



UNIVERSIDAD NACIONAL AUTONOMA
DE MEXICO

FACULTAD DE CIENCIAS

INTRODUCCION A LAS SERIES DE
TIEMPO COINTEGRADAS

T E S I S
QUE PARA OBTENER EL TITULO DE:
A C T U A R I O
P R E S E N T A :
ROMAN HERNANDEZ ALFARO



FACULTAD DE CIENCIAS
UNAM

DIRECTORA DE TESIS: M. EN A. P. MARIA DEL PILAR ALONSO REYES

2006



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:
"Introducción a las Series de Tiempo Cointegradas"

realizado por Hernández Alfaro Roman

con número de cuenta 09820518-5 , quien cubrió los créditos de la carrera de: Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario

M. en A. P. María del Pilar Alonso Reyes

Propietario

M. en C. José Antonio Flores Díaz

Propietario

Act. Jaime Vázquez Alamilla

Suplente

M. en C. Hugo Villaseñor Hernández

Suplente

Dra. Guadalupe Carrasco Licea

Consejo Departamental de Matemáticas

III

Act. Jaime Vázquez Alamilla



CONSEJO DEPARTAMENTAL
DE
MATEMÁTICAS

Un gran hombre me dijo un día:

“Los amigos son ángeles que nos llevan en sus brazos cuando nuestras alas tienen problemas para recordar como volar”

Dedico este trabajo a todos mis ángeles, pero especialmente a dos: mis padres.

Índice general

Introducción	1
1. Series de Tiempo Univariadas	3
1.1. Series de Tiempo	3
1.1.1. Definición	3
1.1.2. Operadores de Rezago y Diferencia	3
1.2. Procesos Estocásticos	5
1.2.1. Procesos Estocásticos Estacionarios	6
1.2.2. Ergodicidad	9
1.2.3. Proceso Estocástico Estacionario Puramente Aleatorio	11
1.2.4. Proceso Estocástico No Estacionario.	12
1.3. Análisis Univariante de Series de Tiempo.	14
1.3.1. Modelos Estacionarios Lineales.	15
1.3.2. Transformaciones de Series No Estacionarias.	24
2. Cointegración	35
2.1. Causalidad en el sentido de Granger.	35
2.2. Modelos de Vectores Autorregresivos (VAR).	40
2.2.1. Definición.	42
2.2.2. Estacionariedad.	43
2.2.3. Estimación y Especificación de un Proceso VAR.	44
2.2.4. Predicción con un Proceso VAR.	47
2.2.5. Extensión de los Vectores Autorregresivos.	49
2.3. Cointegración.	50
2.3.1. Regresiones Espurias.	51
2.3.2. Mecanismo de Corrección de Error (MCE)	52
2.3.3. Cointegración	57
2.3.4. Teorema de Representación de Granger	62
2.3.5. Estimación y Contraste de Relaciones de Cointegración.	63
2.3.6. Pronóstico en Sistemas Cointegrados.	80

3. Aplicaciones	83
3.1. Mercados Eficientes y Cointegración.	83
3.2. La Eficiencia en la Formación de Precios de un Mercado.	84
3.2.1. Los Datos	85
3.2.2. Eficiencia a Largo Plazo en la Formación del Precio.	91
3.2.3. Mecanismo de Ajuste al Precio de Equilibrio.	93
3.2.4. Interpretación de los Resultados Empíricos.	95
3.3. La Eficiencia del Mercado Cambiario entre el Euro, el Peso Mexicano y el Dólar.	96
3.3.1. Los Datos.	97
3.3.2. Interpretación de los Resultados Empíricos.	100
 Conclusiones	 103
 Apéndice	 105
 Bibliografía	 115

Introducción

La Teoría de la Cointegración ha despertado en los últimos años un fuerte interés en el ámbito econométrico, tanto en su parte teórica como práctica. El hecho de que esta estrategia de modelación se centre en las relaciones de equilibrio a largo plazo entre variables, la ha situado como un instrumento fundamental de cara al contraste de las relaciones de equilibrio postuladas por la Teoría Económica, a la vez que permite la estimación de modelos econométricos que incorporan la información que proporcionan estas relaciones. Estos modelos econométricos pueden utilizarse tanto para el análisis estructural y, por tanto, de las relaciones entre variables económicas, como para la predicción de magnitudes de interés o la simulación de políticas alternativas. Esta metodología también es usada en el análisis de la convergencia entre diferentes economías, en la construcción de modelos econométricos regionales, en la explicación del ciclo económico, etc.

Desde las aportaciones de Fuller en la segunda mitad de los años setenta -en los que se establecía las bases para el contraste del orden de integrabilidad-, de la generalización del uso de los modelos de Mecanismo de Corrección de Error, también en la misma época, y de la introducción del concepto de cointegración por Granger en 1981, hasta nuestros días, no ha cesado la aparición de literatura referida al tema, con aportaciones tanto teóricas como aplicadas.

Tanto es así, que hoy difícilmente podrían ser analizados y estudiados modelos macro-económicos, utilizando información temporal, sin tener un conocimiento mínimo de los conceptos de integrabilidad, cointegración y mecanismos de corrección de error. Lo mismo se puede afirmar en otros campos como el de la econometría financiera, regional, etc.

La Teoría de la Cointegración ha supuesto un cambio en la estrategia de modelación dinámica de variables económicas, poniendo de manifiesto las posibles limitaciones de la modelación de series de tiempo basadas en los modelos de función de transferencia, y proporcionando una explicación a los problemas de relaciones espurias entre series de tiempo ya puestos de manifiesto por Yule en el primer cuarto del siglo pasado.

El objetivo de esta tesis es el de realizar una introducción a esta teoría, presentando los conceptos básicos y los principales desarrollos que se han planteado así como una aplicación.

En el capítulo 1 se muestran los resultados sobre la teoría de series de tiempo de una forma abreviada con el fin de dar al lector una introducción apropiada al tema. En el capítulo 2 se desarrolla la teoría de cointegración mostrando los resultados más importantes para el análisis de las series. En el último capítulo se desarrollan dos ejemplos el primero de ellos sobre las relaciones de algunas empresas en el mercado financiera mexicano y el otro enfocado a la interacción de tres monedas: peso mexicano, dólar estadounidense y euro.

Por último se incluye un apéndice que resume el resultado principal de la teoría de cointegración desarrollada en el capítulo 2 y se muestran las conclusiones obtenidas de usar esta teoría en dos ejemplos diferentes pero significativos en la economía.

Capítulo 1

Series de Tiempo Univariadas

En este capítulo se muestran y desarrollan los principales resultados sobre la teoría de *series de tiempo univariadas* y algunos conceptos básicos de procesos estocásticos¹, elementales para los desarrollos. La finalidad del presente capítulo es cimentar las bases que sustentan la teoría de la *Cointegración*.

1.1. Series de Tiempo.

1.1.1. Definición.

Una *serie de tiempo* es una secuencia de datos numéricos igualmente espaciados, cada uno de los cuales se asocia con un instante específico del tiempo, cuya contra partida empírica son las series históricas. Existen numerosos ejemplos: la tasa mensual de desempleo, la medición semanal de la oferta monetaria, los índices diarios de precios de la bolsa de valores, etc. Y desde un punto de vista teórico, una serie de tiempo es una *colección* de variables aleatorias $\{x_t\}$. Una colección de este tipo, ordenada con respecto al tiempo, se conoce como proceso estocástico. Formalmente las series de tiempo son realizaciones de procesos estocásticos $\{x_t \mid t \in T\}$ de variables aleatorias cuyo conjunto de índices T se toma en el tiempo.

1.1.2. Operadores de Rezago y Diferencia.

El operador de rezago o retardo se denota por la letra B^2 , es útil para simplificar la obtención de los parámetros de un rezago. Si x_t es una función del tiempo, el operador de rezago, B , se define como

$$Bx_t = x_{t-1}, \text{ para toda } t.$$

¹La palabra estocástico es de origen griego y significa “relativo al azar”

²Del inglés Backward que significa hacia atrás.

Las potencias del operador se obtienen por aplicaciones sucesivas del operador B.

$$\begin{aligned} B^2 x_t &= B(Bx_t) = B(x_{t-1}) = x_{t-2} \\ B^3 x_t &= B(B^2 x_t) = B(x_{t-2}) = x_{t-3} \\ &\dots \end{aligned}$$

Así que, en forma general

$$B^k x_t = x_{t-k}, \text{ para toda } t \quad k = 0, 1, 2, \dots$$

Cabe señalar que al aplicar el operador B^k una serie que tenía originalmente N observaciones, se reducirá a una con tan sólo $N-k$ observaciones. Ya que al aplicar B^k se está rezagando la serie, $B^0 = I$ (operador identidad) por lo que se tiene que $B^0 x_t = x_t$.

En el análisis de series de tiempo el operador rezago se utiliza en forma de polinomios, esto es

$$x_t + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_k x_{t-k} = x_t + \sum_{i=1}^k \phi_i x_{t-i},$$

es un polinomio de rezago que suele expresarse como $\Phi(B)x_t$, en donde

$$\Phi(B) = 1 + \phi_1 B + \phi_2 B^2 + \dots + \phi_k B^k = 1 + \sum_{i=1}^k \phi_i B^i,$$

o bien

$$\Phi(B) = \sum_{i=0}^k \phi_i B^i \quad \text{si } \phi_0 = 1.$$

También es frecuente trabajar con polinomios de rezago racionales, los que suelen expresarse como cociente, de tal modo que $\Phi(B)$ será un polinomio racional si

$$\Phi(B) = \frac{\Theta(B)}{\Omega(B)} \quad \text{donde,} \quad \begin{aligned} \Theta(B) &= \sum_{i=0}^k \theta_i B^i \quad \text{si } \theta_0 = 1 \\ \Omega(B) &= \sum_{i=0}^k \omega_i B^i \quad \text{si } \omega_0 = 1. \end{aligned}$$

Representar a $\Phi(B)$ por medio de un polinomio racional es equivalente a restringir el número de coeficientes, ϕ_i , del polinomio $\Phi(B)$.

Otro operador de uso frecuente es el de diferencia ∇ que se define por medio de $\nabla x_t = x_t - x_{t-1}$, para toda t .

La relación de ∇ con B , el operador de rezago, es evidente

$$\nabla x_t = x_t - x_{t-1} = x_t - Bx_t = (1 - B)x_t$$

entonces $\nabla x_t = (1 - B)x_t$ o sea que $\nabla = 1 - B$.

De tal modo que se puede obtener una expresión general para ∇^k , aplicando sucesivamente el operador diferencia, pero por la equivalencia $\nabla = 1 - B$ utilizando el Teorema del Binomio de Newton. Entonces

$$\nabla^k x_t = (1 - B)^k x_t.$$

Por lo cual

$$\nabla^k x_t = \sum_{j=0}^k \frac{k!}{j!(k-j)!} (-1)^j x_{t-j}, \quad \text{para toda } t \quad k = 0, 1, 2, \dots$$

El operador ∇ se utiliza para eliminar la tendencia en series y así lograr la estacionariedad, lo que da lugar al concepto de integración.

Una **serie es integrada** cuando fue necesario aplicarle el operador diferencia para convertirla en estacionaria, el grado de integración está relacionado con el número de diferencias que se tomaron para lograr la estacionariedad. De tal forma que si se tiene una serie x_t no estacionaria y al aplicarle la d -ésima diferencia se obtiene la serie $\nabla^d x_t$, y si ésta es estacionaria se dirá que la serie x_t es integrada de orden d y se denota $x_t \sim I(d)$.

1.2. Procesos Estocásticos.

El concepto de proceso estocástico, atribuye dimensión temporal a los elementos del espacio muestral de los comportamientos de una variable aleatoria. Un proceso estocástico $\{x_t\}$, con $\{t \in T\}$, donde T es un conjunto de índices de números reales, se define como un conjunto de variables aleatorias, x_t , ordenadas de acuerdo con el parámetro t , que en el contexto se refiere al tiempo y se denota $\{x_t \mid t \in T\}$.

Como ejemplo de proceso estocástico se tiene el término de perturbación aleatoria en una regresión con series de tiempo. En este caso se trata de un proceso inobservable.

Por la forma en que son registradas las realizaciones u observaciones se puede hablar de **procesos estocásticos continuos y discretos**, de tal modo que un **proceso estocástico continuo** es un conjunto de variables aleatorias, x_t , ordenadas de acuerdo con el parámetro t con $t \in T$, donde T es un intervalo de números reales, sin importar si éste es cerrado o abierto. De un modo análogo un **proceso estocástico discreto** es un conjunto de variables aleatorias, x_t , ordenadas de acuerdo con el parámetro t con $t \in T$, donde T es un conjunto finito o infinito numerable.

Las leyes probabilísticas que rigen cualquier proceso estocástico se describen exhaustivamente y sin ambigüedades mediante su función de distribución de probabilidad (f.d.p.)

conjunta de los vectores de variables aleatorias que se pueden formar con las variables que constituyen el proceso. Sin embargo, siendo T infinito en teoría, la distribución debería poseer dimensión infinita Kolmogorov demostró que esto no es necesario bajo ciertas condiciones de regularidad.

Para muchos fines prácticos, los procesos suelen describirse mediante sus momentos, entre los que destacan los siguientes:

La media de un proceso estocástico se define por

$$\mu_t = E(x_t).$$

La función de autocovarianza se define como

$$\gamma_{t,t+k} = cov(x_t, x_{t+k}) = E\{[x_t - E(x_t)][x_{t+k} - E(x_{t+k})]\}$$

$$\text{entonces } \gamma_{t,t+k} = E[(x_t - \mu_t)(x_{t+k} - \mu_{t+k})] \quad k = 0, \pm 1, \pm 2, \dots$$

A partir de este resultado se obtiene la función de varianza del proceso cuando $k=0$

$$\gamma_{t,t} = var(x_t).$$

Las características de un proceso estocástico, media, varianza y autocovarianza consideran los valores de una misma serie en momentos distintos y estas generalmente son funciones que dependen del tiempo.

Por otra parte, la función de autocorrelación se define mediante la expresión

$$\rho_{t,t+k} = \frac{\gamma_{t,t+k}}{\sqrt{\gamma_{t,t}}\sqrt{\gamma_{t+k,t+k}}}.$$

1.2.1. Procesos Estocásticos Estacionarios.

A los procesos estocásticos se les puede dividir en dos clases: **estacionarios y evolutivos**. De estacionariedad puede hablarse en dos sentidos, uno es amplio y el otro estricto.

En el **sentido estricto** un proceso estocástico es **estacionario** si la distribución conjunta de cualquier conjunto de n variables aleatorias $\{x_{t_1}, x_{t_2}, \dots, x_{t_n}\}$ es la misma que la de $\{x_{t_1+k}, x_{t_2+k}, \dots, x_{t_n+k}\}$ para cualquier n dado, independiente de k . Un proceso que no es estacionario es evolutivo.

La definición de estacionariedad en sentido estricto implica que todos los momentos de la distribución conjunta son independientes de t , es decir, las características del proceso estocástico no sufren alteración al considerar tiempos históricos diferentes. Ésta es, una

restricción demasiado fuerte para imponer a priori a los procesos estocásticos destinados a representar datos reales, ya que éstos violan, sin duda, la hipótesis de estacionariedad estricta.

Un **proceso estacionario en sentido amplio** (o estacionario de segundo orden, o covarianza-estacionario, o débilmente estacionario) es cuando se verifica:

- $E(x_t) = \mu < \infty$ para toda t .
- $cov(x_t, x_{t+k}) = \gamma_k$ sólo depende de $|k|$.

Es decir la media del proceso es constante y la covarianza es sólo función del lapso temporal considerado y no del tiempo histórico. No hay supuestos sobre los momentos de orden superior éstos pueden variar con el tiempo.

Observación.

Si $\{x_t\}$ es estacionario en sentido amplio entonces la $\gamma_0 = var(x_t) = \sigma^2$ para toda t y esto es por que

$$var(x_t) = cov(x_t, x_{t+0}) = cov(x_s, x_{s+0}) = var(x_s)$$

entonces $var(x_t) = var(x_s)$ para toda $t \neq s$.

Por consiguiente la estacionariedad en sentido estricto implica la estacionariedad en sentido amplio, la implicación inversa no siempre se satisface, pero si se cumple cuando la distribución conjunta de todas las variables que integran el proceso es una normal multivariada, ya que la normal se caracteriza por sus primeros dos momentos cosa que no es así para otras distribuciones.

Así pues, debido a que en la práctica es común suponer que la distribución asociada a las series de tiempo es la normal en adelante se hará referencia a procesos estacionarios si cumplen los supuestos de estacionariedad en sentido amplio.

La función de autocorrelación de un proceso estocástico estacionario se define como sigue:

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

Las principales propiedades de la función de autocorrelación de un proceso estocástico estacionario son:

- $\rho_0 = 1$
- $\rho_k = \rho_{-k}$ (La autocorrelación es una función par)
- $|\rho_k| \leq 1$ (No unicidad)

- Es definida positiva³.

Demostración

Sea $\{x_t\}$ un proceso estacionario, entonces tiene media constante y la covarianza sólo depende del lapso k . Además, como la definición de función de autocorrelación es:

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

P.D. $\rho_0 = 1$

$$\begin{aligned} \rho_0 &= \frac{\text{cov}(x_t, x_{t+0})}{\text{var}(x_t)} = \frac{E\{[x_t - E(x_t)][x_{t+0} - E(x_{t+0})]\}}{\text{var}(x_t)} \\ &= \frac{E\{[x_t - \mu][x_{t+0} - \mu]\}}{\text{var}(x_t)} = \frac{E\{[x_t - \mu]^2\}}{\text{var}(x_t)} \\ &\text{entonces } \rho_0 = \frac{\text{var}(x_t)}{\text{var}(x_t)} = 1 \quad \square \end{aligned}$$

P.D. $\rho_k = \rho_{-k}$

Por la definición de autocorrelación basta con ver que $\gamma_k = \gamma_{-k}$ y como $\{x_t\}$ un proceso estacionario se tiene que γ_k sólo depende del lapso temporal k , i.e. la distancia entre t y $t+k$.

$$d(t, t+k) = |(t+k) - t| = |t - (-k) - t| = |t - (t-k)| = d(t-k, t)$$

Por lo cual la distancia de x_t a x_{t+k} es la misma que la distancia de x_{t-k} a x_t entonces

$$\begin{aligned} \text{cov}(x_t, x_{t+k}) &= \text{cov}(x_{t-k}, x_t) \\ \text{entonces } \gamma_k &= \gamma_{-k} \\ \text{por lo tanto } \rho_k &= \rho_{-k} \quad \square \end{aligned}$$

P.D. $|\rho_k| \leq 1$

Por definición $\rho_k = \frac{\gamma_k}{\gamma_0}$

Entonces $|\rho_k| = \left| \frac{\gamma_k}{\gamma_0} \right| = \frac{|\gamma_k|}{\gamma_0}$ ya que $\gamma_0 = \sigma^2$ por lo cual bastará con demostrar que $|\gamma_k| \leq \gamma_0$. Para lo cual se usará la desigualdad de Cauchy-Schwarz que dice: Sean x y y variables aleatorias con segundo momento finito entonces $|E(xy)| \leq \sqrt{E(x^2)E(y^2)}$. Los supuestos de la desigualdad se cumplen por la definición de estacionariedad.

Entonces se tiene:

$$\begin{aligned} |\gamma_k| &= |E\{[x_t - E(x_t)][x_{t-k} - E(x_{t-k})]\}| \\ &= |E\{[x_t - \mu][x_{t-k} - \mu]\}| \leq \sqrt{E\{[x_t - \mu]^2\}E\{[x_{t-k} - \mu]^2\}} \end{aligned}$$

³La matriz de autocorrelación será definida positiva si todos sus valores característicos son positivos, i.e. $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \rho_{(i-j)} > 0$ para cualquier i, j que representen el tiempo y $\alpha \in \Re$

entonces $|\gamma_k| \leq \sqrt{\text{var}(x_t) \text{var}(x_{t-k})} = \sqrt{[\text{var}(x_t)]^2} = |\text{var}(x_t)| = \text{var}(x_t) = \gamma_0$
entonces $|\gamma_k| \leq \gamma_0$ por lo tanto $|\rho_k| \leq 1$ \square

P.D. Que la matriz de autocorrelación es definida positiva.

Una matriz A cuadrada y simétrica es definida positiva si su forma cuadrática es mayor estrictamente que cero, i.e. $X'AX > 0$, con X cualquier vector no nulo de $n \times 1$.

Sea $\{x_t\}$ un proceso estocástico estacionario, por lo cual $E(x_t)$ es una constante, sin pérdida de generalidad $E(x_t) = 0$ y sea $x_t = \sum_{i=1}^k \phi_i x_{t-i}$.

Se tiene que $0 \leq \text{var}(x_t)$ además $\rho_k = \frac{\gamma_k}{\gamma_0}$ por lo cual se requiere que $0 < \text{var}(x_t)$ para que la función de autocorrelación este bien definida.

Entonces

$$\begin{aligned} 0 < \text{var}(x_t) &= E(\{x_t - E(x_t)\}^2) \\ &= E(x_t^2) = E\left(\sum_{i=1}^k \phi_i x_{t-i} \sum_{j=1}^k \phi_j x_{t-j}\right) \\ &= E\left(\sum_{i=1}^k \sum_{j=1}^k \phi_i \phi_j x_{t-i} x_{t-j}\right) \\ &= \sum_{i=1}^k \sum_{j=1}^k \phi_i \phi_j E(x_{t-i} x_{t-j}) \\ &= \sum_{i=1}^k \sum_{j=1}^k \phi_i \phi_j \gamma(i-j) \end{aligned}$$

$$\text{entonces } 0 < \sum_{i=1}^k \sum_{j=1}^k \phi_i \phi_j \gamma(i-j).$$

Y por lo tanto $\sum_{i=1}^k \sum_{j=1}^k \phi_i \phi_j \rho(i-j) > 0$, es decir, la matriz de autocorrelación es definida positiva. \square

1.2.2. Ergodicidad.

En general las series de tiempo con que se trabaja están constituidas por observaciones históricas, es decir, no proceden de la experimentación, por lo cual son irrepetibles y se les

puede ver como una sola realización de un proceso aleatorio multivariado.

Para hacer inferencia estadística en problemas donde interviene una sola variable aleatoria es necesario repetir el experimento, cosa que no se puede hacer en econometría por la naturaleza de las variables empleadas, lo que plantearía un problema de inferencia estadística, a no ser porque se puede demostrar que, cuando un proceso estocástico cumple con ciertas condiciones, es posible estimar consistentemente sus características con una sola realización del mismo. Los procesos que cumplen tales condiciones se denominan ergódicos.

Por ejemplo, **un proceso estocástico estacionario** es ergódico en la media, μ , si es posible estimar consistentemente este parámetro con la media muestral temporal, definida como:

$$\bar{x} = N^{-1} \sum_{t=1}^N x_t.$$

No es la finalidad establecer aquí bajo qué condiciones un proceso es ergódico. Basta con decir que las condiciones casi siempre se cumplen para la clase de procesos que se ocupan (ver Birkhoff, 1931). En este tipo de procesos a medida que el lapso de tiempo crece, la autocorrelación del proceso va desapareciendo. Si el proceso sigue una distribución normal, puede hablarse de una independencia cada vez mayor de las nuevas observaciones respecto a las antiguas, entonces debe cumplirse que el $\lim_{k \rightarrow \infty} \rho_k = 0$, o dicho de otra forma, a medida que el tamaño muestral crece, se dispone siempre de nueva información para estimar los parámetros del proceso, lo que permite construir estimadores consistentes de los parámetros.

Y es debido a la ergodicidad que en la práctica donde se dispone de observaciones, x_t ($t = 1, 2, \dots, N$), relativas a un proceso estocástico que se supone estacionario la media del proceso se estima mediante

$$\bar{x} = N^{-1} \sum_{t=1}^N x_t.$$

Y la función de autocorrelación, ρ_k , se estima mediante la función de autocorrelación muestral, definida como

$$r_k = \frac{\sum_{t=k+1}^{N-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_t (x_t - \bar{x})^2} \quad k = 1, 2, \dots$$

La representación gráfica de r_k , denominada correlograma muestral, constituye el instrumento básico de análisis de series de tiempo, que generaliza el concepto de correlación ordinaria y sirve para evitar la influencia de las unidades de medida.

Barlett (1946) ha proporcionado una expresión para calcular la covarianza aproximada entre r_k y r_{k-s} en el caso de procesos estacionarios normales, supuesto $\rho_k = 0$ para toda $k > K$:

$$cov(r_k, r_{k-s}) \simeq \frac{1}{N} \sum_{-K+s}^K \rho_i \rho_{i-s}.$$

Para N suficientemente grande, si $\rho_k = 0$, r_k sigue aproximadamente una distribución normal. En la práctica se sustituye ρ_i para r_i y para $s = 0$, se obtiene

$$var(r_k) \simeq \frac{1}{N} \left(1 + 2 \sum_1^k r_i^2 \right) \text{ para toda } k > K,$$

expresión que suele usarse para verificar la hipótesis de que $\rho_k = 0$.

A menudo se aplica la fórmula de Barlett la cual es aproximada, para verificar si $\rho_k = 0$ para $k = 1, 2, \dots$, es decir, si la serie es aleatoria. Se suele calcular simplemente el intervalo $\left(\frac{-2}{\sqrt{N}}, \frac{2}{\sqrt{N}} \right)$ y se comprueba si cada r_k cae dentro del mismo, en el caso de que algún r_k caiga fuera, se rechaza la hipótesis de que $\rho_k = 0$, lo que se traduce en que hay evidencia de que la serie no es aleatoria.

1.2.3. Proceso Estocástico Estacionario Puramente Aleatorio.

Es el proceso estacionario fundamental en cuanto que sobre él se construyen el resto de los procesos utilizados en el análisis de series de tiempo. Siendo más frecuente la denominación ruido blanco tomado de la ingeniería. Es la unidad última de análisis. Generaliza el concepto ideal de perturbación aleatoria.

La definición se basa en la función de autocorrelación, siendo puramente aleatorio aquel proceso cuya función de autocovarianza es nula en cualquier retardo ($k = 0$ no es propiamente un retardo.)

Un proceso estacionario puramente aleatorio, o ruido blanco, se define por las siguientes condiciones:

$$\begin{aligned} \mu &= E(x_t) = 0 \\ \gamma_0 &= var(x_t) = \sigma^2 \\ \gamma_k &= cov(x_t, x_{t+k}) = 0. \end{aligned}$$

Teniendo como función de autocorrelación

$$\rho_k = \begin{cases} 1 & \text{si } k = 0 \\ 0 & \text{si } k \neq 0 \end{cases}$$

lo que muestra que no existe correlación alguna (dependencia estadística) sobre los valores desfasados en el tiempo. La información histórica no tiene ningún valor para predecir el

futuro; el proceso no tiene “memoria”.

Si $E(x_t) = \mu + \epsilon_t$ con $t \in T$ donde $\{\epsilon_t\}$ es ruido blanco y la función de autocorrelación muestral esta dada por

$$r_k = \frac{\sum_{t=k+1}^{N-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_t (x_t - \bar{x})^2}.$$

Además se tiene

$$\sqrt{n} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_H \end{pmatrix} \xrightarrow{D} N_H(0, I_H) \quad N_H \left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{pmatrix} \right) \Rightarrow \begin{matrix} \sqrt{n}\rho_1 & \rightarrow & N(0, 1) \\ \sqrt{n}\rho_2 & \rightarrow & N(0, 1) \\ \vdots & & \vdots \\ \sqrt{n}\rho_H & \rightarrow & N(0, 1) \end{matrix}$$

Donde \xrightarrow{D} denota convergencia en distribución. Si se tienen k variables aleatorias independientes que se distribuyen $N(0, 1)$ x_1, x_2, \dots, x_k entonces $x_1^2 + x_2^2 + \dots + x_k^2 \sim \chi_k^2$ por lo cual se tiene que $(\sqrt{n}r_1)^2 + (\sqrt{n}r_2)^2 + \dots + (\sqrt{n}r_H)^2 \sim \chi_H^2$.

Resulta que $Q(H) = n \sum_{i=1}^H \rho_i^2 \xrightarrow{D} \chi_H^2$, Q es el estadístico de Box-Pierce y cuando n “no es muy grande” se calcula de la siguiente forma $Q^*(H) = n(n+2) \sum_{i=1}^H \frac{\rho_i^2}{n-i} \xrightarrow{D} \chi_H^2$.

Si se tienen los datos $\{x_1, x_2, \dots, x_k\}$ y se quiere saber si estos datos provienen de un proceso de ruido blanco se utiliza el estadístico de Box-Pierce cambiando ρ_k por r_k , para realizar la prueba de hipótesis $H_0 : x_t = \mu + \epsilon_t$ donde $\{\epsilon_t\}$ es ruido blanco, es decir, $H_0 : x_t$ es ruido blanco contra $H_a : x_t$ no es ruido blanco. Y el criterio de rechazo es

$$\text{si } \begin{cases} Q < \chi_H^2 & \text{entonces no se rechaza } H_0 \\ Q > \chi_H^2 & \text{entonces se rechaza } H_0. \end{cases}$$

1.2.4. Proceso Estocástico No Estacionario.

El hecho de que la gran mayoría de las series de tiempo con las cuales se trabaja no sean estacionarias, como ya se ha mencionado, no es tan grave ya que hay formas de volver estacionarias a las que en un principio no lo eran. En muchos casos prácticos, la no estacionariedad de las series es homogénea, es decir, que sólo el nivel de la serie es lo que se ve afectado por la no estacionariedad, debido a que existe una tendencia polinomial adaptiva. Si éste es el caso la tendencia se eliminará aplicando el operador diferencia tantas veces como el grado del polinomio para dar lugar a una serie estacionaria.

Un ejemplo de un proceso no estacionario es la caminata aleatoria, que aunque no es homogéneo se vuelve estacionario aplicando el operador diferencia. Una caminata aleatoria se define de la siguiente forma:

$$x_t = x_{t-1} + \epsilon_t,$$

en donde ϵ_t es un proceso ruido blanco. Este modelo se ha utilizado como modelo para las cotizaciones en la Bolsa de Valores. Se comprobará que no es un proceso estacionario.

$$\begin{aligned} x_1 &= x_0 + \epsilon_1 \\ x_2 &= x_1 + \epsilon_2 = x_0 + \epsilon_1 + \epsilon_2 \\ &\dots \\ x_t &= x_0 + \sum_{i=1}^t \epsilon_i, \end{aligned}$$

sin pérdida de generalidad se puede decir que $x_0 = 0$, entonces tenemos

$$\begin{aligned} x_t &= \sum_{i=1}^t \epsilon_i \\ E(x_t) &= E\left(\sum_{i=1}^t \epsilon_i\right) = \sum_{i=1}^t E(\epsilon_i) = 0, \end{aligned}$$

ya que ϵ_t es un proceso ruido blanco.

Obteniendo la varianza

$$\begin{aligned} var(x_t) &= var\left(\sum_{i=1}^t \epsilon_i\right) \\ &= \sum_{i=1}^t var(\epsilon_i) + \sum_{i \neq j} \sum_{j=1}^t cov(\epsilon_i, \epsilon_j) \\ &= \sum_{i=1}^t var(\epsilon_i) + 0 = \sum_{i=1}^t \sigma^2 \end{aligned}$$

$$\text{entonces } var(x_t) = t\sigma^2.$$

Con lo cual se ve que la varianza de una caminata aleatoria es función del tiempo y por tanto no es estacionaria.

Aplicando la primera diferencia al proceso resulta

$$\nabla x_t = x_t - x_{t-1} = \epsilon_t.$$

Y puesto que ϵ_t es un proceso ruido blanco, x_t es un proceso integrado de primer orden ($x_t \sim I(1)$), es decir, ∇x_t es un proceso estacionario. En este proceso, la operación de tomar diferencia tuvo como consecuencia la estabilización de la varianza. En general, la aplicación de diferencias en un proceso no estacionario cualquiera produce, hasta cierto punto, reducción de la varianza. Aún que la “sobrediferenciación” suele producir el efecto contrario.

Al agregar una constante al modelo de caminata aleatoria se obtiene lo que se conoce como caminata aleatoria con deriva $x_t = \delta + x_{t-1} + \epsilon_t$.

Si se tienen los datos $\{w_1, w_2, \dots, w_n\}$ y se quiere saber si éstos provienen de una caminata aleatoria, entonces la hipótesis nula es; $H_0 : \{w_t\}$ es una caminata aleatoria.

Bajo H_0 se tiene

$$\begin{aligned} y_t &= w_t - w_{t-1} \\ &= (\epsilon_1 + \epsilon_2 + \dots + \epsilon_t) - (\epsilon_1 + \epsilon_2 + \dots + \epsilon_{t-1}) = \epsilon_t. \end{aligned}$$

Por lo que se tiene que $\{y_t\}$ es ruido blanco, de tal modo que si no rechazamos que $\{y_t\}$ es ruido blanco entonces $\{w_t\}$ es caminata aleatoria.

1.3. Análisis Univariante de Series de Tiempo.

Dentro del *análisis univariante de series de tiempo* se concibe a cada serie de tiempo, integrada por cuatro componentes:

- Tendencia, o movimientos regulares de la serie a largo plazo.
- Ciclos, que son fluctuaciones a mediano plazo en torno a la tendencia, cuyo período y amplitud presentan cierta regularidad.
- Componente estacional, u oscilaciones a corto plazo de período regular, inferior o igual a un año, y amplitud bastante regular.
- Componente irregular, es la parte residual de la serie cuando se eliminan las componentes anteriores.

Sean:

$$\begin{aligned} x_t &= \text{Serie de tiempo.} \\ T_t &= \text{Tendencia.} \\ C_t &= \text{Ciclo.} \\ S_t &= \text{Componente estacional.} \\ \epsilon_t &= \text{Componente irregular.} \end{aligned}$$

Existen tres modelos de descomposición de la serie $\{x_t\}$ que suelen usarse alternativamente en la práctica, según se comporte la serie analizada y son:

Aditivo $x_t = T_t + C_t + S_t + \epsilon_t,$

Multiplicativo $x_t = T_t * C_t * S_t * \epsilon_t.$

Y el mixto, que puede adoptar, por ejemplo, la siguiente forma:

$$x_t = T_t * C_t * S_t + \epsilon_t.$$

El objetivo del análisis de series de tiempo es estudiar la dinámica o estructura temporal de la información, en otras palabras el objetivo es descubrir el proceso generador de datos (PGD).

1.3.1. Modelos Estacionarios Lineales.

Para analizar y predecir series de tiempo estacionarias bajo el enfoque de Box-Jenkins se utilizan tres tipos de modelos: el autorregresivo (AR), el de medias móviles (MA) y el autorregresivo de medias móviles o mixto (ARMA).

Modelo Autorregresivo de Orden p AR(p).

El caso general de los procesos autorregresivos es el denominado AR(p) y queda descrito por:

$$y_t = (1 - \phi_1 - \phi_2 - \dots - \phi_p) \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t.$$

La cual se puede expresar como:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t \text{ en donde } x_t = y_t - \mu.$$

Haciendo uso del operador rezago B , se puede escribir

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t = \epsilon_t.$$

Definiendo

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \text{ se tiene } \Phi(B)x_t = \epsilon_t.$$

De tal suerte que la condición necesaria y suficiente de estacionariedad para un proceso AR(p) suele expresarse diciendo que las raíces de la ecuación característica $\Phi(Z) = 0$ deben estar fuera del círculo unitario. Este resultado se recoge del siguiente desarrollo.

Expresando AR(p) como $(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t = \epsilon_t$, cuyo polinomio característico es $\Phi(Z) = 1 - \phi_1 Z - \phi_2 Z^2 - \dots - \phi_p Z^p$, se cambia la letra B , para evitar confusión

ya que ésta se ha usado para el rezago.

Se sabe por el teorema fundamental del álgebra que la ecuación

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

puede factorizarse como

$$(1 - \lambda_1 Z)(1 - \lambda_2 Z) \dots (1 - \lambda_p Z) = 0,$$

en donde $\lambda_1, \lambda_2, \dots, \lambda_p$ son las inversas de las raíces de ecuación característica

$$1 - \phi_1 Z - \phi_2 Z^2 - \dots - \phi_p Z^p = 0 \text{ es decir } \lambda_i = \frac{1}{Z_i}, \quad i = 1, 2, \dots, p.$$

De lo anterior se desprende que

$$\begin{aligned} (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) x_t &= \epsilon_t \\ (1 - \lambda_1 B)(1 - \lambda_2 B) \dots (1 - \lambda_p B) x_t &= \epsilon_t \\ x_t &= (1 - \lambda_1 B)^{-1} (1 - \lambda_2 B)^{-1} \dots (1 - \lambda_p B)^{-1} \epsilon_t. \end{aligned}$$

De tal modo que

$$x_t = \prod_{i=1}^p (1 - \lambda_i B)^{-1} \epsilon_t.$$

Para que x_t sea finito, $(1 - \lambda_i B)^{-1}$ debe ser convergente para toda i , y para eso se requiere que $|\lambda_i| < 1$ como condición de estacionariedad. Y así

$$x_t = \prod_{i=1}^p \sum_{j=0}^{\infty} \lambda_i^j B^j \epsilon_t.$$

O bien

$$x_t = \sum_{j=0}^{\infty} \varphi_j B^j \epsilon_t,$$

lo que se llama un proceso de medias móviles de orden infinito $MA(\infty)$. Nótese como el mismo proceso $AR(p)$ puede representarse a través de p parámetros autorregresivos o a través de un número infinito de coeficientes φ asociados a un proceso ruido blanco. Para fines prácticos, resulta preferible trabajar con un número finito que con un número infinito de parámetros, sobre todo si la explicación que se tiene del fenómeno es la misma. A esta idea de economía en el número de parámetros involucrados se le conoce con el nombre de principio de parcimonia.

Para λ_i reales, la estacionariedad requiere que todos los $|\lambda_i| < 1$. Si se admite la existencia de valores de λ complejos, la condición es que λ_i tenga su módulo menor que la unidad, es decir, la condición de estacionariedad es que los λ_i estén dentro del círculo unitario.

Como los λ_i son las inversas de la ecuación característica, la condición de estacionariedad puede expresarse también diciendo que las raíces $\Phi(B) = 0$ deben tener módulo superior a la unidad.

Ahora se procederá a encontrar la función de autocovarianza para posteriormente hallar la función de autocorrelación, para realizar esto se utilizarán las ecuaciones de Yule-Walker, que se obtienen al multiplicar la expresión de un $AR(p)$ por $x_t, x_{t-1}, \dots, x_{t-k}$ para después tomar esperanza, para $k \geq p$

$$\begin{aligned}\gamma_0 &= \phi_1\gamma_1 + \phi_2\gamma_2 + \dots + \phi_p\gamma_p + \sigma^2 \\ \gamma_1 &= \phi_1\gamma_0 + \phi_2\gamma_1 + \dots + \phi_p\gamma_{p-1} \\ \gamma_2 &= \phi_1\gamma_1 + \phi_2\gamma_0 + \dots + \phi_p\gamma_{p-2} \\ &\dots \\ \gamma_k &= \phi_1\gamma_{k-1} + \phi_2\gamma_{k-2} + \dots + \phi_p\gamma_{k-p}\end{aligned}$$

recordando que $\gamma_k = \gamma_{-k}$ se tiene que el sistema es válido para cualquier k .

