomic Forecasting: A Survey", *The Economic Journal*, 99, 28-61.
- Watson, M. (1991). "Using Econometric Models to Predict Recessions", Federal Reserve Bank of Chicago, Economic Perspectives, noviembre-diciembre, 14-25.
- Woitek, U. (1998). "A Note on the Baxter-King Filter", mimeo, Glasgow, Economics Department, University of Glasgow.
- Wold, H. (1938). *A Study in the Analysis of Stationary Time Series*, 2a ed., Uppsala, Almquist and Wicksell.
- Wooldridge, J. (2009). *Introductory Econometrics: A Modern Approach*, 4a ed., South Western College Publishing.
- Zapata, H. y Rambaldi, A. (1997). "Monte Carlo Evidence on Cointegration and Causation", *Oxford Bulletin of Economics and Statistics*, 59, 285-298.
- Zarnowitz, V. y Lambros, L. (1987). "Consensus and Uncertainty in Economic Predictions", *Journal of Political Economy*, 95 (3), 591-621.



APÉNDICE A

Análisis espectral

El análisis espectral tiene que ver con vibraciones y oscilaciones, las cuales, en su forma más pura, son ondulaciones, sin quiebres abruptos que se repiten periódicamente a través del tiempo y que siempre mantienen la misma amplitud y frecuencia de oscilación, como la señal emitida por un diapasón. Estas ondulaciones se conocen como ondas senoidales (o cosenoidales) y son los componentes de prácticamente todas las señales conocidas, como el ruido blanco, la voz o la salida de un modelo ARMA. El análisis espectral separa una señal en las diversas ondas senoidales que la conforman.

Una ilustración familiar de este fenómeno consiste en el paso de un rayo de luz blanca por un prisma que lo separa en componentes de diferentes colores. Cada color corresponde a un tipo de vibración con características particulares. El análisis espectral se refiere a la teoría y técnica que permite descomponer una señal o serie de tiempo (luminosa, sonora, sensorial, económica, etc.) en componentes senoidales de diferentes frecuencias y amplitudes. En el análisis espectral importa menos la forma específica de la serie a lo largo del tiempo y más la identificación de las ondas senoidales presentes.

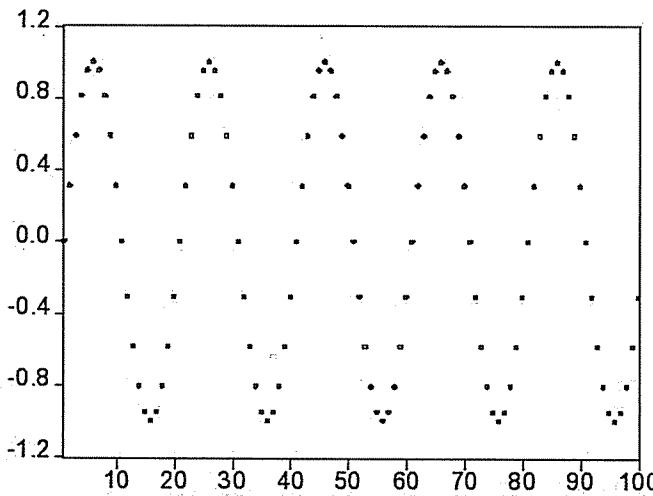
Funciones senoidales

Una función senoidal, como la de la Figura A.1, se representa por:

$$C \operatorname{sen}(\omega t + \phi) \quad (\text{A.1})$$

donde, en el caso discreto, que es el común en series de tiempo, la variable tiempo asume valores enteros, esto es $t = 0, 1, 2, 3, \dots$. Con respecto a los demás parámetros, C es un coeficiente llamado amplitud, positivo o negativo, que multiplica la señal senoidal; ω es la frecuencia angular dada en radianes (por ejemplo π radianes equivale a 180 grados), y ϕ es una constante llamada ángulo de fase que sirve para desplazar la señal en el tiempo, usualmente en una cantidad que se encuentra en el intervalo $-\pi$ y π radianes.

Figura A.1: $\operatorname{sen}(\pi t/10)$



La frecuencia angular ω también se puede escribir como $\omega = 2\pi f$, donde f es la frecuencia dada en ciclos por unidad de tiempo. Ciclo se refiere al parte de la señal que se repite indefinidamente; puede medirse de valle a valle, o pico a pico, o de cualquier punto al mismo punto. La duración del ciclo se define como periodo, el cual está dado en unidades de tiempo por ciclo, y equivale al inverso de la frecuencia f .

Es útil recordar algunas propiedades trigonométricas de las formas senoidales: $\sin(\omega t + 2\pi) = \sin(\omega t)$, $\cos(\omega t + 2\pi) = \cos(\omega t)$, $\sin^2\phi + \cos^2\phi = 1$, $\sin(\omega t + \pi/2) = \cos(\omega t)$, etc. La última propiedad nos dice que las ondas senoidales y cosenoidales son equivalentes, excepto por un ángulo de fase de $\pi/2$ radianes (90 grados), que no afecta la forma de la curva y sólo la desplaza un cuarto de ciclo en el tiempo.

La Figura A.1 muestra 100 observaciones de la función $\sin(\pi t/10)$, donde la amplitud C es unitaria. La curva se dibuja punteada para enfatizar su carácter discreto; esto es, para indicar que, aunque la señal es teóricamente continua, sólo hay observaciones para puntos equidistantes en el tiempo. Sin embargo, aunque las series sean discretas, la costumbre, que se seguirá aquí en adelante, es dibujar las curvas de manera continua uniendo los puntos observados. Cuanto mayor sea el coeficiente que multiplica al tiempo t (por ejemplo, si en lugar de $\pi/10$ fuera $\pi/4$) mayor es la frecuencia, lo que significa que la gráfica tendría más ciclos por unidad de tiempo. Si hiciéramos $C = 5$, por ejemplo, la amplitud de la onda se multiplicaría por cinco; los picos llegarían a $+5$ y los valles a -5 , en lugar de $+1$ y -1 como está en la Figura A.1.

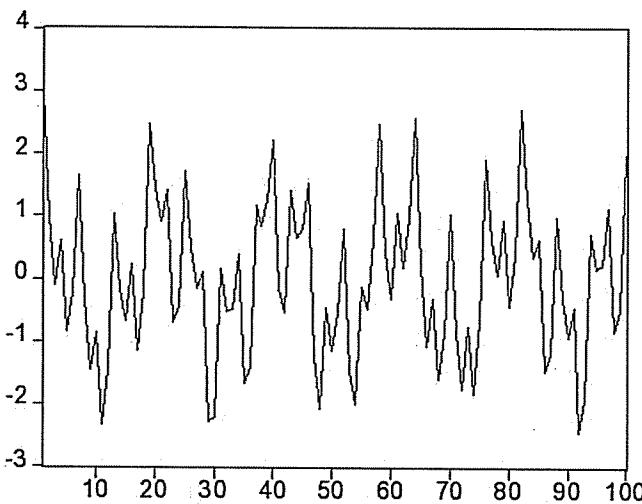
Representación senoidal

La suma de ondas senoidales o cosenoidales, perfectamente periódicas pero de diferentes frecuencias y amplitudes, genera resultados que pueden ser de apariencia muy diferentes a las ondas senoidales individuales que participan en la suma. Esto es cierto tanto para ondas discretas como continuas. Para ilustrar, la Figura A.2 muestra la suma de las tres ondas cosenoidales:

$$\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t) \quad (\text{A.2})$$

Si no supiéramos cómo se generó esta serie, podríamos confundirla con una serie de tiempo observada en la práctica, quizás cambios en el producto interno bruto (PIB), cambios en el precio de algún título bursátil, etc. De hecho, estas series de la vida real están compuestas por una infinidad de señales ondulatorias puras.

El objetivo del análisispectral es hacer lo contrario de lo hecho en la Figura A.2, donde combinamos varias series senoidales de frecuencia y amplitud conocida para obtener una serie de tiempo. En el

Figura A.2: $\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t)$ 

análisis espectral tomamos la serie de tiempo observada y tratamos de encontrar las ondas senoidales (frecuencias y amplitudes) que la componen; se trata de establecer cuáles frecuencias están presentes y en qué proporción. Esto se conoce como trabajar en el dominio de la frecuencia (las variables se grafican contra el eje de la frecuencia). El análisis de series de tiempo tradicional se hace en el dominio del tiempo (las variables se grafican contra el eje del tiempo).

El teorema de representación espectral dice que cualquier proceso estacionario hasta de orden 2 (estacionario en covarianza o débilmente estacionario) puede ser representado por una suma, posiblemente infinita, de señales senoidales y cosenoidales (Hamilton, 1994, capítulo 6). Este enunciado es la contraparte en el dominio de la frecuencia del teorema de representación de Wold utilizado en el dominio del tiempo.

Los orígenes del análisis espectral se encuentran en las series de Fourier, del matemático francés Jean Baptiste Joseph Fourier, quien a comienzos del siglo XIX descompuso señales periódicas en sumas de ondas senoidales y cosenoidales.

Una forma de hacer que la serie de la Figura A.2 se asemeje aún más a una serie de la vida real es sumándole una serie de ruido blanco ε_t , como la de la Figura A.3 (tomada de una normal estándar). La

serie resultante de esta suma se muestra en la Figura A.4.

Figura A.3: ε_t

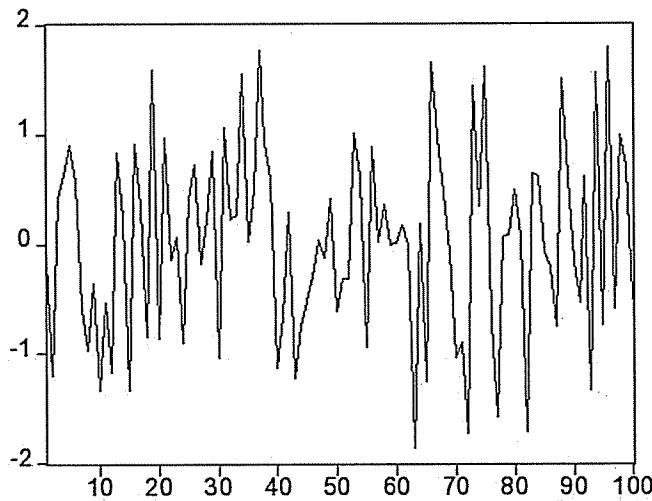
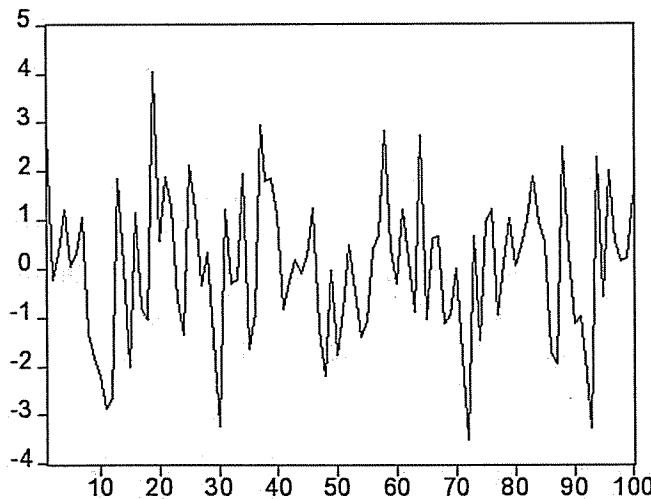


Figura A.4: $\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t) + \varepsilon_t$



Una explicación intuitiva de cómo podemos utilizar funciones, en general, y ondas senoidales, en particular, para representar otras fun-

ciones, se encuentra en Lathi (1968, capítulo 1). En principio, el procedimiento puede verse como una regresión (mínimos cuadrados ordinarios) de la función o serie que se va a explicar contra las funciones que se hayan escogido como explicativas.

Usando K términos podemos expresar una serie de tiempo x_t estacionaria de media cero como:

$$\begin{aligned} x_t &= C_1 \cos(\omega_1 t + \phi_1) + C_2 \cos(\omega_2 t + \phi_2) + \dots + C_K \cos(\omega_K t + \phi_K) + \varepsilon_t \\ &= \sum_{i=1}^K [C_i \cos(\omega_i t + \phi_i)] + \varepsilon_t \end{aligned} \quad (\text{A.3})$$

Donde C_i es la amplitud, ω_i es la frecuencia en radianes, ϕ_i es el ángulo de fase y ε_t es ruido con media 0 que recoge aquella parte de la serie que no es explicada por los K términos utilizados (en teoría, ε_t tiende a 0 si K tiende a infinito).

La expresión (A.3) puede simplificarse para facilitar la estimación, apelando a la identidad trigonométrica:

$$\cos(\omega_i t + \phi_i) = \cos(\omega_i t) \cos(\phi_i) - \sin(\omega_i t) \sin(\phi_i) \quad (\text{A.4})$$

Ello nos permite expresar la serie x_t en términos de senos y cosenos eliminando los ángulos de fase y quedando ésta:

$$x_t = \sum_{i=1}^K (A_i \cos(\omega_i t) + B_i \sin(\omega_i t)) + \varepsilon_t \quad (\text{A.5})$$

Donde $A_i = C_i \cos(\phi_i)$ y $B_i = -C_i \sin(\phi_i)$ son las amplitudes o coeficientes de las ondas.

Notamos que tanto x_t en el lado izquierdo como cada una de las ondas del lado derecho de la expresión son series de tiempo. En otras palabras, podemos imaginar la expresión (A.5) como una regresión donde x_t es la variable dependiente, y los K senos y cosenos son las variables explicativas. Podemos correr x_t contra estas ondas utilizando mínimos cuadrados ordinarios (MCO) y obtener coeficientes estimados \hat{A}_i y \hat{B}_i . Para efectuar la estimación por MCO tendríamos que conocer los ω_i contenidos en x_t con el fin de construir sus variables explicativas.