Para $k = p$ el sistema de ecuaciones es:

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_p \end{bmatrix} = \begin{bmatrix} \gamma_0 & \gamma_1 & \dots & \gamma_{p-1} \\ \gamma_1 & \gamma_0 & \dots & \gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{p-1} & \gamma_{p-2} & \dots & \gamma_0 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix}$$

al dividir entre la varianza se tiene que la función de autocorrelación queda

$$\rho_k = \phi_1\rho_{k-1} + \phi_2\rho_{k-2} + \dots + \phi_p\rho_{k-p}.$$

Y su forma matricial es

$$\begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad \text{o bien, } \rho_p = \mathbf{P}_{pp}\phi_p.$$

En la práctica no se conoce el valor de p . Es conveniente preguntar por las soluciones de $\phi_p = \mathbf{P}_{pp}^{-1}\rho_p$ en el caso en que se especifique un número k de parámetros ϕ_i siendo $k \neq p$.

- Si $k < p$, el sistema de ecuaciones que se genera permite estimar k coeficientes, pero estas estimaciones serían incorrectas (no coinciden con los k primeros coeficientes ϕ_i verdaderos).
- Si $k > p$, el sistema de ecuaciones que se genera tiene $k - p$ ecuaciones redundantes, por lo que proporciona los mismos resultados que $k = p$, es decir, que $\phi_{p+1} = \phi_{p+2} = \dots = \phi_{p+k} = 0$.

Dado que las soluciones de los ϕ_i depende del valor k se utiliza la siguiente notación: ϕ_{ik} es la solución para ϕ_i del sistema de ecuaciones de Yule-Walker de tamaño k . De tal suerte que

$$\text{si } \begin{cases} k < p & \text{entonces } \phi_{ik} \neq \phi_i \\ k = p & \text{entonces } \phi_{ik} = \phi_i \\ k > p & \text{entonces } \begin{cases} \phi_{ik} = \phi_i & \text{si } i \leq p \\ \phi_{ik} = 0 & \text{si } i > p. \end{cases} \end{cases}$$

De todos los coeficientes ϕ_{ik} ($i = 1, 2, \dots, k$) interesa particularmente el k -ésimo, es decir, ϕ_{kk} .

Al ser \mathbf{P}_{pp} definida positiva, se tiene que \mathbf{P}_{pp} es no singular por lo que se puede resolver el sistema de ecuaciones con la regla de Cramer y queda

$$\phi_{kk} = \frac{\text{Det} \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & \rho_1 & \rho_k \end{bmatrix}}{\text{Det} \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_{p-1} \\ \rho_1 & 1 & \cdots & \rho_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{p-1} & \rho_{p-2} & \cdots & 1 \end{bmatrix}}.$$

La sucesión ϕ_{kk} , contemplada como función de k , se denomina Función de Autocorrelación Parcial (FACP).

La FACP mide la correlación entre x_t y x_{t-k} , una vez eliminada la influencia de los rezagos intermedios. Así el coeficiente ϕ_{kk} , mide la correlación entre x_t y x_{t-k} eliminando la influencia de $x_{t-1}, x_{t-2}, \dots, x_{t-k+1}$ sobre x_t y x_{t-k} . Esto se expresa mediante

$$\text{corr}(x_t, x_{t-k} | x_{t-1}, x_{t-2}, \dots, x_{t-k+1}),$$

esto es, considere la serie $\{x_t\}$ para un k dado x_t y x_{t-k} son renombrados como la regresión de sus valores intermedios, es decir,

$$\begin{aligned} x_t &= \alpha_0 + \alpha_1 x_{t-1} + \alpha_2 x_{t-2} + \cdots + \alpha_{k-1} x_{t-k+1} \\ x_{t-k} &= \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \cdots + \beta_{k-1} x_{t-k+1} \end{aligned}$$

en estas regresiones lineales los parámetros se estiman por medio de mínimos cuadrados, obteniendo los correspondientes estimadores de $\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_{k-1}, \hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_{k-1}$, obteniendo los residuales

$$\begin{aligned} R_t &= x_t - (\hat{\alpha}_0 + \hat{\alpha}_1 x_{t-1} + \hat{\alpha}_2 x_{t-2} + \dots + \hat{\alpha}_{k-1} x_{t-k+1}) \\ R_{t-k} &= x_{t-k} - (\hat{\beta}_0 + \hat{\beta}_1 x_{t-1} + \hat{\beta}_2 x_{t-2} + \dots + \hat{\beta}_{k-1} x_{t-k+1}). \end{aligned}$$

Entonces la función de autocorrelación parcial se define como sigue:

$$\phi_{kk} = \frac{\text{cov}(R_t, R_{t-k})}{\sqrt{\text{var}(R_t)}\sqrt{\text{var}(R_{t-k})}}.$$

Y como las ecuaciones de Yule-Walker se obtienen al minimizar

$$E [(x_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_{t-1} - \hat{\alpha}_2 x_{t-2} - \dots - \hat{\alpha}_k x_{t-k})^2],$$

los coeficientes de la FACP son la sucesión de los últimos coeficientes de las autorregresiones sucesivas

$$\begin{aligned} x_t &= \phi_{11} x_{t-1} + \epsilon_t \\ x_t &= \phi_{12} x_{t-1} + \phi_{22} x_{t-2} + \epsilon_t \\ x_t &= \phi_{13} x_{t-1} + \phi_{23} x_{t-2} + \phi_{33} x_{t-3} + \epsilon_t \\ &\dots \\ x_t &= \phi_{1k} x_{t-1} + \phi_{2k} x_{t-2} + \dots + \phi_{kk} x_{t-k} + \epsilon_t. \end{aligned}$$

La función de autocorrelación parcial FACP de los AR, permite identificar el orden del proceso autorregresivo. Obsérvese que, si el proceso autorregresivo es de orden p y se plantean ecuaciones autorregresivas con retardos superiores a p , es de esperar que los coeficientes correspondientes a los retardos superiores a p sean significativamente iguales con cero. Por lo cual si hay evidencia de que $\phi_{kk} = 0$ para todo k mayor que un p dado, entonces el proceso generador de datos es un $\text{AR}(p)$.

Para la FACP muestral, Quenouille (1949) demostró que la varianza de la estimación es $\text{var}(\phi_{kk}) \simeq \frac{1}{N}$, para $k > p$, en donde N es el número de observaciones (las que quedan después de obtener las diferencias adecuadas, si es el caso). Como puede observarse, a diferencia del caso de la FAC la FACP depende exclusivamente del tamaño muestral. Mediante la varianza de ϕ_{kk} se puede verificar si una muestra dada procede de un proceso autorregresivo de orden dado, por ejemplo p comprobando si ϕ_{kk} es significativamente igual a cero para todo $k > p$, para lo cual basta observar si hay algún valor de $|\phi_{kk}|$ ($k > p$) superior al doble del error estándar, $\frac{2}{\sqrt{N}}$. Si se supera esta cota para algún $k > p$, se rechaza la hipótesis de que el orden del proceso es p .

Modelo de Medias Móviles de Orden q MA(q).

El proceso de medias móviles puede contemplarse como una aproximación de cualquier proceso débilmente estacionario, lo que proporciona una buena justificación al modelo de medias móviles. Al ser q finita, la estacionariedad del proceso de medias móviles está siempre garantizada, pero, si en la práctica se está aproximando cualquier proceso estacionario mediante uno de medias móviles, es de esperarse que los θ_i decrezcan a medida que i aumenta, esta condición que parece intuitivamente aceptable, se justifica más adelante.

En general, un proceso MA(q) de media nula se expresa

$$\begin{aligned} x_t &= \epsilon_t - \theta_1\epsilon_{t-1} - \theta_2\epsilon_{t-2} - \dots - \theta_q\epsilon_{t-q} \\ x_t &= (1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q) \epsilon_t \text{ entonces } x_t = \Theta(B)\epsilon_t \\ &\text{donde } \Theta(B) = 1 - \theta_1B - \theta_2B^2 - \dots - \theta_qB^q. \end{aligned}$$

Por la forma que adopta el MA(q) se puede ver que será estacionario, ya que resulta ser un ruido blanco, por lo que ni la media ni la varianza, ni las covarianzas de un proceso MA(q), dependen del tiempo:

$$\begin{aligned} E(x_t) &= \Theta(B)E(\epsilon_t) = 0, \\ \text{var}(x_t) &= \text{var}(\Theta(B)\epsilon_t) = \text{var}(\epsilon_t) + \sum \theta_i^2 \text{var}(B^i\epsilon_{t-i}) + \sum_{i \neq j} \sum \text{cov}(\theta_i\epsilon_i, \theta_j\epsilon_j) \\ \text{var}(x_t) &= (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma^2. \end{aligned}$$

La estacionariedad requiere que la varianza sea finita, que se verifica si

$$(1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) < \infty$$

lo cual se cumple, obviamente, pues, al ser limitado el número de parámetros θ_i la suma converge.

$$\begin{aligned} \text{cov}(x_t, x_{t-k}) &= E(x_t x_{t-k}) \\ &= E[(\epsilon_t - \theta_1\epsilon_{t-1} - \dots - \theta_k\epsilon_{t-k} - \dots - \theta_q\epsilon_{t-q}) \\ &\quad (\epsilon_{t-k} - \theta_1\epsilon_{t-k-1} - \dots - \theta_q\epsilon_{t-k-q})] \\ &= E[\epsilon_t\epsilon_{t-k} - \dots - \theta_1\epsilon_{t-1}\epsilon_{t-k} + \dots - \theta_k\epsilon_{t-k}^2 + \dots \\ &\quad + \theta_1\theta_{k+1}\epsilon_{t-k-1}^2 + \dots + \theta_q^2\epsilon_{t-q}\epsilon_{t-k-q}] \\ &= E[-\theta_k\epsilon_{t-k}^2 + \theta_1\theta_{k+1}\epsilon_{t-k-1}^2 + \dots + \theta_{q-k}\theta_q\epsilon_{t-q}^2] \\ &= -\theta_k E(\epsilon_t^2) + \sum_{i=k+1}^q \theta_{i-k}\theta_i E(\epsilon_t^2) \\ \text{cov}(x_t, x_{t-k}) &= \left[-\theta_k + \sum_{i=k+1}^q \theta_{i-k}\theta_i \right] \sigma^2 \text{ si } k = 1, \dots, q. \end{aligned}$$

De esta expresión se obtiene la función de autocorrelación

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{-\theta + \sum_{i=k+1}^q \theta_{i-k} \theta_i}{1 + \sum_{i=1}^q \theta_i^2} & k = 1, \dots, q \\ 0 & k > q \end{cases}$$

lo que muestra que un proceso MA(q) posee una memoria limitada q períodos, por lo tanto la función de autocorrelación FAC de los MA, permite identificar el orden de un proceso de medias móviles. Davies, Pate y Frost (1974) obtuvieron restricciones sobre los valores posibles de las autocorrelaciones, para un MA(q) que se pueden expresar como:

$$|\rho_k| \leq \begin{cases} \cos\left(\frac{\pi}{M+1}\right) & \text{si } k \text{ es divisor de } q+1 \\ \cos\left(\frac{\pi}{M+2}\right) & \text{si } k \text{ no es divisor de } q+1. \end{cases}$$

En donde M es el mayor entero menor o igual a $\frac{q+1}{k}$.

A menudo se aplica la fórmula de Barlett para verificar si $\rho_k = 0$ para todo $k > q$, es decir, si el proceso MA es de orden q . Como la fórmula es aproximada, se suele calcular el intervalo $\left(\frac{-2}{\sqrt{N}}, \frac{2}{\sqrt{N}}\right)$ y se comprueba que los ρ_t con $k > q$ caen fuera del mismo.

Ya se vio que por la estructura el proceso de medias móviles es estacionario y por tanto su matriz de varianzas-covarianzas definida positiva, entonces sólo falta establecer restricciones sobre los valores de los parámetros para asegurar que el MA(q) provenga de un proceso estocástico, es decir, que su función de autocorrelación sea única; tales restricciones sobre los valores de las θ_i se denominan de invertibilidad.

Ya que un proceso de medias móviles de orden q es de la forma

$$\begin{aligned} x_t &= \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q} \\ x_t &= (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \epsilon_t \\ x_t &= (1 - \omega_1 B) (1 - \omega_2 B) \dots (1 - \omega_q B) \epsilon_t \end{aligned}$$

$$\text{entonces } x_t = \prod_{i=1}^q (1 - \omega_i B) \epsilon_t,$$

en donde $\frac{1}{\omega_i}$ son las raíces de la ecuación característica $1 - \theta_1 Z - \theta_2 Z^2 - \dots - \theta_q Z^q = 0$. Por lo que se tiene

$$\prod_{i=1}^q \left(\frac{1}{1 - \omega_i B} \right) x_t = \epsilon_t \quad \text{entonces} \quad \prod_{i=1}^q \left(\sum_j^{\infty} \omega_i^j B^j \right) x_t = \epsilon_t$$

pero para que esto ocurra basta con que $|\omega_i| < 1$ y si se aceptan valores complejos la condición es que $\frac{1}{\omega_i}$ caigan fuera del círculo unitario.

Y de acuerdo con lo anterior

$$\prod_{i=1}^q (1 + \omega_i B + \omega_i B^2 + \omega_i B^3 + \dots) x_t = \epsilon_t$$

$$x_t + v_1 x_{t-1} + v_2 x_{t-2} + v_3 x_{t-3} + v_4 x_{t-4} + \dots = \epsilon_t$$

$$x_t = \epsilon_t + \sum_{i=1}^{\infty} v_i x_{t-i}.$$

Entonces así como el proceso $AR(p)$ admite una representación $MA(\infty)$, cualquier proceso $MA(q)$ puede expresarse como un $AR(\infty)$. Por lo anterior se dice que un proceso $MA(q)$ es invertible si admite una representación $AR(\infty)$, para lo cual la condición de invertibilidad, es que las raíces de la ecuación característica estén fuera del círculo unitario.

La importancia del concepto de invertibilidad radica en que todo proceso invertible está determinado de manera única por su FAC, lo cual no ocurre con procesos que no son invertibles.

Como ya se explicó el proceso $MA(q)$ es equivalente a un $AR(\infty)$, y, por lo tanto, la función de autocorrelación parcial tiene infinitos términos.

Obsérvese que la fórmula de Batlett, aplicada a este proceso, para el cual $\rho_k = 0$ para $k > q$, es

$$var(r_k) \simeq \frac{1}{N} \left(1 + 2 \sum_1^k r_i^2 \right),$$

resulta de interés para identificar el orden q del proceso de medias móviles a partir de la FAC muestral.

Modelo Autorregresivos de Medias Móviles de Orden p, q ARMA(p, q).

Dicha generalización surge del hecho de que las series de tiempo que se observan en la práctica, muchas veces presentan características tanto de procesos AR como de procesos MA. Además recuérdese que el principio de parcimonia sugiere construir modelos que incluyan el menor número de parámetros, lo cual puede lograrse si intervienen parámetros tanto autorregresivos como de promedios móviles.

Este modelo se define mediante la expresión

$$x_t = \delta + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}.$$

Si el proceso ARMA(p, q) es estacionario, la media, μ , es constante y viene dada por

$$E(x_t) = \mu = \delta + \phi_1(x_{t-1}) + \phi_2(x_{t-2}) + \cdots + \phi_p(x_{t-p}) + (\epsilon_t) \\ - \theta_1(\epsilon_{t-1}) - \theta_2(\epsilon_{t-2}) - \cdots - \theta_q(\epsilon_{t-q}) \\ \mu = \delta + \phi_1\mu + \phi_2\mu + \cdots + \phi_p\mu \quad \text{entonces } \mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \cdots - \phi_p}$$

de modo que para la estacionariedad es necesario que

$$\phi_1 + \phi_2 + \cdots + \phi_p \neq 1.$$

Utilizando el operador rezago, B , el modelo con media nula puede escribirse así:

$$\Phi(B)x_t = \Theta(B)\epsilon_t,$$

en donde $\Phi(B) = 1 - \phi_1B - \phi_2B^2 - \cdots - \phi_pB^p$ y $\Theta(B) = 1 - \theta_1B - \theta_2B^2 - \cdots - \theta_qB^q$.

La ecuación anterior puede escribirse así:

$$x_t = \Phi(B)^{-1}\Theta(B)\epsilon_t$$

por lo que el proceso ARMA(p, q) es equivalente a un proceso de medias móviles de infinitos términos con $p + q$ coeficientes independientes. Y se dirá que es un proceso causal o independiente del futuro.

Al igual que para el proceso autorregresivo, la condición necesaria y suficiente de estacionariedad es que $\Phi(B)^{-1}$ sea convergente, para lo cual las raíces de la ecuación característica, $\Phi(Z) = 0$, deben estar fuera del círculo unitario.

El modelo autorregresivo de medias móviles de orden (p, q) también puede escribirse

$$\Theta(B)^{-1}\Phi(B)x_t = \epsilon_t.$$

Esta expresión muestra que el proceso ARMA(p, q) equivale a un proceso autorregresivo de orden infinito con $p + q$ coeficientes independientes. Para que la ecuación anterior tenga sentido, es preciso que $\Theta(B)^{-1}$ sea convergente, es decir, que las raíces de la ecuación característica de la parte de medias móviles del proceso $\Theta(Z) = 0$ caigan fuera del círculo de radio uno. Ésta es la condición de invertibilidad del proceso, análoga a la ya expresada para el modelo de medias móviles. Las condiciones de invertibilidad y estacionariedad son independientes.

Ahora se encontrarán las funciones de autocorrelación y autocorrelación parcial, ya se vio que el proceso ARMA(p, q) es equivalente a un proceso autorregresivo de orden infinito. Por lo tanto la FACP del proceso ARMA(p, q) tendrá infinitos términos. Análogamente,

se vio que el proceso ARMA(p, q) se comporta como un proceso de medias móviles de infinitos términos, así que su FAC tendrá también infinitos términos.

Una expresión útil de la función de autocovarianza de un proceso ARMA(p, q) de media nula se obtiene multiplicando el proceso por x_{t-k} y tomando esperanza.

La función de autocovarianza es

$$\begin{aligned} \gamma_k = E(x_t x_{t-k}) = & E(\phi_1 x_{t-1} x_{t-k} + \phi_2 x_{t-2} x_{t-k} + \cdots + \phi_p x_{t-p} x_{t-k} \\ & + \epsilon_t x_{t-k} - \theta_1 \epsilon_{t-1} x_{t-k} - \theta_2 \epsilon_{t-2} x_{t-k} - \cdots - \theta_q \epsilon_{t-q} x_{t-k}). \end{aligned}$$

Obsérvese que, para $k > q$, la parte debida a la media móvil se anula en esta expresión, ya que $E(\epsilon_{t-i} \epsilon_{t-j}) = 0$ para toda $i \neq j$. Esto significa que la FAC de un proceso ARMA(p, q) se comporta igual que la de un proceso autorregresivo a partir de $k > q$, es decir, que disminuirá exponencial o senoidalmente, con la posibilidad de alternar signos,

$$\gamma_k = \phi_1 \gamma_{k-1} + \phi_2 \gamma_{k-2} + \cdots + \phi_p \gamma_{k-p}, \quad k > p$$

en cuanto a la FACP, si $p > q$, se comporta como la de un proceso MA después de $p - q$ rezagos, es decir, con decrecimiento exponencial o mediante oscilaciones amortiguadas.

1.3.2. Transformaciones de Series No Estacionarias.

Hasta ahora se han analizado modelos para series estacionarias, pero en la práctica las series observadas rara vez aparecen como la imagen empírica de un proceso estacionario. Dentro del análisis de series de tiempo, se admiten tendencias deterministas. Las transformaciones estacionarias permiten tratar las series observadas dentro de la teoría de los modelos estacionarios.

Las tres formas más frecuentes de no estacionariedad son los cambios en la media, en la varianza y la estacionalidad, es decir la presencia de algún patrón de comportamiento sistemático.

Si la no estacionariedad se debe a que la varianza no es constante, quizá se deba a que en cada punto de observación t , la variable x_t tiene varianza σ_t^2 la cual es función de su media μ_t , de ocurrir esto un argumento derivado del trabajo de Bartlett, conduce a determinar una transformación potencia del tipo

$$G(x_t) = \begin{cases} x_t^\lambda & \text{si } \lambda \neq 0 \\ \log(x_t) & \text{si } \lambda = 0 \end{cases}$$

dicha transformación puede ser útil para estabilizar la varianza de la serie.

Series Con Tendencia y Raíces Unitarias.

Una tendencia determinista podría ser del tipo

$$x_t = \alpha + \beta t,$$

la corrección de una tendencia lineal mediante la primera diferencia, puede resultar en contradicción con el diagnóstico indicado en la relación de la caminata aleatoria. Una tendencia representada por un polinomio de orden bajo, debería ser eliminada mediante un polinomio lineal.

Una tendencia estocástica puede representarse mediante

$$x_t = \delta + x_{t-1} + \epsilon_t,$$

que es una caminata aleatoria con deriva. En este caso, como hemos visto al tratarse de la caminata aleatoria, puede convertirse en estacionaria con la primera diferencia.

Se puede considerar una función más general,

$$x_t = \alpha + \beta t + \gamma x_{t-1} + \epsilon_t,$$

que abarca ambas posibilidades determinista (los dos primeros términos del segundo miembro) y estocástica (los dos últimos).

Si $\beta \neq 0$ y $\gamma = 0$, solo existirá la tendencia determinista, que se estimará por una regresión polinómica. Si $\beta = 0$ y $\gamma = 1$, la tendencia es estocástica. El procedimiento consiste en estimar esta función, y contrastar la significancia de γ mediante la t de Student.

Si se diagnostica una tendencia estocástica, habrá que diferenciar, siendo d el número de veces que se aplicaría esta transformación, haciendo a la serie integrada de orden d . Aun cuando d no tiene a priori límite alguno, hay que evitar el peligro de sobrediferenciar ya que se tiende a incrementar la varianza (el orden sería aquel que estabilizará la varianza), e introduce autocorrelación.

Para verificar si una serie temporal, $x_t \sim I(0)$, es decir, es estacionaria, los métodos utilizados son de dos tipos: los empleados en la metodología Box-Jenkins y los procedimientos basados en contrastes. Los primeros, en los que se basaba el análisis tradicional de series de tiempo consisten el análisis gráfico de la serie y de los correlogramas. En el caso de series estacionarias, la función de autocorrelación decrece rápidamente, al contrario de lo que sucede en presencia de una raíz unitaria.

Estos procedimientos, a pesar de su cómoda y fácil interpretación, no presentan la formalidad requerida y sus resultados pueden en muchos casos ser interpretados discrecional-

mente.⁴ Por ello, además de examinar la varianza de la serie con distintos órdenes de diferenciación, se ha ido planteando diversos contrastes de raíces unitarias entre los más usados están la prueba Durbin-Watson, las diferentes versiones de la prueba Dickey-Fuller, la prueba de Schmidt-Phillips, la prueba de variables instrumentales de Hall y la contrastación de más de una raíz unitaria.

Concretamente sólo se hará la exposición de las pruebas de Dickey-Fuller (DF), Dickey-Fuller Aumentada (DFA) y la contrastación de más de una raíz unitaria.

La **prueba Dickey-Fuller** (DF) parte de que el proceso generador de datos es $x_t = \alpha + \gamma x_{t-1} + \epsilon_t$ en donde $\{\epsilon_t\}$ es ruido blanco.

Para verificar la hipótesis nula de que x_t no es estacionario basta con verificar la hipótesis de que $\gamma = 1$ frente a la alternativa de que $\gamma \neq 1$. En efecto si $\gamma = 1$ el proceso será una caminata aleatoria con deriva que ya se vio que no es estacionaria.

En este caso el modelo puede escribirse

$$(1 - B)x_t = \alpha + \epsilon_t.$$

Obsérvese que la ecuación característica tiene una raíz unitaria, lo que justifica el nombre genérico de este tipo de pruebas. La forma práctica de llevar a cabo la verificación de que $\gamma = 1$ es la siguiente:

$$\begin{aligned} x_t - x_{t-1} &= \alpha + \gamma x_{t-1} - x_{t-1} + \epsilon_t \\ (1 - B)x_t &= \alpha + (\gamma - 1)x_{t-1} + \epsilon_t, \end{aligned}$$

y verificar que el coeficiente de x_{t-1} es igual a cero frente a la hipótesis alternativa de que es menor que cero. Hay un problema en usar la razón t para verificar esta hipótesis, ya que cuando la hipótesis nula ($\gamma - 1 = 0$) es cierta, el estimador mínimo cuadrático de $\gamma - 1$ es sesgado hacia cero que se corresponde con el sesgo hacia dos, del estadístico Durbin-Watson. La consecuencia práctica de este hecho es que una prueba basada en el t de Student tenderá a rechazar incorrectamente la hipótesis nula. La “razón t ” no se distribuye en este caso como una t de Student o una normal al no ser x_t estacionario pero sus valores críticos han sido tabulados por Dickey (1975) utilizando experimentos de Monte Carlo. Fuller (1976) obtuvo la distribución límite apropiada. La distribución de la “razón t ” cambia si no existe el término constante de la regresión.

Si hay un término de tendencia también se producen cambios en la distribución de la “razón t ” que, en este caso, depende del coeficiente de dicho término pero no depende de

⁴Es necesario indicar que la metodología Box-Jenkins también se apoya en las pruebas Box-Pierce y Ljung-Box.

la constante.

Una forma más general de la prueba de Dickey-Fuller se obtiene con el modelo

$$x_t = \alpha + \beta t + \gamma x_{t-1} + \epsilon_t,$$

donde se verifica la hipótesis nula $\gamma = 1$ y $\beta = 0$, frente la alternativa $|\gamma| < 1$. Si la hipótesis nula es correcta la variable x se comportará como una caminata aleatoria. Para ello hay que calcular por mínimos cuadrados ordinarios el modelo denominado sin restricciones y el modelo con restricciones que son respectivamente los que siguen:

$$\begin{aligned}\nabla x_t &= \alpha + \beta t + (\gamma - 1)x_{t-1} + \gamma \nabla x_{t-1} + \epsilon_t \\ \nabla x_t &= \alpha + \gamma \nabla x_{t-1} + \epsilon_t,\end{aligned}$$

el modelo con restricciones es válido ya que la hipótesis nula es $\gamma - 1 = 0$ y $\beta = 0$. A partir de las estimaciones, se obtiene la suma de los errores cuadrados, dadas por D_U (sin restricciones) y D_R (con restricciones). Esto permite construir el estadístico $\mathbf{F}(q, n - k)$,

$$DF = \frac{(n - k)(D_R - D_U)}{qD_U}$$

siendo n el número de observaciones, k los parámetros estimados en la regresión no restringida y q las restricciones.

Los valores críticos tabulados por Dickey-Fuller son más elevados que los de la tabla \mathbf{F} clásica. Si el valor calculado es menor que el tabulado, no se rechaza la hipótesis de caminata aleatoria. Para el contraste puede recurrirse a las tablas de MacKinnon (1990), más completas que las de Dickey-Fuller. Fuller demostró que la distribución del estadístico t es independiente del número de diferencias incluidas.

Al plantear el contraste de Dickey-Fuller se está suponiendo que ϵ_t no está autocorrelacionado, es decir que ϵ_t es ruido blanco. Una solución a este problema es la sugerida por Dickey y Fuller en 1981, consiste en la inclusión en la prueba DF de una estructura de retardos de la variable dependiente que permite capturar la estructura autorregresiva de orden p de ésta, quedando la perturbación lo más incorrelacionada posible. Este contraste se conoce como *prueba de Dickey-Fuller Aumentada* (DFA), partiendo del modelo

AR(p) con tendencia determinista y constante

$$\begin{aligned}
x_t &= \mu + \beta t + \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \epsilon_t \\
x_t &= \mu + \beta t + (\phi_1 + \cdots + \phi_p) x_{t-1} - (\phi_2 + \cdots + \phi_p) x_{t-1} + \cdots \\
&\quad + (\phi_{p-1} + \phi_p) x_{t-p+1} - \phi_p x_{t-p+1} + \phi_p x_{t-p} + \epsilon_t \\
&= \mu + \beta t + (\phi_1 + \cdots + \phi_p) x_{t-1} \\
&\quad - [(\phi_2 + \cdots + \phi_p) (x_{t-1} - x_{t-2}) + \cdots + \phi_p (x_{t-p+1} - x_{t-p})] + \epsilon_t \\
x_t &= \mu + \beta t + \sum_{i=1}^p \phi_i x_{t-1} - \sum_{k=1}^{p-1} \sum_{i=k+1}^p \phi_i \nabla x_{t-k} + \epsilon_t.
\end{aligned}$$

El contraste consiste en estimar ∇x_t

$$\begin{aligned}
\nabla x_t &= x_t - x_{t-1} \\
\nabla x_t &= \mu + \beta t + \left(\sum_{i=1}^p \phi_i - 1 \right) x_{t-1} - \sum_{k=1}^{p-1} \sum_{i=k+1}^p \phi_i \nabla x_{t-k} + \epsilon_t \\
\nabla x_t &= \mu + \beta t + \alpha x_{t-1} + \sum_{i=1}^{p-1} \gamma_i \nabla x_{t-i} + \epsilon_t,
\end{aligned}$$

con p lo suficientemente grande para garantizar que ϵ_t sea aproximadamente ruido blanco. Para que x_t sea estacionaria es preciso que $\beta < 0$, en tanto que x_t será no estacionaria si $\beta = 0$, entonces en la expresión anterior se podría agrupar en el primer miembro todos los términos en ∇x_{t-i} , lo que proporcionaría un polinomio en B que premultiplica a x_t con un factor igual a $\nabla = 1 - B$, lo que implica una raíz unitaria.

La distribución asintótica de los parámetros $\hat{\mu}$, $\hat{\beta}$ y $\hat{\alpha}$ es la misma que en la prueba DF e independiente de los parámetros $\hat{\gamma}_i$ ⁵, los cuales siguen asintóticamente una distribución normal bajo la hipótesis nula de que x_t no es estacionaria. La inclusión de los citados rezagos dependerá de su significación según el estadístico t contrastado contra las tablas de la t -Student⁶.

El número de rezagos de la regresión se elige de manera que los residuos sean ruido blanco. Es posible que esto no se consiga. Si el PGD tiene una representación ARMA, la representación AR debe tener un número infinito de términos. En la práctica sería de esperar que los términos de orden superior tuviesen cada vez menos importancia.

Otra prueba alternativa es la de Phillips (1987), que puede aplicarse a PGD más generales que incluyan esquemas autorregresivos y cierta clase de heteroscedasticidad. Esta

⁵Fuller demostró que los valores críticos del estadístico t no dependen del número de términos rezagados.

⁶También se ha sugerido utilizar algún criterio de información como los de Akaike y Schwarz.

prueba sugiere transformar los estadísticos de la prueba DF para hacerlos compatibles con la presencia de autocorrelación y heteroscedasticidad en el término de perturbación. La idea es utilizar los residuos estimados ϵ_t en la regresión de DF para corregir el estadístico t asociado a los parámetros. De esta forma se obtienen nuevos estadísticos que tienen las mismas distribuciones límite de los estadísticos tabulados por Fuller en 1976. Pero el contraste más acertado suele ser el DFA.

Cuando se trata de contrastar más de una raíz unitaria, por ejemplo contrastar si $x_t \sim I(2)$, es decir, si presenta dos raíces unitarias en su representación autorregresiva. Dickey y Pantula en 1987 sugieren una estrategia de contraste partiendo del orden de integración más alto. Esto es debido a que si la variable es $I(2)$ y se contrasta la H_0 de que es $I(1)$, el modelo bajo la hipótesis alternativa no será estacionario, es decir, estará mal especificado ya que la variable es $I(2)$. Nótese que los contrastes propuestos toman como hipótesis alternativa la total ausencia de raíces unitarias. La idea, por tanto, es contrastar en primer lugar la hipótesis del mayor número de raíces unitarias consideradas, para ir disminuyendo éste a medida que se vaya rechazando su presencia. Por ello, si se quiere contrastar si $x_t \sim I(2)$ los autores sugieren el siguiente procedimiento basado en la prueba DFA:

1. Estimar:

$$\nabla^2 x_t = \alpha_1 \nabla x_{t-1} + \sum_{i=1}^p \gamma_i \nabla^2 x_{t-i} + \epsilon_t,$$

y contrastar:

$$\begin{cases} H_0 : x_t \sim I(2) & (\alpha_1 = 0) \\ H_1 : x_t \sim I(1) & (\alpha_1 < 0) \end{cases}$$

utilizando los valores críticos de la prueba DF.

2. En caso de rechazar la hipótesis nula de existencia de dos raíces unitarias se contrastaría la presencia de una estimando:

$$\nabla^2 x_t = \alpha_1 \nabla x_{t-1} + \alpha_2 x_{t-1} + \sum_{i=1}^p \gamma_i \nabla^2 x_{t-i} + \epsilon_t,$$

y contrastando:

$$\begin{cases} H_0 : x_t \sim I(1) & (\alpha_1 < 0 \text{ y } \alpha_2 = 0) \\ H_1 : x_t \sim I(0) & (\alpha_1 < 0 \text{ y } \alpha_2 < 0) \end{cases}$$

donde ϵ_t es un ruido blanco en todos los casos.

Modelo Autorregresivos Integrados de Medias Móviles de Orden p, d, q ARIMA(p, d, q).

Los *modelos autorregresivos integrados de medias móviles* (ARIMA) pueden ser vistos como una generalización de los modelos ARMA previamente estudiados. Yalom en

1955 sugirió la posibilidad de que un cierto tipo de no estacionariedad mostrado por algunas series de tiempo, podía representarse mediante la simple toma sucesiva de diferencias de la serie original. Esto permite gran flexibilidad de representación a los modelos ARMA, puesto que en realidad lo que se hace al aplicar el operador diferencia ∇^d es eliminar una posible tendencia polinomial de orden d , presente en la serie que se analice. Si la serie original $\{Z_t\}$ la cual no es estacionaria por causa de una tendencia polinomial no determinista a que se le denomina no estacionariedad homogénea de orden d , es posible construir una serie estacionaria $\{x_t\}$, en donde

$$x_t = \nabla^d Z_t \quad \text{para toda } t,$$

es estacionaria.

Si ahora se admite que x_t obedece un proceso ARMA(p, q), es decir, que

$$x_t = \frac{\Theta(B)}{\Phi(B)} \epsilon_t,$$

entonces

$$Z_t = \frac{\Theta(B)}{\nabla^d \Phi(B)} \epsilon_t,$$

y se dice que $\{Z_t\}$ es un proceso ARIMA(p, d, q) (autorregresivo integrado de media móvil). El término “integrado” se refiere a que Z_t se obtiene por el inverso del operador ∇^d , dando como resultado una suma infinita (o una integración); por ejemplo, el inverso del operador ∇ es

$$\nabla^{-1} = (1 - B)^{-1} = (1 + B + B^2 + B^3 + \dots),$$

lo que justifica que el proceso se llame “integrado”.

Este modelo se define mediante la expresión

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p} = \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q},$$

con $x_t = \nabla^d Z_t$ para toda t y ϵ_t ruido blanco.

Ya que el interés primordial se ubica en modelos que sean estacionarios e invertibles, se requiere que las raíces de $\Phi(x) = 0$ y las raíces de $\Theta(x) = 0$ se encuentren fuera del círculo unitario, o bien, si se considera como operador autorregresivo generalizado a;

$$\Psi(B) = \nabla^d \Phi(B),$$

la condición es que d de las raíces de $\Psi(x) = 0$ sean unitarias, mientras que las demás deben estar fuera del círculo unitario. Entonces la expresión se convierte en

$$z_t = \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \dots + \psi_{p+d} Z_{t-p-d} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \dots - \theta_q \epsilon_{t-q}.$$

Y la representación de suma ponderada de choques aleatorios viene dada por

$$Z_t = \frac{\Theta(B)}{\Psi(B)}\epsilon_t = \Gamma\epsilon_t,$$

en donde debido a que

$$\begin{aligned} & (1 - \psi_1 B - \psi_1 B^2 - \dots - \psi_{p+d} B^{p+d}) (1 - \gamma_1 B - \gamma_1 B^2 - \dots) \\ & = (1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_q B^q), \end{aligned}$$

se deduce que las ponderaciones γ deben satisfacer la ecuación en diferencia siguiente (donde B ahora opera sobre el subíndice j)

$$\Psi(B)\gamma_j = \Phi(B)\nabla^d\gamma_j = 0 \quad \text{para } j > \max\{p + d, q\}.$$

Así mismo, si el proceso es invertible, se tiene

$$\Pi(B)Z_t = \frac{\Psi(B)}{\Theta(B)}Z_t = \epsilon_t,$$

con las ponderaciones π que satisfacen

$$\Theta(B)\pi_j = 0 \quad \text{para } j > \max\{p + d, q\}.$$

Por último, nótese que implícitamente se ha supuesto que el nivel del proceso estacionario $\{x_t\}$ es cero, si esto no es así debe considerarse una media constante μw del proceso, lo cual vendría a darle mayor generalidad aun a los modelos ARIMA, o sea, si se considera que

$$\Phi(B)(x_t - \mu w) = \Theta(B)\epsilon_t,$$

se tendría como expresión general de un modelo ARIMA el que sigue

$$\Phi(B)x_t = \theta_0 + \Theta(B)\epsilon_t,$$

con $\theta_0 = \mu w (1 - \phi_1 - \dots - \phi_p)$. La representación asociada con este nuevo parámetro θ_0 es de una tendencia polinomial determinista, que no depende de los choques aleatorios $\{\epsilon_t\}$ y que por lo tanto es perfectamente predecible; debido a que esto es un supuesto, la costumbre es hacer $\theta_0 = 0$, a no ser que el conocimiento del fenómeno en estudio y/o los datos, indiquen lo contrario.

Con el fin de hacer más explícito el papel que juega el parámetro θ_0 , considere la representación

$$Z_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_r t^r + \frac{\Theta(B)}{\nabla^d \Phi(B)}\epsilon_t,$$

en donde las raíces de $\Theta(x) = 0$ y de $\Phi(x) = 0$ se encuentren fuera del círculo unitario. Nótese que este modelo no es estacionario pues contiene una tendencia polinomial determinista, asociada con el operador ∇^d , la cual produce que d raíces del operador autorregresivo

generalizado $\Psi(B) = \nabla^d \Phi(B)$ se encuentren sobre el círculo unitario. Si se desea eliminar a las fuentes de no estacionariedad, se tienen las siguientes opciones:

$$\begin{aligned}\nabla^d Z_t &= \frac{\Theta(B)}{\Phi(B)} \epsilon_t, & \text{si } d > r \\ \nabla^r Z_t &= r! \beta_r + \frac{\Theta(B)}{\Phi(B)} \epsilon_t, & \text{si } d = r \\ \nabla^r Z_t &= r! \beta_r + \frac{\nabla^{r-d} \Theta(B)}{\Phi(B)} \epsilon_t, & \text{si } d < r\end{aligned}$$

en la primera se han eliminado ambas fuentes de no estacionariedad y se obtiene el modelo estacionario e invertible

$$\Phi(B)x_t = \Theta(B)\epsilon_t \text{ con } x_t = \nabla^d Z_t \quad \text{para toda } t, \epsilon_t \text{ ruido blanco y } d \geq 1,$$

en la segunda solamente se ha eliminado la componente no determinista asociada con ∇^d y se llega a que

$$\Phi(B)x_t = \theta_0 + \Theta(B)\epsilon_t,$$

donde $x_t = \nabla^r Z_t$ y $\theta_0 = r! \beta_r (1 - \phi_1 - \dots - \phi_p)$; por último la tercera opción conduce al modelo

$$\Phi(B)x_t = \theta_0 + \nabla^{r-d} \Theta(B)\epsilon_t \text{ con } x_t = \nabla^r Z_t,$$

que no es invertible puesto que $r-d$ raíces del polinomio generalizado de promedios móviles $\nabla^{r-d} \Theta(B)$ están sobre el círculo unitario. Para evitar trabajar con modelos que no son invertibles, lo que se sugiere es aplicar el operador diferencia el número estrictamente necesario para eliminar la componente no determinista no estacionaria, pero sin llegar a provocar que el modelo no sea invertible, es decir, en el tercer caso se tendría

$$\nabla^d Z_t = \alpha_0 + \alpha_1 t + \dots + \alpha_{r-d} t^{r-d} + \frac{\Theta(B)}{\Phi(B)} \epsilon_t \text{ si } d < r.$$

Finalmente, cabe hacer notar una aclaración de carácter notacional, ya que en ocasiones se acostumbra referirse a modelos del tipo ARI o del tipo IMA, esto lo único que indica es que el modelo ARIMA es puramente autorregresivo o puramente de medias móviles, pero integrado.

Modelos Estacionales.

Otra fuente de no estacionariedad en las series de tiempo reales la constituye la estacionalidad. Por una **serie estacional** se entenderá una serie de tiempo que, a parte de contener una tendencia (y/o ciclos) de larga duración, muestra fluctuaciones que se repiten anualmente, quizá con cambios graduales a través de los años.

Una característica fundamental de las series estacionales con observaciones mensuales, es la similitud de comportamiento entre observaciones para el mismo mes en años consecutivos. De igual manera, si la serie estacional es de observaciones trimestrales, existirán similitudes entre los mismos trimestres para años consecutivos. Es importante que aun cuando

la estacionalidad se considera como un fenómeno repetitivo anual, esto no implica que no pueda existir un cierto patrón de comportamiento periódico con duración menor al año; por ejemplo podría existir un período estacional semestral, de tal forma que los meses de diciembre fuesen similares a los meses de junio, además de que los meses de diciembre muestren un comportamiento similar entre sí y de que los meses de junio se comporten también de manera parecida. Por lo anterior, es conveniente hablar genéricamente de un período estacional que comprenda s observaciones contiguas.

La eliminación de las variaciones estacionales, para inducir la estacionariedad, suele hacerse mediante un procedimiento de ajuste, del mismo tipo que el comentado para la tendencia, denominado *diferenciación estacional*.

El operador de *diferenciación estacional* ∇_s^D se define como

$$\begin{aligned}\nabla_s^D x_t &= (1 - B^s)^D x_t \\ &= \sum_{j=0}^D \frac{(-1)^j D!}{j!(D-j)!} x_{t-js}.\end{aligned}$$

Este procedimiento se basa en el modelo clásico aditivo de descomposición. Lo primero que conviene hacer es eliminar la tendencia de la serie, ya que, de otra forma, la diferencia entre los datos relativos al mismo mes (o fracción de año) sería significativa, sin que esto implique evidencia de variaciones estacionales.

Una vez que a la serie original x_t se le aplica D veces el procedimiento de diferenciación estacional se tiene

$$(1 - B^s)^D x_t = Z_t,$$

en donde $s = 4$ para datos trimestrales, 12 para datos mensuales, etc. Es posible que después de la operación de diferenciación estacional, la serie Z_t resulte ser estacionaria (en la práctica, lo sería sólo aproximadamente). Entonces se puede aplicar a Z_t cualquiera de los modelos para series estacionarias, AR(p), MA(q) o ARMA(p, q).

En el último caso, por ejemplo, se tiene:

$$\Phi(B)Z_t = \Theta(B)\epsilon_t,$$

y sustituyendo Z_t , se tiene:

$$\Phi(B)(1 - B^s)^D x_t = \Theta(B)\epsilon_t,$$

es posible que para conseguir la estacionariedad sea necesario aplicar d diferencias junto a la diferenciación estacional. En tal caso, Z_t sería ARIMA(p, d, q), es decir,

$$\Phi(B)(1 - B)^d (1 - B^s)^D x_t = \Theta(B)\epsilon_t.$$

Capítulo 2

Cointegración

En este capítulo se expone el tema principal de la presente tesis; *La Teoría de Cointegración* mostrando y desarrollando los resultados más importantes, con las limitantes propias de los alcances de la misma tesis. El propósito del presente capítulo plantear un panorama lo más completo posible de la teoría de *Cointegración*, de sus beneficios, utilidad y ventajas.

2.1. Causalidad en el sentido de Granger.

Una correlación alta entre dos variables, Y y X , puede deberse a tres tipos de relaciones causa efecto, entre ellas:

- a) que X sea la causa de las variables de Y .
- b) que Y sea la causa de los cambios en X .
- c) que cada una de ellas sea a la vez causa y efecto de la otra.

La **causalidad de Wiener-Granger** o **causalidad en el sentido de Granger**, véase Granger (1969), puede expresarse de forma simplificada como sigue: X causa a Y , en el sentido de Granger, ($X \rightarrow Y$) si Y_t se puede predecir con mayor exactitud utilizando valores pasados de X que sin usarlos, manteniendo igual el resto de la información.

Esta definición sugiere varios comentarios. En primer lugar, hay que observar que usa el concepto de causalidad en el sentido de “preceder” o de predecir, más que de “producir” (un efecto). Cuando se dice que ($X \rightarrow Y$) lo que se está expresando de hecho es que los valores de X “preceden” a los de Y , en el sentido de que anteceden siempre a los de Y y sirven para predecirlos, pero no que necesariamente los valores de X “originen” los valores de Y . Es posible que, por ejemplo, que una tercera variable Z produzca los cambios en Y y en X , sin embargo, ($X \rightarrow Y$). La expresión **causa en el sentido de Granger** no sugiere lo que significa, sería más apropiado hablar de **precedencia**, como propone Leamer.

En segundo lugar, se están considerando tan solo valores pasados de X para predecir Y . En realidad se puede extender esta definición a la causalidad instantánea ($X \Rightarrow Y$). Para ello en vez de considerar sólo valores pasados de X , se tendrían en cuenta valores pasados y presentes de X . Hay que matizar, sin embargo, que, hablando estrictamente, la causalidad instantánea no existe en un contexto de tiempo continuo ya que una causa siempre precede a su efecto aunque sea en un lapso de tiempo muy pequeño. Pero en la práctica el tiempo es siempre una variable discreta, por lo que puede hablarse de causalidad instantánea en el sentido que la relación comienza dentro del mismo período de tiempo unitario en que se origina la causa.

En el contexto de los modelos dinámicos resulta inmediato que la función de transferencia supone, en cierta forma, que X causa a Y . Sin embargo, en éste, como en otros aspectos, el análisis estadístico de las series de tiempo ha tratado de que sean los propios datos quienes refrenden o rechacen este posible nexo de causalidad.

La ya citada causalidad en el sentido de Granger, es definida por el propio Granger en la siguiente forma:

Sea $P(A|B)$ la función de distribución condicional de A dado B , sea Ω_t la representación de toda la información en el universo en el momento t y sea $(\Omega_t - X_t)$ toda la información excepto X_t . Se dirá que X_t no causa a Y_t (o más precisamente que no está temporalmente relacionada con Y_t) si

$$P(Y_{t+1}|\Omega_t) = P(Y_{t+1}|\Omega_t - X_t).$$

En un enfoque dinámico, la causalidad sólo puede entenderse en el sentido de que el pasado causa el presente o el futuro. Sin embargo, los períodos de medición utilizados pueden hacer coincidir, en un sólo momento t , la causa y el efecto (por ejemplo, dentro del mismo mes la acción promocional y su efecto sobre las ventas), y entonces se dirá que existe causalidad instantánea si

$$P(Y_{t+1}|\Omega_{t+1} - Y_{t+1}) \neq P(Y_{t+1}|\Omega_{t+1} - Y_{t+1} - X_{t+1}),$$

es decir, si en el momento $t + 1$ el valor más probable de predicción con toda la información en ese momento de todas las variables excepto la Y , es distinto del que se obtendría desconociendo también X .

Una interdependencia dinámica vendrá dada por una relación causal doble, y no solamente instantánea, entre Y_t y X_t , es decir, Y_t causa a X_t y X_t causa a Y_t .

A niveles más operativos -pero también más limitados- puede considerarse que dados dos modelos lineales correspondientes a las series estacionarias Y_t y X_t :

$$\begin{aligned} a(B)Y_t + b(B)X_t &= n_t \\ c(B)X_t + d(B)Y_t &= \epsilon_t \end{aligned}$$

donde a, b, c y d serán polinomios de rezagos con $a(0) = c(0) = 1$ (generalmente de infinitos términos, por ser cocientes a su vez de polinomios MA y AR) y donde n_t y ϵ_t se supone son ruidos blancos no correlacionados entre sí:

1. X_t no causa a Y_t si $b(B) = 0$.
2. Existe causalidad instantánea si $b(0) \neq 0$.

Para contrastar frecuentemente si $b(B) = 0$, el procedimiento más inmediato es el tradicional contraste \mathbf{F} de nulidad de un subconjunto de parámetros del modelo. Sin embargo, dado que $b(B)$ será, generalmente, un polinomio de infinitos términos de rezagos, es preciso aceptar una parametrización finita que permita su aplicación.

Atendiendo en particular a los residuos de los modelos univariantes de las variantes de las variables Y_t y X_t :

$$\begin{aligned} a_1(B)Y_t &= b_1(B)\epsilon_{1t} \\ a_2(B)X_t &= b_2(B)\epsilon_{2t} \end{aligned}$$

la ausencia o no de correlación entre estos residuos para diferentes retardos puede permitir definir la relación causal entre Y_t y X_t . Así, siendo $\rho_{12}(k)$ el coeficiente de correlación cruzada entre las dos series de ruido blanco ϵ_{1t} y ϵ_{2t} , para un desfase k , las diferentes relaciones de causalidad, según la propuesta de Pierce (1977), pueden deducirse del siguiente cuadro.

<i>Relación de causalidad entre Y y X</i>	<i>Función de correlación cruzada de los residuos</i>
Independencia	$\rho_{12}(k) = 0$ para todo k
Causalidad sólo instantánea	$\rho_{12}(0) \neq 0, \rho_{12}(k) = 0$ para $k \neq 0$
X no causa a Y	$\rho_{12}(k) = 0$ para todo $k < 0$
X causa a Y	$\rho_{12}(k) \neq 0$ para algún $k < 0$
Y causa a X	$\rho_{12}(k) \neq 0$ para algún $k > 0$
Interdependencia dinámica	$\rho_{12}(k) \neq 0$ para algún $k < 0$ y para algún $k > 0$

En los casos extremos, una función de correlación cruzada con valores significativamente distintos de cero, es indicativa de una interdependencia dinámica, mientras que su nulidad para todo k es prueba de la independencia entre Y_t y X_t .

Naturalmente, en la práctica sólo se dispone de las estimaciones muestrales de esos coefi-

cientes de correlación cruzada:

$$r_{12}(k) = \frac{\sum_{t=1}^{N-k} \hat{\epsilon}_{1t} \hat{\epsilon}_{1,t+k}}{\sqrt{\sum_{t=1}^k \hat{\epsilon}_{1t}^2 \sum_{t=1}^k \hat{\epsilon}_{2t}^2}}.$$

pero, bajo la hipótesis de no relación entre las series, puede demostrarse que los coeficientes de correlación muestral se distribuyen asintóticamente como variables normales independientes con media cero y desviación estándar $\frac{1}{\sqrt{N}}$.

De esta forma, se puede contrastar si cada uno de los coeficientes de autocorrelación calculados es o no estadísticamente significativo, alcanzando valores superiores (por ejemplo, a un 95 % de confianza) $a \pm \frac{2}{\sqrt{N}}$. Para un contraste conjunto de nulidad de $\rho_{12}(k)$ para $k > 0$ (es decir, de que Y no causa a X) puede utilizarse el estadístico:

$$Q = N \sum_{k=1}^m r_{12}^2(k),$$

que se distribuye asintóticamente, bajo esta hipótesis nula, como una χ^2 con m grados de libertad.

Si se quiere contrastar la independencia total (nulidad en todo $\rho_{12}(k)$ para cualquier valor positivo o negativo de k) se debe emplear

$$S = N \sum_{k=-m}^m r_{12}^2(k) \sim \chi_{2m+1}^2,$$

siendo habitual hacer $m = 10$, con lo que si la suma de los veintiún coeficientes de correlación al cuadrado no superan $\chi_{21}^2 = 32,7$, valor de las tablas para un nivel de confianza, por ejemplo, del 95 %, no se rechaza, la hipótesis nula de independencia, véase Pierce(1977).

Las pruebas más usadas para verificar la existencia o no de causalidad son la prueba de Granger y la de Sims-DMD. Para describir el primero considérese el siguiente modelo

$$\begin{aligned} Y_t &= \alpha_{10} D_{1t} + \sum_{j=1}^k \alpha_{1j} Y_{t-j} + \sum_{j=1}^k \beta_{1j} X_{t-j} + u_{1t} \\ X_t &= \alpha_{20} D_{2t} + \sum_{j=1}^k \alpha_{2j} X_{t-j} + \sum_{j=1}^k \beta_{2j} Y_{t-j} + u_{2t} \end{aligned}$$

en donde D_{it} son componentes deterministas de las ecuaciones. Sea, en la primera ecuación, $H_0 : \beta_{1j} = 0$, ($j = 1, 2, 3, \dots, k$) y H_1 : no es cierto que $\beta_{1j} = 0$. Si se rechaza H_0 entonces

hay evidencia de que X causa a Y . Para verificar H_0 se utiliza la prueba de los multiplicadores de Lagrange.

Para este fin se haya primero la regresión

$$Y_t = \alpha_{10}D_{1t} + \sum_{j=1}^k \alpha_{1j}Y_{t-j} + u_{1t},$$

se eliminan los regresores que son valores rezagados de X (por la restricción $\beta_{1j} = 0$) y se calculan los correspondientes residuos \hat{u}_t . A continuación se halla la regresión de \hat{u}_t respecto de todos los regresores. Se halla el coeficiente de determinación de esta ecuación, R^2 , y el estadístico $ML = N * R^2$; si la hipótesis nula es cierta, ML se distribuye como una χ_k^2 . Por lo tanto, en la práctica si ML supera el valor crítico de la χ_k^2 para un nivel de significancia dado, se rechaza H_0 , lo que significa que hay evidencia de que X causa a Y .

Si, además de que X causa a Y , tiene sentido pensar que Y causa a X , debe conducirse una segunda prueba, utilizando ahora la segunda ecuación del modelo.

Puede ocurrir que los resultados de la prueba dependan del número de rezagos k adoptados. En general deben usarse tantos rezagos como permitan los grados de libertad, pero si el resultado es sensible a esta especificación no puede depositarse mucha confianza en las conclusiones.

La otra prueba alternativa para verificar el orden de causalidad se basa en un trabajo de Sims (1972) y fue desarrollado por Geweke, Meese y Dent (1983). La idea básica de esta prueba es que el futuro no puede ser causa del presente. Supóngase que se quiere verificar que X causa a Y . Ahora se utilizará la ecuación

$$X_t = \alpha_0 + \sum_{j=1}^k \gamma_j X_{t-j} + \sum_{j=-m}^k \delta_j Y_{t-j} + v_t, \quad j \neq 0$$

en donde el primer término recoge todas las componentes deterministas de X . Obsérvese que las Y comienzan en $-m$. Un retardo negativo significa en realidad un adelanto. Si los valores adelantados de Y no figuran en la ecuación, ésta sería la segunda ecuación del modelo considerando anteriormente. La idea de esta prueba es que si los coeficientes de los valores adelantados de Y en la ecuación son significativos, esto debe interpretarse como que X causa a Y . El motivo de esta interpretación es que resulta más lógico pensar en este orden de causalidad que en el inverso ya que no tiene sentido que el futuro (Y_{t+1}, Y_{t+2}, \dots) cause el presente (X_t) sino al contrario.

El criterio a aplicar en este caso es, nuevamente, el de los multiplicadores de Lagrange. Para ello se halla la regresión restringida omitiendo los términos adelantados de Y (Y_{t+1}, Y_{t+2}, \dots).

La hipótesis H_0 es, $\delta_j = 0$ ($j = -1, -2, \dots, -m$). A continuación se halla la regresión de \hat{v}_t , que son los residuos de la regresión anterior, respecto a todos los regresores y se calcula el coeficiente de determinación R^2 . A partir de aquí se obtiene el estadístico $ML = N * R^2$. Este valor se compara con el valor crítico de la χ_k^2 para un nivel de significación dado. Si ML supera a este valor crítico, se rechaza H_0 , lo que se interpreta como evidencia de que X causa a Y .

Estas pruebas son correctas si Y_t y X_t son estacionarias. En los dos modelos, se ha supuesto que la parte no estacionaria es determinista, por ello se han incluido términos de tendencia y de variables ficticias estacionales, pero es posible que halla que formular la ecuación de verificación en las primeras diferencias de la variable dependiente o en las diferencias estacionales, para tener en cuenta la no estacionariedad de origen estocástico.

2.2. Modelos de Vectores Autorregresivos (VAR).

En la econometría tradicional, por lo que se refiere a los modelos multiecuacionales, las limitaciones más preocupantes implícitas son:

1. Se conoce el orden de la causalidad de las variables que intervienen el modelo.
2. Se sabe que variables hay que omitir en cada ecuación.

El conocer el orden de causalidad de las variables que intervienen en una ecuación significa que se sabe cuales son las variables causa y cual la variable efecto. En un modelo multiecuacional esto implica que se sabe cuáles son las variables endógenas y las exógenas del sistema considerado en su conjunto.

Por otra parte, si se sabe cuales son las variables que intervienen en cada ecuación, entonces se conoce las denominadas “restricciones cero” (los coeficientes de las variables omitidas).

En estas condiciones es posible llevar a cabo el denominado *análisis de la identificación*, que consiste en identificar *a priori*, a la vista de las restricciones cero del sistema, si es posible que existan distintos conjuntos de parámetros estructurales que arrojen los mismos datos con idéntica verosimilitud. Si tal posibilidad existiera sería imposible asignar estimaciones concretas a todos los parámetros estructurales, lo que significa que alguno o algunos se quedarían sin identificar¹.

Sims (1980) afirma que la mayor parte de las restricciones que aparecen en los modelos son falsas y establece las siguientes bases, en las que se asienta la modelización VAR (vector autorregresivo):

¹No es esencial para los fines de la tesis el conocer la naturaleza del problema de la identificación. Véase el manual de J. Johndton, Métodos de Econometría.

1. No hay conocimientos teóricos suficientes para clasificar a las variables en endógenas y exógenas.
2. No es posible a priori establecer restricciones cero.

Lo que Sims propone, en última instancia, es utilizar un sistema de ecuaciones en el que no se distinga entre variables endógenas y exógenas. Cada variable influye en las demás y es influida por todas. Así, en cada ecuación aparece despejada una variable diferente en función de todas y cada una de las demás variables del sistema, siempre rezagadas, y de sus propios rezagos. El sistema así formulado consiste en un conjunto de ecuaciones autorregresivas, que expresa un vector de variables en función de sus propios rezagos, lo que justifica su nombre de modelo con un vector autorregresivo o modelo VAR. Tal sistema de ecuaciones es asimilable a una forma reducida irrestricta, entendiéndose por tal que se deriva de un supuesto sistema estructural sin restricciones cero. Este sistema sería, desde luego, inidentificable, pero el sistema reducido podría utilizarse para predecir y para llevar a cabo simulaciones de política económica. En este contexto, en que no existen variables exógenas, la simulación de política es limitada pues se entiende como tal, el análisis de los efectos, sobre las variables del sistema, de un cambio aleatorio conocido.

Ya que en el segundo miembro de un modelo VAR sólo hay variables rezagadas, cada variable queda determinada exclusivamente en función de la información correspondiente al período precedente a través de su propia ecuación. Esto significa que no existe simultaneidad en la determinación de las variables, por lo que para su estimación puede prescindirse de los métodos especiales para abordar los problemas típicos de los modelos multiecuacionales. Si cada término de error no presenta autocorrelación (cosa que puede conseguirse adoptando para ello el número de rezagos adecuado), las variables del segundo miembro de cada ecuación no estarán correlacionadas con el correspondiente término de error y el método de los mínimos cuadrados ordinarios, aplicado ecuación por ecuación, proporcionará estimadores consistentes de los parámetros.

La aplicación del modelo VAR a la obtención de predicciones es directa ya que como no existen variables exógenas no hay que preocuparse de obtener valores de este tipo de variables para el período de predicción.

Sin embargo, las cosas se complican en la práctica. Para empezar hay que admitir la posibilidad de que existan correlaciones contemporáneas entre perturbaciones de ecuaciones diferentes, lo que afecta tanto a la estimación² del modelo como a su uso para la predicción y la simulación de políticas. Para obviar los inconvenientes derivados de este problema se recurre a la ortogonalización de las perturbaciones, lo que se consigue mediante la diagonalización de la matriz de varianzas y covarianzas contemporáneas.

²En tal caso debe aplicarse el método de estimación para ecuaciones aparentemente no relacionadas. Véase Theil, *Principles of Econometrics*, Wiley, 1971.

Una vez ortogonalizadas las perturbaciones, el modelo puede utilizarse, ecuación por ecuación, para el análisis de políticas. Este análisis hace surgir otros problemas. Uno de ellos es que los resultados pueden depender del orden de las ecuaciones en el sistema. En la práctica se puede decidir el orden de las ecuaciones aplicando una serie de pruebas de causalidad.

El primer paso en la aplicación de esta metodología es la transformación de las series originales en otras estacionarias. Esta transformación es indispensable para que más tarde se pueda ortogonalizar las perturbaciones, pero es también un requisito teórico que impone la propia metodología, que descansa en última instancia en el teorema de descomposición de Wold.

2.2.1. Definición.

Un **vector autorregresivo** de orden p [VAR(p)] de un sistema con K variables $Y_t = (y_{1t}, \dots, y_{Kt})'$ puede ser definido de la siguiente forma:

$$\begin{aligned} y_{1t} &= \pi_{11} + (\pi_{12,1}y_{1,t-1} + \pi_{13,1}y_{2,t-1} + \dots + \pi_{1K+1,1}y_{K,t-1}) + \dots \\ &\quad + (\pi_{12,p}y_{1,t-p} + \dots + \pi_{1K+1,p}y_{K,t-p}) + \epsilon_{1t} \\ &\quad \vdots \\ y_{Kt} &= \pi_{K1} + (\pi_{K2,1}y_{1,t-1} + \pi_{K3,1}y_{2,t-1} + \dots + \pi_{KK+1,1}y_{K,t-1}) + \dots \\ &\quad + (\pi_{K2,p}y_{1,t-p} + \dots + \pi_{KK+1,p}y_{K,t-p}) + \epsilon_{Kt}. \end{aligned}$$

En forma vectorial

$$Y_t = \begin{pmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{k,t} \end{pmatrix} \Pi_0 = \begin{pmatrix} \pi_{11} \\ \pi_{21} \\ \vdots \\ \pi_{K1} \end{pmatrix} \Pi_j = \begin{pmatrix} \pi_{12,j} & \pi_{13,j} & \dots & \pi_{1K+1,j} \\ \pi_{22,j} & \pi_{23,j} & \dots & \pi_{2K+1,j} \\ \vdots & \vdots & \ddots & \vdots \\ \pi_{K2,j} & \pi_{K3,j} & \dots & \pi_{KK+1,j} \end{pmatrix} \epsilon_j = \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \vdots \\ \epsilon_{Kt} \end{pmatrix}$$

$$Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t.$$

El vector ϵ_t tiene media cero, $E(\epsilon_t) = 0$, y la misma matriz (no singular) de varianza-covarianza $\Sigma_\epsilon = E(\epsilon_t \epsilon_t')$ para toda t . Además, ϵ_t y ϵ_s no están correlacionados para $t \neq s$. Un proceso con estas características es frecuentemente llamado **vector ruido blanco** en analogía con la terminología usada para los errores en los procesos univariantes.

Usualmente los parámetros $\Pi_0, \Pi_1, \dots, \Pi_p$ y Σ_ϵ serán desconocidos en la práctica y se deben estimar con los datos disponibles, antes que el proceso pueda ser usado para predecir. Previamente a considerar la estimación del modelo debe discutirse la estacionaridad del proceso VAR. Esta propiedad será útil para derivar las propiedades asintóticas de los estimadores de los parámetros.

2.2.2. Estacionariedad.

En el **Capítulo 1** se definió la estacionariedad para los procesos estocásticos univariantes como una propiedad que asegura que la media y la varianza sean constantes en el tiempo y que la autocovarianza no dependa del tiempo. De manera análoga, la estacionariedad puede definirse para los procesos de vectores autorregresivos.

Un proceso estocástico vectorial se dice estacionario si

- I Todos los vectores aleatorios tienen el mismo vector media $\underline{\mu}$, $E(Y_t) = \underline{\mu}$ para toda t
- II La varianza de todas las variables involucradas es finita, $var(y_{kt}) < \infty$ para toda t con $k = 1, \dots, K$.
- III La matriz de covarianzas de los vectores Y_t y Y_{t+k} que están separados por k períodos sólo depende de k y no de t

$$cov(Y_t, Y_{t+k}) = E \left[(Y_t - \underline{\mu})(Y_{t+k} - \underline{\mu})' \right] = \Gamma_k \quad \text{para toda } t.$$

La última propiedad implica que para $k = 0$ todos los vectores Y_t tienen la misma matriz de varianza-covarianza, esto es, $E \left[(Y_t - \underline{\mu})(Y_t - \underline{\mu})' \right] = \Sigma_Y$ para toda t . Esta condición implica que las series de tiempo en consideración no deben tener tendencia, ni estacionalidad y que la varianza no cambie en el tiempo. Frecuentemente es necesario hacer transformaciones a los datos para asegurar estas propiedades.

Se puede demostrar que un proceso VAR(p) es estacionario si se cumplen las condiciones sobre la media y matriz de covarianza y el polinomio definido por el determinante

$$\text{Det} (I - \Pi_1 z - \Pi_2 z^2 - \dots - \Pi_p z^p),$$

tiene todas sus raíces fuera del círculo complejo unitario. Por ejemplo, para un proceso bivariante VAR(1)

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \pi_{11} \\ \pi_{21} \end{pmatrix} + \begin{pmatrix} \pi_{12,1} & \pi_{13,1} \\ \pi_{22,1} & \pi_{23,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

el determinante será:

$$\begin{aligned} \text{Det} \left[\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} \pi_{12,1} & \pi_{13,1} \\ \pi_{22,1} & \pi_{23,1} \end{pmatrix} z \right] &= \text{Det} \begin{bmatrix} 1 - z\pi_{12,1} & z\pi_{13,1} \\ z\pi_{22,1} & 1 - z\pi_{23,1} \end{bmatrix} \\ &= (1 - z\pi_{12,1})(1 - z\pi_{23,1}) - \pi_{22,1}\pi_{23,1}z^2 \\ &= 1 - (\pi_{12,1} + \pi_{23,1})z + (\pi_{12,1}\pi_{23,1} - \pi_{22,1}\pi_{13,1})z^2 \end{aligned}$$

que es un polinomio de segundo grado.

La propiedad de estacionariedad garantiza para un proceso VAR(1) que $\lim_{n \rightarrow \infty} \Pi_1^n = 0$. Esta propiedad es justo la condición de estabilidad que se usa en el análisis de un sistema de ecuaciones simultáneas.

2.2.3. Estimación y Especificación de un Proceso VAR.

Estimación de un Proceso VAR con Orden p Conocido.

Para estimar los parámetros se considera la k -ésima ecuación del sistema

$$Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \cdots + \Pi_p Y_{t-p} + \underline{\epsilon}_t,$$

que es

$$y_{kt} = \pi_{kt} + \pi_{k2,1} y_{1,t-1} + \cdots + \pi_{kK+1,1} y_{K,t-1} + \cdots + \pi_{k2,p} y_{1,t-p} + \cdots + \pi_{kK+1,p} y_{K,t-p} + \epsilon_{kt}.$$

Se asume que se tienen N observaciones y p valores premuestreados para cada una de las variables, se obtiene el vector

$$Y^k = \begin{pmatrix} y_{k1} \\ y_{k2} \\ \vdots \\ y_{kN} \end{pmatrix}, \quad Y_{-i}^k = \begin{pmatrix} y_{k,1-i} \\ y_{k,2-i} \\ \vdots \\ y_{k,N-i} \end{pmatrix}$$

para $i = 1, \dots, p$ y $k = 1, \dots, K$. En otras palabras Y_{-i}^k contiene las variables del vector Y^k rezagadas i períodos. También se define $\underline{\epsilon}^k = (\epsilon_{k1}, \dots, \epsilon_{kN})'$. Usando esta notación se escribe Y^k como sigue:

$$Y^k = \pi_{k1} \underline{J} + \pi_{k2,1} Y_{-1}^1 + \cdots + \pi_{kK+1,1} Y_{-1}^K + \cdots + \pi_{k2,p} Y_{-p}^1 + \cdots + \pi_{kK+1,p} Y_{-p}^K + \underline{\epsilon}^k,$$

donde \underline{J} es un vector de $N \times 1$ de unos. En forma compacta el sistema se puede escribir como:

$$Y^k = X \Theta_k + \underline{\epsilon}^k,$$

donde

$$\begin{aligned} X &= (\underline{J}, Y_{-1}^1, \dots, Y_{-1}^K, Y_{-2}^1, \dots, Y_{-2}^K, \dots, Y_{-p}^1, \dots, Y_{-p}^K) \\ \Theta_k &= (\pi_{k1}, \pi_{k2,1}, \dots, \pi_{kK+1,1}, \pi_{k2,2}, \dots, \pi_{kK+1,2}, \dots, \pi_{k2,p}, \dots, \pi_{kK+1,p})' \end{aligned}$$

el segundo vector es el vector de coeficientes de la k -ésima ecuación del sistema. Hay que notar que cada una de las K ecuaciones tienen la misma matriz de regresión X . Escribiendo las K ecuaciones como un sistema, se llega a

$$Y = (I_k \otimes X) \Theta + \underline{\epsilon},$$

donde \otimes denota el producto de Kronecker³. Bajo el supuesto de que $\underline{\epsilon}$ es un vector ruido blanco, la matriz de varianza-covarianza de $\underline{\epsilon}$ es $E(\underline{\epsilon}\underline{\epsilon}') = \Sigma_\epsilon \otimes I_N$.

En tales sistemas los estimadores obtenidos por mínimos cuadrados generalizados son idénticos a los estimadores mínimos cuadrados ordinarios. Esto es equivalente a estimar cada ecuación por separado por MCO. Así, sin perder eficiencia en la estimación, se puede estimar cada ecuación por MCO

$$\hat{\Theta}_k = (X'X)^{-1}X'Y^k.$$

Para el sistema completo de estimadores se obtienen

$$\hat{\Theta} = [I_k \otimes (X'X)^{-1}X']Y.$$

Para investigar las propiedades de estos estimadores se asume que las $\underline{\epsilon}_t$ tienen una distribución normal multivariada $N(0, \Sigma_\epsilon)$ y que $\underline{\epsilon}_t$ es independiente de $\underline{\epsilon}_s$ para $t \neq s$. Si, además, Y_t es un proceso estacionario puede ser visto como:

$$\text{p-} \lim_{N \rightarrow \infty} \frac{1}{N} (X'X) = R,$$

es un matriz no singular y

$$\text{p-} \lim_{N \rightarrow \infty} \frac{1}{N} (X'\underline{\epsilon}^k) = 0 \quad k = 1, \dots, K.$$

Entonces,

$$\begin{aligned} \text{p-} \lim \hat{\Theta}_k &= \text{p-} \lim [(X'X)^{-1}X'Y^k] \\ &= \text{p-} \lim [(X'X)^{-1}X'(X\Theta_k + \epsilon^k)] \\ &= \frac{N}{N} \text{p-} \lim [(X'X)^{-1} \{X'X\Theta_k + X'\epsilon^k\}] \\ &= \text{p-} \lim \left[\Theta_k + \left(\frac{X'X}{N} \right)^{-1} \frac{X'\epsilon^k}{N} \right] \\ &= \text{p-} \lim \Theta_k + \text{p-} \lim \left(\frac{X'X}{N} \right)^{-1} \text{p-} \lim \left(\frac{X'\epsilon^k}{N} \right) \\ &= \Theta_k \end{aligned}$$

³El producto de Kronecker o producto directo de una matriz \mathbf{A} de $(M \otimes N)$ y una \mathbf{B} de $(K \otimes L)$ se define como la matriz de $(MK \otimes NL)$

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1N}B \\ a_{21}B & a_{22}B & \cdots & a_{2N}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}B & a_{M2}B & \cdots & a_{MN}B \end{bmatrix}$$

⁴p- l ım Denota el l ımite en probabilidad

por lo tanto cada $\hat{\theta}_k$ es consistente y se sigue que

$$p\text{-}\lim_{N \rightarrow \infty} \hat{\Theta} = \Theta.$$

Además, se asume que los errores del proceso ϵ_t tienen distribución normal, $\hat{\Theta}$ es asintóticamente equivalente al estimador máximo verosímil y por lo tanto asintóticamente eficiente y normalmente distribuido,

$$\sqrt{N}(\hat{\Theta} - \Theta) \xrightarrow{D} N(0, \Sigma_{\hat{\Theta}}).$$

La matriz de varianza-covarianza de la distribución asintótica puede ser vista como

$$\Sigma_{\hat{\Theta}} = \Sigma_{\epsilon} \otimes Q^{-1}.$$

Para estimar esta matriz consistentemente, es necesario un estimador de Σ_{ϵ} que sea consistente y teniendo en cuenta que $p\text{-}\lim_{N \rightarrow \infty} \frac{X'X}{N} = R$, y usando $\left(\frac{X'X}{N}\right)^{-1}$ como un estimador consistente de R^{-1} . Se puede estimar el elemento σ_{ij} de Σ_{ϵ} como:

$$\hat{\sigma}_{ij} = \frac{\left(Y^i - X\hat{\Theta}_i\right)' \left(Y^j - X\hat{\Theta}_j\right)}{N - Kp - 1},$$

donde en el denominador el número de parámetros $Kp+1$ es restado del tamaño muestral N , en cada ecuación. Denotando por $\hat{\Sigma}_{\epsilon}$, la matriz de elementos $\hat{\sigma}_{ij}$, un estimador consistente de $\Sigma_{\hat{\Theta}}$ es

$$\hat{\Sigma}_{\hat{\Theta}} = \hat{\Sigma}_{\epsilon} \otimes \left(\frac{X'X}{N}\right)^{-1}.$$

Hay que notar que éste es un estimador de la matriz de covarianza de $\sqrt{N}(\hat{\Theta} - \Theta)$. Así, como en el caso de los regresores no estocásticos, una aproximación para la matriz de covarianzas de $\hat{\Theta}$ es

$$\hat{\Sigma}_{\epsilon} \otimes (X'X)^{-1}.$$

Hay algunos problemas obvios con la especificación. Uno es que una elección arbitraria del orden del proceso no es satisfactoria. Como resultado, hay muchos parámetros estimados con error estándar muy pequeño por lo cual no son significativamente distintos de cero. Entonces, se sospecha que el modelo contiene algunos parámetros innecesarios. Si tal es el caso, los parámetros estimados así como las predicciones obtenidas para el modelo serán ineficientes.

Selección del Orden para un Proceso VAR.

En el caso univariante se basa la elección de un orden adecuado para un proceso AR en la FACP. Para el caso de los vectores autorregresivos la correlación parcial es una matriz,

elegir el orden del proceso a través de la inspección visual de esta matriz requiere de mucha experiencia. Aún que en este caso la prueba formal es más fácil de usar para decidir la significancia del modelo. Los criterios de Akaike y Schwarz son generalizados para elegir el orden de un VAR.

En el contexto de los vectores autorregresivos los criterios se definen como sigue:

$$AIC(n) = \ln \left[\text{Det} \left(\tilde{\Sigma}_n \right) \right] + \frac{2K^2n}{N}$$

$$BIC(n) = \ln \left[\text{Det} \left(\tilde{\Sigma}_n \right) \right] + \frac{nK^2 \ln(N)}{N}$$

donde K es el número de variables en el sistema, N es el tamaño de la muestra y $\tilde{\Sigma}_n$ es una estimación de la matriz de varianzas-covarianzas de los residuales $\tilde{\Sigma}_\epsilon$ obtenida con un modelo VAR(n). Los elementos de $\tilde{\Sigma}_n$ son calculados como sigue

$$\tilde{\sigma}_{ij} = \frac{\left(Y^i - X\hat{\Theta}_i \right)' \left(Y^j - X\hat{\Theta}_j \right)}{N}$$

la suma de los cuadrados o los productos cruzados es dividida entre el tamaño de muestra y no entre los grados de libertad. Es elegido el modelo de orden p tal que su AIC o BIC sea mínimo.

2.2.4. Predicción con un Proceso VAR.

Hasta el momento se ha permitido asumir que el proceso generador del conjunto de variables es un proceso estocástico conocido. En este caso el pronóstico óptimo es la esperanza condicional dada toda la información hasta el periodo en el que el pronóstico es hecho. Por predicción óptima se entiende que en cada variable el error cuadrático medio (ECM) es minimizado. Si el proceso generador es un VAR(p) conocido del tipo $Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t$ con vector de errores ruido blanco ϵ_t independientes, la esperanza condicional $Y_T(h)$ de Y_{T+h} dado $\Omega_t = \{Y_T, Y_{T-1}, \dots\}$ es fácil de determinar. Y se denota por:

$$Y_T(h) = E(Y_{T+h} | \Omega_T) = \Pi_0 + \Pi_1 E(Y_{T+h-1} | \Omega_T) + \dots + \Pi_p E(Y_{T+h-p} | \Omega_T)$$

$$= \Pi_0 + \Pi_1 Y_T(h-1) + \dots + \Pi_p Y_T(h-p)$$

donde $Y_T(h-i) = Y_{T+h-i}$ para $i \geq h$ y $E(\epsilon_{T+h} | \Omega_T) = 0$. De este modo se puede calcular recursivamente el pronóstico h pasos adelante para $h = 1, 2, \dots$. Por ejemplo, si $p = 1$, de tal modo que Y_t es un VAR(1), se tiene

$$Y_T(1) = \Pi_0 + \Pi_1 Y_T$$

$$Y_T(2) = \Pi_0 + \Pi_1 Y_T(1) = (I + \Pi_1) \Pi_0 + \Pi_1^2 Y_T$$

$$Y_T(3) = \Pi_0 + \Pi_1 Y_T(2) = (I + \Pi_1 + \Pi_1^2) \Pi_0 + \Pi_1^3 Y_T$$

y así sucesivamente.