Supongamos, por el momento, que conocemos los ω_i . Entonces puede mostrarse (Priestley, 1981, capítulo 6) que los coeficientes estimados por MCO serán:

$$\hat{A}_i = \frac{2}{n} \sum_{t=1}^n x_t \cos(\omega_i t) \quad (\text{A.6a})$$

$$\hat{B}_i = \frac{2}{n} \sum_{t=1}^n x_t \sin(\omega_i t) \quad (\text{A.6b})$$

Donde n es el tamaño de la muestra (el número de observaciones de la serie x_t).

La matriz de variables explicativas de la regresión está formada por series senoidales y cosenoidales. Afortunadamente éstas son ortogonales entre sí y permiten obtener expresiones sencillas en la estimación. La ortogonalidad es similar a 0 covarianza, lo que se da si la suma del producto de las series es 0, como en efecto sucede en $\int_0^{2\pi} \cos(i\omega) \sin(j\omega) d\omega = 0$, para cualquier i y j enteros positivos. En consecuencia, como en cualquier regresión cuyas variables explicativas son ortogonales, los coeficientes en las expresiones (A.6a y A.6b) resultan proporcionales a la covarianza entre la variable explicativa y la variable dependiente, lo cual es intuitivamente satisfactorio.

En realidad no conocemos los ω_i , por lo que debemos ensayar una secuencia de valores de ω_i tomados a distancias cortas, esto es, hacemos un barrido de ω_i , y vamos registrando el comportamiento de \hat{A}_i y \hat{B}_i . Este es el principio del estimador denominado periodograma.

El periodograma

La gráfica de $\hat{A}_i^2 + \hat{B}_i^2$ (a veces la gráfica de $\sqrt{\hat{A}_i^2 + \hat{B}_i^2}$) contra ω_i se conoce como periodograma y se denota $I_n(\omega)$. Los valores de $I_n(\omega)$ serán diferentes de 0 cuando ω_i , en el eje horizontal, coincida con una de las frecuencias contenidas en x_t .

Nótese de (A.5) que $\hat{A}_i^2 + \hat{B}_i^2 = C_i^2 \cos^2 \phi + C_i^2 \sin^2 \phi = C_i^2$, de manera que, en efecto, estaríamos graficando una cantidad proporcional a la amplitud C_i de cada onda en (A.3) contra la frecuencia correspondiente. En cierta forma, estamos graficando el contenido de energía o potencia contra la frecuencia, lo que se conoce como el espectro de la serie.

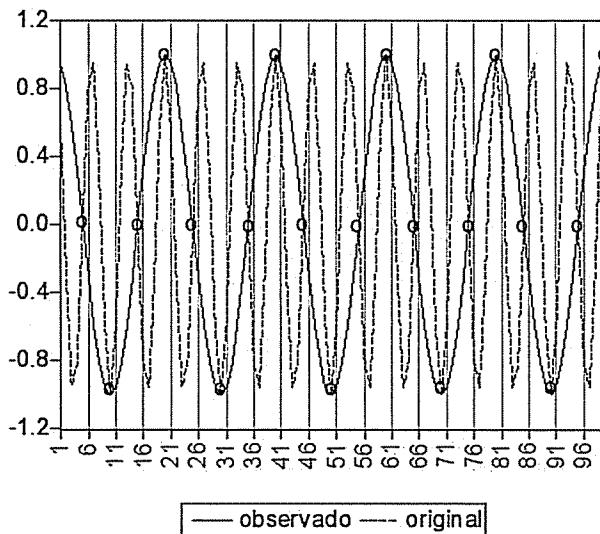
La mayor frecuencia posible (a la derecha en el eje horizontal) será la que corresponda a la llamada *frecuencia de Nyquist*. Si muestreamos la serie x_t cada unidad de tiempo, entonces el periodo mínimo que podemos detectar es de 2 unidades de tiempo ya que, para establecer un ciclo, necesitamos por lo menos detectar (muestrear) un valle y un pico. Si el periodo más corto es de 2 unidades de tiempo por ciclo, la frecuencia máxima será $f = 1/2$ ciclo por unidad de tiempo, o expresada en radianes será $\omega = 2\pi f = 2\pi(1/2) = \pi$.

Lo anterior implica que si contamos con muestras de una señal continua que no contenga frecuencias superiores a $f = 1/2$ ciclo por unidad de tiempo, entonces podemos reconstruir la señal completamente con base en las muestras. Esta es la razón por la cual basta transmitir sólo las muestras (observaciones) de una señal continua, y no la señal completa (lo que resultaría más costoso). Aunque a primera vista parece contraintuitivo, las muestras transmitidas son suficientes para recuperar la totalidad de la señal continua en el punto de recepción (véase, por ejemplo, Lathi, 1968, capítulo 1).

En efecto, una señal discreta puede verse como el muestreo de una señal continua. Y si la señal que estamos muestreando contiene frecuencias superiores a π , que es la máxima que podemos detectar, éstas se contabilizarán como de frecuencia inferior; esto es, para $0 \leq \omega \leq \pi$ las frecuencias $\omega + 2\pi, \omega + 4\pi, \omega + 6\pi, \dots$, se contabilizarán como de frecuencia ω radianes; o para $0 \leq f \leq 1/2$, las frecuencias $f + 1, f + 2, f + 3, \dots$, se contabilizarán como de frecuencia f ciclos por unidad de tiempo.

Este fenómeno se conoce como *aliasing*, y su explicación se ilustra en la Figura A.5, en la cual se muestra la señal original y los puntos muestreados. La clave está en entender que sólo observamos los valores que muestreamos, los cuales dibujan una serie de menor frecuencia que la original. La pérdida de precisión será poca si la frecuencia de muestreo es suficientemente alta como para captar todas las ondas con potencia significativa.

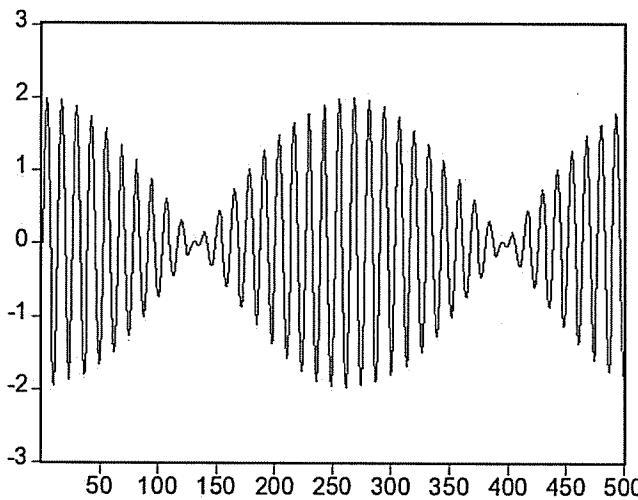
En la práctica, usualmente no tenemos control sobre la frecuencia de muestreo con la cual se recolectan los datos (diarios, mensuales, trimestrales, anuales, etc.) y debemos conjutar si dicha tasa de muestreo deja por fuera frecuencias importantes más altas. De cierta

Figura A.5: *Aliasing*

forma, este fenómeno corresponde al ubicuo problema en estadística de saber si una muestra es representativa de la población.

El efecto llamado *beats* es otro caso de combinación de ondas que se mueven en el mismo espacio o medio que produce una apariencia y efecto diferente a las series individuales. Se trata de la suma de dos ondas con frecuencias muy cercanas entre sí que produce un efecto de modulación que auditivamente se percibe como pulsaciones de baja frecuencia. Suponiendo dos frecuencias cercanas f_1 y f_2 , la frecuencia $(f_1 + f_2)/2$ se escuchará pulsando a la más baja frecuencia $|f_1 - f_2|$. La Figura A.6 ilustra este caso de *beats* o pulsos para 500 observaciones de la señal dada por $\sin(t/2) + \sin(t/2.1)$.

Por razones prácticas, la escogencia usual de frecuencias de estimación es $\omega_i = 2\pi q_i/n$, donde $q_i = 0, 1, 2, 3, \dots, n/2$, de manera que $\omega_i = \{0, \frac{2\pi}{n}, \frac{4\pi}{n}, \frac{6\pi}{n}, \dots, \pi\}$. Esta escogencia asegura que todos los senos y cosenos considerados tengan un número completo de ciclos en la muestra de tamaño n , y que las ordenadas del periodograma sean independientes una de otras, propiedad que resulta muy conveniente para pruebas de hipótesis.

Figura A.6: Beats generados por la señal $\sin(t/2) + \sin(t/2.1)$ 

Aunque el periodograma está definido entre $-\pi$ y π en radianes (o entre -0.5 y 0.5 en ciclos por unidad de tiempo), en razón de su simetría suele mostrarse sólo el lado positivo correspondiente al intervalo $\omega_i \in [0, \pi]$ en radianes (o $[0, .5]$ en ciclos por unidad de tiempo).

Como se mencionó, el periodograma $I_n(\omega)$, para cada ω , es proporcional al cuadrado de los coeficientes en (A.6) de la estimación por MCO, esto es:

$$I_n(\omega) = \Phi \left[\left(\sum_{t=1}^n x_t \cos \omega t \right)^2 + \left(\sum_{t=1}^n x_t \sin \omega t \right)^2 \right] \quad (\text{A.7})$$

Donde Φ es una constante de proporcionalidad. Esta expresión puede escribirse como el cuadrado de la magnitud o módulo del número complejo:

$$= \Phi \left| \sum_{t=1}^n x_t \cos \omega t - i \sum_{t=1}^n x_t \sin \omega t \right|^2 \quad (\text{A.8})$$

Donde $i = \sqrt{-1}$, notación que se mantendrá por ser la tradicional; se espera que no se confunda con el mismo símbolo i cuando éste sea utilizado para denotar algún i -ésimo elemento, subíndice o índice en los límites de una sumatoria.

A su vez, esta ecuación puede expresarse, usando la identidad de Euler $e^{\pm i\omega t} = \cos \omega t \pm i \sin \omega t$, como:

$$= \Phi \left| \sum_{t=1}^n x_t e^{-i\omega t} \right|^2 \quad \text{para} \quad \omega = 0, \frac{2\pi}{n}, \frac{4\pi}{n}, \frac{6\pi}{n}, \dots, \pi \quad (\text{A.9})$$

En la cual la utilización del signo negativo (en lugar del positivo) en la parte imaginaria se hace por convención (véase, por ejemplo, Schwarz y Shaw, 1975, capítulo 2). Una buena explicación de los números imaginarios y complejos en el contexto del análisis espectral se puede encontrar en Loy (2007).

La autocovarianza

Existe una relación muy interesante entre el periodograma y la función de autocovarianza $R(\tau)$ de la serie x_t (para detalles sobre la autocovarianza $R(\tau)$, véase el capítulo 2). Siguiendo a Priestley (1981, p. 399), y recordando que la magnitud al cuadrado de un número complejo es igual al número por su conjugado, la expresión (A.9) puede escribirse como:

$$= \Phi \left[\sum_{t=1}^n x_t e^{-i\omega t} \right] \left[\sum_{\nu=1}^n x_{\nu} e^{i\omega \nu} \right] = \Phi \left[\sum_{t=1}^n \sum_{\nu=1}^n x_t x_{\nu} \cos(t - \nu)\omega \right] \quad (\text{A.10})$$

Donde los términos senoidales se cancelan. Para proceder se define un nuevo índice $\tau = t - \nu$, de manera que τ toma los valores de $1 - n = -(n - 1)$ hasta $n - 1$, y, para τ fijo, t va desde 1 hasta $n - |\tau|$. Así:

$$= \Psi \sum_{\tau=-(n-1)}^{n-1} \left[\frac{1}{n} \sum_{t=1}^{n-|\tau|} x_t x_{t+|\tau|} \right] \cos \tau \omega = \Psi \sum_{\tau=-(n-1)}^{n-1} \hat{R}(\tau) \cos \tau \omega \quad (\text{A.11})$$

Donde Ψ es una constante de proporcionalidad y $\hat{R}(\tau)$ es la autocovarianza estimada bajo el supuesto $E x_t = 0$. Si la autocovarianza refleja una correlación positiva, como en un AR(1) con coeficiente positivo, habrá predominio de frecuencias bajas; si la autocovarianza refleja una

correlación negativa, como en un AR(1) con coeficiente negativo, habrá predominio de frecuencias altas.

Un modelo ARMA puede generar espectros con picos en una o varias frecuencias de resonancia (eigenvalores). De (A.11) se puede inferir que una serie de ruido tiene un espectro plano (independiente de ω), de manera que la energía se distribuye uniformemente en todas las frecuencias.

El espectro

En general los procesos estocásticos encontrados en la práctica no tendrán un juego de frecuencias senoidales fijas. El elemento aleatorio involucrado en las series reales lleva a pensar que presentan un rango de frecuencias alrededor de las que pudieran ser frecuencias principales, de manera que el espectro no será discreto sino continuo, con valores para todo ω_i .

Llamamos $s(\omega)$ a este espectro continuo de una serie x_t , similar a una distribución de probabilidad, donde $s(\omega)d\omega$ se interpreta como la contribución a la energía o potencia total de la serie, contenida en los componentes senoidales, cuyas frecuencias están entre ω y $\omega + d\omega$.

De manera similar a la ecuación (A.11), se puede mostrar (Granger y Newbold, 1986, capítulo 2) que, para series reales y considerando el intervalo $-\pi \leq \omega \leq \pi$, podemos escribir:

$$\begin{aligned} s(\omega) &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R(\tau) e^{-i\omega\tau} = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} R(\tau) \cos \omega\tau \\ &= \frac{\sigma_x^2}{2\pi} + \frac{1}{\pi} \sum_{\tau=1}^{\infty} R(\tau) \cos \omega\tau \end{aligned} \quad (\text{A.12})$$

Donde, a su vez:

$$R(\tau) = E(x_t x_{t-\tau}) = \int_{-\pi}^{\pi} e^{i\omega\tau} s(\omega) d\omega \quad (\text{A.13})$$

Así (A.13) es la función de autocovarianza de x_t . La expresión para $s(\omega)$ también puede escribirse en función de las autocorrelaciones, lo cual se logra dividiendo por la varianza σ_x^2 de la serie x_t . Al igual que la

función de autocorrelación, el espectro puede ayudar en la identificación del tipo de modelo (AR, MA, ARMA) que mejor se ajusta a los datos.