La matriz de error cuadrático medio (ECM) a menudo es usada como una medida de la incertidumbre de predicción. La matriz de ECM h pasos adelante se denota por $\Sigma(h)$, que es,

$$\Sigma(h) = E \left([Y_{T+h} - Y_T(h)] [Y_{T+h} - Y_T(h)]' \right).$$

La predicción $Y_T(h)$ es insesgada si $E[Y_{T+h} - Y_T(h)] = 0$, $\Sigma(h)$ es la matriz de covarianzas de los errores de predicción. Puede mostrarse que la matriz de ECM del VAR(p) ($Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t$) tiene la forma:

$$\Sigma(h) = \Sigma_\epsilon + M_1 \Sigma_\epsilon M_1' + \dots + M_{h-1} \Sigma_\epsilon M_{h-1}' = \Sigma(h-1) + M_{h-1} \Sigma_\epsilon M_{h-1}',$$

donde M_i pueden ser calculadas usando recursivamente las Π_i

$$M_0 = I \text{ y } \sum_{j=1}^{\min(p, i)} \Pi_j M_{i-j} \quad i = 1, 2, \dots$$

Así,

$$\begin{aligned} M_1 &= \Pi_1 \\ M_2 &= \Pi_1 M_1 + \Pi_2 M_0 = \Pi_1^2 + \Pi_2 \\ M_3 &= \Pi_1 M_2 + \Pi_2 M_1 + \Pi_3 M_0 = \Pi_1^3 + \Pi_1 \Pi_2 + \Pi_2 \Pi_1 + \Pi_3, \end{aligned}$$

y así sucesivamente. Para un VAR(1) es fácil ver que $M_i = \Pi_1^i$.

Si se considera que un proceso VAR tiene una distribución conocida, se puede derivar intervalos de probabilidad para la predicción y establecer intervalos de predicción. Por ejemplo, si se considera un proceso VAR con distribución normal, los errores de predicción tendrán también distribución normal.

$$Y_{T+h} - Y_T(h) \sim N(0, \Sigma(h)).$$

Entonces

$$[Y_{T+h} - Y_T(h)]' \Sigma(h)^{-1} [Y_{T+h} - Y_T(h)] \sim \chi_k^2.$$

Este resultado es usado para obtener intervalos de predicción para el vector Y_{T+h} .

Un intervalo de predicción para un sólo componente de Y_{T+h} puede obtenerse de

$$[Y_{T+h} - Y_T(h)]' \Sigma(h)^{-1} [Y_{T+h} - Y_T(h)] \sim \chi_k^2,$$

lo que implica que

$$\frac{y_{k,T+h} - y_{k,T}(h)}{\sigma_k(h)} \sim N(0, 1)$$

donde $y_{k,T+h}$ y $y_{k,T}(h)$ son las k -ésimas componentes de Y_{T+h} y $Y_T(h)$, respectivamente y $\sigma_k(h)$ es la desviación estándar del correspondiente error de predicción. La $\sigma_k(h)$ es la raíz cuadrada del k -ésimo elemento de la diagonal de $\Sigma(h)$.

Si se denota por $z_{(\alpha)}$ al cuantil $(1 - \alpha)$ de la distribución normal estándar

$$\begin{aligned} 1 - \alpha &= P\left(-z_{(\alpha/2)} \leq \frac{y_{k,T+h} - y_{k,T}(h)}{\sigma_k(h)} \leq z_{(\alpha/2)}\right) \\ &= P\left(y_{k,T}(h) - z_{(\alpha/2)}\sigma_k(h) \leq y_{k,T+h} \leq y_{k,T}(h) + z_{(\alpha/2)}\sigma_k(h)\right). \end{aligned}$$

Entonces el intervalo de predicción al $(1 - \alpha)100\%$ para la k -ésima variable del sistema es

$$y_{m,T}(h) \pm z_{(\alpha/2)}\sigma_m(h),$$

o bien

$$\left[y_{m,T}(h) - z_{(\alpha/2)}\sigma_m(h), y_{m,T}(h) + z_{(\alpha/2)}\sigma_m(h) \right].$$

Si se calcula repetidamente un número grande de veces intervalos de predicción de este tipo su interpretación es que el $(1 - \alpha)100\%$ de las veces que se calcule el verdadero valor de Y_{T+h} se encontrará en el intervalo.

2.2.5. Extensión de los Vectores Autorregresivos.

Hasta ahora se han considerado procesos de vectores autorregresivos donde, a priori, todas las variables son vistas como endógenas. Estos modelos son apropiados para pronosticar pero limitan el alcance del análisis de la relación entre todas las variables involucradas.

En algunas situaciones se puede preferir establecer una forma estructural con variables exógenas. De hecho el que todas las variables, a priori, sean vistas como endógenas no es necesariamente un defecto de los modelos VAR, sin embargo, la información de la exogeneidad de las variables casi nunca está disponible pero es controversial. En cada situación se debe realizar una prueba de exogeneidad. Esta prueba puede establecerse en la estructura del proceso VAR. De hecho, la prueba de causalidad en el sentido de Granger puede ser vista como una prueba de exogeneidad. Por ejemplo, en un sistema VAR(1) se fallará al intentar rechazar la restricción $\pi_{22,1} = 0$. Con esta restricción el sistema:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \pi_{11} \\ \pi_{21} \end{pmatrix} + \begin{pmatrix} \pi_{12,1} & \pi_{13,1} \\ \pi_{22,1} & \pi_{23,1} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix}$$

puede escribirse como:

$$\begin{aligned} y_{1t} &= \pi_{11} + \pi_{12,1}y_{1,t-1} + \pi_{13,1}y_{2,t-1} + \epsilon_{1t} \\ y_{2t} &= \pi_{21} + \pi_{23,1}y_{2,t-1} + \epsilon_{2t}. \end{aligned}$$

Asumiendo independencia de las ϵ_{1t} y ϵ_{2t} la variable y_2 puede ser vista como exógena en este sistema bivalente.

Si se sabe a priori que una serie de variables X_t es exógena, esta variable puede fácilmente incorporarse en un modelo VAR(p) para escribirse, por ejemplo,

$$Y_t = BX_t + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t,$$

donde B es la matriz de coeficiente de las variables exógenas y el término de la intersección también está incluido en X_t . De una forma más general, los rezagos de la variable endógena pueden entrar en el modelo además de las variables corrientes X_t . A los modelos de este tipo se les llama modelos ARX o sistemas de ecuaciones dinámicas simultáneas.

Otra extensión del modelo VAR(p) básico $Y_t = \Pi_0 + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t$ se obtiene si se permite que los errores estén correlacionados intertemporalmente, esto es, ϵ_t puede estar correlacionado con ϵ_s para $t \neq s$. Considérese el caso univariante donde se encontró que los errores de los modelos AR eran generados por un proceso de medias móviles de orden finito. Tal proceso puede ser definido en el presente caso multivariante, proceso VAR, con un proceso de errores MA que es llamado proceso de vectores autorregresivos de medias móviles (VARMA) en analogía con el caso univariante. Estos modelos pueden involucrar menos parámetros que los VAR puros y por lo tanto puede proporcionar pronósticos más eficientes.

En las secciones previas sea concentrado el análisis en procesos estacionarios donde la media, la varianza y la autocovarianza son constantes a través del tiempo. El asumir esto excluye, por ejemplo, a variables con tendencia. Como ya se vio algunas veces la estacionariedad puede lograrse con alguna transformación tal como tomar diferencias o logaritmos. Estas transformaciones son en algunos casos también útiles en sistemas multivariantes. En otras palabras, en algunas situaciones será de interés analizar la relación que hay entre la variable original y la transformada porque la transformación puede distorsionar o eliminar la relación original.

2.3. Cointegración.

Los métodos econométricos tradicionales ignoran la no estacionariedad de las variables. La no estacionariedad invalida, en general, los resultados relativos a la distribución de los estimadores y crea problemas tales como el de las *regresiones espurias*.

Si en lugar de trabajar con las variables expresadas en niveles, se hallan regresiones con las primeras diferencias de las variables, la posibilidad de obtener resultados espurios no desaparece, pero disminuye sensiblemente, al tiempo que se reducen los problemas derivados de la multicolinealidad. Estos problemas condujeron a representar los modelos dinámicos

en la forma denominada *mecanismo de corrección de error*.

Con la metodología Box-Jenkins, consolidada al inicio de la década de los años setenta, antes de proceder a la modelización de series de tiempo se les transforma en otras estacionarias. Las predicciones así obtenidas suelen superar a las proporcionadas por los métodos econométricos tradicionales, motivo por el cual el análisis de series de tiempo se llegó a considerar por algunos especialistas como una alternativa, más rigurosa y precisa, a los métodos econométricos. Sin embargo, hoy se puede afirmar que se ha producido una síntesis entre el análisis de las series de tiempo y los métodos econométricos. El análisis de series de tiempo es una herramienta básica de la econometría, aunque la econometría de las series de tiempo es algo más que el análisis de series de tiempo. Una ventaja de los modelos econométricos sobre el análisis de series de tiempo consiste en la posibilidad de verificar la existencia de relaciones a largo plazo y utilizarlas para representar y estimar las relaciones dinámicas a corto plazo. Este tipo de análisis se conoce con el nombre de *cointegración* y consiste el objeto fundamental de la presente sección.

2.3.1. Regresiones Espurias

El problema de las *regresiones espurias* aparece frecuentemente cuando se halla la regresión estática entre series afectadas por tendencias comunes, lo que lleva a encontrar un valor de R^2 elevado, sin que exista realmente una relación de causa-efecto. Cuando se lleva a cabo una regresión espuria, además de un elevado R^2 , suele aparecer un valor pequeño del estadístico, d , Durbin-Watson, indicativo de que los errores de la ecuación están correlacionados positivamente. Esto implica no sólo que los estimadores minimocuadráticos de los coeficientes son ineficientes, sino que los estimadores de los errores estándar son inconsistentes, lo que lleva a incurrir en serios problemas de especificación.

El problema de las regresiones espurias fue analizado por Granger y Newbold (1974) en un celebrado trabajo en donde examinan los resultados de una serie de simulaciones realizadas con regresiones experimentales en las que siempre aparece una serie “explicada” en función de un conjunto de hasta cinco series. Todas las series son de alguno de estos tipos: (a) paseos aleatorios, (b) ARIMA(0, 0, 1) y (c) primeras diferencias (incrementos) de ARIMA(0, 0, 1). De este estudio se pueden extraer las conclusiones siguientes:

1. Cuando todas las variables se expresan en niveles y la serie “explicada” no depende de los regresores (la hipótesis H_0 de independencia es cierta), existen muchas posibilidades de rechazar H_0 . Así, por ejemplo, para un número de regresores iguales o superiores a 3, H_0 se rechaza erróneamente en más del 90 % de los casos.
2. Cuando las variables básicas son paseos aleatorios y figuran en primeras diferencias, el procedimiento usual de verificar la significancia de los coeficientes funciona adecuadamente en la mayoría de los casos, pero cuando se relacionan primeras diferencias

de procesos ARIMA(0, 0, 1), la hipótesis nula se rechaza con una frecuencia del orden del doble de la correcta.

Se puede concluir que cuando se llevan a cabo regresiones entre series de tiempo expresadas en niveles es muy probable encontrar relaciones espurias. Un valor elevado de R^2 (normalmente acompañado de un estadístico de Durbin-Watson bajo) no es indicativo de exista la relación estimada. Cuando se hallan regresiones con las variables expresadas en diferencias la posibilidad de extraer conclusiones erróneas no desaparece, pero disminuye.

Phillips (1986) ha proporcionado una explicación teórica de los resultados obtenidos por Granger y Newbold.

Los resultados espurios no son exclusivos de las regresiones entre series de tiempo independientes. Cuando Y_t y X_t son dos series (no estacionarias) no independientes se tiende también a estimar relaciones en niveles con R^2 más alto y un estadístico d pequeño.

Todos estos resultados podrían llevar a la conclusión de que en las aplicaciones se debe trabajar siempre con las series expresadas en diferencias a fin de reducir los resultados espurios en lo posible y que, incluso de esta forma, no hay garantía de que las pruebas ordinarias funcionen correctamente. Afortunadamente no es así. Hay situaciones un tanto especiales que, es posible detectar, en las que se pueden obtener relaciones correctas entre las variables, tanto a corto como a largo plazo.

2.3.2. Mecanismo de Corrección de Error (MCE)

El *mecanismo de corrección de error* (MCE) consiste en una representación de los modelos dinámicos que se han venido usando en economía aplicada desde las versiones tempranas de Sargan (1964) y Phillips (1957). El interés actual por el MCE en las aplicaciones se desprende del trabajo de Davidson, Hendry, Srba y Yeo (1978), que marca toda una etapa para la práctica de la econometría. El uso de los modelos de MCE se consolida con otros estudios realizados por los seguidores de la escuela de la *London School of Economics*. El MCE combina la presencia de los niveles de las variables, que recogen las relaciones a largo plazo sugeridas por la teoría económica, junto con las diferencias de dichas variables, que captan los desajustes existentes en el corto plazo. De esta forma los modelos MCE permiten modelar tanto las relaciones a largo plazo como la dinámica a corto de las variables. La denominación de MCE se debe a la especificación del modelo en la cual las desviaciones de la relación del largo plazo entre los niveles de las variables funcionan como un “mecanismo” que impulsa a los cambios de las variables a acercarse a su nivel de equilibrio cuando se han alejado de éste. Es decir, se corrigen los errores de equilibrio de periodos anteriores de forma gradual.

Para introducir la representación MCE se considera el sencillo modelo dinámico deter-

minista

$$y_t = \alpha + \omega x_t + \alpha_1 y_{t-1}, \quad |\alpha_1| < 1.$$

La dinámica implica que incluso aunque x permanezca fija, y_t va a seguir cambiando si se parte de una situación inicial de desequilibrio. Este “cambio dinámico característico” o componente de propagación (CP), se obtiene incrementando y admitiendo que x permanece fija:

$$CP = (\nabla y)_{x=cte.} = \alpha_1 \nabla y_{t-1}.$$

Restando y_{t-1} en el segundo miembro y sumando la cantidad equivalente $\alpha + \omega x_{t-1} + \alpha_1 y_{t-2}$ se tiene

$$CP = \alpha + \omega x_{t-1} + (\alpha_1 - 1)y_{t-1},$$

o bien

$$CP = -(1 - \alpha_1) \left(y_{t-1} - \frac{\alpha}{1 - \alpha_1} - \frac{\omega}{1 - \alpha_1} x_{t-1} \right).$$

Dadas unas condiciones exógenas, x_{t-1} , en el período $t - 1$, el correspondiente valor a largo plazo de y , y_{t-1}^* , vendrá dado por la siguiente expresión:

$$y_{t-1}^* = \frac{\alpha}{1 - \alpha_1} + \frac{\omega}{1 - \alpha_1} x_{t-1},$$

que se obtiene, haciendo $y_t = y_{t-1} = y_t^*$ en $y_t = \alpha + \omega x_t + \alpha_1 y_{t-1}$, despejando y_t^* y retardando la ecuación un período.

Así, el segundo factor en $CP = -(1 - \alpha_1) \left(y_{t-1} - \frac{\alpha}{1 - \alpha_1} - \frac{\omega}{1 - \alpha_1} x_{t-1} \right)$ representa la divergencia entre el valor real de la y en el período $t - 1$ y el valor a largo plazo correspondiente a este período, de acuerdo con las condiciones exógenas.

Si se admite que x cambia entre los períodos $t - 1$ y t . En este caso, el cambio total que experimentará y se compondrá de las dos componentes aditivas siguientes:

$$\nabla y_t = \omega \nabla x_t + \alpha_1 \nabla y_{t-1}.$$

Al primer sumando se le puede denominar “componente de impulso”. El segundo es la “componente de propagación”, CP , ya analizada. La primera componente representa el cambio de y inducido por los desequilibrios de corto plazo. Sustituyendo esta última se tiene la representación de corrección de error

$$\nabla y_t = \omega \nabla x_t - (1 - \alpha_1) \left(y_{t-1} - \frac{\alpha}{1 - \alpha_1} - \frac{\omega}{1 - \alpha_1} x_{t-1} \right).$$

La diferencia que figura entre paréntesis se puede contemplar como un “error”, pues es la diferencia entre y_{t-1} y su valor de equilibrio a largo plazo y_{t-1}^* . Cuanto mayor sea este “error”, mayor será el cambio que, en sentido inverso, experimente y (obsérvese el signo).

El término que se ha denominado componente de propagación, CP , opera, pues, como un mecanismo automático de corrección de las desviaciones actuales respecto a la situación de equilibrio a largo plazo, lo que justifica el nombre de corrección de error para esta representación dinámica.

Una ventaja de operar con expresiones de este tipo, en donde las variables vienen expresadas en diferencias, en lugar de hacerlo en niveles, es que se reducen considerablemente los problemas derivados de la colinealidad. En los modelos dinámicos más generales también existen representaciones de corrección de error, como se verá a continuación, en cuyo caso los problemas de multicolinealidad, cuando las variables se expresan en niveles, suelen ser importantes.

Para generalizar la representación MCE, se considerará el modelo dinámico

$$A(B)y_t = \alpha + \Omega(B)x_t + u_t,$$

en donde las raíces de $A(B) = 0$ caen fuera del círculo de radio unitario, como condición de estacionariedad. Este modelo puede escribirse así:

$$y_t = \alpha + \Omega(B)x_t + [1 - A(B)]y_t + u_t.$$

$A(B)$ y $\Omega(B)$ son los siguientes polinomios en el operador de rezago, B , que no poseen factores comunes:

$$\begin{aligned} A(B) &= 1 - \alpha_1 B - \alpha_2 B^2 - \dots - \alpha_n B^n \\ \Omega(B) &= \omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_m B^m. \end{aligned}$$

Operando se obtienen las siguientes identidades algebraicas

$$\begin{aligned} \Omega(B)x_t &= \omega_0 x_t + \omega_1 x_{t-1} + \dots + \omega_m x_{t-m} \\ &= \omega_0 (x_t - x_{t-1}) - (\omega_2 + \omega_3 + \dots + \omega_m) (x_{t-1} - x_{t-2}) \\ &\quad - (\omega_3 + \omega_4 + \dots + \omega_m) (x_{t-2} - x_{t-3}) - \dots \\ &\quad - \omega_m (x_{t-m+1} - x_{t-m}) + (\omega_0 + \omega_1 + \dots + \omega_m) x_{t-1} \\ &= \omega_0 \nabla x_t - \sum_{j=1}^{m-1} \left(\sum_{k=j+1}^m \omega_k \right) \nabla x_{t-j} + \Omega(1)x_{t-1}. \end{aligned}$$

Análogamente,

$$[1 - A(B)]y_t = [1 - A(1)]y_{t-1} - \sum_{j=1}^{n-1} \left(\sum_{k=j+1}^n \alpha_k \right) \nabla y_{t-j}.$$

Con esto el modelo dinámico puede escribirse en la forma siguiente

$$y_t = \alpha + [1 - A(1)]y_{t-1} - \sum_{j=1}^{n-1} \left(\sum_{k=j+1}^n \alpha_k \right) \nabla y_{t-j} + \omega_0 \nabla x_t \\ - \sum_{j=1}^{m-1} \left(\sum_{k=j+1}^m \omega_k \right) \nabla x_{t-j} + \Omega(1)x_{t-1} + u_t.$$

Restando ahora y_{t-1} en los dos miembros se tiene

$$\nabla y_t = - \sum_{j=1}^{n-1} \left(\sum_{k=j+1}^n \alpha_k \right) \nabla y_{t-j} + \omega_0 \nabla x_t - \sum_{j=1}^{m-1} \left(\sum_{k=j+1}^m \omega_k \right) \nabla x_{t-j} \\ - A(1) \left[y_{t-1} - \frac{\alpha}{A(1)} - \frac{\Omega(1)}{A(1)} x_{t-1} \right] + u_t$$

en donde $\frac{\Omega(1)}{A(1)}$ es el multiplicador total. La expresión de arriba es la forma general para del modelo MCE para el caso de dos variables. Y puede generalizarse para m variables, se dirá que un vector $Y_t(m \times 1)$ admite una representación MCE si se puede expresar como:

$$A(B)\nabla Y_t = \underline{\mu} - \Pi Y_{t-1} + \epsilon_t,$$

donde ϵ_t es una perturbación multivariada estacionaria; $A(B)$ es una matriz $(m \times m)$ polinómica en el operador de rezago que cumple: $A(0) = I_m$ y que $A(1)$ tiene todos los elementos finitos; y, finalmente $\Pi \neq 0$.

Sin pérdida de generalidad para el análisis e interpretación de MCE se reducirá al caso de dos variables $I(1)$. En este caso, el MCE vendría dado por:

$$\nabla y_t = \mu_1 + \phi_1(B)\nabla y_{t-1} + \Omega_1(B)\nabla x_{t-1} - \gamma_1 [y_{t-1} - \alpha x_{t-1}] + \epsilon_{1t} \\ \nabla x_t = \mu_2 + \phi_2(B)\nabla y_{t-1} + \Omega_2(B)\nabla x_{t-1} - \gamma_2 [y_{t-1} - \alpha x_{t-1}] + \epsilon_{2t}$$

donde se imponen las siguientes restricciones:

1. el vector $(1, \alpha)'$ es el mismo en ambas ecuaciones;
2. los polinomios $\phi_i(B)$ y $\Omega_i(B)$ no tienen raíces en el círculo unitario;
3. al menos uno de los parámetros $\gamma_i (i = 1, 2)$, conocidos como parámetros de velocidad de ajuste, no es nulo.

Se comprueba como el MCE conjuga la modelación dinámica a corto plazo con la relación de equilibrio a largo: los términos entre corchetes, con las variables en niveles, recogen

la relación a largo plazo entre las variables (relación de cointegración). Este término es el “corrector del error” en el sentido que será distinto de cero únicamente cuando haya alejamiento de la situación de equilibrio, produciendo, en el siguiente periodo, un ajuste hacia dicha relación. Si, por ejemplo, en el momento t se da que $y_t - \alpha x_t < 0$, es decir, que y_t está por debajo de la relación de equilibrio que mantiene respecto a x_t , entonces el término de corrección de error provocará un aumento superior de ∇y_{t+1} a fin de corregir la brecha en la relación de equilibrio. Así, los γ_i reciben el nombre de “parámetros de velocidad del ajuste”, ya que cuanto mayor sea su valor más rápidamente se corregirán los desequilibrios.

El resto de variables incluidas en el MCE son las primeras diferencias y sus rezagos, si son necesarios, de ambas variables. Son, por tanto, variables estacionarias y explican la evolución a corto plazo, es decir, los cambios de un periodo respecto al anterior. Debe tenerse en cuenta que en la relación dinámica a corto plazo pueden incluirse otras variables estacionarias, pero nunca una que no lo sea ya que, debido a sus características, será imposible que contribuya a explicar la evolución de una estacionaria. Dichas variables deben ser exógenas y afectarán a corto plazo pero no a largo (al no estar en la relación de cointegración).

Si la relación a largo plazo implica a m variables, el MCE presentará el mismo número de ecuaciones. Además podría haber más de un vector de Cointegración, por lo que en cada ecuación del MCE habría más de un término de corrección de error (uno para cada vector). Ello complica la estimación del MCE. Por ejemplo, cualquier intento de estimación directa de una ecuación del MCE presentará problemas de multicolinealidad perfecta, ya que las mismas variables en niveles aparecerán en cada vector de cointegración. Si sólo se incluye un término de corrección de error (un vector de cointegración) la estimación de éste será una combinación lineal de todas las relaciones de cointegración existentes.

La tercera restricción que se impone sobre un MCE es que al menos uno de los parámetros de velocidad de ajuste tiene que ser no nulo. Si en una ecuación aparece el término de corrección de error ello significa que la variable dependiente de la misma viene causada por el resto de variables de la relación de cointegración. Así se puede determinar el sentido de causalidad entre las variables: si, por ejemplo, en el modelo ningún γ_i es nulo, entonces y_t y x_t se causan mutuamente; si, en cambio, γ_2 es nulo, x_t causa a y_t pero y_t no causa a x_t .

Por otro lado, en caso de doble causalidad, la existencia de restricciones lineales cruzadas provoca que ninguna de ambas variables pueda considerarse exógena débil para la inferencia sobre los parámetros de la otra. Por ello, la estimación eficiente del MCE requiere procedimientos de información perfecta.

Además si se presentan problemas de simultaneidad entre las variables dependientes del MCE, se requerirá que dicha expresión se estime utilizando variables instrumentales con

procedimientos de estimación simultánea (por ejemplo mínimo cuadrados bivariados o trivariados). La simultaneidad puede estar provocada por incluirse las primeras diferencias sin retardar de las variables, o bien por la correlación contemporánea de los términos de perturbación.

Además, la estimación de un modelo VAR, $A(B)y_t = \epsilon_t$, sin introducir las restricciones dadas por las relaciones de cointegración existentes, no serán eficientes y presentará problemas de inestabilidad al analizar la correspondiente función de respuesta al impulso.

2.3.3. Cointegración

Se ha visto que las regresiones realizadas entre series temporales expresadas en niveles pueden conducir a resultados espurios. Por otra parte, los modelos de mecanismo de corrección de error se han presentado como una forma de representación de las relaciones dinámicas que trata de mitigar este problema, al tiempo que reduce los efectos de la multicolinealidad.

El análisis de cointegración permite, entre otras cosas, detectar si existe la posibilidad de obtener estimaciones correctas, es decir, libres de resultados espurios, de los parámetros que definen las relaciones entre dos o más series, tanto corto como a largo plazo. Además, en los casos en que ello sea posible, de acuerdo con la naturaleza de las series en cuestión, el análisis de la cointegración demuestra que los modelos MCE representan correctamente el comportamiento dinámico de las series y, por tanto, constituye una base adecuada para el análisis empírico.

La integrabilidad es una propiedad dominante, ello significa que la suma o combinación lineal de procesos de distintos ordenes de integración es del mismo orden que el proceso de orden mayor. Es decir, si:

$$z_t = x_t + \alpha y_t \quad \text{con } x_t \sim I(e) \quad y_t \sim I(d)$$

entonces: $z_t \sim I(\max(e, d))$. En términos similares, la combinación lineal de dos procesos con el mismo orden de integración es, en general, de ese orden de integración:⁵

$$z_t = y_t + \alpha x_t \quad x_t, y_t \sim I(d) \quad \text{entonces } z_t \sim I(d).$$

Es decir, la serie resultante presentará una tendencia en varianza resultado de la combinación de las que presentaban las variables originales. La excepción a este caso general es lo que se denomina cointegración.

⁵Debe tenerse en cuenta que la suma de infinitos términos de un mismo orden de integración es de un orden superior. Por ejemplo, se ha visto cómo el paseo aleatorio sin deriva es la suma de todos los choques aleatorios pasados, los cuales son $I(0)$.

Definición

El concepto de cointegración se debe a Engle y Granger (1987). Supóngase dos variables no estacionarias y_t y x_t , entre las que se cree que existe una relación de equilibrio estable a largo plazo. Cabe esperar que, bajo tal supuesto, los residuos de la regresión que explica a y_t en función de x_t sean estacionarios, a pesar de que ninguna de las dos variables del modelo lo sean. Ello sería debido a que su evolución temporal es, en gran medida, común. Ésta es la idea de Cointegración. Y puede ser definido de la siguiente forma:

Sea Y_t un vector de $m \times 1$, las componentes de Y_t están cointegradas de orden d , b y se denota por $Y_t \sim CI(d, b)$, si:

- todas las componentes de Y_t son integradas del mismo orden d , $I(d)$,
- existe un vector α , no nulo, tal que

$$\alpha'Y_t = z_t \sim I(d - b), \quad \text{con } b > 0$$

La relación

$$z_t = \alpha'Y_t,$$

se denomina *relación de cointegración* y el vector α *vector de cointegración*.

Si la relación de equilibrio a largo plazo entre y y x que sugiere la teoría es cierta se puede expresar mediante una relación lineal, se tiene

$$y_t^* = \alpha + \beta x_t,$$

donde y_t^* es el valor de equilibrio a largo plazo de y , correspondiente al valor actual de x . De acuerdo con ello, hay equilibrio en el período t si

$$y_t - y_t^* = 0,$$

es decir, si

$$y_t - (\alpha + \beta x_t) = 0.$$

Ahora bien, como y_t será, en general, distinto del valor de equilibrio, se puede escribir

$$y_t - (\alpha + \beta x_t) = u_t,$$

en donde u_t puede interpretarse como el “error” o desviación entre y_t y su correspondiente valor de equilibrio en el período t . Esta ecuación es equivalente a

$$y_t - \beta x_t = \alpha + u_t,$$

que es, por definición, la relación de cointegración. Definiendo

$$Y_t' = [y_t \quad x_t]$$

$$\alpha' = [1 \quad -\beta] \quad y \quad z_t = \alpha + u_t.$$

Obsérvese que cuando, como en este ejemplo, se conoce cual es el orden de causalidad, la relación de cointegración suele expresarse despejando la variable dependiente. Esto implica que el vector de cointegración α' aparezca normalizado, es decir que el coeficiente de la variable dependiente sea igual a la unidad. En el caso de que α exista no será único. Basta multiplicar el vector por un escalar no nulo para obtener un nuevo vector de cointegración. No obstante, el número de vectores de cointegración linealmente independientes que puede haber entre m variables integradas del mismo orden es $m - 1$. Así, si se normaliza el vector, habrá, como se ha señalado, como máximo $m - 1$ vectores de cointegración distintos. Al número de vectores de cointegración linealmente independientes se le denomina rango de cointegración.

Cuando y_t y $x_t \sim CI(1, 1)$, z_t es estacionaria, por definición, por lo que el “error”, u_t , es también estacionario. Esto significa que las desviaciones de la situación de equilibrio no tienden, por término medio, a ampliarse con el tiempo. Este hecho es consistente con la existencia de una relación de equilibrio a largo plazo entre x y y . Por el contrario, si y_t y x_t son ambas $I(1)$ pero no cointegran, es decir, z_t no es estacionaria, las dos variables se apartarán cada vez más una de la otra con el tiempo.

Se examinará ahora otros aspectos de la definición de cointegración que pueden resultar de interés.

En primer, lugar para la econometría aplicada, el caso más interesante de cointegración es cuando $d = b$, es decir cuando $z_t \sim I(0)$, ya que entonces es cuando se puede identificar los parámetros del vector de cointegración con los coeficientes de una relación a largo plazo entre las variables y aplicar el análisis de la regresión. En este caso es en el que se centrará el análisis.

La importancia estadística del concepto de cointegración estriba en que cuando las variables no estacionarias que aparecen a ambos lados de un modelo de regresión están cointegradas, entonces la estimación de MCO continúa teniendo buenas propiedades. Si, por ejemplo, las variables y_t y x_t son $CI(1, 1)$, entonces al estimar por MCO el modelo

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1, 2, \dots, T$$

la estimación de β coincidirá con la constante de cointegración y los residuos serán estacionarios. El estimador MCO es consistente a pesar de ser las dos variables no estacionarias; más aún, el estimador MCO resulta ser superconsistente, puesto que converge al verdadero

valor de los coeficientes α y β a una tasa $\frac{1}{N}$, en vez de a la tasa habitual⁶, $\frac{1}{\sqrt{N}}$.

La superconsistencia del estimador MCO en la regresión estática se debe a que, para valores del coeficiente β distintos de la constante de cointegración a , el residuo no es estacionario, y su varianza tiende a infinito mientras que, para $\beta = a$, su varianza es finita. Como el estimador MCO “busca” en el espacio paramétrico tratando de minimizar la varianza residual, no es sorprendente que escoja el valor a para el coeficiente β con más facilidad que si las variables no estuviesen integradas.

Por otra parte, dos series no pueden ser cointegradas si no son integradas del mismo orden. Así, por ejemplo, si $y_t \sim I(1)$ y $x_t \sim I(0)$, entonces $z_t \sim I(1)$ y las variables y_t y x_t no son cointegradas. Cuando se relacionan dos series cada una integrada de orden cero, no tiene interés hablar de cointegración.

Cuando se consideran más de dos series de tiempo la situación se puede complicar ya que, al contrario de lo que la definición anterior parece implicar, puede que exista cointegración sin que todas las variables sean integradas del mismo orden. Por ejemplo, supóngase que $y_t \sim I(1)$, $x_t \sim I(2)$ y $v_t \sim I(2)$. Si $[x_t, v_t] \sim CI(2, 1)$, entonces, existirá una relación lineal entre ambas,

$$w_t = \beta_1 x_t + \beta_2 v_t,$$

tal que $w_t \sim I(1)$. En este caso w_t podría cointegrar con y_t , ya que ambas poseen el mismo orden de integración. Si esto es así, se podría escribir

$$y_t = \alpha + \beta w_t + u_t,$$

en donde $u_t \sim I(0)$. Esta situación sólo puede darse cuando x_t y v_t son cointegradas y, además, su orden de integración es superior al de y_t .

Lo que sucede al combinar linealmente las variables del vector Y_t es la cancelación de las componentes no estacionarias de éste, dando lugar a una variable estacionaria. Por ejemplo, para el caso de dos variables, supongamos que las podemos descomponer como:⁷

$$\begin{aligned} x_t &= w_t + \tilde{x}_t & w_t &\sim I(1) \\ & & \tilde{x}_t &\sim I(0) \\ y_t &= \alpha w_t + \tilde{y}_t & \tilde{y}_t &\sim I(0) \end{aligned}$$

⁶Ello quiere decir que la diferencia $\hat{\beta} - \beta$ converge asintóticamente a una normal, incluso cuando se le multiplica por T , lo que no aceptan otros resultados sobre distribuciones límite. Por ejemplo, en el contexto habitual de regresión con variables estacionarias se tiene que $\sqrt{T}(\hat{\beta} - \beta)$ converge a una normal, por lo que $T(\hat{\beta} - \beta)$ no tiene una distribución asintótica bien definida.

⁷Para la descomposición de proceso integrados en una parte $I(1)$ y otra $I(0)$, véase Beveridge y Nelson (1981).

por lo que:

$$y_t - \alpha x_t = \tilde{y}_t - \alpha \tilde{x}_t = z_t \sim I(0),$$

donde w_i es la tendencia estocástica común, y α es el parámetro de cointegración —parámetro que anula la tendencia estocástica en la combinación lineal de las variables $I(1)$ —.

A partir de la anterior descomposición es fácil comprobar que el vector de cointegración normalizado es único en el caso de dos variables: cualquier otro produce series no estacionarias. En el ejemplo anterior si el vector es $(1, \alpha)'$ y se considera $(1, \beta)'$:

$$\begin{aligned} z_t &= y_t - \beta x_t = \alpha w_t + \tilde{y}_t - \beta w_t - \beta \tilde{x}_t \\ &= (\alpha - \beta)w_t + \tilde{y}_t - \beta \tilde{x}_t, \end{aligned}$$

entonces $z_t \sim I(0)$ únicamente si $\beta = \alpha$.

En general, si entre m variables hay r relaciones de cointegración ello significa que hay $m - r$ tendencias estocásticas comunes. Por ejemplo, si $m = 3$ y $r = 1$ un sólo vector de cointegración cancela dos tendencias estocásticas comunes. Si sólo hubiera una tendencia estocástica común bastarían dos variables para obtener una relación de cointegración, y entre las tres variables habría como máximo dos vectores de cointegración linealmente independientes.

Ligado a lo anterior, si dos variables y_t y x_t , ambas $I(1)$ están cointegradas, y una de ellas, por ejemplo x_t , lo está con otra tercera, p_t , entonces y_t y p_t , también lo estarán:

$$\begin{aligned} y_t &= \alpha_1 x_t + z_{1,t} & x_t, y_t &\sim I(1) \\ & & p_t &\sim I(1) \\ x_t &= \alpha_2 p_t + z_{2,t} & z_{1,t}, z_{2,t} &\sim I(0) \end{aligned}$$

entonces

$$\begin{aligned} y_t &= \alpha_1 \alpha_2 p_t + \alpha_1 z_{2,t} + z_{1,t} = \alpha_3 p_t + z_{3,t} \\ \text{con } z_{1,t} + \alpha_1 z_{2,t} &= z_{3,t} \sim I(0) \text{ y } \alpha_3 = \alpha_1 \alpha_2. \end{aligned}$$

Además si y_t y x_t cointegran y_{t-i} y x_t también lo hacen (pues y_t cointegra con y_{t-i}). Si existe una relación de cointegración entre los niveles de las variables se puede esperar que también la haya entre los logaritmos naturales de las misma. En cambio, la situación inversa no se cumple.

Por otro lado, si y_t y x_t son $CI(1, 1)$ entonces, o bien y_t causa a x_t , o bien la inversa, x_t causa a y_t , en el sentido de causalidad de Granger, pudiendo haber causalidad en ambas direcciones.

Cuando los datos presentan estacionalidad, el concepto de cointegración es más complicado. En este caso no, existe un sólo orden de integración, pues, junto a la posibilidad de integración ordinaria (a frecuencia cero), es posible que se presente algún tipo de integración estacional (es decir, a frecuencia de 1, 2 o más ciclos por año), dependiendo de la naturaleza y la periodicidad de los datos. Entonces, dos series y_t y x_t pueden no ser cointegradas por alguno de estos motivos: (a) porque sus órdenes de integración a la frecuencia cero sean diferentes; (b) porque difieran sus órdenes de integración a las frecuencias estacionales; (c) porque teniendo los mismos órdenes de integración a las correspondientes frecuencias, no exista un vector de cointegración.

2.3.4. Teorema de Representación de Granger

La relación formal entre el mecanismo de corrección de errores y las relaciones de cointegración la establece el **Teorema de Representación de Granger**. Este teorema prueba que un sistema de variables cointegradas puede representarse de tres formas diferentes: vectores autorregresivos (VAR), mecanismo de corrección de errores, y medias móviles multivariadas. El teorema establece el isomorfismo entre las representaciones.

El Teorema de Representación de Granger, en una versión simple, dice que:

- Si las N componentes de un vector de series de tiempo, X_t , son $CI(1, 1)$ de rango de cointegración r , existe una representación Mecanismo de Corrección de Errores válido para representar el Proceso Generador de Datos.
- Si el PGD de un conjunto de variables admite una representación MCE, ésta está cointegrada.

Esta conexión entre cointegración y MCE puede entenderse intuitivamente considerando el caso sencillo de que existan dos series $x_t, y_t \sim CI(1, 1)$. En este caso, z_t representará, los errores o desviaciones del equilibrio. De acuerdo con los MCE los errores de equilibrio de un período operan como mecanismo corrector del equilibrio en el período siguiente, desviando a la variable dependiente en el sentido opuesto al error en una cantidad proporcional al mismo. Dado que las variables cointegradas, por definición, no se desvían por término medio de su nivel de equilibrio durante largos períodos de tiempo, este hecho pone en evidencia que el MCE está presente en este caso, pues de lo contrario las series evolucionarían libremente y, por ser no estacionarias, tenderían a divergir cada vez más entre sí.

En la práctica esto significa que la evidencia de residuos estacionarios en la regresión estática entre x_t e y_t (o entre y_t y x_t) implica la existencia de un MCE. La representación dinámica en forma de MCE queda así justificada y puede utilizarse para estimar los parámetros relacionados con la dinámica a corto plazo sin temor a resultados espurios.

El análisis puede ser extendido para incluir componentes estacionales. Además el teorema puede ser extendido véase Hylleberg y Mizon (1989) para incorporar varias representaciones adicionales de sistemas cointegrados. Entre éstos están el de Bewley y el de representación de tendencias comunes (el último debido a Stock y Watson 1988).

Engle y Granger en 1987 establecieron un isomorfismo entre cointegración y mecanismo de corrección de errores. Con el propósito de examinar cointegración en un sistema de ecuaciones, se derivaron los resultados, formulando el sistema en la forma de MCE, comenzando esta vez por la representación de medias móviles del proceso. De este sistema se obtendrá el número de relaciones de cointegración, basándose en el método propuesto por Johansen en 1988, será el estimador máximo verosímil de r . Esto permitirá realizar una prueba de hipótesis concerniente a la dimensión del espacio de cointegración, y establecer un “valor central” de α

2.3.5. Estimación y Contraste de Relaciones de Cointegración

Debe tenerse en cuenta que, dada la relación mostrada entre cointegración y modelos de MCE, las estimaciones de relaciones de cointegración que no consideren todo el MCE (toda la información) no serán eficientes. Este caso es tratado al analizar la primera etapa del procedimiento bietápico de Engle y Granger (1987). En dicho procedimiento, para contrastar las relaciones de cointegración y para la estimación del MCE, se estima directamente la relación de cointegración y, posteriormente, se modeliza el MCE. Ésta ha sido la vía tradicional de estimación y contraste de relaciones de cointegración. Esta estrategia presenta problemas de eficiencia en la primera etapa al no considerar toda la información disponible.

Otra vía de modelizar y contrastar series cointegradas sería estimar directamente el MCE. No obstante, esta alternativa no es simple, ya que la expresión

$$\begin{aligned}\nabla y_t &= \mu_1 + \phi_1(B)\nabla y_{t-1} + \Omega_1(B)\nabla x_{t-1} + \gamma_1[y_{t-1} - \alpha x_{t-1}] + \epsilon_{1t} \\ \nabla x_t &= \mu_2 + \phi_2(B)\nabla y_{t-1} + \Omega_2(B)\nabla x_{t-1} + \gamma_2[y_{t-1} - \alpha x_{t-1}] + \epsilon_{2t}\end{aligned}$$

es un sistema de ecuaciones con restricciones sobre el vector de cointegración, el cual aparece en todas las ecuaciones. Además, puede haber más de un vector si las variables implicadas son más de dos. Por tanto, su estimación no es sencilla y mucho menos cuando hay implicadas muchas variables y más de un vector de cointegración. Una excepción se da cuando todas las variables excepto una de las implicadas en la relación de largo plazo son exógenas débiles.

Una forma eficiente de estimar el vector de cointegración sin necesidad de estimar todo el MCE, es la denominada estimación dinámica por mínimos cuadrados ordinarios. En caso de que haya más de un vector de cointegración y las variables no sean débilmente exógenas

un procedimiento útil es el procedimiento máximo verosímil con información completa de Johansen.

Estimación Bietápica de Engle y Granger

Engle y Granger en 1987 proponen la estimación en dos pasos de los modelos que involucran variables cointegradas. El procedimiento consiste en estimar en un primer paso la relación de cointegración realizando la regresión estática de las variables en niveles. En el segundo, paso se estima el MCE introduciendo los residuos de la relación de cointegración estimada, desfasados un periodo. Ambos pasos sólo requieren de MCO, y puede mostrarse que los resultados son consistentes para todos los parámetros. En particular, los estimadores de las parámetros en el primer paso convergen en probabilidad a una tasa T mientras que los elementos del vector de los términos de corrección de error, en el segundo paso, convergen asintóticamente a la tasa usual de \sqrt{T} .

Este procedimiento es conveniente porque la dinámica necesita ser especificada hasta que el mecanismo de corrección de error debe ser estimado (si bien esto puede ser sensible, como se verá más adelante). Se puede ilustrar esto usando un modelo simple sin intersección al origen.

Una importante implicación de la teoría de series x_t integradas de orden uno es que la varianza de ∇x_t es asintóticamente insignificante con relación a la de x_t . Suponiendo que alguna relación dinámica ligue las series $x_t, y_t \sim I(1)$, y que estas dos series sean cointegradas. Considérese la regresión estática de y_t sobre x_t ,

$$y_t = \alpha x_t + v_t.$$

Ahora v_t contiene toda la dinámica omitida, pero esta puede ser reparametrizada en términos de $\nabla x_{t-j}, \nabla y_{t-m}$ y $(y_{t-r} - \alpha x_{t-r})$, para $j, m, r > 0$, que son todas $I(0)$ si existe la cointegración. Así, α es estimada consistentemente por la regresión a pesar de la omisión de toda la dinámica. De hecho,

$$\begin{aligned} \hat{\alpha} &= \left(\sum_{t=1}^T x_t y_t \right) \left(\sum_{t=1}^T x_t^2 \right)^{-1} \\ &= \alpha + \left(\sum_{t=1}^T x_t v_t \right) \left(\sum_{t=1}^T x_t^2 \right)^{-1} \end{aligned}$$

$\{v_t\}$ es $I(0)$ bajo la cointegración pero $\{x_t\}$ es $I(1)$.

A medida que t tiende a infinito, $\sum x_t^2$ tiende a infinito y $\hat{\alpha}$ tiende a α independientemente de $\sum x_t v_t$, que se ve superado por el crecimiento de $\sum x_t^2$, a una tasa de T y no a la

tasa usual de \sqrt{T} . Converge rápido asintóticamente y esto es usado por Engle y Granger como base de su estimación bietápica. Igualmente para el caso inverso:

$$\begin{aligned}x_t &= \frac{1}{\alpha} y_t + v'_t \\ \hat{\alpha}' &= \frac{1}{\alpha} + \frac{\sum y_t v'_t}{\sum y_t^2}.\end{aligned}$$

La misma conclusión anterior se puede extender a las variables cointegradas pero medidas con error: las estimaciones siguen siendo consistentes si los errores son $I(0)$ por lo que no influyen en la relación a largo plazo entre variable $I(1)$. Por ejemplo, si x_t se mide con error:

$$x_t^* = x_t + v_t \text{ con } x_t \sim I(1) \text{ y } v_t \sim I(0),$$

y la relación de cointegración es $y_t = \alpha x_t + z_t$ y se estima:

$$\begin{aligned}y_t &= \alpha^* x_t^* + z_t^* = \alpha^* (x_t + v_t) + z_t^* \\ &= \alpha^* x_t + z_t^* + \alpha^* v_t = \alpha^* x_t + z_t''.\end{aligned}$$

Z_t'' será estacionario, y $\hat{\alpha}^*$ tenderá al valor poblacional α que es el único que produce unos residuos estacionarios⁸.

De tal suerte que los parámetros convergen al valor poblacional a una velocidad superior (al aumentar la muestra), que las estimaciones con variables estacionarias. Este hecho se debe a que para el verdadero valor α , los residuos son estacionarios y para cualquier otro valor de α , los residuales son $I(1)$, no estacionarios y, por tanto, con varianza muy elevada. Al ser la estimación MCO la que minimiza la varianza residual, ésta encontrará “muy deprisa” este único valor que minimiza dicha varianza.

Esto es llamado teorema de superconsistencia de Stock (1987) y puede ser enunciado formalmente como sigue.

Supóngase que x_t satisface $(1 - B)x_t = C(B)e_t$ con $C(B) = C(1) + (1 - B)C^*(B)$, donde $C^*(B)$ es absolutamente sumable⁹, la perturbación tiene momento de orden cuarto finito, y x_t es $CI(1, 1)$ con r vectores de cointegración (incorporados en la matriz α) satisfaciendo únicamente,

$$\alpha' M = 0; \quad \alpha = q + Q\theta,$$

⁸No obstante, el que una de las variables se mida con error aleatorio, aunque éste sea estacionario, afectará a la estimación de los parámetros de la relación dinámica a corto plazo dada por las variables estacionarias que aparecen en el MCE.

⁹La serie infinita $\{c_j\}_1^\infty$ se dice que es absolutamente sumable si $\sum |c_j| < \infty$. Para que la matriz $C^*(B)$ sea absolutamente sumable, la condición es que $\sum \|C_j^*\| < \infty$.

entonces¹⁰

$$p\text{-lím } T^{1-\delta}(\hat{\theta} - \theta) = 0 \text{ para } \delta > 0.$$

Así, en lugar de que converja a una tasa de \sqrt{T} , como un proceso estacionario, los estimadores mínimo cuadrados convergen a una tasa de T . Este teorema y la representación por mecanismo de corrección de error de sistemas cointegrados pueden ser asociados para dar el siguiente teorema.

Teorema de Engle y Granger. La estimación bietápica de una ecuación de un sistema de corrección de error con un vector de cointegración, obtenido al tomar la estimación $\hat{\alpha}$ de α de la regresión estática en lugar del verdadero valor para estimar el MCE en la segunda etapa, tiene la misma distribución límite con el estimador máximo verosímil que usando el verdadero valor de α . El método de mínimos cuadrados en la segunda etapa proporciona estimadores consistentes del error estándar.

Bosquejo de la Demostración del Teorema de Engle y Granger

La siguiente es una prueba del teorema para el caso bivariado. Considere las estimaciones de β y γ en las dos ecuaciones dadas por

$$\begin{aligned}\nabla y_t &= \beta \nabla x_t + \gamma(y_{t-1} - \alpha x_{t-1}) + \epsilon_t, \\ \nabla y_t &= \beta \nabla x_t + \gamma(y_{t-1} - \hat{\alpha} x_{t-1}) + \epsilon_t^*.\end{aligned}$$

y_t y x_t son cointegradas y ambas $I(1)$ con el parámetro de cointegración dado por α . El MCE es estimado usando el verdadero valor del parámetro de cointegración en la primera ecuación de arriba, mientras en la segunda α es sustituida por $\hat{\alpha}$, donde $\hat{\alpha}$ es derivada de la regresión estática de y_t sobre x_t . También $\epsilon_t^* = \epsilon_t + \gamma(\hat{\alpha} - \alpha)x_{t-1}$. Sea $z_t = y_t - \alpha x_t$.

Es necesario mostrar que la distribución asintótica del estimador $\hat{\beta}$ y $\hat{\gamma}$, de β y γ respectivamente, son los mismos sin importar si se usa α o $\hat{\alpha}$.

En un modo estándar, asumiendo adecuadamente los valores iniciales, se tiene de la primera ecuación

$$T^{1/2} \begin{bmatrix} (\hat{\beta} - \beta) \\ (\hat{\gamma} - \gamma) \end{bmatrix} = \begin{bmatrix} T^{-1} \sum_{t=1}^T (\nabla x_t)^2 & T^{-1} \sum_{t=1}^T \nabla x_t z_{t-1} \\ T^{-1} \sum_{t=1}^T \nabla x_t z_{t-1} & T^{-1} \sum_{t=1}^T z_{t-1}^2 \end{bmatrix}^{-1} \begin{bmatrix} T^{-1/2} \sum_{t=1}^T \nabla x_t \epsilon_t \\ T^{-1/2} \sum_{t=1}^T z_{t-1} \epsilon_t \end{bmatrix}.$$

Los estimadores derivados de la segunda ecuación también están dados arriba pero reemplazando z_t y ϵ_t por \hat{z}_t y ϵ_t^* . De este resultado se deduce la demostración si las siguientes condiciones son ciertas:

¹⁰Los elementos de q y Q típicamente son todos ceros y unos, definiendo un coeficiente en cada columna de α como la unidad y definiendo la rotación si $r > 1$. $M = p\text{-lím } E(T^{-2} \sum x_t x_t')$.

$$(i) \text{ p-}\lim T^{-1} \sum_{t=1}^T \nabla x_t z_{t-1} = \text{p-}\lim T^{-1} \sum_{t=1}^T \nabla x_t \hat{z}_{t-1};$$

$$(ii) \text{ p-}\lim T^{-1} \sum_{t=1}^T z_{t-1}^2 = \text{p-}\lim T^{-1} \sum_{t=1}^T \hat{z}_{t-1}^2,$$

(iii) las distribuciones asintóticas de $T^{-1/2} \sum_{t=1}^T z_{t-1} \epsilon_t$ y $T^{-1/2} \sum_{t=1}^T \hat{z}_{t-1} \epsilon_t^*$ son las mismas;

(iv) las distribuciones asintóticas de $T^{-1/2} \sum_{t=1}^T \nabla x_t \epsilon_t$ y $T^{-1/2} \sum_{t=1}^T \nabla x_t \epsilon_t^*$ son las mismas;

En $\nabla y_t = \beta \nabla x_t + \gamma(y_{t-1} - \alpha x - t - 1) + \epsilon_t$ se asume que $\{\epsilon_t\}$ es un proceso de inovaciones tal que $E(\nabla x_t \epsilon_t) = 0$.

Primero hay que notar que, por las propiedades de las series $I(0)$ y $I(1)$, las siguientes expresiones no son explosivas ni degeneradas cuando T tiende a infinito, las series tienden a cero:

$$(a) T^{-2} \sum_{t=1}^T x_{t-1}^2;$$

$$(b) T^{-1} \sum_{t=1}^T x_{t-1} \hat{z}_{t-1}; \quad T^{-1} \sum_{t=1}^T x_{t-1} \nabla x_t; \quad T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_t;$$

$$(c) T(\hat{\alpha} - \alpha).$$

Segundo,

$$z_{t-1} = \hat{z}_{t-1} + (\hat{\alpha} - \alpha)x_{t-1}.$$

Usando esta ecuación,

$$\begin{aligned} \text{p-}\lim T^{-1} \sum_{t=1}^T \nabla x_t z_{t-1} &= \text{p-}\lim T^{-1} \sum_{t=1}^T \nabla x_t \hat{z}_{t-1} + \text{p-}\lim T^{-1} (\hat{\alpha} - \alpha) \sum_{t=1}^T \nabla x_t x_{t-1} \\ &= \text{p-}\lim T^{-1} \sum_{t=1}^T \nabla x_t \hat{z}_{t-1} + \text{p-}\lim T^{-1} \{ [T(\hat{\alpha} - \alpha)] T^{-1} \sum_{t=1}^T \nabla x_t x_{t-1} \}. \end{aligned}$$

El resultado (i) se sigue de (b) y (c). También

$$\begin{aligned} \text{p-}\lim T^{-1} \sum_{t=1}^T z_{t-1}^2 &= \text{p-}\lim \left(T^{-1} \sum_{t=1}^T \hat{z}_{t-1}^2 + (\hat{\alpha} - \alpha) T^{-1} \sum_{t=1}^T x_{t-1}^2 + 2(\hat{\alpha} - \alpha) T^{-1} \sum_{t=1}^T x_{t-1} \hat{z}_{t-1} \right) \\ &= \text{p-}\lim T^{-1} \sum_{t=1}^T \hat{z}_{t-1}^2 + \text{p-}\lim T^{-1} \left\{ [T(\hat{\alpha} - \alpha)]^2 T^{-2} \sum_{t=1}^T x_{t-1}^2 \right\} \\ &\quad + 2\text{p-}\lim T^{-1} \left\{ T(\hat{\alpha} - \alpha) T^{-1} \sum_{t=1}^T x_{t-1} \hat{z}_{t-1} \right\}. \end{aligned}$$

El resultado (ii) se sigue de (a), (b) y (c).

Finalmente

$$\begin{aligned}
T^{-1/2} \sum_{t=1}^T z_{t-1} \epsilon_t &= T^{-1/2} \sum_{t=1}^T \hat{z}_{t-1} \epsilon_t + T^{-1/2} [T(\hat{\alpha} - \alpha)] T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_t \\
&= T^{-1/2} \sum_{t=1}^T \hat{z}_{t-1} \epsilon_t^* - \gamma T^{-1/2} [T(\hat{\alpha} - \alpha)] T^{-1} \sum_{t=1}^T \hat{z}_{t-1} x_{t-1} \\
&\quad + T^{-1/2} [T(\hat{\alpha} - \alpha)] T^{-1} \sum_{t=1}^T x_{t-1} \epsilon_t.
\end{aligned}$$

Por (b) y (c), las últimas dos expresiones del lado derecho convergen a una tasa de $T^{1/2}$. De esto sigue el resultado (iii), y (iv) se prueba análogamente para:

$$T^{-1/2} \sum_{t=1}^T \nabla x_t \epsilon_t = T^{-1/2} \sum_{t=1}^T \nabla x_t \epsilon_t^* - \gamma T^{-1/2} [T(\hat{\alpha} - \alpha)] T^{-1} \sum_{t=1}^T \nabla x_t x_{t-1}.$$

Estimación Directa de la Relación de Cointegración

La estimación en niveles de una relación entre variables integradas, normalmente produce unos residuos que no son estacionarios, un R^2 elevado y se tiende a aceptar como significativo el parámetro asociado al regresor. Es decir, se está ante el caso de una regresión espuria entre las variables, debiendo tomar diferencias o una estructura dinámica (retardos) para realizar la inferencia correctamente. En cambio, si un conjunto de variables están cointegradas -existe una combinación lineal entre ellas de un orden de integrabilidad menor- puede realizarse la regresión mínimo cuadrática (en el sentido de no incurrir en relaciones espurias), al obtenerse unos residuos estacionarios. Esto pone de manifiesto la utilidad de la teoría de cointegración a la hora de discriminar entre relaciones espurias y relaciones reales entre variables.

Para una mayor simplicidad se revisará el caso bivariado haciendo referencia, cuando sea oportuno al caso de más variables. Así, si $y_t, x_t \sim CI(1, 1)$ y se puede plantear la regresión

$$y_t = \mu + \alpha x_t + \epsilon_t,$$

ya que para un determinado valor del parámetro de cointegración, α , los residuos obtenidos son estacionarios. La estimación mínimo cuadrática ordinaria, al minimizar la varianza residual, estimará consistentemente este único parámetro que conduce a unos residuos estacionarios.

En la expresión anterior, aunque estén involucradas variables $I(1)$, no se trata de una relación espuria ya que:

- los residuos son estacionarios para un determinado valor de α y, por tanto, el estadístico Durbin Watson (DW) será significativamente distinto de cero al no haber una raíz unitaria en ϵ_t ;
- hay una relación de causalidad entre las variables cuando se da cointegración (y no de casualidad como en el caso de las relaciones espurias).

La estimación MCO de la regresión de cointegración $y_t = \mu + \alpha x_t + \epsilon_t$ presenta las siguientes características:

1. Es una estimación *superconsistente* (Stock, 1987).
2. La consistencia de los estimadores no se ve alterada por el hecho de que los regresores de la regresión de cointegración estén correlacionados con el término de perturbación. Es más, se puede ocupar cualquiera de las variables cointegradas como variable dependiente y la estimación sigue siendo consistente. La variable utilizada como variable dependiente implica que el vector de cointegración estimado esté normalizado a que el parámetro de ésta tome el valor unitario.
3. La estimación del parámetro α es sesgada, sobretodo en muestras pequeñas. Estos sesgos en muestras pequeñas se deben a la autocorrelación que presenta ϵ_t (Phillips, 1988). Esta correlación es debida a que ϵ_t recoge toda la dinámica del MCE no especificada en la regresión de cointegración. Este sesgo no tiene una distribución normal ni media no nula, pero desaparece cuando el tamaño muestral tiende a infinito. Asimismo, el R^2 tiende a uno cuando la muestra aumenta, aunque es sesgado. Ello es debido a que, en el modelo de regresión lineal simple, el producto de las pendientes de la regresión de cada variable sobre la otra es igual al coeficiente de determinación. En el caso de variables cointegradas este producto debe ser uno. En caso contrario, ello muestra la presencia de sesgos en la estimación de la pendiente. Por ello, Banerjee et al. (1986) sugieren utilizar $(1 - R^2)$ como una medida del sesgo cometido. De este modo, si una regresión de cointegración presenta un R^2 no muy elevado, debe interpretarse con cautela la estimación obtenida.

El hecho de que el R^2 de la regresión de cointegración estimada no sea igual a cero significa que la estimación estática no es invariante a la normalización elegida, es decir, a la variable que se especifica como dependiente.

4. La estimación MCO no es completamente eficiente al no considerar el resto de información disponible, es decir, todo el MCE.
5. Los t -estadísticos de los parámetros del vector de cointegración están sesgados y son inconsistentes. Por tanto, la inferencia sobre los parámetros estimados no se puede hacer de manera tradicional. Este resultado es consecuencia de dos hechos:

- a) En la regresión estática suele haber una considerable autocorrelación residual, lo que lleva a la inconsistencia de la estimación de los errores estándar de los parámetros.
 - b) El hecho de estimar con variables integradas supone que, en general, la distribución de los parámetros estimados no sea normal. Una excepción se presenta cuando el regresor es débilmente exógeno y, por tanto, no correlacionado con ϵ_t . En este caso, la distribución asintótica de su parámetro estimado será normal.
6. Si las variables implicadas en la relación de cointegración son más de tres es de esperar una fuerte colinealidad entre los regresores. Ello es así ya que, para que haya una relación de cointegración, las variables han de evolucionar conjuntamente a largo plazo. La eliminación de una de las variables explicativas en la regresión de cointegración, a fin de reducir la multicolinealidad, conducirá a resultados inconsistentes al no poder obtener residuos estacionarios. Con ello se constata que la multicolinealidad, más que un problema, es una característica inherente a las variables cointegradas.

Por otro lado, dada la autocorrelación de la perturbación en la regresión de cointegración, podría pensarse en estimar por Cochrane-Orcutt a fin de corregir ésta. Sin embargo, como señalan Engle y Granger en 1987, dicho procedimiento producirá estimaciones inconsistentes. También afirma que la regresión, en primeras diferencias, de una variable sobre la otra tampoco es consistente (se estaría omitiendo, como variable, el término de corrección de error).

En resumen, se podría decir que la estimación MCO de la regresión de cointegración proporciona, de forma sencilla, unos parámetros superconsistentes, aunque sesgados y no eficientes, sobre los que no se puede hacer inferencia pero que permitiría contrastar si existe una raíz unitaria en los residuos estimados (que son consistentes). Una forma eficiente de estimar una relación de cointegración que permite realizar inferencia sobre los parámetros estimados es la estimación dinámica de la relación de cointegración por mínimos cuadrados ordinarios.

Contrastes de Cointegración Sobre los Residuos

Una forma sencilla de contrastar una relación de cointegración entre variables consiste en analizar si los residuos de la regresión de cointegración presentan un orden de integración menor que el de las variables involucradas. Para el caso de variables $I(1)$ el contraste consistirá en determinar si los residuos presentan una raíz unitaria (no cointegración) o, lo que es lo mismo, determinar si son o no estacionarios.

Para ello se pueden utilizar los contrastes de raíces unitarias presentadas anteriormente. Sin embargo, los valores críticos utilizados en dichos contrastes para determinar el orden de integración de una variable no son aplicables para el caso de la cointegración. Ello es

debido a que la estimación por MCO proporciona los residuos de menor varianza. Así, éstos tienden a aparecer lo más estacionarios posible aunque no haya cointegración (no sean estacionarios). Por ello, los valores críticos usuales en los contrastes de cointegración tenderán a rechazar la hipótesis de no estacionariedad de los residuos más a menudo que al nivel de significancia nominal. Es decir, los valores críticos de los contrastes de integración están sesgados a la baja (en valor absoluto) cuando se aplican a los contrastes de los residuos de cointegración. Ello hace necesario utilizar otros valores críticos para los contrastes de cointegración que se basan en los residuos de la regresión estática. Además, dichos valores dependen del número de variables integradas introducidas en la regresión estática: cuantas más variables se introducen en la regresión, más estacionarios parecerán los residuos. Por ello, los valores críticos para los contrastes de no cointegración se deben corregir al alza en función del número de variables introducidas en la regresión estática.

Se mostrarán dos contrastes propuestos en Engle y Granger, de los que existen valores críticos tabulados: el basado en el Durbin-Watson de la regresión estática y el Dickey Fuller Aumentado sobre los residuos de dicha regresión. En ambos la hipótesis nula es $H_0 : \epsilon_t \sim I(1)$ (no cointegración).

El rechazo de la existencia de una relación de cointegración entre las variables integradas utilizadas en la regresión estática deben interpretarse como que la relación estimada es de tipo espurio.

Contraste DW Sobre los Residuales de Cointegración (DWRC). El DWRC, fue sugerido por Sargan y Bhargava en 1983, se calcula de la misma forma que el estadístico Durbin-Watson y está dado por

$$DWRC = \frac{\sum_{t=2}^T (\hat{\epsilon}_t - \hat{\epsilon}_{t-1})^2}{\sum_{t=1}^T \hat{\epsilon}_t^2},$$

donde $\hat{\epsilon}_t$ denota los residuales MCO de la regresión de cointegración.

La hipótesis nula que se plantea es $H_0 : DW = 0$ (no cointegración) en la estimación de $y_t = \mu + \alpha x_t + \epsilon_t$, es decir, la existencia de una raíz unitaria en los residuos, si el estadístico DWRC es significativamente mayor que cero. Este contraste, que parece bastante potente, se basa en que $DWRC = 2(1 - \rho)$. El DWRC se compara con los valores críticos de las tablas que aparecen en Engle y Granger (1987) para el caso de dos variables. Si el DWRC estimado es inferior a los valores críticos tabulados no se podrá rechazar la H_0 . Una ventaja de este contraste es que es invariante frente a la posible inclusión de constantes y tendencias en el modelo, sin que por ello varíen sus valores críticos. El problema de este

contraste es que generalmente $\hat{\epsilon}$ sigue un esquema AR(p) mientras que el contraste DWRC considera únicamente un esquema AR(1).

Contraste Dickey Fuller Sobre los Residuales de Cointegración DFRC. El contraste del tipo DF o DFA sobre la regresión de cointegración, consiste en estimar por MCO: $\nabla \hat{\epsilon}_t = \delta \hat{\epsilon}_{t-1} + \sum_{i=1}^p \theta_i \nabla \hat{\epsilon}_{t-i} + e_t$, donde $\hat{\epsilon}_t$ son los residuos MCO de la regresión de cointegración, y el número de retardos p se escoge suficientemente grande como para que e_t sea ruido blanco.

Los retardos de los residuos diferenciados se introducen para eliminar la autocorrelación que presenten éstos. Como sugieren Phillips y Oularis, p debería aproximarse por $T^{1/3}$, para el contraste de raíces unitarias Dickey y Fuller cuando la variable sigue un proceso ARIMA($p, 1, q$) de orden desconocido. Ello es debido a que los residuos de la regresión de cointegración son combinaciones lineales de los residuos del PGD (MCE) y, por tanto, en general no seguirán procesos AR finitos pero sí ARMA.

Se plantea la hipótesis de no cointegración contrastando la $H_0 : \delta = 0$ frente a la $H_a : \delta < 0$, de cointegración. Los valores críticos están tabulados en Engle y Granger (1987) para una relación entre dos variables en una muestra de cien observaciones. Engle y Yoo en 1987 amplían los valores críticos para más de dos variables y Mackinnon aumenta con superficies de respuesta los valores críticos del contraste DFARC para cualquier tamaño de muestra y hasta seis variables, permitiendo la inclusión de una tendencia determinista.

La razón para esta disparidad es que $\hat{\epsilon}_t$ es una variable “generada” mediante un procedimiento que ha escogido $\hat{\mu}$ y $\hat{\alpha}$ de modo que se minimice la varianza residual, es decir, la varianza de $\hat{\epsilon}_t$ y, por tanto que parezca tan estacionario como sea posible. En consecuencia, si se utilizan los valores críticos de los contrastes DF o DFA para este caso, se rechazará la hipótesis nula de no estacionariedad con demasiada frecuencia. Para evitar este sesgo los valores críticos deben aumentarse ligeramente. Otro factor que influye sobre los valores críticos apropiados es el número de variables explicativas, pues una combinación lineal de una larga lista de variables explicativas no estacionarias pueden parecer estacionarias, incluso si las variables no están cointegradas.

Conviene señalar que los contrastes de cointegración sobre los residuos presentan una baja potencia ante valores de parámetro autorregresivo de los residuos cercanos, pero inferiores, a la unidad. Engle y Granger mostraron que, cuando el PGD de los errores de la regresión de cointegración es un AR(1) con el parámetro autorregresivo igual a 0.9, el poder de DWRC, DFRC y FDARC al 5 % los valores críticos son 20 %, 15 % y 11 % respectivamente.

Una vez estimada y contrastada la relación de cointegración entre un conjunto de variables, se tratará de estimar el modelo de MCE correspondiente.

Estimación del Mecanismo de Corrección del Error

Engle y Granger demostraron que una vez estimado por MCO el vector de cointegración (regresión estática), los resultantes parámetros del MCE (PGD) pueden ser estimados consistentemente introduciendo los residuos de la regresión estática rezagada un periodo $\hat{\epsilon}_{t-1}$, en el MCE. Así, en la segunda etapa del proceso de estimación y contraste de relaciones de cointegración, se estimaría el MCE introduciendo $\hat{\epsilon}_{t-1}$ en lugar del vector de cointegración. Los autores demostraron que estos estimadores bietápicos tienen como distribuciones límite las resultantes de la estimación máximo verosímil que utilice el auténtico valor del vector de cointegración. Ahora bien, bajo los supuestos de:

- a) existencia de un único vector de cointegración, y
- b) exogeneidad débil de las variables condicionantes de los modelos,

el procedimiento de estimación se simplifica. Éste consiste en añadir una nueva etapa a la estimación bietápica anterior en la que se corrigen los parámetros estimados en la primera etapa, haciéndolos asintóticamente equivalentes a los máximo verosímiles, además de proporcionar un conjunto de errores estándar que permite calcular los estadísticos t de los parámetros del vector de cointegración. El procedimiento de estimación en tres etapas de Engle y Yoo es expuesto a continuación con un ejemplo. Supóngase que el PGD:

$$\nabla y_t = \beta \nabla x_t + \gamma [y_{t-1} - \mu - \alpha x_{t-1}] + \epsilon_t$$

$$\nabla x_t = v_t \text{ con } \begin{pmatrix} \epsilon_t \\ v_t \end{pmatrix} \sim iidN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon & 0 \\ 0 & \sigma_v \end{pmatrix} \right)$$

donde x_t es exógena débil para los parámetros de la primera ecuación. Las tres etapas de este procedimiento serían:

- 1ª Es la misma que en la estimación bietápica, obteniendo los parámetros y los residuos contrastando la estacionariedad de estos últimos.
- 2ª Estimar por MCO:

$$\nabla y_t = \beta \nabla x_t + \gamma \hat{\epsilon}_{t-1} + u_t,$$

obteniendo $\hat{\gamma}$ y \hat{u}_t . Ello equivale a la segunda etapa del procedimiento de Engle y Granger.

- 3ª Estimar por MCO:

$$\hat{u} = \delta(-\hat{\gamma}x_t) + w_t,$$

y corregir el parámetro $\hat{\alpha}$ estimado en la primera etapa haciendo $\tilde{\alpha} = \hat{\alpha} + \hat{\delta}$. El error estándar de $\tilde{\alpha}$ es el de $\hat{\delta}$ por lo que ahora ya es posible la inferencia sobre el parámetro del vector de cointegración.

Procedimiento Máximo Verosímil de Johansen

El procedimiento máximo verosímil con información completa de Johansen tiene una serie de ventajas frente a los restantes métodos, como son el contrastar simultáneamente el orden de integración de las variables y la presencia de relación de cointegración entre ellas; estimar todos los vectores de cointegración, sin imponer a priori que únicamente hay uno; y no verse afectado por la endogeneidad de las variables implicadas en la relación de cointegración. Por estos motivos, así como por su implementación en algunos paquetes informáticos, se convierte en una alternativa cada vez más utilizada frente a otros métodos de estimación y contraste. No obstante, el procedimiento de Johansen también impone restricciones de factores comunes en el sistema.

El procedimiento parte de la modelación de vectores autoregresivos, sugerida por Sims en 1980, en la que todas las variables se consideran endógenas. Formalmente, sea el modelo VAR(p):

$$Y_t = \mu + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t,$$

donde Y_t es un vector columna de orden $(m \times 1)$, m es el número de variables del modelo μ es un vector de constantes, y ϵ_t es un vector de perturbaciones aleatorias idéntica e independientemente distribuidas con media nula y matriz de varianzas y covarianzas Ω .

Por otro lado, se puede escribir como:

$$\nabla Y_t = \mu + \Gamma_1 \nabla Y_{t-1} + \dots + \Gamma_{p-1} \nabla Y_{t-p+1} + \Pi Y_{t-p} + \epsilon_t,$$

donde:

$$\begin{aligned} \Gamma_i &= -I + \Pi_1 + \dots + \Pi_i \quad i = 1, \dots, p-1 \\ \Pi &= -I + \Pi_1 + \dots + \Pi_p \end{aligned}$$

y la matriz Π , de orden $(m \times m)$, contiene la información sobre la relación a largo plazo entre las variables, llamándose también matriz de impactos. La última expresión es la de un MCE en forma matricial. Debe notarse que para que la expresión esté equilibrada es necesario que ΠY_{t-p} sea $I(0)$, lo cual implica que la matriz Π recoge la relación de cointegración.

Dado el rango de Π igual con r , puede demostrarse que:

- a) Si $r = 0$, Π es una matriz nula, por lo que la expresión únicamente presentaría variables en primeras diferencias y las variables del vector Y_t serían $I(1)$. No existiría, por tanto, ninguna combinación lineal de variables no estacionarias que fuera $I(0)$. Es decir, no existiría ninguna relación de cointegración.
- b) Si $r = m$, el proceso multivariado Y_t sería estacionario. Intuitivamente, esto sería debido a que entre m variables sólo puede haber como máximo $m - 1$ vectores de

cointegración que formen una base en el espacio de cointegración. Es decir, en el espacio generado por las columnas de α , donde α es la matriz de vectores de cointegración, $\alpha'Y_t$ es $I(0)$, el rango de α es igual a $m - 1$, y todas las variables de Y_t son $I(1)$. En este caso, existiría una tendencia común en varianza entre las variables de Y_t que originaría la no estacionariedad univariante de las variables. Por tanto, habría $m - 1$ vectores de cointegración linealmente independientes que carecerían esa tendencia común. Así, Y_t será estacionario si $\Pi_{m \times m}$ tiene rango completo, ya que esta matriz recoge las relaciones (vectores) de cointegración (relaciones a largo plazo), y entre m variables sólo puede haber como máximo $m - 1$ vectores de cointegración linealmente independientes.

- c) Si $0 < r < m$, estará entre las dos situaciones anteriores, por lo que habrá r relaciones de cointegración. Desde otro punto de vista, $Y_t = \mu + \Pi_1 Y_{t-1} + \dots + \Pi_p Y_{t-p} + \epsilon_t$ se puede escribir como $\Pi(B)Y_t = \mu + \epsilon_t$. El sistema será no estacionario si $\det[\Pi(1)] = \det[I - \Pi_1 - \dots - \Pi_p] = \det[-\Pi] = 0$, lo cual implica que el rango de Π es igual a $r < m$. El rango de Π mostrará al número de columnas linealmente independientes de esta matriz (vectores de cointegración).

Por otro lado si $r > 0$, Π puede reescribirse como el producto de dos matrices de dimensión $(m \times r)$, $\Pi = \gamma\alpha'$, siendo α la matriz de vectores cointegración -por lo que $\alpha'Y_t$ sería estacionario y representaría el término de corrección de error-, y siendo γ la matriz de parámetros que mediría el ajuste de ∇Y_t respecto de los errores de desequilibrio. Es decir, recogería la velocidad del ajuste de cada vector de cointegración en cada ecuación. De esta forma queda clara la afirmación que el modelo de arriba es un MCE. Hay que destacar que al sustituir Π por $\gamma\alpha'$ se está ante una sobreparametrización del modelo. De esta manera, no se puede estimar una única α , sino que lo único que se podrá estimar es una base del espacio generado por las columnas de ésta.

Bajo el supuesto que la expresión $\nabla Y_t = \mu + \Gamma_1 \nabla Y_{t-1} + \dots + \Gamma_{p-1} \nabla Y_{t-p+1} + \Pi Y_{t-p} + \epsilon_t$ es el PGD, $\mu = 0$, $0 < r < m$, los parámetros son constantes, $Y_{t-p} \dots Y_0$ están dados y $\epsilon_t \sim IN(0, \Omega)$, la estimación máximo verosímil de Π puede obtenerse a partir de las siguientes etapas:

1. Estimar por MCO los sistemas de ecuaciones

$$\begin{aligned}\nabla Y_t &= \Gamma_{01} \nabla Y_{t-1} + \dots + \Gamma_{0p-1} \nabla Y_{t-p+1} + R_{0t} \\ Y_{t-p} &= \Gamma_{11} \nabla Y_{t-1} + \dots + \Gamma_{1p-1} \nabla Y_{t-p+1} + R_{pt}\end{aligned}$$

para obtener los residuos R_{0t} y R_{pt} .

2. Calcular los momentos de segundo orden de todos los residuos:

$$S_{ij} = \frac{\sum_{t=1}^T R_{it} R'_{jt}}{T} \quad i, j = 0, p$$

donde S_{ij} es una matriz cuadrada de $(m \times m)$

3. Johansen demuestra que la estimación máximo verosímil de la matriz de vectores de cointegración α , imponiendo la restricción de normalización $\alpha' S_{pp} \alpha = I$, se obtiene a partir del cálculo de los valores propios de $(S_{p0} S_{00}^{-1} S_{0p})$ respecto a S_{pp} ,¹¹ es decir, las λ_i , $i = 1, \dots, m$, tales que:

$$|\lambda S_{pp} - S_{p0} S_{00}^{-1} S_{0p}| = 0.$$

Una vez obtenidas las lambdas éstas se ordenarían de mayor a menor de forma que $\lambda_1 > \lambda_2 > \dots > \lambda_m$.

4. Para contrastar la hipótesis nula de que hay como máximo r vectores de cointegración frente a la alternativa de que hay m , $r \leq m$, el contraste de razón de verosimilitud viene dado por:

$$\lambda_{traza}(r) = -2 \ln Q = -T \sum_{i=r+1}^m (1 - \lambda_i),$$

que sigue una distribución que puede aproximarse por $c\chi_{(f)}^2$ donde $c = 0,85 - \frac{0,58}{f}$, y χ_f^2 es una distribución χ^2 con $f = 2(m - r)^2$ grados de libertad. Este estadístico se denomina estadístico de la traza.

Un estadístico alternativo para contrastar la significancia del r -ésimo valor propio mayor, λ_r , es:

$$\lambda_{max}(r) = -T \ln(1 - \lambda_r).$$

Los valores críticos de ambos estadísticos se encuentran en Johansen (1988) y Osterwald-Lenum (1992; tabla 0). Debe señalarse que las distribuciones de los estadísticos dependen del número de relaciones de cointegración, por lo que los valores críticos varían en función de éstas.