El periodograma es el estimador de $s(\omega)$, de la misma manera que el histograma puede ser considerado como un estimador de la función de densidad de probabilidad.

Propiedades de muestreo

El periodograma es un estimador asintóticamente insesgado de las frecuencias del espectro, pero es inconsistente en el sentido de que su varianza no disminuye a medida que el tamaño de la muestra aumenta. El problema radica en que la estimación para cada frecuencia es prácticamente independiente de las demás, y estamos estimando tantas autocorrelaciones como observaciones (esto se puede ver de (A.11), al recordar que $R(\tau) = R(-\tau)$), de manera que la gráfica del periodograma contra ω resulta errática y volátil.

Una solución es suavizar el periodograma utilizando filtros o ventanas similares al uso del promedio móvil centrado en el dominio del tiempo; de hecho, se utiliza un método similar para estimar funciones de densidad de probabilidad (*kernel estimation*). Debe advertirse, sin embargo, que suavizar el periodograma tiene su costo, ya que en muestras finitas puede generar sesgo en la estimación (y en infinitas si el suavizado se hace mal).

Para cada i se promedian los estimativos de $I_n(\omega)$ para frecuencias cercanas a ω_i , a la izquierda y a la derecha, a fin de darle mayor peso a $I_n(\omega_i)$ y menos a las demás frecuencias. La forma de las ponderaciones configura una ventana o *kernel* que se mueve a lo largo del eje ω y produce el periodograma suavizado. Hay muchas formas de ventanas, las hay rectangulares (como la de una distribución uniforme), triangulares, en forma de carpa, o no lineales, y otras (para ejemplos de ventanas y su diseño se puede consultar Stoica y Moses, 1997).

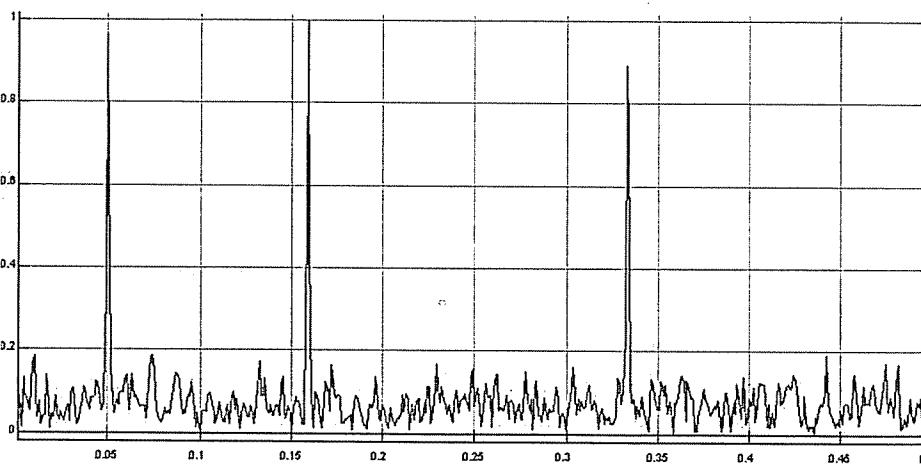
El *software* para el análisis espectral usualmente presenta una ventana (y su ancho) como *default*, pero permite al usuario escoger entre una gran variedad de opciones. Bajo condiciones generales, la estimación suavizada para cada ω_i es consistente y converge en distribución a una normal independientemente de la ventana escogida (aunque la escogencia de la ventana influye en la eficiencia estadística), lo cual permite realizar pruebas de hipótesis.

En muchos casos, las series encontradas en la práctica suelen ser no estacionarias; por ello es necesario estimar y eliminar la tendencia, diferenciar o tomar cambios porcentuales para obtener estacionariedad. En el caso de las series estacionales, éstas pueden verse como la suma de dos espectros, uno que corresponde al proceso sin el efecto estacional y otro que corresponde al efecto estacional. En general, el espectro de una suma de procesos estocásticos estacionarios independientes de media cero es la suma de sus espectros.

Ejemplos

Para la estimación de los espectros se utilizó el programa SigView de SignalLab, versión 1.98. Todos los espectros, excepto el de la Figura A.7, fueron suavizados por medio de la ventana Hanning, que viene como *default*.

Figura A.7: Espectro sin suavizar de $\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t) + \varepsilon_t$

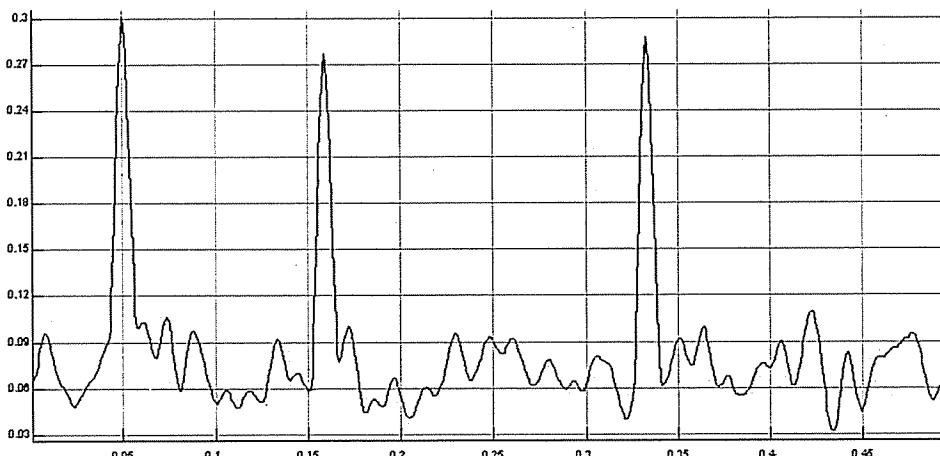


En la Figura A.7 se observa el espectro estimado sin suavizar para 100 observaciones de la serie $\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t) + \varepsilon_t$ (de las cuales sólo se habían mostrado 100 en la Figura A.4). Como es de esperarse, se ven tres picos pronunciados. El primero corresponde a la frecuencia más baja, 0.05 ciclos por unidad de tiempo, generada por el término $\cos(\pi t/10)$, la cual se despeja de $2\pi f = \pi/10$ o $f = 1/20 = 0.05$.

El segundo pico corresponde al término $\cos(t)$ y se encuentra cerca de $f = 1/2\pi = 0.159$. La frecuencia más alta, cerca de $f = 1/3 = 0.33333$, corresponde al término $\cos(2\pi t/3)$.

El espectro del término de ruido ε_t incluido en la ecuación de la serie refleja energía repartida de manera más o menos uniforme en todas las frecuencias, el cual se suma al espectro de las otras tres señales cosenoidales de la serie. La Figura A.8 muestra el espectro suavizado de la Figura A.7. Aunque esencialmente contiene la misma información, la amplitud de los picos varía, porque ahora las ordenadas del espectro corresponden a un promedio de puntos adyacentes según las ponderaciones de la ventana utilizada.

Figura A.8: Espectro suavizado de $\cos(\pi t/10) + \cos(2\pi t/3) + \cos(t) + \varepsilon_t$



La Figura A.9 muestra una serie comúnmente utilizada como *benchmark* en el análisis de series de tiempo. Se trata de la serie de linceos atrapados anualmente entre 1821 y 1934 en el noroeste canadiense, en el distrito del río Mackenzie. La serie, que puede consultarse en Campbell y Walker (1977), es obviamente periódica y evoca el tema de la interacción entre la presa y su predador. La Figura A.10 muestra el espectro para la serie de linceos cuyo eje horizontal de frecuencia está

marcado en ciclos por año. El pico se observa cerca de 0.11 ciclos por año, lo que equivale a un periodo un poco superior a 9 años.

Figura A.9: Linces atrapados anualmente (*lynx trappings*)

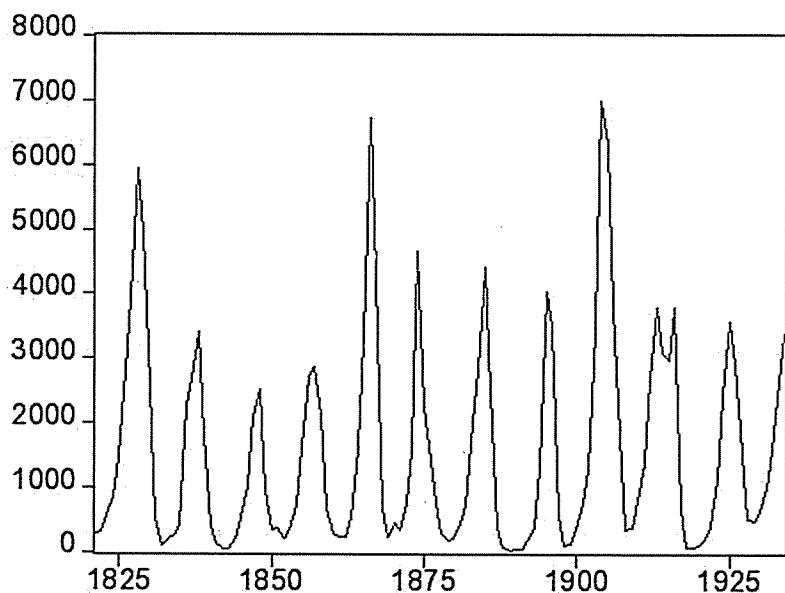
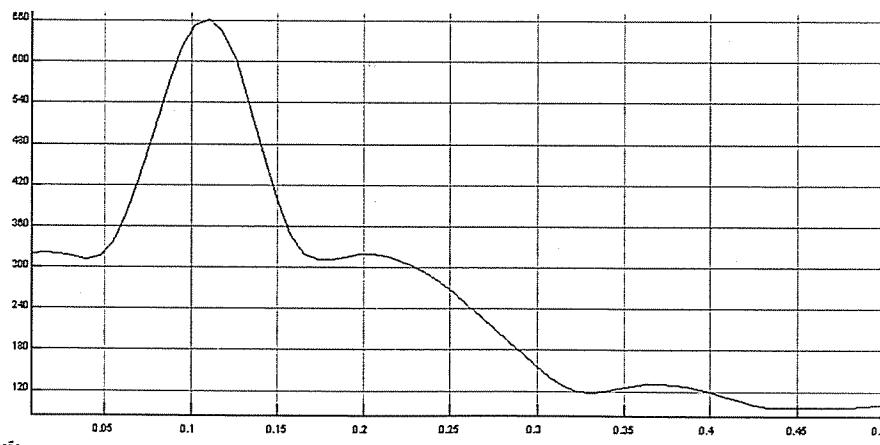
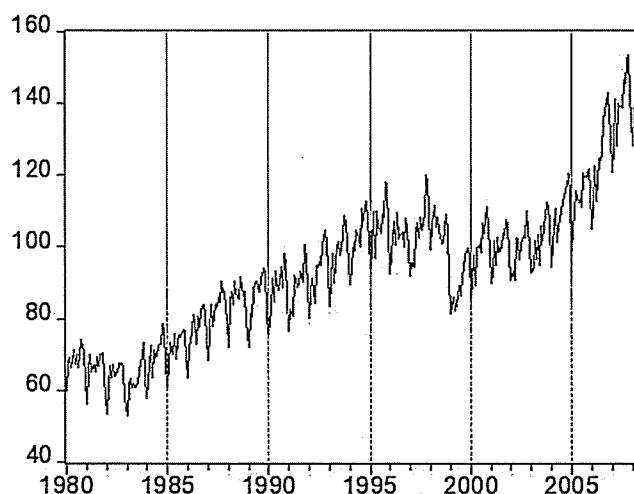


Figura A.10: Espectro de la serie de linces atrapados anualmente



La Figura A.11 muestra la serie de producción real mensual de la muestra mensual manufacturera publicada por el Departamento Administrativo Nacional de Estadística (DANE), sin trilla, desde 1980. La serie es no estacionaria con evidentes periodicidades. La Figura A.12 muestra el cambio porcentual en la serie, y la Figura A.13, su espectro. Un tema de interés es si existen componentes en la serie que puedan reflejar la presencia de ciclos económicos, esto es, de componentes con períodos de varios años. La frecuencia más baja que se observa en el espectro está alrededor de 0.033 ciclos por mes o un periodo cercano a los 30 meses, lo que podría sugerir un ciclo económico de 2.5 años. Sin embargo, su potencia (magnitud) es relativamente baja.

Figura A.11: Producción real mensual manufacturera



El primer componente significativo está a una frecuencia de 0.082 ciclos por mes, que representa el ciclo anual. Le siguen componentes a frecuencias aproximadamente armónicas de 0.17 ciclos por mes o períodos de 6 meses, de 0.25 ciclos por mes o períodos de 4 meses, de 0.34 o períodos de tres meses, y dos frecuencias de 0.42 y 0.47 ciclos por mes, con períodos entre 2 y 2.5 meses.

Podría obtenerse una primera aproximación de la serie del cambio porcentual en la producción real manufacturera (siguiendo la ecuación [A.5] a través de una suma de señales senoidales y cosenoideales de

Figura A.12: Variación porcentual en la producción real mensual manufacturera

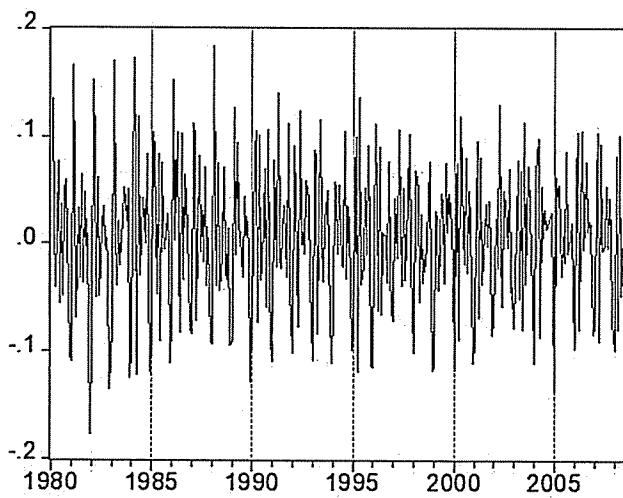
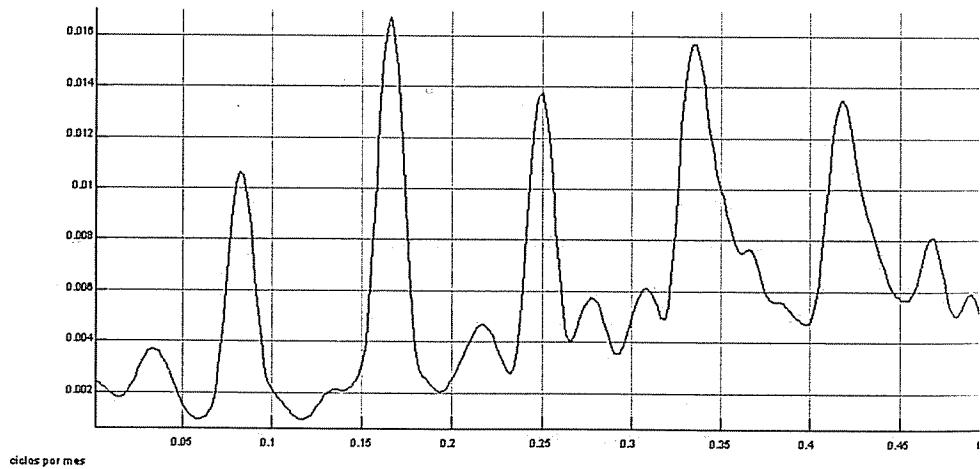


Figura A.13: Espectro de la variación porcentual en la producción real manufacturera



diferentes amplitudes y períodos cercanos a los 12, 6, 4, 3 y 2.3 meses, o podrían incluirse algunas de estas frecuencias como variables adicionales en una estimación tradicional de esta serie en el domino del tiempo.

Referencias

- Campbell, M. J. y Walker, A. M. (1977). "A Survey of Statistical Work on the MacKenzie River Series of Annual Canadian Lynx Trappings for the Years 1821-1934 and a New Analysis", *Journal of the Royal Statistical Society*, 140 (4), 411-431.
- Granger, C. W. y Newbold, P. (1986). *Forecasting Economic Time Series*, New Jersey, Academic Press.
- Hamilton, J. (1994). *Time Series Analysis*, New Jersey, Princeton University Press.
- Lathi, B. P. (1968). *Communication Systems*, New York, John Wiley and Sons.
- Loy, G. (2007). *Musimathics: The Mathematical Foundations of Music*, Boston, MIT Press.
- Montenegro, Á. (2007). *Series de Tiempo*, Bogotá, Facultad de Economía de la Universidad Javeriana.
- Priestley, M. B. (1981). *Spectral Analysis and Time Series*, New Jersey, Academic Press.
- Schwarz, M. y Shaw, L. (1975). *Signal Processing: Discrete Spectral Analysis, Detection, and Estimation*, New York, McGraw Hill.
- Stoica, P. y Moses, R. (1997). *Introduction to Spectral Analysis*, New Jersey, Prentice Hall.

APÉNDICE B

Filtro de Kalman

El filtro de Kalman es un algoritmo para procesar datos de manera iterativa. No requiere reprocesar todas las observaciones anteriores cada vez que se actualiza. Sus fundamentos se desarrollaron en ingeniería y su presentación aún conserva gran parte de la notación y motivación original. El filtro sirve para estimar y predecir el movimiento de una variable que no observamos directamente, pero cuyo efecto medimos, contaminado por ruido, a través de otras variables.

La literatura económica reciente muestra un creciente interés, teórico y práctico, en la utilización del filtro de Kalman como técnica econométrica. Greenslade, Pierse y Saleheen (2003) lo aplican para estimar la tasa de desempleo que no acelera la inflación (Nairu); Akerlof *et al.* (2000) lo utilizan para terciar en la controversia sobre la curva de Phillips; Engel y Kim (1999) lo emplean para identificar el componente transitorio y permanente de la tasa de cambio; Garratt y Hall (1996) lo usan para construir un indicador de la actividad económica; Tanizaki (1993) lo combina con una variable dependiente binaria para estimar una función de exceso de demanda de dinero; Crafts, Leybourne y Mills (1989) lo aplican a la cliometría y concluyen, contrario a la sabiduría popular, que la tendencia del producto no cayó en el siglo XIX en Inglaterra y Francia; Fama y Gibbons (1982) lo utilizan para estimar la

tasa de interés real *ex ante*, la cual es no observable, y LeRoy y Waud (1977) lo usan para investigar la relación entre los instrumentos y los objetivos de la política monetaria.

En este documento intentamos una derivación del filtro que, por su notación y método, puede resultar más accesible a los economistas y cuyo requisito es el conocimiento de econometría matricial básica. La palabra *filtrar* en ingeniería equivale a la palabra *estimar* en econometría. Una versión anterior de este texto se encuentra en Montenegro (1991). Además del texto original de Kalman (1960) y de Kalman y Bucy (1961), algunas referencias sobre este tema son Maybeck (1971), Granger y Newbold (1986), Jones (1985), Harvey (1989), Meinhold y Singpurwalla (1983), quienes presentan un enfoque bayesiano de la derivación del filtro, y Hamilton (1994). En Priestley (1981) se derivan las ecuaciones utilizando una técnica conocida como análisis de factores (*factor analysis*). Así mismo, internet también es una buena fuente sobre el tema.

El filtro de Kalman produce estimativos que se actualizan con la llegada de cada nueva observación y puede modelar sistemas cuyos parámetros cambian a través del tiempo. Una diferencia con el método de regresión, más ampliamente conocido por los economistas, es que el filtro de Kalman puede dar más o menos peso al estado actual o reciente de un proceso, mientras que la regresión generalmente promedia sobre todo el periodo de la muestra. Difiere también de los modelos de regresión de coeficientes variables y estocásticos (por ejemplo, Hildreth-Houck o Cooley-Prescott) en que permite una mayor generalidad en la especificación estocástica de los parámetros. En espíritu es similar a los métodos de suavizado exponencial con tendencia (por ejemplo, Holt-Winters) excepto que, además, permite la actualización de los parámetros.

Estructura del modelo estado-espacio

El modelo estado-espacio (*state-space*) dentro del cual se enmarca la formulación del filtro supone la existencia de una o más variables, o parámetros variables, con los cuales podemos describir completamente el estado en que se encuentra algún sistema en un momento dado; por

eso se llaman variables *de estado*. El conjunto de variables de estado se conoce como el *vector de estado* y lo denotamos β_t .

El modelo supone que β_t no es observable directamente y que en su lugar observamos y_t , que puede ser un vector o un escalar según el caso, donde y_t es una combinación lineal de las variables de estado β_t más un error de medición ε_t (por ejemplo, β_t puede ser la posición de un misil enemigo que no se observa directamente, sino a través de las mediciones de un radar contaminadas por ruido, y_t). Según esto, la ecuación que relaciona las variables es:

$$y_t = X_t \beta_t + \varepsilon_t \quad (\text{ecuación de medición}) \quad (B.1)$$

Donde la matriz X_t se supone conocida y el error de medición ε_t se distribuye normalmente, con media 0 y varianza conocida σ_ε^2 , esto es, $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$. A esta ecuación también se le pueden agregar variables exógenas. Se postula, además, que conocemos la forma como β_t cambia a través del tiempo:

$$\beta_t = T_t \beta_{t-1} + d_t + G_t w_t \quad (\text{ecuación de transición}) \quad (B.2)$$

Donde la matriz de transición T_t , el vector d_t de controles externos y la matriz G_t se suponen conocidos y w_t se distribuye normalmente con media cero y varianza conocida σ_w^2 , esto es, $w_t \sim N(0, \sigma_w^2)$. La especificación de los errores puede generalizarse de manera que σ_ε^2 y σ_w^2 dependan del tiempo, en cuyo caso se les agregaría un subíndice t . Notamos que X_t y T_t tienen subíndice de tiempo y son conocidas.

La ecuación (B.1), llamada de medición (o de observaciones), y la ecuación (B.2), llamada de transición (o del sistema), constituyen la formulación general del modelo de estado-espacio. El filtro de Kalman fue diseñado para estimar β_t en este modelo mediante un proceso iterativo.

En el caso univariado, donde y_t es un escalar, tendríamos que X_t sería un vector transpuesto; β_t , un vector, y ε_t , un escalar.

La ecuación (B.1) es similar al modelo lineal clásico y la ecuación (B.2) corresponde a un proceso markoviano, ambos con coeficientes variables. Aunque no es evidente a primera vista, la formulación del modelo estado-espacio es una generalización de varias especificaciones, como el ARMA(p, q) o Box y Jenkins, modelo de suavizado exponencial

del tipo Holt-Winters y el modelo de regresión. El filtro de Kalman, como generalización de estos modelos, puede realizar muchas de sus funciones (para una ampliación y demostración de este punto, véase Watson, 1983).

Por ejemplo, la formulación estado-espacio para el modelo de regresión con coeficientes constantes sería:

$$y_t = x'_t \beta_t + \varepsilon_t \quad (\text{ecuación de medición})$$

$$\beta_t = \beta_{t-1} \quad (\text{ecuación de transición})$$

Donde x_t es un vector y donde definimos $T_t = 1$, $d_t = 0$, $w_t = 0$. Consideremos el siguiente modelo AR(2):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

Y su formulación estado-espacio:

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \quad (\text{ecuación de medición})$$

$$\begin{pmatrix} y_{t+1} \\ y_t \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{t+1} \\ 0 \end{pmatrix} \quad (\text{ecuación de transición})$$

Consideremos el siguiente modelo ARMA(2, 1):

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + w_t + g w_{t-1}$$

Tal modelo puede formularse en estado-espacio como sigue:

$$y_t = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + g w_t \end{pmatrix} \quad (\text{ecuación de medición})$$

$$\begin{pmatrix} y_t \\ \phi_2 y_{t-1} + g w_t \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \phi_2 y_{t-2} + g w_{t-1} \end{pmatrix} + \begin{pmatrix} 1 \\ g \end{pmatrix} w_t \quad (\text{ec. de trans.})$$

Donde definimos:

$$\beta_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + g w_t \end{pmatrix}, \quad T_t = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix}, \quad G_t = \begin{pmatrix} 1 \\ g \end{pmatrix}$$

$$X_t = (1, 0), \quad d_t = 0, \quad \varepsilon_t = 0$$

Alternativamente, el modelo ARMA(2, 1) puede representarse de la siguiente forma:

$$y_t = \phi_1 + \beta_{1t} + \phi_4\beta_{2t} \quad (\text{ecuación de medición})$$

$$\begin{pmatrix} \beta_{1t} \\ \beta_{2t} \end{pmatrix} = \begin{pmatrix} \phi_2 & \phi_3 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \beta_{1t-1} \\ \beta_{2t-1} \end{pmatrix} + \phi_5 \begin{pmatrix} w_t \\ 0 \end{pmatrix} \quad (\text{ecuación de transición})$$

Al reemplazar β_{1t} y β_{2t} de la ecuación de transición en la de medición y pasando ϕ_1 al lado izquierdo, tenemos:

$$\begin{aligned} y_t - \phi_1 &= \phi_2\beta_{1t-1} + \phi_3\beta_{2t-1} + \phi_5w_t + \phi_4\phi_2\beta_{1t-2} + \phi_4\phi_3\beta_{2t-2} \\ &\quad + \phi_4\phi_5w_{t-1} \\ &= \phi_2 \left(\beta_{1t-1} + \frac{\phi_3}{\phi_2} \beta_{2t-1} \right) + \phi_5w_t + \phi_4\phi_2 \left(\beta_{1t-2} + \frac{\phi_3}{\phi_2} \beta_{2t-2} \right) \\ &\quad + \phi_4\phi_5w_{t-1} \end{aligned}$$

Finalmente, notando de la ecuación de medición que las expresiones en paréntesis representan rezagos de la variable del lado izquierdo, obtenemos el modelo ARMA(2, 1) deseado:

$$= \phi_2(y_{t-1} - \phi_1) + \phi_4\phi_2(y_{t-2} - \phi_1) + \phi_5w_t + \phi_4\phi_5w_{t-1}$$

Otros ejemplos de representaciones de estado-espacio y de formulaciones alternativas pueden consultarse en Judge *et al.* (1985, p. 981), Chatfield (1989), Harvey (1982), Schwartz y Shaw (1975) y Hamilton (1994).

Ecuaciones del filtro de Kalman

Definimos como I_{t-1} el conjunto de información en el tiempo $t-1$, el cual incluye la información que tengamos sobre el modelo más la información contenida en las observaciones $y_{t-1}, y_{t-2}, y_{t-3}, \dots$. Similamente, I_t será el conjunto de información en el tiempo t , el cual

incluye la información sobre el modelo más la información contenida en $y_t, y_{t-1}, y_{t-2}, \dots$

El problema de estimación de la formulación estado-espacio se reduce básicamente a lo siguiente: observamos la variable y_t , pero queremos β_t , la cual no observamos y debemos estimar con base en la información disponible I_t .

Para resolver este problema apelamos a un conocido resultado de la teoría multivariada (Priestley 1981, p. 76; Granger y Newbold, 1986, p. 124; Lindgren, 1976, p. 476), el cual, para el caso que nos ocupa, dice que el estimador Φ , que minimiza el error cuadrado medio, $E [\beta_t - \Phi(I_t)]^2$, es $\Phi(I_t) = E(\beta_t|I_t)$. En otras palabras, la mejor forma de estimar β_t , de manera que el error cuadrado medio sea mínimo, es encontrando el valor esperado de β_t dado I_t . Los pasos requeridos para encontrar $E(\beta_t|I_t)$ se desarrollan a continuación.

Supongamos que el valor inicial β_0 se toma de alguna distribución normal independiente de ε_t y w_t . Sabemos que una combinación lineal de variables aleatorias normales, dependientes o independientes, es normal (véase Priestley, 1981, p. 91), lo cual nos permite afirmar que β_1 será normal, porque según la ecuación (B.2) resulta de una combinación lineal de variables normales; igualmente para β_2, β_3, \dots y, en general, para β_t .