5. Una vez contrastado el rango de cointegración se obtendría una estimación de α . Las columnas de α serán los vectores propios asociados a cada λ_i . Así la matriz α se estima a partir de:

$$(\lambda_i S_{pp} - S_{p0} S_{00}^{-1} S_{0p}) \hat{\alpha}_i = 0 \quad i = 1, \dots, r$$

donde $\hat{\alpha}_i$ es la i -ésima columna de $\hat{\alpha}$. La obtención de una estimación consistente de γ , Π y de Ω se obtienen de:

$$\hat{\gamma} = S_{0p} \hat{\alpha} \quad \hat{\Pi} = \hat{\gamma} \hat{\alpha} \quad \hat{\Omega} = S_{00} - \hat{\gamma} \hat{\gamma}'.$$

¹¹También demuestra que el problema de maximizar de verosimilitud concentrada en Π se reduce a encontrar las correlaciones canónicas entre Y_{t-p} y ∇Y_t corregidas por los efectos de la presencia de valores retardados de ∇Y_t , es decir, las correlaciones canónicas entre R_0 y R_p . Por tanto, es un “problema de valores propios”.

La idea intuitiva que hay detrás de este procedimiento de estimación es que se deben encontrar las combinaciones lineales del vector Y_t que estén correlacionadas al máximo con ∇Y_t . La secuencia de contrastes sería empezar planteando la $H_0 : r = 0$ (no cointegración) frente una alternativa de $r = 1$, utilizando uno de los estadísticos anteriormente citados. En caso de rechazar esta hipótesis, se contrastaría la $H_0 : r = 1$ frente a la alternativa de $r = 2$, y así sucesivamente hasta el momento en que no se rechace H_0 , o bien hasta que se tuviera que aceptar la hipótesis alternativa de $r = m$ (es decir, que todas las variables son estacionarias) ¹²

Cabe señalar que hasta el momento se ha supuesto que el modelo no presenta ningún componente determinista. En caso de que éstos estén presentes, las distribuciones de los estadísticos se ven modificadas (aun que no sus expresiones) como ocurría en los contrastes de raíz unitaria.

Considerando ahora la inclusión de términos deterministas en el modelo, el vector de constantes, μ , puede entrar de dos formas distintas en el modelo: sin restringir, es decir, como una constante en cada ecuación, lo cual implica una tendencia determinista en los niveles de las variables; y restringido, de forma que sólo apareciera dentro de los mecanismos de corrección de error, es decir, bajo la hipótesis de que $\mu = \gamma\alpha'_0$, donde γ es la matriz de $(m \times r)$ de parámetros de velocidad de ajuste y α_0 es de dimensión $(r \times 1)$. En este último caso el vector Y_t no presentaría tendencias deterministas y las constantes sólo aparecerían en las relaciones a largo plazo entre las variables.

Estimación y Contraste cuando el Término Determinista entra sin Restricciones en el MCE. La estimación de este caso es prácticamente igual a la del caso en que no están presentes términos deterministas en el modelo, comentado anteriormente. Con respecto a aquél sólo cambia la primera etapa. En la misma se introduciría como regresor adicional un vector de unos con el fin de recoger el efecto de las constantes. Si en el modelo se hubiesen introducido variables ficticias estacionarias de suma anual nula, éstas se introducirían también en esta etapa. El resto de las etapas son calculadas igual, cambiando sólo los valores críticos de referencia (Johansen y Juselius (1990) u Osterwald-Lenum (1992)).

Estimación y Contraste cuando el Término Determinista entra Restringido en el MCE. En este caso la hipótesis nula consiste en que $\Pi = \gamma\alpha'$ y $\mu = \gamma\alpha'_0$ donde α_0 es un vector de $(r \times 1)$ de constantes. Es decir el vector de constantes μ es proporcional a la matriz γ . Así, se puede escribir:

$$\gamma\alpha'Y_{t-p} + \mu = \gamma\alpha'Y_{t-p} + \gamma\alpha_0 = \gamma\alpha^*Y_{t-p}^*,$$

¹²En caso de incertidumbre sobre la aceptación o rechazo de algunas de las hipótesis señaladas puede ser de utilidad analizar la representación gráfica de los residuos resultantes de las relaciones estimadas, que deben mostrar un comportamiento estacionario en caso de ser relaciones de cointegración.

donde $Y_t^* = (Y_t', 1)'$ y $\alpha^* = (\alpha', \alpha_0)'$. De esta forma, la constante tan sólo aparece en el término de corrección de error, es decir, en las relaciones de cointegración, y no implica que las variables estén afectadas por una tendencia determinista, como sucede en el caso de que el vector de constantes no aparezca restringido.

En este caso el procedimiento de estimación y contraste consistiría en las siguientes etapas:

1. Estimar por MCO los sistemas de ecuaciones

$$\begin{aligned}\nabla Y_t &= \Gamma_{01} \nabla Y_{t-1} + \cdots + \Gamma_{0p-1} \nabla Y_{t-p+1} + R_{0t}^* \\ Y_{t-p}^* &= \Gamma_{11} \nabla Y_{t-1} + \cdots + \Gamma_{1p-1} \nabla Y_{t-p+1} + R_{pt}^*\end{aligned}$$

para obtener los vectores de residuos R_{0t}^* y R_{pt}^* .

2. Calcular los momentos de segundo orden de todos los residuos:

$$S_{ij}^* = \frac{\sum_{t=1}^T R_{it}^* R_{jt}^{*'}}{T} \quad i, j = 0, p.$$

3. Resolver:

$$|\lambda^* S_{pp}^* - S_{p0}^* S_{00}^{*-1} S_{0p}^*| = 0,$$

y ordenar los valores propios de mayor a menor.

4. Contrastar la hipótesis nula de que hay como máximo r vectores de cointegración frente a la alternativa de que hay m , $r \leq m$, con:

$$\lambda_{traza}^*(r) = -2 \ln Q = -T \sum_{i=r+1}^m (1 - \lambda_i^*).$$

Los valores críticos en este caso a parecen en Johansen y Juselius (1990) y en Osterwald-Lenum (1992).

5. Como anteriormente, las columnas de α serán los vectores propios asociados a cada λ_i^* ($i = 1, \dots, r$). La obtención de una estimación consistente de γ , Π y de Ω se obtienen de las misma forma que anteriormente.

A fin de comprobar si el vector de constantes μ entra de forma restringida o no en el modelo, Johansen (1992) sugiere realizar la secuencia de contraste partiendo del supuesto de que el vector μ está restringido a aparecer sólo en el término de corrección de error. Cada vez que se rechace la H_0 se contrastaría la misma bajo el supuesto de que μ no está restringido. Si esta hipótesis también es rechazada se contrastaría un rango de cointegración mayor de nuevo bajo el supuesto de que μ está restringido, y así sucesivamente. Los valores críticos de los estadísticos de la traza y el valor propio máximo bajo ambos supuestos se encuentran

en Osterwald-Lenum (1992).

En resumen, el procedimiento máximo verosímil con información completa de Johansen permite la estimación simultánea de todos los vectores de cointegración sin imponer a priori restricciones sobre el número de vectores de cointegración ni sobre la exogeneidad de las variables.

Contraste de Cointegración del Estadístico de la Traza Multivariada de Phillips y Oularis. Phillips y Oularis analizan la distribución asintótica de varios contrastes de cointegración basados en los residuales de la regresión estática. Entre éstos se encuentra la prueba DFA sobre los residuos, así como sus correcciones semiparamétricas que son equivalentes a la sugeridas por Phillips y Perron para el contraste DF cuando hay autocorrelación residual.

Los autores demuestran que dichos contrastes no son invariantes a la variable elegida para normalizar el vector de cointegración. Es decir, que depende de la variable elegida como dependiente en la regresión estática. Por ello, proponen un contraste invariante ante la normalización, denominado estadístico de la traza multivariada. La hipótesis nula es la de cointegración y la prueba consiste en contrastar la significación de dicho estadístico con las tablas elaboradas por los propios autores.

Las etapas para construir el contraste son:

1. Dado un vector ($m \times 1$) de variables Y_t , estimar por MCO el sistema VAR(1):

$$Y_t = \pi Y_{t-1} + u_t,$$

y estimar el vector de residuos u_t .

2. Obtener una estimación consistente de matriz de varianzas y covarianzas de los residuos u_t estimados dada por:

$$\hat{\Omega} = \frac{\sum_{t=2}^T \hat{u}_t \hat{u}_t'}{T} + \frac{1}{T} \sum_{j=1}^l \left(1 - \frac{j}{1+l}\right) \sum_{t=j+1}^T (\hat{u}_t \hat{u}_{t-j}' + \hat{u}_{t-j} \hat{u}_t').$$

3. Calcular la matriz M_{yy} mediante:

$$M_{yy} = \frac{\sum_{t=1}^T Y_t Y_t'}{T}.$$

4. Obtener el estadístico de la traza multivariada que viene dado por:

$$P_y = T * \text{traza}(\hat{\Omega}M_{yy}^{-1}).$$

5. Comprobar la significancia de dicho estadístico con la tabla elaborada por Phillips y Oularis (1990). Cuando hay cointegración, dicho estadístico diverge, por lo que para altos valores del mismo se tenderá a rechazar la hipótesis de no cointegración. Los valores críticos del contraste dependen del número de variables incluidas en el vector de cointegración. Los autores tabulan valores para un máximo de seis variables.

En caso de incluir una constante en el modelo, la primera etapa se estimaría incluyendo un vector de constantes, y en la tercera etapa se calcularía la matriz de momentos M_{yy} utilizando las desviaciones de Y_t respecto a su vector de medias. Los valores críticos también fueron tabulados por Phillips y Oularis (1990).

2.3.6. Pronóstico en Sistemas Cointegrados

El pronóstico univariado de variables integradas presenta la característica de que la varianza del error de pronóstico crece con el horizonte de pronóstico. Sin embargo, siguiendo Engle y Yoo (1987), los pronósticos derivados de sistemas cointegrados tienen la propiedad de que existen combinaciones lineales de dichos pronósticos que son idénticamente nulas, independientemente del punto de origen de los pronósticos. Ello es debido a que el pronóstico a largo plazo viene dado exactamente por el vector de cointegración. Además la varianza de pronóstico de dichas combinaciones lineales se mantiene finita, mientras que para cualquier otra combinación lineal la varianza tiende a infinito a medida que el horizonte de pronóstico se incrementa.

Así por ejemplo, dada una representación por MCE con una relación de cointegración $y_t = \alpha'X_t$ la mejor predicción a largo plazo hecha en el momento T de $y_{T+h|T}$ (con h suficientemente lejano) condicionada a X_{T+h} es $y_{T+h|T} = \alpha'X_{T+h}$, la cual tendrá varianza finita aunque h tienda a infinito.

Por otro lado, mediante un experimento de simulación Monte Carlo, Engle y Yoo comparan la capacidad de pronóstico, utilizando el error cuadrático medio, de un modelo de MCE estimado por el procedimiento bietápico de Engle y Granger frente a la estimación de un modelo VAR no restringido. Los resultados obtenidos arrojan que los VAR irrestrictos son un poco mejor para pronosticar en el corto plazo (cuatro periodos). A partir de este horizonte el MCE se muestra superior al VAR. Ello se debe a la restricción que se impone sobre el MCE en el largo plazo. Por otro lado, la mejor capacidad de pronóstico del VAR no restringido en el corto plazo podría deberse a los sesgos de estimación de la primera etapa del procedimiento de Engle y Granger. Además puede suponerse que procedimientos más eficientes de estimación del MCE serán superiores, en cuanto a capacidad de pronóstico, a

los dos tratados en la simulación por Engle y Yoo.

El hecho de que la varianza de los errores del pronóstico de la combinación de cointegración se mantiene finita no resuelve el problema del pronóstico a largo plazo con variables integradas. El siguiente ejemplo ilustra la dificultad. Considere el proceso

$$x_t = \pi_0 + \pi x_{t-1} + \epsilon_t,$$

donde $|\pi| < 1$. Entonces, por repetidas sustituciones, el pronóstico de h pasos adelante del tiempo t , denotado por: $\hat{x}_{t+h|t}$, está dado por

$$\hat{x}_{t+h|t} = \pi_0(1 - \pi)^{-1}(1 - \pi^h) + \pi^h x_t.$$

Cuando h tiende a infinito, $\hat{x}_{t+h|t}$ tiende a $\pi_0(1 - \pi)^{-1}$, que es la media incondicional del proceso. Este argumento, cuando se aplica a variables estacionarias tal como $\alpha'x_t$ o ∇x_{it} (donde $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$), implica que el sistema de ecuaciones, si se reescribe enteramente en términos de variables $I(0)$, pierde la capacidad de pronosticar los valores futuros basados en los pasados. Cuando el horizonte de pronóstico aumente, el mejor pronóstico es la media incondicional. Trabajar en los niveles de variables $I(1)$ es igual de problemático, el pasado es ahora más informativo, pero el error de pronóstico tiene varianza creciente con h .

Capítulo 3

Aplicaciones

En este capítulo se desarrollan dos ejemplos el primero de ellos sobre las relaciones de algunas empresas en el mercado financiera mexicano y el otro enfocado a la interacción de tres monedas: peso mexicano, dólar estadounidense y euro. La finalidad del capítulo es mostrar la utilidad que presenta la aplicación de la teoría de *Cointegración* en la práctica, para demostrar la veracidad o la ausencia de esta de la teoría económica, aplicándola en series de datos reales.

3.1. Mercados Eficientes y Cointegración

Un mercado se considera eficiente cuando no pueden obtenerse ganancias extraordinarias con base a distintos conjuntos de información y donde, por tanto, no existen oportunidades de arbitraje (Fama, 1970 y Dwyer y Wallace, 1992). La interpretación de esta definición no es sin embargo sencilla. Por un lado, diversos estudios sobre el tipo de cambio (Alexander y Johnson, 1992, Copeland, 1991, Karfakis y Parikh, 1994, Crowder, 1994 y MacDonald y Taylor, 1989) sostienen que un mercado es eficiente en el caso donde las series no estén cointegradas¹. Ello se debe a que la presencia de cointegración implica una relación de causalidad, a través del teorema de representación de Engle y Granger (1987), que contradice la idea de que las series no pueden ser pronosticadas con el conjunto de información disponible (Granger, 1986 y Hakkio y Rush, 1989). En este contexto, se destaca como relevantes, en todo caso, las causas que originan el rechazo de la presencia de cointegración entre las series tales como una prima de riesgo, el uso ineficiente de la información (Hakkio y Rush, 1989 y Crowder, 1994) o el cambio estructural en las relaciones entre las series (Septhon y Larsen, 1991).

¹Se excluyen los casos en donde se analiza el mercado actual y el de futuros, ya que las pruebas respectivas se basan en el análisis directo de una relación de cointegración entre las series (Hakkio y Rush, 1989).

Por el contrario, autores como Dwyer y Wallace (1992), Ferré y Hall (2002), sugieren que es compatible la presencia de cointegración con un mercado eficiente. Esto es, al definir a un mercado eficiente, como aquel donde no existen oportunidades de arbitraje, entonces se abandona la equivalencia entre ineficiencia y cointegración. Así, la eficiencia no implica que no puedan predecirse las series, sino que no existan oportunidades de arbitraje (Baffes, 1994).

Ferré y Hall (1992), Dwyer y Wallace, (1992) y Fama (1991), argumentan que un mercado eficiente puede incluir la propiedad de cointegración, ya que esto no implica necesariamente que se obtengan ganancias extraordinarias, no obstante que su comportamiento pueda ser predecible. Asimismo, Dwyer y Wallace (1992) sostienen que la eficiencia no implica que los precios de los activos no son predecibles, ya que ello no tiene relación directa e inmediata con la existencia de ganancias de arbitraje, rompiendo de esta forma la equivalencia directa entre ineficiencia y cointegración. En todo caso, existen situaciones en donde la cointegración es consistente con la propiedad de eficiencia y, por ello, se requieren condiciones adicionales que implican imponer restricciones a la cointegración para analizar la eficiencia del mercado (Dwyer y Wallace, 1992).

3.2. La Eficiencia en la Formación de Precios de un Mercado

A continuación se examina la eficiencia de los mercados desde el punto de vista del uso que se hace de la información, con arreglo a la teoría de Fama (1970). En un mundo sin información perfecta, el mercado no alcanza automáticamente el precio de equilibrio. Sin embargo, esta circunstancia es compatible con una cualidad que este autor denomina “eficiencia en precios”. Un mercado se califica eficiente, aunque nunca acierte plenamente con el precio de equilibrio, si cada par de operadores cierra sus transacciones a precios que pueden prever los demás con cierta aproximación.

El marco teórico utilizado es una adaptación de la Hipótesis de Mercado Eficiente (*EMH, Efficient Market Hypothesis*), atribuida a la Escuela de Economía Financiera de Chicago. Fama convierte todos los precios pactados en el mercado en predicciones del precio de equilibrio, separadas de éste por un error que se distribuye aleatoriamente, pero mantiene el supuesto de que el precio fijado es único. Dado que en la mayoría de mercados no hay un precio “más representativo” que los demás, y, la información relativa a los precios en el mercado objeto de estudio se da mediante la publicación de un precio máximo y un precio mínimo se desarrolla una teoría *ad hoc* sobre el registro de precios distintos basada en el concepto de cointegración. Granger (1986) garantiza que si dos variables están cointegradas, debe existir un proceso de ajuste que evita que los errores con respecto al equilibrio a largo

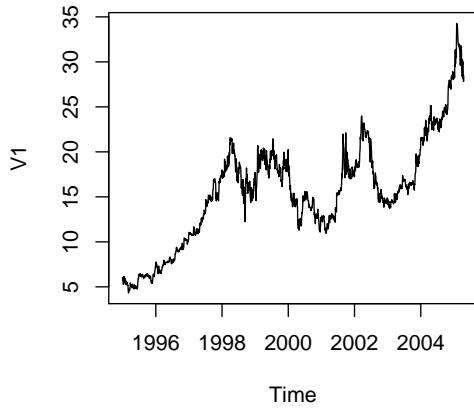
plazo sean cada vez mayores. Por tanto, en un mercado eficiente, los distintos precios a los que se cierran transacciones conforme se reciben órdenes compradoras y vendedoras deben estar cointegrados. En concreto, con el análisis de cointegración unido a la especificación de mecanismos de corrección de errores (MCE) se clasificarán nueve acciones de la Bolsa Mexicana de Valores en un contexto de precios múltiples según su eficiencia.

3.2.1. Los Datos

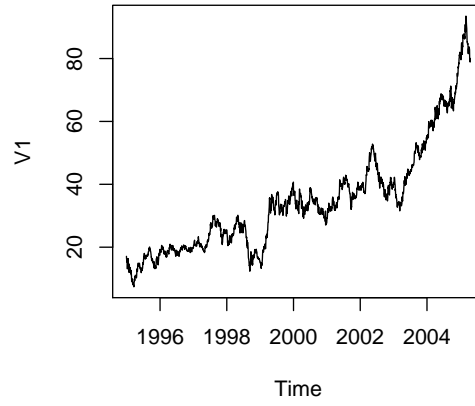
Se emplearán los precios máximos y mínimos de acciones que cotizan en la Bolsa Mexicana de Valores (BMV) así como el Índice de Precios y Cotizaciones de la BMV (IPC). La información comprende datos diarios desde el 30 de diciembre de 1994 al 21 de abril de 2005, para las acciones de: Bimbo, Cemex, Femsa, Grupo Carso, Grupo Modelo, Televisa, Telmex y Walmex; para Grupo México del 8 de febrero de 1995 al 21 de abril de 2005 y para el IPC desde el 30 de diciembre de 1994 al 1 de abril de 2005. Estas acciones fueron seleccionadas por que en el periodo analizado formaron parte de la base de acciones de referencia para el cálculo del IPC principal indicador del mercado mexicano. Las estimaciones econométricas y estadísticas se realizaron en E-views 4.1.

<i>Nombre de la Empresa</i>	<i>Clave de Cotización</i>	<i>Serie</i>	<i>Descripción</i>	<i>Fecha de Constitución</i>	<i>Fecha de Listado en la BMV</i>
Grupo Bimbo S.A. de C.V.	BIMBO	A	Acciones Industriales	15/06/1966	19/02/1980
Cemex S.A. de C.V.	CEMEX	CPO	Acciones Industriales	20/01/1931	05/01/1976
Fomento Económico Mexicano S.A. de C.V.	FEMSA	UBD	Acciones Industriales	20/10/1986	03/09/1987
Grupo Carso S.A. de C.V.	GCARSO	A1	Acciones Industriales	22/10/1980	19/06/1990
Grupo México S.A. de C.V.	GMEXICO	B	Acciones Industriales	02/09/1999	12/10/2000
Grupo Modelo S.A. de C.V.	GMODELO	C	Acciones Industriales	21/11/1991	16/02/1994
Telefonos de México S.A. de C.V.	TELMEX	L	Acciones Servicios	23/12/1920	06/02/1951
Grupo Televisa S.A.	TLEVISA	CPO	Acciones Servicios	19/12/1990	10/12/1991
Wal-Mart de México S.A. de C.V.	WALMEX	V	Acciones Comerciales	25/10/1965	25/10/1974

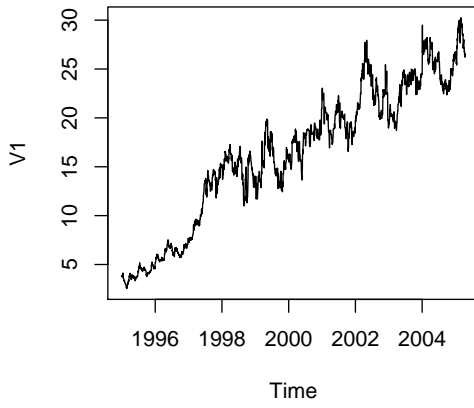
BIMBO A



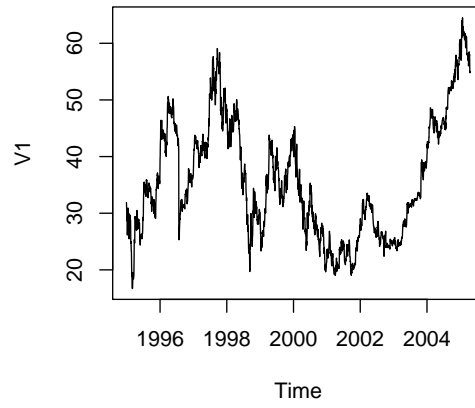
CEMEX CPO



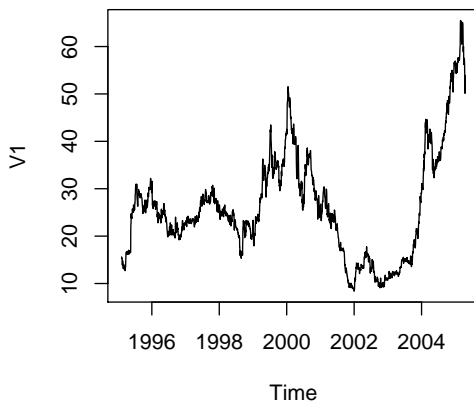
FEMSA UBD



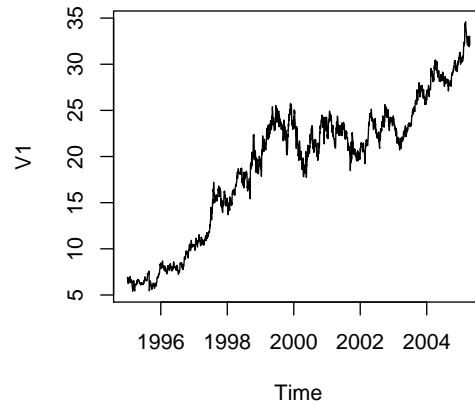
GCARSO A1

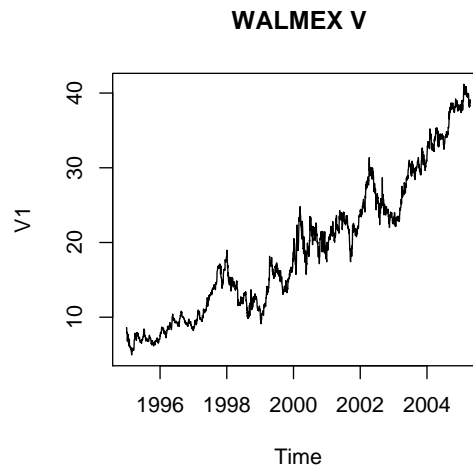
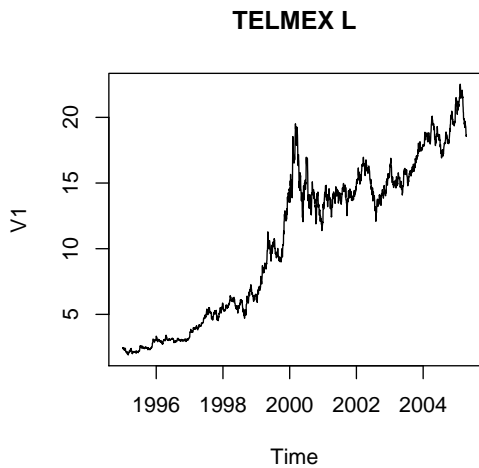
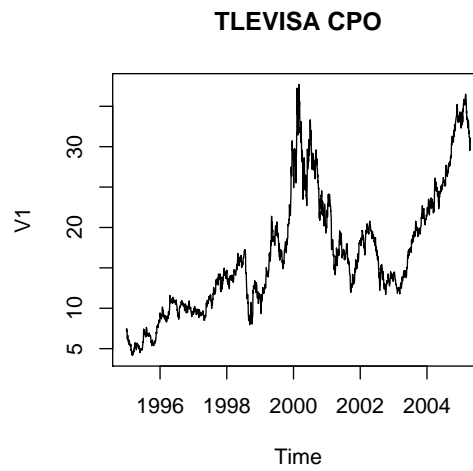
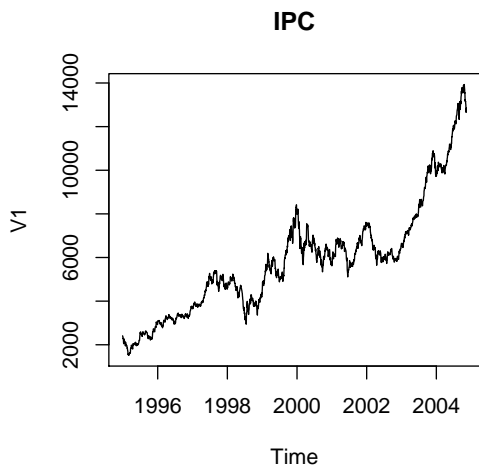


GMEXICO B



GMODELO C





Dado que en la formulación de modelos de series de tiempo se requiere el estudio previo de las propiedades de estacionariedad de las series individuales, y como las gráficas revelan la ausencia de estacionariedad, se trabajará con las series de precios transformadas en logaritmos. Según Hamilton, la transformación logarítmica (en minúsculas) mitiga la amplitud de las fluctuaciones permitiendo alcanzar más fácilmente la estacionariedad en varianza tras tomar la primera diferencia. También la primera diferencia de una variable en logaritmos es aproximadamente igual al cambio porcentual de la variable original y ello significa que si la serie en logaritmos es integrada de orden uno, la tasa de variación de la serie original es estacionaria. Por otra parte, el uso de variables en logaritmos permite interpretar como elasticidades los coeficientes estimados. Se han comprobado las propiedades de estacionariedad mediante los pertinentes contrastes de raíces unitarias.

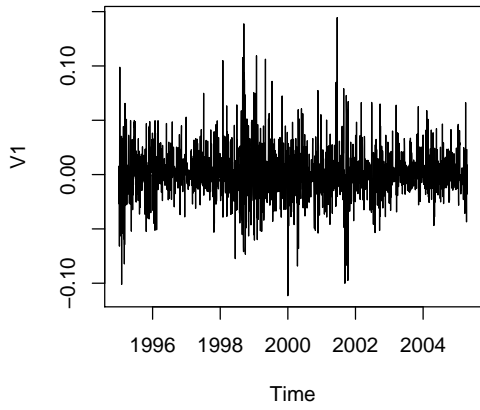
En concreto, se ha llevado a cabo la identificación del orden de integración de las variables mediante la prueba de Dickey-Fuller y la prueba de Phillips-Perron.

Pruebas de raíces unitarias				
Variable	Máximos		Mínimos	
	DFA	PP(1)	DFA	PP(1)
<i>bimbo</i>	1.1089(1)	1.3178	1.0851(2)	1.2672
$\nabla bimbo$	-47.801(0)*	-47.699*	-46.994(0)*	-46.963*
<i>cemex</i>	0.9958(1)	1.0224	0.9098(0)	0.9506
$\nabla cemex$	-45.157(0)*	-44.976*	-49.327(0)*	-49.282*
<i>femsa</i>	1.0323(1)	1.1323	1.0468(1)	1.1672
$\nabla femsa$	-47.771(0)*	-47.628*	-48.950(0)*	-48.873*
<i>gcarso</i>	0.1952(1)	0.2724	0.1622(1)	0.1919
$\nabla gcarso$	-44.445(0)*	-44.284*	-48.494(0)*	-48.387*
<i>gmexico</i>	.6494(1)	0.6781	0.6331(1)	0.6596
$\nabla gmexico$	-45.616(0)*	-45.475*	-46.991(0)*	-46.955*
<i>gmodelo</i>	1.4825(2)	1.8459	1.2329(2)	1.8031
$\nabla gmodelo$	-32.684(2)*	-49.365*	-40.150(1)*	-54.568*
<i>ipc</i>	1.8553(3)	1.8448	1.7713(2)	1.7256
∇ipc	-26.908(2)*	-39.720*	-35.549(1)*	-39.580*
<i>telmex</i>	1.5939(2)	1.6122	1.4474(2)	1.5083
$\nabla telmex$	-38.255(1)*	-43.775*	-38.526(1)*	-48.367*
<i>tlevisa</i>	0.6960(1)	0.7703	0.5877(1)	0.7106
$\nabla tlevisa$	-44.495(0)*	-44.210*	-43.954(0)*	-43.497*
<i>walmex</i>	1.1212(2)	1.1662	1.0944(2)	1.1494
$\nabla walmex$	-37.202(1)*	-46.523*	-37.455(1)*	-45.915*

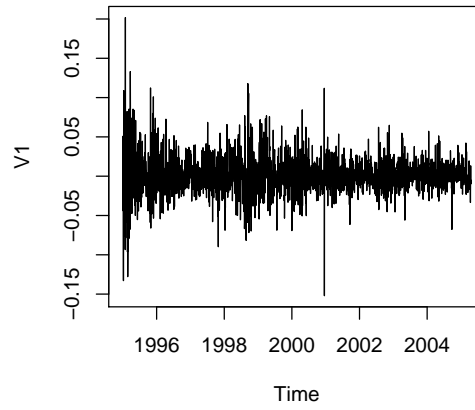
Notas: * indica el rechazo de la hipótesis nula al 1 % de significancia. El valor crítico al 1 % para la prueba Dickey-Fuller Aumentada y Phillips-Perron, de acuerdo con los valores calculados por MacKinnon, es -2.5658 considerando un modelo sin constante y sin tendencia. Los valores entre paréntesis representan el número de rezagos utilizados en la prueba.

Todas las series de precios de las acciones así como la del IPC, son integradas de orden uno, $I(1)$, al 1 % de significancia, lo que quiere decir que tras aplicar la primera diferencia las series se vuelven estacionarias, las gráficas de la primera diferencia del logaritmo de las series se muestran a continuación.

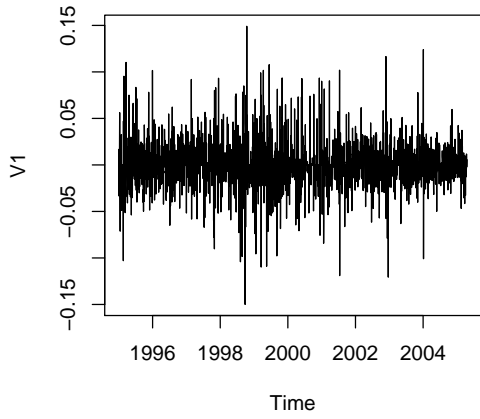
Dif bimbo a



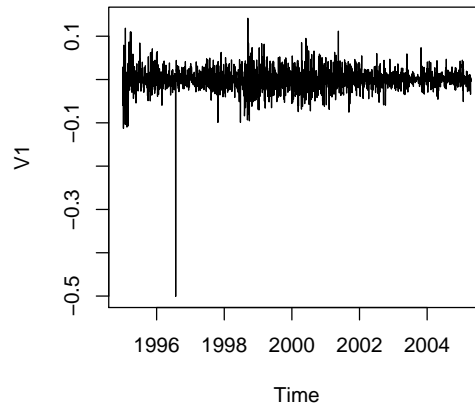
Dif cemex cpo



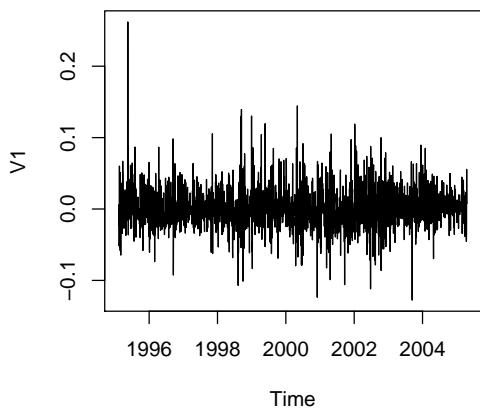
Dif femsa ubd



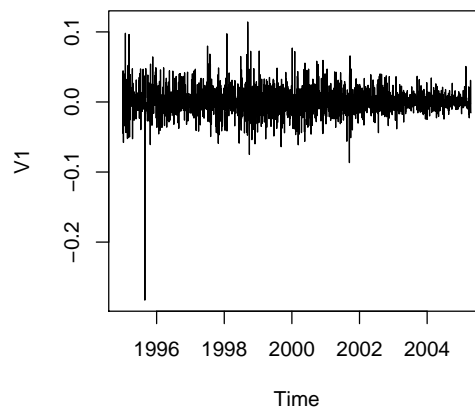
Dif gcarso a1

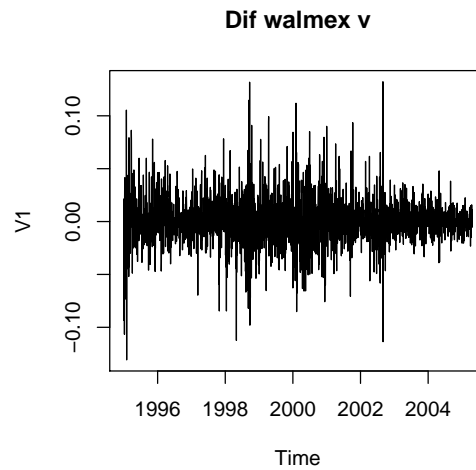
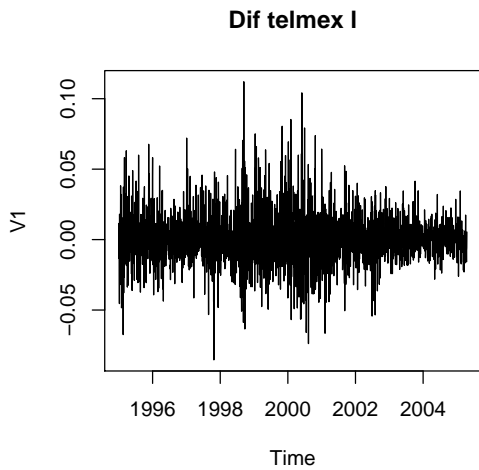
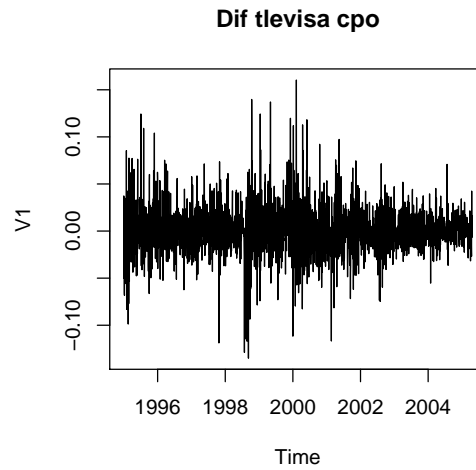
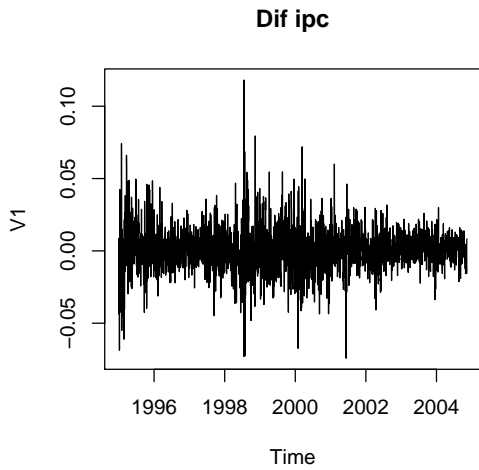


Dif gmexico b



Dif gmodelo c





La dualidad de precios en cada una de las acciones así como en el IPC permite analizar la eficiencia en mercados cuyas transacciones se sabe que cierran entre un precio máximo y un precio mínimo.

Tras haber comprobado la falta de estacionariedad de todas las series de precios se realiza un análisis de cointegración. El Teorema de Representación de Granger garantiza que debe existir un mecanismo de corrección de error en forma de un vector autorregresivo si las variables están cointegradas. Por esta razón, se especifica el proceso de ajuste que evita el crecimiento de los errores con respecto al equilibrio a largo plazo en las acciones cuyas parejas de precios logran una relación de cointegración significativa.

3.2.2. Eficiencia a Largo Plazo en la Formación del Precio.

La interdependencia que existe se muestra a través de la especificación para cada acción de la regresión mínimo cuadrática:

$$P_t^M = a^* + b^*P_t^X + u_t,$$

donde P_t^M es la cotización mínima, P_t^X la cotización máxima y a^* la constante. En general, la regresión mostrará indicios de autocorrelación, o lo que es lo mismo, los residuos u_t no serán independientes entre sí. Ahora bien, este hecho podrá ser consecuencia de dos supuestos distintos de partida: 1) el mercado es eficiente en el sentido “débil” de Fama, o, alternativamente, 2) la diferencia entre el error cometido en el precio máximo y el error cometido en el precio mínimo, respecto del precio de equilibrio no tiende a ser constante a largo plazo sino que depende del tiempo. En este segundo caso, el mercado no se podrá llamar “eficiente”, en el sentido de Fama. Por otra parte, el valor de b , calculado teóricamente, es la unidad, mientras que el de b^* podrá alejarse de la unidad. Éstas son las razones básicas por la que la regresión mínimo cuadrática de los precios máximos sobre los mínimos es un procedimiento inapropiado para estudiar la relación entre ambos. Los mismos motivos justifican el empleo de un procedimiento alternativo, especialmente diseñado para analizar la relación entre series cuya regresión no supera la prueba de independencia de los residuos. Tal procedimiento es la cointegración.