Por ser sus componentes normales e independientes, el vector (β_t, ε_t) tendrá una distribución normal conjunta dada por el producto de las distribuciones marginales de β_t y ε_t . Además, cualquier otra combinación lineal de este vector también tendrá una distribución normal conjunta (véase Kendall y Stuart, 1963, p. 350; Lindgren, 1976, p. 472). En consecuencia, el vector (y_t, β_t) , mostrado aquí como una combinación lineal de β_t y ε_t ,

$$\begin{pmatrix} y_t \\ \beta_t \end{pmatrix} = \begin{pmatrix} X_t & M \\ N & 0 \end{pmatrix} \begin{pmatrix} \beta_t \\ \varepsilon_t \end{pmatrix}$$

(donde M y N son matrices o vectores unitarios usados para ilustrar), tendrá una distribución normal conjunta dado I_{t-1} de la forma:

$$\begin{pmatrix} \beta_t \\ y_t \end{pmatrix}_{I_{t-1}} \sim N \left\{ \begin{pmatrix} E(\beta_t|I_{t-1}) \\ E(y_t|I_{t-1}) \end{pmatrix}, \begin{pmatrix} \text{VAR}(\beta_t|I_{t-1}) & \text{COV}(\beta_t, y_t|I_{t-1}) \\ \text{COV}(y_t, \beta_t|I_{t-1}) & \text{VAR}(y_t|I_{t-1}) \end{pmatrix} \right\} \quad (\text{B.3})$$

Estrictamente hablando, los términos llamados VAR y COV son matrices de varianza-covarianza para el caso general vectorial. Los términos de (B.3) se explican a continuación. Al tomar el valor esperado de (B.2) y (B.1) tenemos respectivamente que:

$$E(\beta_t|I_{t-1}) = T_t E(\beta_{t-1}|I_{t-1}) + d_t \quad (\text{B.4})$$

$$E(y_t|I_{t-1}) = X_t E(\beta_t|I_{t-1}) \quad (\text{B.5})$$

Si recordamos la manipulación cuadrática de matrices y que la varianza de una suma de variables aleatorias no correlacionadas es la suma de sus varianzas, encontramos de las ecuaciones (B.2) y (B.1), respectivamente, que:

$$\text{VAR}(\beta_t|I_{t-1}) = T_t \text{VAR}(\beta_{t-1}|I_{t-1}) T_t' + G_t \sigma_w^2 G_t' \quad (\text{B.6})$$

$$\text{VAR}(y_t|I_{t-1}) = X_t \text{VAR}(\beta_t|I_{t-1}) X_t' + \sigma_\varepsilon^2 \quad (\text{B.7})$$

Para encontrar las expresiones de los dos términos llamados COV en (B.3) apelamos a un conocido resultado sobre la covarianza de dos variables aleatorias que, junto con (B.5), nos permite escribir:

$$\begin{aligned} \text{COV}(y_t, \beta_t|I_{t-1}) &= E(y_t \beta_t'|I_{t-1}) - E(y_t|I_{t-1}) E(\beta_t'|I_{t-1}) \\ &= E(y_t \beta_t'|I_{t-1}) - X_t E(\beta_t|I_{t-1}) E(\beta_t'|I_{t-1}) \end{aligned} \quad (\text{B.8})$$

Por otro lado, posmultiplicando (B.1) por β_t' y tomando expectativas y apelando a otro conocido resultado, esta vez, sobre la varianza de dos variables aleatorias, obtendremos:

$$\begin{aligned} E(y_t \beta_t'|I_{t-1}) &= E[(X_t \beta_t \beta_t' + \varepsilon_t \beta_t')|I_{t-1}] = X_t E(\beta_t \beta_t'|I_{t-1}) \\ &= X_t [\text{VAR}(\beta_t|I_{t-1}) + E(\beta_t|I_{t-1}) E(\beta_t'|I_{t-1})] \\ &= X_t \text{VAR}(\beta_t|I_{t-1}) + X_t E(\beta_t|I_{t-1}) E(\beta_t'|I_{t-1}) \end{aligned} \quad (\text{B.9})$$

De manera que (B.8) queda:

$$\begin{aligned} \text{COV}(y_t, \beta_t|I_{t-1}) &= X_t \text{VAR}(\beta_t|I_{t-1}) + X_t E(\beta_t|I_{t-1}) E(\beta_t'|I_{t-1}) \\ &\quad - X_t E(\beta_t|I_{t-1}) E(\beta_t'|I_{t-1}) \\ &= X_t \text{VAR}(\beta_t|I_{t-1}) \end{aligned} \quad (\text{B.10})$$

Igualmente, la otra covarianza será –notando que $\text{VAR}(\beta_t|I_{t-1})$ es una matriz simétrica y por tanto igual a su transpuesta–:

$$\text{COV}(\beta_t, y_t|I_{t-1}) = \text{VAR}(\beta_t|I_{t-1})X'_t \quad (\text{B.11})$$

Con lo cual completamos la explicación de los términos de (B.3). hora, utilizando otro resultado de la teoría multivariada (Pristley, 1981, pp. 84-85; Anderson, 1958, pp. 28-29; Harvey, 1989, p. 165), procedemos a encontrar la función de distribución condicional de la variable β_t .

Según este resultado, cuando dos variables, en este caso β_t y y_t , tienen una distribución normal conjunta como la descrita en la ecuación (B.3), entonces la distribución de β_t dado y_t (la más reciente medición) I_{t-1} , será:

$$\beta_t|y_t, I_{t-1} \sim$$

$$\begin{aligned} & \mathcal{N}\left\{ E(\beta_t|I_{t-1}) + \text{COV}(\beta_t, y_t|I_{t-1}) [\text{VAR}(y_t|I_{t-1})]^{-1} [y_t - E(y_t|I_{t-1})], \right. \\ & \left. \text{VAR}(\beta_t|I_{t-1}) - \text{COV}(\beta_t, y_t|I_{t-1}) [\text{VAR}(y_t|I_{t-1})]^{-1} \text{COV}(y_t, \beta_t|I_{t-1}) \right\} \end{aligned} \quad (\text{B.12})$$

Notamos que decir β_t dado y_t e I_{t-1} es lo mismo que decir β_t dado ; así el “mejor” estimador de β_t que veníamos buscando corresponde la media de la distribución normal de la expresión (B.12), esto es:

$$E(\beta_t|I_t) = E(\beta_t|I_{t-1}) + \text{COV}(\beta_t, y_t|I_{t-1}) [\text{VAR}(y_t|I_{t-1})]^{-1} [y_t - E(y_t|I_{t-1})] \quad (\text{B.13})$$

Éste tendrá una varianza que corresponde a la varianza de la distribución normal descrita en (B.12), esto es:

$$\text{AR}(\beta_t|I_t) = \text{VAR}(\beta_t|I_{t-1}) - \text{COV}(\beta_t, y_t|I_{t-1}) [\text{VAR}(y_t|I_{t-1})]^{-1} \text{COV}(y_t, \beta_t|I_{t-1}) \quad (\text{B.14})$$

Al reemplazar los valores indicados en (B.5), (B.7), (B.10) y (B.11) (B.13) y (B.14), obtenemos las ecuaciones llamadas de actualización:

$$E(\beta_t|I_t) = E(\beta_t|I_{t-1}) + K_t [y_t - X_t E(\beta_t|I_{t-1})] \quad (\text{B.15})$$

$$\text{VAR}(\beta_t|I_t) = \text{VAR}(\beta_t|I_{t-1}) - K_t X_t \text{VAR}(\beta_t|I_{t-1}) \quad (\text{B.16})$$

Donde:

$$K_t = \text{VAR}(\beta_t|I_{t-1}) X'_t [X_t \text{VAR}(\beta_t|I_{t-1}) X'_t + \sigma_\varepsilon^2]^{-1} \quad (\text{B.17})$$

Las cuales, junto con (B.4) y (B.6), llamadas de predicción, y repetidas aquí, conforman el filtro de Kalman:

$$E(\beta_t|I_{t-1}) = T_t E(\beta_{t-1}|I_{t-1}) + d_t \quad (\text{B.4})$$

$$\text{VAR}(\beta_t|I_{t-1}) = T_t \text{VAR}(\beta_{t-1}|I_{t-1}) T'_t + G_t \sigma_w^2 G'_t \quad (\text{B.6})$$

El proceso iterativo del filtro se inicia en $t = 0$ a partir de $E(\beta_0|I_0)$ y $\text{VAR}(\beta_0|I_0)$, dados inicialmente. Con estos valores encontramos $E(\beta_1|I_0)$ y $\text{VAR}(\beta_1|I_0)$ de (B.4) y (B.6); luego, una vez conocida la observación y_1 , de (B.15) y (B.16) encontramos $E(\beta_1|I_1)$ y $\text{VAR}(\beta_1|I_1)$.

Ahora con $E(\beta_1|I_1)$ y $\text{VAR}(\beta_1|I_1)$ iniciamos la segunda iteración haciendo la predicción $E(\beta_2|I_1)$ y su varianza $\text{VAR}(\beta_2|I_1)$ de (B.4) y (B.6), y así sucesivamente. Los valores iniciales de $E(\beta_0|I_0)$ y $\text{VAR}(\beta_0|I_0)$ deben escogerse de la manera menos arbitraria posible, apelando a la experiencia del investigador y al conocimiento que se tenga sobre el sistema bajo estudio, sea este económico, físico, etc.

El investigador puede escoger $E(\beta_0|I_0)$ y luego escoger $\text{VAR}(\beta_0|I_0)$ de manera que refleje su propia incertidumbre acerca de ese valor inicial $E(\beta_0|I_0)$. Otra posibilidad, particularmente atractiva para los economistas, es empezar utilizando las estimaciones de una regresión basada en las observaciones disponibles. Obviamente la primera iteración del filtro de Kalman se hace con la próxima observación disponible.

La expresión K_t de (B.17) se conoce como la ganancia del filtro de Kalman. Quizás intuitivamente sea más clara de la ecuación (B.13), donde observamos que la parte que corresponde a K_t : $\text{COV}(\beta_t, y_t|I_{t-1}) [\text{VAR}(y_t|I_{t-1})]^{-1}$, es el coeficiente de una regresión de $\beta_t - E(\beta_t|I_{t-1})$ como variable dependiente contra $y_t - E(y_t|I_{t-1})$ como variable explicativa. El vector K_t tiene un componente para cada variable de estado β_t , componente que, en general, tiende a ser proporcional a la varianza de la variable de estado correspondiente.

Si se trata de modelar una serie no estacionaria no basta, por ejemplo, con hacer que algunos elementos de T_t en (B.2) sean mayores que 1, sino que, si es del caso, es importante especificar varianzas que aumenten con el tiempo.

Ejemplo 1

Supongamos que se tienen datos agregados de ventas totales y del salario total pagado a vendedores en alguna actividad comercial y que estas variables se relacionan a través de la siguiente ecuación:

$$s_t = b_t + c_t \nu_t + \varepsilon_t \quad (\text{B.18})$$

Donde s_t es el salario total pagado en el tiempo t , b_t es el básico, ν_t el volumen de ventas totales, c_t es la comisión sobre ventas y ε_t es un término de error. En terminología del filtro de Kalman, esta sería una ecuación de medición.

Las variables de estado que se van a estimar son b_t y c_t , las cuales supone siguen la siguiente ecuación de transición:

$$\begin{pmatrix} b_t \\ c_t \end{pmatrix} = \begin{pmatrix} b_{t-1} \\ c_{t-1} \end{pmatrix} + \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} \quad (\text{B.19})$$

Donde $w'_t = (w_{1t}, w_{2t})$ es un vector de ruido. Utilizando la terminología del texto, podemos decir que $d_t = 0$ y que:

$$X_t = \begin{pmatrix} 1 & \nu_t \end{pmatrix} \quad T_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad G_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad (\text{B.20})$$

La serie s_t fue construida a partir de 10 observaciones de ε_t tomadas de una distribución normal con media cero. Los datos se muestran en Cuadro B.1. Observamos que en $t = 6$ tanto b_t como c_t duplican su valor pasando de 2 a 4 y de 0.1 a 0.2 respectivamente.

Cuadro B.1: Cifras para el ejemplo 1

t	s_t	b_t	c_t	ν_t	ε_t	$b_t(\text{est})$	$c_t(\text{est})$
1	3.65	2	0.1	20	-0.35	1.6	0.07
2	5.91	2	0.1	40	-0.09	1.8	0.10
3	5.58	2	0.1	30	0.58	2.1	0.10
4	6.42	2	0.1	45	-0.08	2.0	0.10
5	7.51	2	0.1	55	0.01	2.0	0.10
6	7.35	4	0.2	15	0.35	5.2	0.09

Continúa

Continación

7	8.97	4	0.2	25	-0.03	5.7	0.10
8	14.37	4	0.2	50	0.37	5.5	0.17
9	5.77	4	0.2	10	-0.23	4.3	0.18
10	11.17	4	0.2	35	0.17	4.4	0.19

Este es un salto inusualmente grande para series de tiempo que siguen la ecuación (B.19), pero se utiliza aquí con el objeto de ilustrar la habilidad del filtro de Kalman para monitorear y ajustarse a estos cambios. Se espera que en la práctica las variaciones en las series sean proporcionalmente menores y que, así mismo, sea mayor la convergencia del filtro. Adicionalmente se supuso que:

$$\sigma_w^2 = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.001 \end{pmatrix}, \quad \sigma_\epsilon^2 = 0.5$$

Los valores de $E(\beta_0|I_0)$ y $\text{VAR}(\beta_0|I_0)$ escogidos para la iniciación fueron:

$$E(\beta_0|I_0) = \begin{pmatrix} 1 \\ 0.05 \end{pmatrix}, \quad \text{VAR}(\beta_0|I_0) = \begin{pmatrix} 0.5 & 0 \\ 0 & 0.001 \end{pmatrix}$$

Al partir de estos valores la estimación de b_t y de c_t está dada por (B.15) para cada iteración. El ejercicio numérico descrito fue procesado utilizando el programa Eviews 5. Los resultados se incluyen en las últimas dos columnas del Cuadro B.1 y se muestran en la Figura 1 para el salario básico y en la Figura B.2 para la comisión sobre ventas, junto con los respectivos valores verdaderos de b_t y de c_t . Se observa que la convergencia es relativamente rápida; siendo dos o tres iteraciones suficientes para acercar la estimación al valor verdadero.

La estimación por mínimos cuadrados ordinarios de (B.18) arroja un b_t estimado de 4.71 y un c_t estimado de 0.091, en contraste con los resultados finales del filtro de Kalman, que son 4.36 y 0.193, respectivamente.

Figura B.1: Salario básico

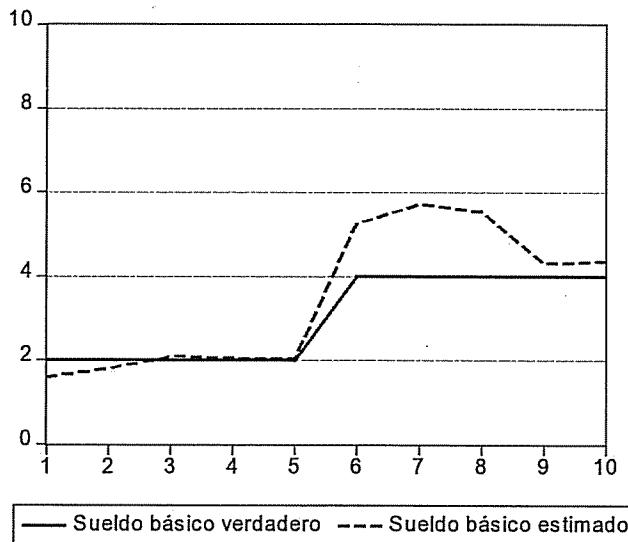
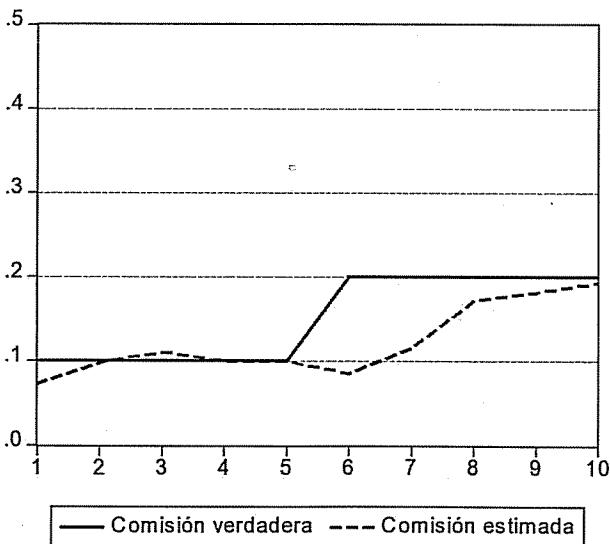


Figura B.2: Comisión sobre ventas



Ejemplo 2

La siguiente ilustración utiliza series mensuales de empleo en Colombia desde enero del 2000 hasta octubre del 2004; específicamente, las tasas de desempleo y de subempleo para 13 ciudades. Las ecuaciones del modelo son:

$$u_t = \beta_{1t}u_{t-3} + \beta_{2t}u_{t-12} + \phi sub_{t-12} + \varepsilon_t \quad (\text{ecuación de medición})$$

$$\beta_{1t} = \beta_{1t-1} + w_{1t}$$

$$\beta_{2t} = \beta_{2t-1} + w_{2t} \quad (\text{ecuación de transición})$$

Donde β_{1t} y β_{2t} son las variables de estado, u_t es la tasa de desempleo, sub_t es la tasa de subempleo, ϕ es un coeficiente y ε_t , w_{1t} y w_{2t} son ruido blanco. La descripción del modelo en lenguaje Eviews y el resultado de la estimación se muestran en el Cuadro B.2. Los coeficientes $C(1)$ y $C(2)$ en el Cuadro B.2 corresponden a la estimación de las varianzas. La forma adoptada por Eviews, por ejemplo $var = \exp(c(1))$, asegura varianzas positivas.

Cuadro B.2: Modelo en Eviews y resultado de la estimación

```
u = sv1*u(-3) + sv2*u(-12)+c(3)*sub(-12)+[var=0.1]
@state sv1 = sv1(-1)+[var=exp(c(1))]
@state sv2 = sv2(-1)+[var=exp(c(2))]
```

S-space: Kalman				
Method: Maximum likelihood (Marquardt)				
Date: 12/02/04 Time: 19:24				
Sample: 2000:01 2004:10				
Included observations: 58				
Valid observations: 46				
Convergence achieved after 14 iterations				
	Coefficient	Std. Error	z-Statistic	Prob.
C(1)	-7.032217	2.237703	-3.142605	0.0017
C(2)	-7.404593	3.120243	-2.373082	0.0176

Continúa

Continuación

C(3)	0.093652	0.058302	1.606322	0.1082
	Final State	Root MSE	z-Statistic	Prob.
SV1	-0.054542	0.101106	-0.539451	0.5896
SV2	0.771058	0.098310	7.843105	0.0000
Log likelihood	-75.17988	Akaike info criterion	3.399125	
Parameters	3	Schwarz criterion	3.518385	
Diffuse priors	2	Hannan-Quinn criter.	3.443801	

Referencias

- Akerlof, G.; Dickens, W.; Perry, G.; Bewley, T. y Blinder, A. (2000). "Near-Rational Wage and Price Setting and the Long-Run Phillips Curve", *Brookings Papers on Economic Activity*, 1, 1-60.
- Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*, New York, John Wiley.
- Chatfield, C. (1989). *The Analysis of Time Series: An Introduction*, New York, Chapman and Hall.
- Crafts, N.; Leybourne, S. y Mills, T. (1989). "The Climacteric in Late Victorian Britain and France: A Reappraisal of the Evidence", *Journal of Applied Econometrics*, 4 (2), 103-117.
- Engel, C. y Kim, C. (1999). "The Long-Run U.S./U.K. Real Exchange Rate", *Journal of Money, Credit and Banking*, 31 (3, Part 1), 335-356.
- Fama, E. y Gibbons, M. (1982). "Inflation, Real Returns, and Capital Investment", *Journal of Monetary Economics*, 9, 297-323.
- Garratt, A. y Hall, S. (1996). "Measuring Underlying Economic Activity", *Journal of Applied Econometrics*, 11 (2), 135-151.
- Fanger, C. W. y Newbold, P. (1986). *Forecasting Economic Time Series*, New Jersey, Academic Press.

- Greenslade, J.; Pierse, R. y Saleheen, J. (2003). *A Kalman Filter Approach to Estimating the UK Nairu*, Working Paper 179, London, The Bank of England.
- Hamilton, J. (1994). *Time Series Analysis*, Princeton, Princeton University Press.
- Harvey, A. C. (1982), "The Kalman Filter and its Applications in Econometrics and Time Series Analysis", *Methods of Operations Research*, 44, 3-18.
- (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge, Cambridge University Press.
- Jones, R. (1985). "Time Series Analysis with Unequally Spaced Data", en E. Hannan, P. Krishnaiah y M. Rao, *Handbook of Statistics* (vol. 5), New York, Elsevier.
- Judge, G. et al. (1985). *The Theory and Practice of Econometrics*, New York, John Wiley.
- Kalman, R. E. (1960). "A New Approach to Linear Filtering and Prediction Problems", Transactions of ASME, series D, *Journal of Basic Engineering*, 82, 35-45.
- y Bucy, R. (1961). "New Results in Linear Filtering and Prediction Theory", Transactions of ASME, series D, *Journal of Basic Engineering*, 83, 95-108.
- Kendall, M. y Stuart, A. (1963). *The Advanced Theory of Statistics* (vol. 1), 2a ed., New York, Hafner Publishing.
- LeRoy, S. y Waud, R. (1977). "Applications of the Kalman Filter in Short-Run Monetary Control", *International Economic Review*, 18 (1), 195-207.
- Lindgren, B. (1976). *Statistical Theory*, New York, Macmillan.
- Maybeck, P. (1971). *Stochastic Models, Estimation and Control* (vol. 1), New Jersey, Academic Press.
- Meinhold, R. y Singpurwalla, N. (1983). "Understanding the Kalman Filter", *The American Statistician*, 37 (2), 123-127.
- Montenegro, Á. (1991). "El Filtro Kalman: derivación y práctica", *Desarrollo y Sociedad*, 28, 79-91.

- 0
- riestley, M. B. (1981). *Spectral Analysis and Time Series*, New Jersey, Academic Press.
- chwarz, M. y Shaw, L. (1975). *Signal Processing: Discrete Spectral Analysis, Detection and Estimation*, New York, McGraw-Hill.
- mizaki, H. (1993). "Kalman Filter Model with Qualitative Dependent Variables", *The Review of Economics and Statistics*, 75 (4), 747-752.
- atson, P. K. (1983). "Kalman Filtering as an Alternative to Ordinary Least Squares – some Theoretical Considerations and Empirical Results", *Empirical Economics*, 8, 71-85.

APÉNDICE C

Redes neurales artificiales

Las redes neurales artificiales son modelos estadísticos no lineales utilizados principalmente para la clasificación y predicción de datos y variables. Intentan hacer mímica de algunos mecanismos de procesamiento de información que ocurren en el sistema nervioso de los organismos biológicos, bajo la convicción de que, siendo producto de la selección natural, dichos mecanismos deben ser efectivos y eficientes.

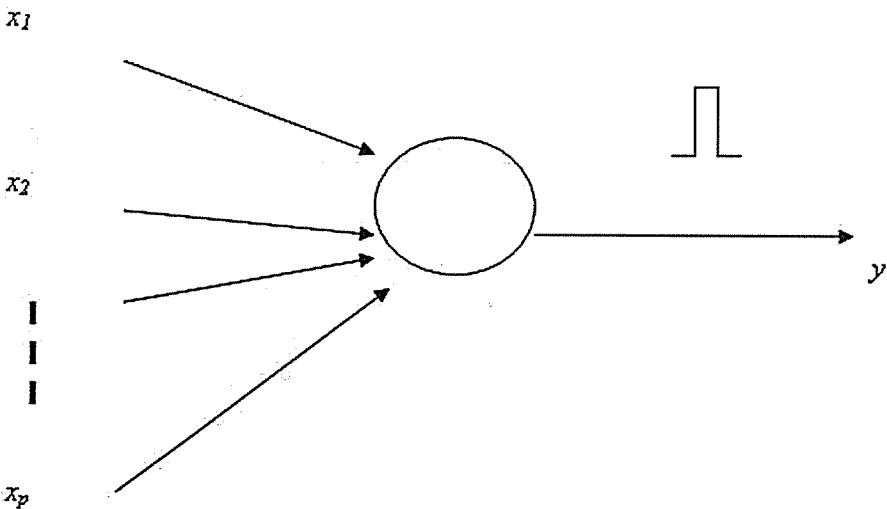
Todos los modelos, ya sean lineales o no lineales, tienen como objeto servir de aproximaciones útiles de la realidad, y nunca tienen la pretensión de sustituirla. De manera similar, las redes neurales artificiales son una simplificación útil de las redes neurales biológicas.

La unidad básica de procesamiento de las redes neurales biológicas es la neurona. En términos generales, una neurona es una célula con tres partes principales: el cuerpo central, que tiene el núcleo; unas extensiones o filamentos, llamados dendritas, que reciben las señales o estímulos que llegan a la célula desde los sensores o transductores, y una extensión de salida, llamada axón, que transporta el resultado del procesamiento de los datos hacia las dendritas de otras células o hacia actuadores (músculos, por ejemplo).

Si la suma ponderada de las señales o estímulos de entrada sobrepasa cierto nivel de umbral, la neurona dispara una señal de salida por el

ón de magnitud fija. Las señales enviadas hacia la siguiente neurona hacen contacto en los puntos llamados *sinapses*, entre el axón de una neurona y las dendritas de la siguiente; en estos puntos la señal se pondera, esto es, se puede dejar pasar, atenuar o hasta inhibir. En otras palabras, podemos imaginar una neurona como una unidad con múltiples entradas x_1, x_2, \dots, x_p , ponderadas, y una salida binaria y (activada o no activada), como se muestra en la Figura C.1.

Figura C.1: Modelo de una neurona



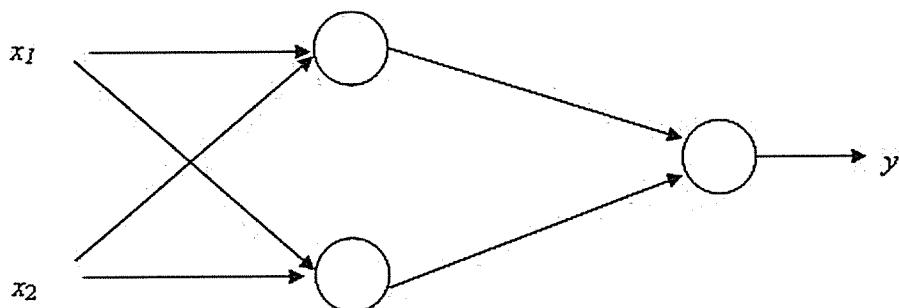
En la práctica, con el propósito de simplificar el manejo matemático de la función neuronal, aproximamos la salida binaria por una salida suavizada, por ejemplo, al intervalo $[0,1]$, como en un modelo logit $\frac{e^x}{1+e^x}$, o al intervalo $[-1,1]$, como en una función de tangente hiperbólica $\frac{e^x - e^{-x}}{e^x + e^{-x}}$.

Si bien en las redes biológicas cada neurona puede tener miles de conexiones con otras células, en las redes artificiales tales conexiones entre neuronas son pocas, por las limitaciones existentes en las técnicas de estimación y en el número de datos disponibles.