La conclusión a que se llega no es novedosa. Sin embargo, plantea la necesidad de cumplir ciertas etapas rutinarias en la investigación empírica, al efecto de adaptarla con todo el rigor posible al planteamiento teórico. Por ello, se efectúa la regresión mínimo cuadrática de los precios máximos sobre los mínimos en todas las series de acciones. Esto permite realizar un prueba de independencia de los residuos (Durbin Watson) y comprobar que todas las series revelan la presencia de autocorrelación, condición necesaria pero no suficiente para que el mercado en cuestión pueda ser calificado como de eficiencia débil en el sentido de Fama.

Pruebas de autocorrelación

Variable	Número de observaciones	Estadístico Durbin Watson
<i>bimbo</i>	2690	1.258484*
<i>cemex</i>	2690	1.148246*
<i>femsa</i>	2690	1.330172*
<i>gcarso</i>	2690	0.990304*
<i>gmexico</i>	2662	1.396045*
<i>gmodelo</i>	2690	1.176536*
<i>ipc</i>	2577	0.998541*
<i>telmex</i>	2690	1.068436*
<i>tlevisa</i>	2690	0.996269*
<i>walmex</i>	2690	1.035403*

Notas: * indica el rechazo de la hipótesis de no autocorrelación al 1% de significancia. El valor crítico al 1% para la prueba Durbin Watson, de acuerdo con los valores calculados para una variable explicativa y 100 observaciones, son: $D_L = 1.65$ y $D_U = 1.69$

Así, se comprueba la autocorrelación en los residuos de la regresión de las series de precios máximos y mínimos de todas las series de las acciones, y como el estadístico Durbin Watson de todos los residuos de las regresiones están entre 0 y 1.65 se dice que la autocorrelación es positiva.

El siguiente paso es comprobar la estacionariedad de los residuos.

Pruebas de estacionariedad de los residuales		
Variable	DFA	PP(1)
<i>bimbo</i>	-15.1105(4)	-48.8431
<i>cemex</i>	-19.8339(2)	-43.2438
<i>femsa</i>	-27.5930(1)	-41.1203
<i>gcarso</i>	-8.7020(9)	-46.7132
<i>gmexico</i>	-22.8163(2)	-42.0302
<i>gmodelo</i>	-11.6810(7)	-47.3532
<i>ipc</i>	-9.7502(8)	-43.9629
<i>telmex</i>	-10.6982(6)	-45.9885
<i>tlevisa</i>	-10.0580(8)	-46.3163
<i>walmex</i>	-8.4701(10)	-47.0071

Notas: Los valores críticos aparecen en Engle y Granger, y son diferentes de los de Fuller. La razón para esta disparidad es que los residuales son una variable “generada” mediante un proceso que escoge a los parámetros de modo que se minimice la varianza residual. Para evitar este sesgo los valores críticos deben aumentar ligeramente. El valor crítico al 1% para la prueba Dickey-Fuller Aumentada y Phillips-Perron, de acuerdo con los valores calculados por Engle y Granger, es -3.77 considerando un modelo sin constante y sin tendencia. Los valores entre paréntesis representan el número de rezagos utilizados en la prueba.

Como el estadístico DFA, es menor (es decir, mayor en valor absoluto) que el valor crítico tabulado, la hipótesis nula, es decir, que la serie no es estacionaria es rechazada, y se puede afirmar que las dos series son cointegradas de orden (1,1). A este procedimiento sigue el de Johansen, cuyos resultados entre los mínimos y los máximos se presentan en el siguiente Cuadro.

Pruebas de cointegración bivariada

Variable	Número de observaciones	$\hat{\lambda}_{Traza}$		$\hat{\lambda}_{Max}$		Vectores de cointegración
		$r = 0$	$r \leq 1$	$r = 0$	$r \leq 1$	
<i>bimbo</i>	2690	219.2476*	2.657561	216.5900*	2.657561	1
<i>cemex</i>	2690	256.1927*	1.022982	255.1697*	1.022982	1
<i>femsa</i>	2690	309.4098*	3.927654**	305.4822*	3.927654**	2
<i>gcarso</i>	2690	178.9310*	5.519216**	173.4118*	5.519216**	2
<i>gmexico</i>	2662	320.3320*	1.603370	318.7287*	1.603370	1
<i>gmodelo</i>	2690	231.7056*	2.211260	229.4944*	2.211260	1
<i>ipc</i>	2577	182.1478*	0.791358	181.3565*	0.791358	1
<i>telmex</i>	2690	184.0710*	2.348914	181.7221*	2.348914	1
<i>tlevisa</i>	2690	175.5369*	1.805944	173.7309*	1.805944	1
<i>walmex</i>	2690	168.2794*	1.029149	167.2502*	1.029149	1

Notas: * y ** indican rechazo de la hipótesis nula al 1% y 5% de significancia respectivamente. Los valores críticos al 1% y 5% para el estadístico de la traza son: 20.04 y 15.41 para $r = 0$, 6.65 y 3.76 para $r \leq 1$, los valores críticos al 1% y 5% para el estadístico de máxima verosimilitud son: 18.63 y 14.07 para $r = 0$, 6.65 y 3.76 para $r \leq 1$

En este caso, la hipótesis nula de existencia de un vector de cointegración ($r = 1$) se acepta en todas las series de acciones tanto por el estadístico de la traza como por el de máxima verosimilitud, además en el caso de femsa y grupo carso la existencia de a lo más un vector de cointegración puede ser rechazada al 5% por ambos estadísticos, pero se tomará el criterio de rechazar al 1%. El hallazgo de un vector de cointegración entre precios máximos y mínimos sostiene la hipótesis de eficiencia débil del mercado, en el sentido de Fama, que a partir de aquí se denominará, por afán de brevedad, eficiencia a largo plazo del mercado, o bien “eficiencia del mercado” simplemente.

3.2.3. Mecanismo de Ajuste al Precio de Equilibrio.

La existencia de una relación de cointegración entre los precios máximos y los precios mínimos es una condición necesaria pero no suficiente para verificar la existencia de un mercado puro de eficiencia “débil” en el sentido de Fama. Su confirmación mediante la prueba DFA en los residuos o Johansen permite aceptar la hipótesis de existencia de una relación de cointegración, pero no ilustra el ajuste de las series de máximos y mínimos a la misma. En este sentido, hay que apuntar que el procedimiento de Johansen es el primer paso en la comprobación de existencia de cointegración y que se exige la significancia del MCE para hablar con certeza de variables cointegradas. Por ello, el análisis se completa con la especificación de los errores cometidos respecto a dicha relación de equilibrio.

Si al menos uno de los parámetros de velocidad de ajuste es distinto de cero en el MCE, P_t^M y P_t^X están cointegradas. El mecanismo corrector de errores concilia la modelización dinámica a corto plazo con el equilibrio a largo plazo, de modo que el término de velocidad de ajuste mide el error de previsión de los agentes en el sentido de que será distinto de cero cuando haya alejamientos de la situación de equilibrio, produciéndose en el siguiente

te período un ajuste hacia dicha relación. Si en la estimación del MCE ninguno de los parámetros de velocidad de ajuste resulta significativo las variables no están cointegradas, el término de corrección es cero en ambas direcciones, y se habla de mercados totalmente abiertos. En este caso, el modelo se debe plantear sólo en términos de variables diferenciadas, sin que exista una relación a largo plazo identificable entre ellas que influya en la dinámica a corto plazo.

Para conocer la dirección de la relación, si es que existe, se realizó la prueba de causalidad de Granger a todas las series encontrando que la hipótesis nula de que los mínimos no causan a los máximos se rechaza con un α muy proximo a cero, lo mismo ocurre en el caso de la hipótesis de que los máximos no causan a los mínimos en el sentido de Granger, lo cual indica una causalidad en ambos sentidos en las acciones analizadas así como para el IPC.

Modelos de Corrección de Error					
Variable Dependiente	Variable	Coefficiente	Error Estándar	Estadístico t	Valor p
$\nabla bimbo_t^x$	$\nabla bimbo_{t-1}^m$	0.154597	0.024090	6.417424	0.0000
	$\nabla bimbo_{t-2}^m$	-0.062973	0.018693	-3.368844	0.0008
	$\nabla bimbo_{t-1}^x$	0.068034	0.026704	2.547773	0.0109
	res_{t-1}	0.204570	0.025801	7.928750	0.0000
$\nabla bimbo_t^m$	$\nabla bimbo_{t-1}^m$	0.071028	0.023574	3.012945	0.0026
	$\nabla bimbo_{t-1}^x$	0.205646	0.025160	8.173601	0.0000
	res_{t-1}	-0.361502	0.024833	-14.55742	0.0000
$\nabla cemea_t^x$	$\nabla cemea_{t-1}^m$	0.051985	0.021736	2.391643	0.0168
	$\nabla cemea_{t-1}^x$	0.131651	0.025950	5.073343	0.0000
	res_{t-1}	0.130147	0.022656	5.744531	0.0000
$\nabla cemea_t^m$	$\nabla cemea_{t-1}^m$	-0.067156	0.024214	-2.773501	0.0056
	$\nabla cemea_{t-2}^m$	-0.091539	0.018015	-5.081319	0.0000
	$\nabla cemea_{t-1}^x$	0.417488	0.029628	14.09109	0.0000
	res_{t-1}	-0.316685	0.025765	-12.29137	0.0000
$\nabla femsa_t^x$	$\nabla femsa_{t-2}^m$	-0.055124	0.018932	-2.911641	0.0036
	$\nabla femsa_{t-1}^x$	0.161872	0.020366	7.947979	0.0000
	res_{t-1}	0.276483	0.026415	10.46700	0.0000
$\nabla femsa_t^m$	$\nabla femsa_{t-2}^m$	-0.046645	0.018048	-2.584434	0.0098
	$\nabla femsa_{t-1}^x$	0.193707	0.019416	9.976917	0.0000
	res_{t-1}	-0.373761	0.025181	-14.84278	0.0000
$\nabla gcarso_t^x$	$\nabla gcarso_{t-2}^m$	-0.048480	0.016750	-2.894375	0.0038
	$\nabla gcarso_{t-1}^x$	0.179601	0.019406	9.254926	0.0000
	res_{t-1}	0.195011	0.022274	8.754935	0.0000
$\nabla gcarso_t^m$	$\nabla gcarso_{t-2}^m$	-0.065786	0.018965	-3.468806	0.0005
	$\nabla gcarso_{t-1}^x$	0.242757	0.021973	11.04807	0.0000
	res_{t-1}	-0.288453	0.025221	-11.43716	0.0000

Modelos de Corrección de Error (Continuación)

Variable Dependiente	Variable	Coficiente	Error Estándar	Estadístico t	Valor p
$\nabla gmexico_t^x$	$\nabla gmexico_{t-1}^m$	0.143246	0.024852	5.763986	0.0000
	$\nabla gmexico_{t-1}^x$	0.110901	0.026493	4.185974	0.0000
	res_{t-1}	0.248947	0.027738	8.974982	0.0000
$\nabla gmexico_t^m$	$\nabla gmexico_{t-2}^m$	-0.080540	0.018214	-4.421782	0.0000
	$\nabla gmexico_{t-1}^x$	0.286988	0.019971	14.37042	0.0000
	res_{t-1}	-0.331238	0.023315	-14.20724	0.0000
$\nabla gmodelo_t^x$	$\nabla gmodelo_{t-1}^m$	0.150945	0.017651	8.551900	0.0000
	$\nabla gmodelo_{t-2}^x$	-0.122378	0.019522	-6.268660	0.0000
	res_{t-1}	0.107075	0.021210	5.048247	0.0000
$\nabla gmodelo_t^m$	$\nabla gmodelo_{t-2}^m$	-0.103871	0.018093	-5.741004	0.0000
	$\nabla gmodelo_{t-1}^x$	0.253547	0.021223	11.94693	0.0000
	res_{t-1}	-0.355797	0.022819	-15.59217	0.0000
∇ipc_t^x	∇ipc_{t-1}^x	0.084322	0.027810	3.032100	0.0025
	∇ipc_{t-1}^m	0.216308	0.025536	8.470745	0.0000
	res_{t-1}	0.122814	0.024606	4.991223	0.0000
∇ipc_t^m	∇ipc_{t-1}^x	0.331160	0.029786	11.11787	0.0000
	∇ipc_{t-1}^m	0.090700	0.027351	3.316173	0.0009
	res_{t-1}	-0.293819	0.026355	-11.14858	0.0000
$\nabla telmex_t^x$	$\nabla telmex_{t-1}^x$	0.079979	0.027544	2.903646	0.0037
	$\nabla telmex_{t-2}^x$	-0.165783	0.019674	-8.426613	0.0000
	$\nabla telmex_{t-1}^m$	0.161303	0.025382	6.355003	0.0000
	res_{t-1}	0.074762	0.023813	3.139493	0.0017
$\nabla telmex_t^m$	$\nabla telmex_{t-1}^x$	0.381387	0.030165	12.64325	0.0000
	$\nabla telmex_{t-1}^m$	-0.094595	0.026133	-3.619802	0.0003
	$\nabla telmex_{t-2}^m$	-0.123417	0.018255	-6.760728	0.0000
	res_{t-1}	-0.270571	0.025541	-10.59363	0.0000
$\nabla tlevisa_t^x$	$\nabla tlevisa_{t-1}^x$	0.088397	0.028454	3.106665	0.0019
	$\nabla tlevisa_{t-1}^m$	0.131274	0.026488	4.956028	0.0000
	res_{t-1}	0.148241	0.026359	5.623940	0.0000
$\nabla tlevisa_t^m$	$\nabla tlevisa_{t-1}^x$	0.357887	0.019797	18.07822	0.0000
	$\nabla tlevisa_{t-2}^m$	-0.089695	0.018241	-4.917169	0.0000
	res_{t-1}	-0.260974	0.023468	-11.12058	0.0000
$\nabla walmex_t^x$	$\nabla walmex_{t-1}^x$	0.066413	0.026571	2.499453	0.0125
	$\nabla walmex_{t-1}^m$	0.125618	0.025032	5.018243	0.0000
	res_{t-1}	0.166523	0.023989	6.941664	0.0000
$\nabla walmex_t^m$	$\nabla walmex_{t-1}^x$	0.307717	0.019584	15.71299	0.0000
	$\nabla walmex_{t-2}^m$	-0.105667	0.018207	-5.803820	0.0000
	res_{t-1}	-0.238495	0.021198	-11.25095	0.0000

Notas: Con x se denota la serie de precios máximos y con m la de precios mínimos, la variable res_t son los residuales de la correspondiente ecuación de cointegración

3.2.4. Interpretación de los Resultados Empíricos.

La prueba de MCE evidencia la existencia de mecanismos de corrección de errores en ambas direcciones. Es decir, ocurre que ambas series muestran una propensión a mantener

constante la diferencia entre los valores de ambas. A lo que se denomina MCE de doble dirección. Según lo expuesto, la aplicación sucesiva de las dos pruebas, primero la de cointegración y luego la de MCE, arroja los siguientes resultados:

Ambas series son $I(1)$ y, además, están cointegradas; existe MCE de doble dirección. Esto supone que la variabilidad en la predicción de errores y en el número de transacciones pueden descartarse como origen de la autocorrelación en la regresión de una serie sobre la otra; una evidencia favorable, en suma, a la hipótesis de que el mercado es eficiente en el sentido de Fama.

Se denominan mercados (totalmente) cerrados a aquellos en que precios máximos y mínimos están cointegrados y existe doble MCE, que es precisamente el caso de las acciones y el IPC ya que el término de velocidad de ajuste es significativo para todas las variables y además para ambos modelos con una significancia del 99%. El número de transacciones se mantiene de manera regular de un periodo a otro, y lo mismo ocurre con la destreza de los operadores para predecir precios de equilibrio. Los operadores tienden a ser los mismos; con toda probabilidad, no es fácil adquirir la condición de “miembro del mercado”. De ahí que se hablará de “mercados cerrados”. Tampoco es casualidad que sean estos mercados los que muestran eficiencia plena en el uso de la información.

En suma, la contrastación empírica permite clasificar a las acciones según su grado de eficiencia de mercado. De la investigación se deduce que el mercado accionario mexicano está altamente profesionalizado (doble MCE).

Todas las series de acciones así como el IPC consideradas encajan en la noción de “eficiencia débil” de Fama. Los resultados obtenidos han permitido explicar el grado de profesionalidad de una serie de acciones que cotizan en la BMV así como al IPC como su principal indicador, teniendo en cuenta su grado de eficiencia en el uso de la información.

De acuerdo con los resultados obtenidos el mercado financiero mexicano es clasificado como un mercado plenamente eficiente en el sentido “débil” de Fama.

3.3. La Eficiencia del Mercado Cambiario entre el Euro, el Peso Mexicano y el Dólar

Durante el último trimestre del año 2002 y los primeros meses del 2003 se observó una caída del dólar con respecto al euro, junto con un comportamiento singular en las trayectorias de éstas divisas y el peso mexicano. En efecto, la devaluación del dólar en relación al euro ha estado acompañada de una devaluación del peso frente al dólar. La explicación más extendida sobre este comportamiento es que la incertidumbre y los riesgos cambiarios,

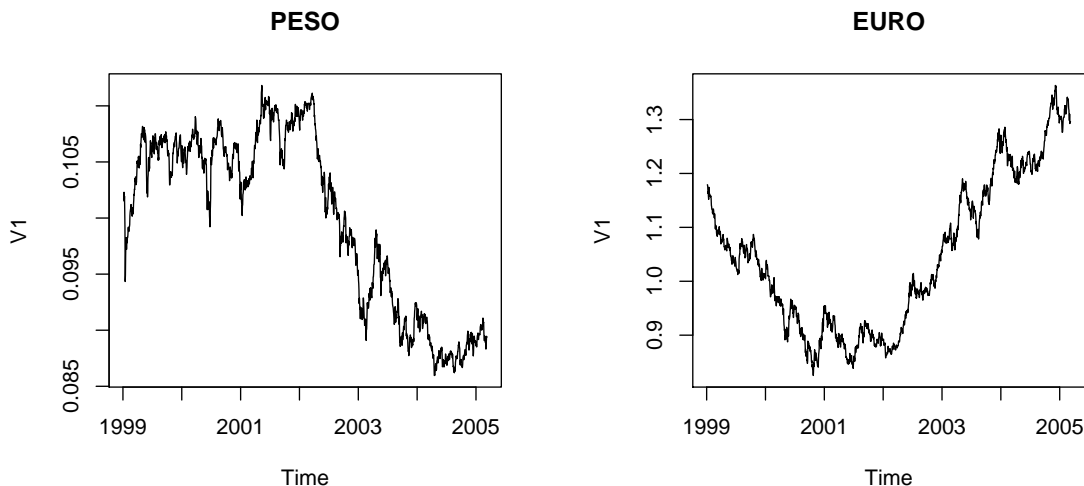
en el contexto internacional actual, se tradujeron en la salida de capitales de la esfera del dólar hacia el euro u otras monedas. Este flujo se inició con la salida de capitales de la moneda más débil, en este caso el peso, que se tradujo entonces en su devaluación, en una proporción similar a la del dólar con respecto al euro.

Este comportamiento de los mercados cambiarios, junto con la explicación referida, pone en el centro del debate la existencia de eficiencia en el mercado del tipo de cambio entre estas tres monedas y en particular para el caso mexicano. En efecto, la evolución del tipo de cambio en la economía mexicana es ciertamente una variable fundamental, en la medida en que se encuentra fuertemente relacionada con la trayectoria de otras variables como la tasa de inflación, el tipo de interés nominal o la balanza comercial (Carstens, y Werner, 1999). Así, el régimen cambiario, a partir de 1995, se modificó pasando de una flotación controlada a una determinación del tipo de cambio por las fuerzas del mercado. En este sentido, verificar la hipótesis de eficiencia en el caso mexicano adquiere una mayor relevancia. Lo que quiere decir que la validez de esta hipótesis indica que el tipo de cambio en México responde adecuadamente a las fuerzas de la libre competencia, y por tanto no existen factores o información adicional que sugieran la presencia de imperfecciones en el mercado que puedan traducirse en ganancias extraordinarias. En este sentido, el objetivo de esta sección es analizar la presencia de eficiencia en el mercado cambiario entre las divisas mencionadas. La selección del dólar y el euro como monedas de referencia se debe básicamente a su importancia en el mercado internacional.

3.3.1. Los Datos

La información utilizada incluye datos diarios del 4 de enero de 1999 al 1 de abril de 2005 de los tipos de cambio del peso (Pt) y el euro (Et), ambos con referencia al dólar, considerando el dato de venta para el cierre del día, reportado en las estadísticas financieras internacionales del FMI. Letras minúsculas representan el logaritmo de las series. Las estimaciones econométricas y estadísticas se realizaron en E-views 4.1.

El análisis del orden de integración de las series se realizó con las pruebas estándar de raíces unitarias de Dickey Fuller Aumentada y de Phillips-Perron. La especificación de estas pruebas correspondió al procedimiento “de lo general a lo específico”, estimando en principio regresiones con constante y tendencia, analizando además su significancia estadística. El número de retardos (k) en la prueba ADF fue seleccionado de acuerdo al criterio de información bayesiana.



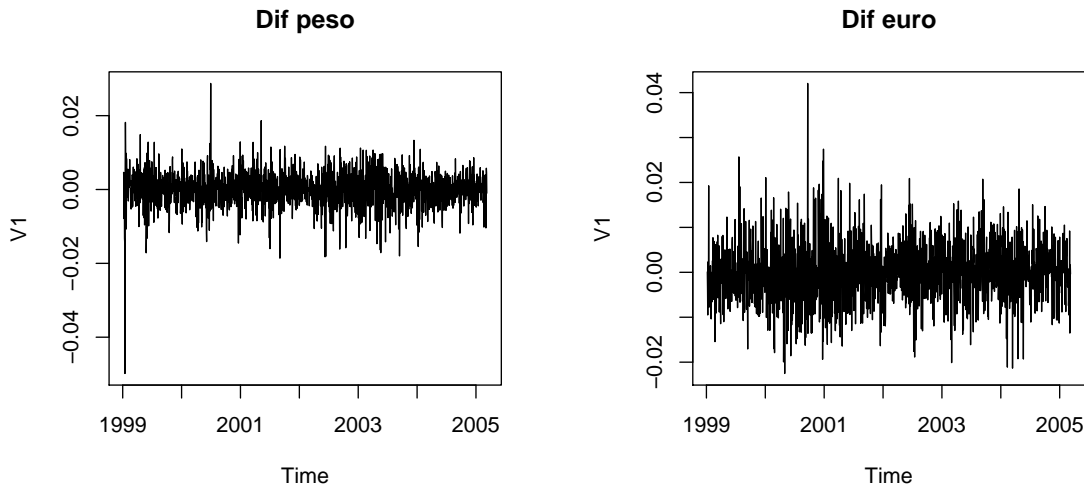
Las pruebas de hipótesis de raíces unitarias se muestran en el siguiente cuadro, donde se indica que el tipo de cambio nominal del peso mexicano, respecto al dólar es una serie no estacionaria (I(1)) atendiendo a las pruebas ADF, PP por otra parte, los resultados sobre el euro-dólar sugieren también, que la serie es I(1). Esta evidencia es consistente con la hipótesis de que los tipos de cambio nominales tienen normalmente una raíz unitaria y siguen una caminata aleatoria (Meese y Singleton, 1982, Baillie y Bollerslev, 1989, Alexander y Johnson, 1992 y MacDonald y Taylor, 1989).

Así, en el caso donde la evolución del tipo de cambio puede aproximarse como una senda aleatoria, entonces la mejor predicción que puede realizarse es el valor previo con un coeficiente igual a uno, lo que se conoce como la condición de insesgamiento (Diebold, Gardeazabal e Yilmaz, 1994).

Pruebas de raíces unitarias

Variable	DFA			PP		
	A	B	C	A	B	C
p_t	-2.6031(24)	-0.8956(24)	0.6106(24)	-2.6154	-0.8995	0.6122
∇p_t	-37.322(24)*	-37.322(24)*	-37.324(24)*	-37.225*	-37.226*	-37.228*
e_t	-2.4152(24)	-0.5736(24)	-0.4652(24)	-2.4096	-0.5102	-0.4000
∇e_t	-42.032(24)*	-41.909(24)*	-41.919(24)*	-42.075*	-41.921*	-41.931*

Notas: * indica rechazo de la hipótesis nula al 1% de significancia. Los valores críticos al 1% para la prueba Dickey-Fuller Aumentada y Phillips-Perron, de acuerdo con los valores calculados por MacKinnon, son de -3.9637 incluyendo constante y tendencia (modelo A), -3.4342 únicamente la constante (modelo B) y -2.5663 sin constante y sin tendencia (modelo C)



El número de retardos del modelo de vectores autoregresivos, utilizado para el procedimiento de Johansen, se seleccionó de acuerdo a los criterios de información de Akaike (AIC) y Schwarz (BIC), en donde se observa que el modelo VAR puede especificarse considerando 1 u 8 retardos. En este sentido, se utilizó el número de retardos correspondiente el criterio de BIC.

Criterios de Akaike y Schwarz para determinar número de rezagos

	Número de rezagos											
	1	2	3	4	5	6	7	8	9	10	11	12
AIC	-14.91	-14.91	-14.90	-14.90	-14.89	-14.89	-14.89	-14.95	-14.94	-14.94	-14.94	-14.93
BIC	-14.89	-14.87	-14.85	-14.84	-14.82	-14.80	-14.79	-14.83	-14.81	-14.80	-14.78	-14.77

El procedimiento de Johansen (1988) rechaza la hipótesis nula de cero vectores de cointegración, tanto por el estadístico de la traza, como por el de la raíz característica máxima, indicando la presencia de una relación de cointegración entre los tipos de cambio nominal peso-dólar y el euro-dólar. Ello sugiere que existe una relación estable entre el peso, el euro y el dólar en el largo plazo.

Estadísticos de la prueba del procedimiento de Johansen

Valores característicos	Hipótesis No EC	$\hat{\lambda}_{max}$	λ_{max} 95 %	$\hat{\lambda}_{Traza}$	λ_{Traza} 95 %
0.011817	Ninguna*	19.13851	18.96	29.41921	25.32
0.006365	A lo más 1	10.28070	12.25	10.28070	12.25

Notas: $\hat{\lambda}_{max}$ = estadístico de la raíz característica máxima. $\hat{\lambda}_{Traza}$ = estadístico de la traza. (*) Rechazo de la hipótesis nula al 5% de significancia. Número de retardos utilizados en el VAR 1.

Así, la hipótesis de eficiencia se puede comprobar estimando la ecuación con y sin restricciones:

$$\begin{aligned}\nabla p_t &= \alpha + \gamma \nabla e_t - (p_{t-1} - \phi e_{t-1}) + v_{1t} \\ \nabla p_t &= \alpha + \phi \nabla e_t - (p_{t-1} - \phi e_{t-1}) + v_{2t}\end{aligned}$$

Las estimaciones por máxima verosimilitud con información completa (FIML, por sus siglas en inglés) del modelo sin restricciones y con restricciones se presentan en los cuadros siguientes.

Estimaciones por FIML de la ecuación sin restricción

<i>Parámetro</i>	<i>Coficiente</i>	<i>Error</i>	<i>Estadístico</i>	<i>Valor</i>
		<i>Estándar</i>	<i>t</i>	<i>p</i>
α	-2.287142	0.001243	-1840.304	0.0000
γ	-0.613618	0.149354	-4.108479	0.0000
ϕ	-0.523227	0.010440	-50.11840	0.0000

Nota: función logarítmica de máxima verosimilitud = 2943.164.

Estimaciones por FIML de la ecuación con restricción

<i>Parámetro</i>	<i>Coficiente</i>	<i>Error</i>	<i>Estadístico</i>	<i>Valor</i>
		<i>Estándar</i>	<i>t</i>	<i>p</i>
α	-2.287139	0.001242	-1841.352	0.0000
ϕ	-0.523420	0.010428	-50.19426	0.0000

Nota: función logarítmica de máxima verosimilitud = 2942.962.

El contraste de razón de máxima verosimilitud para ambas estimaciones se distribuye como una $\chi^2_{(m)}$ con m grados de libertad, dados por el número de restricciones. De este modo el contraste de hipótesis se define como:

$$LR = 2 \times (2943,164 - 2942,962) = 0,404.$$

El valor crítico al 5% de significancia estadística de una $\chi^2_{(1)} = 3,84$, indica que no puede rechazarse la hipótesis nula de que $\gamma = \phi$. Ello implica el no rechazo de la hipótesis de eficiencia en el mercado cambiario entre el peso y el euro, tomando como referencia al dólar.

3.3.2. Interpretación de los Resultados Empíricos.

El conjunto de las estimaciones realizadas indican entonces que los mercados cambiarios conjuntos del peso mexicano, y el euro, con referencia al dólar, son eficientes. Esto es, las restricciones en los parámetros obtenidas por FIML no son rechazadas por los datos. Este resultado indica que no existen opciones de arbitraje ni de ganancias extraordinarias en

estos mercados.

En esta sección se utiliza una nueva prueba desarrollada por Ferré y Hall (2002) para analizar la eficiencia del mercado cambiario entre el peso mexicano, el euro y el dólar. Esta prueba difiere de la propuesta realizada por Granger (1986), en donde la presencia de cointegración entre dos series de tipo de cambio implica que el mercado es ineficiente, ya que existe un mecanismo de corrección de errores que indica una relación de causalidad entre las series. La nueva prueba planteada por Ferré y Hall (2002) rompe el vínculo directo que se había establecido entre cointegración e ineficiencia en el mercado, como consecuencia de la relación que se establecía entre capacidad de pronóstico e ineficiencia. De este modo, el concepto de eficiencia del mercado corresponde al caso donde los tipos de cambio incluyen, en todo momento, el conjunto de la información disponible y, por lo tanto, no pueden realizarse ganancias extraordinarias. En este sentido, las pruebas de cointegración son insuficientes para argumentar a favor de la eficiencia en el mercado.

Los resultados obtenidos indican que el mercado cambiario entre el peso y el euro con respecto al dólar es eficiente. Esto es, las restricciones impuestas en los parámetros, en el modelo con dinámica limitada, no rechazan la hipótesis de eficiencia en el mercado. Ello indica que no existe la posibilidad de arbitraje o de ganancias extraordinarias, y que en particular el peso mexicano no está sujeto a ataques especulativos en la triangulación entre tipos de cambio. Así, la aplicación de un régimen cambiario basado en las libres fuerzas del mercado, en donde no existen problemas de ineficiencia, resulta ser confiable y conducir a la inexistencia de ganancias especulativas. En este sentido, es importante reconocer que la percepción de las virtudes del mercado cambiario mexicano, por parte de las autoridades monetarias, se sustenta con este análisis dándole mayor confiabilidad.

Conclusiones

El propósito de la econometría es utilizar datos, métodos de inferencia estadística y modelos estructurales y descriptivos para la resolución de problemas económicos. La aplicación y desarrollo de métodos econométricos en finanzas es resiente y va a la par con el estudio de los mercados y productos financieros. Uno de los temas que más han ocupado a los econométricos financieros ha sido el de la eficiencia de mercado, la aplicación de este trabajo tuvo como objetivo práctico medir la eficiencia de mercados utilizando la metodología alternativa de cointegración introducida por Engle y Granger, la cual permite trabajar con series no estacionarias.

La metodología de Cointegración ofrece un procedimiento que cumple con varias características importantes: permite distinguir entre regresiones espurias y regresiones válidas, en el sentido que representan una relación estable de largo plazo entre las variables, con mecanismos de ajuste que tienden a disminuir las discrepancias que se presenten; permite combinar la metodología de series de tiempo con información de teorías económicas de equilibrio de largo plazo, con lo cual se eliminan muchas de las objeciones que se hacen a cada una de estas metodologías tomadas por separado; permite la mezcla de información de distinta periodicidad, por ejemplo, la ecuación de Cointegración podría hacerse con datos anuales y la de corrección de errores con información mensual; es relativamente fácil de aplicar, su uso consiste en la estimación de varias ecuaciones por mínimos cuadrados ordinarios, la dificultad principal estriba en la teoría estadística que esta por detrás de las pruebas, teoría que es mucho más difícil que la teoría usual.

Los resultados muestran que las series en el periodo observado, están cointegradas y además existe una relación de causalidad entre las variables que a su vez admiten una representación por medio de un mecanismo de corrección de error, la conclusión importante de este trabajo es que las acciones seleccionadas, el IPC y el tipo de cambio peso-euro con el dólar como referencia son mercados eficientes en forma débil de Fama, tal y como lo muestran los resultados.

El que los mercados sean eficientes, implica que los precios actuales reflejan toda la información y éstos pueden cambiar sólo en respuesta de nueva información, por lo tanto no es posible predecir o pronosticar un cambio en precios porque éste es un fenómeno aleatorio.

Lo cual no elimina que puedan llegar a existir desequilibrios en el corto plazo, que pueden ser provocados por la utilización de mecanismos privilegiados de acceso a la información relevante de algunos agentes económicos para conseguir utilidades extraordinarias, en otras palabras, el costo de la información en esos momentos no es igual a cero como lo estipularía el esquema teórico diseñado por Eugene Fama, esto es ocasionado por gente sin ética que en ocasiones crean efectos ficticios en los mercados por ser choques de fuertes magnitudes; sin embargo, no son trascendentes porque en el largo plazo, los mercados tienden a estabilizarse.

Es importante recalcar que los resultados y sus implicaciones son válidas en el periodo de análisis así como que no existe un modelo verdadero para cada mercado o activo financiero, se tiene que recordar que un modelo es una representación simplificada de la realidad, por lo que hay que estar conscientes de que un modelo conlleva necesariamente errores de especificación.

Los procedimientos estadísticos han sido válidos pero han requerido hipótesis teóricas y supuestos concretos, respecto a cuestiones, en principio, más simples, de la teoría económica como es el análisis de precios. Este es uno de los resultados que se deducen de la reinterpretación del modelo de Fama, como la referencia más extendida del concepto eficiencia en precios.

La metodología esta todavía en desarrollo, hace falta bastante trabajo, por ejemplo, en la estimación y análisis de Cointegración no lineal. En resumen se trata de una teoría que parece muy adecuada para una buena cantidad de problemas que se presentan en economía y finanzas.

Apéndice

En seguida se presenta la demostración el Teorema de Representación de Granger, tomando la representación VAR como punto de partida, esta demostración fue tomada de Banerjee et al. (1993).

Sea $X_t \sim I(1)$ un vector con n componentes y con tendencia determinista en la media en cada una de sus componentes. Supóngase que el sistema puede escribirse como un VAR de orden k finito

$$X_t = \mu + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \pi_k X_{t-k} + \epsilon_t, \quad t = 1, 2, \dots, T$$

donde ϵ_t es un vector ruido blanco y los primeros k datos $X_{1-k}, X_{1-k+1}, \dots, X_0$ son fijos, el modelo puede ser reescrito como un MCE

$$\begin{aligned} X_t &= \mu + (X_{t-1} + X_{t-2} + \cdots + X_{t-k}) - (X_{t-1} + X_{t-2} + \cdots + X_{t-k}) \\ &\quad + \pi_1 (X_{t-2} + X_{t-3} + \cdots + X_{t-k}) - \pi_1 (X_{t-2} + X_{t-3} + \cdots + X_{t-k}) \\ &\quad + \pi_2 (X_{t-3} + X_{t-4} + \cdots + X_{t-k}) - \pi_2 (X_{t-3} + X_{t-4} + \cdots + X_{t-k}) + \cdots \\ &\quad + \pi_{k-1} (X_{t-k}) - \pi_{k-1} (X_{t-k}) + \pi_1 X_{t-1} + \pi_2 X_{t-2} + \cdots + \pi_k X_{t-k} + \epsilon_t \\ X_t - X_{t-1} &= \mu - (X_{t-1} - X_{t-2}) - (X_{t-2} - X_{t-3}) - \cdots - (X_{t-k+1} - X_{t-k}) - X_{t-k} \\ &\quad + \pi_1 [(X_{t-1} - X_{t-2}) + (X_{t-2} - X_{t-3}) + \cdots + (X_{t-k+1} - X_{t-k})] + \pi_1 X_{t-k} \\ &\quad + \pi_2 [(X_{t-2} - X_{t-3}) + (X_{t-3} - X_{t-4}) + \cdots + (X_{t-k+1} - X_{t-k})] + \pi_2 X_{t-k} \\ &\quad + \cdots + \pi_{k-1} [X_{t-k+1} - X_{t-k}] + \pi_{k-1} X_{t-k} + \pi_k X_{t-k} + \epsilon_t \\ \nabla X_t &= \mu + (-I_n + \pi_1) \nabla X_{t-1} + (-I_n + \pi_1 + \pi_2) \nabla X_{t-2} + \cdots \\ &\quad + (-I_n + \pi_1 + \pi_2 + \cdots + \pi_{k-1}) \nabla X_{t-k+1} \\ &\quad + (-I_n + \pi_1 + \pi_2 + \cdots + \pi_k) X_{t-k} + \epsilon_t \\ &= \mu + \Gamma_1 \nabla X_{t-1} + \Gamma_2 \nabla X_{t-2} + \cdots + \Gamma_{k-1} \nabla X_{t-k+1} + \pi X_{t-k} + \epsilon_t \\ \text{entonces} \quad \nabla X_t &= \mu + \sum_{i=1}^{k-1} \Gamma_i (1-B) B^i X_t + \pi X_{t-k} + \epsilon_t. \end{aligned}$$

Esta expresión puede describirse como:

$$\begin{aligned} \nabla X_t &= \mu + \sum_{i=1}^{k-1} \Gamma_i (1-B) B^i X_t + \pi X_{t-k} + \epsilon_t \\ (1-B) X_t &= \mu + \sum_{i=1}^{k-1} \Gamma_i (1-B) B^i X_t + \pi X_{t-k} + \epsilon_t \\ \mu + \epsilon_t &= (1-B) X_t - \sum_{i=1}^{k-1} \Gamma_i (1-B) B^i X_t - \pi X_{t-k} \end{aligned}$$

$$\begin{aligned}\mu + \epsilon_t &= \left[(1-B)I_n - \sum_{i=1}^{k-1} \Gamma_i (1-B)B^i - \pi B^k \right] X_t \\ \Pi(B)X_t &= \mu + \epsilon_t,\end{aligned}$$

donde

$$\begin{aligned}\Pi(B) &= (1-B)I_n - \sum_{i=1}^{k-1} \Gamma_i (1-B)B^i - \pi B^k \\ \Gamma_i &= -I_n + \pi_1 + \pi_2 + \cdots + \pi_i \quad i = 1, 2, \dots, k \\ \Gamma_k &= \pi = \Pi(-1).\end{aligned}$$

Además

$$\begin{aligned}\Pi(B)X_t &= (1-B)X_t - \sum_{i=1}^{k-1} \Gamma_i (1-B)B^i X_t - \pi X_{t-k} \\ &= -\pi X_{t-k} + (1-B)^{-1} \left[(1-B)I_n - \sum_{i=1}^{k-1} \Gamma_i (1-B)B^i - \pi B^k + \pi B^k \right] \nabla X_t \\ &= -\pi X_{t-k} + (1-B)^{-1} [\Pi(B) - \Pi(1)B^k] \nabla X_t\end{aligned}$$

$$\text{entonces} \quad \Pi(B)X_t = -\pi X_{t-k} + \Psi(B)\nabla X_t = \mu + \epsilon_t,$$

donde $\Psi(B) = (1-B)^{-1} (\Pi(B) - \Pi(1)B^k) = I_n - \sum_{i=1}^{k-1} \Gamma_i B^i$. Además $-\Psi = -\Psi(1)$.