Una red neural artificial típica consiste de capas de neuronas que procesan y transportan la información de la entrada a la salida. Todas las neuronas de la primera capa reciben los datos de entrada, los ponderan por sus coeficientes de entrada, restan el umbral (la constante del *logit*) y pasan el resultado a cada una de las neuronas de la capa

siguiente. La salida final puede hacerse a través de una sola neurona o de manera lineal, ponderando y sumando las salidas de las neuronas de la última capa, más una constante si se considera necesario (véase Figura C.2). Las capas ubicadas entre la entrada y la salida se conocen como capas ocultas (*hidden layers*).

Figura C.2: Red con dos neuronas ocultas y dos variables de entrada



Aunque existe una gran variedad de alternativas de diseño de redes, algunas con múltiples salidas (por ejemplo, para análisis multivariado) y otras con realimentación de salida a entrada, la red típica suele tener una salida y una o dos capas escondidas con un número de neuronas cercano al número de variables de entrada.

El tema de la realimentación, tan importante en las redes biológicas, puede incorporarse en las redes artificiales en el hecho de reestimar el modelo con la llegada de nuevas observaciones. Esta reestimación, que obviamente se hace comparando la salida con la realidad observada, modifica los parámetros del modelo para ajustarse a las nuevas observaciones. De otra forma, la incorporación de ecuaciones de realimentación explícitas en el modelo introduciría complejidades adicionales en la estimación. Alternativamente, se puede introducir la realimentación usando rezagos de la variable de salida como variables de entrada, lo cual resulta muy apropiado cuando se trabaja con series de tiempo.

Aun cuando es cierto que una neurona biológica dispara un pulso de altura (amplitud) igual ante cualquier estímulo que sobrepase su umbral, también es cierto que si el estímulo es intenso, la neurona disparará con mayor frecuencia. En otras palabras, la transmisión neuronal se hace en frecuencia modulada (FM) que, como se sabe, ofrece una mayor nitidez y protección contra la interferencia y los ruidos del ambiente;

sto contrasta con la transmisión en amplitud modulada (AM), como memoria, por ejemplo, hacer que la altura o amplitud de los pulsos sea proporcional a la intensidad del estímulo, la cual es más sensible a la interferencia externa.

Como se mencionó, la dificultad de modelar con neuronas binarias ha llevado a sustituirlas por funciones, cuya salida es limitada pero no binaria, como son los logits. Esto, en efecto, hace que en las redes artificiales la salida de las neuronas sea proporcional dentro de un rango de nivel de entrada –un poco más AM que FM–. Nuestros modelos no son en tiempo real, razón por la cual tiene poco sentido hablar de FM; sin embargo, esta diferencia con las redes biológicas podría generalizarse en el sentido de que la amplitud de la salida de los logits se interprete como proporcional a la frecuencia.

specificación

Una red neural puede verse como una caja negra con varias entradas, por un lado, y una salida, por el otro. A diferencia del modelo clásico de regresión, el interés no radica en la magnitud de los coeficientes individuales o en que sus signos sean los esperados, sino que se trata de un sistema no lineal redundante donde los coeficientes individuales pueden tener poco sentido. Lo que importa es que funcione.

La redundancia se manifiesta en que las variables de entrada se conectan a todas o varias de las neuronas de la primera capa, las cuales su vez van pasando la información en paralelo a la siguiente capa. Sin embargo, aunque en el diseño tradicional todas las neuronas de la primera capa reciben todas las entradas, también podría diseñarse una red de manera que algunas neuronas de la primera capa tengan entradas diferentes.

Algunos autores sostienen que las neuronas se especializan en reaccionar a estímulos específicos y no responden a otros –por ejemplo, ciertas neuronas sólo responden al dulce, otras únicamente responden al sabor salado, etc.–; sin embargo, estudios recientes (Smith y Sargolskee, 2001) indican que una misma neurona puede responder a

varios sabores y que el mensaje final transmitido depende de los patrones de actividad a través de un gran conjunto de neuronas.

Una red sencilla con dos variables de entrada x_1 y x_2 , una capa escondida con dos neuronas y una neurona de salida, como la mostrada en la Figura C.2, tendría la siguiente ecuación:

$$y = \frac{e^{\beta_0 + \beta_1 \frac{e^{\alpha_{01} + \alpha_{11}x_1 + \alpha_{21}x_2}}{1 + e^{\alpha_{01} + \alpha_{11}x_1 + \alpha_{21}x_2}} + \beta_2 \frac{e^{\alpha_{02} + \alpha_{12}x_1 + \alpha_{22}x_2}}{1 + e^{\alpha_{02} + \alpha_{12}x_1 + \alpha_{22}x_2}}}}{1 + e^{\beta_0 + \beta_1 \frac{e^{\alpha_{01} + \alpha_{11}x_1 + \alpha_{21}x_2}}{1 + e^{\alpha_{01} + \alpha_{11}x_1 + \alpha_{21}x_2}} + \beta_2 \frac{e^{\alpha_{02} + \alpha_{12}x_1 + \alpha_{22}x_2}}{1 + e^{\alpha_{02} + \alpha_{12}x_1 + \alpha_{22}x_2}}}}$$

Donde α_{01} , α_{02} y β_0 son los niveles de umbral (la constante) de cada neurona. Como puede observarse, la salida puede escribirse como un *logit* de *logits*, lo cual da una expresión extremadamente no lineal, especialmente si el número de capas y neuronas aumenta.

Existen muchas lecturas que sirven como introducción al tema de las redes neurales artificiales; por ejemplo, las siguientes que se encuentran en internet, Frohlich (1999), Statsoft (1997), Galkin (2001) y Kolarik y Gottfried (1997). En temas económicos, podemos citar, entre otros, a Franses y Draisma (1997), Kuan y White (1994), Swanson y White (1997) y Tkacz y Hu (1999).

Estimación

Una vez especificado el modelo, la estimación consiste en escoger los valores de las ponderaciones (coeficientes) y el umbral (constante) para cada neurona (*logit*). Por la naturaleza no lineal de la ecuación, la estimación de las redes neurales debe hacerse por algún método de optimización numérico. Muchos de ellos son, en esencia, métodos de mínimos cuadrados no lineales.

Uno de los métodos más populares es el llamado de propagación hacia atrás (*back propagation o delta method*), el cual consiste en partir de coeficientes y umbrales iniciales (que pueden ser arbitrarios), en calcular el error entre la salida estimada y el valor verdadero y en usar este error para devolverse por la red ajustando los coeficientes y umbrales de las neuronas de las capas a través de un algoritmo algo *ad hoc*. Igual se hace para la segunda observación de la muestra, luego para la tercera, y así sucesivamente, hasta completarla. Luego se repite el proceso hasta lograr un juego de coeficientes que minimicen la

suma de errores al cuadrado. Un ejemplo numérico puede consultarse en Frohlich (1999). Existe *software* especializado en el mercado para implementar este algoritmo de propagación hacia atrás.

Sin embargo, si el tamaño de la red y la muestra lo permiten, la estimación puede realizarse por medio del *software* de mínimos cuadrados no lineales, disponible en la mayoría de los programas estadísticos. La estimación requiere paciencia, ya que no siempre converge fácilmente, pues depende crucialmente de los valores iniciales escogidos.

La estimación puede requerir un buen número de iteraciones (100 o 1000 son comunes), de manera que los criterios de parada del algoritmo deben ajustarse apropiadamente o, en su defecto, debe reiniciarse la iteración continuamente.

Evaluación

El objetivo de las redes neurales no es el modelaje estructural, ni siquiera la forma reducida, sino la predicción, por lo cual se tolera algún grado de redundancia (o multicolinealidad en el lenguaje de los modelos lineales). La capacidad predictiva es más importante en la evaluación que las pruebas t sobre coeficientes individuales, los cuales con frecuencia no resultan todos significativos. El R^2 , la desviación estándar del error, la verosimilitud estimada y criterios como Akaike y Schwarz sirven de guía, a la par que las pruebas de significancia individual.

Como siempre, se recomienda utilizar parte de la muestra para validación cruzada; esto es, usar por ejemplo el 70 % de la muestra para estimación y el resto para probar y evaluar la capacidad predictiva. Deben calcularse medidas de predicción, como la raíz del error cuadrado medio o el error cuadrado medio absoluto. La estabilidad del modelo puede estimarse omitiendo sucesivamente algunos períodos finales de la muestra o agregando observaciones que no hayan sido utilizadas antes en la estimación.

El hecho de que la estimación converja no implica que sea imposible mejorarla. Con frecuencia, arrancar de otros valores iniciales, al menos en algunos coeficientes, puede mejorar la estimación. Tampoco el hecho de que la estimación dé un buen ajuste en la muestra, necesariamente, garantiza un buen desempeño predictivo.

Predicción

Las conocidas dificultades para la predicción con modelos no lineales aplican de igual manera al caso de las redes neurales artificiales. En los modelos lineales con ruido aditivo de media 0, por ejemplo, un autorregresivo sencillo:

$$x_t = \alpha x_{t-1} + \varepsilon_t$$

La predicción realizada en el tiempo n para T periodos en el futuro será la expectativa de la variable para $n+T$, condicionada al conjunto de información en el tiempo n . Por ejemplo, para $T=2$ la predicción será:

$$Ex_{n+2} = E[\alpha x_{n+1} + \varepsilon_{n+2}] = E[\alpha(\alpha x_n + \varepsilon_{n+1}) + \varepsilon_{n+2}] = \alpha^2 x_n$$

Donde x_n es la última observación de la muestra.

Por la naturaleza lineal de la expresión, es posible tomar la expectativa de los errores, la cual es cero y su efecto desaparece de la predicción.

En contraste, en un modelo no lineal el error de los períodos anteriores entra de diferentes formas no lineales en la expresión del modelo de manera que no siempre será posible tomar su valor esperado y eliminar su efecto de la predicción. Para ilustrar, tomemos un *logit* sencillo:

$$x_t = \frac{1}{1 + e^{-\alpha x_{t-1}}} + \varepsilon_t$$

Donde la predicción realizada para $T=2$ será el valor esperado condicional:

$$\begin{aligned} Ex_{n+2} &= E\left[\frac{1}{1 + e^{-\alpha x_{n+1}}} + \varepsilon_{n+2}\right] = E\left[\frac{1}{1 + e^{-\alpha\left[\frac{1}{1 + e^{-\alpha x_n}} + \varepsilon_{n+1}\right]}} + \varepsilon_{n+2}\right] \\ &= E\left[\frac{1}{1 + e^{-\alpha\left[\frac{1}{1 + e^{-\alpha x_n}} + \varepsilon_{n+1}\right]}}\right] \end{aligned}$$

Sin embargo, nótese que no da igual a:

$$\frac{1}{1 + e^{-\alpha\left[\frac{1}{1 + e^{-\alpha x_n}}\right]}}$$

De manera que pasar por alto el efecto del error en los modelos no lineales, como se acostumbra en los lineales, tiende a sesgar la predicción. Obsérvese que el sesgo se presenta para $T \geq 2$, ya que para $T = 1$ no alcanzan a acumularse los errores desconocidos.

Algunos autores sugieren abstenerse de realizar predicciones para períodos lejanos. Otros sugieren (por ejemplo, Lundbergh y Terasvirta, 2000) realizar varias predicciones, al estilo de un experimento de Monte Carlo, adicionando en cada predicción una o varias observaciones de suido, según el horizonte de predicción, tomadas de un generador de números aleatorios o de la "población" formada por los residuos estimados (método *bootstrap*), y luego promediar las predicciones o, mejor, dibujar el histograma para cada horizonte.

Ejemplos

Ilustraremos las redes neurales artificiales con estimaciones realizadas en Eviews, por mínimos cuadrados no lineales, utilizando el método de Marquardt, que es una modificación del método de Gauss-Newton. Nótese que en Eviews la función que se escribe $@logit(x)$ quiere decir el *logit* de lo que esté en paréntesis, de manera que este comando representa una neurona. Por ejemplo: $@logit(\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2)$, representa la expresión:

$$\frac{e^{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2}}{1 + e^{\alpha_0 + \alpha_1 x_1 + \alpha_2 x_2}}$$

Esta equivale a una neurona con dos entradas, x_1 y x_2 , y un umbral igual a α_0 . Utilizando este comando se pueden interconectar varias neuronas y escribir la ecuación para la salida de una red, como se muestra a continuación. Los ejemplos presentados comparan muy favorablemente con los mejores modelos lineales para las mismas variables. Otros ejercicios empíricos en el área económica se encuentran en la literatura citada.

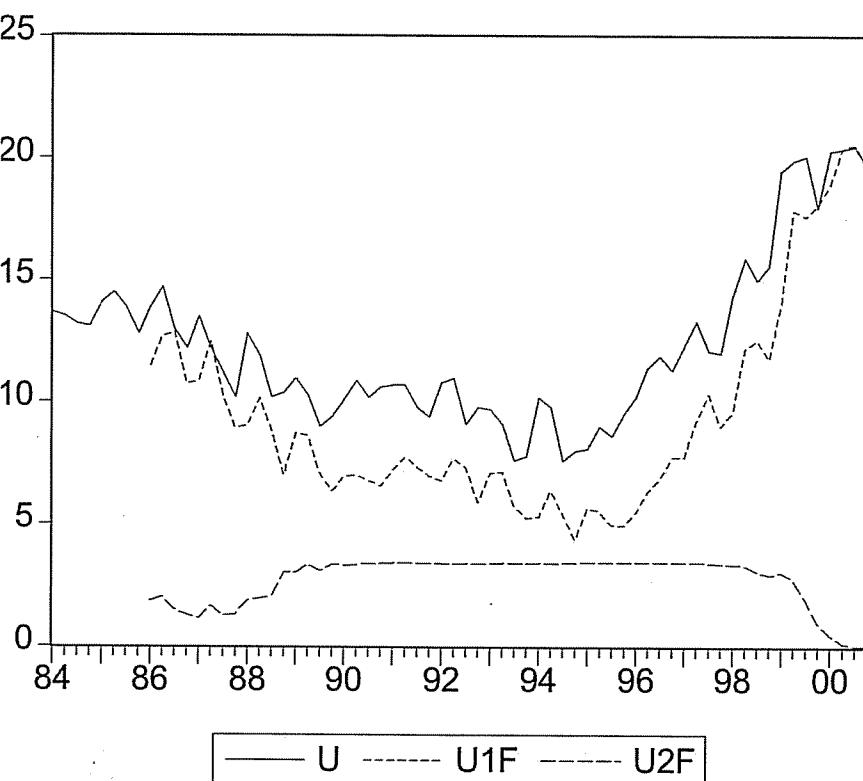
Recordamos que no existe una teoría establecida ni resultados estadísticos de aplicación general para redes neurales artificiales, como los hay para el caso de los modelos lineales, en especial para el caso de la regresión; de manera que es un campo en desarrollo que admite cierta exploración y experimentación.

El Cuadro C.1 muestra un modelo para la tasa trimestral de desempleo U en siete ciudades, utilizando rezagos uno, cuatro, seis y ocho como variables de entrada. La salida del modelo es la suma ponderada de la salida de dos neuronas ocultas. La Figura C.3, que acompaña al Cuadro C.1, muestra los componentes del modelo: la serie original U, la salida de la primera neurona del modelo U1F y la salida de la segunda U2F. El objetivo es poder observar la manera cómo las neuronas se “reparten” el trabajo. Puede verse que la segunda neurona permanece la mayor parte del tiempo cerca de saturación, pero varía de manera importante en los extremos del periodo muestral.

Cuadro C.1: Modelo para la tasa trimestral de desempleo U

Dependent Variable:	U			
Method:	Least Squares			
Date:	01/26/01	Time:	18:03	
Sample(adjusted):	1986:1	2000:4		
Included observations:	60	after adjusting endpoints		
Convergence achieved after	68	iterations		
U=C(1)*@LOGIT(C(2)*U(-1)+C(3)*U(-4)+C(7))				
+C(4)*@LOGIT(C(5)*U(-6) +C(6)*U(-8)+C(8))				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	25.27105	2.708139	9.331517	0.0000
C(2)	0.128401	0.023311	5.508126	0.0000
C(3)	0.113710	0.034776	3.269822	0.0019
C(7)	-3.429905	0.524035	-6.545186	0.0000
C(4)	3.417180	1.304881	2.618766	0.0115
C(5)	-0.839668	0.485130	-1.730810	0.0894
C(6)	-0.698513	0.576454	-1.211741	0.2311
C(8)	20.79585	12.53401	1.659154	0.1031
R-squared	0.922073	Mean dependent var		12.13000
Adjusted R-squared	0.911583	S. D. dependent var		3.589358
S.E. of regression	1.067294	Akaike info criterion		3.091695
Sum squared resid	59.23401	Schwarz criterion		3.370941
Log likelihood	-84.75085	Durbin-Watson stat		1.738770

Figura C.3: Serie original U y contribución de cada neurona

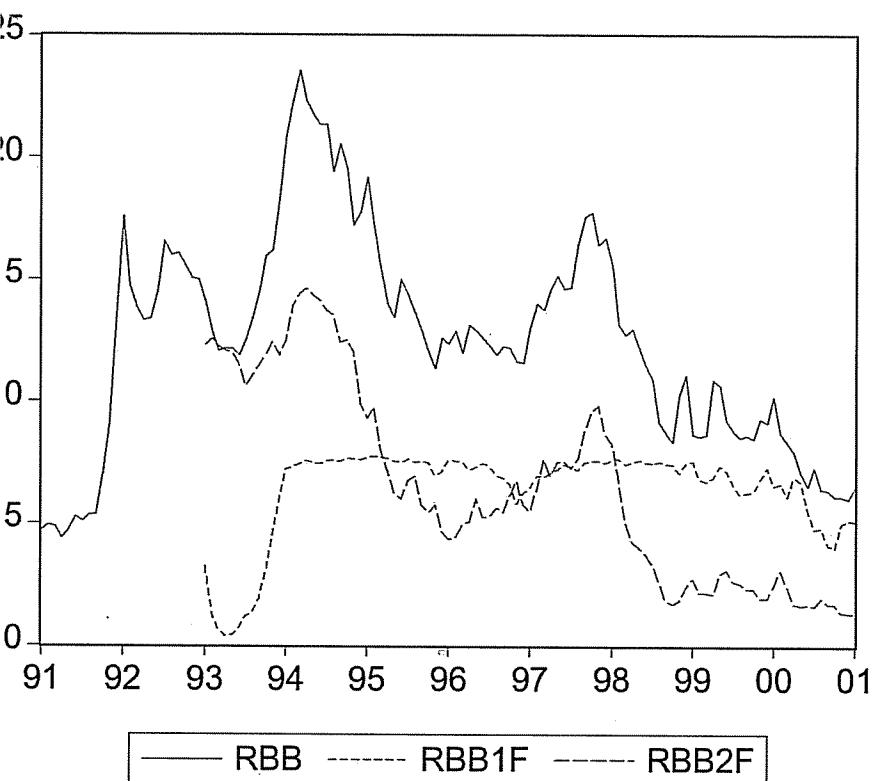


El Cuadro C.2 muestra un modelo para el índice real mensual de precios de la Bolsa de Bogotá (RBB), en función de sus propios rezagos, no, doce y veinticuatro, y del primer rezago del índice de precios al por mayor (IPP). La Figura C.4 muestra la serie original RBB y la contribución de cada neurona, RBB1F y RBB2F, a la salida.

Cuadro C.2: Modelo para el índice real mensual de precios
de la Bolsa de Bogotá

Dependent Variable: RBB				
Method: Least Squares				
Date: 02/17/01 Time: 18:40				
Sample(adjusted): 1993:01 2001:01				
Included observations: 97 after adjusting endpoints				
Convergence achieved after 48 iterations				
$RBB = C(1) * @LOGIT(C(2) * RBB(-1) + C(3) * RBB(-12)$				
$+ C(4) * RBB(-24) + C(9)$				
$* IPP(-1) + C(10)) + C(5) * @LOGIT(C(6) * RBB(-1)$				
$+ C(7) * RBB(-12) + C(8)$				
$* RBB(-24) + C(11) * IPP(-1) + C(12))$				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	7.761804	1.760581	4.408660	0.0000
C(2)	0.438009	0.182039	2.406129	0.0183
C(3)	0.358055	0.197154	1.816120	0.0729
C(4)	0.176492	0.065189	2.707403	0.0082
C(9)	0.086315	0.035895	2.404667	0.0184
C(10)	-17.46103	6.853172	-2.547875	0.0126
C(5)	15.84571	2.696772	5.875806	0.0000
C(6)	0.166143	0.025214	6.589398	0.0000
C(7)	-0.108060	0.042376	-2.549997	0.0126
C(8)	-0.053535	0.023012	-2.326356	0.0224
C(11)	-0.034711	0.014146	-2.453689	0.0162
C(12)	2.289703	1.767312	1.295585	0.1986
R-squared	0.957726	Mean dependent var	12.97459	
Adjusted R-squared	0.952255	S.D. dependent var	4.208079	
S.E. of regression	0.919487	Akaike info criterion	2.785362	
Sum squared resid	71.86382	Schwarz criterion	3.103883	
Log likelihood	-123.0900	Durbin-Watson stat	1.650784	

Figura C.4: Serie original RBB y contribución de cada neurona



El Cuadro C.3 muestra un modelo para la inflación porcentual mensual P utilizando una capa oculta con cuatro neuronas y rezagos uno, dos, doce y veinticuatro de P y el rezago dos de PFOOD, la inflación mensual de alimentos, como variables de entradas. El que, según se lee en la salida, la convergencia se haya alcanzado en una iteración sólo se refiere al número de iteraciones desde la última arrancada, y esconde el hecho de que antes de la última arrancada del programa de optimización ya se había realizado un gran número de iteraciones.

Cuadro C.3: Modelo para la inflación porcentual mensual

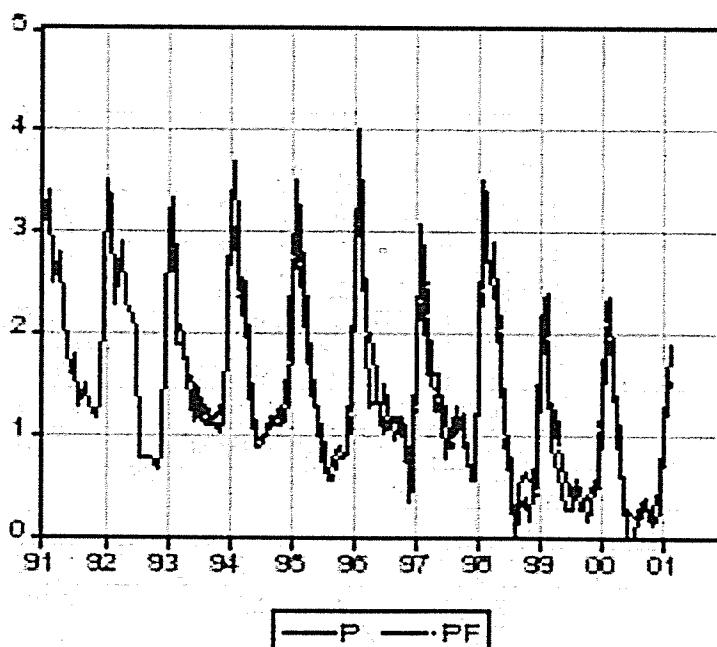
Dependent Variable: P				
Method: Least Squares				
Date: 03/11/01 Time: 11:03				
Sample(adjusted): 1993:01 2001:02				
Included observations: 98 after adjusting endpoints				
Convergence achieved after 1 iterations				
$P = C(1) * @LOGIT(C(2)*P(-1) + C(4)*PFOOD(-2) + C(5)*P(-24) + C(6)) + C(8)$				
$* @LOGIT(C(9)*P(-1) + C(11)*PFOOD(-2) + C(12)*P(-24) + C(13))$				
$+ C(15)* @LOGIT(C(16)*P(-1) + C(17)*P(-12) + C(18)*P(-24) + C(19))$				
$+ C(20)* @LOGIT(C(21)*P(-1) + C(22)*P(-12) + C(23)*P(-24) + C(24))$				
$+ C(25)$				
	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.998223	0.201902	4.944108	0.0000
C(2)	219.0753	21.52328	10.17853	0.0000
C(4)	75.81948	19.60520	3.867315	0.0002
C(5)	173.9004	15.27686	11.38325	0.0000
C(6)	-1123.770	37.37223	-30.06965	0.0000
C(8)	-4.758524	3.166842	-1.502609	0.1370
C(9)	0.331767	0.261245	1.269944	0.2079
C(11)	0.197856	0.147953	1.337291	0.1851
C(12)	-0.771356	0.506079	-1.524181	0.1316
C(13)	1.690513	0.691306	2.445391	0.0168
C(15)	-1.594159	0.452032	-3.526647	0.0007
C(16)	-2.414954	1.070296	-2.256341	0.0269
C(17)	-2.123755	0.966296	-2.197831	0.0310

Continúa

Continuación

C(18)	0.828407	0.540183	1.533567	0.1292
C(19)	3.078906	1.429685	2.153556	0.0344
C(20)	-0.554763	0.211204	-2.626674	0.0104
C(21)	-11.66533	11.97170	-0.974409	0.3329
C(22)	10.27910	10.95351	0.938430	0.3510
C(23)	12.04022	11.86746	1.014558	0.3135
C(24)	-12.61991	12.47649	-1.011495	0.3149
C(25)	5.493597	2.551100	2.153423	0.0344
R-squared	0.900042	Mean dependent var	1.338776	
Adjusted R-squared	0.874079	S.D. dependent var	0.893578	
S.E. of regression	0.317089	Akaike info criterion	0.728139	
Sum squared resid	7.741984	Schwarz criterion	1.282061	
Log likelihood	-14.67881	Durbin-Watson stat	2.283271	

Figura C.5: Serie original P y predicción PF



A diferencia de las figuras anteriores, la Figura 5 muestra la predicción del modelo para la inflación mensual porcentual PF contra la serie original P . Mientras que la muestra va desde enero de 1991 hasta febrero de 2001, la predicción (por los rezagos) se inicia en el primer mes de 1993. La raíz del error cuadrado medio de la predicción es 0.28, el error medio absoluto es 0.22 y el coeficiente de Theil es 0.088.

Referencias

- Franses, P. H. y Driisma, G. (1997). "Recognizing Changing Seasonal Patterns using Artificial Neural Networks", *Journal of Econometrics*, 81, 273-280.
- Frohlich, J. (1999). *Neural Networks with Java*, recuperado el 11 de febrero de 2010, de <http://rfhs8012.fh-regensburg.de/saj39122/jfroehl/diplom/e-13-text.html>
- Galkin, I. (2001). *Crash Introduction to Artificial Neural Networks*, recuperado el 11 de febrero de 2010, de <http://ulcar.uml.edu/iag/CS/Intro-to-ANN.html>
- Kolarik, T. y Gottfried, R. (1997). *Time Series Forecasting Using Neural Networks*, recuperado el 11 de febrero de 2010, de <http://portal.acm.org/citation.cfm?id=190290&dl=GUIDE&coll=GUIDE&CFID=75856190&CFTOKEN=49128686>
- Kuan, C. M. y White, H. (1994). "Artificial Neural Networks: An Economic Perspective", *Econometric Reviews*, 13, 1-19.
- Lundbergh, S. y Terasvirta, T. (2000). "Forecasting with Smooth Transition Autoregressive Models", SSE Working Paper Series in Economics and Finance, 390, Stockholm, School of Economics.
- Smith, D. y Margolskee, R. (2001). "Making Sense of Taste", *Scientific American*, marzo, 32-39.
- Statsoft (1997). *Neural Networks* [Electronic Textbook], recuperado el 11 de febrero de 2010, de <http://www.cs.kuleuven.ac.be/~kurtd/STATS/stneunet.html>
- Swanson, N. y White, H. (1997). "A Model Selection Approach to Real Time Macroeconomic Forecasting Using Linear Models and

Artificial Neural Networks”, *Review of Economics and Statistics*, 79, 540-50.

Kacz, G. y Hu, S. (1999). “Forecasting GDP Growth Using Artificial Neural Networks”, Working Paper 99-3, Ottawa, Department of Monetary and Financial Analysis, Bank of Canada.