La componente ortogonal P_\perp de una matriz P de rango q y dimensión $n \times q$ ($0 < q < n$) se define como sigue:

- i P_\perp es de dimensión $n \times (n - q)$;
- ii $P'_\perp P = 0_{(n-q) \times q}$, $P' P_\perp = 0_{q \times (n-q)}$;
- iii P'_\perp tiene rango $n - q$.

Para empezar se deben realizar ciertas suposiciones

Suposición A1. El polinomio característico,

$$\Pi(z) = (1-z)I_n - \sum_{i=1}^{k-1} \Gamma_i (1-z)z^i - \pi z^k,$$

tiene sus raíces iguales o estrictamente mayores a uno; esto es, $|\Pi(z)| = 0$ lo cual implica que, $|z| > 1$ o bien $z = 1$.

Suposición A2. La matriz π de $n \times n$ tiene rango reducido $r < n$ y por lo tanto se puede expresar como el producto de dos matrices de $n \times r$, γ y α , donde γ y α tienen rango r . Así $\pi = \gamma\alpha'$.

Suposición A3. La matriz de $(n - r) \times (n - r)$, $\gamma'_\perp \Psi \alpha_\perp$ tiene rango completo $n - r$.

La suposición A1 garantiza que la no estacionariedad de X_t puede ser removida tomando diferencia. A2 descarta que X_t sea un proceso estacionario. Si π tiene rango completo (esto es, si $|\Pi(z)|$ no tiene raíces unitarias), entonces se tiene $X_t = \frac{\mu + \epsilon_t}{\Pi(B)}$, lo cual implicaría que X_t fuera estacionaria. Esto también declara, en la forma autorregresiva, que el sistema tiene r vectores de cointegración linealmente independientes. A la luz de la suposición A2, $\gamma\alpha'$ proporciona una transformación de la matriz π (y por tanto una combinación lineal de las x_{it} que es estacionaria). La suposición A3, esencialmente, asegura que X_t es de orden no mayor a 1. Bajo los supuestos declarados arriba, pueden probarse los siguientes resultados:

R1 ∇X_t es estacionario.

R2 $\alpha' X_t$ es estacionario.

R3 $E(\nabla X_t) = \alpha_{\perp} (\gamma'_{\perp} \Psi \alpha_{\perp})^{-1} \gamma'_{\perp} \mu$.

R4 $E(\alpha' X_t) = -(\gamma' \gamma)^{-1} \gamma' \mu + (\gamma' \gamma)^{-1} (\gamma' \Psi \alpha_{\perp}) (\gamma'_{\perp} \Psi \alpha_{\perp})^{-1} \gamma'_{\perp} \mu$.

R5 ∇X_t tiene una representación de medias móviles dada por

$$\nabla X_t = C(B)(\mu + \epsilon_t).$$

R6 $C(1) = \alpha_{\perp} (\gamma'_{\perp} \Psi \alpha_{\perp})^{-1} \gamma'_{\perp}$ tiene rango $n - r$.

R7 $\alpha' C(1) = 0_{r \times n}$

$C(1) \gamma = 0_{n \times r}$

R8 $X_t = X_0 + C(1) \sum_{i=1}^t \epsilon_i + \tau t + S_t$,

donde $C(B) = C(1) + (1 - B)C_1(B)$, $\tau = C(1)\mu$, X_0 es un vector de constantes de integración, y $S_t = C_1(B)\epsilon_t$.

Demostración Multiplicando $-\pi X_t + \Psi(B)\nabla X_t = \mu + \epsilon_t$ por γ' y γ'_{\perp} respectivamente para obtener las ecuaciones

$$\begin{aligned} -\gamma' \gamma \alpha' X_t + \gamma' \Psi(B) \nabla X_t &= \gamma' (\mu + \epsilon_t), \\ \gamma'_{\perp} \Psi(B) \nabla X_t &= \gamma'_{\perp} (\mu + \epsilon_t), \end{aligned}$$

usando la descomposición $\pi = \gamma \alpha'$ y el resultado $\gamma'_{\perp} \gamma = 0_{(n-r) \times r}$. La matriz π no es invertible, por lo tanto el sistema dado arriba no puede invertirse directamente para expresar las x_{it} en términos de las ϵ_{it} . Para obtener un sistema invertible, se definen dos nuevas variables, $w_t = (\alpha' \alpha)^{-1} \alpha' X_t$ y $v_t = (\alpha'_{\perp} \alpha_{\perp})^{-1} \alpha'_{\perp} \nabla X_t$. Luego, se definen las matrices $\bar{\alpha} = \alpha(\alpha' \alpha)^{-1}$ y $\bar{\alpha}_{\perp} = \alpha_{\perp}(\alpha'_{\perp} \alpha_{\perp})^{-1}$. Sea $R = (\alpha, \alpha_{\perp})$ una matriz de $n \times n$ y rango n . Entonces $R(R'R)^{-1}R' = I_n$ y por lo tanto $(\alpha \bar{\alpha}' + \alpha_{\perp} \bar{\alpha}'_{\perp}) = I_n$. Así,

$$\nabla X_t = (\alpha \bar{\alpha}' + \alpha_{\perp} \bar{\alpha}'_{\perp}) \nabla X_t = \alpha \nabla w_t + \alpha_{\perp} v_t.$$

Sustituyendo esta expresión en el sistema de ecuaciones

$$\begin{aligned} -(\gamma' \gamma)(\alpha' \alpha) w_t + \gamma' \Psi(B) \alpha \nabla w_t + \gamma' \Psi(B) \alpha_{\perp} v_t &= \gamma' (\mu + \epsilon_t), \\ \gamma'_{\perp} \Psi(B) \alpha \nabla w_t + \gamma'_{\perp} \Psi(B) \alpha_{\perp} v_t &= \gamma'_{\perp} (\mu + \epsilon_t), \end{aligned}$$

Donde el primer término del lado izquierdo de la primera ecuación necesita escribirse primero como $-(\gamma' \gamma)(\alpha' \alpha)(\alpha' \alpha)^{-1} \alpha' X_t$. Las ecuaciones para w_t y v_t pueden escribirse ahora en forma autorregresiva como

$$\tilde{A}(B) (w'_t, v'_t)' = (\gamma, \gamma_{\perp})' (\mu + \epsilon_t),$$

con

$$\tilde{A}(z) = \begin{bmatrix} -(\gamma' \gamma)(\alpha' \alpha) + \gamma' \Psi(z) \alpha (1 - z) & \gamma' \Psi(z) \alpha_{\perp} \\ \gamma'_{\perp} \Psi(z) \alpha (1 - z) & \gamma'_{\perp} \Psi(z) \alpha_{\perp} \end{bmatrix}.$$

Para $z = 1$, el determinante de esta matriz es:

$$|\tilde{A}(1)| = (-1)^r |\gamma' \gamma| |\alpha' \alpha| |\gamma'_{\perp} \Psi \alpha_{\perp}|,$$

que es diferente de cero por los supuestos A2 y A3. Entonces $z = 1$ no es una raíz. Para $z \neq 1$, a través de operaciones algebraicas se llega a la siguiente expresión de la matriz $\tilde{A}(z)$

$$\tilde{A}(z) = (\gamma, \gamma_{\perp})' \Pi(z) [\alpha, \alpha_{\perp}(1-z)^{-1}].$$

Para mostrar esto, en $\tilde{A}(z)$ se sustituye $\Psi(z)$ en términos de $\Pi(z)$ y $\Pi(1) = -\pi$, se usa la descomposición $\pi = \gamma\alpha'$ y la condición de ortogonalidad $\gamma'_{\perp}\gamma = \alpha'_{\perp}\alpha = 0_{(n-r) \times r}$. Para $z \neq 1$, por lo tanto el determinante de $\tilde{A}(z)$ es

$$|\tilde{A}(z)| = |(\gamma, \gamma_{\perp})| |\Pi(z)| |(\alpha, \alpha_{\perp})| (1-z)^{-(n-r)},$$

donde se tiene que usar el resultado de que el determinante de una matriz obtenida al multiplicar $n-r$ columnas (o renglones) de una matriz de $n \times n$ por una constante es el determinante de la matriz original multiplicado por la constante elevado a la potencia $n-r$. Así para $z \neq 1$, $|\tilde{A}(z)| = 0$ si y sólo si $|\Pi(z)| = 0$. Por el supuesto A1, si se excluye $z = 1$, sólo las raíces restante del determinante estarán fuera del círculo unitario.

Esto muestra que todas las raíces de $|\tilde{A}(z)| = 0$ están fuera del disco unitario. Entonces el sistema

$$\begin{aligned} -(\gamma'\gamma)(\alpha'\alpha)w_t + \gamma'\Psi(B)\alpha\nabla w_t + \gamma'\Psi(B)\alpha_{\perp}v_t &= \gamma'(\mu + \epsilon_t), \\ \gamma'_{\perp}\Psi(B)\alpha\nabla w_t + \gamma'_{\perp}\Psi(B)\alpha_{\perp}v_t &= \gamma'_{\perp}(\mu + \epsilon_t), \end{aligned}$$

es invertible y w_t y v_t pueden ser dadas las distribuciones iniciales tal que se vuelvan estacionarias. Por tanto en $\nabla X_t = \alpha_{\perp}v_t + \alpha\nabla w_t$, la estacionariedad de v_t y w_t implica la estacionariedad de ∇X_t . Además, como $\alpha'X_t = (\alpha'\alpha)w_t$, entonces $\alpha'X_t$ es también estacionario. Esto completa la prueba de (R1) y (R2). Para probar (R3) y (R4), hay que notar que

$$(w'_t, v'_t)' = \tilde{A}(B)^{-1}(\gamma, \gamma_{\perp})'(\mu + \epsilon_t).$$

Así, $E[(w'_t, v'_t)'] = \tilde{A}(1)^{-1}(\gamma, \gamma_{\perp})'\mu$

$$\tilde{A}(1) = \begin{bmatrix} -(\gamma'\gamma)(\alpha'\alpha) & \gamma'\Psi(1)\alpha_{\perp} \\ 0 & \gamma'_{\perp}\Psi(1)\alpha_{\perp} \end{bmatrix}.$$

Y su inversa es

$$\tilde{A}(1)^{-1} = \begin{bmatrix} -(\alpha'\alpha)^{-1}(\gamma'\gamma)^{-1} & (\alpha'\alpha)^{-1}(\gamma'\gamma)^{-1}\gamma'\Psi(1)\alpha_{\perp}(\gamma'_{\perp}\Psi(1)\alpha_{\perp})^{-1} \\ 0 & (\gamma'_{\perp}\Psi(1)\alpha_{\perp})^{-1} \end{bmatrix}.$$

Así

$$\begin{aligned} E(w_t) &= -(\alpha'\alpha)^{-1}(\gamma'\gamma)^{-1}\gamma'\mu + (\alpha'\alpha)^{-1}(\gamma'\gamma)^{-1}\gamma'\Psi(1)\alpha_{\perp}(\gamma'_{\perp}\Psi(1)\alpha_{\perp})^{-1}\gamma'_{\perp}\mu, \\ E(v_t) &= (\gamma'_{\perp}\Psi(1)\alpha_{\perp})^{-1}\gamma'_{\perp}\mu. \end{aligned}$$

Se tiene que $E(\nabla X_t) = \alpha_{\perp}E(v_t) + \alpha E(\nabla w_t)$. Notando que $E(\nabla w_t) = 0$ ya que

$$\begin{aligned} E[\gamma'_{\perp}\Psi(1)\alpha\nabla w_t + \gamma'_{\perp}\Psi(1)\alpha_{\perp}v_t] &= E[\gamma'_{\perp}(\mu + \epsilon_t)] \\ \gamma'_{\perp}\Psi(1)\alpha E[\nabla w_t] + \gamma'_{\perp}\Psi(1)\alpha_{\perp}E[v_t] &= \gamma'_{\perp}E[\mu] + \gamma'_{\perp}E[\epsilon_t] \\ \gamma'_{\perp}\Psi(1)\alpha E[\nabla w_t] + \gamma'_{\perp}\Psi(1)\alpha_{\perp}(\gamma'_{\perp}\Psi(1)\alpha_{\perp})^{-1}\gamma'_{\perp}\mu &= \gamma'_{\perp}\mu \\ \gamma'_{\perp}\Psi(1)\alpha E[\nabla w_t] + \gamma'_{\perp}\mu &= \gamma'_{\perp}\mu \\ \gamma'_{\perp}\Psi(1)\alpha E[\nabla w_t] &= 0 \\ E[\nabla w_t] &= 0. \end{aligned}$$

Entonces se tiene que

$$\begin{aligned} E(\nabla X_t) &= \alpha_{\perp} E(v_t) \\ E(\nabla X_t) &= \alpha_{\perp} (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \gamma'_{\perp} \mu. \end{aligned}$$

Esto prueba (R3). Ahora

$$\begin{aligned} E(\alpha' X_t) &= E[(\alpha' \alpha) w_t] \\ &= (\alpha' \alpha) E[w_t] \\ &= (\alpha' \alpha) [-(\alpha' \alpha)^{-1} (\gamma' \gamma)^{-1} \gamma' \mu + (\alpha' \alpha)^{-1} (\gamma' \gamma)^{-1} \gamma' \Psi(1) \alpha_{\perp} (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \gamma'_{\perp} \mu] \\ &= -(\gamma' \gamma)^{-1} \gamma' \mu + (\gamma' \gamma)^{-1} \gamma' \Psi(1) \alpha_{\perp} (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \gamma'_{\perp} \mu. \end{aligned}$$

Esto completa la prueba de (R4). Recordando

$$(w'_t, v'_t)' = (\tilde{A}(B))^{-1} [(\gamma, \gamma_{\perp})'(\mu + \epsilon_t)].$$

Pero

$$\begin{aligned} \nabla X_t &= \alpha \nabla w_t + \alpha_{\perp} v_t = [\alpha(1 - B), \alpha_{\perp}] (w'_t, v'_t)' \\ &= [\alpha(1 - B), \alpha_{\perp}] (\tilde{A}(B))^{-1} [(\gamma, \gamma_{\perp})'(\mu + \epsilon_t)] \\ &= C(B)(\mu + \epsilon_t), \end{aligned}$$

donde $C(B) = [\alpha(1 - B), \alpha_{\perp}] (\tilde{A}(B))^{-1} [(\gamma, \gamma_{\perp})']$. Esto completa la prueba de (R5).

Para probar (R6), hay que notar que

$$C(1) = (0_{n \times (n-r)}, \alpha_{\perp}) (\tilde{A}(1))^{-1} [(\gamma, \gamma_{\perp})'].$$

Sustituyendo $(\tilde{A}(1))^{-1}$

$$\begin{aligned} C(1) &= (0, \alpha_{\perp}) \begin{bmatrix} -(\alpha' \alpha)^{-1} (\gamma' \gamma)^{-1} & (\alpha' \alpha)^{-1} (\gamma' \gamma)^{-1} \gamma' \Psi(1) \alpha_{\perp} (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \\ 0 & (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \end{bmatrix} \begin{bmatrix} \gamma' \\ \gamma'_{\perp} \end{bmatrix} \\ C(1) &= \alpha_{\perp} (\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1} \gamma'_{\perp} \end{aligned}$$

como se requiere. Las matrices α_{\perp} , γ'_{\perp} y $(\gamma'_{\perp} \Psi(1) \alpha_{\perp})^{-1}$ tienen rango $n - r$ usando las suposiciones A2, A3, y la definición de los componentes ortogonales α_{\perp} y γ'_{\perp} . Así, $C(1)$ tiene rango $(n - r)$. Esto completa la prueba de (R6). Nótese que $E(\nabla X_t) = C(1)\mu = \tau$.

(R7) se sigue inmediato de (R6).

Finalmente, para probar (R8), primero se escribe $C(B) = C(1) + (1 - B)C_1(B)$. Así, por (R5),

$$\begin{aligned} \nabla X_t &= C(1)\mu + C(1)\epsilon_t + (1 - B)C_1(B)\mu + (1 - B)C_1(B)\epsilon_t \\ &= C(1)\mu + C(1)\epsilon_t + C_1(B)(\mu - \mu) + (1 - B)C_1(B)\epsilon_t \\ &= C(1)\mu + C(1)\epsilon_t + (1 - B)C_1(B)\epsilon_t. \end{aligned}$$

Integrando esta expresión

$$\begin{aligned} \nabla X_t &= C(1)\mu + C(1)\epsilon_t + (1 - B)C_1(B)\epsilon_t \\ \sum_{i=1}^t (1 - B)X_i &= C(1) \sum_{i=1}^t \mu + C(1) \sum_{i=1}^t \epsilon_i + C_1(B) \sum_{i=1}^t (1 - B)\epsilon_i \\ \sum_{i=1}^t (X_i - X_{i-1}) &= C(1)t\mu + C(1) \sum_{i=1}^t \epsilon_i + C_1(B) \sum_{i=1}^t (\epsilon_i - \epsilon_{i-1}) \end{aligned}$$

$$\begin{aligned}
(X_1 - X_0) + \cdots + (X_t - X_{t-1}) &= C(1)\mu t + C(1) \sum_{i=1}^t \epsilon_i + C_1(B) [(\epsilon_1 - \epsilon_0) + \cdots + (\epsilon_t - \epsilon_{t-1})] \\
X_t - X_0 &= C(1)\mu t + C(1) \sum_{i=1}^t \epsilon_i + C_1(B) (\epsilon_t - \epsilon_0)
\end{aligned}$$

Si $\epsilon_0 = 0$ y $S_t = C_1(B)\epsilon_t$.

$$X_t = X_0 + C(1) \sum_{i=1}^t \epsilon_i + C(1)\mu t + S_t \quad \square$$

como se requiere. Esto completa la demostración del teorema.

Interpretación de los Resultados del Teorema de Representación de Granger

Primero, se puede ver por (R2) y (R8) respectivamente que, mientras X_t es no estacionaria (porque contiene el error $\sum_{i=1}^t \epsilon_i$), $\alpha' X_t$ es estacionario. De hecho, $\alpha' X_t$ proviene de una serie de combinaciones de las $x - it$ cointegradas.

Segundo, a pesar de la presencia del término de la intersección en el proceso generador de X_t , no hay tendencia lineal en las combinaciones de cointegración. Por (R6) y (R8), la tendencia en el proceso X_t desaparece si $\gamma'_\perp \mu = 0_{(n-r) \times 1}$.

Tercero, (R6) es la condición necesaria para que un proceso sea integrado de primer orden. Si esta matriz no es de rango completo, $|\tilde{A}(B)| = 0$ se tendrá una raíz igual con uno y se puede extraer una raíz unitaria adicional del sistema, lo que conduce a un sistema de variables $I(2)$.

Cuarto, $C(1)$ es una matriz cuadrada n -dimensional que tiene rango $n - r$. Por lo tanto, partiendo del supuesto de una matriz π de rango reducido en la representación autorregresiva se obtiene una matriz $C(1)$ con rango reducido en la representación de medias móviles. Como se mostrará adelante, esto es posible al ir en la otra dirección y obtener el resultado que una matriz $C(1)$ con rango $(n - r)$ implica, para un sistema cointegrado, una matriz de la forma π con rango r . en efecto, hay una interesante dualidad entre la singularidad del “impacto” de la matriz π para la representación autorregresiva y la singularidad del impacto de la matriz $C(1)$ para la representación de medias móviles.

Quinto, si $\gamma'_\perp \mu = 0_{(n-r) \times 1}$, entonces μ está en el espacio ortogonal de γ_\perp y por tanto en el espacio de γ . Así, μ puede escribirse como $\gamma\beta_0$ donde β_0 es un vector arbitrario de $r \times 1$. De la expresión para $E(\alpha' X_t)$ en (R4), se nota que $E(\alpha' X_t) = -\beta_0$, y la constante entra en el sistema únicamente vía término de corrección de error. Esto puede verse más claramente al escribir como

$$\begin{aligned}
\nabla X_t &= \mu + \sum_{i=1}^{k-1} \Gamma_i (1 - B) B^i X_t + \pi X_{t-k} + \epsilon_t \\
&= \gamma\beta_0 + \sum_{i=1}^{k-1} \Gamma_i (1 - B) B^i X_t + \gamma\alpha' X_{t-k} + \epsilon_t \\
&= \sum_{i=1}^{k-1} \Gamma_i (1 - B) B^i X_t + \gamma(\beta_0 + \alpha' X_{t-k}) + \epsilon_t.
\end{aligned}$$

Si la restricción no se satisface, el término de intersección entra en el sistema al mismo tiempo como término de corrección de error y como un componente de crecimiento autónomo. En seguida se presentará el procedimiento de Johansen de máxima verosimilitud para estimar las relaciones de cointegración, el tratamiento de la constante es importante en la determinación del procedimiento de estimación y los valores críticos que se usan para inferencia.

Ahora se retoma la representación de un sistema cointegrado en su forma autorregresiva o (equivalentemente) en forma de mecanismo de corrección de error. Donde $\{\nabla X_t\}$ es un proceso estacionario (posiblemente) con término de intersección, se puede expresar como un modelo multivariado de medias móviles usando el teorema de descomposición de Wold (1954)

$$\nabla X_t = C(B)(m + e_t),$$

donde $e_t \sim IID(0, \Omega)$; B es nuevamente el operador de rezago, y $C(B)$ es una matriz polinomial de B dada por

$$C(B) = \sum_{i=0}^{\infty} C_i B^i, \quad C_0 = I_n.$$

El efecto acumulativo o total de $C(B)$ está dado por

$$C(1) = \left(I_n + \sum_{i=1}^{\infty} C_i \right) < \infty,$$

donde las C_i de nuevo obedecen un decaimiento exponencial. Usando $C(1)$, se puede reescribir $C(B)$ como

$$\begin{aligned} \sum_{i=0}^{\infty} C_i B^i &= C_0 + C_1 B + C_2 B^2 + \dots \\ &= C_0 + (C_1 - C_1) + \dots + C_1 B + (C_2 - C_2) B + \dots \\ &= (C_0 + C_1 + \dots) + (-C_1 - C_2 - \dots) B^0 + (-C_2 - C_3 - \dots) B^1 \\ &\quad - (-C_1 - C_2 - \dots) B^1 + (-C_3 - C_4 - \dots) B^2 - (-C_2 - C_3 - \dots) B^2 + \dots \\ &= I_n + \sum_{i=1}^{\infty} C_i + \sum_{i=0}^{\infty} \left(- \sum_{j=i+1}^{\infty} C_j \right) B^i - \sum_{i=0}^{\infty} \left(- \sum_{j=i+1}^{\infty} C_j \right) B^{i+1} \\ &= I_n + \sum_{i=1}^{\infty} C_i + \sum_{i=0}^{\infty} C_i^* B^i - B \sum_{i=0}^{\infty} C_i^* B^i \\ &= I_n + \sum_{i=1}^{\infty} C_i + C^*(B) - B C^*(B) \\ &= C(1) + (1 - B) C^*(B) \end{aligned}$$

donde $C^*(B) = \sum_{i=0}^{\infty} C_i^* B^i$ y $C_i^* = - \sum_{j=i+1}^{\infty} C_j$ así que $C_0^* = I_n - C(1)$. Nótese que la existencia de estas matrices está garantizada por la condición de decaimiento exponencial. Así

$$\begin{aligned} \nabla X_t &= C(B)(m + e_t) \\ \nabla X_t &= C(1)m + C(1)e_t + C^*(B)\nabla e_t \\ (\nabla X_t - \mu) &= C(1)e_t + C^*(B)\nabla e_t \end{aligned}$$

donde $\mu = C(1)m$.

Las suposiciones necesarias para derivar la representación autorregresiva del proceso están dadas abajo.

Suposición B1. El polinomio característico,

$$C(z) = \sum_{i=1}^{\infty} C_i z^i,$$

tiene raíces iguales o estrictamente mayores a uno; esto es, $|C(z)| = 0$ lo cual implica que $|z| > 1$ o $z = 1$.

Suposición B2. La matriz $C(1)$ tiene rango reducido $n - r$ y es por lo tanto expresable como el producto de dos matrices de $n \times (n - r)$ ϕ y η , en donde ϕ y η tiene rango $n - r$. Así, $C(1) = \phi\eta'$.

Suposición B3. La matriz $\phi'_{\perp} C^*(1)\eta_{\perp}$ de $r \times r$ tiene rango completo r .

Los supuestos B1 – B3 son análogos a los supuestos A1 – A3 de la primera parte. Dados los resultados sobre $C(1)$ es natural que se requiera que $C(1)$ sea de rango reducido y tenga rango $n - r$. Además, $n = r$ implica que $C(1)$ es idénticamente la matriz nula. Así, $(\nabla X_t - \mu) = C^*(B)\nabla e_t$, lo cual implica, después de la integración, que X_t es integrada (a lo más) de orden 0. Por el supuesto B3 y eliminando la posibilidad de que $C^*(B)$ tenga raíces sobre el círculo unitario, así X_t no puede ser integrado de orden -1. En ambos casos, se tiene una contradicción del supuesto de que las componentes de X_t son $I(1)$.

Para derivar la representación autorregresiva, se multiplica $\nabla X_t - \mu = C(1)e_t + C^*(B)\nabla e_t$ por ϕ' y ϕ'_{\perp} respectivamente para obtener las ecuaciones

$$\begin{aligned}\phi' \phi \eta' e_t + \phi' C^*(B) \nabla e_t &= \phi' (\nabla X_t - \mu) \\ \phi'_{\perp} C^*(B) \nabla e_t &= \phi'_{\perp} (\nabla X_t - \mu),\end{aligned}$$

usando la descomposición $C(1) = \phi\eta'$ y el resultado $\phi'_{\perp} \phi = 0_{r \times (n-r)}$. La matriz $C(1)$ no es invertible, por lo tanto el sistema dado arriba no puede invertirse directamente para expresar las x_{it} en términos de las e_{it} . Para obtener un sistema invertible, se definen dos nuevas variables, $W_t = (\eta'\eta)^{-1}\eta'e_t$ e $y_t = (\eta'_{\perp}\eta_{\perp})^{-1}\eta'_{\perp}\nabla e_t$. Luego, se definen las matrices $\bar{\eta} = \eta(\eta'\eta)^{-1}$ y $\bar{\eta}_{\perp} = \eta_{\perp}(\eta'_{\perp}\eta_{\perp})^{-1}$. Y como en la parte anterior

$$\nabla e_t = (\eta\bar{\eta}' + \eta_{\perp}\bar{\eta}'_{\perp})\nabla e_t = \eta_{\perp}y_t + \eta\nabla W_t.$$

Sustituyendo esta expresión en el sistema de ecuaciones

$$\begin{aligned}(\phi'\phi)(\eta'\eta)W_t + \phi' C^*(B)\eta\nabla W_t + \phi' C^*(B)\eta_{\perp}y_t &= \phi' (\nabla X_t - \mu) \\ \phi'_{\perp} C^*(B)\eta\nabla W_t + \phi'_{\perp} C^*(B)\eta_{\perp}y_t &= \phi'_{\perp} (\nabla X_t - \mu).\end{aligned}$$

Y pueden escribirse ahora en forma autorregresiva como

$$\tilde{D}(B) (W'_t, y'_t)' = (\phi, \phi_{\perp})' (\nabla X_t - \mu),$$

con

$$\tilde{D}(z) = \begin{bmatrix} (\phi'\phi)(\eta'\eta) + \phi' C^*(z)\eta(1-z) & \phi' C^*(z)\eta_{\perp} \\ \phi'_{\perp} C^*(z)\eta(1-z) & \phi'_{\perp} C^*(z)\eta_{\perp} \end{bmatrix}.$$

Para $|z| > 1$, el determinante de esta matriz es:

$$|\tilde{D}(1)| = |\phi'\phi| |\eta'\eta| |\phi'_{\perp} C^*(1)\eta_{\perp}|,$$

que es diferente de cero por los supuestos B2 y B3. Así $\tilde{D}(z)$ no tiene una raíz en uno. Para $z \neq 1$

$$\tilde{D}(z) = (\phi, \phi_{\perp})' C(z) [\eta, \eta_{\perp}(1-z)^{-1}],$$

Para mostrar esto, se sustituye $C^*(z)$ en $\tilde{D}(z)$ en términos de $C(z)$ y $C(1) = \phi\eta'$, y la condición de ortogonalidad $\phi'_{\perp} \phi = \eta'_{\perp} \eta = 0_{r \times (n-r)}$. Para $z > 1$,

$$|\tilde{D}(z)| = |(\phi, \phi_{\perp})| |C(z)| |(\eta, \eta_{\perp})| (1-z)^{-r}.$$

Así para $z > 1$, $|\tilde{D}(z)| = 0$ si y sólo si $|C(z)| = 0$. Excluye $z = 1$, por el supuesto B1 solamente restan las raíces del determinante que están fuera del círculo unitario.

Todas las raíces de $|\tilde{D}(z)| = 0$ están por lo tanto fuera del disco unitario y el sistema

$$\begin{aligned}(\phi'\phi)(\eta'\eta)W_t + \phi' C^*(B)\eta\nabla W_t + \phi' C^*(B)\eta_{\perp}y_t &= \phi' (\nabla X_t - \mu) \\ \phi'_{\perp} C^*(B)\eta\nabla W_t + \phi'_{\perp} C^*(B)\eta_{\perp}y_t &= \phi'_{\perp} (\nabla X_t - \mu),\end{aligned}$$

es invertible. Así

$$(W'_t, y'_t)' = \tilde{D}(B)^{-1}(\phi, \phi_\perp)'(\nabla X_t - \mu).$$

Además hay que notar que

$$\tilde{D}(1) = \begin{bmatrix} (\phi'\phi)(\eta'\eta) & \phi' C^*(1)\eta_\perp \\ 0 & \phi'_\perp C^*(1)\eta_\perp \end{bmatrix}.$$

Y su inversa es

$$\tilde{D}(1)^{-1} = \begin{bmatrix} (\eta'\eta)^{-1}(\phi'\phi)^{-1} & -(\eta'\eta)^{-1}(\phi'\phi)^{-1}\phi' C^*(1)\eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1} \\ 0 & (\phi'_\perp C^*(1)\eta_\perp)^{-1} \end{bmatrix}.$$

Por la definición de ∇e_t ,

$$\begin{aligned} \nabla e_t = \eta \nabla W_t + \eta_\perp y_t &= [\eta(1-B), \eta_\perp](W'_t, y'_t)' \\ &= [\eta(1-B), \eta_\perp] \left(\tilde{D}(B) \right)^{-1} [(\phi, \phi_\perp)'(\nabla X_t - \mu)] \\ &= F(B)(\nabla X_t - \mu), \end{aligned}$$

donde $F(B) = [\eta(1-B), \eta_\perp] \left(\tilde{D}(B) \right)^{-1} [(\phi, \phi_\perp)']$.

Integrando queda

$$F(B)X_t = F(1)t\mu + e_t + X_0,$$

donde X_0 es una constante de integración. Para obtener el valor de $F(1)$, hay que notar que

$$F(1) = (0_{n \times (n-r)}, \eta_\perp) \left(\tilde{D}(1) \right)^{-1} [(\phi, \phi_\perp)'].$$

Al sustituir $(\tilde{D}(1))^{-1}$ queda $F(1) = \eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1}\phi'_\perp$. Así, se renombra a $\mu = C(1)m = (\phi\eta')m$, $F(1)\mu = 0_{n \times 1}$. La representación autorregresiva, en su forma final, está dada por lo tanto por

$$F(B)X_t = X_0 + e_t.$$

Varias de las derivaciones que aparecen arriba son de notarse, particularmente con respecto a la matriz $F(1)$. Primero, $F(1)C(1) = C(1)F(1) = 0_n$. Este resultado sigue de sustituir $F(1)$ por $\eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1}\phi'_\perp$ y $C(1)$ por $\phi\eta'$ y usando la condición de ortogonalidad. Esto enfatiza la dualidad entre el impacto de la matriz en la representación MA, dada por $C(1)$, y el impacto de la matriz en la representación AR, dada por $F(1)$. El espacio nulo del primero es el rango del espacio del último y viceversa.

Segundo el isomorfismo de $F(1)$ con la matriz $\gamma\alpha'$ de la sección anterior puede demostrarse. Notando que

$$\text{i } C(1)\eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1} = 0_{n \times r};$$

$$\text{ii } \phi'_\perp C(1) = 0_{r \times n}.$$

Ambas $\eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1}$ y ϕ_\perp son matrices de rango r y dimensión $n \times r$, así definiendo ϕ'_\perp como α' y $\eta_\perp(\phi'_\perp C^*(1)\eta_\perp)^{-1}$ como γ , se tiene $F(1) = \gamma\alpha'$, que es una matriz de $n \times n$ con rango r y es isomórfica a π . Esto es natural al definir $\phi'X_t$ como las combinaciones de cointegración de las x_{it} . Integrando

$\phi'_\perp C^*(B)\nabla e_t = \phi'_\perp(\nabla X_t - \mu)$ se muestra que $\phi'_\perp X_t$ no contiene una componente de la forma $\sum_{i=1}^t e_i$. Además

por la ortogonalidad de μ con ϕ_\perp , las combinaciones de cointegración no contienen tendencia. Ambos resultados coinciden exactamente con los correspondientes resultados para $\alpha'X_t$ de la sección anterior.

Tercero, si $\tilde{D}(B)$ no es de rango completo, sería posible extraer otra raíz unitaria en la representación

$$\tilde{D}(z) = \begin{bmatrix} (\phi'\phi)(\eta'\eta) + \phi' C^*(z)\eta(1-z) & \phi' C^*(z)\eta_\perp \\ \phi'_\perp C^*(z)\eta(1-z) & \phi'_\perp C^*(z)\eta_\perp \end{bmatrix}$$

y el sistema sería $I(0)$ en lugar de $I(1)$, como se supone originalmente. El supuesto B3 es claro ahora. Finalmente usando el resultado de que el rango de $F(1)$ es r , es posible escribir el modelo en la forma de mecanismo de corrección de errores como

$$\nabla X_t = H(B)\nabla X_{t-1} + X_0 + F(1)X_{t-k} + e_t,$$

donde $F(1)$, como π en la sección anterior es una matriz de rango r y por tanto puede descomponerse en dos matrices de $n \times r$, cada una de rango r . Los pasos involucrados para ir de la forma final del modelo autorregresivo a la forma de un sistema de MCE están dados en la página 105, con π jugando el papel de $F(1)$.

Con esto se demuestra el isomorfismo entre los modelos de medias móviles, los modelos de corrección de error y la representación autorregresiva de los procesos de cointegración.

Bibliografía

- [1] Álvarez, Vázquez Nelson (1998). “Econometría (Addenda)”. Segunda Edición. Universidad Nacional de Educación a Distancia, Madrid, España.

- [2] Amemiya, Takeshi (1996). “Advanced Econometrics”. Harvard University Press, E.U.A.

- [3] Banerjee, Dolado, Galbraith y Hendry (1993). “Co-Integration, Error-Correction, and the Econometric Analysis of Non-Stationary Data”. Oxford University Press, New York.

- [4] Barlett, M, S.: “On the Theoretical Specification of Sampling Properties of Autocorrelated Time Series.” Journal of the Royal Statistical Society. Series B, 8, 1946.

- [5] Blatt, Mogens (1995). “Applied Time Series Analysis”. Documento de trabajo Núm. VI, Colegio de México, México.

- [6] Davidson, J.H., Hendry, D.H., Srba, F. y Yeo, S.: “Econometric Modelling of the Aggregate Time-Series ‘Relationship between Consumers’ Expenditure and Income in the United Kingdom”, The Economic Journal, 88, 1978.

- [7] Dhrymes, Phoebus (1998). “Time Series, Unit Roots, and Cointegration”. Academic Press, San Diego California.

- [8] Dickey, D.A.: Hypothesis Testing for Non-Stationary Time Series. Sin publicar. Universidad del Estado de Iowa, 1975.

- [9] Enders, Walter (1996). “Rats Handbook for Econometric Time Series”. John Wiley & Sons Inc., New York.

- [10] Engle,R.F y Granger,C.W.J.:“Cointegration and Error Correction: Representation. Estimation and Testing”, *Econometrica*, 55, 1987.
- [11] Fuller, W.A.: *Introduction to Statistical Time Series*. New York, Wiley, 1976.
- [12] Geweke, J., Meese, R. y Dent, W.: “Comparing Alternative Tests of Causality in Temporal Systems” , *Journal of Econometrics*, 23, 1983.
- [13] Granger, C.W.J.: “Investigating Causal Relatios by Econometric Models and Cross-Spectral Methods” *Econometrita*, 37, 1969.
- [14] Granger,C.W.J. y Newbold,P.:“Spurious Regressions in Econometrics”, *Journal of Econometrics*, 3,1974.
- [15] Greene, William H. (1997). “Econometric Analysis”. Tercera Edición. Prentice Hall, New Jersey. E.U.A.
- [16] Gujarati, Damodar N. (1992). “Econometría”. Segunda Edición. Mc Graw Hill, México.
- [17] Guerrero, Victor M. (1991). “Análisis Estadístico de Series de Tiempo Económicas”. Universidad Autónoma Metropolitana, México.
- [18] Hatanaka, Michio (1996). “Time Series Based Econometrics Unit Roots and Cointegration”. Oxford University Press, New York.
- [19] Mackinnon, J.G.: “Critical values for Coingrating Test”, Mimeo Queens University, Ontario, 1990.
- [20] Maddala G. S. (1992). “Introduction to Econometrics”. Segunda Edición. Prentice Hall, E.U.A.
- [21] Novales, Cinia Alfonso (1993). “Econometría”. Segunda Edición. McGraw-Hill. Madrid, España.

- [22] Otero, José Ma. (1993). "Econometría Series Temporales y Predicción". Editorial AC. Madrid, España.
- [23] Phillips, P.C.B: "Time Series Regression whit a Unit Root", *Econometrita*, 55, 1987.
- [24] Phillips,A.W.: "Stabilization Policy and the Time Forms of Lagged Responses", *Economic Journal*, 67, 1957.
- [25] Phillips,P.C.B.: "Understanding Suprious Regressions in Econometrics", *Journal of Econometrics*, 33, 1986.
- [26] Pierce, D.A., y Haugh, L.D.: "Causality in temporal systems. Characterization and a survey", *Journal of Econometrics*, vol. 5, 1977.
- [27] Pulido, San Roman Antonio (1993). "Modelos Econométricos". Cuarta Edición. Ediciones Piramide. Madrid, España.
- [28] Quenouille, M.H.: "Aproximate Test of Correlation in Time-Series". *Journal of de the Royal Statistical Society, Series B*, 11,1. 1949.
- [29] Sargan,J.D.: "Wages and Prices in de United Kingdom: A Study in Econometric Methodology". En *Econometric Analysis for National Planning*; P.E.Hart,G.Mills y J.N.Whuttaker, Londres, Butterworths, 1964.
- [30] Sims, C.A.: "Macroeconomics and Reality" *Econometrica*, 48, 1980.
- [31] Sims. C. A.: "Money, Income and Causality", *American Economic Review*, 62. 1972.